UNIVERSIDADE CATÓLICA PORTUGUESA

Using Predictive Machine Learning Models to Analyse Weather Resilience and Robustness in U.S Airports

Eduardo Fonseca

Católica Porto Business School

2025

UNIVERSIDADE CATÓLICA PORTUGUESA

Using Predictive Machine Learning Models to Analyse Weather Resilience and Robustness in U.S Airports

Master's Final Assignment – Dissertation

Presented to *Universidade Católica Portuguesa*
to obtain a Master's Degree in Management – Business Analytics

by

Eduardo Fonseca

Under supervision of

Dr Jiabin Luo

Católica Porto Business School, Universidade Católica Portuguesa

September of 2025

Declaration

I declare that I have personally prepared this report and that it has not in whole or in part been submitted for any other degree or qualification. Nor has it appeared in whole or in part in any textbook, journal or any other document previously published or produced for any purpose. The work described here is my/our own, carried out personally unless otherwise stated. All sources of information, including quotations, are acknowledged by means of reference, both in the final reference section and at the point where they occur in the text.

Acknowledgements

I want to give some special thanks and personal acknowledgement to those who without, this journey would not have been possible, or not nearly as worth it.

For my parents for their constant support (and constant worrying). For my grandparents, Tita and Zezinho, for always having a hot meal and some wise words ready when I need them most. For my girlfriend, Maria, for always being patient and a source of motivation. For my friends, who helped keep my sanity, and here I highlight the “Pinocos” and the “Big 4”. For all my study buddies Wei, Arianne and António, that helped me brainstorm and keep my work on track. And lastly, to all family, friends and colleagues whose names go unmentioned but not forgotten.

Abstract

Flight delays pose a major challenge to the aviation sector, disrupting operations, reducing efficiency, and undermining passenger experience. This thesis examines airport resilience through the lens of robustness, defined as the ability to sustain operational performance under adverse weather conditions. A dataset combining meteorological information from Meteostat/NOAA and airline performance metrics from Sage Data was constructed for six major U.S. airports (ATL, DEN, DFW, JFK, LAX, ORD) covering 2010–2025.

Machine learning models, including linear regression, random forest, and gradient boosting, were applied to predict on-time flight performance and quantify the impact of meteorological variables such as precipitation, wind speed, and temperature extremes. Tree-based models outperformed linear models, with random forest achieving the highest predictive accuracy. Crucially, results reveal substantial variation in robustness across airports: Denver (DEN) and Atlanta (ATL) exhibited the highest resilience, maintaining more stable on-time performance despite weather disruptions, whereas New York (JFK) and Chicago (ORD) were most sensitive, with delays strongly linked to adverse conditions. Dallas/Fort Worth (DFW) and Los Angeles (LAX) showed intermediate robustness.

By integrating predictive modelling with the conceptual framework of resilience, this thesis advances both academic understanding and practical assessment of airport performance. The comparative insights provide an evidence base for targeted strategies to mitigate vulnerabilities and strengthen resilience within the U.S. aviation system.

Keywords: Airport Resiliency; Robustness; Weather Disruptions; Predictive Modelling; Machine Learning

Resumo

Os atrasos nos voos representam um grande desafio para o setor da aviação, perturbando as operações, reduzindo a eficiência e prejudicando a experiência dos passageiros. Esta tese examina a resiliência dos aeroportos através da lente da robustez, definida como a capacidade de manter o desempenho operacional em condições meteorológicas adversas. Foi construído um conjunto de dados combinando informações meteorológicas da Meteostat/NOAA e métricas de desempenho das companhias aéreas da Sage Data para seis grandes aeroportos dos EUA (ATL, DEN, DFW, JFK, LAX, ORD), cobrindo o período de 2010 a 2025.

Modelos de aprendizagem automática, incluindo regressão linear, Random Forest e Gradient Boosting, foram aplicados para prever o desempenho dos voos em termos de pontualidade e quantificar o impacto de variáveis meteorológicas, tais como precipitação, velocidade do vento e temperaturas extremas. Os modelos “tree based” superaram os modelos lineares, com a Random Forest a atingir a maior precisão preditiva. Fundamentalmente, os resultados revelam uma variação substancial na robustez entre os aeroportos: Denver (DEN) e Atlanta (ATL) exibiram a maior resiliência, mantendo um desempenho mais estável em termos de pontualidade, apesar das perturbações meteorológicas, enquanto Nova Iorque (JFK) e Chicago (ORD) foram os mais sensíveis, com atrasos fortemente ligados a condições adversas. Dallas/Fort Worth (DFW) e Los Angeles (LAX) mostraram uma robustez intermédia.

Ao integrar a modelagem preditiva com a estrutura conceitual da resiliência, esta tese promove tanto o entendimento acadêmico quanto a avaliação prática do desempenho dos aeroportos. As perceções comparativas fornecem uma base de evidências para estratégias direcionadas para mitigar vulnerabilidades e fortalecer a resiliência dentro do sistema de aviação dos EUA.

Palavras-Chave: Resiliência de Aeroportos; Robustez; Perturbações Climáticas; Modelagem Preditiva; Machine Learning

Content

Table of Figures

List of Tables

1. Introduction

Air transport underpins global connectivity but remains highly vulnerable to disruptions, particularly those caused by adverse weather. Flight delays are one of the most visible and costly outcomes, reducing efficiency, affecting passenger experience, and creating ripple effects across the wider network.

Resilience has therefore become a key concern in aviation, with robustness—the ability to sustain operations under adverse conditions—central to assessing airport performance. While delays have been studied extensively, there is limited comparative research on how major airports in different climates withstand weather-related disruptions.

This thesis addresses that gap by analysing six of the busiest U.S. hubs—Atlanta (ATL), Denver (DEN), Dallas/Fort Worth (DFW), John F. Kennedy (JFK), Los Angeles (LAX), and Chicago O'Hare (ORD)—over the period 2010–2025. Using a data-driven approach, it evaluates airport robustness by linking meteorological variables with operational outcomes through predictive modelling.

The study pursues four objectives:

1. To build harmonised datasets combining weather and performance variables.
2. To explore patterns of delay across airports and climate contexts.
3. To apply and compare machine learning models to quantify weather impacts.
4. To assess and rank the relative robustness of airports to weather disruptions.

The scope is limited to weather-related factors, excluding other causes of delay. Methodologically, the thesis applies a structured pipeline of data preparation, exploratory analysis, feature selection, and model evaluation.

2. Literature review

What is Resilience?

Resilience is a multifaceted concept that has evolved significantly, moving from its ecological origins into becoming a key concept in infrastructure and transportation systems. It fundamentally refers to “the ability of a system, community, or society exposed to hazards to resist, absorb, accommodate to, and recover from the effects of a hazard in a timely and efficient manner, including through the preservation and restoration of its essential basic structures and functions” (UNISDR, 2009). In the context of transportation, this translates to the ability of a transportation system to absorb disturbances, maintain its basic structure and function, and recover to a required level of service within an acceptable time and costs after being affected by disruptions (Wan et al., 2018).

Evolution of the Term Resilience

The term "resilience" originates from the Latin word "resiliere," meaning "to spring back" or "to rebound" (Bešinović, 2020). Its conceptual foundation in systems theory was laid by Holling (1973), who introduced it in theoretical ecology to differentiate it from stability. Holling's (1973) concept emphasized a system's capacity to persist and absorb changes or shocks without shifting to a fundamentally different state of behavior

Over time, the concept of resilience broadened, finding applications in diverse fields such as psychology, economics, and engineering (Hosseini et al., 2016). Its emergence in transportation studies gained prominence in the late 1990s (Ahmed & Dey, 2020), notably after the Kobe earthquake, which highlighted the vulnerability of road networks. Murray-Tuite (2006) was among the first to specifically define resilience within the transportation context, outlining ten dimensions, including redundancy, diversity, efficiency, autonomous components, strength, collaboration, adaptability, mobility, safety, and the ability to recover quickly.

Research in transportation resilience has grown steadily, reflecting its increasing popularity and recognition as a vital aspect of system performance (Pan et al., 2021). This shift reflects an expansion of safety research from traditional risk management and security concerns towards a more holistic focus on resilience and sustainability

(Wan et al., 2018). While a single, universally accepted definition of transportation resilience remains elusive, various definitions share commonalities, typically highlighting the system's ability to cope with abnormal conditions and recover effectively (Pan et al., 2021).

Application of Resilience

Resilience in transportation systems encompasses both proactive planning (designing systems to be inherently resilient) and reactive measures (responding to and recovering from disruptions) (Bešinović, 2020). This involves a comprehensive understanding of how systems perform across different phases: pre-disruption, during disruption, and post-disruption (Sun et al., 2020).

Key applications of resilience in transportation include:

• **Disruption Management**: Resilience aims to improve system performance during disturbances and minimize losses. This includes ensuring effective services even in normal conditions while preparing for disruptions (Bešinović, 2020).

• **Planning and Design**: Beyond mere concrete defenses, building resilient transportation systems involves adopting resilient working practices, planning, and operational buffers (Bešinović, 2020).

• **Critical Component Analysis**: Identifying the most critical components within a network allows for the strategic allocation of limited resources. (Zhou et al., 2019).

• **Addressing Climate Change and Extreme Weather Events (EWE)**: Resilience strategies are vital for mitigating the impacts of increasing threats like sea level rise, floods, and storms on transportation infrastructure (Ahmed & Dey, 2020).

• **Demand-Centered Solutions**: Research is increasingly focusing on demand-centered and community resilience, recognizing that understanding user behavior and diverse responses to disruption strategies is essential (Pan et al., 2021).

• **Interdependency of Critical Systems**: Recognizing that transportation systems are interconnected with other critical infrastructures (e.g., water, telecommunications, power grids) is vital. A multidisciplinary approach is needed to assess and develop early warning systems for disruptions that can propagate across these interdependent systems (Ahmed & Dey, 2020).

• **Cost-Benefit Justification**: Quantifying resilience allows for cost-benefit analyses, which can justify investments in infrastructure improvements and suitable control measures (Wan et al., 2018).

• **Intelligent Transportation Systems (ITS)**: Machine Learning (ML) approaches are becoming fundamental in ITS for tasks such as perception, prediction, and management. This includes applications like optimizing traffic light management, allocating resources in vehicular networks, and detecting network safety hazards (Yuan et al., 2022).

Measurement of Resilience

Measuring resilience typically involves two main steps: defining appropriate metrics and then applying evaluation approaches to calculate these metrics (Sun et al., 2020).

Resilience metrics can generally be categorized into three types:

• **Topological Metrics**:

These metrics focus on the structural properties of the transportation network, often disregarding the dynamic aspects of system performance. They are derived from graph theory and include measures such as average shortest paths, betweenness centrality and efficiency. (Zhou et al., 2019).

Topological approaches require limited data, are mathematically solid, and are suitable for comparing different network structures relatively quickly. But often fail to realistically replicate system behavior during disruptions (Bešinović, 2020).

• **Performance-Based Metrics**:

These metrics are designed to assess a system's resilience by evaluating its performance over the entire period affected by a disaster, rather than just at specific points in time (Zhou et al., 2019).

Common performance-based metrics include:

- Degradation of system quality over time (Mattsson & Jenelius, 2015);
- Time-dependent ratio of recovery to loss (Zhou et al., 2019);
- Expected fraction of demand satisfied (Zhou et al., 2019);
- Economic resilience (Cox et al., 2011);
- Travel time indicators (Pan et al., 2021);

▪ Lost service days (Sun et al., 2020);

▪ Metrics that combine performance loss with resource expenditures required for recovery (Zhou et al., 2019).

• **Attributes-Based Metrics**:

These focus on specific properties or "attributes" of resilience, such as the R4 framework (Leobons et al., 2019; Mattsson & Jenelius, 2015). The "R4 framework," originally proposed by Bruneau et al. (2003), is widely used to describe the four core properties of resilience:

1. **Robustness:** It is generally defined as the ability to withstand or absorb disturbances and remain intact when exposed to disruptions (Wan et al. 2018). In transportation systems, the level of robustness helps determine the initial performance loss during a disruption (Zhou, Wang & Yang, 2019).

2. **Redundancy:** This indicates the ability of certain components of a system to take over the functions of failed components without adversely affecting the overall performance. In transportation, redundancy is typically viewed as the existence of optional routes between origins and destinations (Wan et al., 2018).

3. **Resourcefulness:** This property refers to the availability of resources such as materials, supplies, and crews, necessary to restore functionality (Adams et al., 2012). In transportation systems, resourcefulness represents the amount of available repair units in post-disaster operations (Zhou, Wang & Yang, 2019).

4. **Rapidity:** This emphasizes the speed with which a system returns to a state of normal function after a disturbance (Adams et al., 2012).

Additional Key Characteristics

In addition to the R4 framework, other characteristics are frequently identified as vital to system resilience:

• **Recoverability (Recovery):** Often discussed the most in resilience research, recoverability is defined as the ability of a network to recover functionality in a timely manner (Baroud et al., 2014).

• **Reliability:** Defined as the probability that a network remains operative given a disruption event, acting as a pre- or post-disruption metric (Barker et al., 2013)

• **Adaptability (or Flexibility):** This reflects the flexible ability to respond to new pressures or to reconfigure resources and cope with uncertainties aftershocks (Wan et al., 2018).

• **Vulnerability:** Although often viewed as the opposite of resilience, vulnerability captures the susceptibility to damage or perturbation (Wan et al., 2018).

• **Preparedness:** Refers to measures prepared before a disruption happens to enhance resilience by reducing potential negative impacts (Jin et al., 2014).

• **Survivability:** Generally defined as the ability to withstand sudden disturbances to functionality while meeting original demands (Faturechi & Miller-Hooks, 2014).

Approaches for Quantifying Resilience

Various methodological approaches are employed to estimate and quantify transportation resilience:

• **Data-Driven Approach**:

This approach directly utilizes historical data (e.g., traffic patterns, passenger ridership, weather data) to assess system performance changes in different scenarios It is particularly useful for ex-post analyses of past disruptions and often incorporates statistical methods (Bešinović, 2020; Zhou et al., 2019).

Data-driven methods can provide quick assessments without requiring explicit modeling of the system's inherent mechanisms, although it requires well-structured and documented data, and increasingly relies on Artificial Intelligence (AI) applications for information extraction and future disruption prediction (Bešinović, 2020).

• **Simulation Approach**:

Simulation models evaluate network performance using both topological and system/performance indicators (e.g., delay, passenger loads) within stochastic environments (Zhou et al., 2019). They can capture dynamic effects of disruptions and modeling network impacts based on various disruption distributions (Pan et al., 2021).

• **Optimization Approach**:

Mathematical optimization models are used to quantify or improve network resilience, often focusing on identifying the most critical elements or optimal post-event recovery actions (Pan et al., 2021).

• **Fuzzy Logic**:

This approach is useful when explicit mathematical models are unavailable. It uses real numbers (0 to 1) to represent the truth of variables, allowing for the inclusion of relative attributes for a given concept (Zadeh, 2008).

• **Bayesian Networks**:

These models are used to quantify resilience by considering absorptive, adaptive, and restorative capacities. They can effectively model and rank the influence of various variables on a system's resilience (Hosseini et al., 2016).

• **Integrated/Hybrid Approaches**:

To overcome the limitations of individual methods, researchers often combine different approaches. This allows for a more comprehensive assessment and enhancement of resilience in complex transportation systems (Bešinović, 2020).

Emerging Developments

Recent work has refined how resilience is measured in transportation systems. Chen (2024) argues that both structural and functional dimensions must be considered, noting that multimodal and intercity networks cannot be understood through topology alone. Similarly, Zhang (2024) introduces a simulation-based framework that captures dynamic recovery processes, addressing the limitations of static or purely structural models.

Scholars have also highlighted contextual and practical aspects. Bergantino (2024) shows that resilience differs across regions, emphasizing the influence of geographic context. In parallel, Cordero (2024) presents a performance measure–based framework intended to guide planning and decision-making, while Nipa et al. (2023) develop a decision-support tool built on 21 validated dimensions—including organizational, structural, and financial factors—to produce composite resilience scores for highway projects.

Machine Learning

Machine learning refers to algorithms that learn from data (experience **E**) to improve performance on a task **T** as measured by a performance metric **P**, without being explicitly programmed for every rule (Mitchell, 1997).

Linear Regression Models

Linear regression models assume a linear relationship between predictor variables and a response variable. They are often extended with regularization to improve robustness and interpretability.

- ***Ordinary Least Squares (OLS)*** estimates coefficients by minimizing the residual sum of squares. While easy to interpret and analytically solvable, OLS suffers from sensitivity to outliers, instability under multicollinearity, and invalid statistical inference if heteroskedasticity is ignored (though the coefficients themselves remain unbiased) (Wooldridge, 2012).
- ***Ridge Regression*** addresses multicollinearity by adding an L2L_2L2 penalty, shrinking coefficients towards zero without eliminating them. This reduces variance at the cost of some bias, yielding more stable predictions (Hoerl and Kennard, 1970).
- ***LASSO*** (Tibshirani, 1996) extends regression by adding an L1L_1L1 penalty, shrinking some coefficients exactly to zero and thus performing variable selection. It is useful for producing sparse, interpretable models, though in cases of highly correlated predictors it tends to select only one variable from a group (Zou and Hastie, 2005).

Non-Linear / Ensemble Models

Ensemble methods combine multiple weaker learners—often decision trees—into stronger predictive models that capture complex, non-linear relationships.

- ***Random Forests*** (Breiman, 2001) aggregate many randomized decision trees, each trained on bootstrap samples and a subset of features. The ensemble reduces overfitting, improves accuracy, and provides variable importance measures (Biau and Scornet, 2015).

- ***Gradient Boosting*** (Friedman, 2001) builds trees sequentially, each correcting the errors of the previous ensemble using gradient descent on a loss function. This iterative focus on hard-to-predict instances yields highly accurate models.

Research Gap

As supported by the literature above, in this work we want to look at resilience in the context of flight delays. We will compare the robustness of six different airports by equating robustness – as the ability to withstand disturbances- to how each airport reacts to weather effects. To this end we will use machine learning, taking a data-driven approach to quantifying this attribute of resilience.

3. Methodology

3.1. Data Collection

According to the Federal Aviation Administration (FAA, 2025) the six airports chosen are the ones with the biggest passenger and flights turnover in the United States, making them better to analyze using machine learning better to spot trend in the datasets, due to more available data. These six airports all also fall under different U.S. climate regions defined by Karl and Koss (1984), with ATL falling under “Southeast”, DEN falling under "Southwest”, DFW falling under "South”, JFK falling under "Northeast”, LAX falling under "West” and “ORD” falling under “Ohio Valley”. The difference in climate conditions will allow for the airports and flight delay times to be differently affected by weather conditions.

3.1.1. Data Sources

For the data using in this thesis there were 2 main sources, the “Performance Variables” of the airports were taken from the Sage Data´s Airline Performance Database (this information being provided to them by the Bureau of Transportation Statistics), meanwhile the “Weather Variables” were taken from the Meteostat database (the information being provided to them by the National Oceanic and Atmospheric Administration). Both these sources also had a “Temporal Variable” - “date”. Even if technically an index variable, it is worth mentioning since it will be important later, in merging all data.

Table 1 differentiates the two kinds of variables mentioned above and their source.

Category	**Variable**	**Source**
Weather Variables	tavg (average temperature)	Meteostat (NOAA)
	tmin (minimum temperature)	Meteostat (NOAA)
	tmax (maximum temperature)	Meteostat (NOAA)
	prcp (precipitation)	Meteostat (NOAA)

	snow (snowfall)	Meteostat (NOAA)
	wdir (wind direction)	Meteostat (NOAA)
	wspd (average wind speed)	Meteostat (NOAA)
	wpgt (wind peak gust)	Meteostat (NOAA)
	pres (sea-level air pressure)	Meteostat (NOAA)
	tsun (total sunshine duration)	Meteostat (NOAA)
Performance Variables	% On-Time Departures	Sage Data – Airline Performance Database
	% On-Time Arrivals	Sage Data – Airline Performance Database
	Total Delay Minutes (Departures)	Sage Data – Airline Performance Database
	Total Delay Minutes (Arrivals)	Sage Data – Airline Performance Database
	NAS Delay Minutes (Departures)	Sage Data – Airline Performance Database
	NAS Delay Minutes (Arrivals)	Sage Data – Airline Performance Database
	15+ Minutes Delays (Departures)	Sage Data – Airline Performance Database
	15+ Minute Delays (Arrivals)	Sage Data – Airline Performance Database
Temporal Variable	date (daily record index)	Common across all datasets

Table 1 - Starting Variables and Sources

The raw files were all extracted in Excel (xlsx) and a total of 54 files were downloaded and organized in structured folders by variable type and airport.

The “Performance Variables” were not aggregate when downloaded, as such, each of the 6 airports had 8 different excel files associated with them, for a total of 48 excel files.

As for the “Weather Variables”, due to a website limitation it was impossible to download the information related to an airport for the chosen time period (01/01/2010 until 01/07/2025) all at once, as such 2 files were downloaded for each of the 6 chosen airports, for a total of 12 excel files. Each airport had 2 weather files associated to them, both with all the “Weather Variables”, but the first contained information for the time period between 01/01/2010 and 01/01/2019, meanwhile the second file had the period between 02/01/2019 and 01/07/2025.

3.1.2. Merging Files

For the merge of the files, 3 python scripts were used.

The first one, “JoiningWeatherDatabases.py” (Annex 1), concatenates the “Weather Variables” files for each of the airports, creating one file for each airport for the time period 01/01/2010 until 01/07/2025 in chronological order, having the “date” variable function as an index variable. This reduces our 12 “Weather Variables” files to 6.

The second script, “Formatting_Performance_Metrics.py” (Annex 2) focused on formatting the “Performance Variables” files. While the excel sheet of Meteostat was only the dataframe containing the weather data, the sheets from Sage Data had extra information, unimportant for our analysis, that should be removed before merging databases.

The last script, “Merging.py” (Annex 3), merged all the variables for each airport, using “date” as the common denominator and an outer join, so no data was lost in the process. Of our original 54 files (48 after running “JoiningWeatherDatabase.py”) this merger left us with our final 6 files, one for each of our studied airports.

3.1.3. Final Datasets

On our 6 final datasets, each has 19 columns and 5661 rows. All datasets cover the time frame between 01/01/2009 and 01/07/2025, and some basic sanity checks showed no duplicates in any of the datasets, but exactly 1032 missing values in each dataset.

As for the variables, we can see their data types and a brief explanation of them in table 2, to serve as an introduction as we explore them further on the EDA.

Variable	Explanation	Data Type
date	Temporal variable on a DD-MM-YY format	datetime64[ns]
tavg	Average temperature for the day (in Celsius)	float64
tmin	Minimum temperature of the day (in Celsius)	float64
tmax	Maximum temperature of the day (in Celsius)	float64
prcp	Total precipitation for the day (in mm)	float64
snow	Depth of snow on that day (in mm)	int64
wdir	Direction from where the wind comes (in degrees)	int64
wspd	Average wind speed of the day (in km/h)	float64
wpgt	Peak wind gust of the day (in km/h)	int64
pres	Air pressure at sea-level (in hPa)	float64
tsun	Total sunshine duration during that day (in minutes)	int64
Total Delay (Minutes) from the Airline Performance - Departures Database	Total delay of departing flights on that airport during that day, measured in minutes	float64

Total Delay (Minutes) from the Airline Performance - Arrivals Database	Total delay of arriving flights on that airport during that day, measured in minutes	float64
NAS Delay (Minutes) from the Airline Performance - Departures Database	Delays in departing flights attributable to the National Aviation System (NAS) on the airport, during that day, measured in minutes	float64
NAS Delay (Minutes) from the Airline Performance - Arrivals Database	Delays in arriving flights attributable to the National Aviation System (NAS) on the airport, during that day, measured in minutes	float64
15 Minute or More Delay from the Airline Performance - Departures Database	Number of delays higher, or equal to 15 minutes in departing flights on the airport, during that day	float64
15 Minute or More Delay from the Airline Performance - Arrivals Database	Number of delays higher, or equal to 15 minutes in arriving flights on the airport, during that day	float64
On-Time % from the Airline Performance - Departures Database	Percentage of departing flights that were on-time (no delay registered) on the airport, during that day	float64
On-Time (%) from the Airline Performance - Arrivals Database	Percentage of arriving flights that were on-time (no delay registered) on the airport, during that day	float64

Table 2 - Starting Variable Types

3.2. Machine Learning Pipeline

The machine learning pipeline used can be seen in the code in Appendix 4. It starts with obtaining and loading the data, followed by some basic sanity checks, then the Exploratory Data Analysis (EDA), where we then determine what needs to be done in Data Cleaning and Transformation, before Training and Testing the models and getting our results.

3.3. Models Used

3.3.1. Train-Test Split

Before moving to our models, we only need to separate our data between Train data, used to train the models, and Test data, used to measure the performance of the model.

Since we are working with a time series and want to predict the delays we used a time-based split. For the split itself we did 80/20, with the first 80% of data being used to train and the last 20% to test, which has been empirically shown to be the best division for the split (Joseph, 2022).

3.3.2. Baseline Model

For the baseline modelling stage, we started with Ordinary Least Squares (OLS) Linear Regression. It provides an interpretable benchmark without penalization, allowing a direct understanding of the relationship between predictors and the target variable.

3.3.3. Regularized Models

After OLS we selected two models that introduced regularizations, to see how our data would perform, these were Ridge and Lasso Regression.

Ridge Regression uses an L2 penalty to shrink coefficient magnitudes, improving stability in the presence of multicollinearity or noisy features. Lasso Regression applies an L1 penalty, which can reduce some coefficients to exactly zero, effectively performing feature selection. This sequence of models allows for a systematic comparison between a simple, fully interpretable baseline (OLS) and progressively

regularized variants that may improve predictive performance while mitigating overfitting.

For Ridge we chose alpha equal to 1.0, being a widely accepted baseline, with even scikit-learn defaulting to 1.0. Meanwhile for Lasso we chose alpha equal to 0.01, going with a smaller value to ensure the model doesn´t over shrink the values, since unlike Ridge, Lasso can zero the coefficients, discarding some of our features.

3.3.4. Tree-Based Models

In addition to the linear modelling family, we selected two non-linear ensemble methods to extend the baseline comparison: Random Forest Regression (RF) and Gradient Boosting Regression (GB). Both approaches are tree-based algorithms capable of capturing complex, non-linear relationships and interactions between variables without the need for explicit feature transformations. RF builds an ensemble of decision trees using bootstrap sampling and random feature selection, averaging their predictions to reduce variance and improve robustness. GB, in contrast, constructs trees sequentially, with each tree correcting the residual errors of the previous ones, thereby often achieving higher accuracy through bias reduction.

4. Exploratory Data Analysis (EDA)

For the visualization of the data, we combined all dataframes in one, adding an “Airport” variable, an object with the airport´s code. This was done to aid the visualization of the data and used only on the EDA step.

4.1. Descriptive Statistics

For the descriptive statistics we will first visualize the “Performance Variables” by Airport, and then by variable, then we´ll look at the “Weather Variables” making the same differentiation.

4.1.2. Performance Variables by Airport

4.1.2.1. ATL

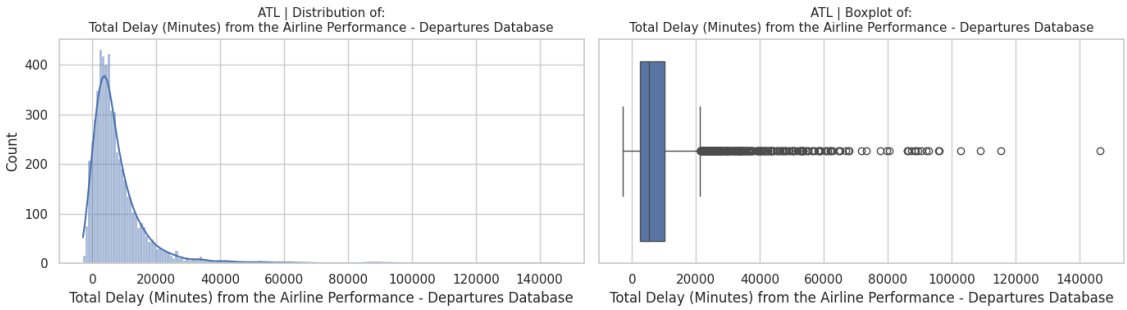


Figure 1 - ATL: Distribution of Total Delay (Departures)

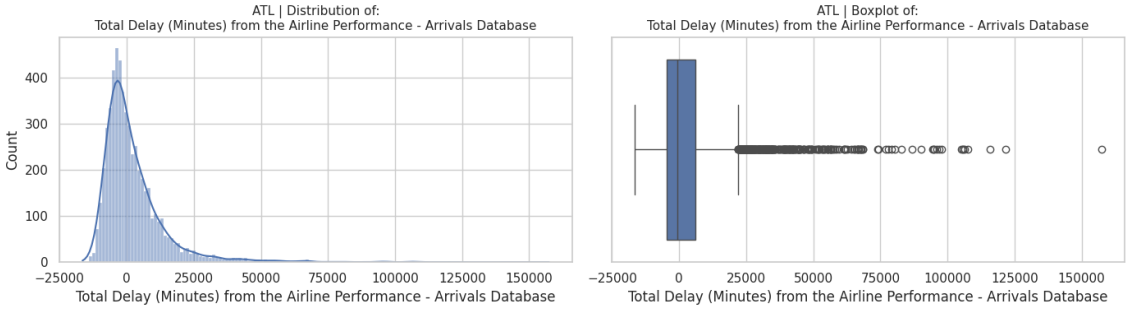


Figure 2 - ATL: Distribution of Total Delay (Arrivals)

The variable *Total Delay (Departures)* at ATL displays a highly right-skewed distribution on *Figure 1*, with most observations concentrated below 20,000 minutes and a long tail extending toward extreme values. A very similar pattern emerges for the variable *Total Delay (Arrivals) in Figure 2*. This consistency between departures and arrivals suggests that the airport is affected by systemic disruptions that propagate across both inbound and outbound operations.

The boxplot confirms the presence of numerous outliers, some exceeding 100,000 minutes, this can be explained by days with exceptional conditions, like extreme weather or system failures – leading to high delays for both departing and arriving flights.

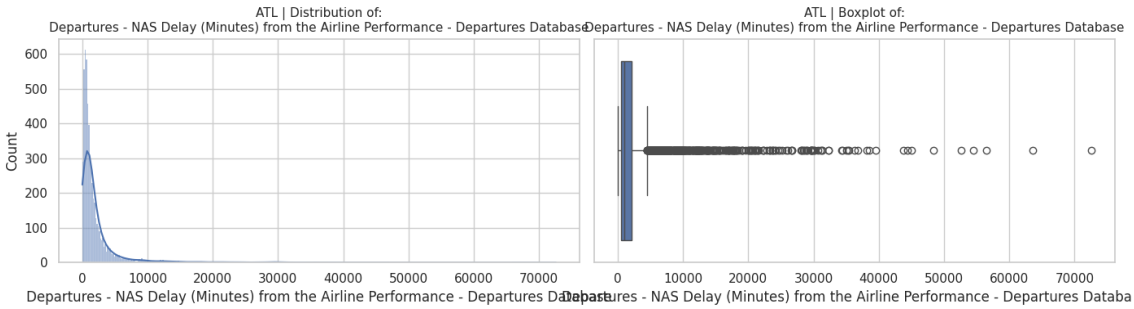


Figure 3 - ATL: Distribution of NAS Delay (Departures)

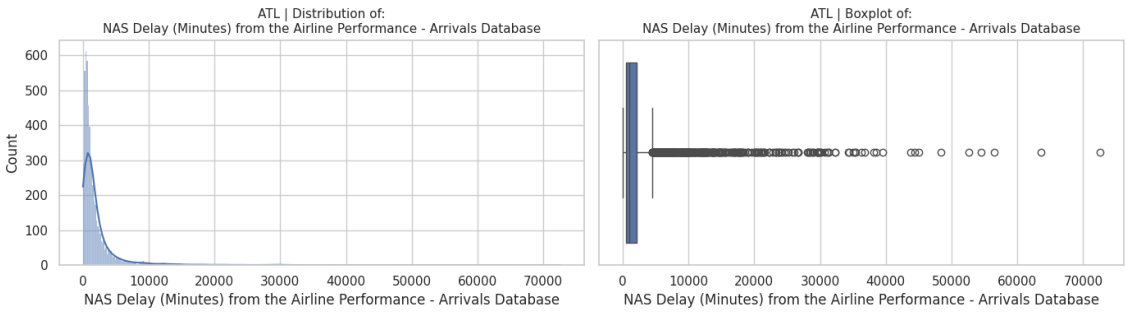


Figure 4 - ATL: Distribution of NAS Delay (Arrivals)

The variables *NAS Delay Departures and Arrivals* show in both *Figure* 3 and 4 a pronounced right skew, with most delays concentrated at lower values but accompanied by several extreme outliers above 70,000 minutes. The boxplots once again confirm this, showing a similar distribution to Total Delay.

Generally, we can see less delays when compared to the Total Delay variable (both for departures as for arrivals), which makes sense when considering NAS Delays are only one of the delay categories included in Total Delay, albeit the broadest one.

It should also be noted here the similarities how similar the NAS arrivals and departures delay graphs are, showing there could be a duplication of the data.

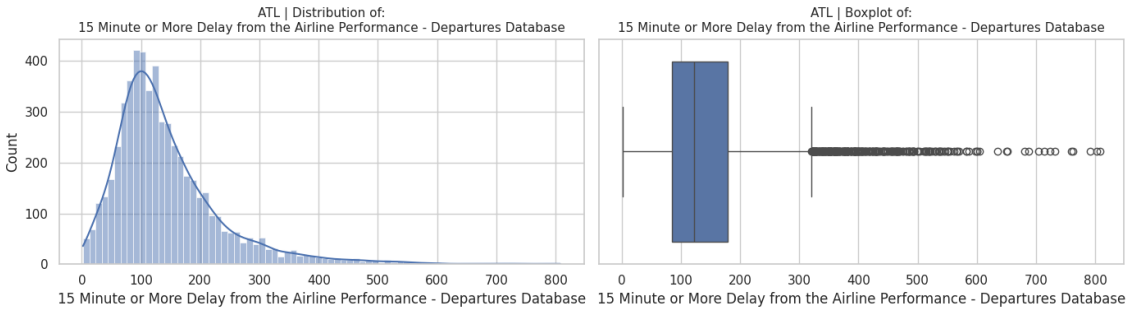


Figure 5 - ATL: Distribution of 15 Min or More Delay (Departures)

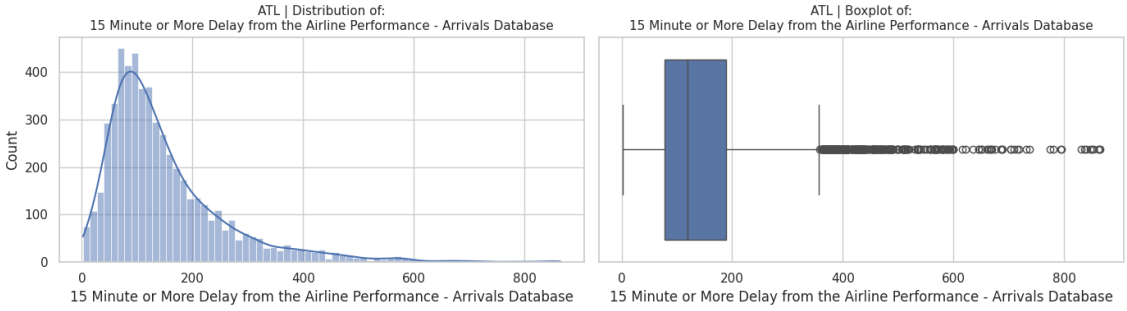


Figure 6 - ATL: Distribution of 15 Min or More Delay (Arrivals)

Figures 5 and 6 show both *15 Minute or More Delay Departures* and *Arrivals* as having right-skewed distributions, with most delays concentrated around 100–200 and a long tail of outliers exceeding 800.

Although this variable still has a skewed distribution, we can see that the removal of "smaller" delays helps to slightly flatten the distribution, even if the distribution is still far from normal.

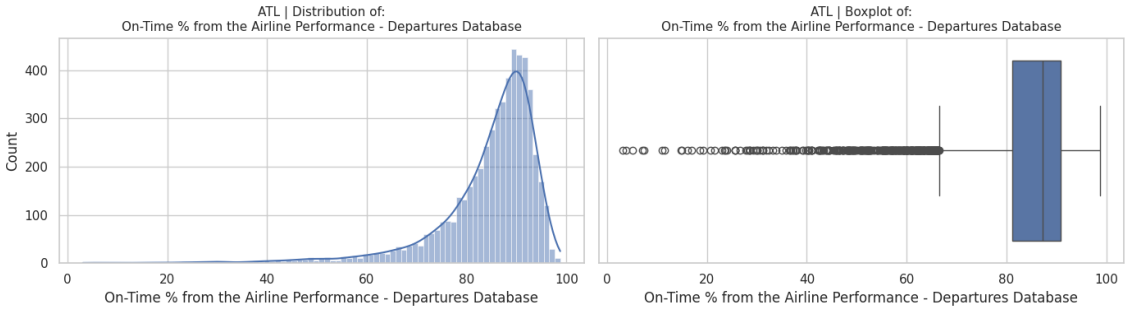


Figure 7 - ATL: Distribution of On-Time % (Departures)

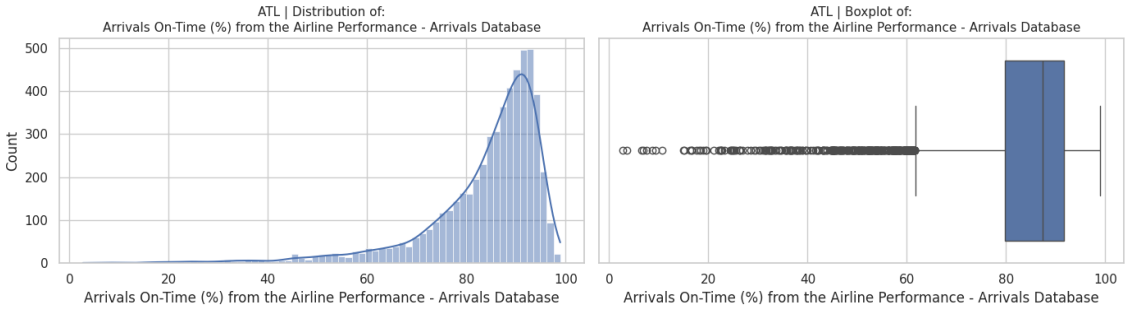


Figure 8 - ATL: Distribution of On-Time % (Arrivals)

The variables *On-Time Departures* and *Arrivals* both display strongly left-skewed distributions, with most values concentrated above 70%. The boxplots confirm high consistency in operational performance, with only limited occurrences of lower on-time percentages. The similarity between departures and arrivals suggests that punctuality patterns are broadly aligned across both traffic flows.

This type of distribution shows that, at least in ATL, the delays come only from a small percentage of flights on most day, which is to be expected.

4.1.2.2. DEN

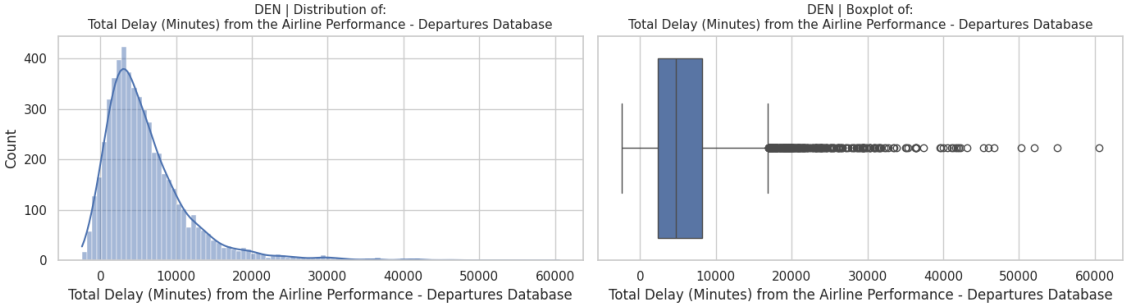


Figure 9 - DEN: Distribution of Total Delay (Departures)

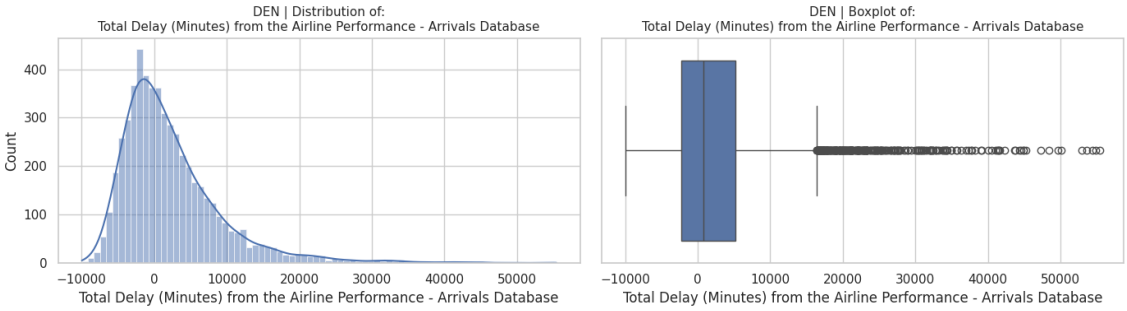


Figure 10 - DEN: Distribution of Total Delay (Arrivals)

The variables *Total Delay Departures* and *Arrivals* show, both in *Figures* 9 and 10, right-skewed distributions, with most values concentrated below 10,000 minutes but extending into several extreme outliers above 50,000 minutes. The boxplots confirm a dense clustering of moderate delays and a long tail of less frequent but severe disruptions, which appear in both departures and arrivals with similar intensity.

When comparing DEN to ATL, the overall shape of distributions is consistent. However, ATL shows a higher number of very extreme cases, with some delays surpassing 100 000 minutes, where DEN´s maximum delays remain below 60 000. This might suggest that ATL might have a bigger systematic vulnerability in terms of delays than DEN.

The arrival distribution also showcases clearly one oddity with the Total Delay variable, that although present in other graphs can be more easily visualized here: negative total delay times. At first glance this could either be a mistake of the data or simply the result of considering early flights as “negative delays” on the sum of Total Delays. We will explore and explain this further when visualizing this variable across all airports.

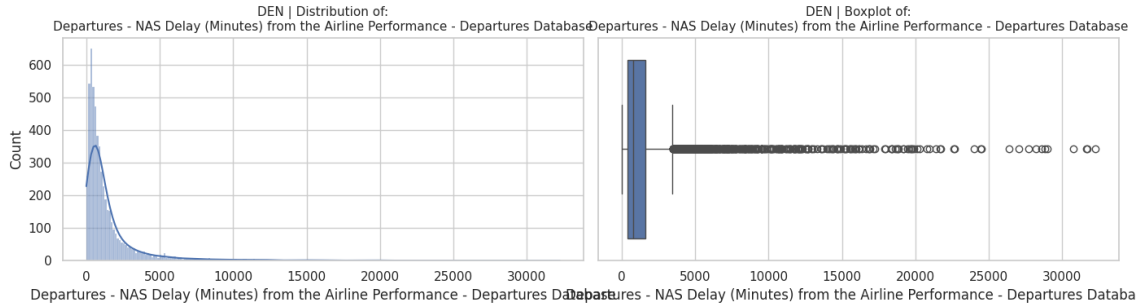


Figure 11 - DEN: Distribution of NAS Delay (Departures)

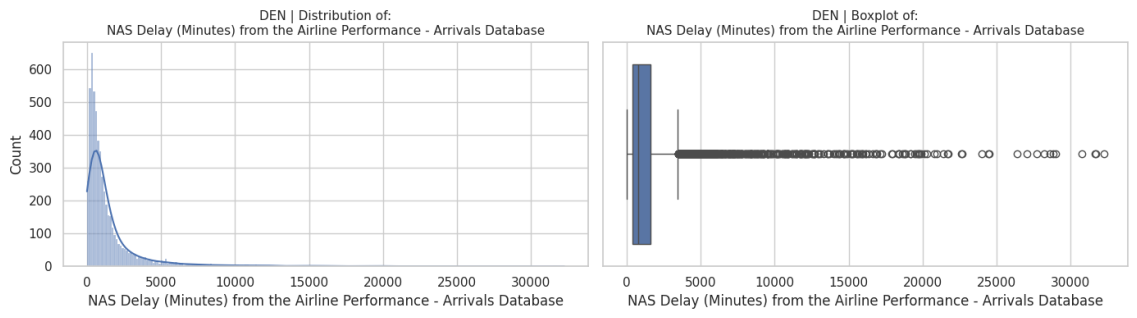


Figure 12 - DEN: Distribution of NAS Delay (Arrivals)

At DEN, in *Figures* 11 and 12 the variables *NAS Delay Departures* and *Arrivals* show strongly right-skewed distributions, with most delays concentrated at very low values

under 5,000 minutes. The boxplots confirm this clustering, while also highlighting the presence of several outliers extending over 30,000 minutes.

