**How to reliably predict airline delays**

Business Problem, Opportunity, and Impact

Overview of industry, business, or problem

The airline industry provides air transportation for passengers and cargo by using aircraft. Flight delays and cancellations can significantly affect customer experience and cause financial loss to the business. The total cost of delays from 2016-2019 was over US$23 billion.

Define the specific problem that should be solved

Assist customers in making better decisions when booking flights and help businesses mitigate financial loss due to delays.

* Predict flights that will be delayed allowing passengers to avoid those flights if they chose.
* Allow businesses to do root-cause analysis on flights with habitual delays as well as apply mitigation strategies to avoid negative customer experiences related to delays.

How to accomplish the problem

* Analyze flight delay data: detecting seasonality, airport and connection flights patterns and any correlation.

Why does this problem matter?

Delays have costs airlines over US$23 billion. Additionally, negative customer experiences further erode a company’s reputation and future profit by loss of business.

Potential Audience

* Anyone who is booking flight arrangements (for self or others) would be interested. This information would allow them to make informed decisions regarding flights and/or airports.
* Airline executives could benefit by identifying frequently delayed flights and do deep dives into cause and effects.

First glance at the Dataset

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Datasets | Sources | Description | Topics | Related Topics |
| Airline Delay and Cancellation Data, 2009-2018 | <https://www.kaggle.com/yuanyuwendymu/airline-delay-and-cancellation-data-2009-2018>  Original source is OST website, which has data from 1987-now | * US flights * 2009-2018 * Has a column for departure delay, a column where 2% of entries have missing data for this variable * Has unique carrier code and the flight number. | Transportation, operations |  |
| January Flight Delay Prediction | <https://www.kaggle.com/divyansh22/flight-delay-prediction>  Original source is Bureau of Transportation Statistics | * US flights * Jan 2019- Jan 2020 * Data is very complete; however it is unclear if this dataset has any delays. There are not flag columns indicating it was delayed and there is no column for departure delay. The csv files are named ‘Jan\_2019\_ontime’ and ‘Jan\_2020\_ontime.csv’ * This dataset has the unique carrier code and the tail number, but not the flight number. |  |  |
| Feb 2020 US Flight Delay | <https://www.kaggle.com/rowhitswami/feb-2020-us-flight-delay>  Original source is Bureau of Transportation Statistics | * US flights * Feb 2020 * Has a flag for departures that were more than 15 minutes late, but not how long the departure was * Departure time and the departure delay flag are missing data for 1% of entries * Has unique carrier code but not the actual flight number. |  |  |

Methods and Models

Milestones

Timeline

Concerns

* 2019-2020 data may capture delays caused by a rare event (COVID-19). Using these points may impact the prediction. While the entries sometimes have information on the cause of the delay, this is often not available.