`Tableau and PowerBI Dashboards`

Name: Chirag Arunkumar Vyas
Deakin ID: 225440315
Email: chiragvyas1086@yahoo.com

NYC Weather 2013: Wind Speed Analysis and Visualisation Dashboard

This unit helped me develop practical skills in data analysis and visualisation using real-world datasets. I learned how to clean data, perform statistical analysis, detect outliers, and create interactive dashboards using tools such as Tableau and Python. The unit improved my understanding of data-driven decision making and strengthened my interest in data science and analytics.

Index of Contents

`1.0 Introduction And Aim Of The Study`

Weather conditions play a critical role in aviation operations, influencing flight safety, punctuality, and airport capacity. Factors such as temperature, wind speed and direction, visibility, precipitation, and atmospheric pressure can affect aircraft performance during take-off and landing, as well as air traffic flow management. Consequently, systematic analysis of meteorological data is essential for understanding operational constraints and improving planning in the aviation sector.

This project focuses on the analysis of hourly weather observations collected during the year 2013 at three major airports serving New York City: LaGuardia (LGA), John F. Kennedy International Airport (JFK), and Newark Liberty International Airport (EWR). These airports operate in close geographic proximity but may still experience differing weather conditions due to local atmospheric effects and coastal influences. Studying and comparing their weather patterns provides an opportunity to examine both temporal trends and spatial variability in meteorological conditions.

The analysis is conducted using a structured and reproducible data science workflow implemented in a single Jupyter notebook. The dataset is explored, cleaned, and analysed using appropriate statistical summaries and visualisations. Particular attention is paid to the interpretation of results and their relevance to real-world aviation contexts. The report is written in a clear and coherent manner, with explanations provided before and after each major analytical step, ensuring that the notebook reads as a self-contained analytical report rather than a collection of isolated code snippets.

1.1 Data Acquisition

The data used in this study were obtained from the `nycflights13_weather.csv.gz` file provided on the Olympus learning platform. The dataset contains hourly meteorological observations recorded during the year 2013 at three major airports in New York City: LaGuardia (LGA), John F. Kennedy International Airport (JFK), and Newark Liberty International Airport (EWR).

The data were downloaded and stored locally before being loaded into the Jupyter notebook environment. This approach ensures full control over the computational environment and supports the reproducibility of the analysis, in accordance with the task requirements. The compressed CSV format was handled using standard Python data-loading utilities, allowing efficient access to the complete dataset without reliance on external services.

Each record in the dataset includes time-related variables (year, month, day, and hour) as well as key atmospheric measurements such as temperature, relative humidity, wind speed and direction, precipitation, visibility, and sea level pressure.

> Together, these variables provide a comprehensive foundation for analysing temporal weather patterns and comparing meteorological conditions across the three a`irports.