In comparison to ATL, the overall distributional form remains consistent, yet ATL displays more extreme outliers, reaching beyond 70,000 minutes. This once again suggests that while both airports face NAS-related vulnerabilities, the intensity of these disruptions is higher at ATL, highlighting a comparatively lower resilience in absorbing and mitigating such large-scale operational shocks.

Once again, we can also see that the similarities between the departures and arrivals graphs seem to be two strong. We will analyse the rest of the airports, but this seems to point to the NAS delays variable being duplicated.

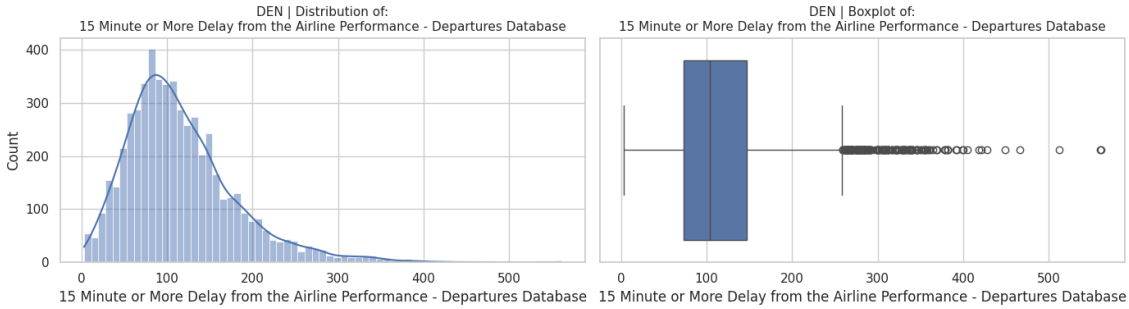


Figure 13 - DEN: Distribution of 15 Min or More Delay (Departures)

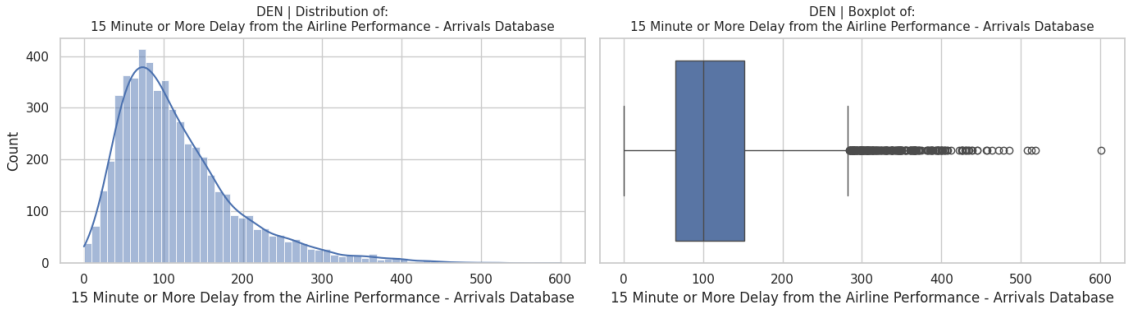


Figure 14 - DEN: Distribution of 15 Min or More Delay (Arrivals)

At DEN, in *Figures* 13 and 14, the variables *15 Minute or More Delay Departures* and *Arrivals* show right-skewed distributions, with most observations concentrated around 100 minutes. The boxplots confirm this central clustering, while also revealing several outliers extending beyond 500 minutes.

Compared to ATL, DEN's distributions are broadly consistent in shape, but ATL exhibits a slightly wider spread and more extreme maximum values, with some delays

surpassing 800 minutes. Continuing our suspicion about DEN´s possible higher robustness.

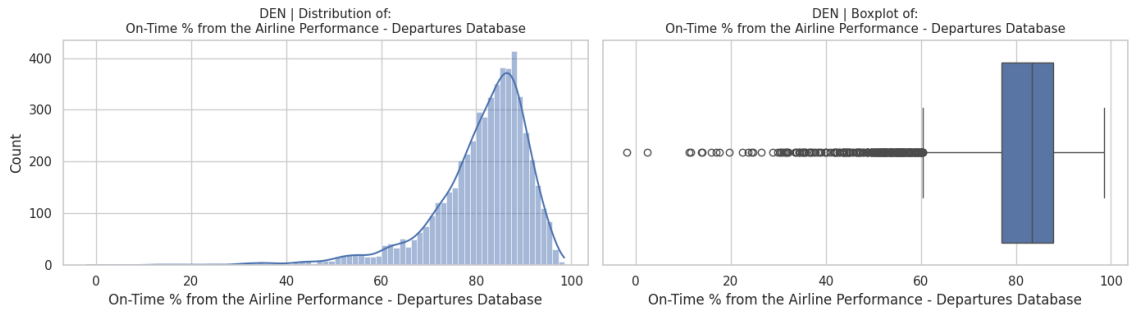


Figure 15 - DEN: Distribution of On-Time % (Departures)

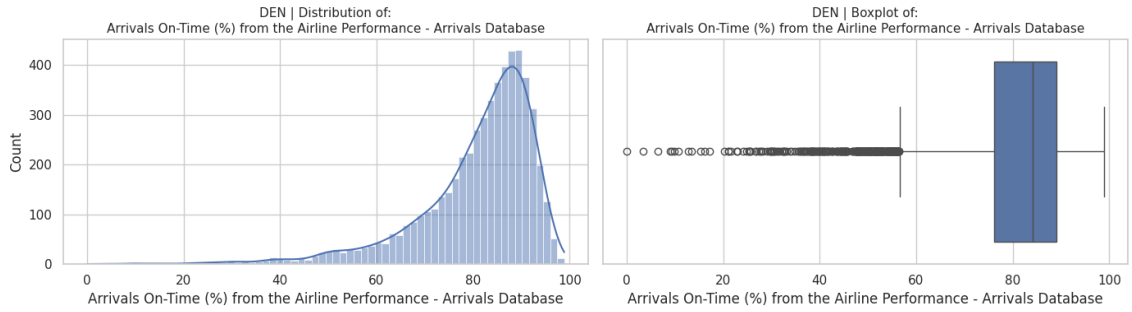


Figure 16 - DEN: Distribution of On-Time % (Arrivals)

At DEN, in *Figures* 15 and 16 the variables *On-Time Departures* and *Arrivals* show left-skewed distributions, with most values concentrated between 75% and 95%. The boxplots indicate consistent operational performance, though some outliers fall well below 60%, highlighting occasional disruptions.

Compared to ATL, DEN's on-time distributions follow a similar pattern but with slightly broader dispersion, as evidenced by a greater spread of outliers. While both airports maintain high overall punctuality, ATL shows a marginally tighter clustering around higher values, suggesting stronger consistency in sustaining on-time performance.

4.1.2.3. DFW

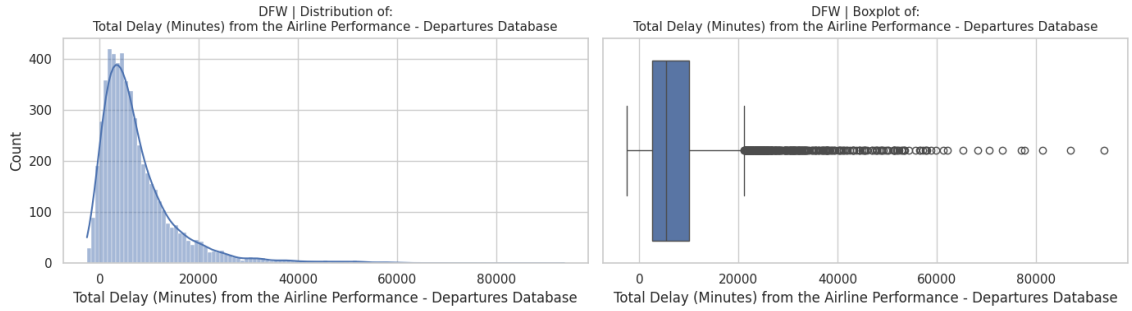


Figure 17 - DFW: Distribution of Total Delay (Departures)

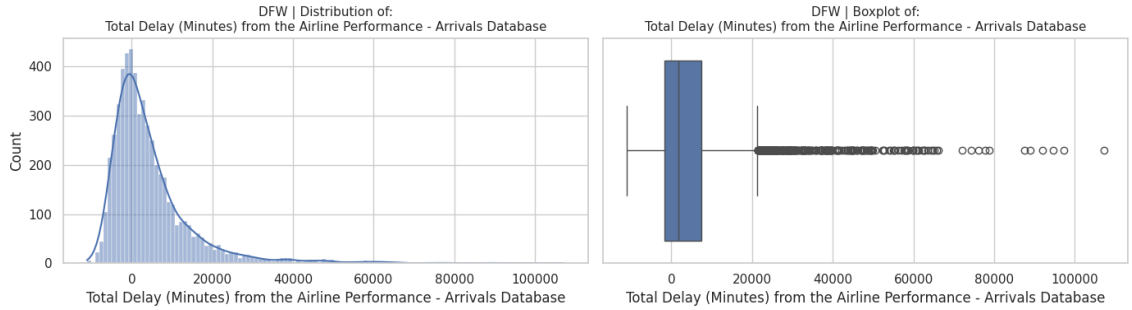


Figure 18 - DFW: Distribution of Total Delay (Arrivals)

At DFW, in *Figures* 17 and 18, the variables *Total Delay Departures* and *Arrivals* are right-skewed, with most values concentrated below 20,000 minutes and several outliers extending beyond 80,000 minutes.

Compared with ATL and DEN, DFW sits between the two in terms of severity. ATL exhibits the most extreme outliers, exceeding 100,000 minutes, while DEN's maximum delays are typically below 60,000 minutes. DFW shows higher extremes than DEN but not as pronounced as ATL.

Once again, the quantity of negative values for total delays should be noted.

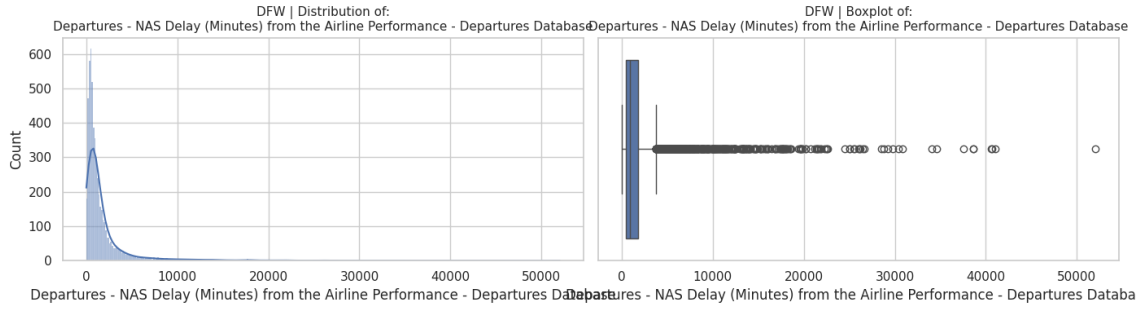


Figure 19 - DFW: Distribution of NAS Delays (Departures)

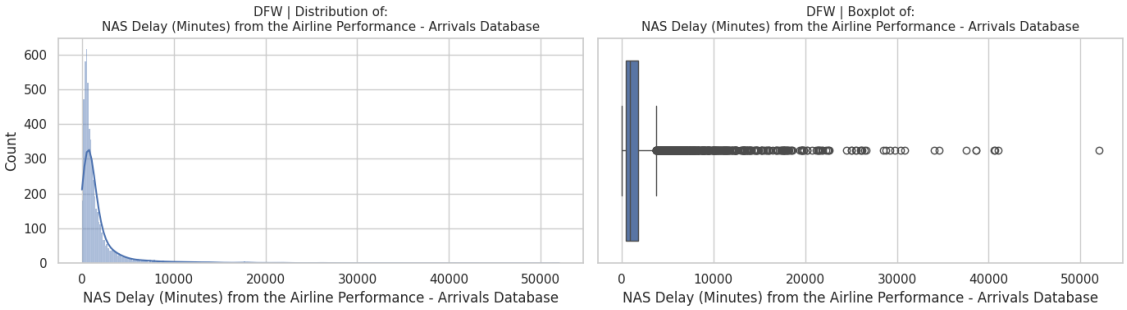


Figure 20 - DFW: Distribution of NAS Delays (Arrivals)

At DFW, in *Figures* 19 and 20, the variables *NAS Delay Departures* and *Arrivals* show sharply right-skewed distributions, with most delays concentrated under 5,000 minutes. The boxplots reveal a dense cluster of low-level disruptions but also highlight reaching beyond 50,000 minutes, affecting both arrivals and departures in a similar way.

When compared to ATL and DEN, DFW again lies in between. ATL shows the most extreme NAS delay outliers, surpassing 70,000 minutes, while DEN's maximum values rarely exceed 30,000 minutes. DFW's distribution extends further than DEN's but not as severely as ATL's.

Again, in DFW we see the risk of data duplication in the NAS Delays, we will keep checking in the remaining airports, but this is something that will be addressed when we reach the data manipulation phase.

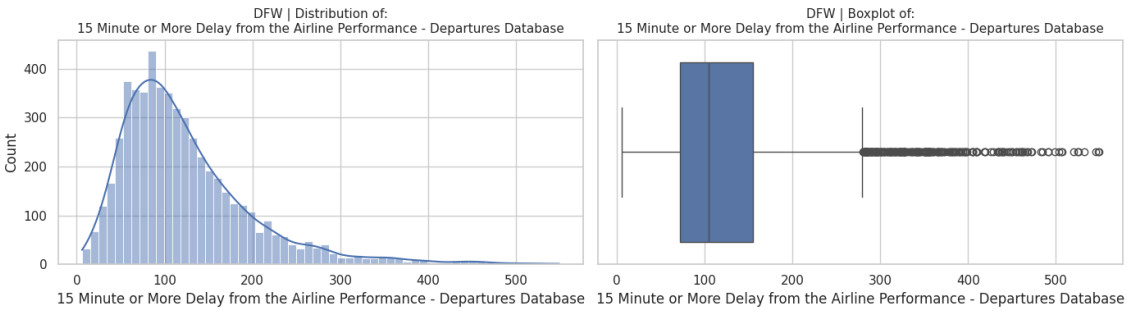


Figure 21 - DFW: Distribution of 15 Min or More Delay (Departures)

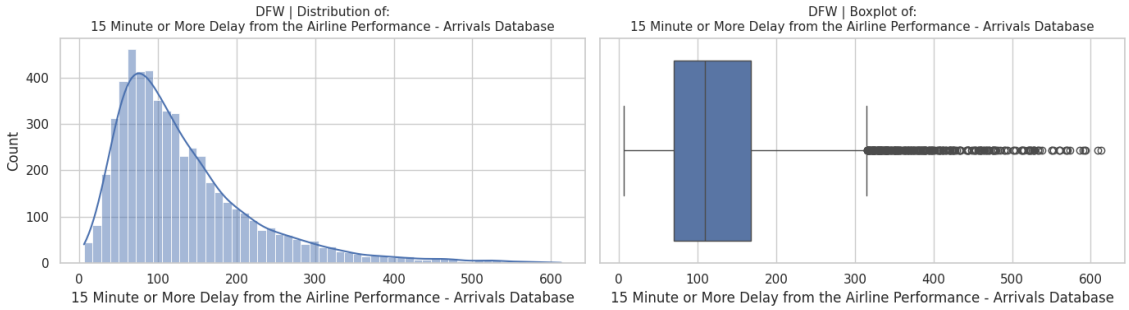


Figure 22 - DFW: Distribution of 15 Min or More Delay (Arrivals)

At DFW, in *Figures* 21 and 22, the variables *15 Minute or More Delay Departures* and *Arrivals* both show right-skewed distributions, with most delays clustered between 50 and 150 minutes. The boxplots highlight this concentration but also reveal outliers extending beyond 500 minutes.

Compared to ATL and DEN, the distributional shape remains similar, but the extent of the outliers at DFW is more aligned with ATL than with DEN. While DEN's maximum values are somewhat lower, ATL and DFW both display longer tails, suggesting they are more exposed to extended delay episodes.

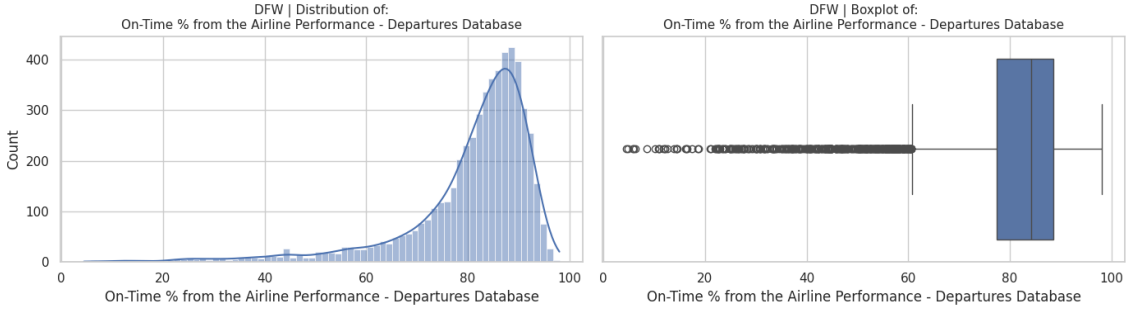


Figure 23 - DFW: Distribution of On-Time % (Departures)

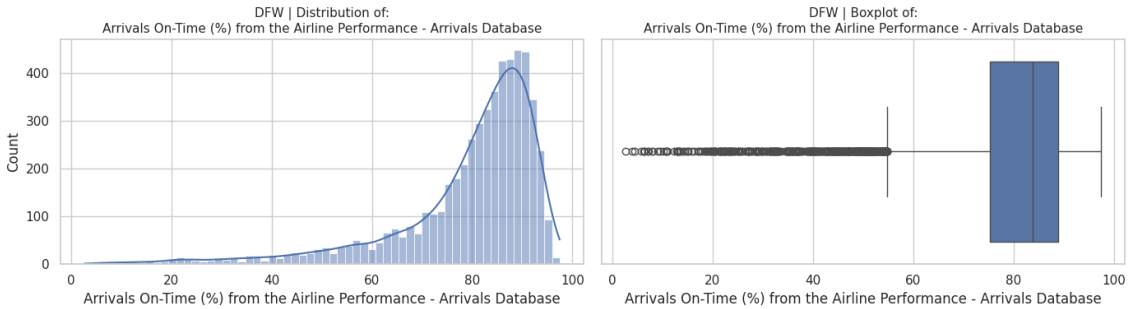


Figure 24 - DFW: Distribution of On-Time % (Arrivals)

At DFW, in *Figures* 23 and 24, the variables *On-Time Departures* and *Arrivals* show left-skewed distributions, with most values concentrated between 75% and 95%. The boxplots confirm consistent performance, though a notable number of outliers fall below 60%, reflecting occasional sharp declines in punctuality.

When compared with ATL and DEN, DFW's distributions are broadly similar, but it shows a slightly greater spread of low outliers than ATL. DEN also displays more dispersion, though with fewer extreme cases. This positions ATL as the most consistent in sustaining high punctuality, with DFW somewhat less robust, and DEN falling in between.

4.1.2.4. JFK

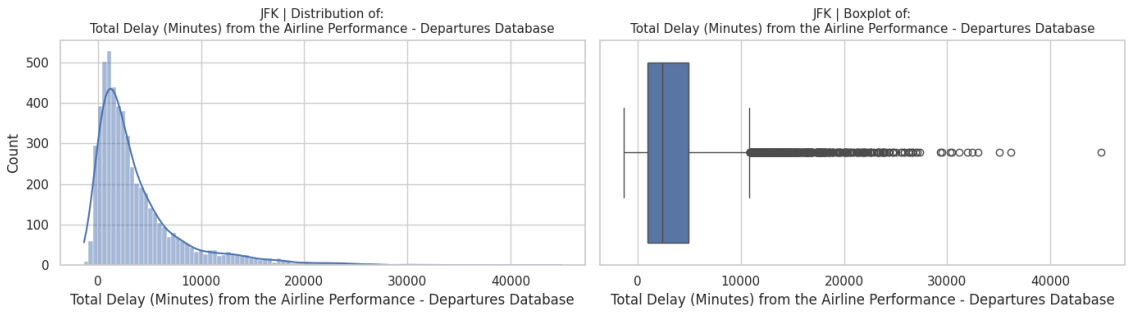


Figure 25 - JFK: Distribution of Total Delay (Departures)

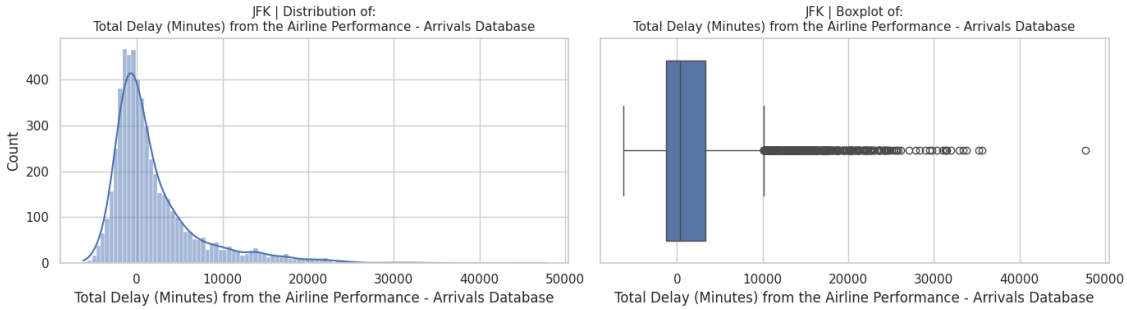


Figure 26 - JFK: Distribution of Total Delay (Arrivals)

At JFK, in *Figures* 25 and 26, the variables *Total Delay Departures* and *Arrivals* are right-skewed, with most delays below 10,000 minutes and several outliers extending up to around 40,000 minutes.

Compared with ATL, DEN, and DFW, JFK shows lower maximum values, as its outliers rarely exceed 40–50,000 minutes. ATL remains the most extreme, surpassing 100,000 minutes, while DEN and DFW occupy an intermediate position. This suggests that,

relative to these hubs, JFK experiences fewer exceptionally long delay events, pointing to comparatively stronger resilience against extreme disruptions.

Just like in previous airports, we can observe negative total delay times. This will be handled in the data manipulation stage.

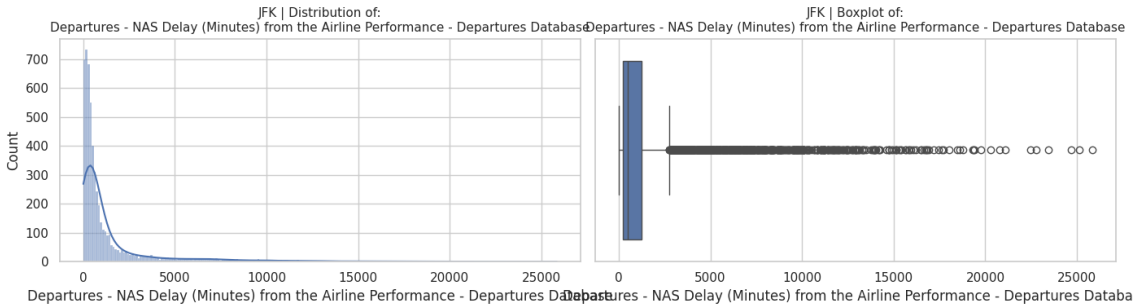


Figure 27- JFK: Distribution of NAS Delay (Departures)

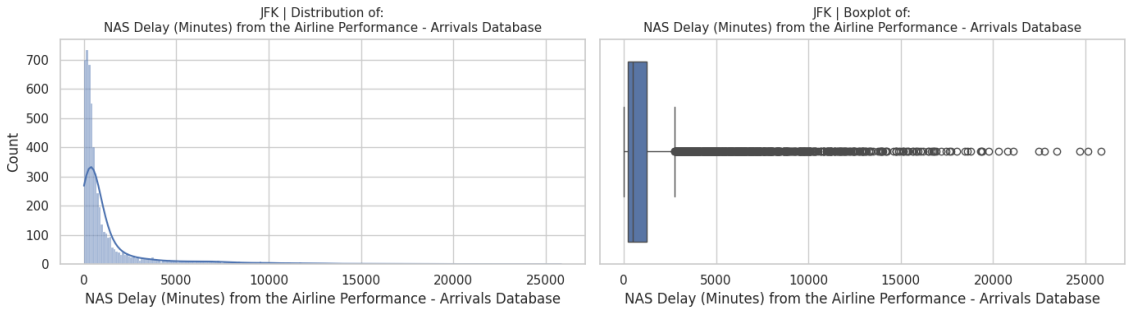


Figure 28 - JFK: Distribution of NAS Delay (Arrivals)

At JFK, in *Figures* 27 and 28, the variables *NAS Delay Departures* and *Arrivals* are highly right-skewed, with most delays concentrated under 5,000 minutes. The boxplots show that while the bulk of disruptions remain modest, occasional outliers extend beyond 25,000 minutes.

Compared with ATL, DEN, and DFW, JFK shows the least extreme NAS-related disruptions, as its outliers remain lower than those of the other hubs.

Just like in the previous airports, NAS Delays also shows signs of duplication in JFK.

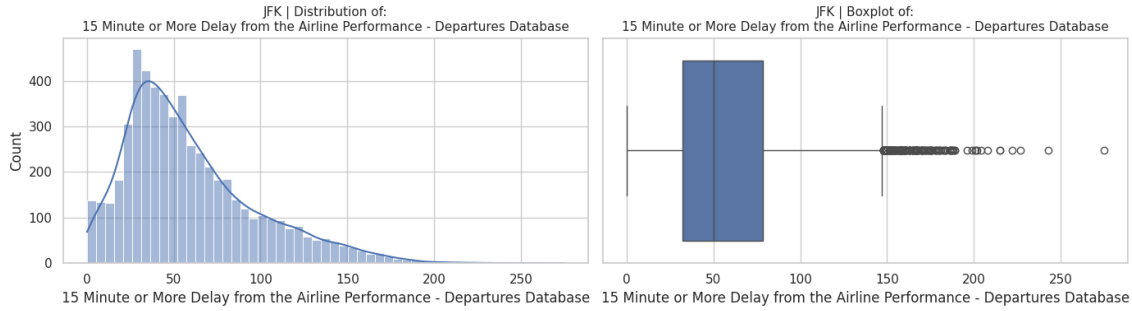


Figure 29 - JFK: Distribution of 15 Min or More Delay (Departures)

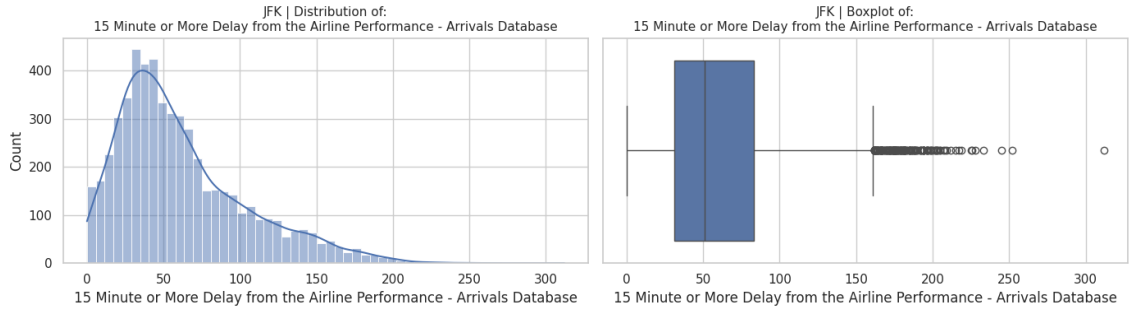


Figure 30 - JFK: Distribution of 15 Min or More Delay (Arrivals)

At JFK, in *Figures* 29 and 30, the variables *15 Minute or More Delay Departures* and *Arrivals* show right-skewed distributions, with most values clustered around 40–70 minutes. The boxplots confirm this central concentration, while also displaying several outliers above 200 minutes.

Compared with ATL, DEN, and DFW, JFK exhibits shorter tails and fewer extreme outliers, as delays rarely exceed 300 minutes. This again suggests that JFK is comparatively less exposed to prolonged delay events.

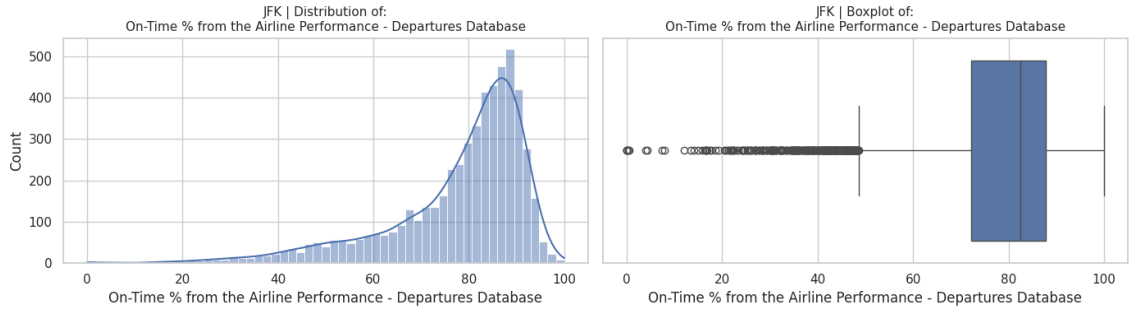


Figure 31 - JFK: Distribution of On-Time % (Departures)

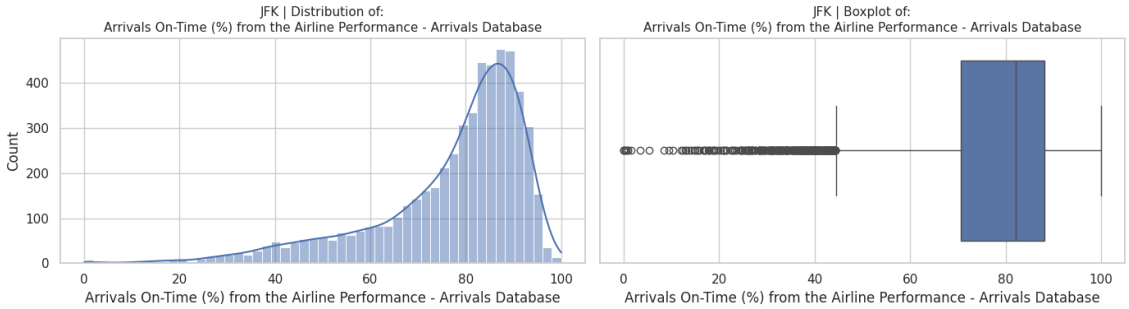


Figure 32 - JFK: Distribution of On-Time % (Arrivals)

At JFK, in *Figures* 31 and 32, the variables *On-Time Departures* and *Arrivals* show left-skewedness, with most values concentrated between 75% and 95%. The boxplots confirm this central clustering, though some outliers fall well below 60%.

Compared with ATL, DEN, and DFW, JFK shows a slightly wider spread of values, including more instances of lower punctuality. This positions JFK as somewhat less consistent in maintaining high on-time performance relative to the other hubs.

4.1.2.5. LAX

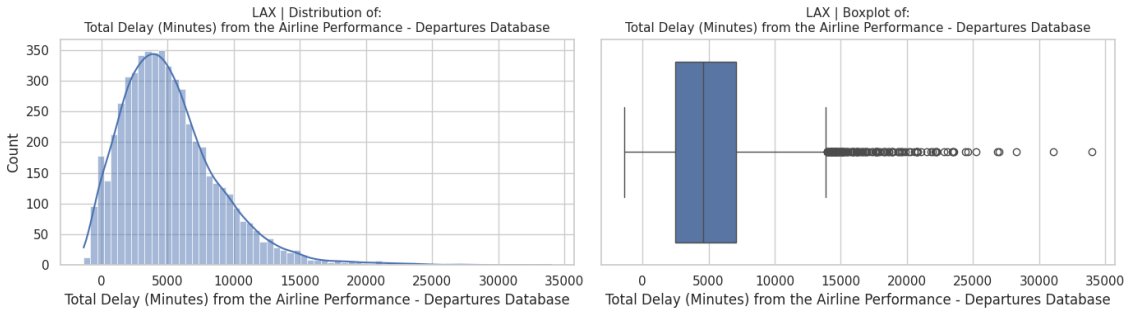


Figure 33 - LAX: Distribution of Total Delay (Departures)

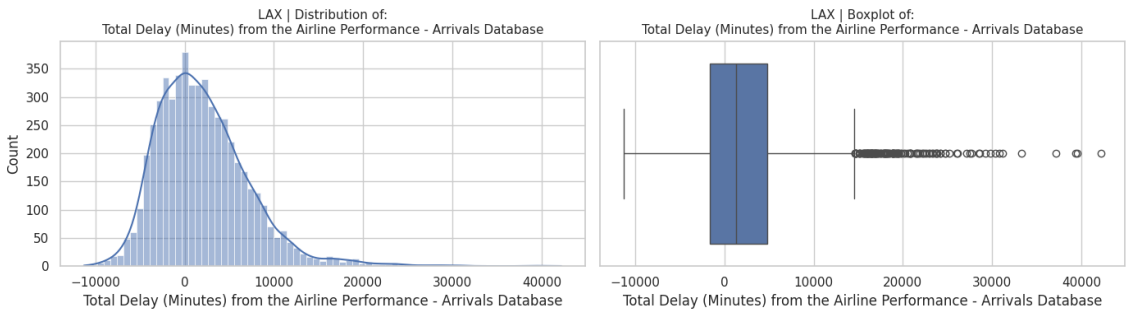


Figure 34 - LAX: Distribution of Total Delay (Arrivals)

At LAX, in *Figures* 33 and 34, the variables *Total Delay Departures* and *Arrivals* are right-skewed, with most values concentrated under 10,000 minutes. The boxplots show a strong clustering around this range, though a few outliers extend beyond 30,000–40,000 minutes.

Compared with ATL, DEN, DFW, and JFK, LAX exhibits less extreme maximum values than ATL and DFW but higher than those observed at DEN and JFK. Suggesting moderate resilience with some vulnerability to prolonged delay events.

The presence of negative total delay times seems even more prevalent here, although these values are surprising, they do seem to be consistent across all airports.

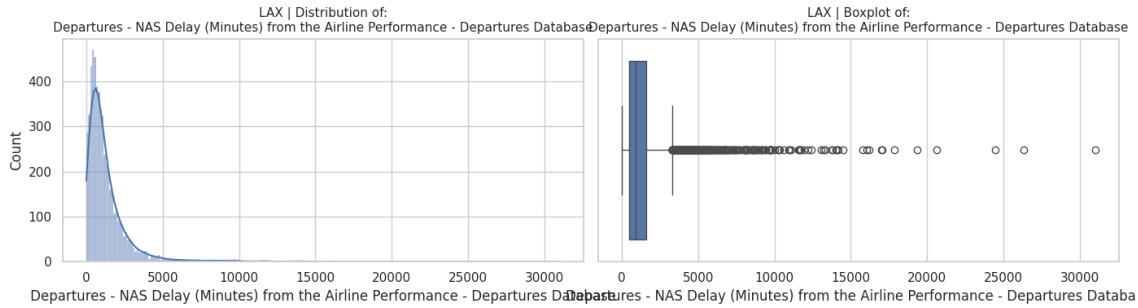


Figure 35 - LAX: Distribution of NAS Delay (Departures)

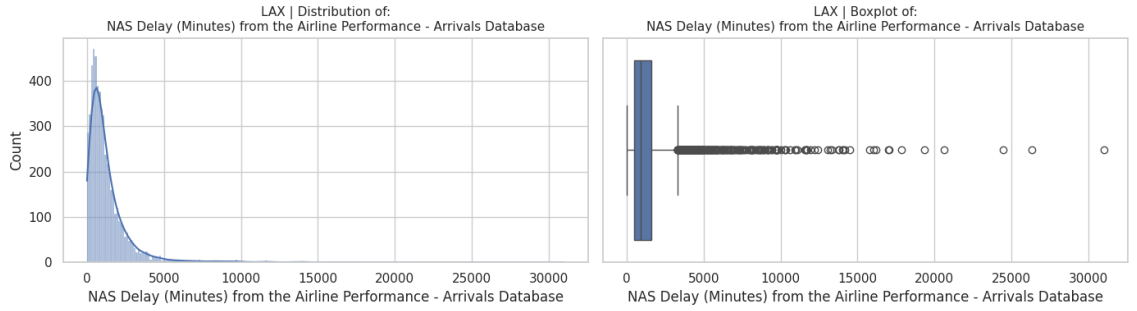


Figure 36 - LAX: Distribution of NAS Delay (Arrivals)

At LAX, in Figures 35 and 36, the variables *NAS Delay Departures* and *Arrivals* display highly right-skewed distributions, with most delays falling below 5,000 minutes. The boxplots highlight this clustering but also show outliers surpassing 30,000 minutes.

When compared to ATL, DEN, DFW, and JFK, LAX falls again closer to the middle. It experiences more extreme NAS-related delays than JFK, but generally less than ATL and DFW.

The already acknowledged data duplication issue is once again noticeable.

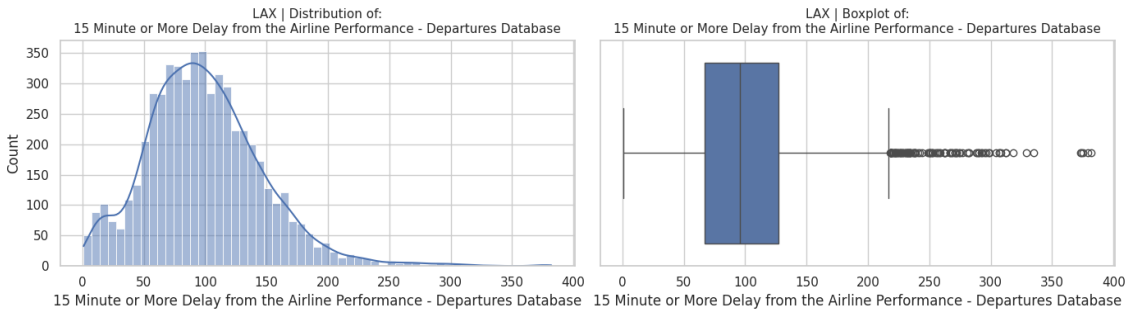


Figure 37 - LAX: Distribution of 15 Min or More Delay (Departures)

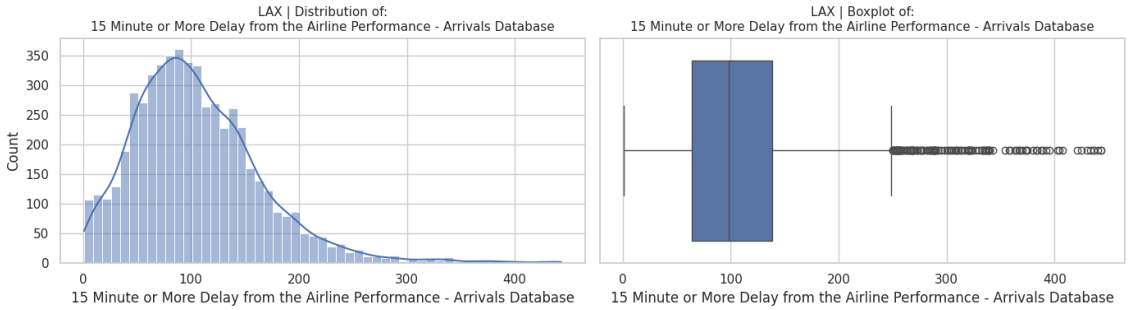


Figure 38 - LAX: Distribution of 15 Min or More Delay (Arrivals)

At LAX, in *Figures* 37 and 38, the variables *15 Minute or More Delay Departures* and *Arrivals* show right-skewedness, with most delays concentrated between 70 and 120 minutes. The boxplots highlight this clustering, while also showing outliers surpassing 400 minutes.

Compared with ATL, DEN, DFW, and JFK, LAX shows a distribution more contained than ATL and DFW, but broader than JFK, similarly to DEN.

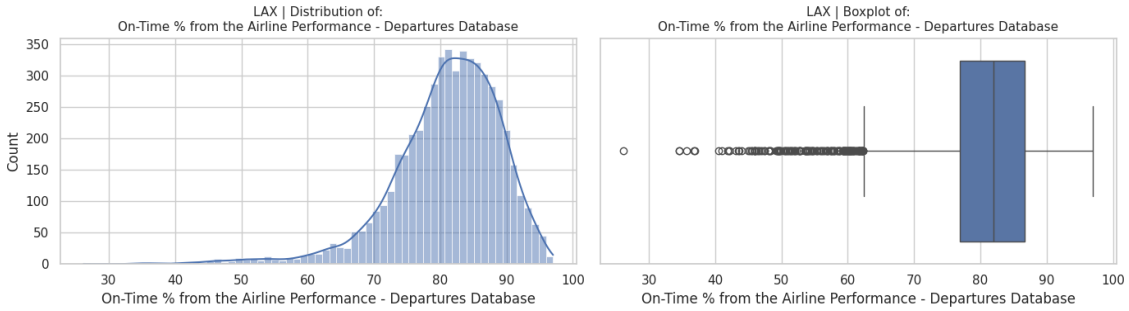


Figure 39 - LAX: Distribution of On-Time % (Departures)

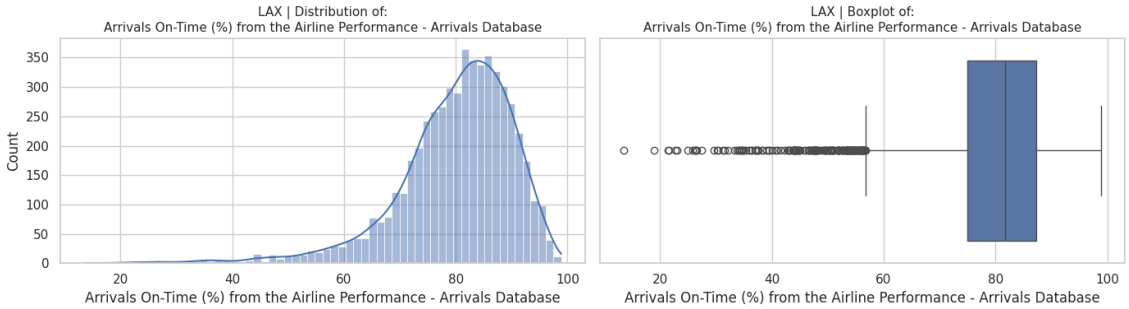


Figure 40 - LAX: Distribution of On-Time % (Arrivals)

At LAX, in *Figures* 39 and 40, the variables *On-Time Departures* and *Arrivals* show left-skewed distributions, with most values concentrated between 75% and 90%. The boxplots confirm this consistency, though occasional outliers drop below 50%.

Compared with ATL, DEN, DFW, and JFK, LAX demonstrates slightly stronger clustering around higher values, suggesting more stable performance. While ATL remains the most consistent overall, LAX outperforms DEN, DFW, and JFK in terms of limiting the spread of low outliers.

