**DSA210 Project Final Report**

**Motivation: Why are you working on this project?**

My main motivation for working on this project stems from my aspiration to become a pilot. I have always been deeply interested in aviation, particularly how external conditions like weather affect flight operations. Understanding the dynamics between weather conditions and flight activity is crucial in aviation safety and efficiency, which makes this project both relevant and personally meaningful.

**Data Source: Where did you get this data? How did you collect it?**

This project uses two types of data:

* **Flight data** was collected for both Istanbul Airport (IST) and Sabiha Gökçen Airport (SAW) using FlightRadar24, a publicly available source. The data includes scheduled and actual flight counts on a daily basis.
* **Weather data** was collected from Visual Crossing Weather, which provides historical weather information. The dataset includes variables such as temperature, humidity, wind speed, wind gust, precipitation, cloud cover, visibility, and a general weather condition description.

All data spans from February 1st to mid-May 2025 and was preprocessed using Python (Pandas and NumPy) for consistency and analysis.

**Data Analysis: Techniques used, different stages of the analysis**

The data analysis was conducted in several stages:

1. **Data Preprocessing**:
   * Cleaning and formatting date columns.
   * Merging weather and flight datasets by date.
   * Creating additional boolean features (e.g., rainy, snowy, windy) for classification-based analysis.
2. **Exploratory Data Analysis (EDA)**:
   * Correlation matrices and heatmaps were used to explore relationships between weather variables and flight data.
   * Scatter plots were generated to visualize the relationship between specific weather features (e.g., wind speed, temperature) and the number of actual flights.
3. **Statistical Testing**:
   * Paired t-tests were applied to test differences in flight numbers under various weather conditions (e.g., rainy vs. clear days).
4. **Machine Learning**:
   * Predictive models were built to estimate actual flight counts based on weather conditions using:
     + **Linear Regression**
     + **Random Forest Regressor**
     + **XGBoost Regressor**
   * Performance was evaluated using metrics like RMSE (Root Mean Squared Error) and R² score.

**Findings: What are the interesting findings that you found in this project?**

* Weather conditions, especially precipitation and snow, have a noticeable impact on flight operations. Both Istanbul airports showed reduced actual flights and increased deviation from scheduled flights on days with rain or snow.
* Temperature and cloud cover showed weaker correlations, suggesting they are less critical individually in affecting flight counts.
* Wind speed had a less conclusive effect on flight operations. While high wind days showed some disruption, statistical tests and visualizations did not reveal a consistent pattern strong enough to confirm a solid relationship.
* Machine learning models, particularly Random Forest and XGBoost, outperformed Linear Regression, suggesting that non-linear relationships play an important role in predicting flight numbers based on weather.

**Limitations and Future Work: What could be done better? Do you have any future plans about your project?**

* **Limitations**:
  + The dataset does not include wind direction, which is a crucial factor for determining crosswinds, a known disruptor of airport operations.
  + Delay durations and cancellation data were not available, which limited the scope of impact analysis.
  + Only two airports from the same city were considered, which may reduce the generalizability of the conclusions.
* **Future Work**:
  + Future versions of the project could include wind direction data to investigate the effects of crosswinds on flight schedules.
  + Expanding the dataset to include delay durations, cancellation reasons, and more airports would allow for broader and deeper analysis.
  + Further improvement could be made by integrating weather severity indexes or aviation-specific weather warnings to enhance prediction models.