Project Proposal

组员：陈嘉懿、陈琪颖、金相恒、张天成（首字母顺序排列）

Project problem

Civil aviation is an important part of the transportation system, but due to a number of factors including, but not limited to, weather, military control, major events, emergencies, etc., civil flights have a high probability of being delayed for a long period of time, which may last for hours or even longer. This causes a lot of trouble for passengers' travel plans.

In order to minimize the disturbance caused by unexpected flight delays, we intend to predict flight delays based on flight data. The detailed problem set is formulated below:

1. Predict flight delays based on all information before planes land on the runway (wheel on time).
   This task is set to be the benchmark task, which should be the easiest one, for most emergencies take place before the wheel on time. Accuracy is the key criteria of this task.
2. Predict flight delays based on all information before planes leave the runway (wheel off time).
   In view of the short time between landing and actual arrival, it is not very meaningful to give a prediction after the aircraft has landed, the second task is to give a prediction of the length of the delay of the aircraft based on the information already available before the aircraft takes off (including the information known to the tower at the time of takeoff, such as military control, major events, etc.).
3. Use interpretable time series modeling to gain insights from segmentation of different airports/departure flights, etc.
   Use an interpretable time series model, forecast data from different airports/flights separately and analyze the differences in the importance of different features in each group, resulting into insights after other necessary analyses.

Dataset

The U.S. Department of Transportation's (DOT) Bureau of Transportation Statistics tracks the on-time performance of domestic flights operated by large air carriers. dataset

Methods plan to use

We plan to base our method on the paper Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting, in which a interpretable transformer structure temporal model is introduced. This model has build-in interpretability and state-of-art prediction accuracy.