4.1.2.6. ORD

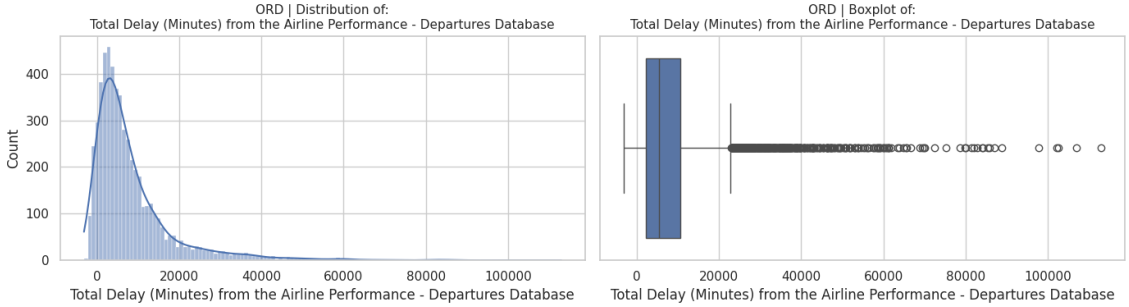


Figure 41 - ORD: Distribution of Total Delay (Departures)

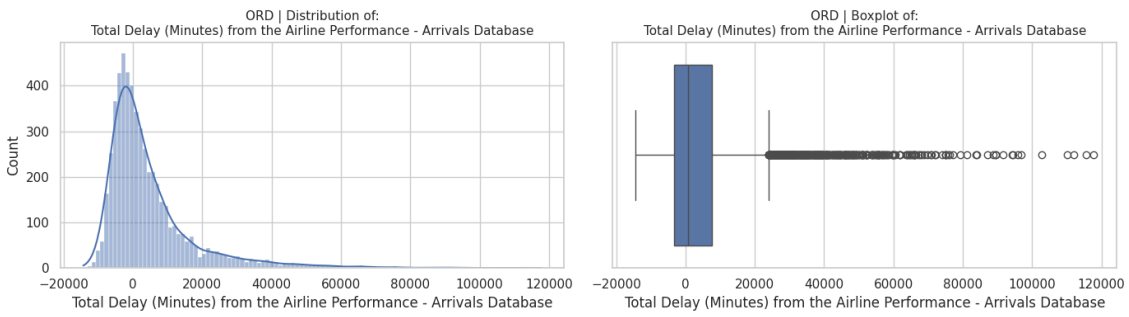


Figure 42 - ORD: Distribution of Total Delay (Arrivals)

At ORD, in *Figures* 41 and 42, the variables *Total Delay Departures* and *Arrivals* show right-skewedness, with most values clustered below 15,000 minutes. The boxplots highlight this central concentration but also show many outliers, some extending beyond 100,000 minutes, particularly on the arrivals side.

Compared with ATL, DEN, DFW, JFK, and LAX, ORD shows a profile closer to ATL and DFW, with extreme outliers exceeding 100,000 minutes. DEN, JFK, and LAX present lower maximums, rarely surpassing 50,000 minutes. This positions ORD among the most exposed airports in terms of extreme total delay events, suggesting comparatively weaker resilience against prolonged disruptions.

This also confirms that the data for all airports contains negative Total Delay times, meaning it is most likely not a measuring error, but product of the way total delay is calculated.

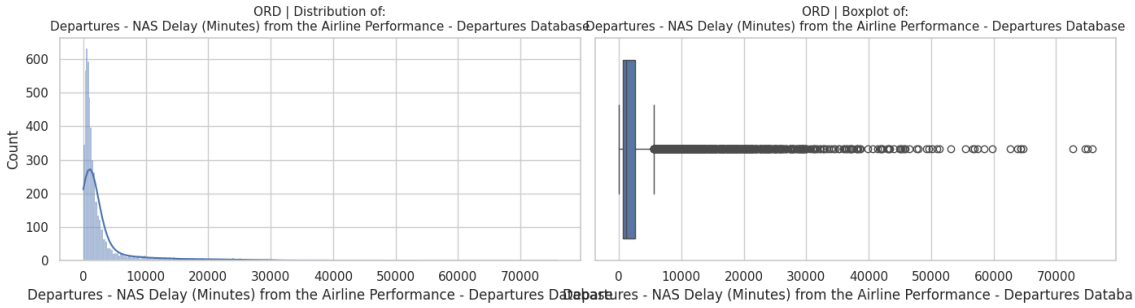


Figure 43 - ORD: Distribution of NAS Delay (Departures)

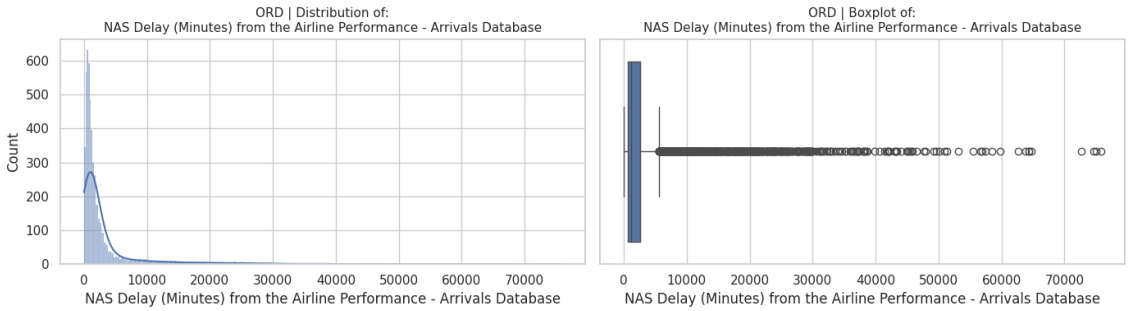


Figure 44 - ORD: Distribution of NAS Delay (Arrivals)

At ORD, in *Figures* 43 and 44, the variables *NAS Delay Departures* and *Arrivals* are strongly right-skewed, with most delays falling under 5,000 minutes. The boxplots reveal a dense concentration of moderate values, alongside numerous outliers, some surpassing 70,000 minutes.

Compared with ATL, DEN, DFW, JFK, and LAX, ORD resembles ATL and DFW in exhibiting the most severe outliers. This positions ORD among the most unstable hubs when it comes to NAS-related disruptions, suggesting lower resilience relative to most of its peers.

With this we can confirm that for all airports NAS delays shows evidence of duplication of the data.

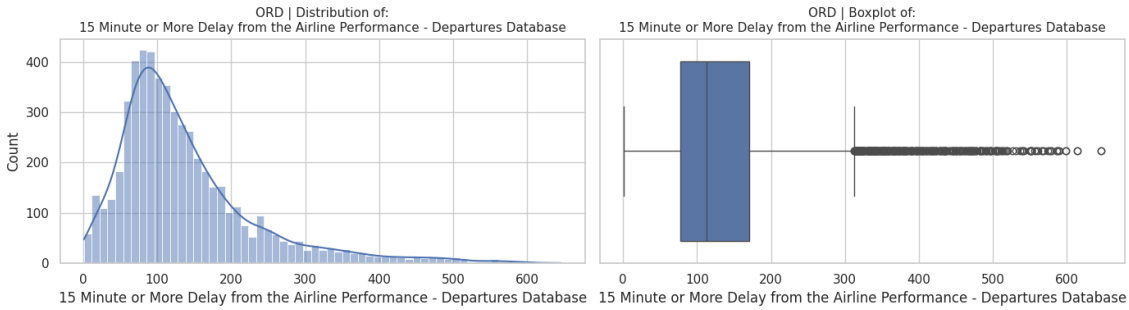


Figure 45 - ORD: Distribution of 15 Min or More Delay (Departures)

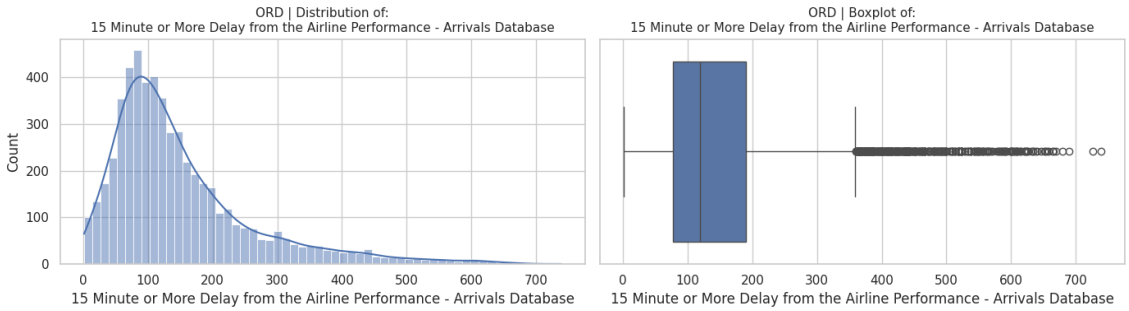


Figure 46 - ORD: Distribution of 15 MIN or More Delay (Arrivals)

At ORD, in *Figures* 45 and 46, the variables *15 Minute or More Delay Departures* and *Arrivals* show right-skewed distributions, with most delays clustered between 80 and 150 minutes. The boxplots confirm this concentration but also reveal a long tail of outliers, some exceeding 700 minutes.

Compared with ATL, DEN, DFW, JFK, and LAX, ORD more closely resembles ATL and DFW, where extreme values surpass 600 minutes.

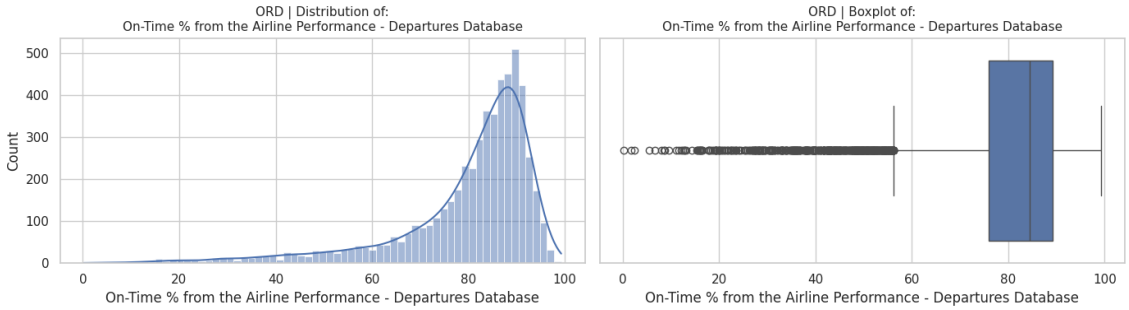


Figure 47 - ORD: Distribution of On-Time % (Departures)

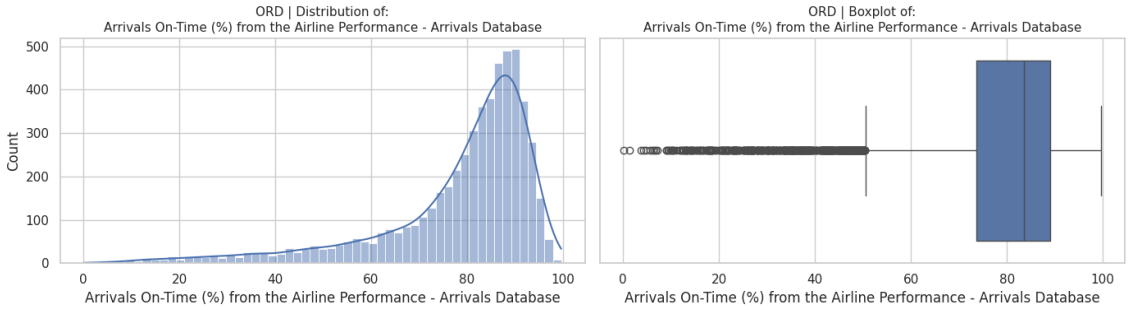


Figure 48 - ORD: Distribution of On-Time % (Arrivals)

At ORD, in *Figures* 47 and 48, the variables *On-Time Departures* and *Arrivals* are left-skewed, with most values concentrated between 75% and 95%. The boxplots confirm this clustering but also show a large spread of outliers below 60%.

Compared with ATL, DEN, DFW, JFK, and LAX, ORD demonstrates more dispersion and a higher frequency of low outliers. ATL and LAX remain the most consistent performers, showing tighter clustering around higher percentages, while JFK, DFW, and DEN fall in intermediate positions. This places ORD among the less efficient airports in maintaining punctuality.

4.1.3. Performance Variables by Variable

4.1.3.1. Total Delays

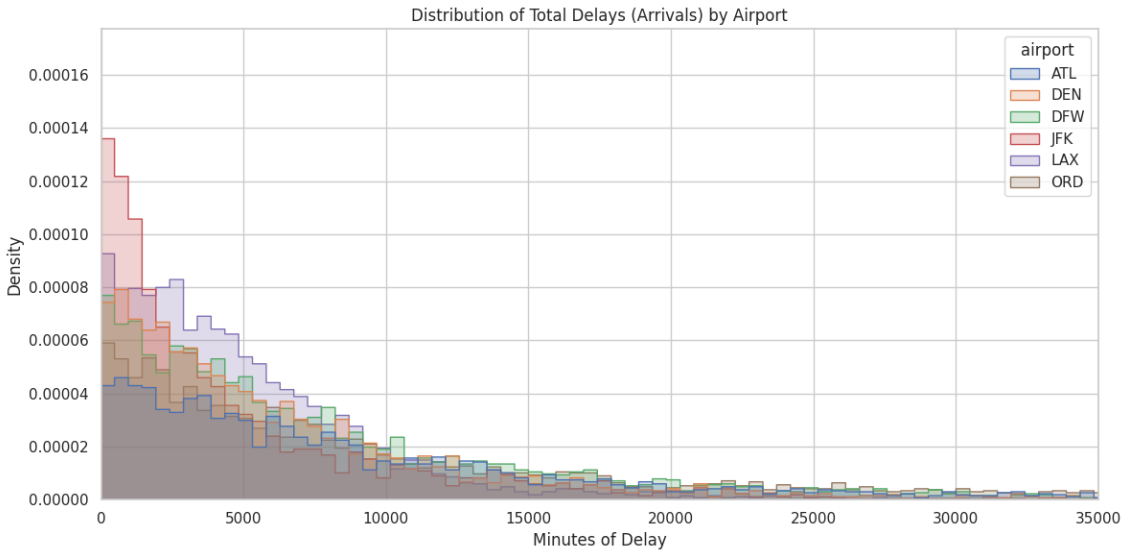


Figure 49 - Distribution of Total Delays (Arrivals) by Airport

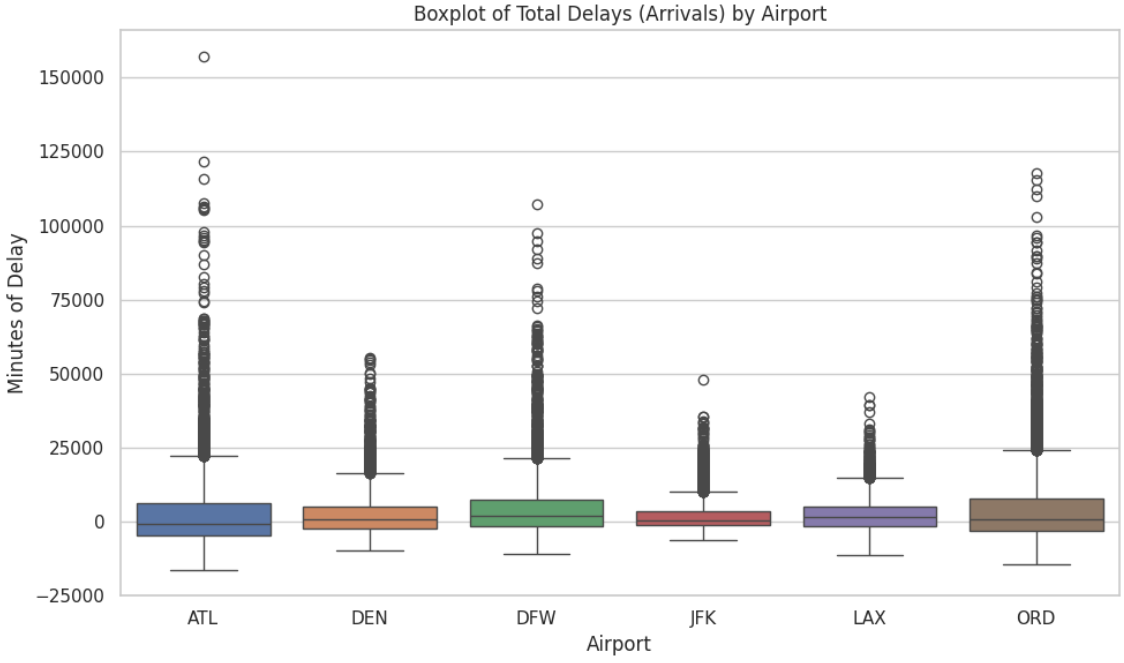


Figure 50 - Boxplots of Total Delays (Arrivals) by Airport

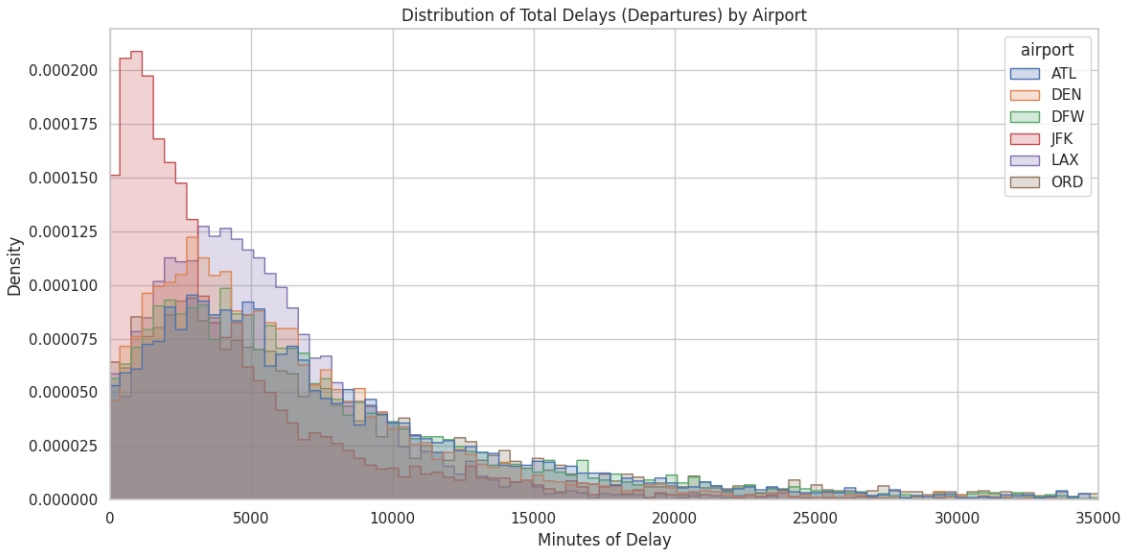


Figure 51 - Distribution of Total Delays (Departures) by Airport

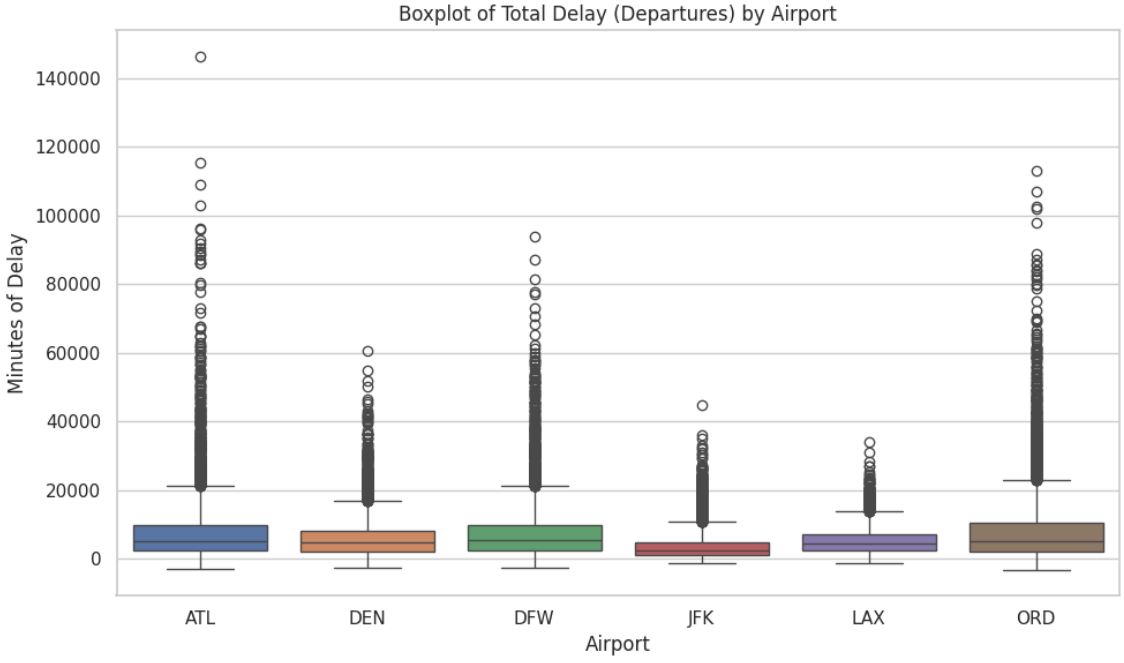


Figure 52 - Boxplot of Total Delay (Departures) by Airport

In *Figures* 49 and 51, distributions of *Total Delay* for arrivals and departures are consistently right skewed across all airports, with most values concentrated at lower ranges and a long tail of extreme delays. The boxplots (*Figures* 50 and 52) highlight significant variation in the scale of outliers: ATL, ORD, and DFW record the most severe extremes, with some events exceeding 100,000 minutes, whereas JFK consistently shows the lowest maximums, rarely surpassing 50,000 minutes. DEN and

LAX occupy an intermediate position, with distributions extending beyond 30,000 minutes but not reaching the extremes observed at the busiest hubs.

From the boxplots we can see that the lowest values of total delays are consistently negative across all airports and both in departures and arrivals, even if the negative occurrences are more accentuated in arrivals. These negative values are due to the “early” flights being used in the calculation of total delay, in the form of “negative delays”, this also explains why they occur more commonly in arrivals, since early arrivals are more common than early departures, since early departures depend not only on air control approval but also on all passengers being boarded ahead of time.

4.1.3.2. NAS Delay

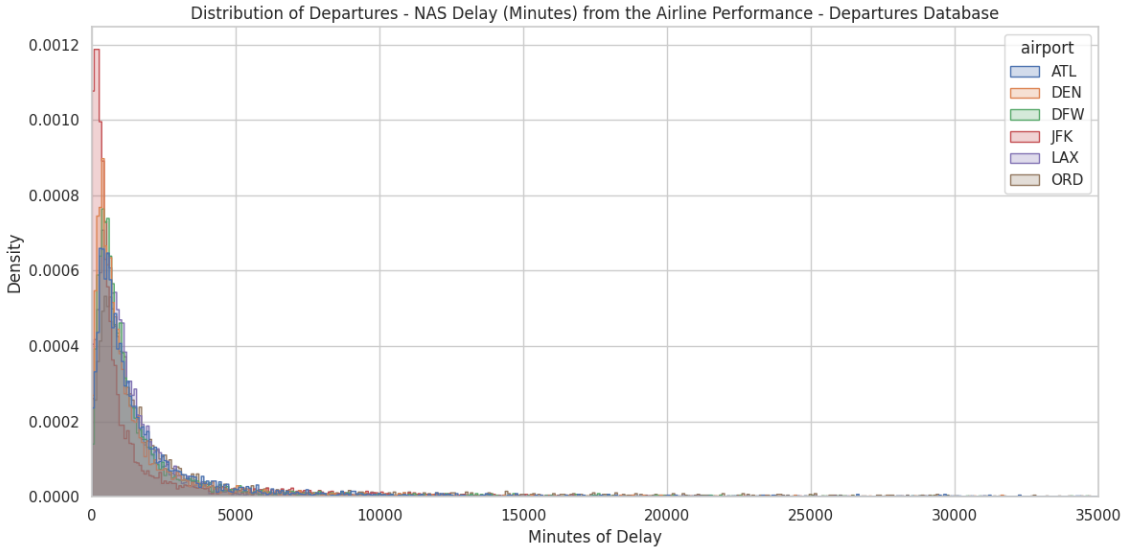


Figure 53 - Distribution of NAS Delay (Departures) by Airport

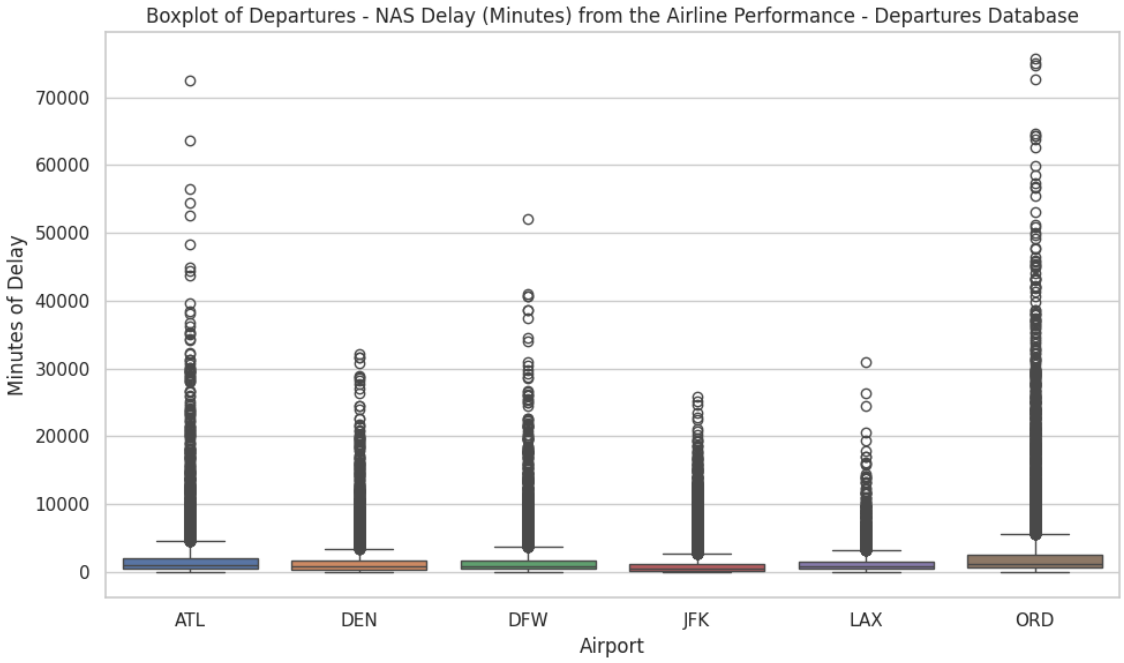


Figure 54 - Boxplot of NAS Delay (Departures) by Airport

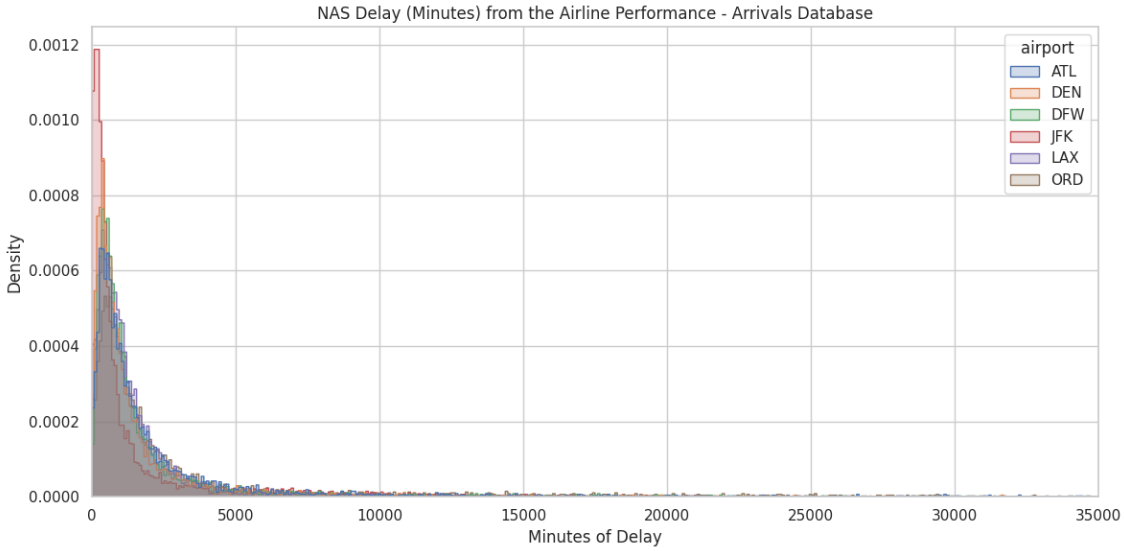


Figure 55 - Distribution of NAS Delay (Arrivals) by Airport

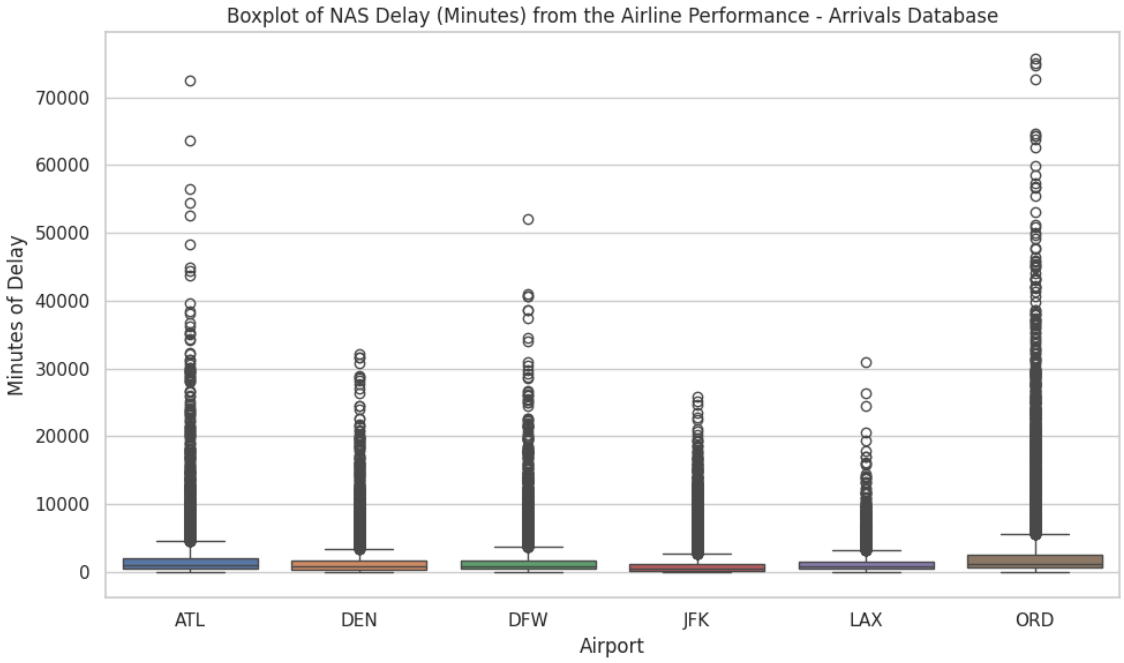


Figure 56 - Boxplot of NAS Delay (Arrivals) by Airport

In *Figures* 53 and 55, the distributions of *NAS Delay* for both departures and arrivals are sharply right skewed across all airports, with most values concentrated below 5,000 minutes. However, the boxplots (*Figures* 54 and 56) reveal substantial differences in the scale of extreme outliers. ORD and ATL exhibit the most severe cases, with delays exceeding 70,000 minutes, followed by DFW, which reaches above 50,000 minutes. DEN and LAX show intermediate patterns, rarely surpassing 30,000 minutes, while JFK consistently records the lowest extremes, remaining under this threshold.

As for the similarity between the departure and arrival distribution, we can attribute this to a duplication error, since realistically it makes no sense for National Aviation System delays to be the same in arriving and departing flights. This makes this variable an easy contender for dropped in the feature selection stage, since it would be too unreliable to use as the dependent variable.

4.1.3.3. 15 Minute or More Delay

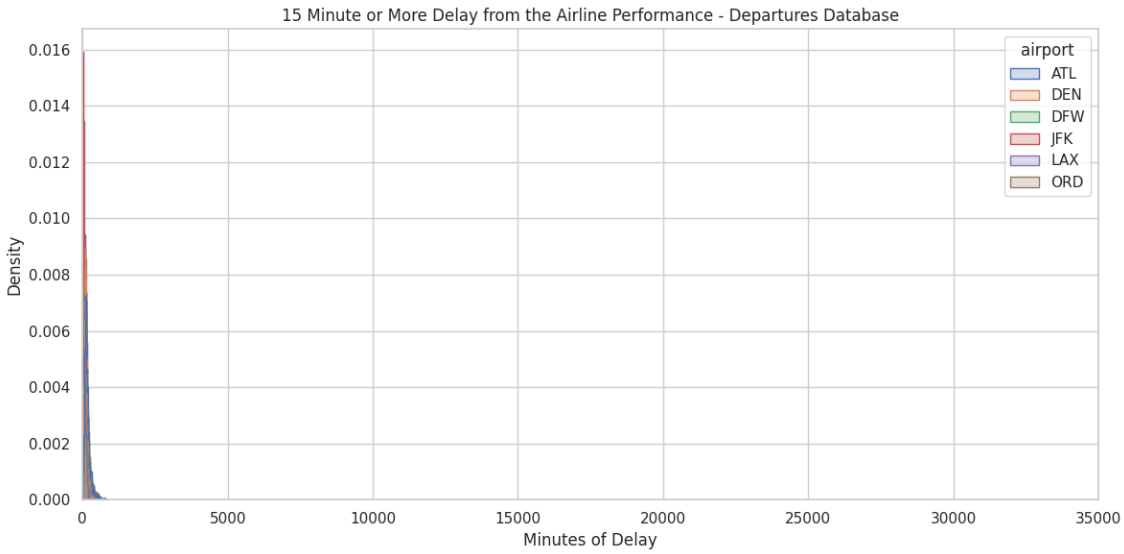


Figure 57 - Distribution of 15 Min or More Delay (Departure) by Airport

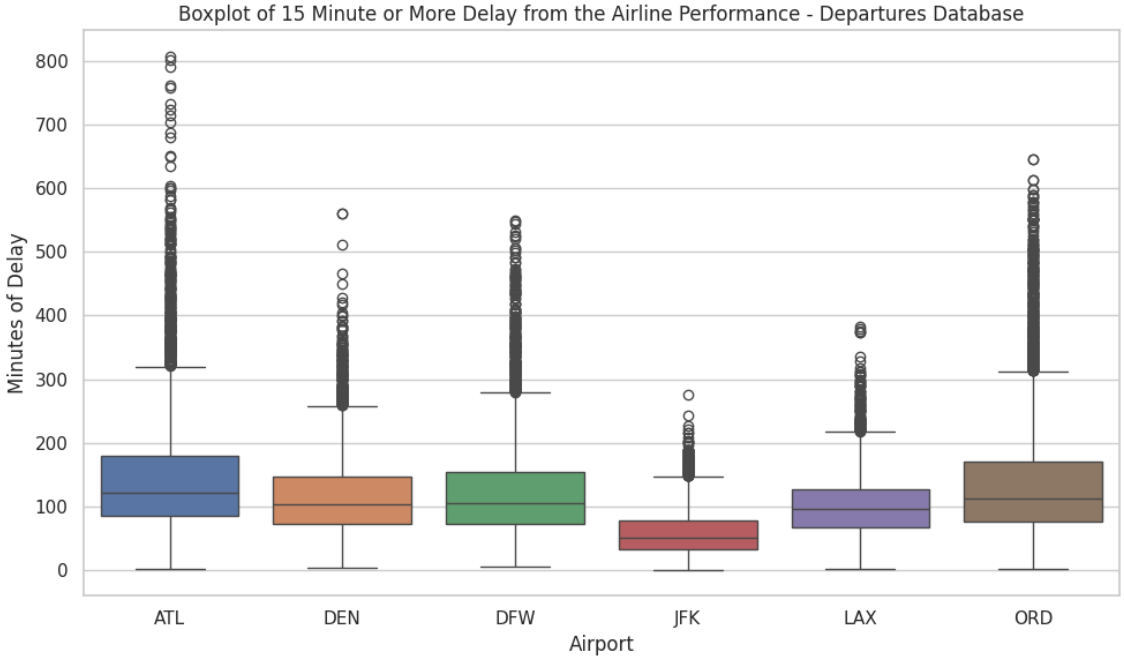


Figure 58 - Boxplot of 15 Min or More Delay (Departure) by Airport

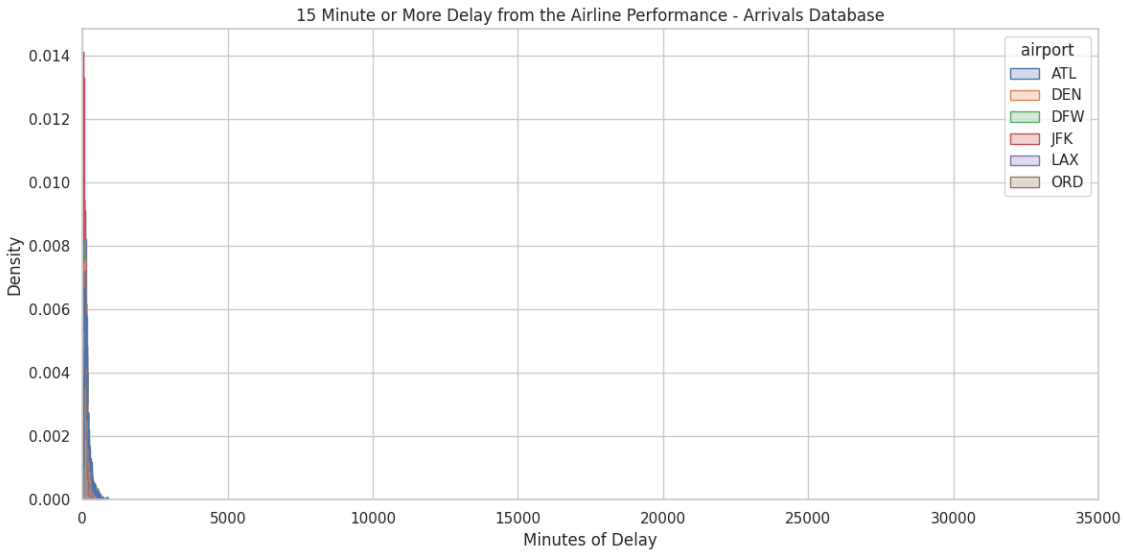


Figure 59 - Distribution of 15 Min or More Delay (Arrivals) by Airport

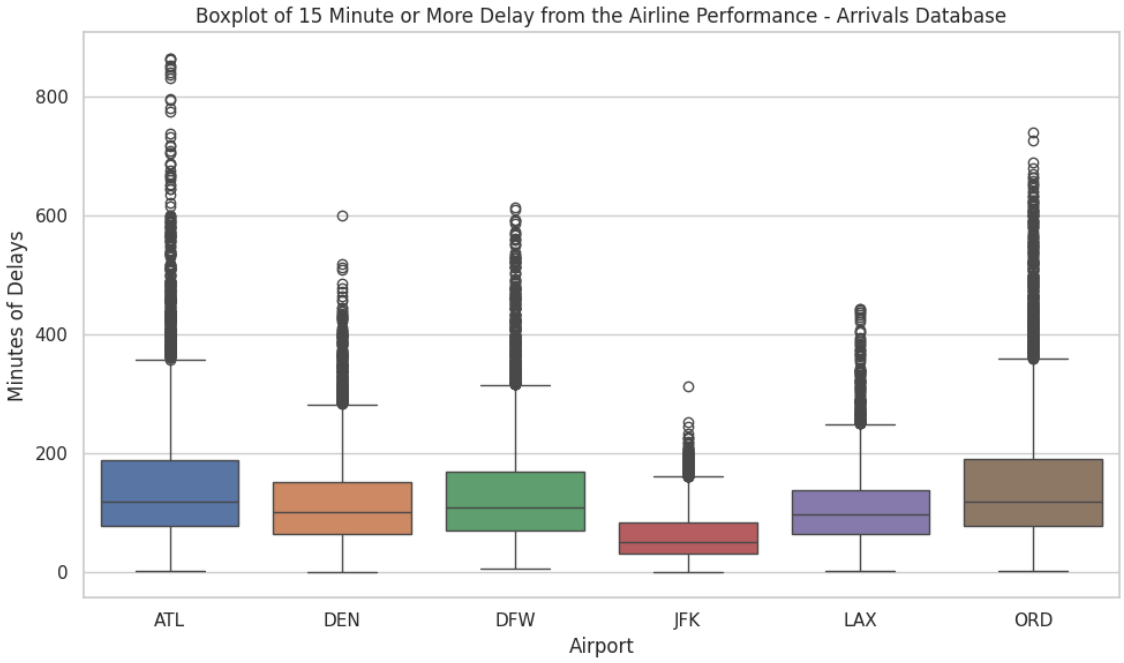


Figure 60 - Boxplot of 15 Min or More Delay (Arrivals) by Airport

In *Figures* 57 and 59, the distributions of *15 Minute or More Delay* for both departures and arrivals are right-skewed, with most values clustered between 80 and 150 minutes across airports. The boxplots (*Figures* 58 and 60) reveal that while this central range is consistent, the scale of outliers varies considerably. ATL, ORD, and DFW show the most severe extremes, with delays surpassing 800 minutes, whereas JFK rarely exceeds 300 minutes. DEN and LAX fall in between, with maximums generally capped at 500–600 minutes.

