FlyOnTime: A Travel Assistant with Flight Delay Tracking and Predictions (Final Report)

Team 27: Jiancong Gao, Zhengyang Chen, Jingyi Li, Yi Li, Yilun Zha, Ziqi Tang

1. Introduction and Motivation

Flight delay leads to long-time waiting and uncertainty, and various delay reasons may result in stress and frustration on travelers. However, such problems can be informed in advance or avoided by investigating previous delays and scheduling flight trips accordingly.

When planning a flight trip, common factors like price, dates, times can be easily filtered or sorted on official websites of airlines and travel agencies. However, to the best of our knowledge, there is no such function provided by them which implements the interface of delay rates for end users.

2. Problem definition

So, to provide a more informative and more user-friendly travel assistant of flight delay, we define our problem with the following two parts:

1) **Visualizing** historical flight delay information for users to acknowledge delay statistics at different airports and its daily changes;
2) Building a delay **prediction** model for custom flight recommendation.

3. Literature Survey on Recent Research

Flight delays: According to Michael(2010), flight delays make significant contributions to the negative experience of flight trips and will impact passengers' consequential trips and cost. Global Survey(2019) shows that travelers nowadays are struggling with delayed flights because of the lack of information. However, according to Efthymioube et al.(2018), even just be warned of possible delays is helpful for travelers and can improve their satisfaction. This motivates the development of effective and convenient tools for monitoring and predicting flight delays.

Flight delay prediction: Hansen(2002) proposes an analysis method on flight delay of LAX using deterministic queueing. Abdelghany et al.(2004) present a delay projection model on large-scale flight schedule. Peterman et al.(2004) use censored regression to analyze the delay impact. Juan Jose Rebollo et al.(2014) predict departure delays through Random Forest algorithms and achieve high accuracy. Yu et al.(2019) build neural network models that handle large datasets and compute multi-factor which will impact the flight delays. Thiagarajan(2017) based on limited flight delay datasets to train delay predict model. We leverage and adapt such machine learning models into our tasks.

Group of researchers also investigates the delay analysis in systematic aspect. Rosen(2002) indicates the delay propagation from flight to entire infrastructure system. Janic(2003) presents an assessment model for delay and economic consequences, and Wu(2005) explores the inherent delays in flight schedule under multiple influence factors. Useful information is retrieved from those papers for our interface design.

Deshpande et al.(2012) found that flight delay reasons may include aircraft not ready, weather impact, and airspace congestion. Therefore, we integrate related datasets containing the above conditions and build prediction models based on all possible related features.

UI and visualization: In aspects of UI and visualization techniques, Xiaoye(2017) uses different variables for flight delay to build various visualizations to create a prediction model for flight delay time. On the other hand, Gennady's paper(2005) provides many helpful ways to create interaction with geo-spatial data. Sabrina(2018) proposes many different cartograms for geo-statistic data. The usage of colors and radius as scale for different variables is very inspiring. Kai's paper(2018) creates many interactive visualizations that show the delay of flight of the U.S. and run decision trees on his visualization. In addition, Powell et al.(2018) present semantic web techniques in D3 and geodata interactions, Cruz et al.(2018) also suggest a framework of geospatial and temporal data integration using D3. Both papers have limitations in specified domain knowledge where we could try to expand in a more generic application.

4. Proposed Methods

(a). Innovation List

To better inform travellers, we are going to build our FlyOnTime project with several innovations:

1) We design an interactive web interface for analyzing and visualizing the flight delays within the U.S. Compare with other existing website, our project visualized flight delays in a more interactive and informative way with many interacting designs as details are introduced in part (d).
2) We build a prediction model for flight delays with various features, which is a brand-new engine for custom flight search and recommendation that helps passengers to plan ahead with potential delay. Details can be found in part (c).