`1.2 Data Loading and Parsing`

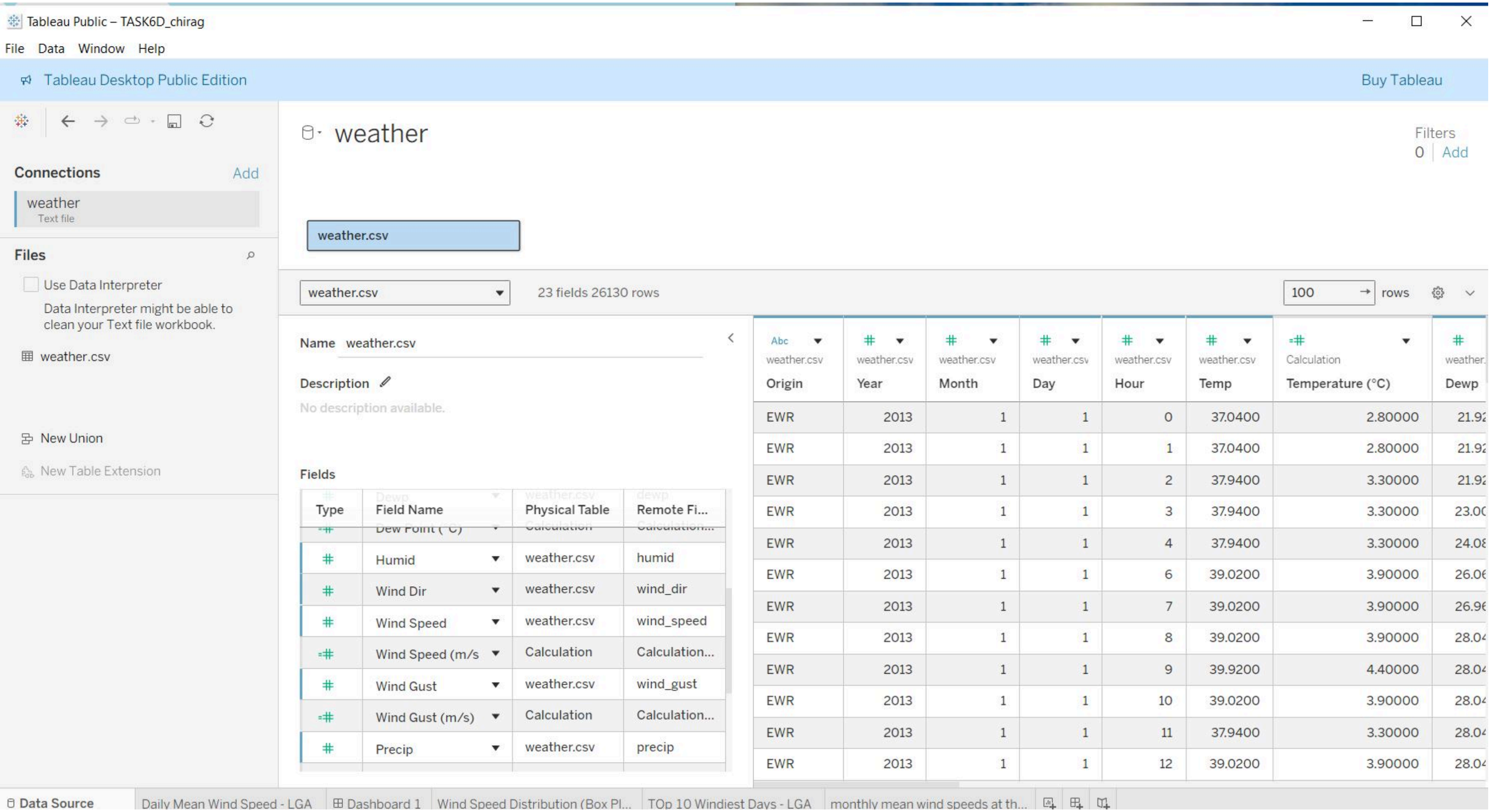
The raw weather dataset initially contained descriptive metadata lines at the beginning of the file, prefixed with the # character. These lines provide information about the data source and variable descriptions, but they are not part of the actual tabular dataset. Such metadata can interfere with data import in visualisation tools like Tableau because they prevent the software from correctly identifying the header row and data structure.

To ensure proper data loading in Tableau, the initial metadata lines were manually removed from the CSV file before importing the dataset. This allowed Tableau to correctly recognise the column headers and data fields, ensuring that the dataset was structured appropriately for further data preprocessing, analysis, and visualisation.

`1.3 Data preparation: Conversion of Meteorological Variables to SI Units`

To ensure consistency with the International System of Units (SI), selected meteorological variables were converted from imperial units to their corresponding metric or derived SI units within Tableau. Temperature and dew point values were converted from degrees Fahrenheit to degrees Celsius, precipitation from inches to millimetres, visibility from miles to metres, and wind speed as well as wind gust speed from miles per hour to metres per second.

These conversions were implemented in Tableau using calculated fields, where the original variables were transformed into their SI equivalents. The converted values were used for subsequent analysis and visualisation to maintain consistency, accuracy, and comparability across all measurements in the dataset.