4.1.3.4. On-Time

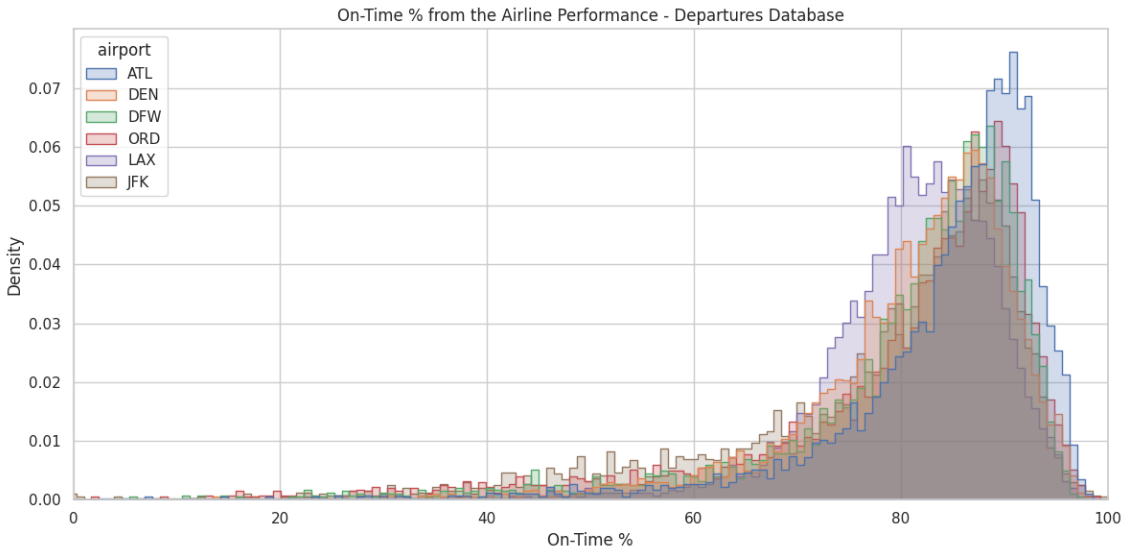


Figure 61 - Distribution of On-Time % (Departures) by Airport

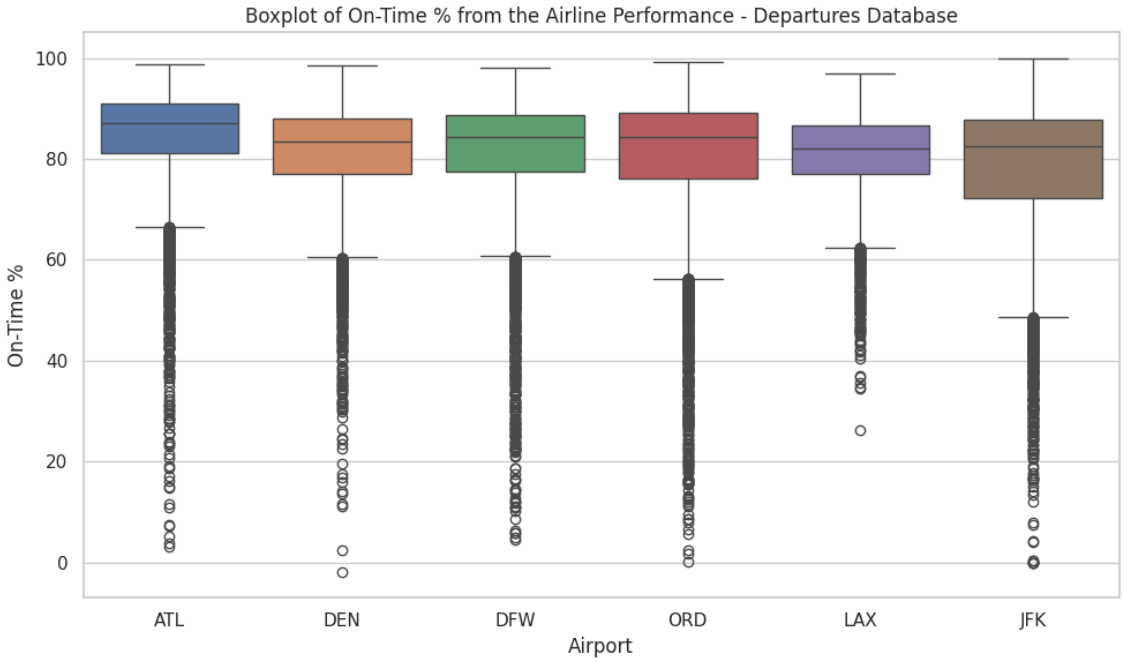


Figure 62 - Boxplot of On-Time % (Departures) by Airport

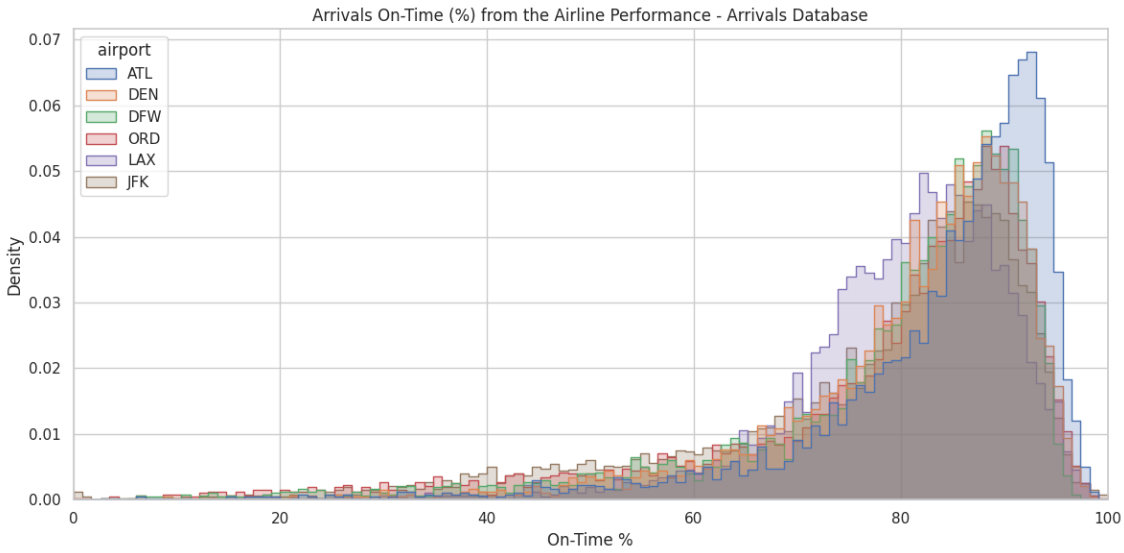


Figure 63 - Distribution of On-Time % (Arrivals) by Airport

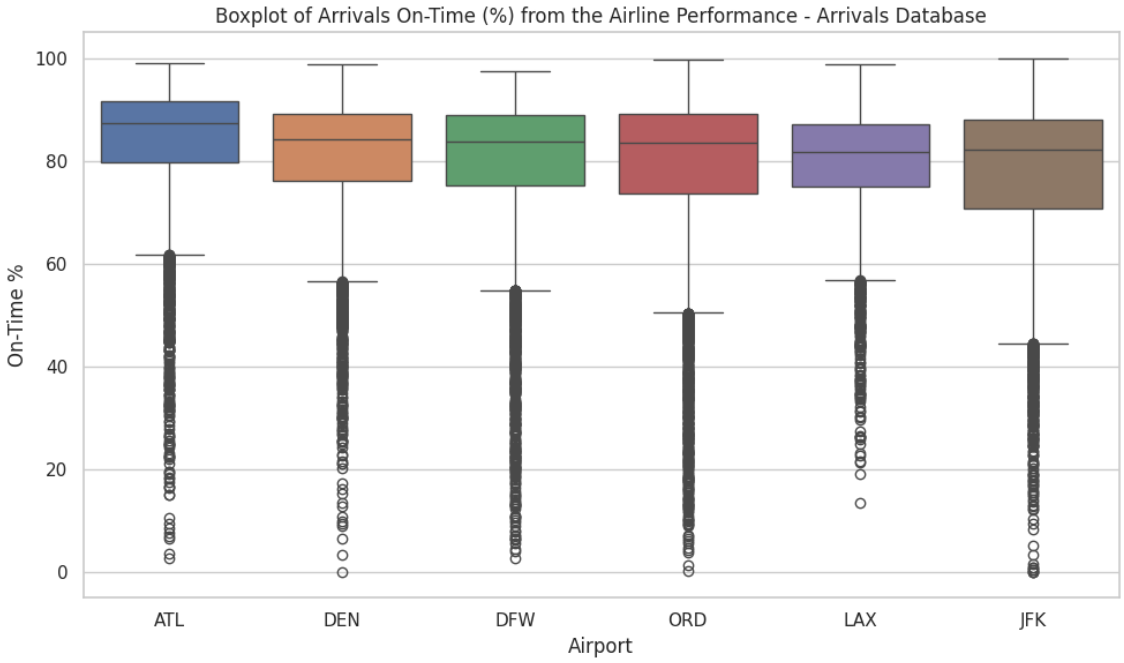


Figure 64 - Boxplot of On-Time % (Arrivals) by Airport

In *Figures* 61 and 63, the distributions of *On-Time* for both departures and arrivals are left-skewed, with most values concentrated between 75% and 95%. The boxplots (*Figures* 62 and 64) confirm that all airports maintain a relatively high level of punctuality, but notable differences emerge in the spread of outliers. ATL and LAX show the tightest clustering around higher values, indicating consistently strong performance. DEN and DFW exhibit wider distributions with a moderate number of low outliers, while ORD and especially JFK display the largest spread, with frequent values dropping well below 60%.

4.1.4. Weather Variables by Airport

4.1.4.1. ATL

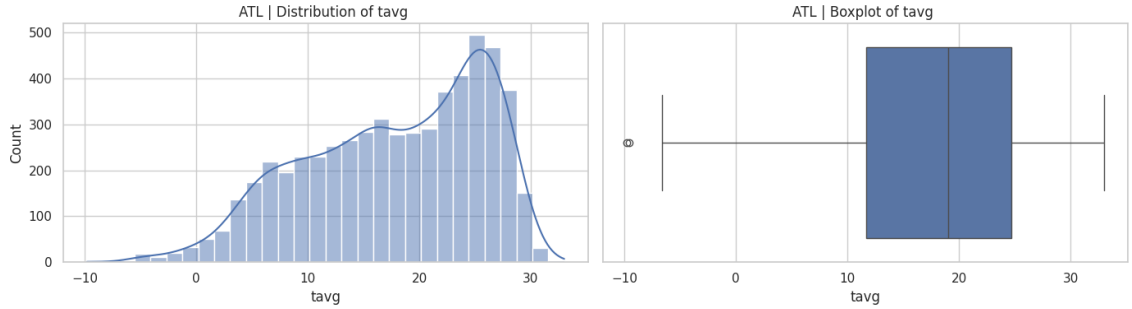


Figure 65 - Distribution and Boxplot of Average Temperature in ATL

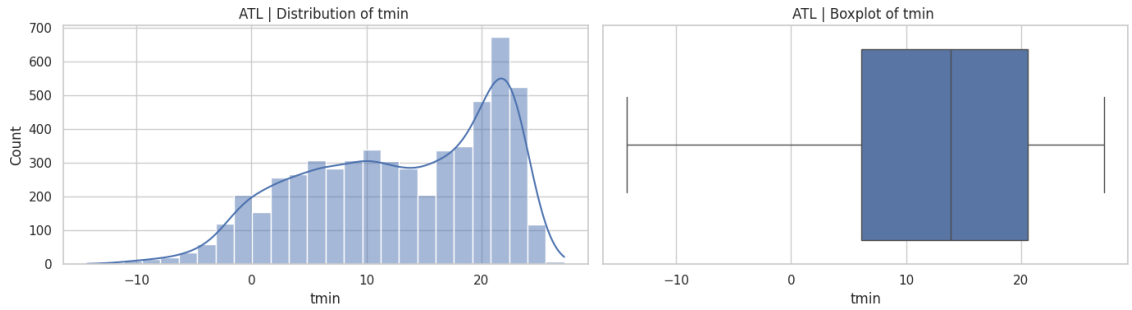


Figure 66 - Distribution and Boxplot of Minimum Temperature in ATL

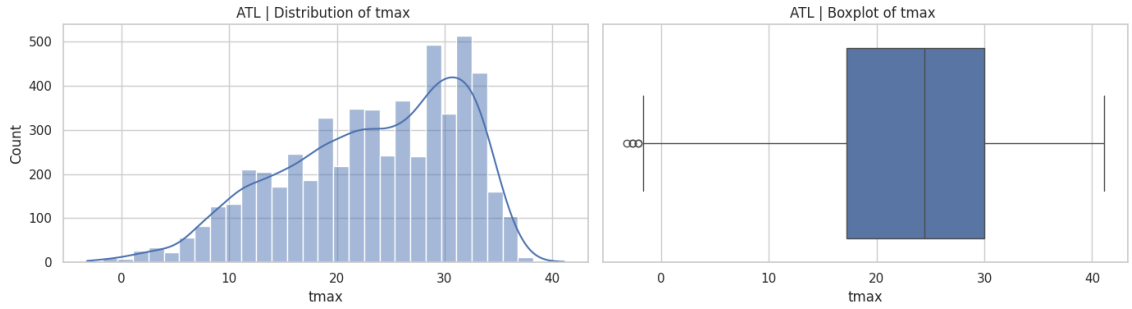


Figure 67 - Distribution and Boxplot of Maximum Temperature in ATL

In *Figures* 65-67, at ATL, the distributions of *average temperature (tavg)*, *minimum temperature (tmin)*, and *maximum temperature (tmax)* all display clear seasonal variation ranges. *tavg* is centred around 15–20°C, with values typically spanning from near 0°C in winter to above 30°C in summer. *tmin* extends below freezing, reaching under –10°C in extreme cases, while *tmax* often surpasses 30°C, occasionally exceeding 40°C. The boxplots confirm this spread, with limited outliers, suggesting that extreme temperature events are relatively infrequent but present.

These distributions highlight the moderate subtropical climate of Atlanta, where average conditions remain temperate, yet the airport is periodically exposed to cold winter spells and hot summer peaks. Such variability is relevant for resilience analysis, as both extreme heat and extreme cold can contribute to operational disruptions, though they occur less frequently than more moderate conditions in ATL.

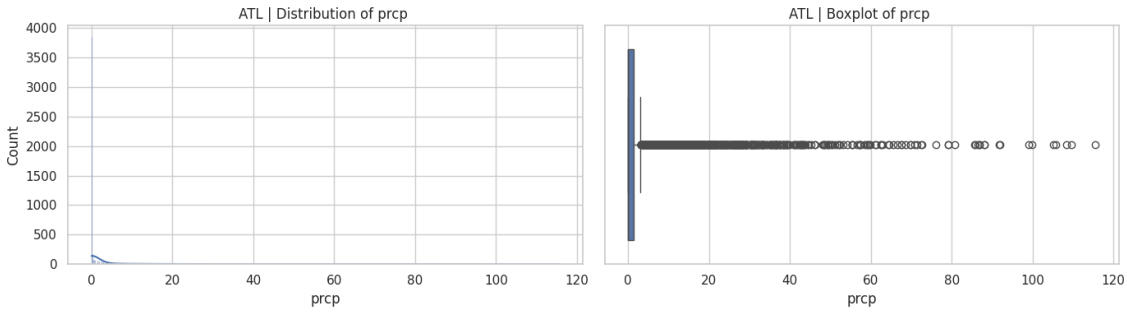


Figure 68 - Distribution and Boxplot of Precipitation in ATL

According to *Figure* 68, at ATL, the distribution of *precipitation (prcp)* is highly skewed, with most days recording very low or no rainfall. The boxplot confirms this, showing a dense concentration near zero and a long tail of outliers exceeding 100 mm.

This pattern aligns with Atlanta's humid subtropical climate, where light or moderate rainfall is common throughout the year, but severe thunderstorms and intense rain events also occur, particularly during the summer months. Such extremes, although infrequent, have an outsized effect on airport operations, testing resilience during peak storm activity.

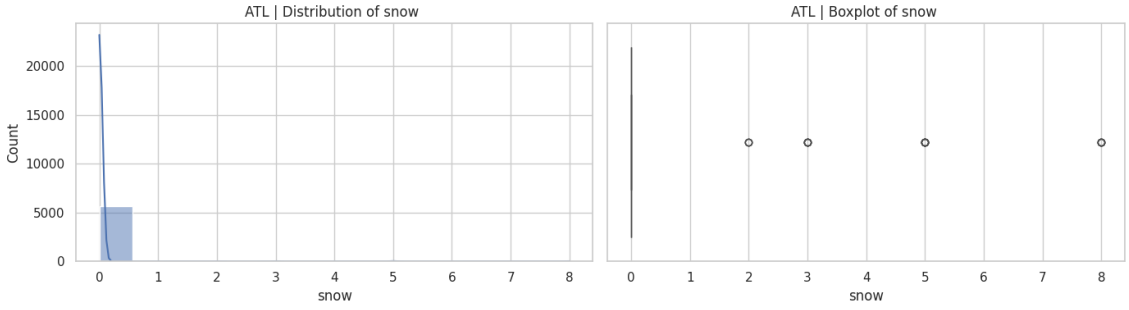


Figure 69 - Distribution and Boxplot of Snow Depth in ATL

According to *Figure* 69, at ATL, the distribution of *snow depth* is highly concentrated at zero, with only a small number of days showing measurable snowfall. The boxplot

reinforces this, as most values are absent of snow, while a few outliers indicate days with accumulations reaching up to 8 cm.

This pattern is consistent with Atlanta's climate, where winters are generally mild and snowfall is infrequent. When snow does occur, it often leads to disproportionate operational challenges because the airport and surrounding city are less prepared for such events.

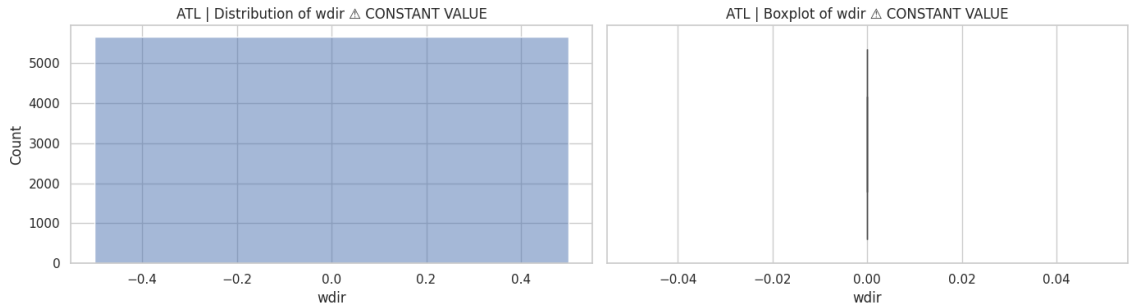


Figure 70 - Distribution and Boxplot of Wind Direction in ATL

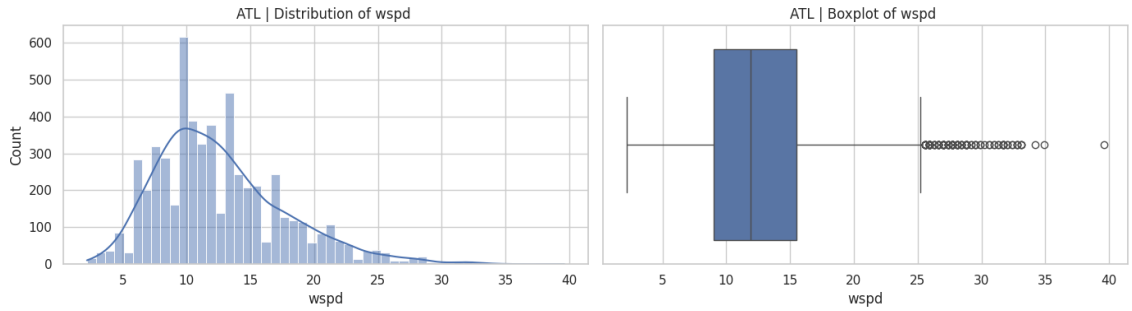


Figure 71 - Distribution and Boxplot of Wind Speed in ATL

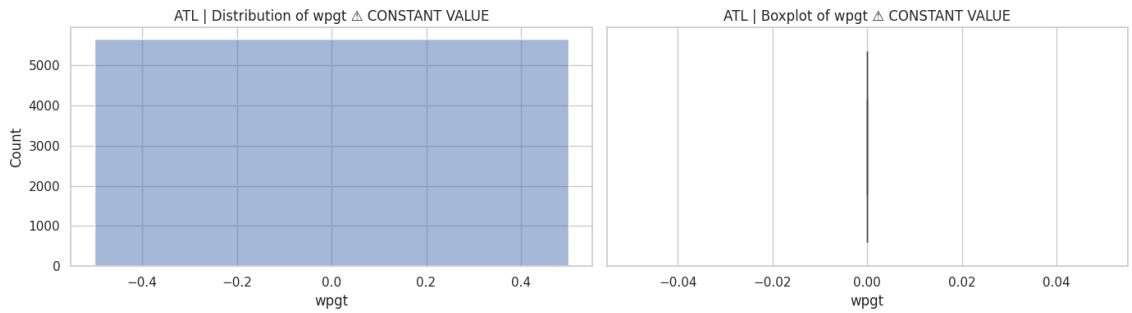


Figure 72 - Distribution and Boxplot of Wind Peak Gust in ATL

At ATL, the variables *wind direction (wdir)* and *wind peak gust (wpgt)* (*Figures* 70 and 72) appear constant in the dataset, suggesting missing or incomplete variation, and therefore cannot provide meaningful insights. By contrast, *wind speed (wspd)* (*Figure*

71) shows a more interpretable distribution, centred between 8–15 km/h, with occasional outliers reaching nearly 40 km/h. The boxplot confirms that most observations fall within a moderate range, while high-wind events are relatively rare but present. This behaviour reflects Atlanta's inland location, where prevailing winds are generally moderate and extreme gusts are less common compared to coastal airports.

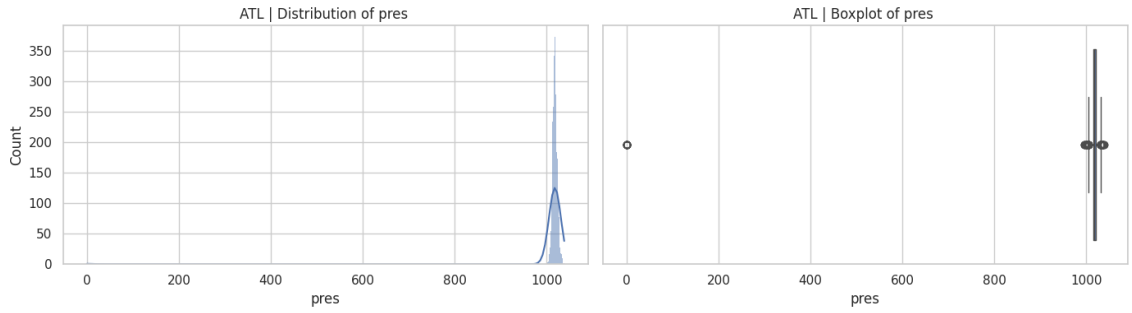


Figure 73 - Distribution and Boxplot of Atmospheric Pressure in ATL

At ATL, the distribution of *atmospheric pressure (pres)* (*Figure* 73) is tightly clustered around 1000 hPa, with only minor variation across the dataset. The boxplot confirms this concentration, with very few outliers, most of which likely represent anomalies or recording errors rather than genuine atmospheric extremes.

This stability is consistent with Atlanta's inland, low-elevation geography, where pressure values tend to remain relatively steady compared with coastal or high-altitude airports.

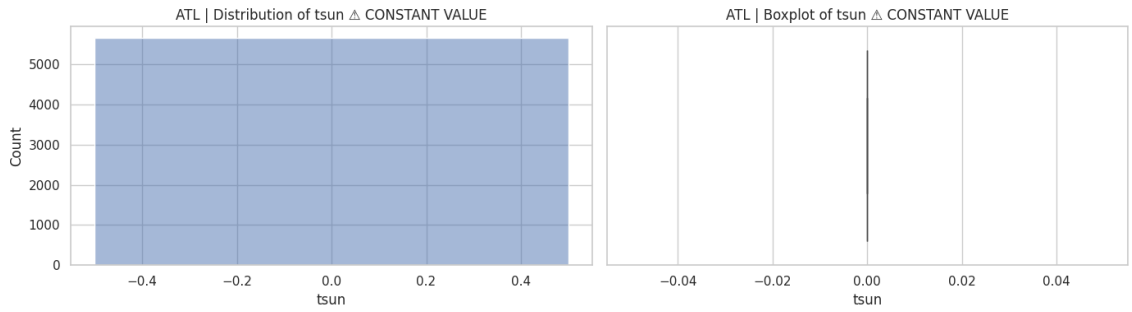


Figure 74 - Distribution and Boxplot of Total Sunshine Time in ATL

At ATL, the *total sunshine duration (tsun)* variable appears constant in the dataset, with no variation across the observed period. The histogram and boxplot (*Figure* 74) both confirm a flat distribution at zero, suggesting that sunshine data was either not recorded or not available for this station.

4.1.4.2. DEN

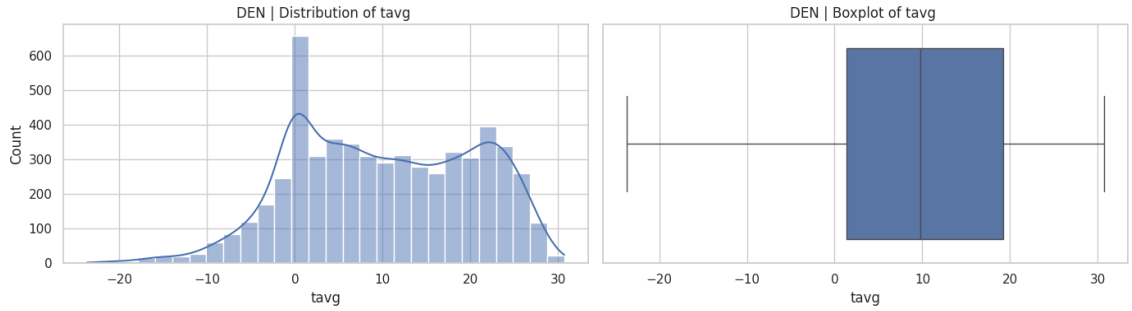


Figure 75 - Distribution and Boxplot of Average Temperature in DEN

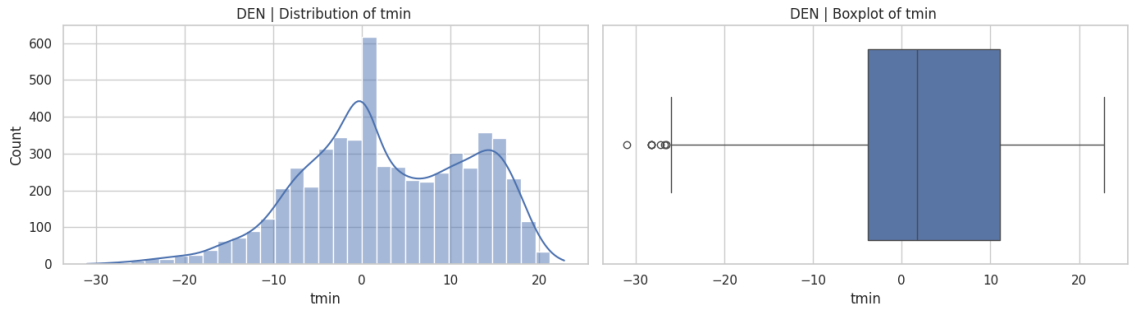


Figure 76 - Distribution and Boxplot of Minimum Temperature in DEN

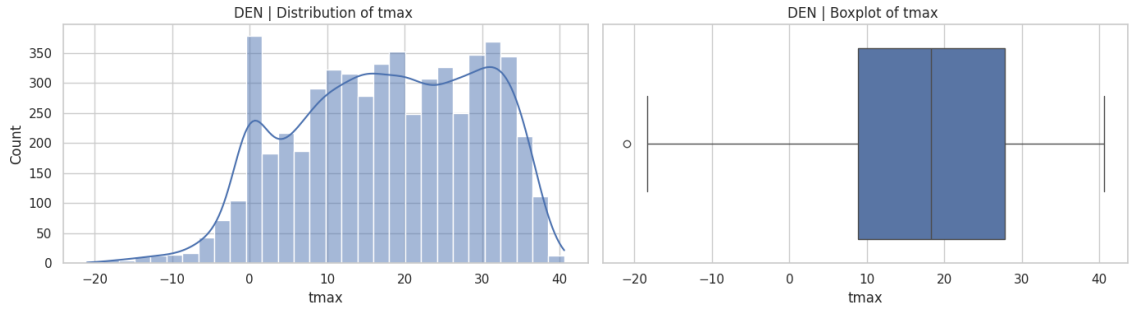


Figure 77 - Distribution and Boxplot of Maximum Temperature in DEN

At DEN, the distributions of *tavg*, *tmin*, and *tmax* (Figures 75-77) show a much wider spread than at ATL. *tavg* typically falls between 0°C and 20°C, with cold-season values extending below –20°C and summer highs surpassing 35°C. The boxplots highlight this broad variability, with numerous outliers at both extremes, particularly in *tmin* where sub-zero values are frequent.

This reflects Denver's continental climate and high elevation, characterised by cold winters, hot summers, and significant diurnal variation. The large range of temperatures suggests that DEN is more regularly exposed to weather extremes than

ATL, requiring greater operational adaptability to both freezing and heat-related disruptions.

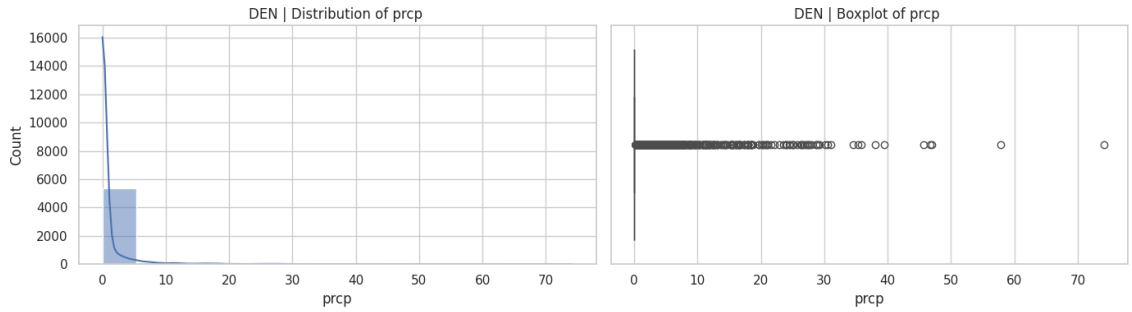


Figure 78 - Distribution and Boxplot of Precipitation in DEN

At DEN, the distribution of *prcp* (*Figure* 78) is highly skewed, with most days showing little or no measurable rainfall. The boxplot confirms this clustering near zero, with occasional outliers reaching above 40 mm.

This behaviour aligns with Denver's semi-arid continental climate, which is characterised by relatively low annual precipitation but punctuated by short, intense storms, especially during late spring and summer. These sporadic heavy downpours can create short-term but significant disruptions for the airport.

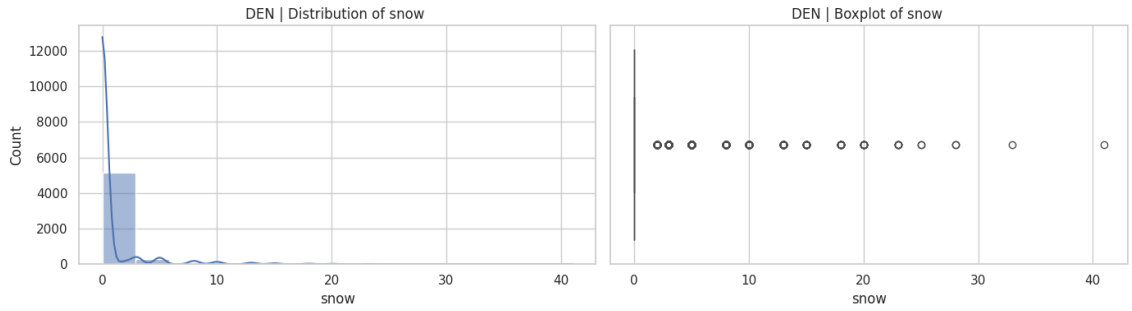


Figure 79 - Distribution and Boxplot of Snow Depth in DEN

At DEN, the distribution of *snow depth (Figure 79)* is dominated by zeros, reflecting many snow-free days, but with a noticeable tail of observations showing significant accumulations. The boxplot highlights a series of outliers, with snow depths occasionally surpassing 20 cm.

This is consistent with Denver's climate and high elevation, which bring cold winters with frequent snow events. Unlike ATL, where snowfall is rare, DEN is routinely

exposed to snowstorms that can heavily impact airport operations, from runway clearance requirements to flight cancellations.

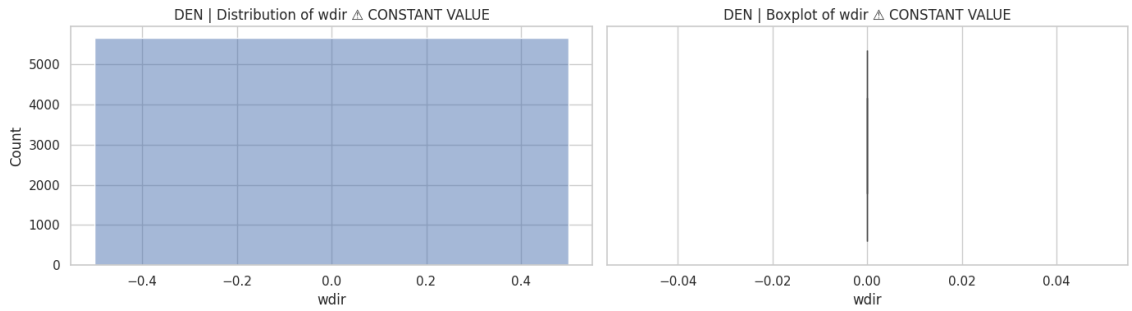


Figure 80 - Distribution and Boxplot of Wind Direction in DEN

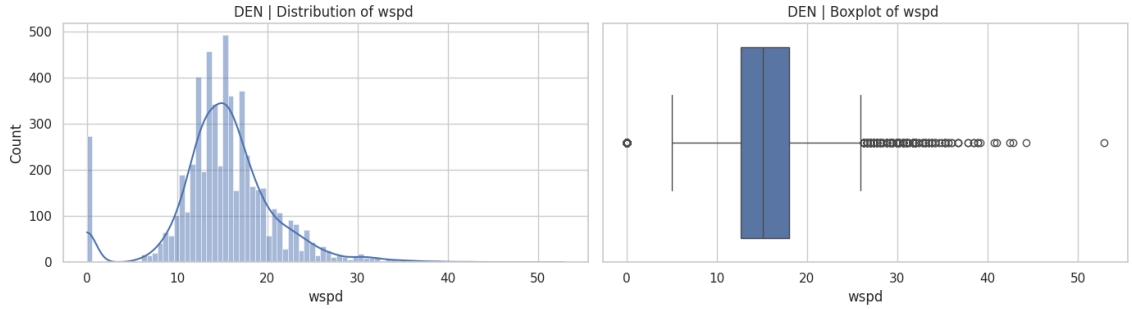


Figure 81 - Distribution and Boxplot of Wind Speed in DEN

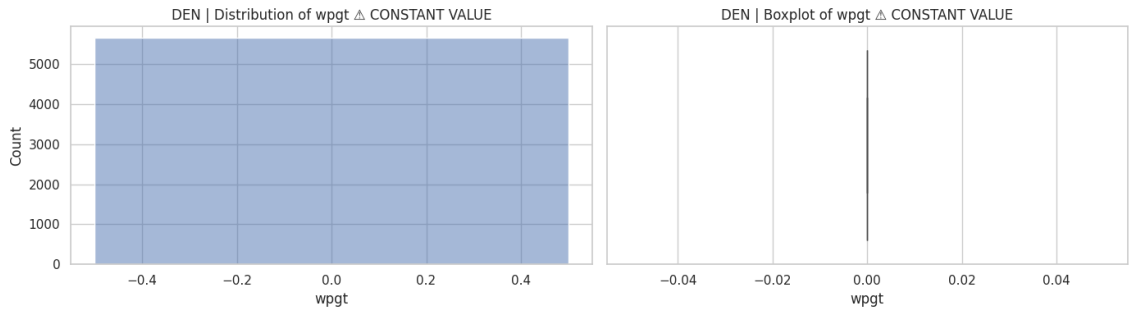


Figure 82 - Distribution and Boxplot of Wind Peak Gust in DEN

At DEN, both *wdir* and *wpgt* (*Figures* 80 and 82) once again appear constant in the dataset, suggesting missing or unrecorded variation, which limits their interpretive value. Meanwhile, *wspd* (*Figure* 81) shows a meaningful distribution, centred between 10–20 km/h, with outliers exceeding 50 km/h.

This behaviour is consistent with Denver's geographical location on the high plains near the Rocky Mountains, where the city is exposed to strong and variable winds due to local topography.

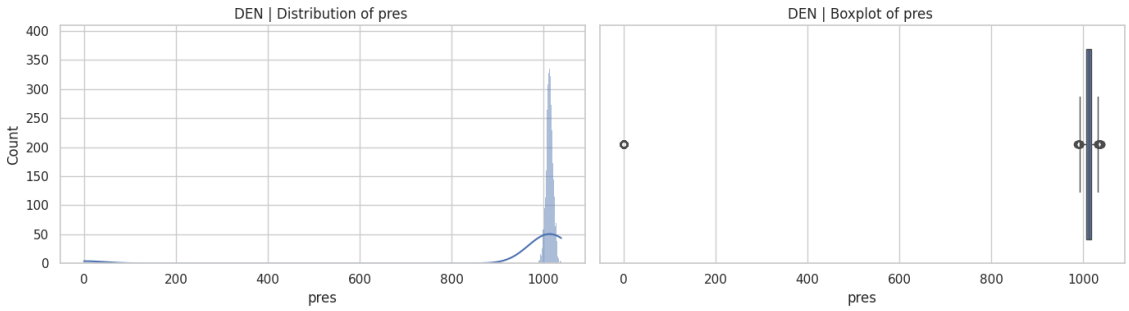


Figure 83 - Distribution and Boxplot of Atmospheric Pressure in DEN

At DEN, the distribution of *pres* (*Figure* 83) is tightly clustered around 1000 hPa, with very limited variability. The boxplot confirms this narrow range, with only a few scattered outliers, some of which may represent measurement errors rather than genuine atmospheric extremes.

This pattern is expected given Denver's high-elevation location (over 1,600 metres above sea level), where baseline pressure is naturally lower than at sea level but remains relatively stable over time.

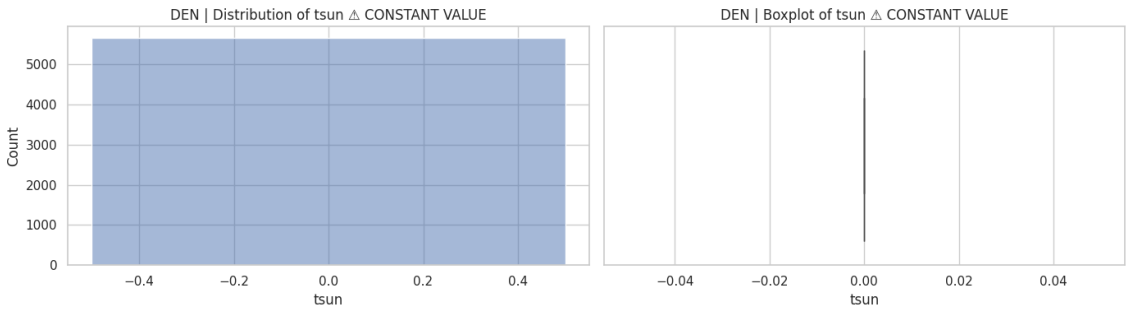


Figure 84 - Distribution and Boxplot of Total Sunshine Duration in DEN

At DEN, tsun is once again constant across the dataset, with no variation recorded. The histogram and boxplot (*Figure* 84) both confirm this flat distribution, suggesting either missing data or a lack of reporting for this indicator.

4.1.4.3. DFW

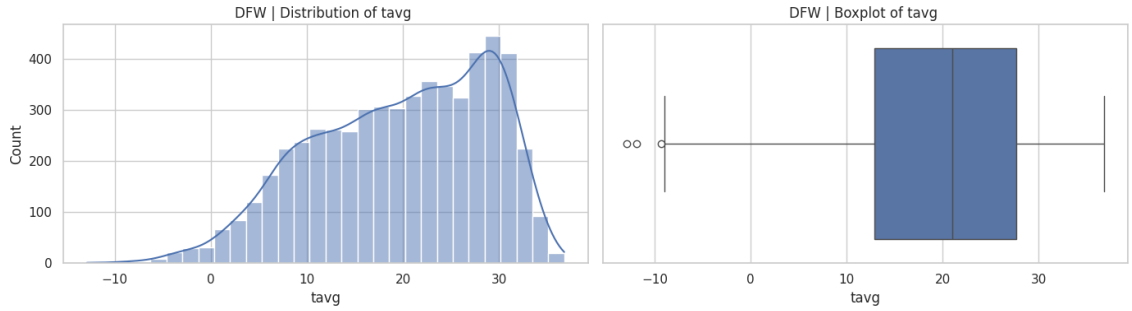


Figure 85 - Distribution and Boxplot of Average Temperature in DFW

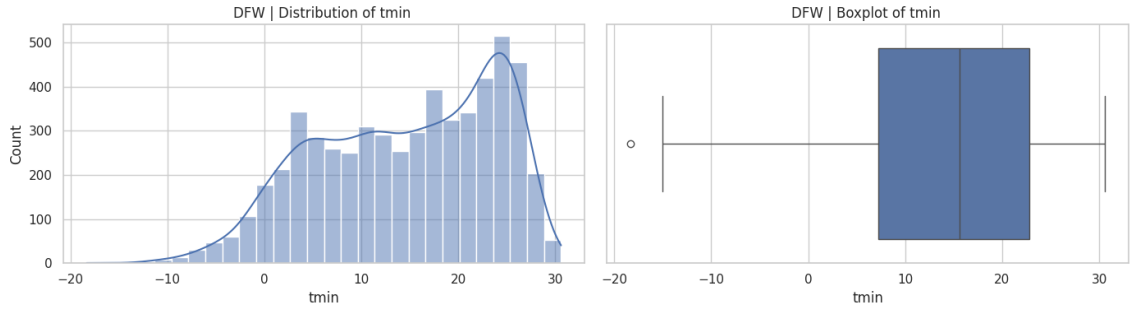


Figure 86 - Distribution and Boxplot of Minimum Temperature in DFW

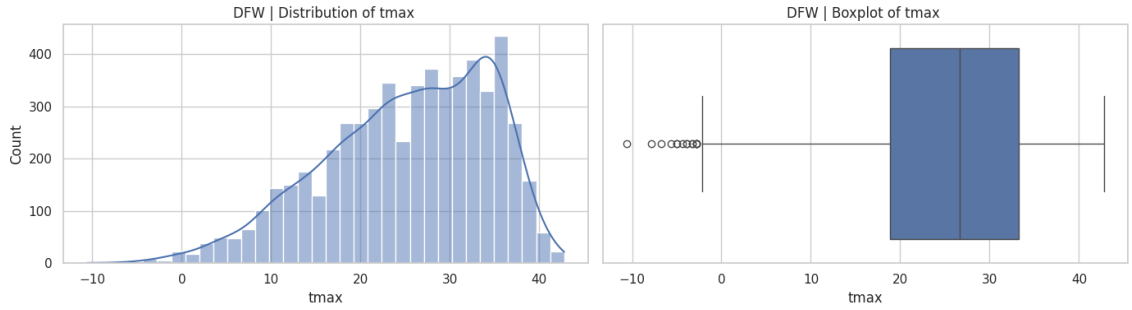


Figure 87 - Distribution and Boxplot of Maximum Temperature in DFW

At DFW, the distributions of *tavg*, *tmin*, and *tmax* (*Figures* 85-87) reveal a broad seasonal range. *tavg* typically span from around 5°C in winter to above 30°C in summer. *tmin* occasionally fall below freezing, reaching close to –15°C, while *tmax* frequently surpass 35°C, with extremes nearing 40°C. The boxplots confirm this variability, showing a consistent spread with limited outliers.

This behaviour reflects Dallas–Fort Worth's humid subtropical climate, characterised by hot summers and mild but occasionally cold winters. The marked summer heat is particularly relevant for airport resilience, as prolonged high temperatures can stress aircraft performance and infrastructure.

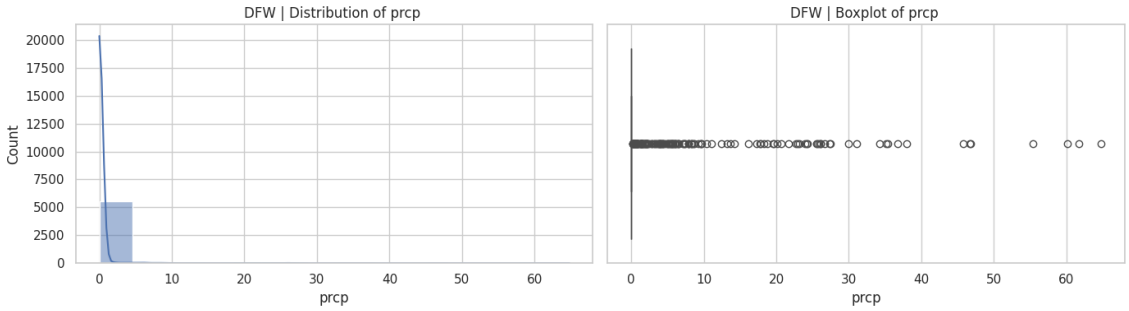


Figure 88 - Distribution and Boxplot of Precipitation in DFW

At DFW, the distribution of *prcp* (*Figure* 88) is highly skewed, with most days registering little or no rainfall. The boxplot confirms this, showing a strong clustering near zero and a series of outliers extending above 60 mm, representing days of intense rainfall.

This pattern is consistent with the region's humid subtropical climate, where precipitation is frequent but often occurs in bursts, particularly during spring and summer thunderstorms.

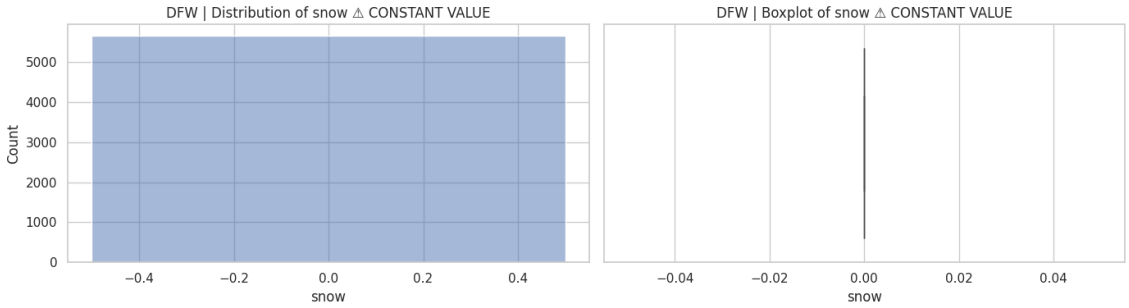


Figure 89 - Distribution and Boxplot of Snow Depth in DFW

At DFW, the *snow depth* variable is constant at zero, with no variation recorded across the dataset. The histogram and boxplot (*Figure* 89) both confirm this flat distribution, indicating that snowfall was not observed during the recorded period or that the data was not captured.