(b). Data collection and Data cleaning

Data collection The dataset that we use to visualize flight status and predict delay is obtained from Reporting Carrier On-Time Performance[1] (1987-present) from the Bureau of Transportation Statistics, United States Department of Transportation. This dataset is large,

[1] https://www.transtats.bts.gov

considering there are 2,530,994 domestic flights in total and 910,931 flights have a delay for arrival for year 2018. This dataset contains information from time, airline, airport, flight details, delay info and five cause of delay. A brief overview can be found in Table 1.

We use data from Jan. 2018 to Dec. 2018 and filter the data with condition that both departure airport and arrival airport are in the top 30 busiest airport in 2018.[2] Because some cities/areas have multiple airports, e.g. JFK/EWR/LGA for New York City, so for better visual presentation, we combine those airports together and show results according to city in web interface.

Table 1 Brief Overview of Categories and Examples of BTS Dataset

Category	Sub-Category	Sample
Time Period	YEAR	2019
	MONTH	1
	FL/DATE	1/19/19
Airline	OP_CARRIER	DL
Origin	ORIGIN	DTW
	ORIGIN_CITY	Detroit, MI
Destination	DEST	LGA
	DEST_CITY	New York, NY
Departure Performance	DEP_TIME	1546
	DEP_DELAY / min	26
Arrival Performance	ARR_TIME	1832
	ARR_DELAY /min	19
Cause of Delay	WEATHER_DELAY /min	218

[2] https://en.wikipedia.org/wiki/List_of_the_busiest_airports_in_the_United_States

Data cleaning We mutated and modified raw data in order to import them into the prediction model. For example, flight time in this dataset is formatted as an integer like“1005”, which means 10:05. Thus, we use OpenRefine to perform the data cleaning job including the exclusion of NA, splitting columns, parsing time into the correct format, etc,.

Data Analysis In general, the average delay/in advance time for each flight is 5.5 minutes. However, we have to notice that 15.6% of the flights arrive with a delay longer than 15 minutes, and 2.5% of the flights are delayed for more than 2 hours (Figure 1). Among all airports, BOS is the most delayed airport, while PDX is the least delayed one (Figure 2). From Monday to Sunday, we find that delay time is higher on weekdays than weekends (Figure 3). From morning to night, the delay time increases significantly (Figure 4). We believe it is related to cumulative delay and the delay of previous flight. Also for delay reason, surprisingly, weather only accounts for 4% of the total delay, the three most common delay reasons are Carrier Delay, National Air System Delay and Late Aircraft Delay (Figure 5).

