**DSA 210 Project Proposal**

**How Weather Conditions Affect Flight Cancellations**

**1. Why This Project?**

Flight cancellations can be frustrating for passengers and costly for airlines. While some cancellations are due to operational issues, weather conditions play a major role in determining whether a flight takes off as planned. This project aims to explore how different weather factors contribute to flight cancellations. By analyzing planned vs tracked flights alongside weather data, I hope to uncover key patterns and even build a model to predict cancellations in advance.

**2. Where Will the Data Come From?**

I will use publicly available datasets that provide:

* **Flight Data:** Information on planned vs. tracked flights from official airport records.
* **Weather Data:** Temperature, wind speed, precipitation, humidity, visibility, and cloud cover from sources like NOAA or OpenWeather API.

By merging these datasets based on time and location (airport codes), we can examine how different weather conditions influence flight cancellations.

**3. How Will We Collect the Data?**

* Gather flight data from online repositories or airport databases.
* Obtain historical weather data for the same timeframes and locations.
* Clean and preprocess the datasets to ensure they align correctly.

**4. What’s the Plan for Analysis?**

Our approach includes several steps:

1. **Exploratory Data Analysis (EDA):**
   * Identify trends and patterns in cancellations.
   * Visualize how weather conditions vary over time.
   * Measure correlations between specific weather events and flight cancellations.
2. **Hypothesis Testing:**
   * Use statistical methods to confirm which weather conditions are the strongest predictors of cancellations.
3. **Building a Predictive Model:**
   * Train a machine learning model (e.g., Logistic Regression, Random Forest, or XGBoost) to estimate the likelihood of flight cancellations based on weather conditions.
   * Evaluate model accuracy using metrics like Precision, Recall, and F1-score.
4. **Visualizing the Findings:**
   * Use heatmaps, scatter plots, and bar charts to make insights clear and easy to understand.

**5. What Do We Expect to Find?**

* Discover which weather conditions (such as heavy rain, snow, or strong winds) are most likely to cause flight cancellations.
* Identify peak times or seasons when cancellations are more frequent.
* Develop a model that predicts cancellations based on real-time weather data.

**6. Limitations and Future Possibilities**

* Some cancellations may be due to operational issues, not just weather, which might affect our model’s accuracy.
* Some weather data might be incomplete or unavailable for certain flights.
* Future work could involve expanding the model to multiple airports and incorporating live weather updates for real-time cancellation predictions.

**7. Tools & Technologies**

* **Programming Language:** Python
* **Libraries:** Pandas, NumPy, Seaborn, Matplotlib, Scikit-learn
* **Deployment:** A GitHub repository containing all code, a README.md for instructions, and a requirements.txt file for dependencies.

**8. Project Timeline**

* **March 14:** Submit project proposal on GitHub.
* **April 18:** Collect and clean the data, perform initial analysis.
* **May 23:** Train and evaluate the machine learning model.
* **May 30:** Submit the final project.

By the end of this project, I hope to provide insights that help airlines and passengers better prepare for potential disruptions caused by weather conditions.