This aligns with the airport's climatic context: Dallas–Fort Worth has a humid subtropical climate, where snowfall is extremely rare.

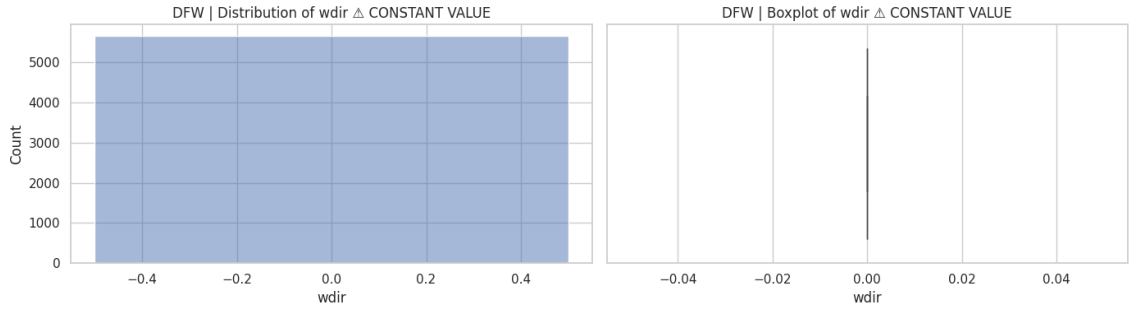


Figure 90 - Distribution and Boxplot of Wind Direction in DFW

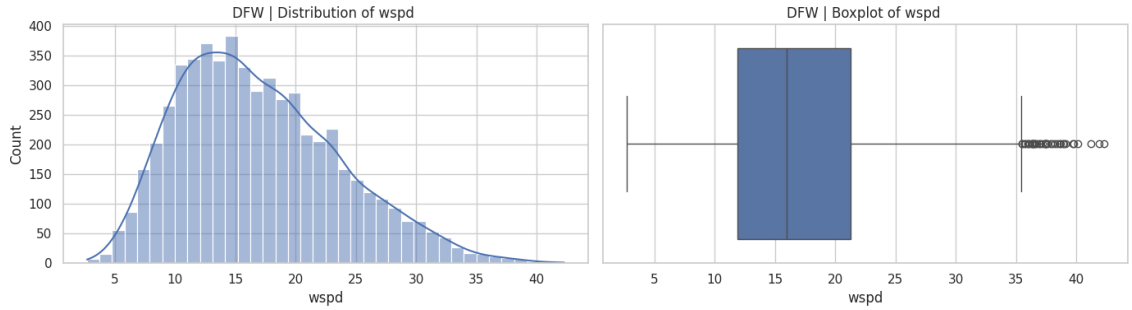


Figure 91 - Distribution and Boxplot of Wind Speed in DFW

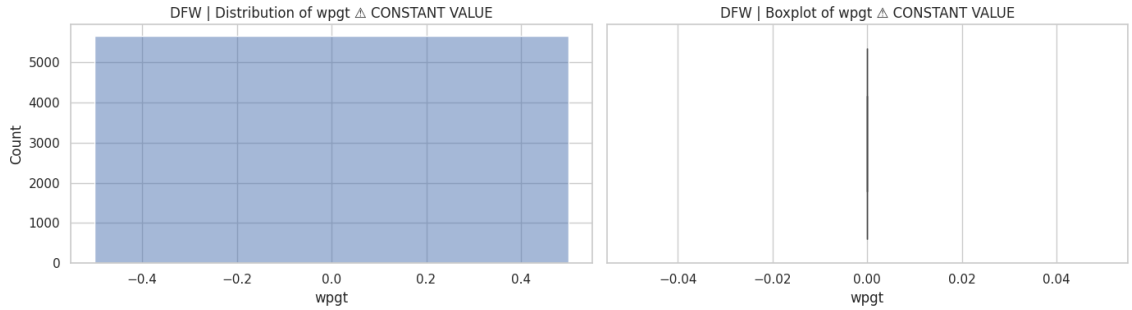


Figure 92 - Distribution and Boxplot of Wind Peak Gust in DFW

At DFW, *wdir* and *wpgt* (*Figures* 90 and 92) continue constant across the dataset, providing no usable variation for analysis. *wspd* (*Figure* 91) shows a meaningful distribution, with most values ranging between 10–20 km/h and occasional outliers surpassing 40 km/h.

This reflects the regional climate of north-central Texas, where prevailing winds are generally steady but can strengthen during thunderstorms and seasonal weather fronts.

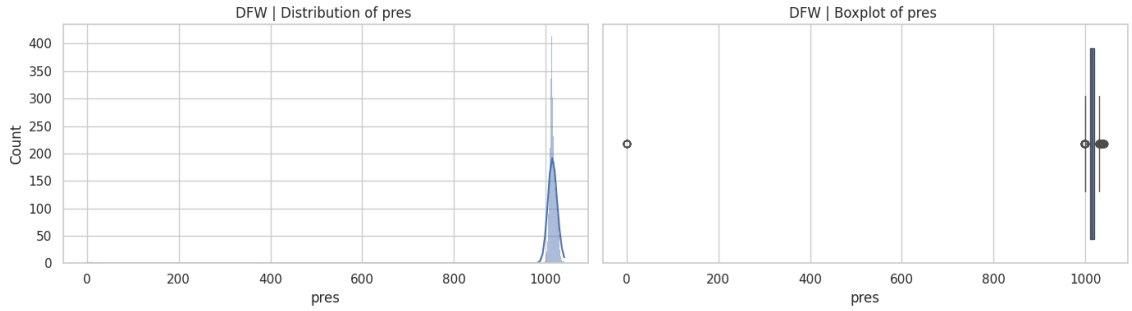


Figure 93 - Distribution and Boxplot of Atmospheric Pressure in DFW

At DFW, the distribution of *pres* (*Figure* 93) is once again tightly centred around 1000 hPa, with only minimal variability across the dataset. The boxplot reflects this stability, showing a narrow interquartile range and a few scattered outliers at 0, just as observed in the last airports.

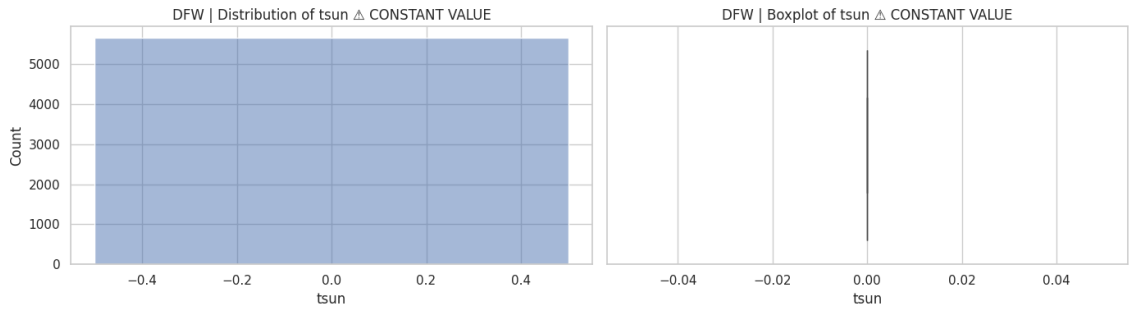


Figure 94 - Distribution and Boxplot of Total Sunshine Duration in DFW

At DFW, *tsun* is again constant at zero, showing no variation across the dataset. Both the histogram and boxplot (*Figure* 94) confirm this flat distribution, suggesting either missing data or an error.

4.1.4.4. JFK

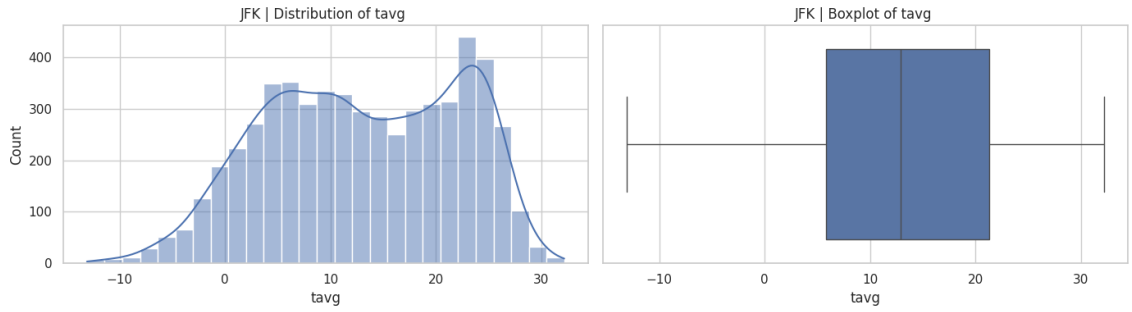


Figure 95 - Distribution and Boxplot of Average Temperature in JFK

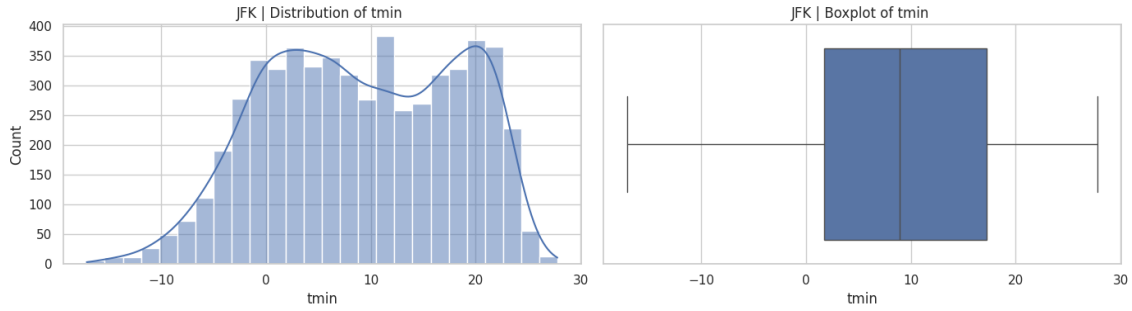


Figure 96 - Distribution and Boxplot of Minimum Temperature in JFK

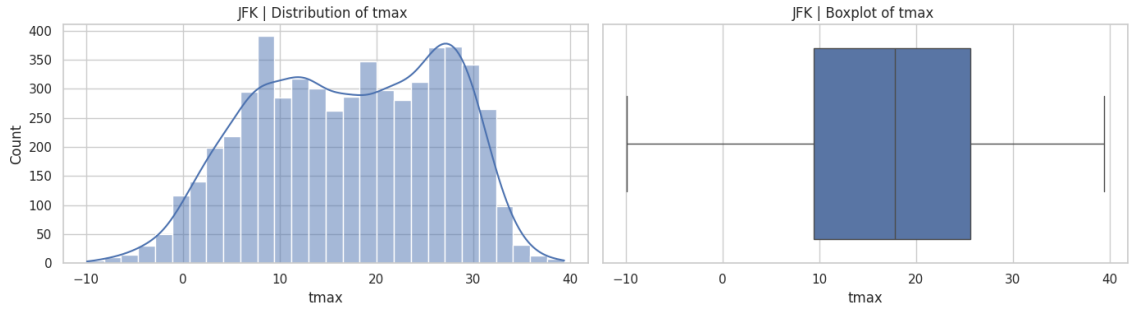


Figure 97 - Distribution and Boxplot of Minimum Temperature in JFK

At JFK, the distributions of *tavg*, *tmin*, and *tmax* (*Figures* 95-97) display a wide seasonal range, with averages clustering between 5°C and 20°C, minimum values occasionally dropping below –10°C, and maximum values often reaching above 30°C.

This reflects New York City's humid subtropical climate with strong seasonal contrasts—cold winters, warm springs, and hot, humid summers. The regular exposure to both freezing and heat extremes means that JFK's operations must remain resilient to a variety of temperature-related disruptions.

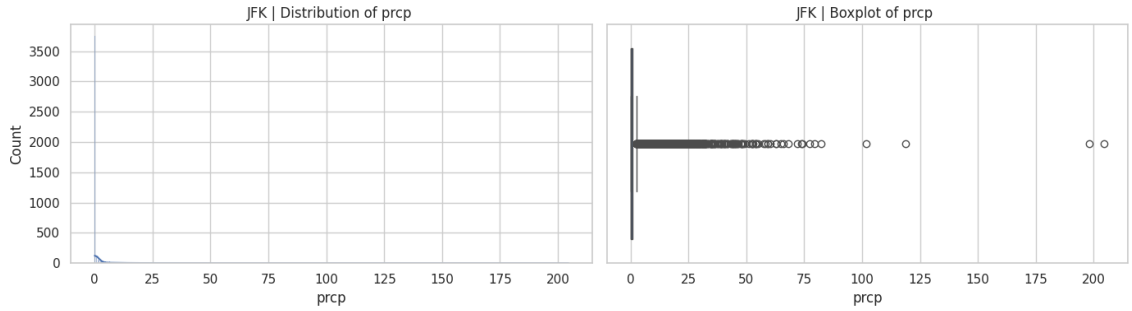


Figure 98 - Distribution and Boxplot of Precipitation in JFK

At JFK, the distribution of *prcp* (*Figure* 98) is highly skewed toward zero, with most days experiencing little or no rainfall. The boxplot highlights many outliers, with occasional extreme events surpassing 100 mm and even reaching close to 200 mm.

This aligns with the climate of New York City, where rainfall is distributed evenly throughout the year but can occur in heavy episodes.

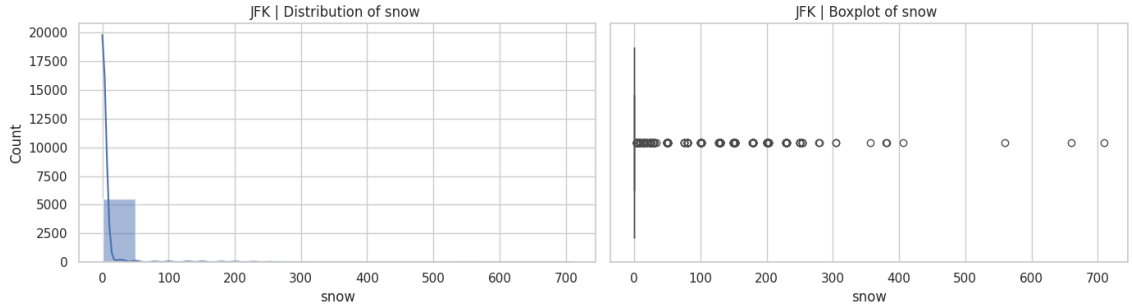


Figure 99 - Distribution and Boxplot of Snow Depth in JFK

At JFK, the distribution of *snow depth* (*Figure* 99) is extremely skewed, with most observations clustered at zero, indicating snow-free conditions. However, the boxplot reveals numerous outliers, with some extreme cases surpassing 500 mm and even nearing 700 mm.

This reflects New York's continental influence, where winters frequently bring snow, though the intensity varies considerably by year. While average snowfall is moderate compared to more northern cities, heavy snowstorms can occur and severely disrupt airport operations.

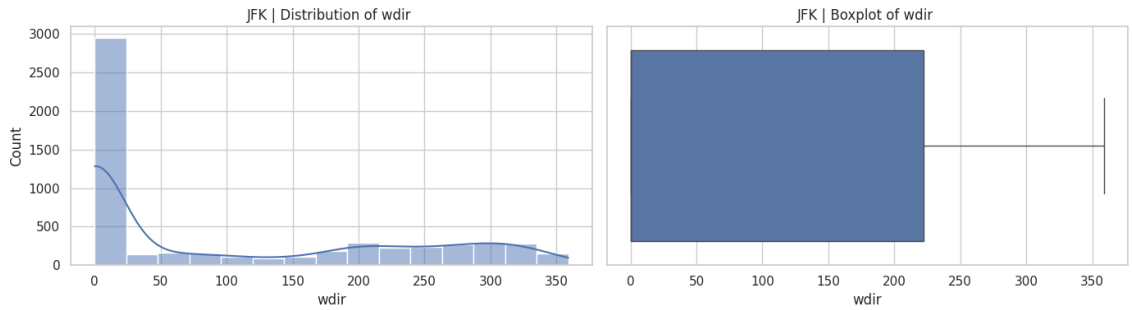


Figure 100 - Distribution and Boxplot of Wind Direction in JFK

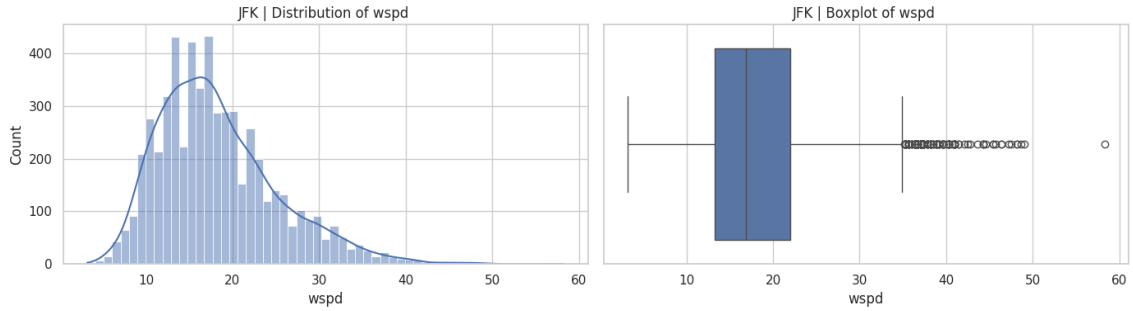


Figure 101 - Distribution and Boxplot of Wind Speed in JFK

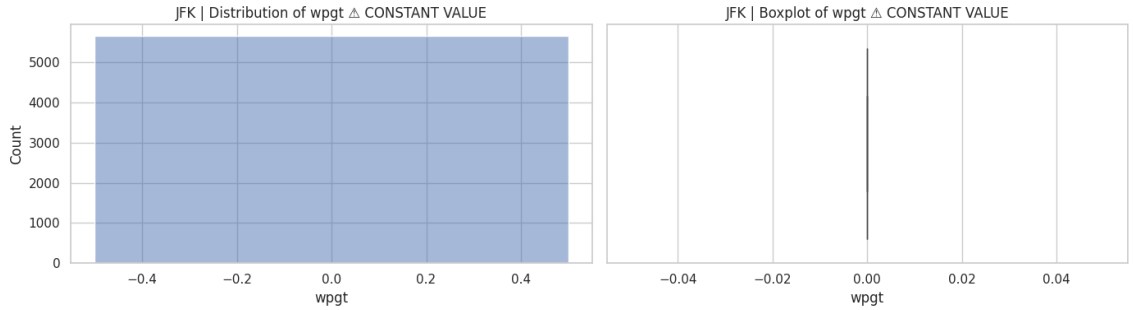


Figure 102 - Distribution and Boxplot of Wind Peak Gust in JFK

At JFK, *wdir* (*Figure* 100) displays a uniform spread across all angles, though some clustering at lower values may suggest seasonal prevalence of certain directions, or possible misreadings, as in the airports before.

The distribution of *wspd* (Figure 101) shows a concentration between 10–20 km/h, with a long right tail indicating occasional stronger winds reaching up to 60 km/h. These conditions are consistent with JFK's coastal location.

However, *wpgt* (*Figure* 102) is constant in this dataset, limiting its analytical value.

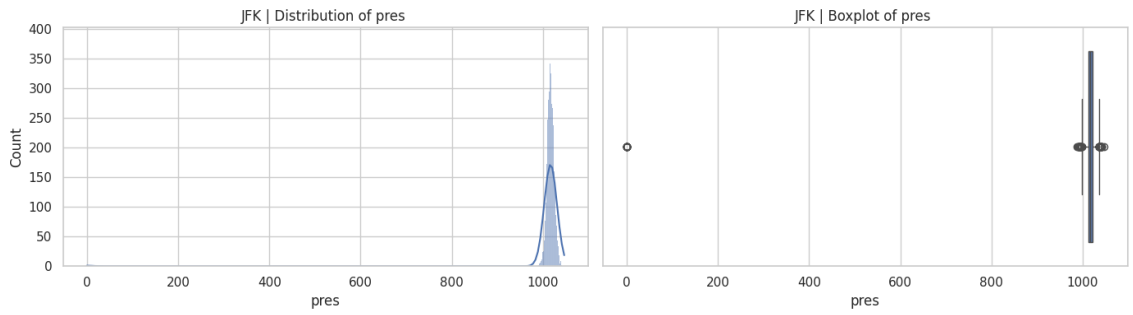


Figure 103 - Distribution and Boxplot of Atmospheric Pressure in JFK

At JFK, *pres* (*Figure* 103) is like the other airports, centering around 1000 hPa, with extreme cases at 0 most likely representing scanning errors.

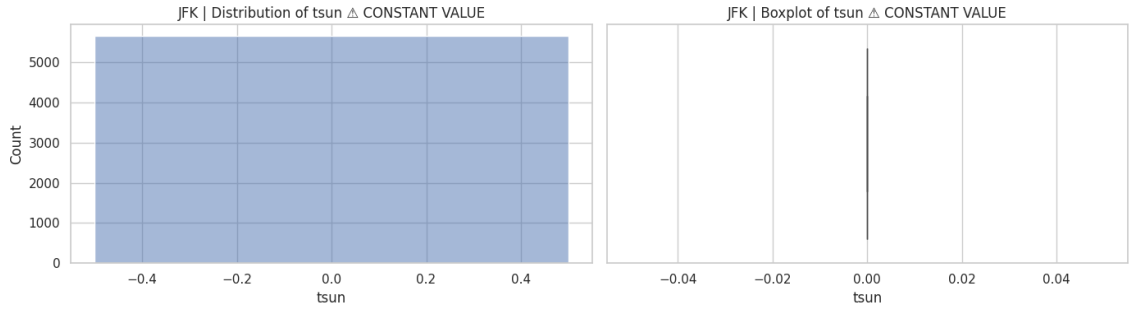


Figure 104 - Distribution and Boxplot of Total Sunshine Duration in JFK

Just as before *tsun* is constant at zero, showing no variation across the dataset. Both the histogram and boxplot confirm this flat distribution.

4.1.4.5. LAX

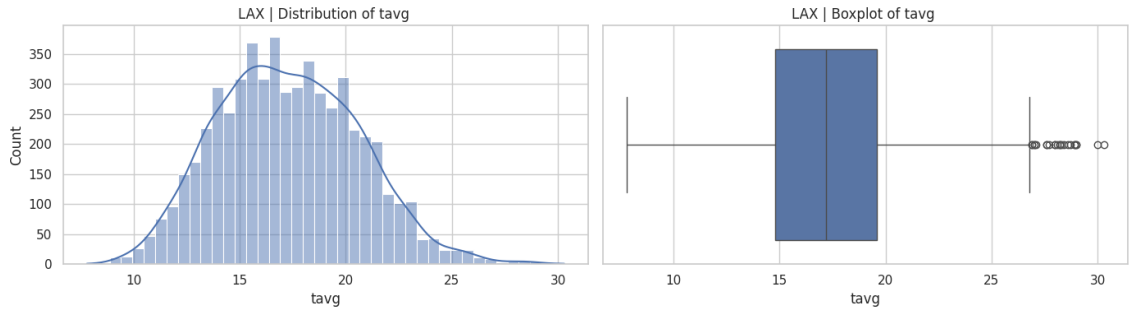


Figure 105 - Distribution and Boxplot of Average Temperature in LAX

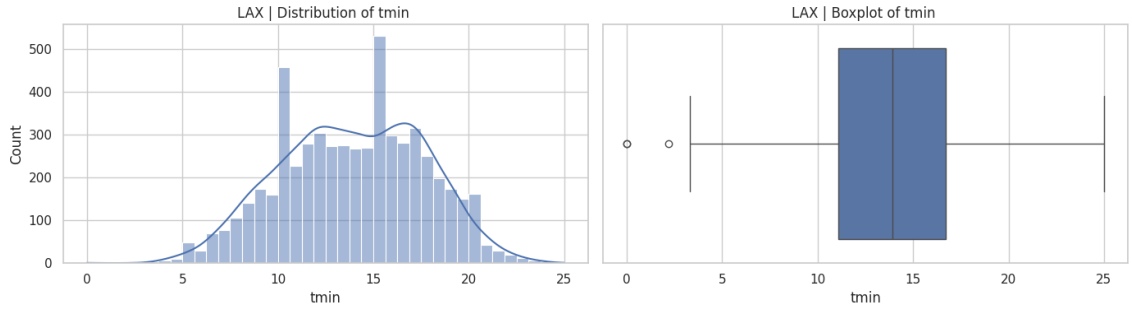


Figure 106 - Distribution and Boxplot of Minimum Temperature in LAX

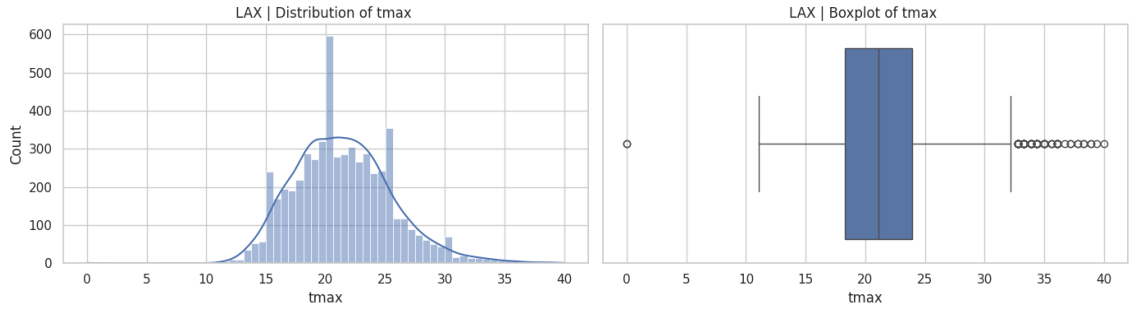


Figure 107 - Distribution and Boxplot of Maximum Temperature in LAX

LAX's temperature variables (tavg, tmin, and tmax) (*Figures* 105-107) show a narrower range compared to airports such as DEN or JFK, reflecting the milder climate of Southern California. *tavg* clusters between 15°C and 20°C, while *tmin* and *tmax* also show limited variation.

These results align with Los Angeles' Mediterranean climate, characterised by mild winters and warm, dry summers. The relatively low variability in temperatures indicates more predictable operational conditions.

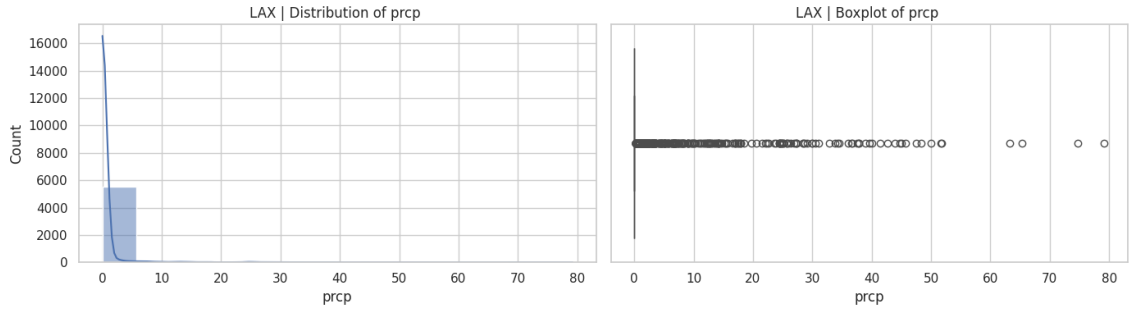


Figure 108 - Distribution and Boxplot of Precipitation in LAX

For LAX, prcp is heavily concentrated around very low values, with the histogram (*Figure* 108) showing a strong peak near zero and a rapid decline as amounts increase.

This distribution is consistent with Los Angeles' semi-arid to Mediterranean climate, where dry conditions dominate for most of the year, punctuated by sporadic but sometimes heavy rain events, often associated with winter storms or El Niño years.

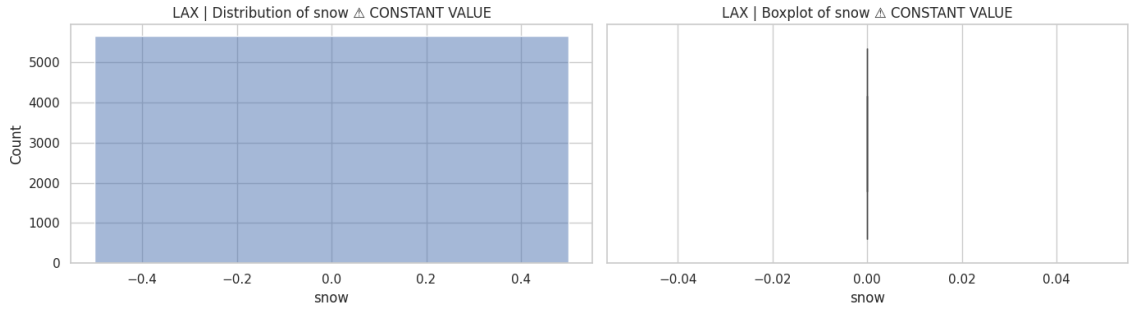


Figure 109 - Distribution and Boxplot of Snow Depth in LAX

Looking at the snow distribution for LAX (Figure 109), it is constant at zero, with no variation at all. Both the histogram and the boxplot confirm that snowfall does not occur at Los Angeles International Airport, which is fully expected, given Los Angeles' climate.

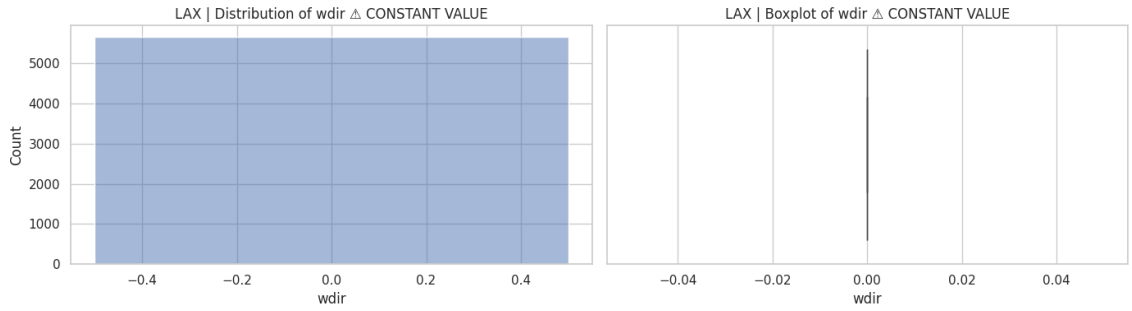


Figure 110 - Distribution and Boxplot of Wind Direction in LAX

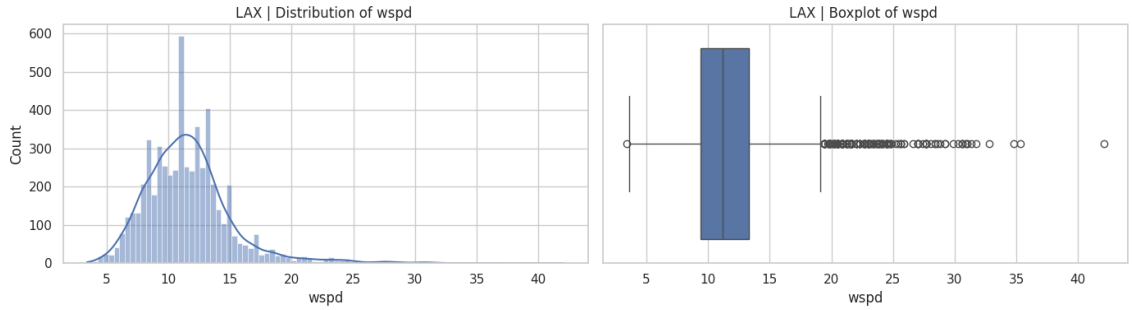


Figure 111 - Distribution and Boxplot of Wind Speed in LAX

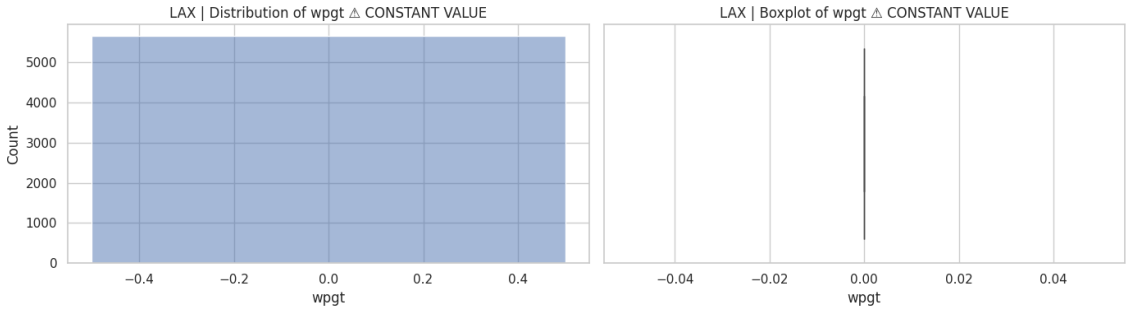


Figure 112 - Distribution and Boxplot of Wind Peak Gust in LAX

At LAX, the plots for wdir and wpgt (*Figures* 110 and 112) show constant values across the dataset, just like in ATL, DEN and DFW. This indicates issues in data collection or reporting for these variables, making them unusable for meaningful analysis. Meanwhile, *wspd* (*Figure* 111) demonstrates a clear distribution, with most values clustering around 10 to 15 km/h.

This pattern is consistent with the typical weather at LAX, which is located near the Pacific Ocean and experiences steady, mild coastal winds. However, the lack of usable direction and gust data limits a more detailed assessment of wind variability at this location.

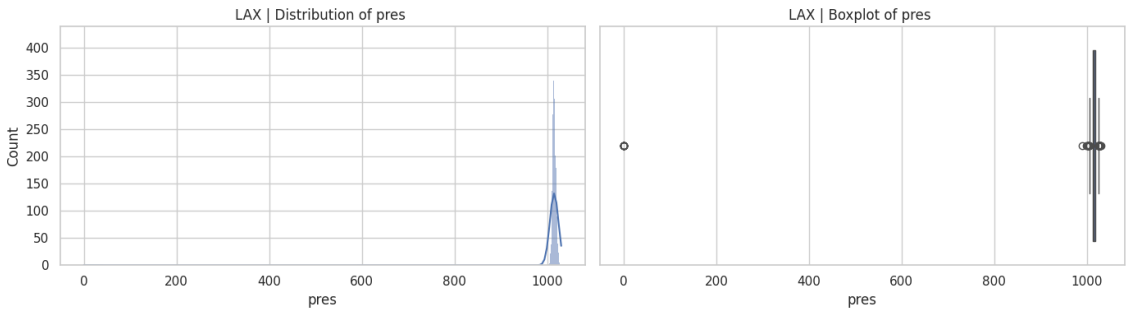


Figure 113 - Distribution and Boxplot of Atmospheric Pressure in LAX

The plots for pres at LAX show a highly concentrated distribution around 1010–1020 hPa, with very little variability aside from a few outliers. These outliers are equal to 0, and most likely represent capture errors or misreadings.

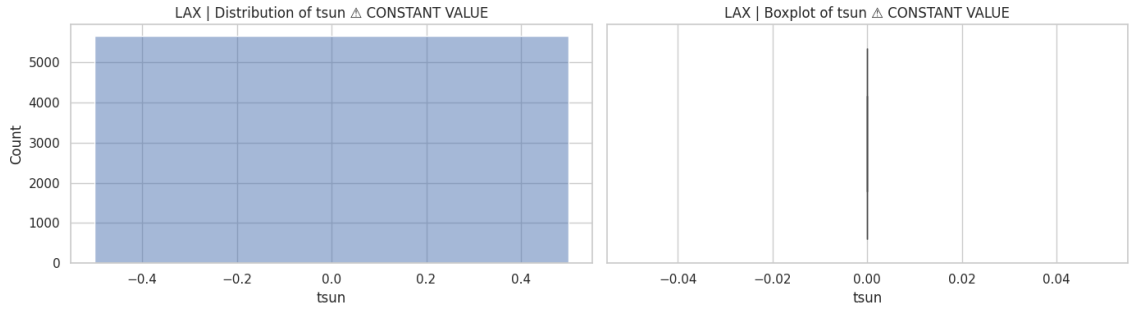


Figure 114 - Distribution and Boxplot of Total Sunshine Duration in LAX

The plots for tsun at LAX (Figure 114) indicate that this variable is recorded as a constant value (zero) throughout the dataset, with no variability, reflecting that no actual sunshine duration data was collected for this station.

4.1.4.6. ORD

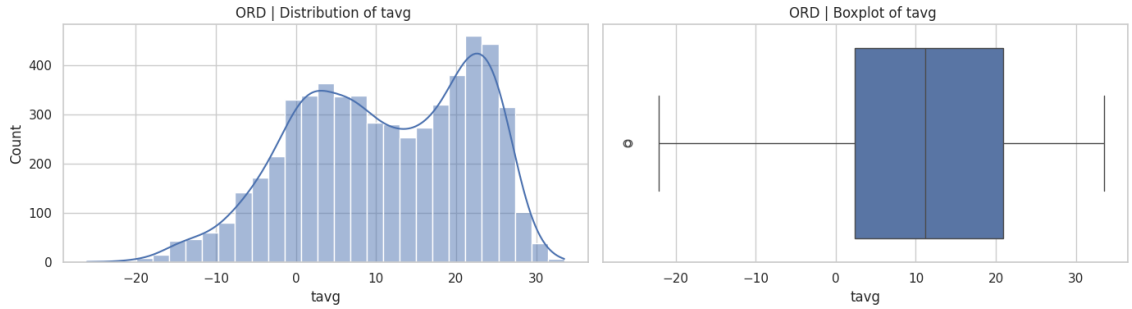


Figure 115 - Distribution and Boxplot of Average Temperature in ORD

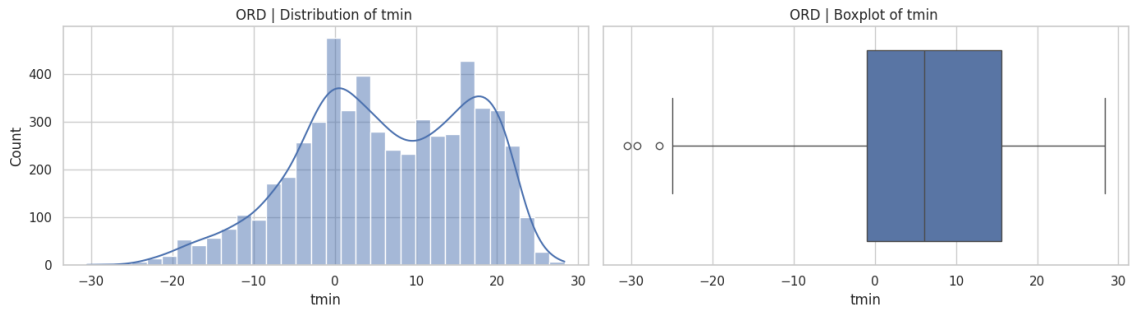


Figure 116 - Distribution and Boxplot of Minimum Temperature in ORD

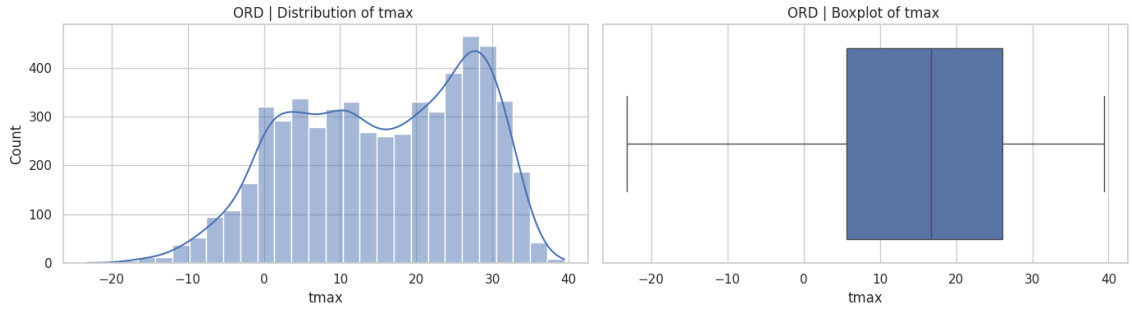


Figure 117 - Distribution and Boxplot of Maximum Temperature in ORD

In ORD, *tavg* spans from well below freezing to above 30°C, with two distinct peaks in the density curve reflecting the contrast between cold winters and warm summers. *tmin* frequently drop below zero, with outliers showing occasional extreme cold, while *tmax* often reach the upper 20s to 30s during summer, with some extremes approaching 40°C. The boxplots (*Figures* 115-117) confirm this wide variability, showing broader interquartile ranges compared with airports like LAX.

Such variability is consistent with Chicago's Midwest location, where winters are cold and snowy, and summers are hot and humid.

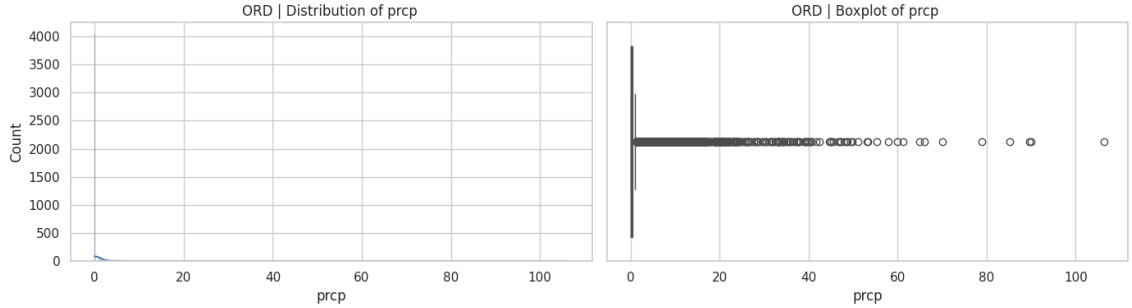


Figure 118 - Distribution and Boxplot of Precipitation in ORD

The *prcp* distribution at ORD (*Figure* 118) shows that most days record very low or no rainfall, with the histogram heavily concentrated near zero. However, the boxplot reveals numerous outliers extending to values above 100 mm, indicating occasional heavy rain events.

This precipitation pattern aligns with Chicago's climate, where rainfall is spread evenly throughout the year but occasionally intensified by severe thunderstorms during summer and early autumn, or heavy rainfalls associated with frontal systems in spring.

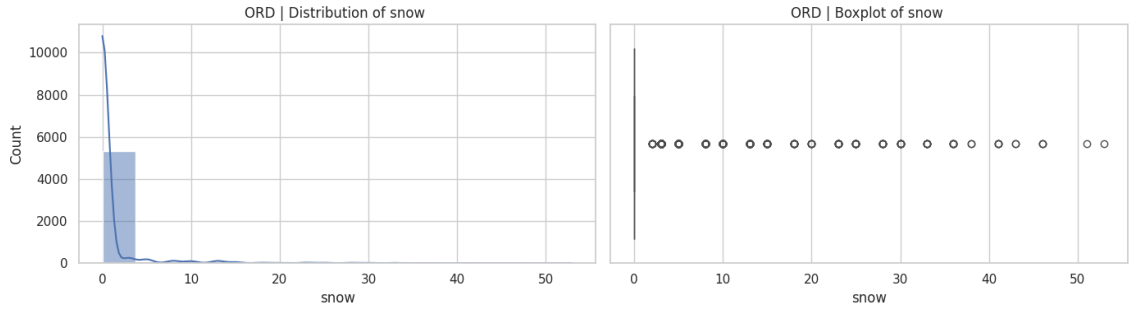


Figure 119 - Distribution and Boxplot of Snow Depth in ORD

The snow distribution at ORD (*Figure* 119) is highly skewed, with most days showing no snowfall, as expected. The histogram shows occasional events with snowfall amounts exceeding 10 mm, and the boxplot highlights many outliers, with extreme cases surpassing 50 mm. This indicates that while snowfall is relatively infrequent, significant snowstorms do occur and can have a severe operational impact.

This pattern reflects Chicago's well-known winter climate, characterised by cold temperatures and frequent snow events.