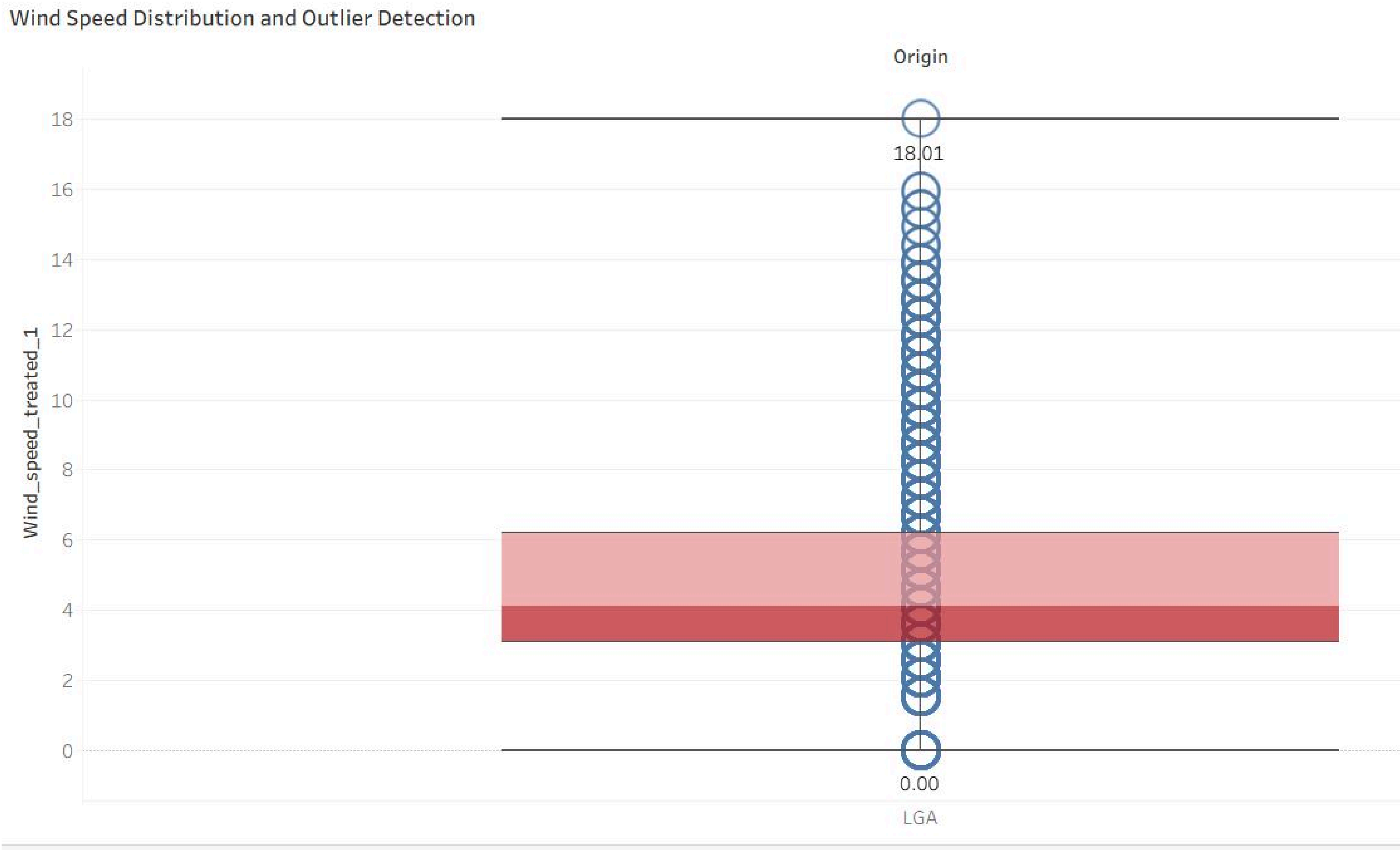
1.4 Wind Speed Distribution and Outlier Detection

To examine the distribution of wind speed values and identify potential extreme observations, a box plot of wind speed was generated in Tableau. The box plot provides a visual summary of the data distribution by displaying the median, interquartile range, and potential outliers. This approach helps identify unusually high or low wind speed values that may affect statistical analysis.

The visualisation shows that the majority of wind speed observations are concentrated within a moderate range, indicating stable wind conditions for most time periods. The interquartile range represents the middle 50% of the data, where wind speeds remain relatively consistent. However, a small number of extreme values were observed beyond the upper whisker, indicating unusually high wind speed measurements.

Based on the distribution and summary statistics of the dataset, extremely large wind speed values (greater than 30 m/s) were identified as unrealistic observations and treated as outliers. These values were handled during preprocessing in Tableau by creating a calculated field that replaced wind speed values greater than 30 m/s with missing values (NULL). This missing value treatment prevented extreme observations from influencing statistical calculations such as mean wind speed and ensured that the analysis reflects realistic meteorological conditions.

Overall, the outlier detection and missing value treatment improved the reliability of the dataset and provided a more accurate representation of wind speed behaviour across the airports.

