Proposal 19 : Analysis of Airline Delay and Cancellation

Problem:
Analyse flight delays and cancellations to develop a scoring system.

Motivation and Real World Application:
By offering passengers with more information about flights like potential delaying time and the likelihood of cancellation, passengers can schedule trips depending on prior data to reduce or avoid the aforementioned issues. In addition, analyzing the data for different airlines and airports can help passengers estimate a more suitable time or season to travel.

These are the concerns we would like to address:

1. By analysing airline information from previous years, can we identify any trends between flight delays/cancellations and airlines? Are there other factors, such as destinations or weather, that cause these issues?
2. During inevitable conditions, such as weather conditions, delays or cancellations are undeniable. Which parameters cause the most impact on delays/cancellations?
3. Based on our analysis, can we provide a scoring system such that customers could make use of the most suitable airline provided depending on their tolerance for delay/cancellations? (E.g. Are they willing to risk a 2-hour delay vs 6-hour delay?)

Dataset:
The original source of our dataset is on the website of Bureau Transposition Statistics (https://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=236&DB_Short_Name=On-Time). It contains the on-time performance of USA domestic airlines from 1987 to present. It contains airline identifier, airports identifier, information about flight cancellations and flight delays, how long is the delay, reasons that caused the delay or cancellation, etc.

Proposed Solution
There are two sections: 1. a rough overview of flight performance depending on parameters to give an “absolute ranking” of airlines, and 2. a detailed analysis that allows customers to make decisions based on their preferences.

1. Rough overview of airlines: We hope to identify and quantify any correlations between delays and cancellations with airlines, airports, travel seasons, geographical locations and other parameters by statistical analysis.
2. Detailed analysis of each airline: The next layer of analysis will be dependent on the finer details: e.g. delay intervals (e.g. every 15 min interval), cancellation, number of diverted landings, and delays due to diverted flights to create a scoring system, weighted by these parameters. We hope this detailed analysis can allow the audience to make their own cost-benefit analysis, instead of focusing on the "absolute ranking in step 1".
3. If time allows, we would like to make use of a predictive model to identify the best airline for a specific season or destination. This can potentially be done with a regression model.

Project steps

Step	Estimated completion time
Data extraction	1 week
Data analysis for rough overview of airlines ranking and visualization	1 week
Parameter visualization for each airline e.g. intervaled delay times, flight diversions and associated delays	1-2 weeks
Correlation analysis between delay/cancellation with parameters e.g. airline, weather/seasons, location	1-2 weeks
Develop scoring system for passengers based on tolerance for delays/cancellations	1 week
Potential data prediction: regression analysis, parameter determination and scoring	If time allows