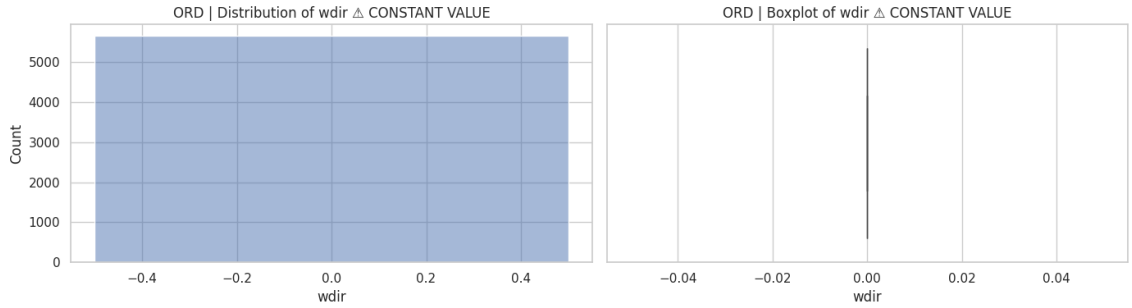


Figure 120 - Distribution and Boxplot of Wind Direction in ORD

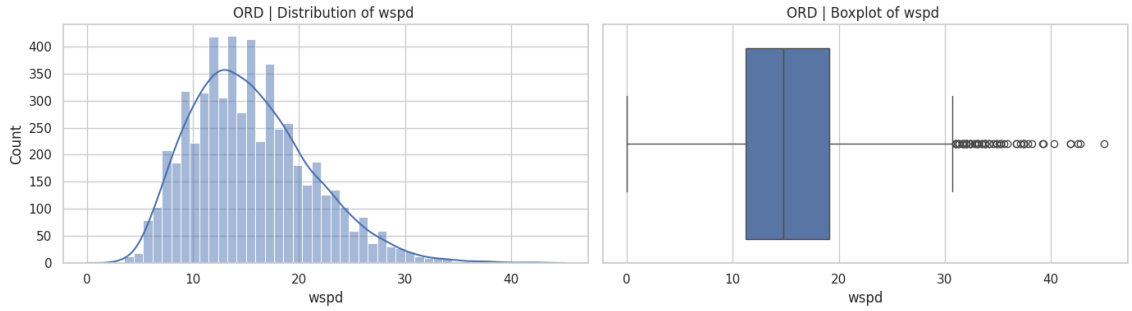


Figure 121 - Distribution and Boxplot Wind Speed in ORD

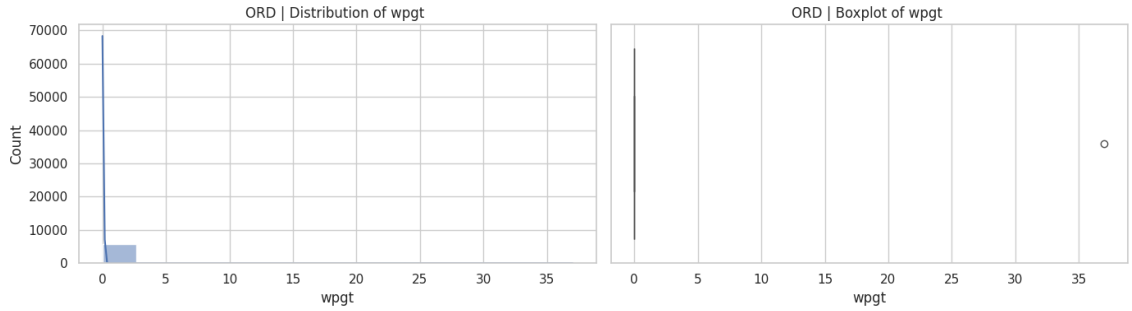


Figure 122 - Distribution and Boxplot Wind Peak Gust in ORD

The wind-related variables at ORD (*Figures* 120-122) reveal mixed results. *wdir* appears as a constant value, suggesting missing or unrecorded data, providing no usable insights for analysis. On the other hand, *wspd* shows a wide distribution centred around 15 km/h, with numerous outliers reaching up to 40 km/h.

wpgt is largely constant, except for rare extreme values, such as an isolated outlier close to 37 km/h. This pattern suggests that while gust data is limited in variability but like in JFK it is not fully constant.

These results align with the expected climate of Chicago, where ORD is located. The city is known for its variable and often strong winds, earning it the nickname “The Windy City.” While the constant values in wind direction and peak gusts may reflect limitations in the dataset, the variability in sustained wind speed matches Chicago’s climatology.

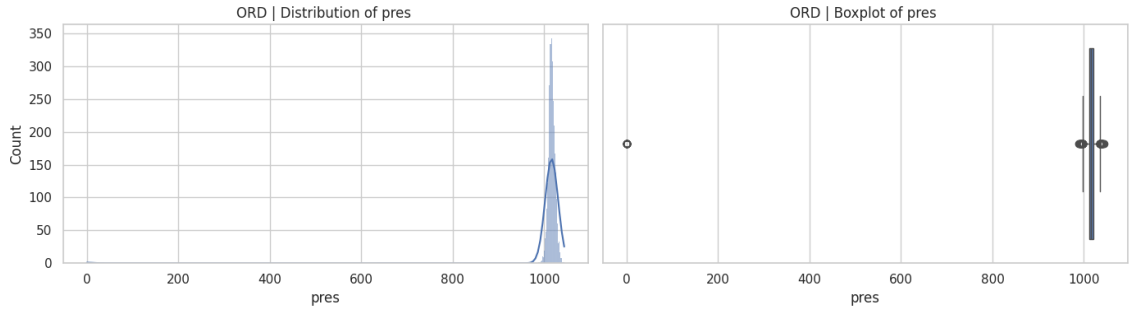


Figure 123 - Distribution and Boxplot of Atmospheric Pressure in ORD

In ORD the distribution of pres (*Figure* 123) shows, once again, a very tight clustering around 1000 hPa. With only minor deviations, apart from isolated outliers equal to zero, probably due to reading errors.

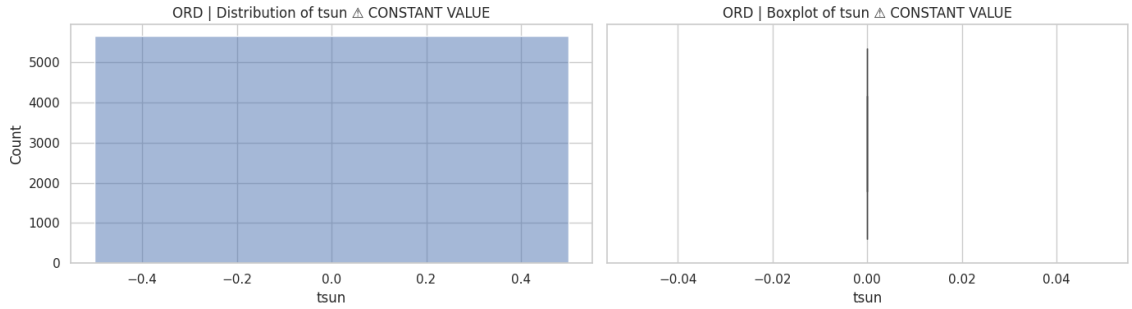


Figure 124 - Distribution and Boxplot of Total Sunshine Duration in ORD

The plots for *tsun* at ORD (Figure 124) reveal a constant value across the dataset, with no variability captured in either the histogram or the boxplot. As a result, like in the airports before, the *tsun* data does not provide any meaningful insights.

4.1.5. Weather Variables by Variable

4.1.5.1. Temperature Variables

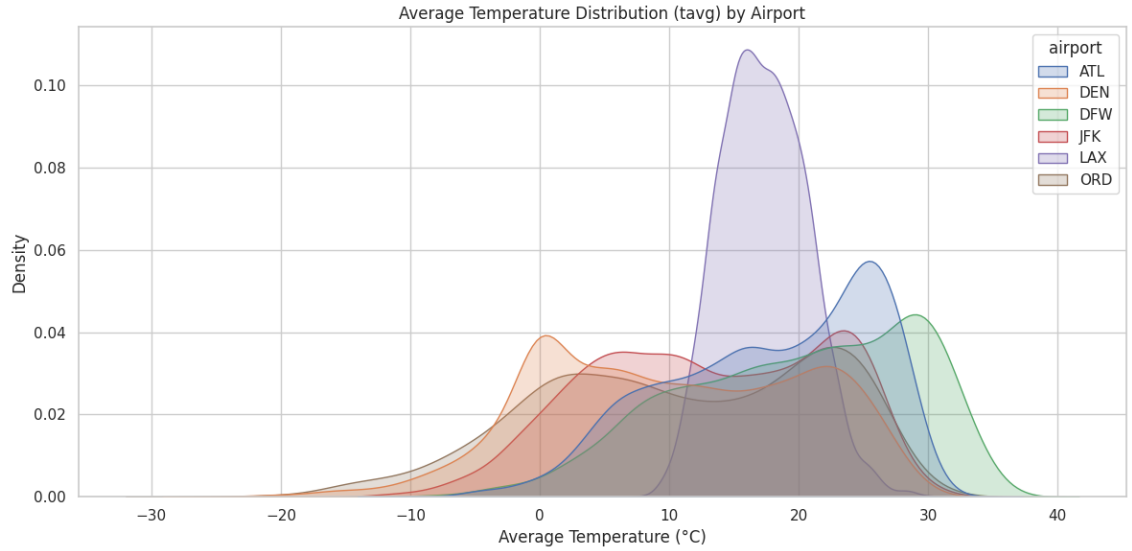


Figure 125 - Distribution of Average Temperature by Airport

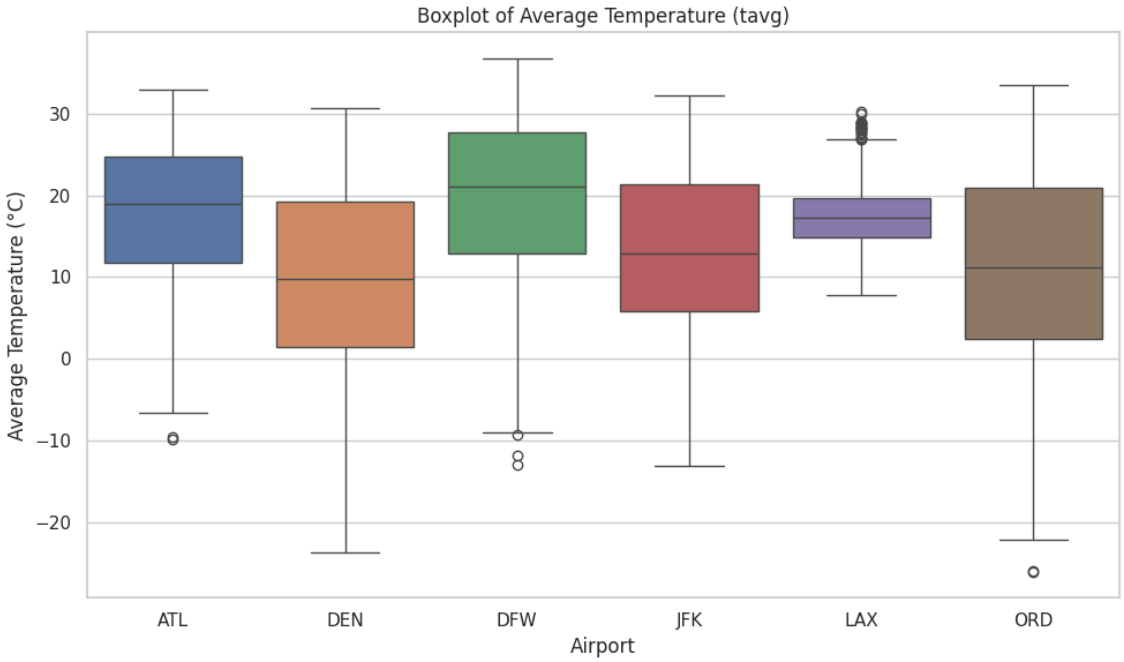


Figure 126 - Boxplot of Average Temperature by Airport

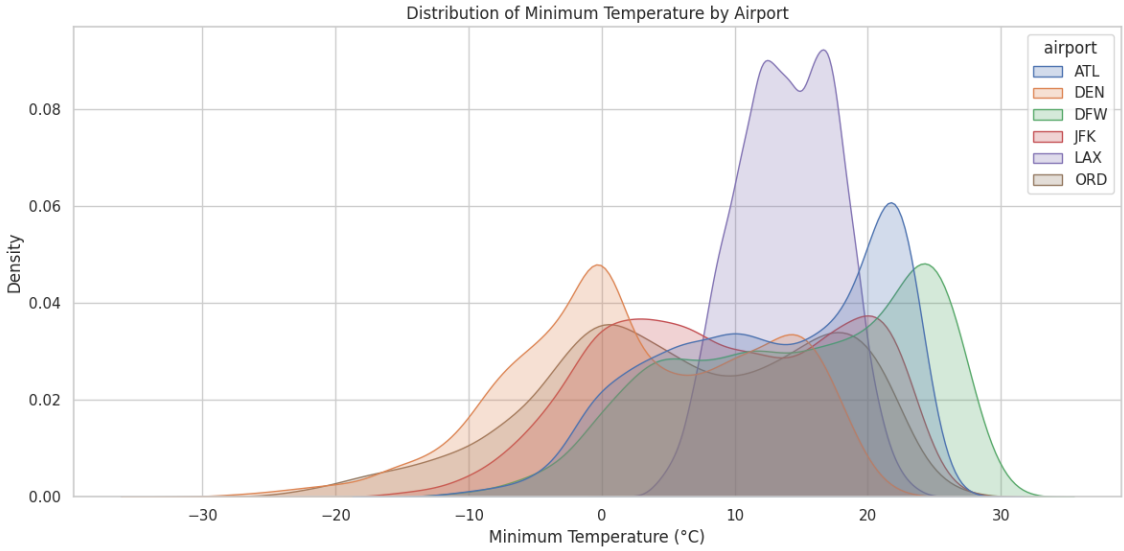


Figure 127 - Distribution of Minimum Temperature by Airport

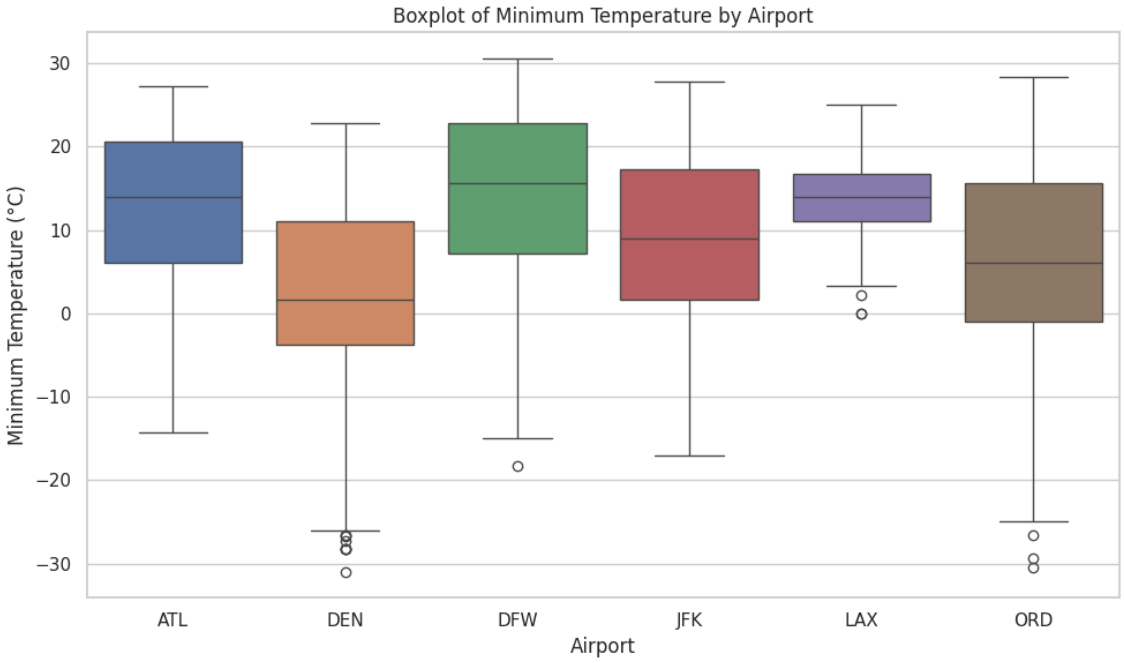


Figure 128 - Boxplot of Minimum Temperature by Airport

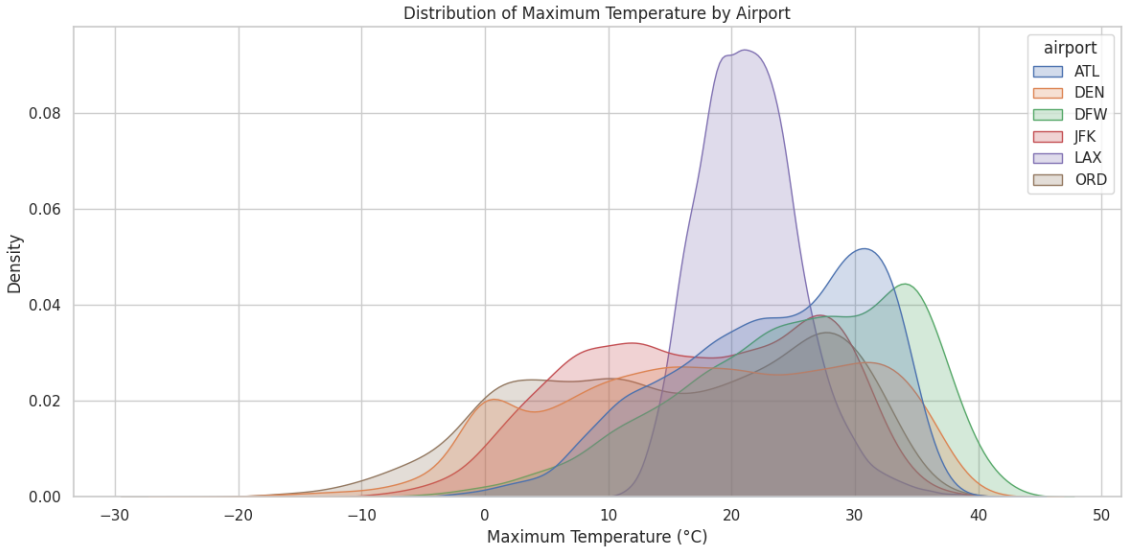


Figure 129 - Distribution of Maximum Temperature by Airport

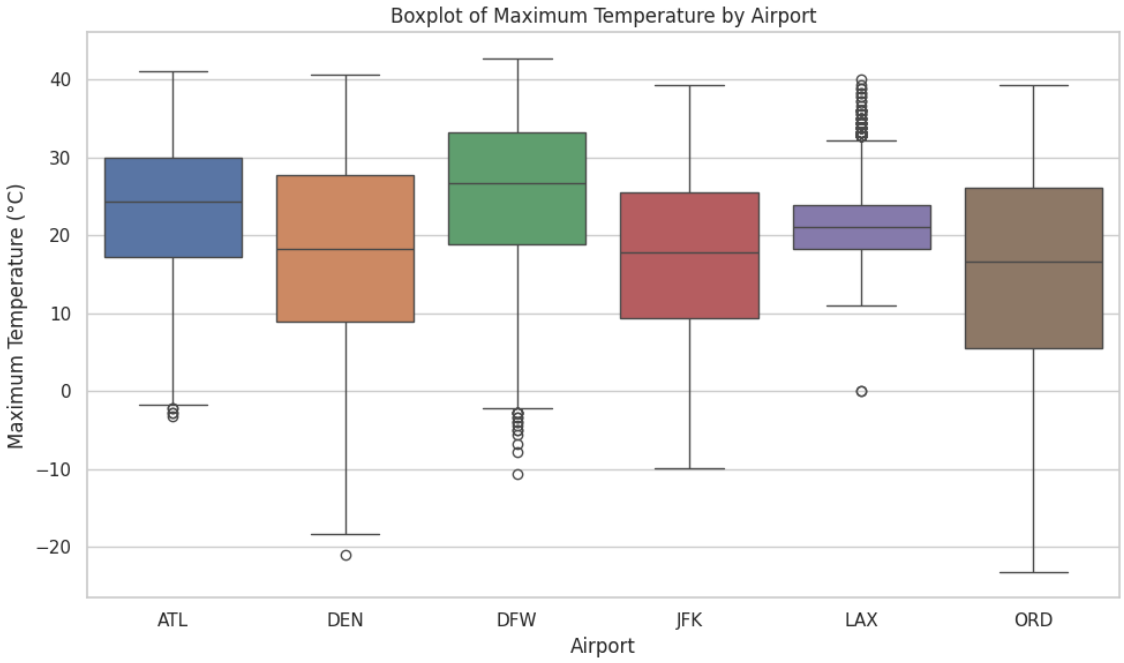


Figure 130 - Distribution of Maximum Temperature by Airport

The density plots and boxplots (Figures 125-130) show clear regional differences in temperature distributions. LAX stands out with the narrowest range, concentrated between 10 °C and 25 °C, reflecting its mild coastal climate. DEN and ORD exhibit the widest ranges, with minimums dropping below –20 °C and maximums exceeding 35 °C, highlighting strong seasonal contrasts. ATL and DFW show broader distributions skewed toward higher temperatures, while JFK's range is more balanced but still includes significant winter lows. Outliers are particularly visible in ORD and DEN, consistent with extreme cold spells.

These results align closely with the climatological expectations for each location. LAX, on the Pacific coast, benefits from a stable Mediterranean climate with minimal variation. Denver (DEN), located in the Rocky Mountains, experiences continental extremes, from very cold winters to hot summers. Chicago (ORD), influenced by its Midwestern continental climate and Lake Michigan, also shows a wide span, with sharp seasonal contrasts. Atlanta (ATL) and Dallas–Fort Worth (DFW) reflect the warm and humid subtropical conditions of the Southeast and South-Central U.S., with consistently high averages but less severe cold. New York (JFK), with its coastal position, displays moderate extremes compared to inland cities, though cold winters remain evident.

4.1.5.2. Precipitation

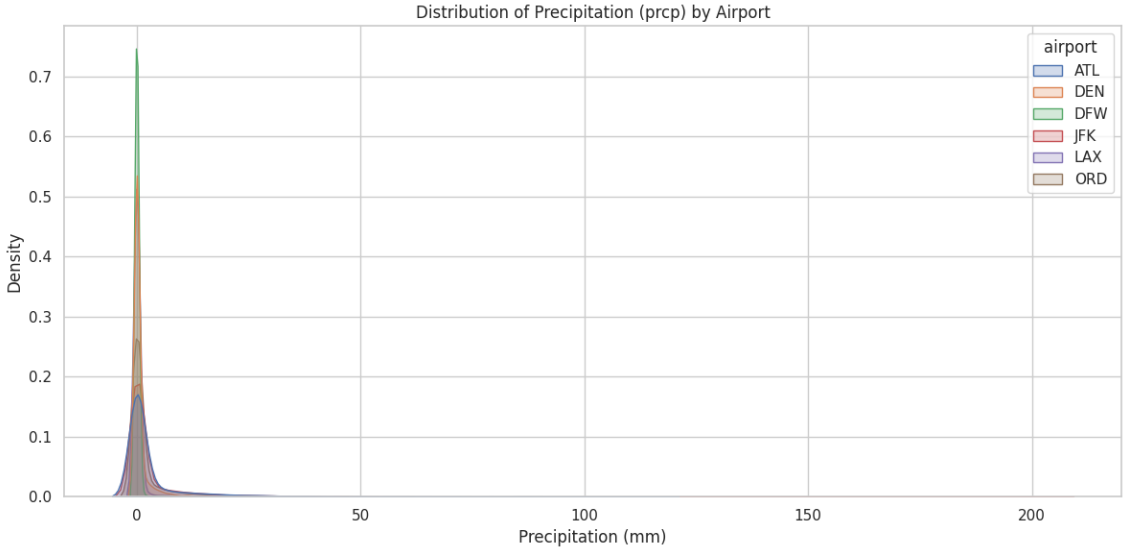


Figure 131 - Distribution of Precipitation by Airport

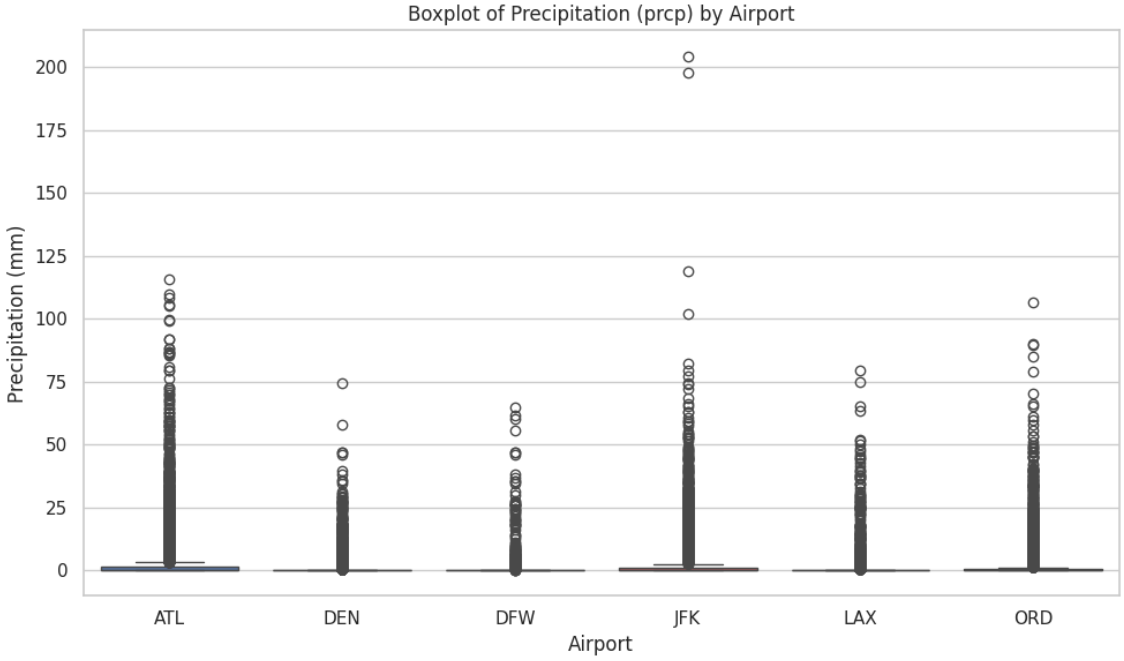


Figure 132 - Boxplot of Precipitation by Airport

Precipitation across all airports is heavily skewed toward low or no rainfall (*Figures* 131 and 132), with occasional extreme outliers. JFK and ATL show the highest extremes, consistent with their coastal and humid subtropical climates prone to storms, while ORD also records frequent heavy rainfall typical of the Midwest. In contrast, DEN and DFW exhibit lower precipitation due to their semi-arid and continental influences, and LAX reflects Southern California's dry Mediterranean climate with infrequent but

sometimes intense winter rains. Overall, the results align well with the expected climatic profiles of each location.

4.1.5.3. Snow Depth

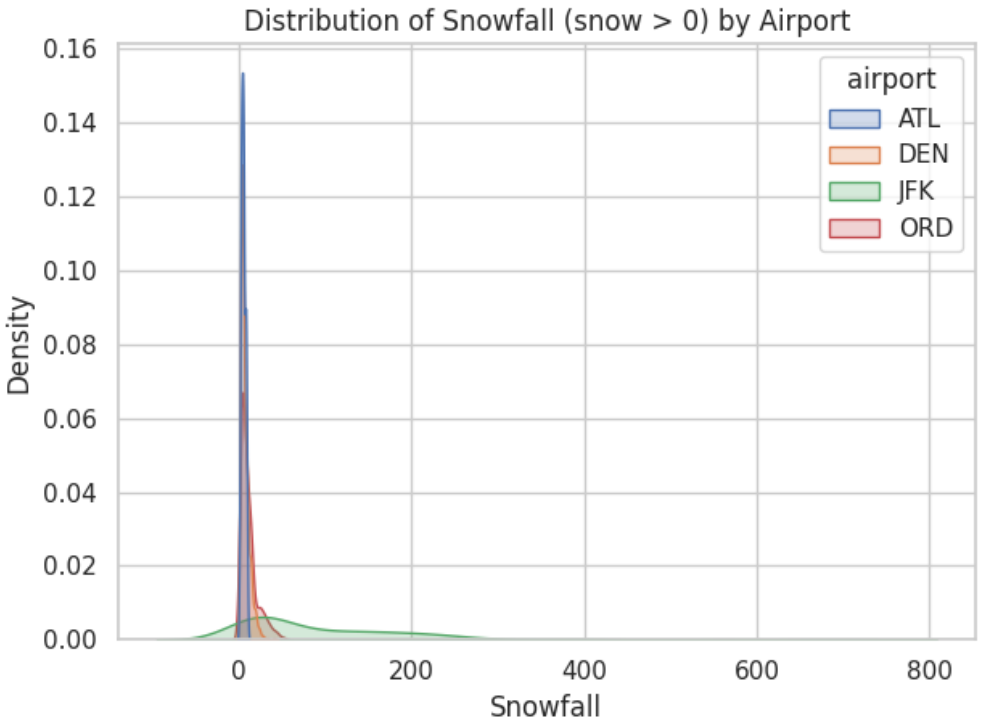


Figure 133 - Distribution of Snow Depth (when >0) by Airport

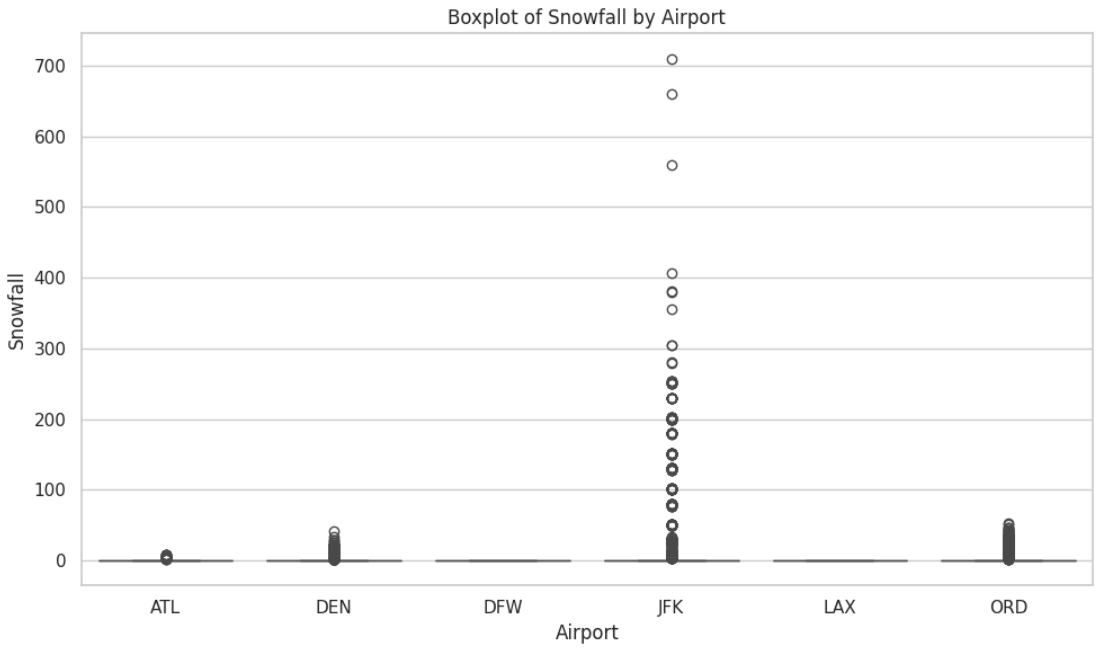


Figure 134 - Boxplot of Snow Depth by Airport

Snow Depth (*Figures* 133 and 134) is negligible at ATL and DFW, while ORD, DEN, and JFK show frequent snow, with JFK displaying extreme outliers from coastal blizzards. These results match local climates, with southern airports rarely experiencing snow and northern airports facing regular to severe winter events.

4.1.5.4. Wind Variables

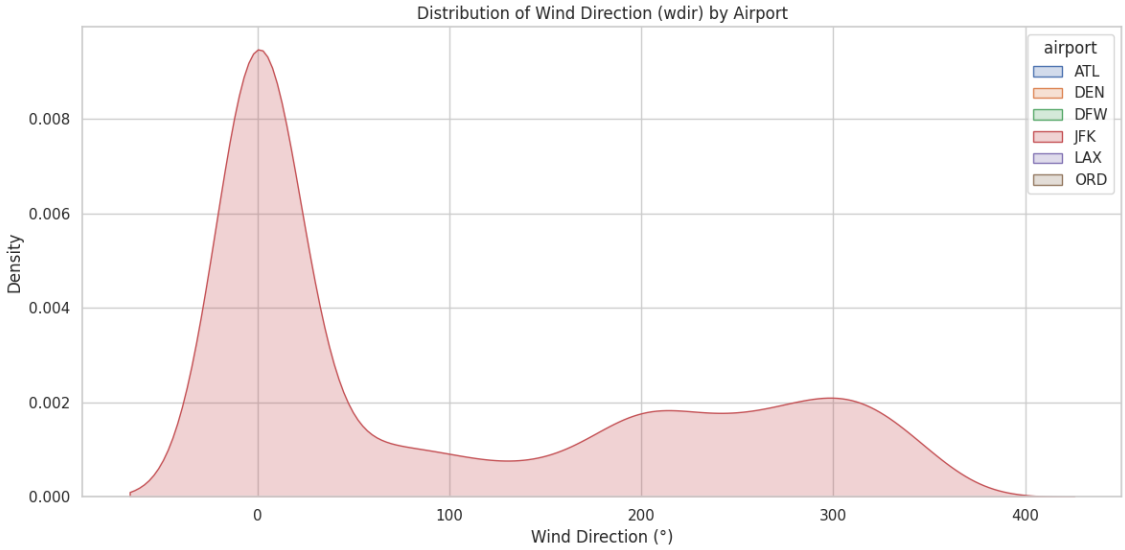


Figure 135 - Distribution of Wind Direction by Airport

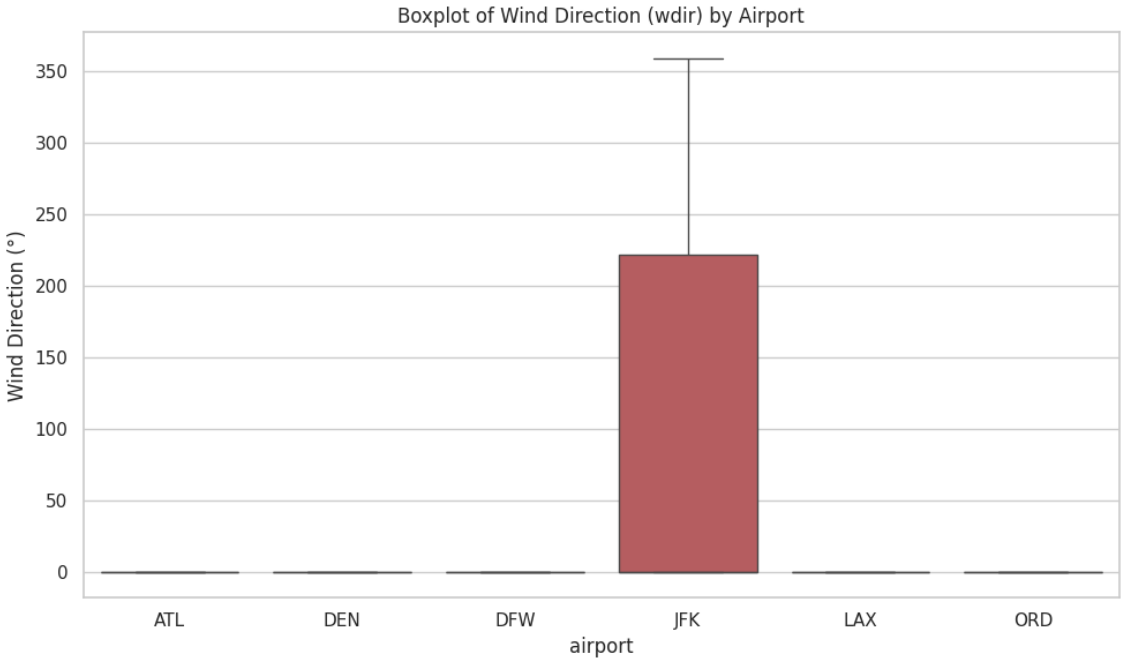


Figure 136 - Boxplot of Wind Direction by Airport

From *wdir* (*Figures* 135 and 136) we can see that only JFK shows any real distribution, showing that, as we suspected from the airport analysis, this variable probably had some capturing issues across the other airports, making it uninformative for our analysis.

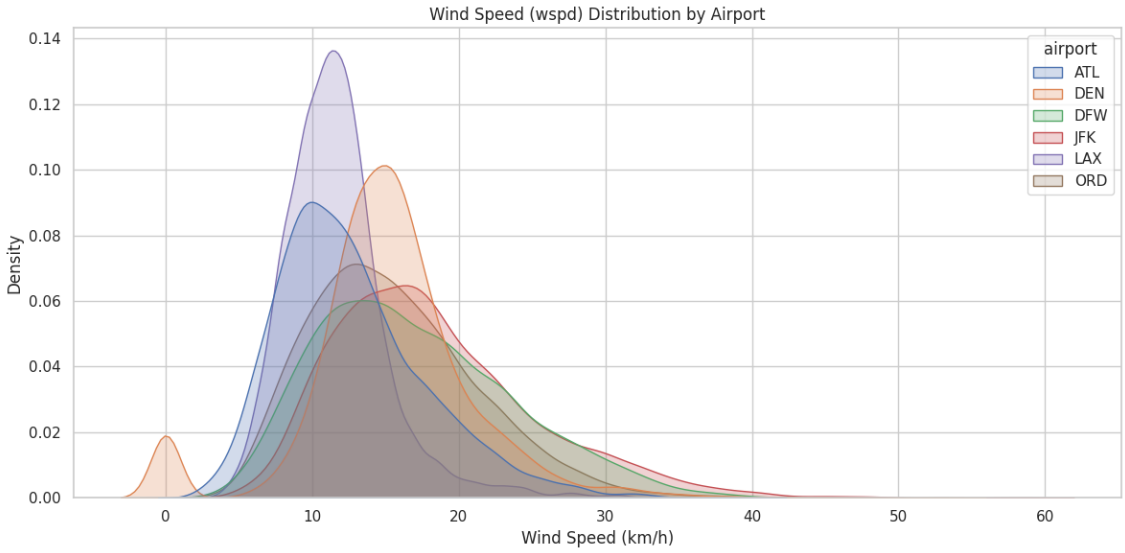


Figure 137 - Distribution of Wind Speed by Airport

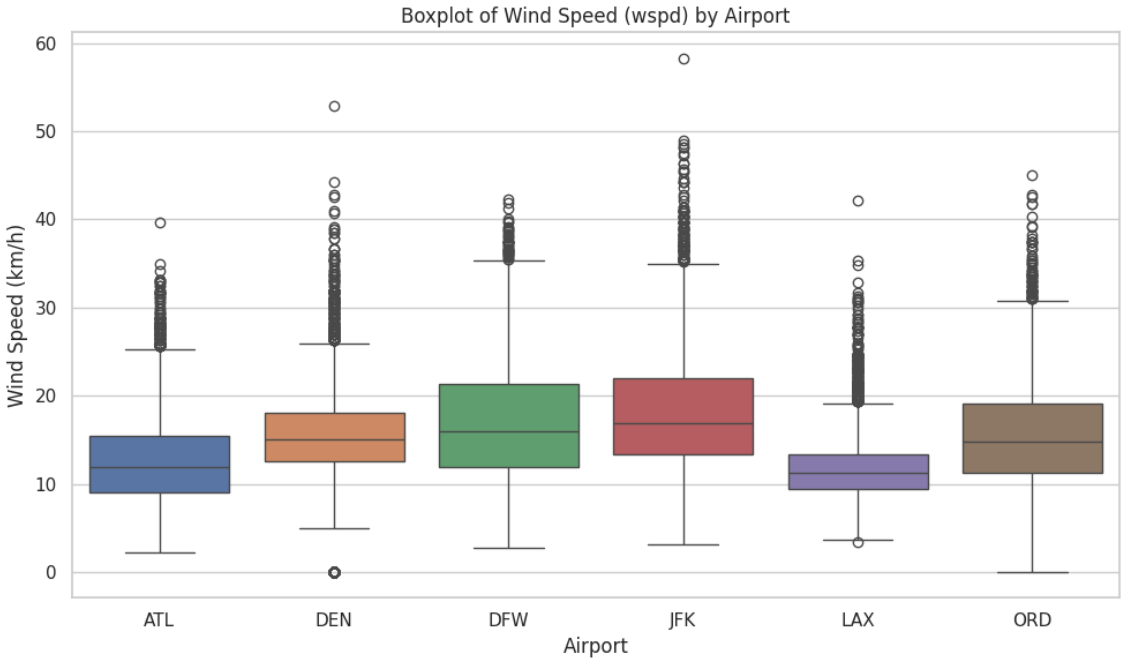


Figure 138 - Boxplot of Wind Speed by Airport

The *wspd* distributions (*Figures* 137 and 138) show that most airports cluster around moderate values, typically between 10 and 20 km/h, though variability exists across locations. JFK, DFW, and ORD display broader distributions and higher outliers, indicating occasional stronger winds, while ATL and LAX are more concentrated at the

lower end, reflecting generally calmer conditions. The boxplots confirm these trends, with LAX exhibiting the narrowest spread and JFK and ORD having the widest, alongside more extreme high-wind events.

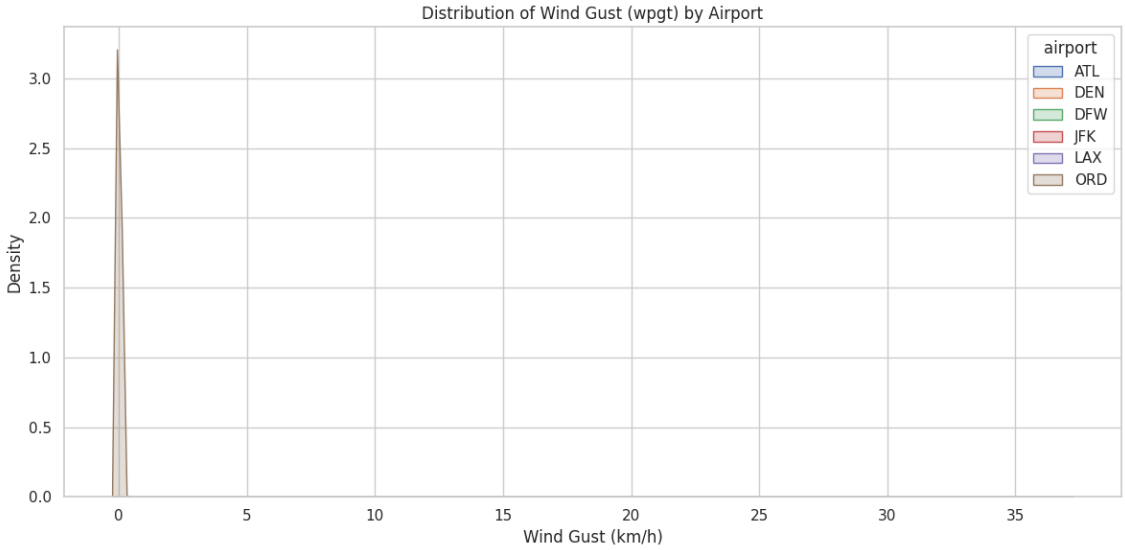


Figure 139 - Distribution of Wind Peak Gust by Airport

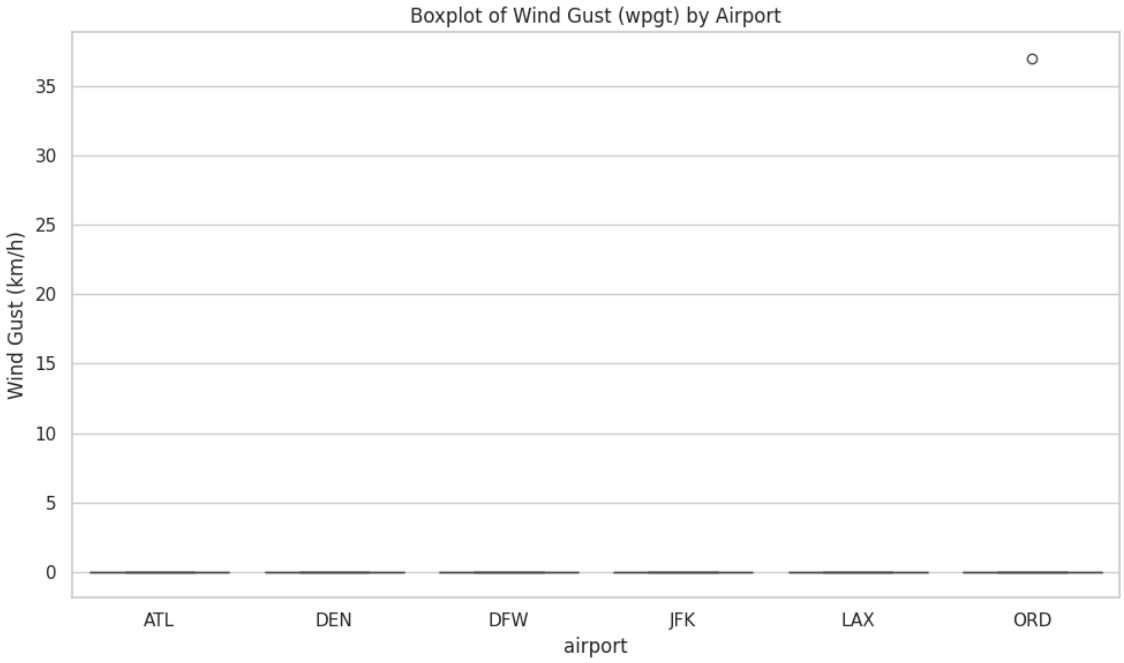


Figure 140 - Distribution of Wind Peak Gust by Airport

Once again virtually no information can be obtained from *wpgt* (*Figures* 139 and 140). This can probably be due to capturing or information storage errors, but since a similar thing happens with wind direction, perhaps the issue is due to the wind measuring equipment.