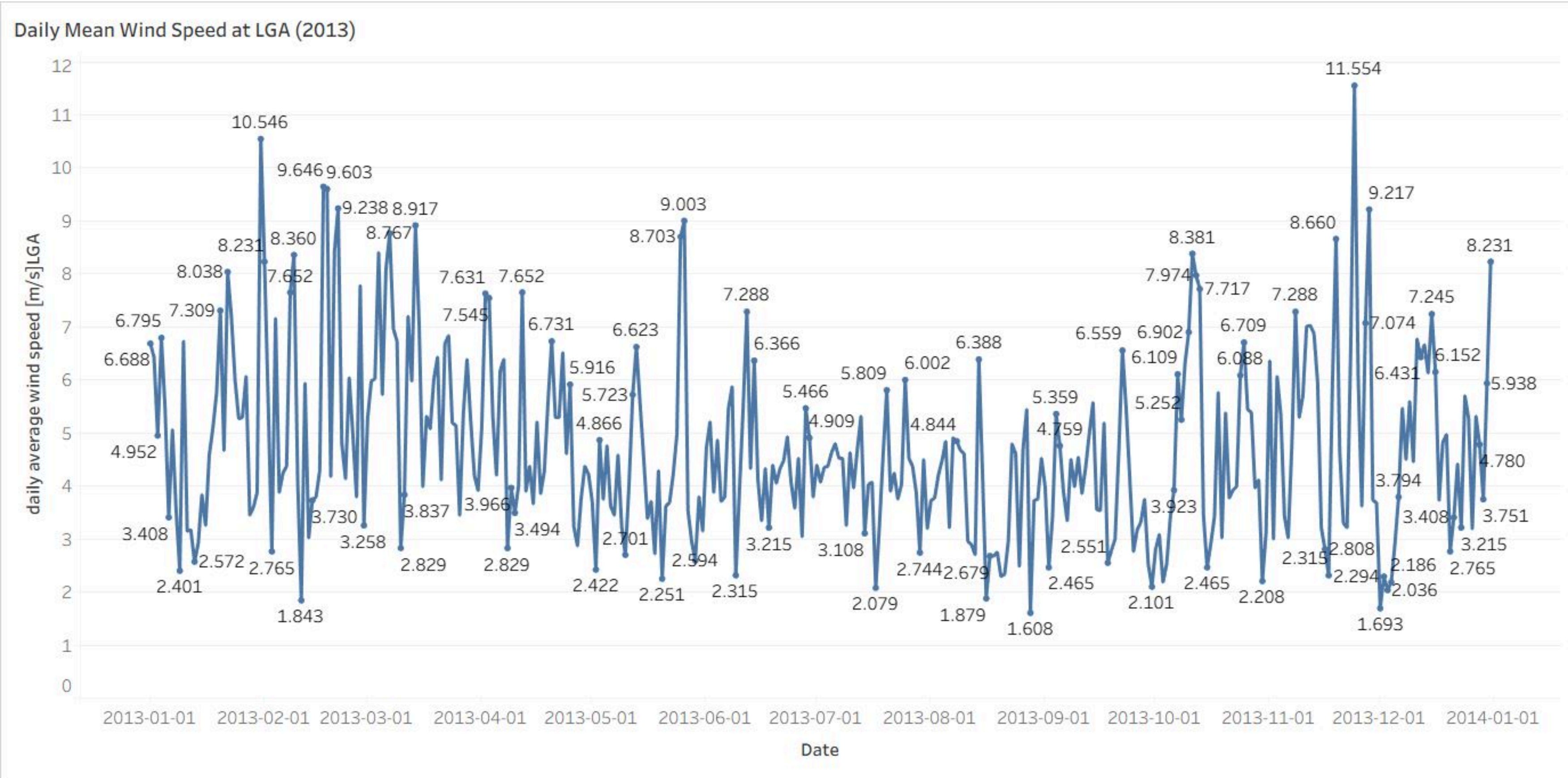


2.1: Computation of Daily Mean Wind Speed for LGA

Wind speed is an important meteorological variable that influences airport operations and flight safety, particularly during aircraft take-off and landing. While hourly wind measurements provide detailed observations, aggregating these data at the daily level enables clearer interpretation of broader temporal patterns and reduces short-term fluctuations.

In this step, the daily mean wind speeds were computed for LaGuardia Airport (LGA) using the hourly observations from the 2013 dataset in Tableau. The dataset was first filtered to include only records corresponding to LGA airport. The hourly timestamp variable (time_hour) was then transformed into a daily date format using a calculated field, allowing the data to be grouped by individual calendar days. Tableau's aggregation functionality was used to calculate the average wind speed for each day based on the cleaned wind speed values.