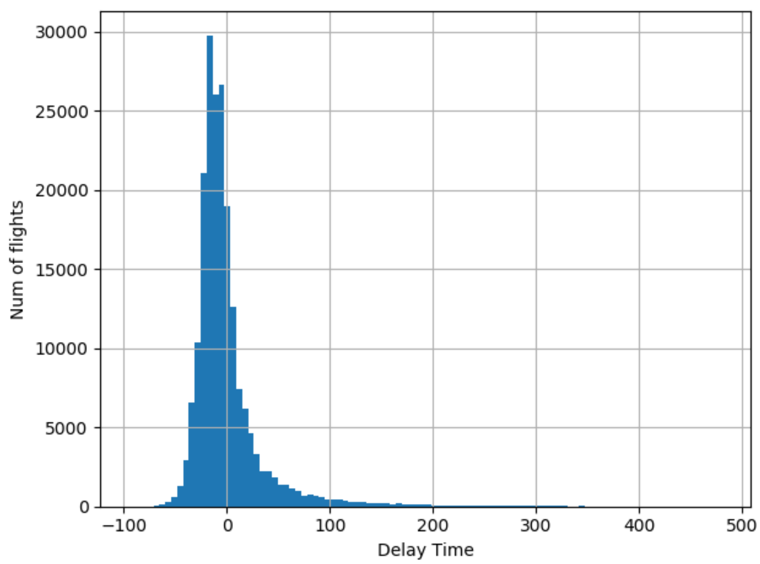

Figure 1: Delay Time Histogram of Jan. 2018

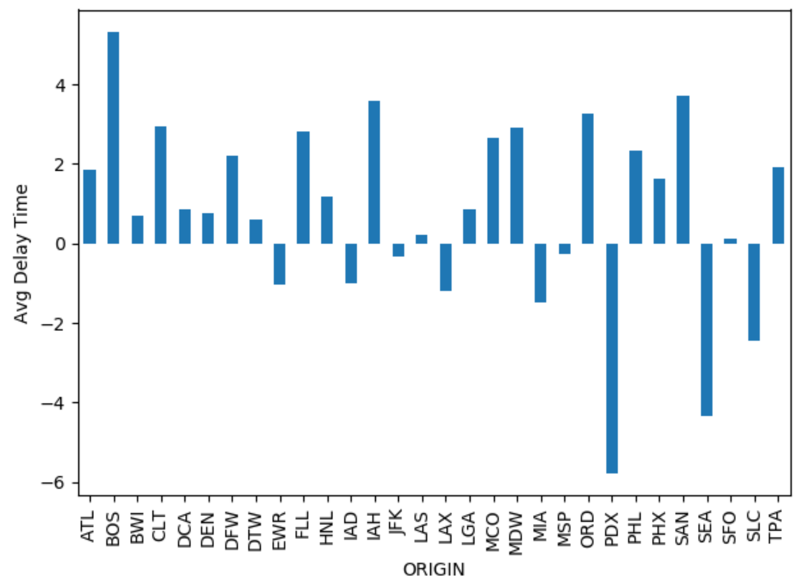

Figure 2: Average Delay Time Comparison for each Origin Airport (Jan. 2018)