4.1.5.5. Atmospheric Pressure

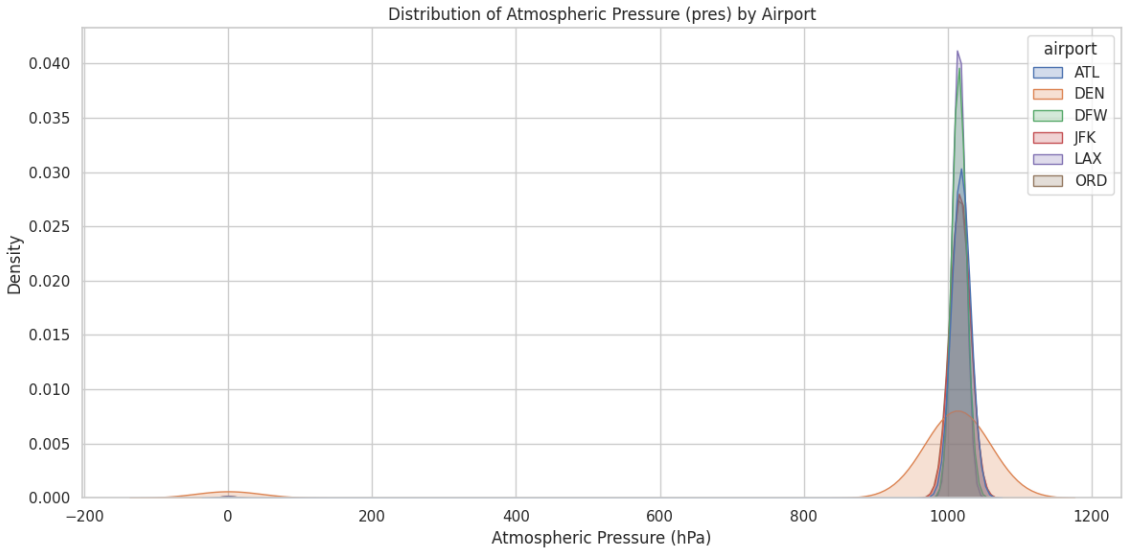


Figure 141 - Distribution of Atmospheric Pressure by Airport

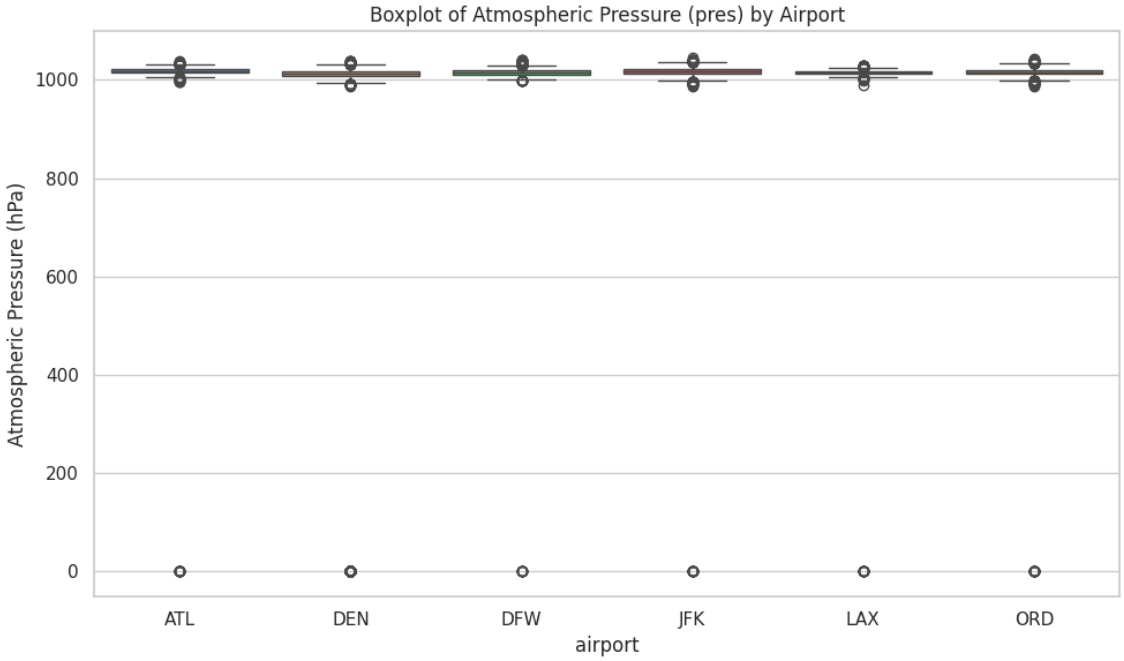


Figure 142 - Boxplot of Atmospheric Pressure by Airport

The distribution of *pres* (*Figures* 141 and 142) across airports shows values clustered around 1000 hPa, with minimal variability, except for Denver (DEN), which displays a slightly lower distribution reflecting its high-altitude location. Outliers near zero are likely erroneous data points rather than true observations and we will deal with them during the data cleaning and manipulation step.

4.1.5.6. Sunshine Duration

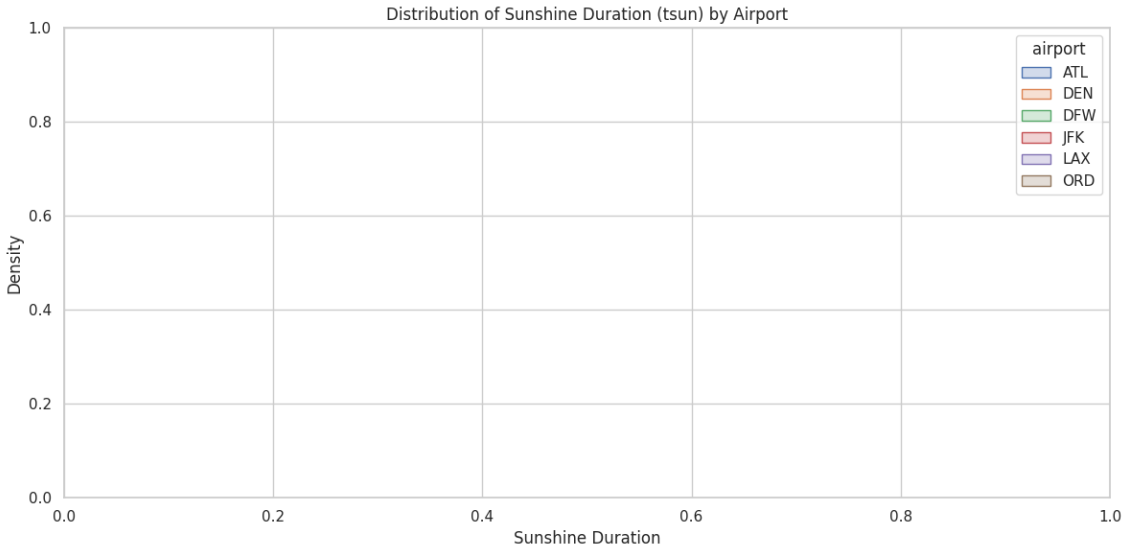


Figure 143 - Distribution of Sunshine Duration by Airport

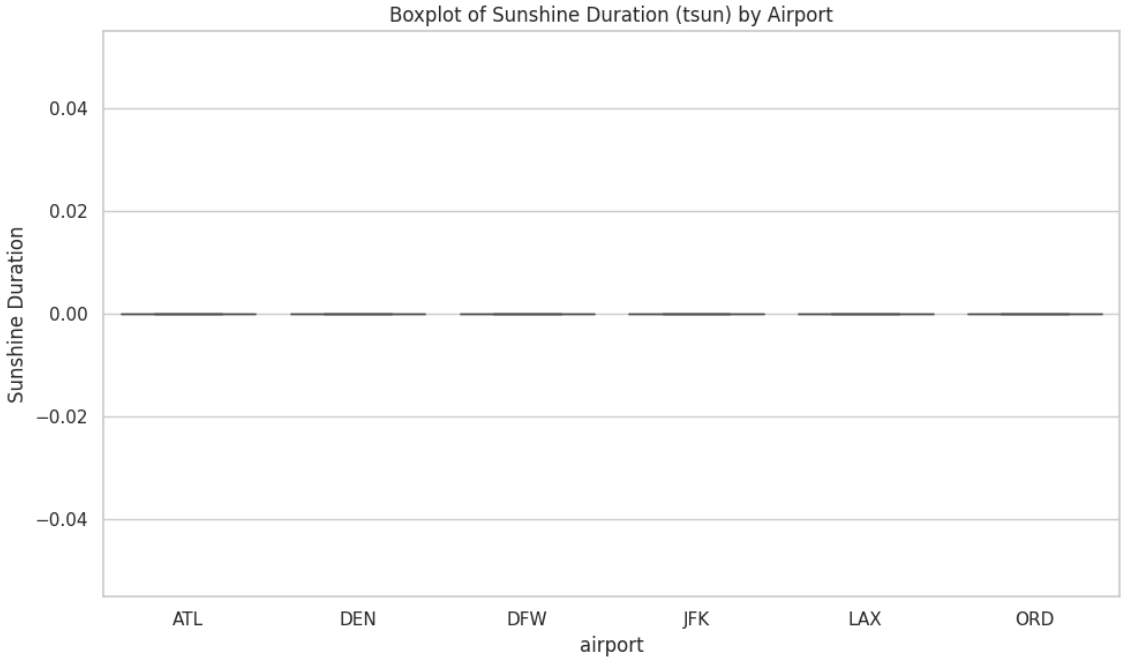


Figure 144 - Boxplot of Sunshine Duration by Airport

Again, tsun shows no interpretable data across all airports (*Figures* 143 and 144). This is most likely due to capturing or data availability issues: many U.S. weather stations, especially those operated at airports, do not consistently measure or report sunshine duration. Instead, proxies such as cloud cover or solar radiation data are sometimes used.

4.2. Time Series Exploration

4.2.1. Weekly Trends

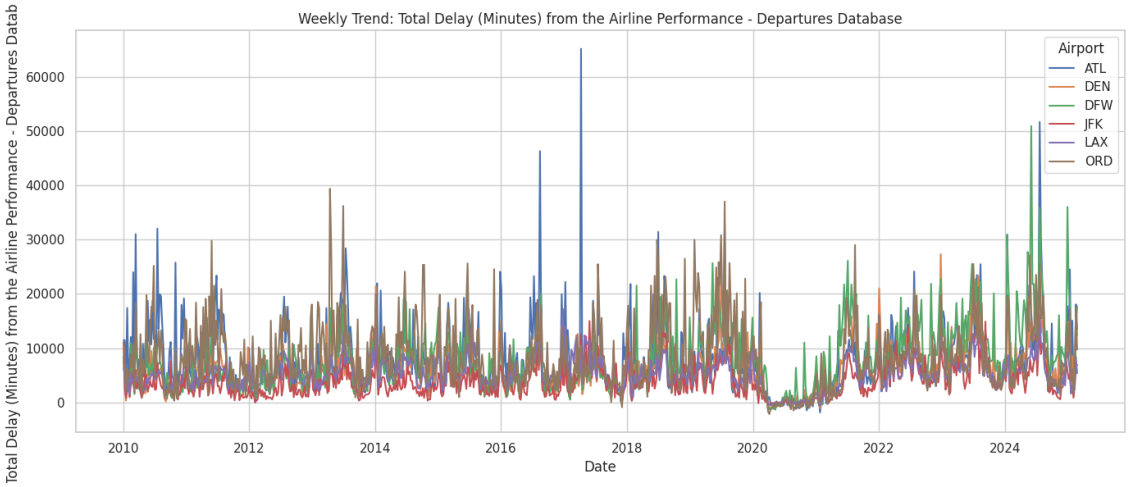


Figure 145 - Time Series: Weekly Trends of Total Delay (Departure) by Airport

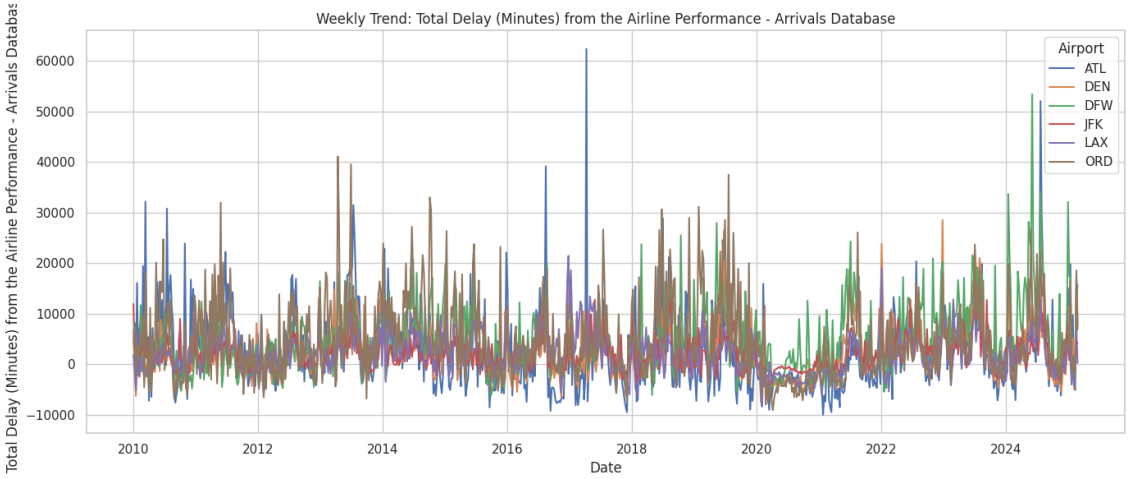


Figure 146 - Time Series: Weekly Trends of Total Delay (Arrival) by Airport

This weekly time series, both for departures and arrivals (*Figures* 145 and 146), reveals strong signs of seasonality across all airports. Peaks in delay minutes tend to occur around summer months and end-of-year (holiday season), which are well-known periods of heightened air travel demand. Conversely, valleys appear more consistently during late winter and early spring.

These patterns align with expected aviation behaviour: summer brings thunderstorms in the Southeast (impacting ATL, DFW), while winter snow and ice heavily affect northern hubs (ORD, DEN, JFK). LAX shows comparatively less extreme variability,

reflecting its milder climate, but still follows the seasonal passenger-demand cycles. Overall, seasonality emerges as a factor seemingly influencing delay trends.

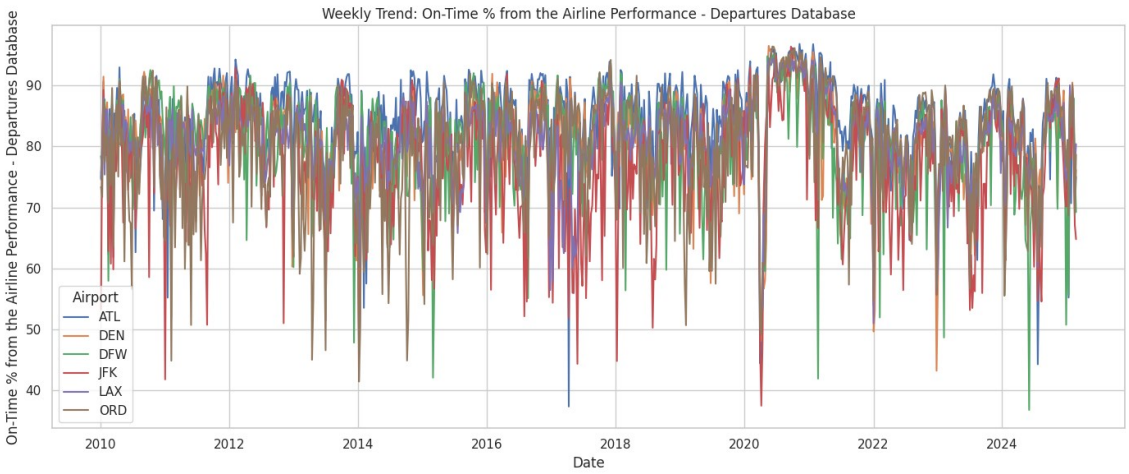


Figure 147 - Time Series: Weekly Trends of On-Time % (Departure) by Airport

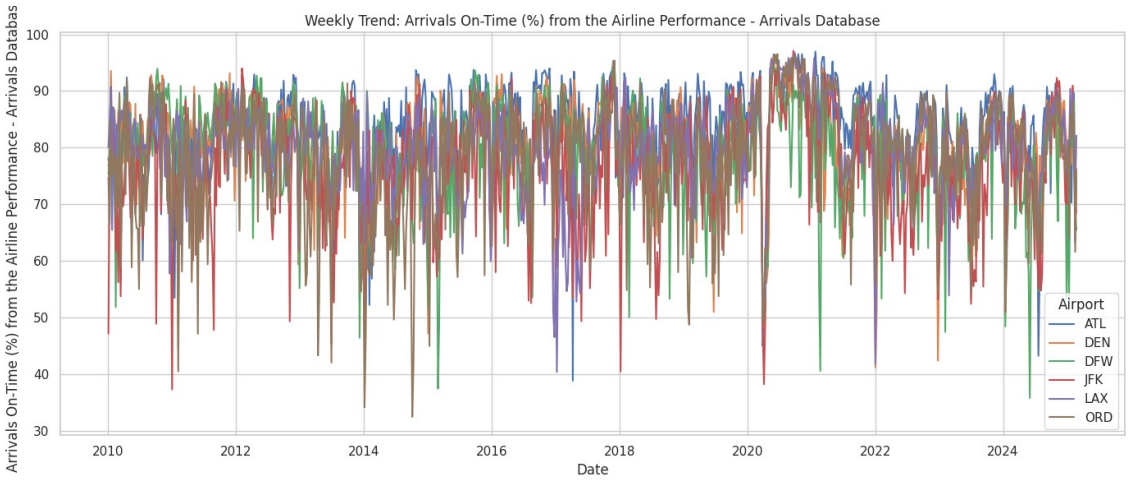


Figure 148 - Time Series: Weekly Trends of On-Time % (Arrival) by Airport

These weekly trends of *On-time* flights (*Figures* 147 and 148) show a consistent seasonal rhythm across all six airports. Peaks of higher punctuality are often observed during the spring and autumn months, while dips are more frequent in the winter and summer periods, coinciding with disruptive weather conditions (snowstorms in northern hubs like ORD, JFK, and DEN, and thunderstorms or extreme heat in ATL and DFW). LAX shows less pronounced seasonality, with on-time performance remaining relatively stable, reflecting its milder climate.

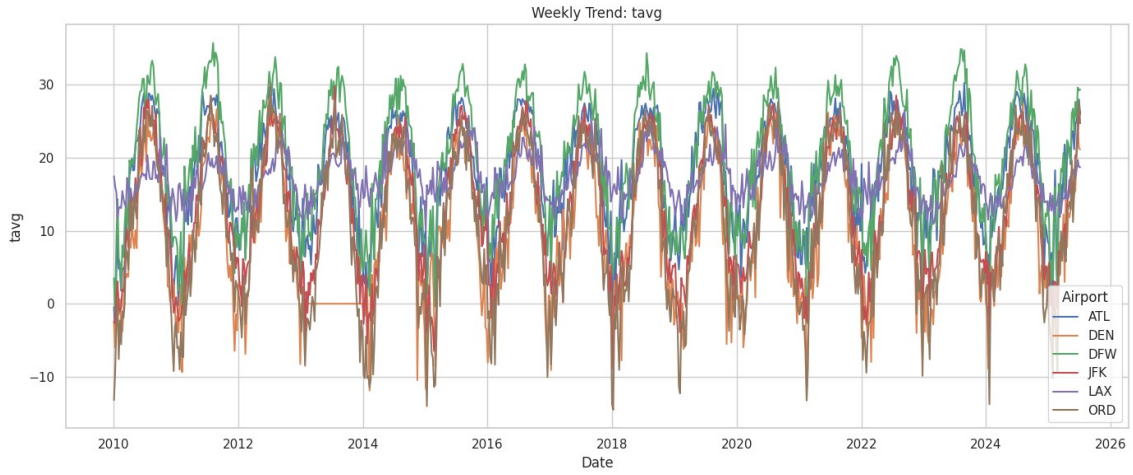


Figure 149 - Time Series: Weekly Trends of Average Temperature by Airport

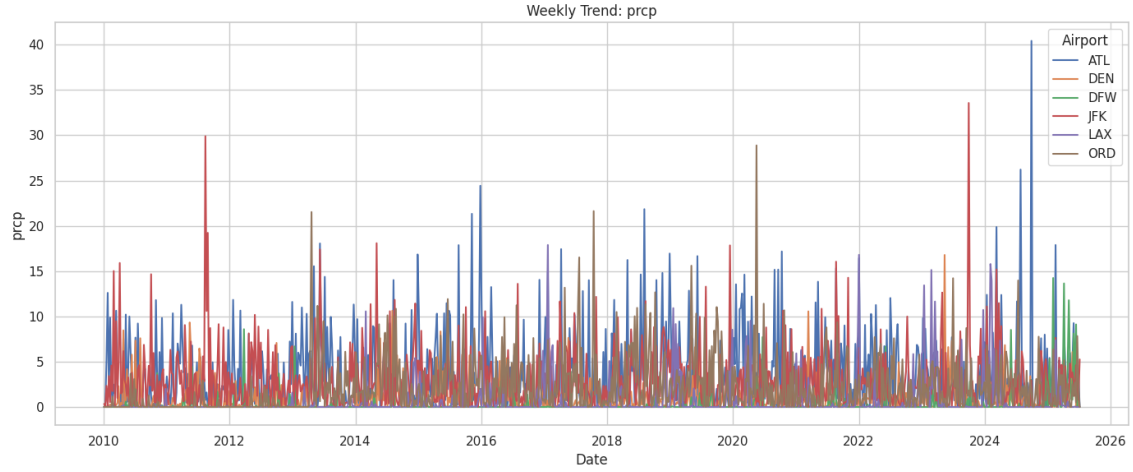


Figure 150 - Time Series: Weekly Trends of Average Precipitation by Airport

For *tavg* (*Figure* 149), as expected there is a strong, regular yearly cycle at all airports, with peaks in the summer months and troughs in the winter months. This seasonality is highly consistent and matches expected climate patterns — warmer summers and colder winters — though the amplitude differs: DEN, ORD, and JFK show sharper swings between winter and summer, while LAX remains more stable with milder fluctuations.

The prcp distribution patterns (*Figure* 150) are much more irregular compared to temperature. However, some seasonality still emerges: ATL and JFK, for example, tend to show more frequent precipitation peaks during summer and fall (linked to thunderstorms and tropical systems), whereas DEN and ORD display higher spikes in spring and winter, reflecting snowfall and transitional weather systems. LAX shows fewer but distinct peaks, possibly associated with winter storms.

2.2.2. Monthly Trends

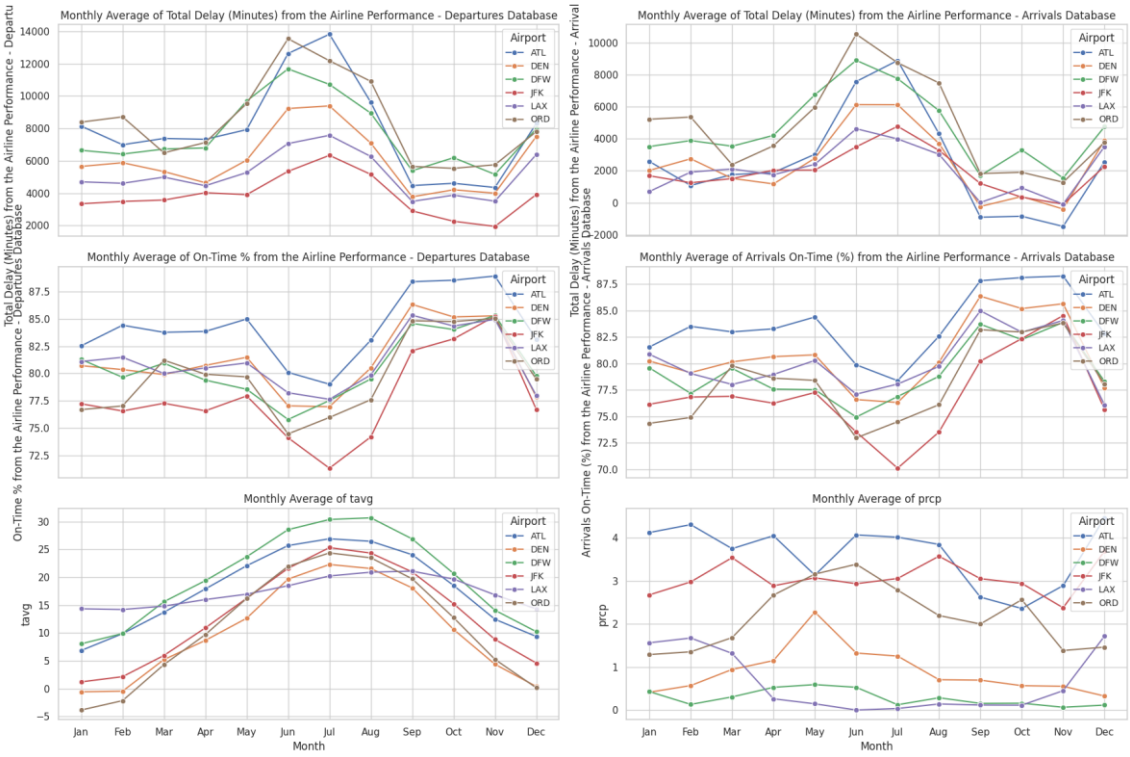


Figure 151- Time Series: Monthly Trends for Relevant Variables by Airport

For *Total Delays*, both departure and arrival delays (*Figure* 151) tend to peak in summer, with ORD, JFK, and ATL showing the strongest increases. This aligns with the well-known summer travel peak and more frequent convective storms. On-time performance mirrors this pattern inversely, dipping in summer and recovering in autumn and winter. Notably, JFK consistently performs worse than the others, particularly in July and August.

tavg follows the expected seasonal cycle, rising sharply from winter into summer and peaking in July before declining. *prcp* shows less uniformity across airports: ATL maintains high rainfall through summer, JFK and ORD peak in summer and early fall, while DEN and DFW remain relatively drier but with variability linked to storm seasons. LAX stands out with consistently low precipitation, reflecting its semi-arid climate.

4.2. Correlation Analysis

The correlation heatmaps (*Appendix* 5) reveal marked differences in how weather variables influence operational performance across the six airports. As expected,

temperature measures (average, minimum, and maximum) exhibit very strong internal correlations across all locations, yet their direct association with delays remains weak. Instead, precipitation, snow, and wind appear to play a more critical role, albeit with varying intensity depending on the airport.

For DEN and ORD, *snow* emerges as a particularly influential factor. Both airports show a clear negative correlation between snow and temperature, alongside mild positive correlations with *total delays*. This reflects the operational challenges posed by winter weather, where snow accumulation can disrupt runways, ground operations, and schedules. DEN further demonstrates a notable relationship between *wpgt* and temperature, underscoring the airport's unique exposure to altitude and severe winter conditions. By contrast, DFW and ATL exhibit limited weather–delay correlations, with precipitation and snow showing only weak associations with performance metrics.

JFK and LAX occupy an intermediate position. At JFK, both *precipitation* and *snow* demonstrate stronger correlations with delays compared to Atlanta or Dallas, consistent with the city's vulnerability to heavy rain and winter storms. At LAX, snow is irrelevant, but precipitation stands out as the most significant weather factor.

5. Data Cleaning and Transformation

Next, we will clean and transform the data, to ensure it is suitable for modelling. We will also drop uninformative variables that will not benefit our models and choose only one performance variable to use as our dependent variable. All these steps can be observed in the code (*Appendix* 4).

5.1. Handling Outliers

To detect outliers in our dataset we computed the mean standard deviation of each variable, and the variables with the most variance were the Total Delay variables, both arrivals and departures, making these candidates for outlier removal, since Total Delays could reach tens of thousands of minutes due to extreme events in some days.

Other variables with a high mean standard deviation were justifiable to keep as is, for example the Air Pressure variable, that had the third highest mean standard deviation value, is a weather variable with a valid, limited meteorological range, and as such its outliers were kept.

One other variable whose outliers were taken out was the Precipitation variable, since most days in all airports were equal to 0, but some were over 100 (mm), creating a highly right-skewed variable.

Due to how heavily skewed the data is we didn´t use the standard deviation method to cut the outliers, since this method would mislabel normal high values of Precipitation as outliers (since it is 0 most days). But instead, we used quantile-based clipping, which also doesn't assume normality and works well with long-tailed distributions. We used this method to clip values above the 99^{th} percentile for the two Total Delay and the Precipitation variables.

5.2. Logically invalid values

We then shifted our focus to dealing with invalid values that we noticed during the EDA and decide now what to do with them. These instances were, cases of negative departure and arrival delays, pressure values of 0 and one instance of the percentage of on-time departing flights being <0 or >100.

The negative departure and arrival delays are due to these values being a sum of the total delays of that day, and early flights counting as negative delays. This makes these variables less reliable to act as the dependent variable, but removing all instances of negative delays would mean removing 2 116 rows for departure and 14 688 rows for arrival, across all six airports. So instead, we replaced the negative delay values with 0, equalling negative delays to no delays, even if that isn´t necessarily the case.

For the air pressure of 0 we had a total of 491 cases across all six airports, this probably being a result of a faulty reading or sensor, since this value makes no sense and the rest of the distribution checks out for hPa (the SI unit for atmospheric pressure). To resolve this we decided to replace all instances of 0 with NaN

Lastly, we had to deal with the case of the On-Time Departing flight percentage being outside of its supposed range. This could probably be due to a mistake calculating this value or handling the data, so just like the last case, we dealt with it by replacing it with NaN.

5.3. Log-Transformation

To decide which variables should be log-transformation, we calculated their mean skewness. The result was the variables with the most meaningful skews being the snow depth, precipitation, and both total delays variables. We log-transformed all of these except for the snow depth variable, due to its sparse and low values which would make the transformation not add much value.

5.4. Dummy Variables

To make use of our temporal variable, in date, we created dummy variables for the month and the day of the week of each day. That way we can use them in our models,

since a standard date-time variable would be unreadable and cause an error, making our models stronger and more accurate.

5.5. Feature Selection

To start our feature selection, we must deal with uninformative columns, dropping them entirely from our datasets, to reduce dimensionality in our models. Since having a high number of features will make our modelling more challenging, and these won´t add much to the explanatory power of our models.

The uninformative variables dropped were *Total Sunshine Duration*, *Wind Direction* and *Wind Peak Gust*, due to being mostly constant or NaNs, with most airports had trouble recording these values. Then we dropped the *NAS Delay* variables, due to the issue of duplication of the Arrivals and Departure databases, making the data unusable.

Before continuing we also needed to narrow our target variable to one. *NAS Delay* is already off the table, and *Total Delay* had to be log-transformed and had the "negative delays" issue – this leaves us with *On-Time (%) Flights and 15 Minutes or More Delays*. Of the two, *On-Time (%) Flights* is the more informative one, as it considers all types of delays we are looking to study, not just longer ones.

To include both Departures and Arriving flights we created a composite "*OnTime_Average*" variable, using the average of both for each day, as the chosen target variable.

Table 3 shows the final variables used in our modelling, after feature selection.

Dependent Variable (Y)		
	OnTime_Average	Average of percentage of departing and arriving flights that were on-time
Independent Variables (X)		
	tavg	Average Temperature (Celsius)

	tmin	Minimum Temperature (Celsius)
	tmax	Maximum Temperature (Celsius)
	prcp	Total Precipitation (mm)
	snow	Depth of Snow (mm)
	wspd	Wind Speed (km/h)
	pres	Atmospheric Pressure (hPa)
	month_x	Eleven binary variables representing the month, with x ranging from 2 to 12 (February to December). January is the reference month.
	dow_x	Six binary variables representing the day of the week, where x is the first three letters of the current day of the week. Sunday is the reference day.

Table 3 - Final Variables

5.6. NaNs and Missing Values

All our missing values are already marked as NaNs, this includes 129 rows for each airport, due to missing data on the performance variables database (whose values between 22-02-25 until 01-07-25), and the invalid values we turned into NaNs.

Lastly, since redundant columns were already dropped in Feature Selection, we dropped all rows with any NaN values, leaving only observations without any missing or null values to use in our models.

6. Results

6.1. Model Performance Overview

6.1.1. OLS

	Airport	n_train	n_test	R^2	MAE	RMSE
0	ATL	4405	1102	0.129149	6.674249	10.499134
1	DEN	4113	1029	0.060516	7.862669	11.362580
2	DFW	4416	1105	0.068586	8.843449	13.582348
3	JFK	4405	1102	0.160737	8.764009	12.626378
4	LAX	4415	1104	0.044695	5.690992	7.703031
5	ORD	4404	1101	0.154288	7.799666	11.117422

Table 4 - OLS Model Output

We can see from *Table* 4 that the model has low explanatory power, explaining only around 16% of the variance in on-time performance at best, in the case of JFK´s R squared, and only around 4% at worst, in the case of LAX´s R squared. The Mean Absolute Error (MAE), ranges from 5.7 in LAX, to 8.8 in DFW, and the Root Mean Squared Error (RMSE) stands between 7.7 and 13.6 with LAX and DFW having the lowest and highest values respectively once again.

Overall, the OLS baseline shows the linear relationships between weather variables and the delays as weak, or this relationship is driven by non-linear factors that OLS cannot capture. Even so, we by looking at the variation among the airports, we can see that JFK and ORD show more explained variance than other airports, possibly due to effects of the snow, rain and seasonal storms. Meanwhile, LAX low explanatory power might be due to the milder climate variations and lack of snow.

6.1.2. Ridge and Lasso

	Airport	Model	R^2	MAE	RMSE
0	ATL	Linear	0.129149	6.674249	10.499134
1	ATL	Ridge	0.129104	6.674413	10.499402
2	ATL	Lasso	0.128823	6.674672	10.501099
3	DEN	Linear	0.060516	7.862669	11.362580
4	DEN	Ridge	0.060517	7.862672	11.362578
5	DEN	Lasso	0.060470	7.863210	11.362864
6	DFW	Linear	0.068586	8.843449	13.582348
7	DFW	Ridge	0.068586	8.843419	13.582346
8	DFW	Lasso	0.068534	8.842489	13.582726
9	JFK	Linear	0.160737	8.764009	12.626378
10	JFK	Ridge	0.160735	8.764015	12.626392
11	JFK	Lasso	0.160709	8.763875	12.626588
12	LAX	Linear	0.044695	5.690992	7.703031
13	LAX	Ridge	0.044702	5.690982	7.703000
14	LAX	Lasso	0.045100	5.690675	7.701399
15	ORD	Linear	0.154288	7.799666	11.117422
16	ORD	Ridge	0.154290	7.799658	11.117413
17	ORD	Lasso	0.154381	7.799107	11.116811

Table 5 - Ridge, Lasso and OLS Outputs

As seen in *Table* 5, the variations in explanatory power are minimal across airports for both Ridge and Lasso, with JFK and ORD maintaining the highest R squared at 16.1% and 15.4% respectively, with changes only in the fourth or fifth decimal place.

The measures for error maintain no observable enhancement, both for MAE and RMSE, except in specific cases, like the MAE in JFK lowering 0.0002 with Lasso.

Unlike it could be expected the models perform almost identically, meaning regularization adds little value in this scenario, suggesting the predictors are not harmed by multicollinearity and are already relatively stable on the OLS.

6.1.3. Random Forest (RF)

	Airport	Model	R^2	MAE	RMSE
0	ATL	Random Forest	0.270253	6.191947	9.610979
1	DEN	Random Forest	0.331463	6.673959	9.585067
2	DFW	Random Forest	0.242358	8.093229	12.249988
3	JFK	Random Forest	0.333314	7.952285	11.253569
4	LAX	Random Forest	0.070416	5.695409	7.598623
5	ORD	Random Forest	0.381408	6.541321	9.508132

Table 6- Random Forest Model Output

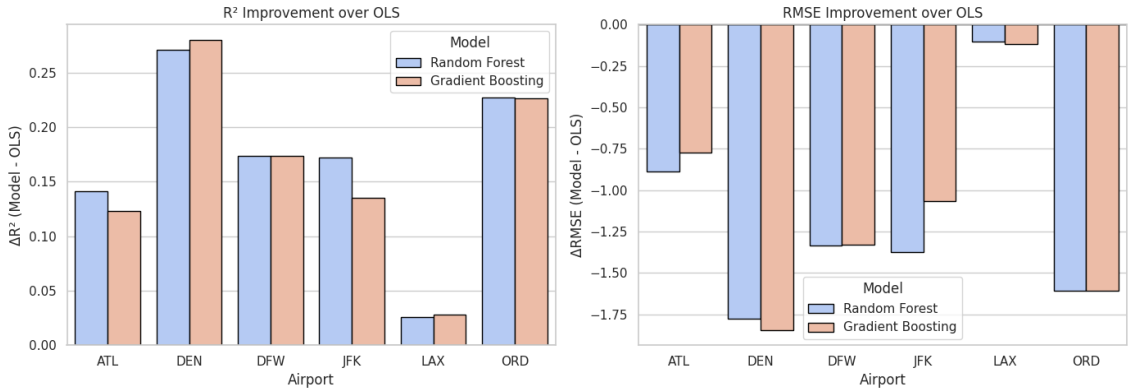


Figure 152- RF and GB Improvement Over OLS

Looking at *Table* 6 and *Figure* 153 we can see that RF shows a big jump in explanatory power, roughly double those of the linear models. ORD, DEN and JFK stand out as the best R squares of random forests with significant improvements, ATL and DFW show moderate improvement and LAX remains weak, possibly confirming the delays being less weather driven and more operations related.

Errors also drop meaningfully compared to OLS, with LAX continuing to have the lowest MAE and RMSE, while DFW, followed by JFK possess the highest.

RF clearly outperforms the linear models by capturing non-linear and interaction effects in the data. And while LAX remains weakly modelled, DFW also remains challenging as it still shows high MAE and RMSE in spite improvement, possibly reflecting a more complex interaction between weather and delays.

6.1.4. Gradient Boosting (GB)

	Airport	Model	R²	MAE	RMSE
0	ATL	Gradient Boosting	0.252471	6.214432	9.727368
1	DEN	Gradient Boosting	0.340736	6.663765	9.518357
2	DFW	Gradient Boosting	0.242190	7.963979	12.251340
3	JFK	Gradient Boosting	0.296371	8.156241	11.561165
4	LAX	Gradient Boosting	0.073034	5.689412	7.587914
5	ORD	Gradient Boosting	0.381089	6.568404	9.510583

Table 7 - Gradient Boosting Model Output

Table 7 shows that GB has a somewhat similar performance to RF in terms of explanatory power. ORD and DEN stand again as the best modelled, having the highest R squared, with similar values to RF, JFK and ATL are slightly weaker but still

much stronger than linear models (*Figure* 152), DFW stands similar and LAX remains the lowest R squared.

The errors are consistent with RF, with LAX still performing the best in these parameters, possibly artificially due to low variability in weather-driven delays.

GB and RF perform very similarly, with different airports performing better in each of the two models.

6.2. Cross-Airport Insights

```
=== Best Performing Model per Airport ===
  Airport                  Model      R²  ΔR²_vs_OLS     MAE     RMSE
0     ATL      **Random Forest**  0.2703      0.1411  6.1919   9.6110
1     DEN  **Gradient Boosting**  0.3407      0.2802  6.6638   9.5184
2     DFW      **Random Forest**  0.2424      0.1738  8.0932  12.2500
3     JFK      **Random Forest**  0.3333      0.1726  7.9523  11.2536
4     LAX  **Gradient Boosting**  0.0730      0.0283  5.6894   7.5879
5     ORD      **Random Forest**  0.3814      0.2271  6.5413   9.5081
```

Table 8 - Best Performing Models of Each Airport

From *Table* 8 we can see the best performing model for each airport. As stated, before the non-linear models were the ones with the best performance, mainly RF, having the best explanatory power in ATL, DFW, JFK and ORD, while GB has the best performance in DEN and ORD.