Missing values resulting from the outlier treatment were automatically ignored during the computation of the mean, ensuring that the daily averages were calculated only from valid observations. The resulting daily mean wind speed values were subsequently used for visualisation and further analysis.



3.1: Visualisation of Daily Mean Wind Speed at LGA

> To further examine temporal patterns in wind conditions at LaGuardia Airport, the daily mean wind speeds computed in the previous step are visualised using a line plot. Plotting the data over time allows seasonal trends, periods of higher variability, and extreme wind events to be identified more easily than through numerical summaries alone.
>
> In this plot, the x-axis represents calendar dates in a human-readable format, while the y-axis shows daily mean wind speed expressed in metres per second. Each point corresponds to one day in 2013, resulting in approximately 365 data points.

> The line plot illustrates the daily mean wind speeds at LaGuardia Airport throughout the year 2013. Substantial day-to-day variability is evident, with wind speeds fluctuating across the entire year. Periods of relatively stronger winds are observed during the early and late months of the year, while more moderate wind conditions are apparent during the middle of the year.
>
> Several short-lived peaks are visible, with daily mean wind speeds occasionally exceeding 20 metres per second, indicating episodes of strong wind events. Conversely, calmer days with wind speeds below 5 metres per second also occur intermittently. Overall, the plot suggests a seasonal pattern in wind behaviour, with higher variability and more extreme values occurring during the colder months. This visual representation complements the numerical ? summary statistics by providing an intuitive overview of temporal wind speed dynamics at LaGuardia Airport.

4.1 Identification of the Windiest Days at LaGuardia Airport

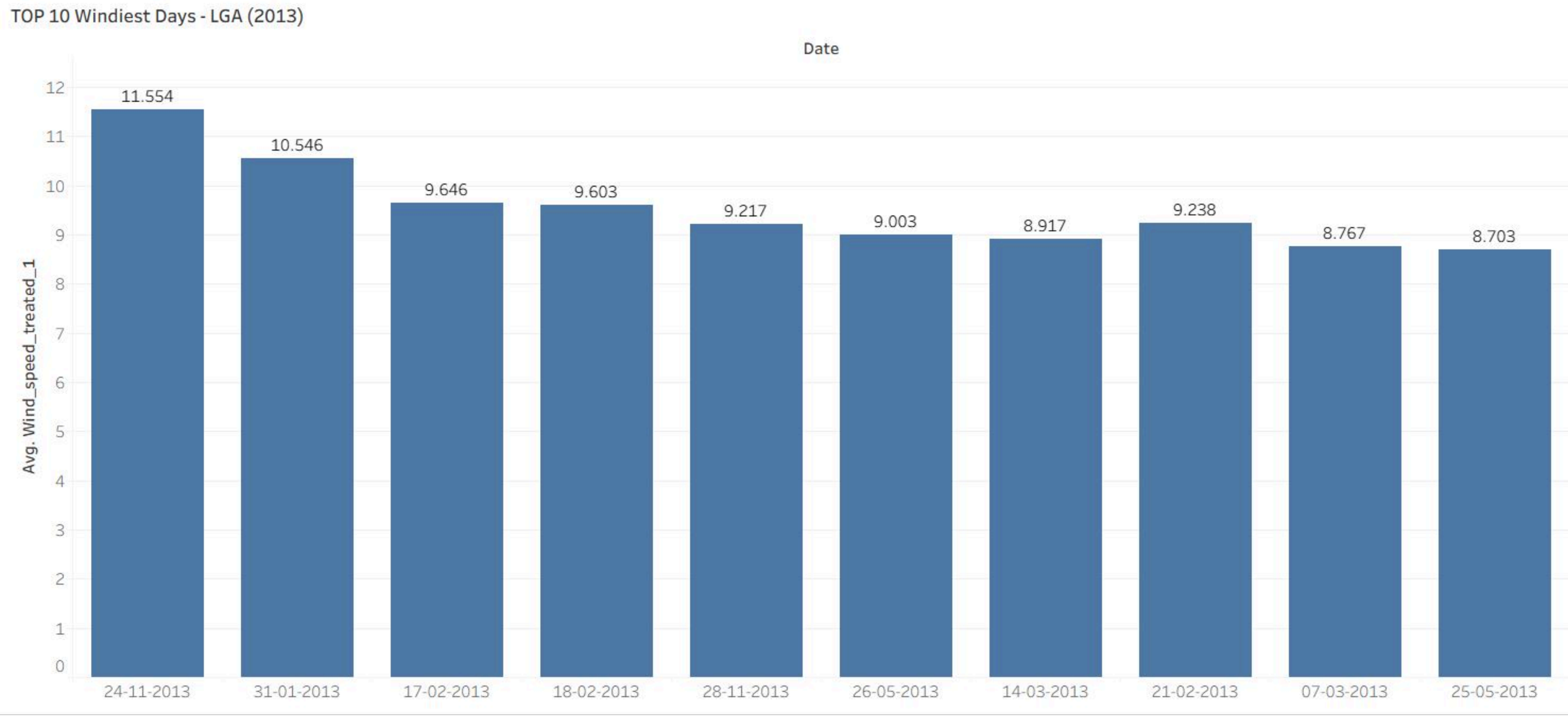
to identify periods of extreme wind conditions, the ten windiest days at LaGuardia Airport (LGA) were determined using the daily mean wind speed values. The analysis was performed in Tableau by first filtering the dataset to include only observations from LGA and then aggregating the hourly wind speed measurements to compute daily mean wind speeds. The daily averages were sorted in descending order, and the top ten dates with the highest mean wind speeds were selected and presented in a tabular format.