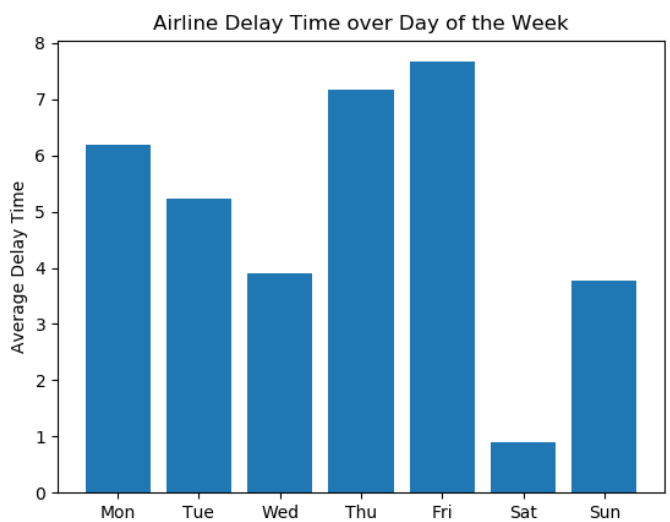

Figure 3: Average Delay Time Comparison over Day of the Week

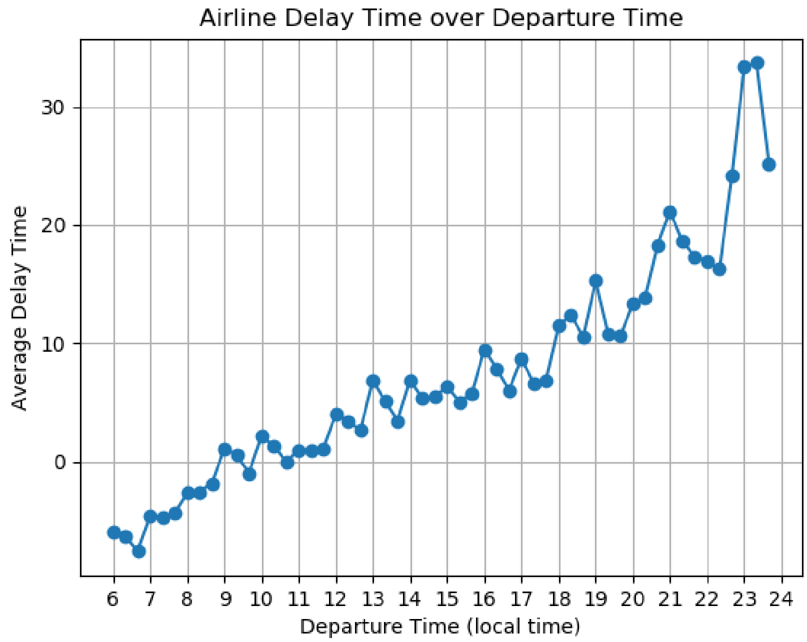

Figure 4: Average Delay Time Comparison over Departure Time (6:00-24:00)

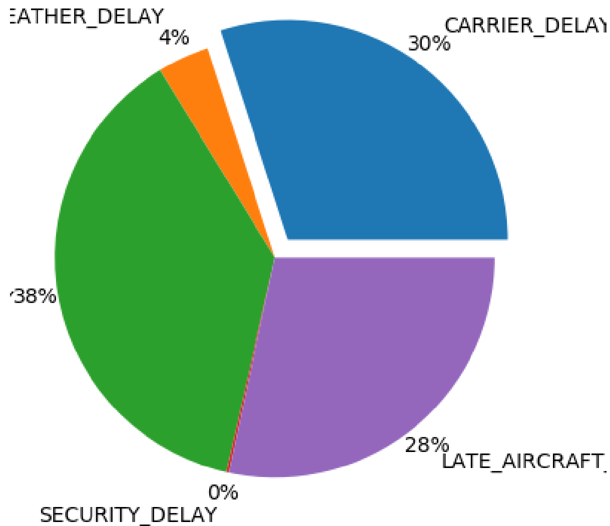

Figure 5: Delay Reason Proportion

(c). Delay Time Prediction Model

(1). Benchmark

Flight Delay can be measured by multiple ways, including but not limited to 1) raw delay time, reparented by real number; 2) delay time binned into certain ranges, represented by categories; and 3) whether delay a certain amount of time, represented by boolean variable. Different end user may prefer different measurement. In this project, we develop classification models to predict the future delay probabilities in terms of the third benchmark.

(2). Dataset split

The flight delay data is in time-series in nature and we need to predict future delay using the historical data. To simulate the actual usage and better measure the model generalizability, we split our dataset by flight date to construct the training (60%), validation (20%) and hold-out set (20%).

(3). Input feature

We combine temporal, geographical, and carrier specific data as input features. Including flight month and day of the week, departure time, origin and destination cities, flight distance, number of stops, and Carrier. Continues features are directly used and categorical features are one-hot encoded.

(4). Prediction model

We have evaluated many machine learning models, including Linear Regression, Decision Tree, Random Forest, Logistic Regression, Naïve Bayes, and SVM for delay prediction. Random Forest classifier was chosen as our final prediction model.

Random Forest is an ensemble approach which consists of many individual decision trees. The training process involves bagging to build each decision tree using subset and sub features sampled from the entire dataset. The prediction is made by voting from all decision trees. Random forest has such advantages that benefit our task: Easy handling high dimensional features, automatically generate feature importance, handles missing features, and can be implemented for parallel computing.

Model hyper-parameters are tuned by grid search and 5-fold cross validation. In the final model, we used 500 decision trees with a maximum depth of 50, the minimum number of samples in each leaf is 5. The quality of split is measured by entropy and information gain. To cure the dataset imbalance problem, the class weight was set to be 1: 2 for negative and positive delays.

(d). User Interface

For this project, we use D3 to achieve our web-based user interface. To be specific, our user interface is made up of three main panels.

(1) Filter panel

On the top left is the Filter panel which allows users to interactively give input to the program. Categories of input include month, date, take-off and landing time, airlines and locations. It serves as a hub for the entire program because it receives raw data from the dataset, transmit filtered data into other panels and the prediction model.