	Airport	Best Model	Variable	Impact (%)
3	ATL	Random Forest	prcp	30.45
6	ATL	Random Forest	pres	15.98
2	ATL	Random Forest	tmax	14.09
0	ATL	Random Forest	tavg	14.04
5	ATL	Random Forest	wspd	12.41
1	ATL	Random Forest	tmin	10.14
4	ATL	Random Forest	snow	2.89
10	DEN	Gradient Boosting	prcp	30.90
8	DEN	Gradient Boosting	tmin	18.99
12	DEN	Gradient Boosting	wspd	14.93
9	DEN	Gradient Boosting	tmax	11.91
7	DEN	Gradient Boosting	tavg	10.54
13	DEN	Gradient Boosting	pres	7.23
11	DEN	Gradient Boosting	snow	5.50
14	DFW	Random Forest	tavg	23.34
20	DFW	Random Forest	pres	21.44
19	DFW	Random Forest	wspd	20.93
16	DFW	Random Forest	tmax	18.42
15	DFW	Random Forest	tmin	14.46
17	DFW	Random Forest	prcp	1.42
18	DFW	Random Forest	snow	0.00
24	JFK	Random Forest	prcp	28.31
27	JFK	Random Forest	pres	15.60
26	JFK	Random Forest	wspd	14.97
21	JFK	Random Forest	tavg	13.68
22	JFK	Random Forest	tmin	13.34
23	JFK	Random Forest	tmax	10.69
25	JFK	Random Forest	snow	3.40
31	LAX	Gradient Boosting	prcp	33.34
33	LAX	Gradient Boosting	wspd	19.65
28	LAX	Gradient Boosting	tavg	13.48
34	LAX	Gradient Boosting	pres	13.09
30	LAX	Gradient Boosting	tmax	11.69
29	LAX	Gradient Boosting	tmin	8.75
32	LAX	Gradient Boosting	snow	0.00
38	ORD	Random Forest	prcp	21.37
41	ORD	Random Forest	pres	17.10
40	ORD	Random Forest	wspd	16.99
37	ORD	Random Forest	tmax	16.20
35	ORD	Random Forest	tavg	13.60
36	ORD	Random Forest	tmin	12.17
39	ORD	Random Forest	snow	2.57

Table 9- Feature Importance of Weather Variables Across Airport´s Best Models

Table 9 shows the importance of the weather variables for each of the best performing models of each airport. In all airports except DFW, perhaps due to its dryer climate, Precipitation stands as the most impactful weather variable on delays. Meanwhile, in all

airports snow depth appears as the least impactful variable, possibly due to the rarity of snowfall.

```
=== Weather Robustness per Airport ===
Airport        Best Model  R²_Full  R²_NoWeather  ΔR² (Weather Contribution)  RMSE_Full  RMSE_NoWeather  ΔRMSE (Weather Contribution)
    ATL     Random Forest   0.2944        0.0367                      0.2577     9.4508         11.0426                        1.5919
    DEN Gradient Boosting   0.3727        0.0112                      0.3615     9.2847         11.6568                        2.3722
    DFW     Random Forest   0.2846       -0.0447                      0.3294    11.9035         14.3850                        2.4814
    JFK     Random Forest   0.3455        0.0561                      0.2894    11.1506         13.3903                        2.2397
    LAX Gradient Boosting   0.0600        0.1271                     -0.0671     7.6411          7.3632                       -0.2779
    ORD     Random Forest   0.3286        0.0626                      0.2660     9.9058         11.7047                        1.7989
```

Table 10 - Model Performance Comparison with and without Weather Variables

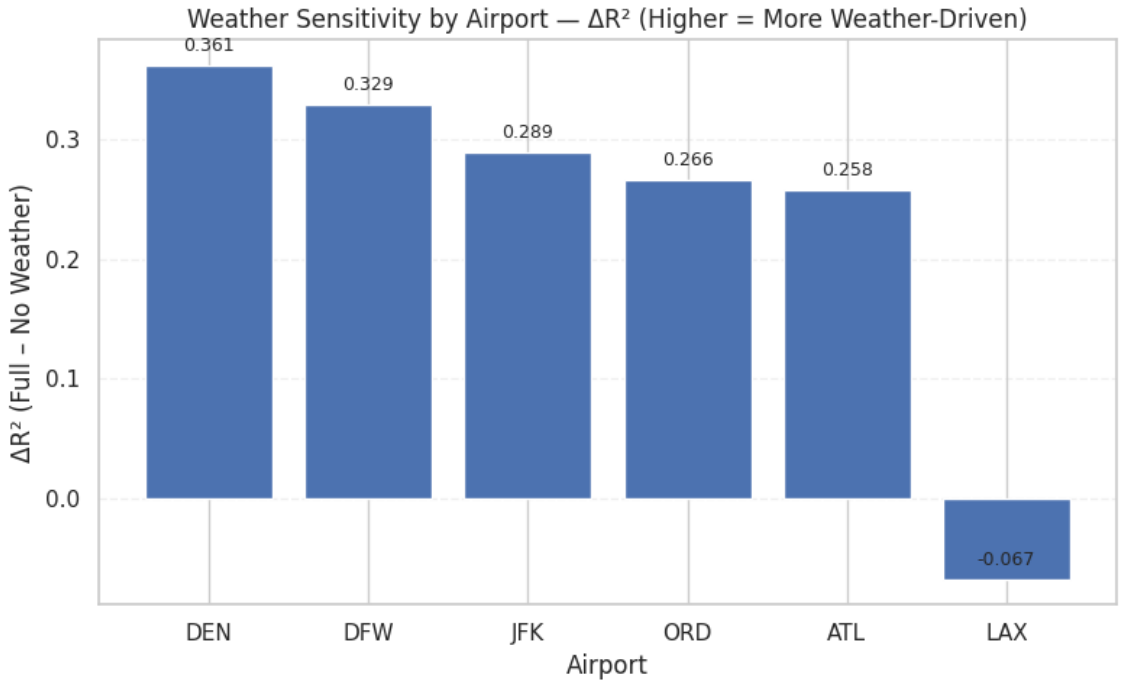


Figure 153 - Weather Sensitivity by Airport (R^2)

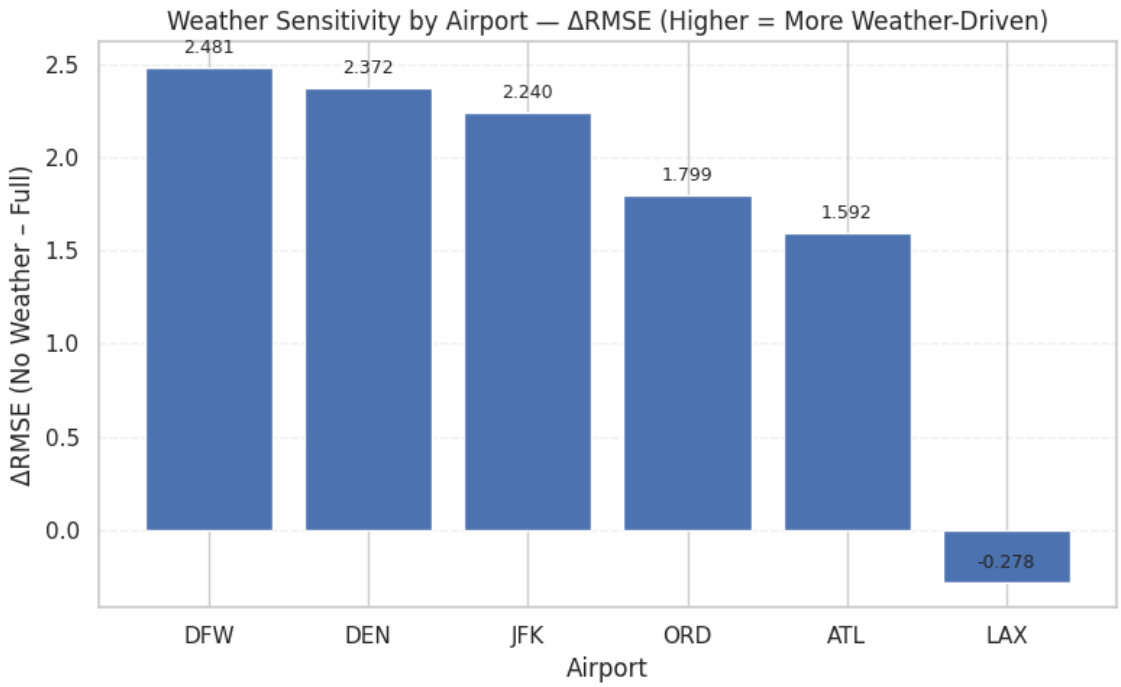


Figure 154 - Weather Sensitivity by Airport (RMSE)

By modelling the best performing models for each airport with only the non-weather variables (date related variables) we can see how the weather variables contribute to the delays. By doing this instead of just modelling with just the weather variables we are using a stronger model as reference, since we are dealing with relatively low values of R squared.

By doing this we can determine which airports are more sensitive to weather changes on their delays. In terms of resilience, the less sensitive to weather changes, the more robust the airports are, since the base percentage of delays isn't as affected by disruptive weather events.

On *Table* 10 and *Figures* 153 and 154 we can see exactly that, meaning that the most robust airport would be LAX, with the weather variables having a negative effect on explanatory power and in RMSE.

Following LAX, with a big difference in sensitivity, we have ATL, ORD and JFK airports, in which both R squared and RMSE variations agree in having the second, third and fourth highest resilience (or lowest weather sensitivity) of our six airports.

Lastly, we have DEN and DFW, while DEN has the highest weather sensitivity, it was DFW that registered the biggest change in RMSE when not using weather variables, confirming that the relatively high error rates of this value are caused by the relationship with the weather variables.

Conclusion

By equating robustness to weather sensitivity in flight delay variation and using machine learning modelling techniques we concluded that of our six airports LAX was our most robust one, followed by ATL, ORD, JFK, and lastly DFW closely followed by DEN.

The results lined up with climate expectations, with more robust airports being in areas with less extreme weather conditions. With *precipitation* being the most impactful features across most airports, with *average temperature* and *atmospheric pressure* also providing consistent relevant impact. Factors like these might be helpful to take into consideration when building/expanding airports, since as climate changes effects deepen extreme weather conditions become more common, so choosing the right geography becomes key, no not risk the airport´s resilience.

However, with more data accessible it would be possible to use more features for modelling (different and more specific weather variables) and using more historical data - since it grows each passing day - better predictive models will be produced and applied to a wider array of airports. A wider range of models can also be tested with models like k-means or more complex ones like neural networks, as to get closer to reality and better understand the relationship between transport resilience/robustness and weather variability.

AI Generative Declaration

During the preparation of my written work/thesis, Using Predictive Machine Learning Models to Analyse Weather Resilience and Robustness in U.S Airports, ChatGPT and NoteBookLM were used for the following tasks: Code debugging and organization, summarizing academic papers, reference formatting, language polishing and checking terminology consistency. After using these tools, I reviewed and edited the content as necessary, and I take full responsibility for the content of the work presented.

I also declare that I am aware of and respect the Artificial Intelligence Rules of Conduct of Católica Porto Business School.

References

Adams, T. M., Bekkem, K. R., & Toledo-Durán, E. J. (2012). Freight resilience measures. *Journal of Transportation Engineering, 138*(11), 1403–1409. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000415

Ahmed, S., & Dey, K. (2020). Resilience modeling concepts in transportation systems: A comprehensive review based on mode, and modeling techniques. *Journal of Infrastructure Preservation and Resilience, 1*(1), 8. https://doi.org/10.1186/s43065-020-00008-9

Barker, K., Ramirez-Marquez, J. E., & Rocco, C. M. (2013). Resilience-based network component importance measures. *Reliability Engineering & System Safety, 117,* 89–97. https://doi.org/10.1016/j.ress.2013.03.012.

Baroud, H., Barker, K., & Ramirez-Marquez, J. E. (2014). Importance measures for inland waterway network resilience. *Transportation Research Part E: Logistics and Transportation Review, 62,* 55–67. https://doi.org/10.1016/j.tre.2013.11.010.

Bergantino, A. S. (2024). Assessing transport network resilience: Empirical insights from spatially grounded analyses. *Transport Reviews.* Advance online publication. https://doi.org/10.1080/01441647.2024.2322434.

Bešinović, N. (2020). Resilience in railway transport systems: a literature review and research agenda', *Transport Reviews*, 40(4), pp. 457–478. https://doi.org/10.1080/01441647.2020.1728419.

Biau, G. and Scornet, E. (2016). A random forest guided tour, *TEST*, 25(2), pp. 197–227. https://doi.org/10.1007/s11749-016-0481-7.

Breiman, L. (2001). Random forests, *Machine Learning*, 45(1), pp. 5–32. https://doi.org/10.1023/A:1010933404324.

Bruneau, M., Chang, S.E., Eguchi, R.T., Lee, G.C., O'Rourke, T.D., Reinhorn, A.M., Shinozuka, M., Tierney, K., Wallace, W.A. and von Winterfeldt, D. (2003). A framework to quantitatively assess and enhance the seismic resilience of communities, *Earthquake Spectra*, 19(4), pp. 733–752. https://doi.org/10.1193/1.1623497.

Chen, X. (2024). Resilience measurement and analysis of intercity public and multimodal transport networks, *Transportation Research Part C: Emerging Technologies*, 165, 104702. https://doi.org/10.1016/j.trd.2024.104202.

Cordero, F. (2024). Performance measure–based framework for evaluating transport system resilience, *Transportation Research Record*, 2678(5), pp. 890–902. https://doi.org/10.1177/03611981231190396.

Cox, A., Prager, F. and Rose, A. (2011). Transportation security and the role of resilience: A foundation for operational metrics, *Transport Policy*, 18(2), pp. 307–317. https://doi.org/10.1016/j.tranpol.2010.09.004.

Faturechi, R. and Miller-Hooks, E. (2014). Travel time resilience of roadway networks under disaster, *Transportation Research Part B: Methodological*, 70, pp. 47–64. https://doi.org/10.1016/j.trb.2014.08.007.

Friedman, J.H. (2001). Greedy function approximation: A gradient boosting machine, *Annals of Statistics*, 29(5), pp. 1189–1232. https://doi.org/10.1214/aos/1013203451.

Holling, C.S. (1973). Resilience and stability of ecological systems, *Annual Review of Ecology and Systematics*, 4, pp. 1–23. https://doi.org/10.1146/annurev.es.04.110173.000245.

Hoerl, A.E. and Kennard, R.W. (1970). Ridge regression: Biased estimation for nonorthogonal problems, *Technometrics*, 12(1), pp. 55–67. https://doi.org/10.1080/00401706.1970.10488634.

Hosseini, S., Barker, K. and Ramirez-Marquez, J.E. (2016). A review of definitions and measures of system resilience, *Reliability Engineering & System Safety*, 145, pp. 47–61. https://doi.org/10.1016/j.ress.2015.08.006.

Jin, J.G., Tang, L.C., Sun, L. and Lee, D.H. (2014). Enhancing metro network resilience via localized integration with bus services, *Transportation Research Part E: Logistics and Transportation Review*, 63, pp. 17–30. https://doi.org/10.1016/j.tre.2014.01.002.

Joseph, V. R. (2022). Optimal ratio for data splitting. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *15*(4), 531-538. https://doi.org/10.48550/arXiv.2202.03326.

Karl, T.R. and Koss, W.J. (1984). Regional and national monthly, seasonal, and annual temperature weighted by area, 1895–1983. *Historical Climatology Series 4-3.* Asheville, NC: National Climatic Data Center, U.S. Department of Commerce.

Leobons, C.M., Barcellos, V., Campos, G. and De Mello, R.A. (2019). Assessing urban transportation systems resilience – A proposal of indicators, *Transportation Research Procedia*, 37, pp. 322–329. https://doi.org/10.1016/j.trpro.2018.12.199.

Mattsson, L.-G. and Jenelius, E. (2015). Vulnerability and resilience of transport systems – A discussion of recent research, *Transportation Research Part A: Policy and Practice*, 81, pp. 16–34. https://doi.org/10.1016/j.tra.2015.06.002.

Mitchell, T.M. (1997). Machine learning. New York, NY: McGraw-Hill. ISBN 978-0070428072.

Murray-Tuite, P.M. (2006). A comparison of transportation network resilience under simulated system optimum and user equilibrium conditions, *Proceedings of the 2006 Winter Simulation Conference*, pp. 1398–1405. IEEE. https://doi.org/10.1109/WSC.2006.323240.

Nipa, T.J., Chowdhury, A.G. and Rahman, M.M. (2023). Development of a decision-making system measuring the resilience level of highway projects, *Journal of Infrastructure Preservation and Resilience*, 4(1), 17. https://doi.org/10.1186/s43065-023-00084-7.

Pan, S., Yan, H., He, J. and He, Z. (2021). Vulnerability and resilience of transportation systems: A recent literature review, *Physica A: Statistical Mechanics and its Applications*, 581, 126235. https://doi.org/10.1016/j.physa.2021.126235.

Sun, W., Bocchini, P. and Davison, B.D. (2020). Resilience metrics and measurement methods for transportation infrastructure: the state of the art, *Sustainable and Resilient Infrastructure*, 5(3), pp. 168–199. https://doi.org/10.1080/23789689.2018.1448663.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), pp. 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

UNISDR (2009). Terminology on disaster risk reduction. Geneva: United Nations International Strategy for Disaster Reduction.

Wan, C., Yang, Z., Zhang, D., Yan, X. and Fan, S. (2018). Resilience in transportation systems: a systematic review and future directions, *Transport Reviews*, 38(4), pp. 479–498. https://doi.org/10.1080/01441647.2017.1383532.

Wooldridge, J.M. (2012). Introductory econometrics: A modern approach. 5th edn. Mason, OH: South-Western Cengage Learning.

Yuan, T., DaRocha, W., Rothenberg, C.E., Obraczka, K., Barakat, C. and Turletti, T. (2022). Machine learning for next-generation intelligent transportation systems: A survey, *Transactions on Emerging Telecommunications Technologies*, 33(4), e4427. https://doi.org/10.1002/ett.4427.

Zadeh, L.A. (2008). Is there a need for fuzzy logic?, *Information Sciences*, 178(13), pp. 2751–2779. https://doi.org/10.1016/j.ins.2008.02.012.

Zhang, L. (2024). Analysis of multi-modal public transportation system resilience using performance-based dynamic simulations, *Transport Policy*, 145, pp. 45–57. https://doi.org/10.1016/j.tranpol.2024.04.006.

Zhou, Y., Wang, J. and Yang, H. (2019). Resilience of Transportation Systems: Concepts and Comprehensive Review, *IEEE Transactions on Intelligent Transportation Systems*, 20(12), pp. 4262–4276. https://doi.org/10.1109/TITS.2018.2883766.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2), pp. 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x.

Federal Aviation Administration (2025). *CY2024 ACAIS Preliminary enplanements at U.S. airports, by state*. Washington, DC: FAA. Available at: https://repository.library.noaa.gov/view/noaa/10238 (Accessed: 20 June 2025).

Appendices

Appendix 1 – JoiningWeatherDatabases.py (Self-made)

```
#Importing necessary libraries

import pandas as pd


#Loading our ATL_Weather excel files into pandas DataFrames

ATL_Weather_2010_2019 = pd.read_excel('Weather_ATL_01012010_01012019.xlsx')

ATL_Weather_2019_2025 = pd.read_excel('Weather_ATL_02012019_01072025.xlsx')


#Concatenate the DataFrames using the 'date' column

ATL_Weather = pd.concat([ATL_Weather_2010_2019, ATL_Weather_2019_2025],
axis=0, ignore_index=False)


#There should be no missing values, but just in case let´s handle them by filling NaN
values

ATL_Weather.fillna(0, inplace=True)


#Let´s see the first few rows of the merged DataFrame

print(ATL_Weather.head())


#Saving the DataFrame to a new Excel File

ATL_Weather.to_excel('ATL_Weather.xlsx',index=False)


#-------------------------------------------------------------------------#
```

```
#Loading our DEN_Weather excel files into pandas DataFrames

DEN_Weather_2010_2019 =
pd.read_excel('Weather_DEN_01012010_01012019.xlsx')

DEN_Weather_2019_2025 =
pd.read_excel('Weather_DEN_02012019_01072025.xlsx')


#Concatenate the DataFrames using the 'date' column

DEN_Weather = pd.concat([DEN_Weather_2010_2019, DEN_Weather_2019_2025],
axis=0, ignore_index=False)


#There should be no missing values, but just in case let´s handle them by filling NaN
values

DEN_Weather.fillna(0, inplace=True)


#Let´s see the first few rows of the merged DataFrame

print(DEN_Weather.head())


#Saving the DataFrame to a new Excel File

DEN_Weather.to_excel('DEN_Weather.xlsx',index=False)


#------------------------------------------------------------------------------#


#Loading our DFW_Weather excel files into pandas DataFrames

DFW_Weather_2010_2019 =
pd.read_excel('Weather_DFW_01012010_01012019.xlsx')

DFW_Weather_2019_2025 =
pd.read_excel('Weather_DFW_02012019_01072025.xlsx')
```

```
#Concatenate the DataFrames using the 'date' column
DFW_Weather = pd.concat([DFW_Weather_2010_2019, DFW_Weather_2019_2025], axis=0, ignore_index=False)

#There should be no missing values, but just in case let´s handle them by filling NaN values
DFW_Weather.fillna(0, inplace=True)

#Let´s see the first few rows of the merged DataFrame
print(DFW_Weather.head())

#Saving the DataFrame to a new Excel File
DFW_Weather.to_excel('DFW_Weather.xlsx',index=False)

#-------------------------------------------------------------------------#

#Loading our JFK_Weather excel files into pandas DataFrames
JFK_Weather_2010_2019 = pd.read_excel('Weather_JFK_01012010_01012019.xlsx')
JFK_Weather_2019_2025 = pd.read_excel('Weather_JFK_02012019_01072025.xlsx')

#Concatenate the DataFrames using the 'date' column
JFK_Weather = pd.concat([JFK_Weather_2010_2019, JFK_Weather_2019_2025], axis=0, ignore_index=False)

#There should be no missing values, but just in case let´s handle them by filling NaN values
JFK_Weather.fillna(0, inplace=True)
```

```
#Let´s see the first few rows of the merged DataFrame
print(JFK_Weather.head())

#Saving the DataFrame to a new Excel File
JFK_Weather.to_excel('JFK_Weather.xlsx',index=False)

#-------------------------------------------------------------------------#

#Loading our LAX_Weather excel files into pandas DataFrames
LAX_Weather_2010_2019 = pd.read_excel('Weather_LAX_01012010_01012019.xlsx')
LAX_Weather_2019_2025 = pd.read_excel('Weather_LAX_02012019_01072025.xlsx')

#Concatenate the DataFrames using the 'date' column
LAX_Weather = pd.concat([LAX_Weather_2010_2019, LAX_Weather_2019_2025],
axis=0, ignore_index=False)

#There should be no missing values, but just in case let´s handle them by filling NaN
values
LAX_Weather.fillna(0, inplace=True)

#Let´s see the first few rows of the merged DataFrame
print(LAX_Weather.head())

#Saving the DataFrame to a new Excel File
LAX_Weather.to_excel('LAX_Weather.xlsx',index=False)
```

```
#-------------------------------------------------------------------------#

#Loading our ORD_Weather excel files into pandas DataFrames

ORD_Weather_2010_2019 =
pd.read_excel('Weather_ORD_01012010_01012019.xlsx')

ORD_Weather_2019_2025 =
pd.read_excel('Weather_ORD_02012019_01072025.xlsx')

#Concatenate the DataFrames using the 'date' column

ORD_Weather = pd.concat([ORD_Weather_2010_2019, ORD_Weather_2019_2025],
axis=0, ignore_index=False)

#There should be no missing values, but just in case let´s handle them by filling NaN
values

ORD_Weather.fillna(0, inplace=True)

#Let´s see the first few rows of the merged DataFrame

print(ORD_Weather.head())

#Saving the DataFrame to a new Excel File

ORD_Weather.to_excel('ORD_Weather.xlsx',index=False)
```

Appendix 2 – Formatting_Performance_Metrics.py (Self-made)

```
import os

import pandas as pd
```

```
# Folders to process
Folders = [
    r"C:\Users\Edu\Desktop\Base de Dados Tese\% On-time - Arrivals Database",
    r"C:\Users\Edu\Desktop\Base de Dados Tese\% On-time - Departures Database",
    r"C:\Users\Edu\Desktop\Base de Dados Tese\15 Minute or More Delay - Arrivals Database",
    r"C:\Users\Edu\Desktop\Base de Dados Tese\15 Minutes or More Delay - Departures Database",
    r"C:\Users\Edu\Desktop\Base de Dados Tese\NAS Delay Time (minutes) - Arrivals Database",
    r"C:\Users\Edu\Desktop\Base de Dados Tese\NAS Delay Time (minutes) - Departures Database",
    r"C:\Users\Edu\Desktop\Base de Dados Tese\Total Delay Time (minutes) - Arrivals Database",
    r"C:\Users\Edu\Desktop\Base de Dados Tese\Total Delay Time (minutes) - Departures Database",
]

for Folder in Folders:
    for File in os.listdir(Folder):
        if File.endswith(".xlsx") and not File.startswith("~$"):
            File_Path = os.path.join(Folder, File)

            try:
                # Read the first 9 rows (no headers)
                head = pd.read_excel(File_Path, header=None, nrows=9)
```

```
        # Get value from B1 and use it to replace B9
        b1_value = head.iloc[0, 1]  # Cell B1
        # The first 8 rows (B1 to B8) will be skipped in the final file

        # Read the rest of the file with actual headers (starting at row 9)
        df = pd.read_excel(File_Path, skiprows=8)

        # Keep only first 5532 rows (which corresponds to Excel row 5540)
        df = df.iloc[:5532]

        # Rename the second column using the B1 value
        if df.shape[1] > 1:
            old_col_name = df.columns[1]
            df.rename(columns={old_col_name: b1_value}, inplace=True)

        # Save the updated file
        df.to_excel(File_Path, index=False)
        print(f" Processed: {File}")

    except Exception as e:
        print(f" Error in {File}: {e}")
```

Appendix 3 – Merging.py (Self-made)

```
import pandas as pd
import os
```

```
# List of airport codes
airports = ['ATL', 'DEN', 'DFW', 'JFK', 'LAX', 'ORD']

# Base folder paths
base_path = "C:/Users/Edu/Desktop/Base de Dados Tese"
output_path = os.path.join(base_path, "BASES_DE_DADOS_COMBINADAS")

# Function to process one airport
def merge_airport_data(airport_code):
    try:
        weather = pd.read_excel(f"{base_path}/Weather Info/{airport_code}_Weather.xlsx")

        files = {
            "TotalDelayDepartures": f"Total Delay Time (minutes) - Departures Database/Total Delay (Minutes) from the Airline Performance - Departures Database ({airport_code}).xlsx",
            "TotalDelayArrivals": f"Total Delay Time (minutes) - Arrivals Database/Total Delay (Minutes) from the Airline Performance - Arrivals Database ({airport_code}).xlsx",
            "NASDelayDepartures": f"NAS Delay Time (minutes) - Departures Database/Departures - NAS Delay (Minutes) from the Airline Performance - Departures Database ({airport_code}).xlsx",
            "NASDelayArrivals": f"NAS Delay Time (minutes) - Arrivals Database/NAS Delay (Minutes) from the Airline Performance - Arrivals Database ({airport_code}).xlsx",
            "FifteenDelayDepartures": f"15 Minutes or More Delay - Departures Database/15 Minute or More Delay from the Airline Performance - Departures Database ({airport_code}).xlsx",
```

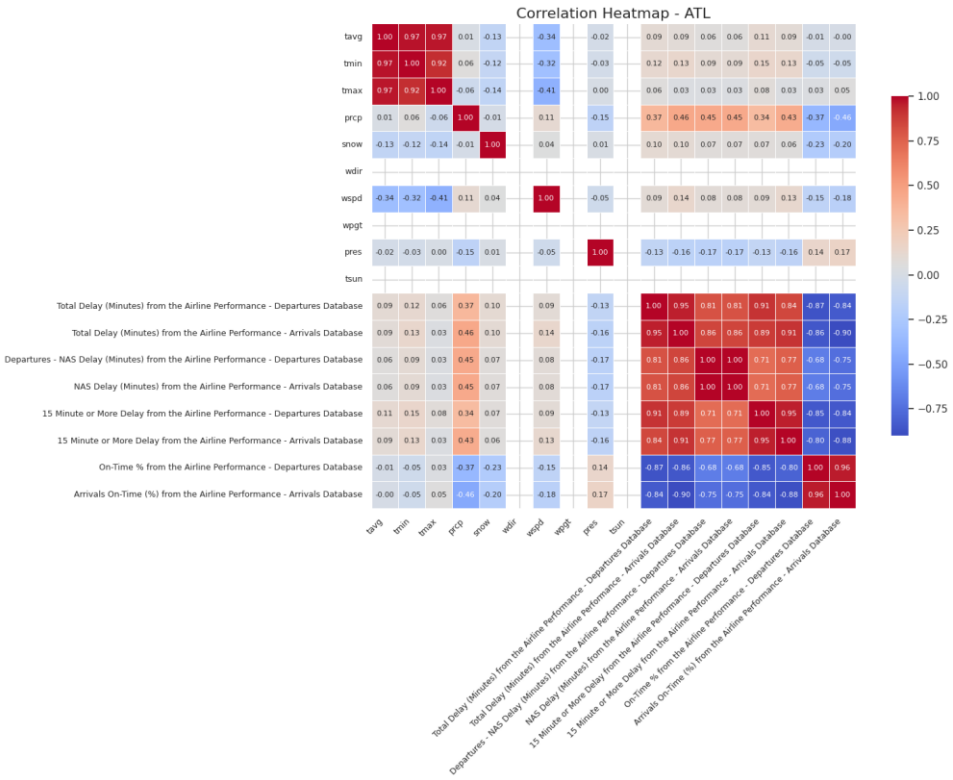
```
        "FifteenDelayArrivals": f"15 Minute or More Delay - Arrivals Database/15 Minute or More Delay from the Airline Performance - Arrivals Database ({airport_code}).xlsx",
        "PercentageOnTimeDepartures": f"% On-time - Departures Database/On-Time _ from the Airline Performance - Departures Database ({airport_code}).xlsx",
        "PercentageOnTimeArrivals": f"% On-time - Arrivals Database/Arrivals On-Time (_) from the Airline Performance - Arrivals Database ({airport_code}).xlsx",
    }

    dfs = {}
    for key, rel_path in files.items():
        full_path = os.path.join(base_path, rel_path)
        df = pd.read_excel(full_path)
        df.columns = df.columns.str.strip()  # Clean column names
        if "Time" in df.columns:
            df.rename(columns={"Time": "date"}, inplace=True)
        df['date'] = pd.to_datetime(df['date'], format="%m/%d/%Y", errors='coerce')
        df.dropna(subset=['date'], inplace=True)
        dfs[key] = df

    # Weather dataframe
    weather['date'] = pd.to_datetime(weather['date'], errors='coerce')
    weather.dropna(subset=['date'], inplace=True)

    # Merge all dataframes
    merged = weather
    for key in dfs:
        merged = pd.merge(merged, dfs[key], on='date', how='outer')
```

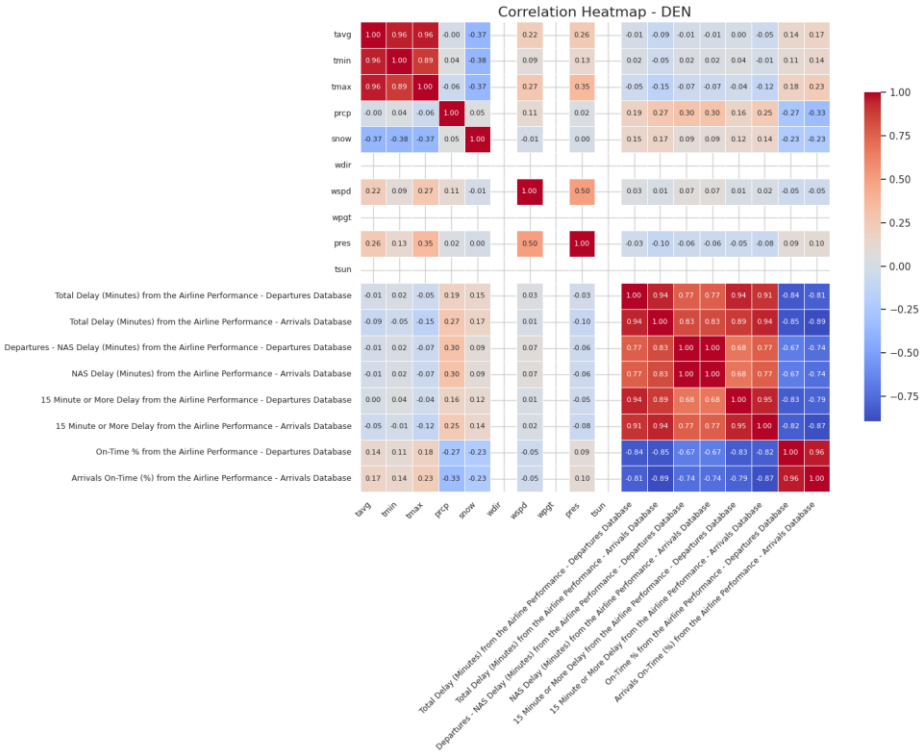
```
        merged.sort_values(by='date', inplace=True)

        # Save result
        output_file = os.path.join(output_path, f"{airport_code}.xlsx")
        merged.to_excel(output_file, index=False)
        print(f" {airport_code} merged and saved.")

    except Exception as e:
        print(f" Error processing {airport_code}: {e}")

# Run for all airports
for airport in airports:
    merge_airport_data(airport)
```

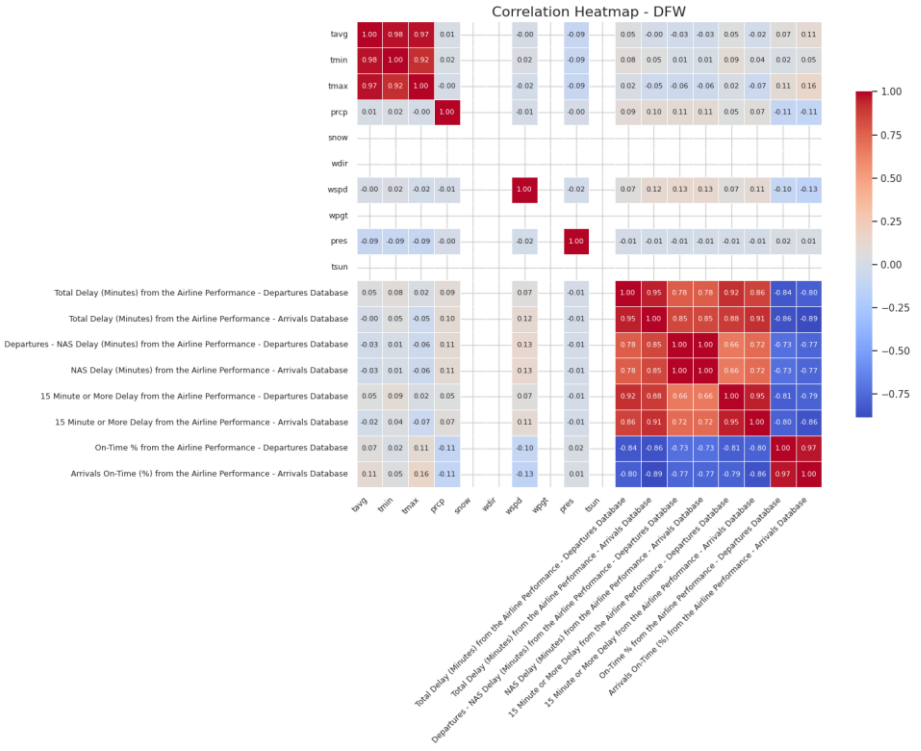
Appendix 4 – Python Code (Self-Made)

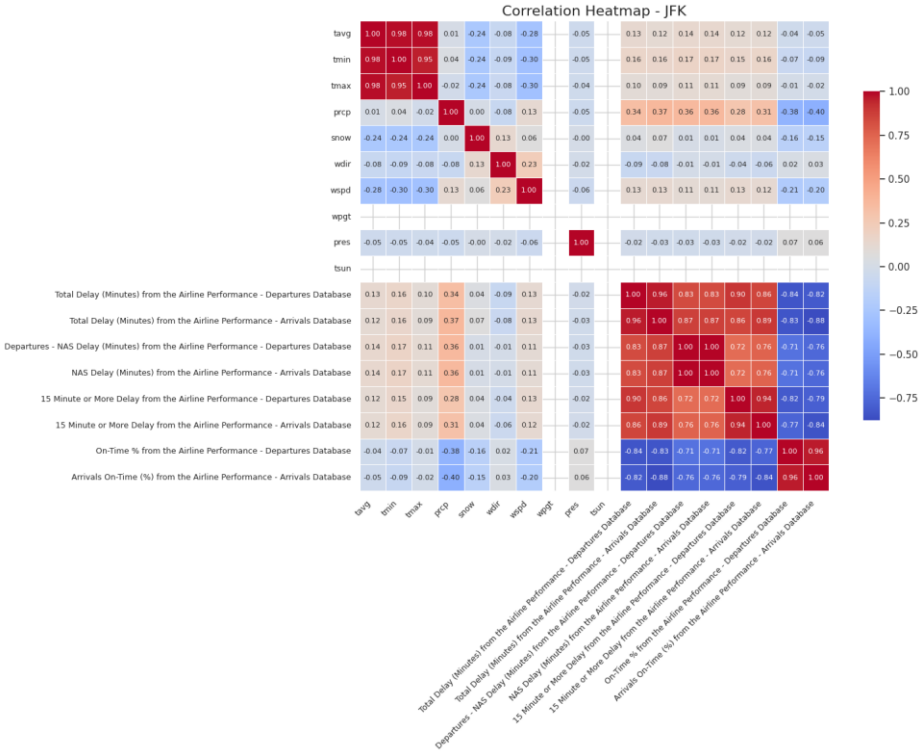
https://colab.research.google.com/drive/1uJKSMUKgyUb7jzKLvzCQ3x6W3_qPuPsn?usp=sharing (Accessible only as viewer)

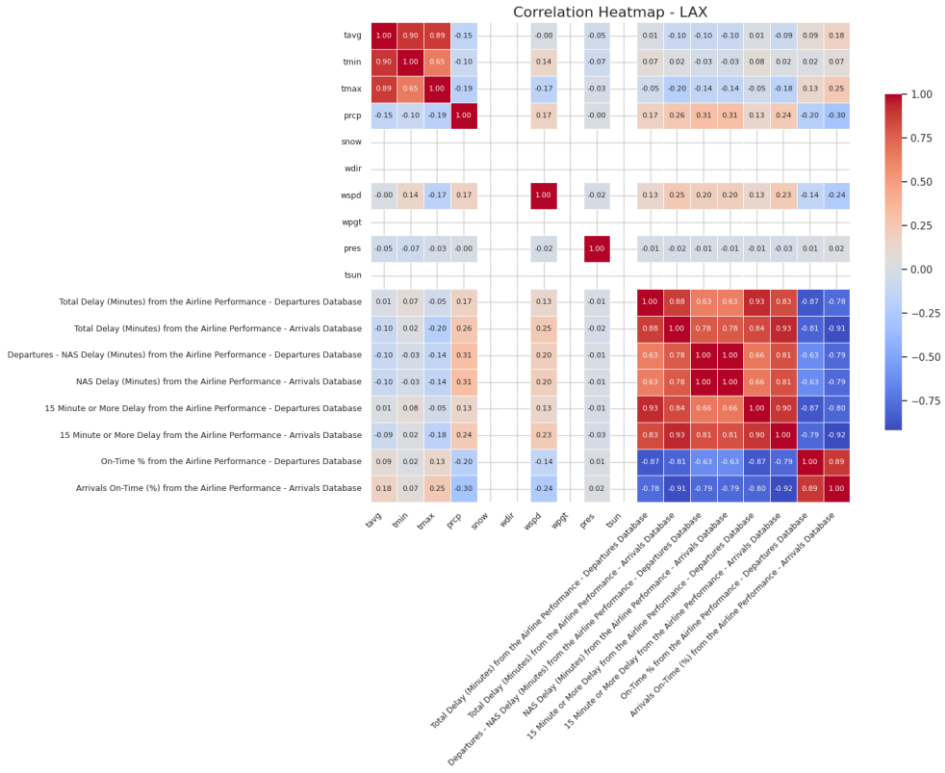
Appendix 5 – Correlation Heat Maps (Self-Made in Python)

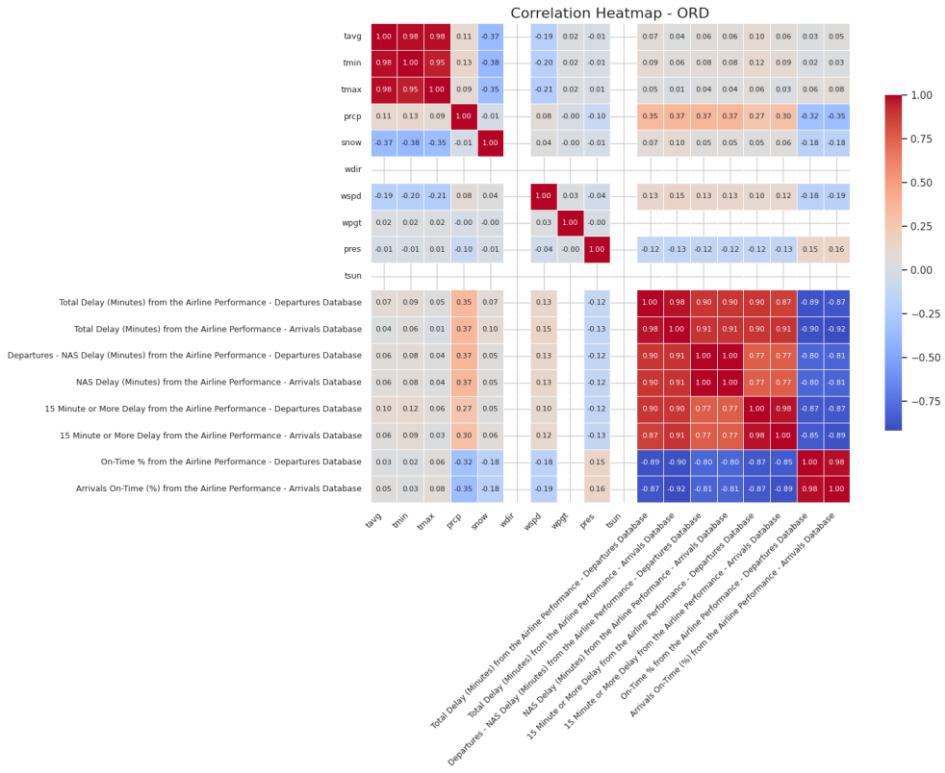

Correlation Heatmap - ATL
tavg
tmin
tmax
prcp
snow
wdir
wspd
wpgt
pres
tsun
Total Delay (Minutes) from the Airline Performance - Departures Database
Total Delay (Minutes) from the Airline Performance - Arrivals Database
Departures - NAS Delay (Minutes) from the Airline Performance - Departures Database
NAS Delay (Minutes) from the Airline Performance - Arrivals Database
15 Minute or More Delay from the Airline Performance - Departures Database
15 Minute or More Delay from the Airline Performance - Arrivals Database
On-Time % from the Airline Performance - Departures Database
Arrivals On-Time (%) from the Airline Performance - Arrivals Database
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75


Correlation Heatmap - DEN
tavg
tmin
tmax
prcp
snow
wdir
wspd
wpgt
pres
tsun
Total Delay (Minutes) from the Airline Performance - Departures Database
Total Delay (Minutes) from the Airline Performance - Arrivals Database
Departures - NAS Delay (Minutes) from the Airline Performance - Departures Database
NAS Delay (Minutes) from the Airline Performance - Arrivals Database
15 Minute or More Delay from the Airline Performance - Departures Database
15 Minute or More Delay from the Airline Performance - Arrivals Database
On-Time % from the Airline Performance - Departures Database
Arrivals On-Time (%) from the Airline Performance - Arrivals Database
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75


Correlation Heatmap - DFW
tavg
tmin
tmax
prcp
snow
wdir
wspd
wpgt
pres
tsun
Total Delay (Minutes) from the Airline Performance - Departures Database
Total Delay (Minutes) from the Airline Performance - Arrivals Database
Departures - NAS Delay (Minutes) from the Airline Performance - Departures Database
NAS Delay (Minutes) from the Airline Performance - Arrivals Database
15 Minute or More Delay from the Airline Performance - Departures Database
15 Minute or More Delay from the Airline Performance - Arrivals Database
On-Time % from the Airline Performance - Departures Database
Arrivals On-Time (%) from the Airline Performance - Arrivals Database
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75


Correlation Heatmap - JFK
tavg
tmin
tmax
prcp
snow
wdir
wspd
wpgt
pres
tsun
Total Delay (Minutes) from the Airline Performance - Departures Database
Total Delay (Minutes) from the Airline Performance - Arrivals Database
Departures - NAS Delay (Minutes) from the Airline Performance - Departures Database
NAS Delay (Minutes) from the Airline Performance - Arrivals Database
15 Minute or More Delay from the Airline Performance - Departures Database
15 Minute or More Delay from the Airline Performance - Arrivals Database
On-Time % from the Airline Performance - Departures Database
Arrivals On-Time (%) from the Airline Performance - Arrivals Database
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75


Correlation Heatmap - LAX
tavg
tmin
tmax
prcp
snow
wdir
wspd
wpgt
pres
tsun
Total Delay (Minutes) from the Airline Performance - Departures Database
Total Delay (Minutes) from the Airline Performance - Arrivals Database
Departures - NAS Delay (Minutes) from the Airline Performance - Departures Database
NAS Delay (Minutes) from the Airline Performance - Arrivals Database
15 Minute or More Delay from the Airline Performance - Departures Database
15 Minute or More Delay from the Airline Performance - Arrivals Database
On-Time % from the Airline Performance - Departures Database
Arrivals On-Time (%) from the Airline Performance - Arrivals Database
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75


Correlation Heatmap - ORD
tavg
tmin
tmax
prcp
snow
wdir
wspd
wpgt
pres
tsun
Total Delay (Minutes) from the Airline Performance - Departures Database
Total Delay (Minutes) from the Airline Performance - Arrivals Database
Departures - NAS Delay (Minutes) from the Airline Performance - Departures Database
NAS Delay (Minutes) from the Airline Performance - Arrivals Database
15 Minute or More Delay from the Airline Performance - Departures Database
15 Minute or More Delay from the Airline Performance - Arrivals Database
On-Time % from the Airline Performance - Departures Database
Arrivals On-Time (%) from the Airline Performance - Arrivals Database
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75