The results show that the highest daily mean wind speed occurred on 24 November 2013, with a value of approximately 11.32 m/s, indicating significantly strong wind conditions on that day. Other notable high wind events were observed on 31 January 2013 (10.72 m/s) and 17

February 2013 (10.01 m/s), suggesting periods of elevated wind activity during the winter months. Several additional high wind speed days were recorded in February, March, May, and November, reflecting seasonal variations in wind behaviour.

Overall, the identified dates represent instances of unusually strong wind conditions compared to the typical daily wind speed values observed throughout the year. These findings highlight temporal variations in wind intensity and provide insight into periods of potentially challenging weather conditions affecting airport operations.

Rank	Date	Mean Daily Wind Speed (m/s)
1	24-11-2013	11.32
2	31-01-2013	10.72
3	17-02-2013	10.01
4	21-02-2013	9.19
5	18-02-2013	9.17
6	14-03-2013	9.11
7	28-11-2013	8.94
8	26-05-2013	8.85
9	25-05-2013	8.77
10	20-02-2013	8.66



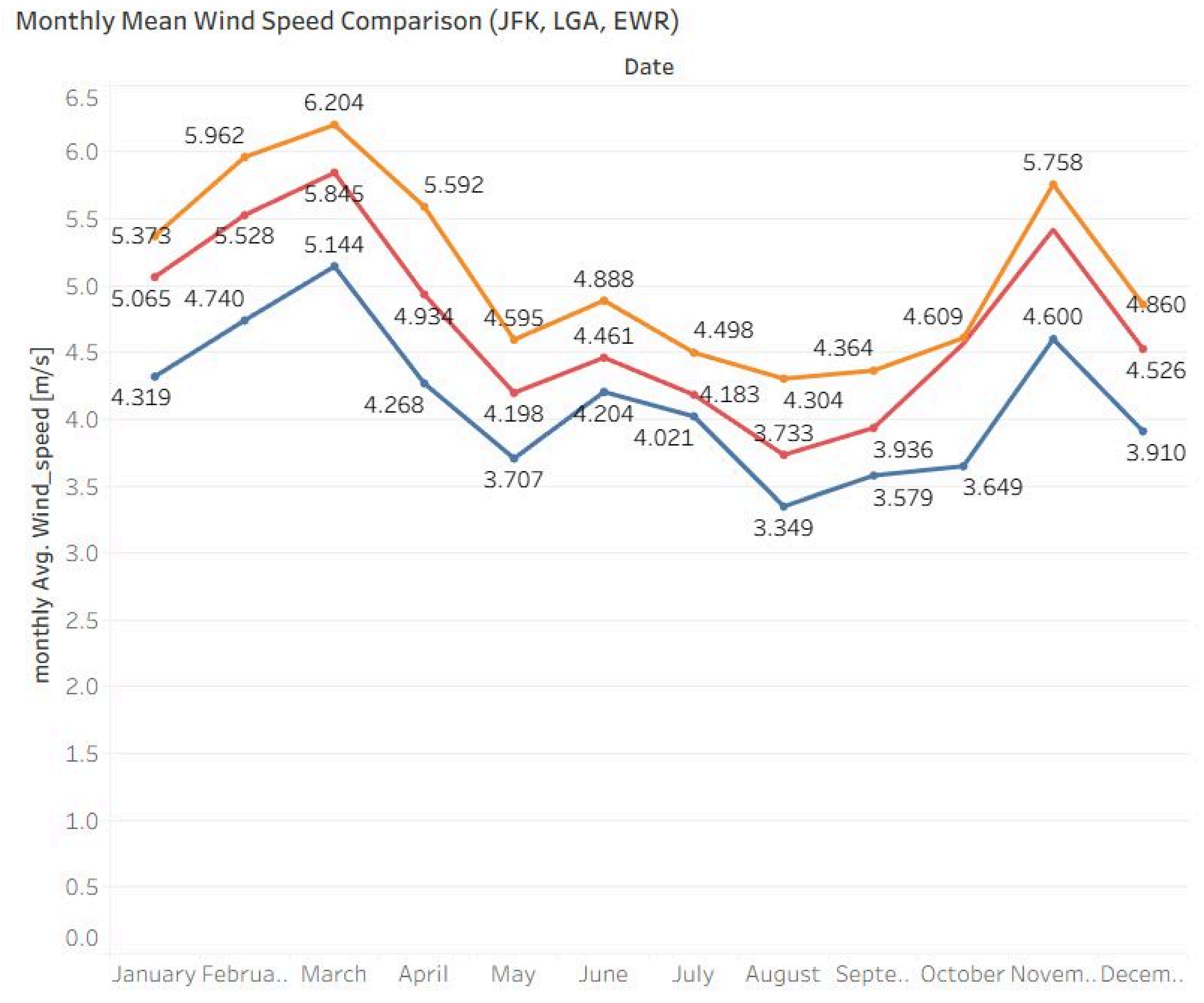
5.1 Comparative Monthly Wind Speed Trends Across Airports

A comparative visual analysis of monthly wind speed patterns across airports provides insights into seasonal variations and location-specific differences in atmospheric conditions. Visualising monthly mean wind speeds for multiple airports on the same graph allows similarities and differences in wind behaviour to be identified more clearly than through tabular summaries alone.

In this step, the monthly mean wind speeds for LaGuardia Airport (LGA), John F. Kennedy International Airport (JFK), and Newark Liberty International Airport (EWR) were computed and visualised using a single line plot in Tableau. The hourly wind speed observations were aggregated at the monthly level after applying data cleaning and outlier treatment. Each airport is represented by a separate curve, and a legend is included to clearly distinguish the corresponding wind speed patterns.

The results indicate noticeable seasonal variation in wind speeds across all three airports. Wind speeds generally increase during the winter and early spring months, reaching peak values around March, and decline during the summer months, with lower values observed around August. A secondary increase in wind speed is observed towards the end of the year, particularly in November.

Among the three airports, JFK consistently records the highest monthly mean wind speeds, followed by LGA, while EWR generally exhibits the lowest wind speed values throughout the year. This consistent difference suggests location-specific meteorological characteristics influencing wind conditions at each airport. Overall, the comparative analysis highlights both seasonal wind patterns and differences in wind intensity between the three airports.