In terms of D3 objects, we plan to create several checkboxes and scroll bars which receives input from users, transferring parameter to other panels and the prediction model. Interaction events mainly consists of mouse actions, where dynamic data information retrieval methods are to be defined. Precision of time intervals are to be set in hours, where thresholds of query can be handled.

In addition, basic text matching or fuzzy search will also be a technical issue. Implementation of these functionality will be done in front-end, a most junior approach will be maintaining the preset list of key airport names and assign onChange event to the input box.

(2)**Visualization panel**

On the bottom is the Visualization panel, which is the main visual part of our project which generates an interactive and visual friendly map based on the input from the filter panel. It is based on U.S. States topoJSON in D3. When users set up different filters, the vis panel will able to automatically respond to data changes and give updated output. In particular, we allow users to take a closer look at the performance of specific airport by clicking on the ring chart. A series of mouse events in D3 will be used to achieve that goal.

Data visualization in this part will have interaction with the filter panel, while filtered data is retrieved, it will be rescaled and delivered to plotting functions and will present them in explicit shapes - the scaled circle with percentage representing the delay rate and the link indicating the routing of airlines.

(3) Information Panel

The Information Panel on the right receives data from both filter panel and the prediction model to show detailed information about any specific airport(Airport Details) or flight route(Delay-based Flight Recommendation). Airport Details mode contains some basic D3 objects like text, rectangle in gradient colors.

While the panel size is limited, it will be essential to filter out the less important information, hence a sorting will be applied and limits are preset in D3. Plus, interaction between info panel and other two panels are to be implemented - filtered data will be transmitted from filter panel as input dataset, and when user hovering over the visualization panel, the information will be added a second filter accordingly.

(4) Prediction Panel

The prediction panel on the bottom also receives data from the filter panel. Through the prediction model to predict flights with the lowest delay rate, and recommend to the user with low delay rate flight's number and its corresponding company. These information are visualized as a horizontal bar chart (percentage of delay rate and flight numbers are x and y axis) as shown in Figure 5.

The top five to top ten flights with the lowest delay rate are displayed in ascending order. Also, because of the user's possible choices from the filter panel, such as not selecting a start month and a sufficiently specific takeoff and landing time, when the prediction model returns flights with top ten low delay rates, it is possible to return same flight multiple times. In this case, we will combine the information of the same flight and choose to only visualize the flight information with the lowest latency among them.