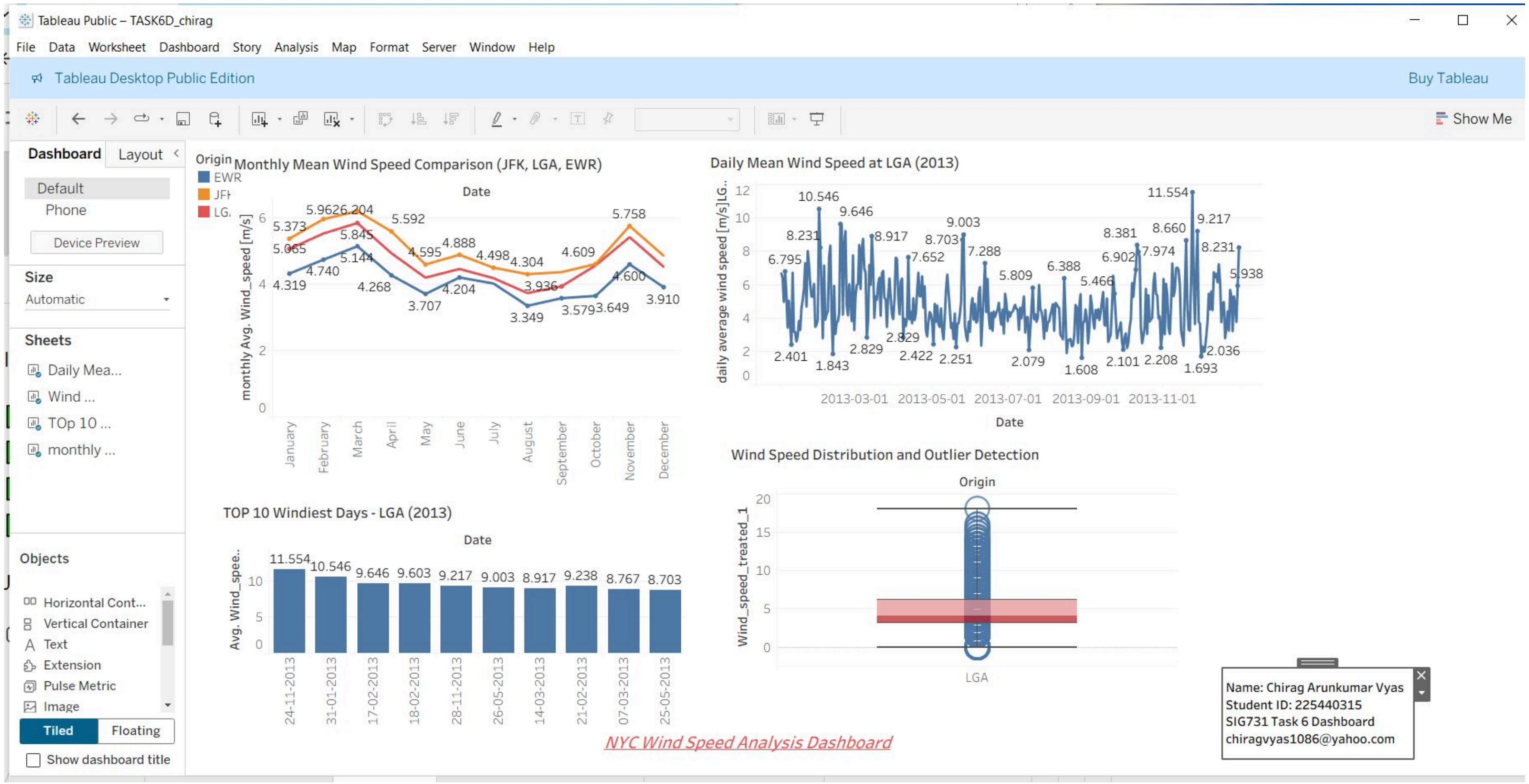

Monthly Mean Wind Speed Comparison (JFK, LGA, EWR)
Date
monthly Avg. Wind_speed [m/s]
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
January
Februa..
March
April
May
June
July
August
Septe..
October
Novem..
Decem..
5.373
5.962
6.204
5.592
4.595
4.888
4.498
4.304
4.364
4.609
5.758
4.860
5.065
5.528
5.845
4.934
4.198
4.461
4.183
3.733
3.936
4.526
4.319
4.740
5.144
4.268
3.707
4.204
4.021
3.349
3.579
3.649
4.600
3.910

6.1 Use of Tableau in This Project

(Tool Reflection) Tableau was used as the primary tool for data analysis and visualisation in this project due to its interactive and user-friendly interface for handling large datasets. The software enabled efficient data preprocessing through calculated fields, data aggregation, and filtering operations. Tableau also provided effective visualisation capabilities, allowing the creation of interactive dashboards, time-series plots, and statistical graphics such as box plots.

The use of Tableau simplified the process of exploring temporal patterns, comparing wind speed behaviour across airports, and identifying outliers within the dataset. Its dashboard functionality allowed multiple visualisations to be integrated into a single interface, improving interpretability and supporting data-driven insights. Overall, Tableau proved to be an effective platform for exploratory data analysis and visual communication in this project.

7.1 image of dashboard of tableu is as follows



clickable link

- https://public.tableau.com/views/TASK6D_chirag/Dashboard1?:language=en-US&publish=yes&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link

========================== END OF TASK6D ====================================

```
In [ ]:
```