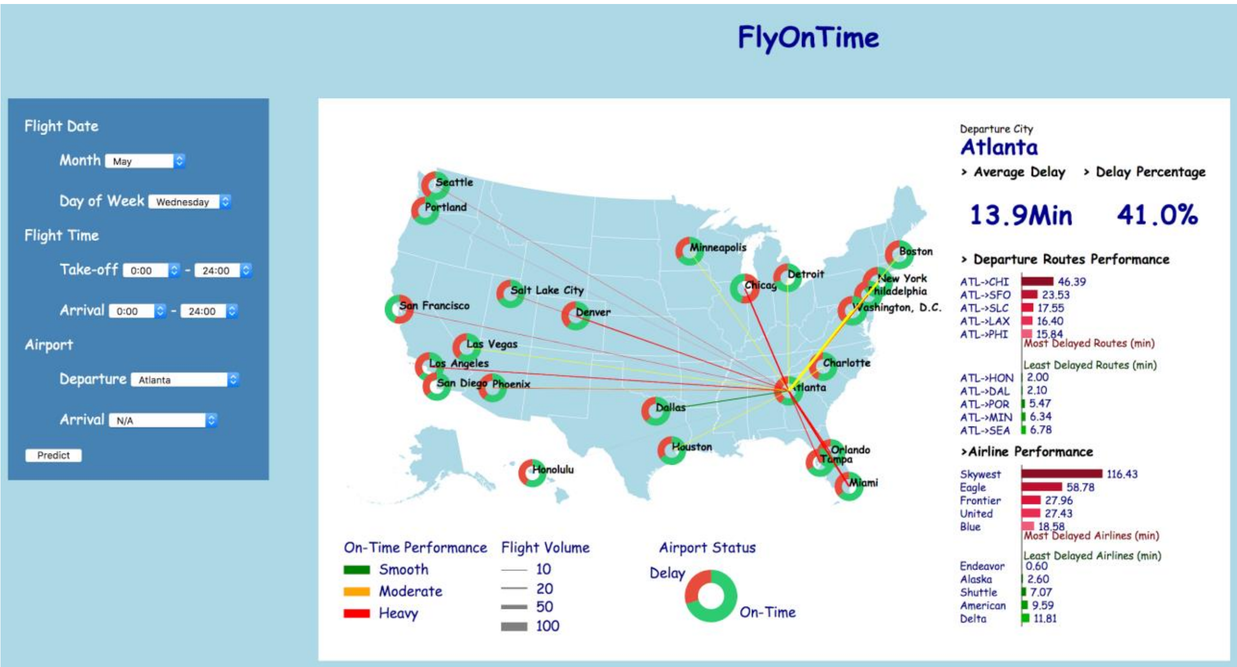

Figure 3: Interface of ‘Airport Details’

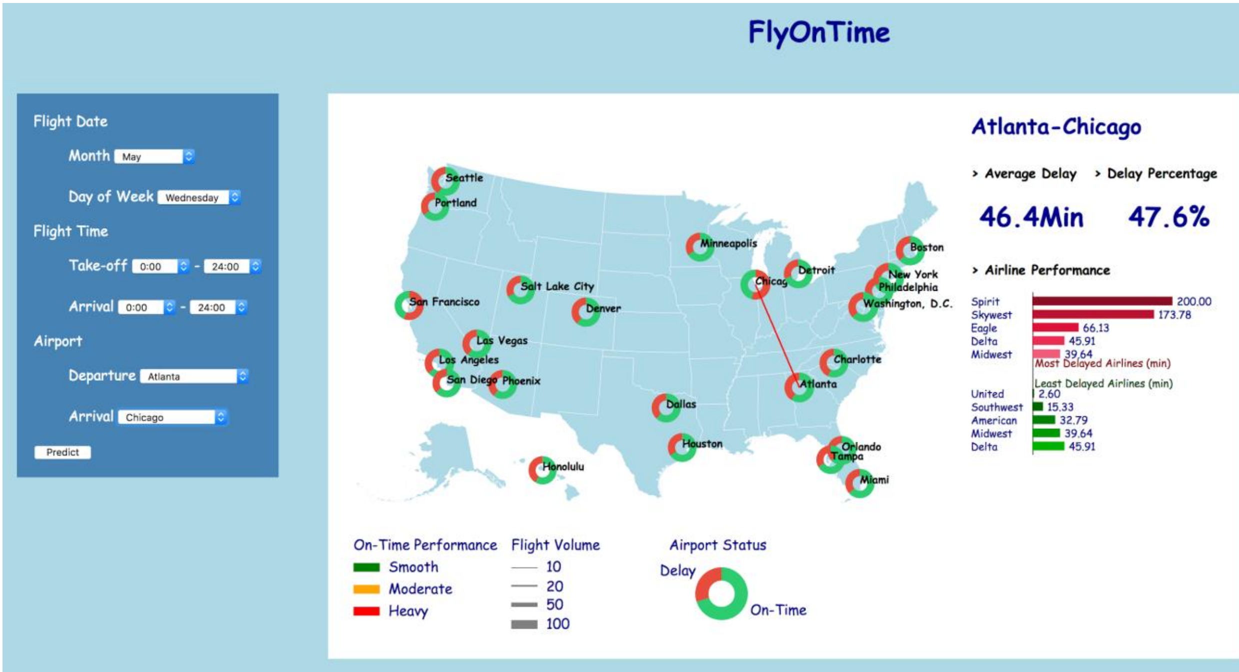

Figure 4: Interface of 'Delay-based Route Analysis'

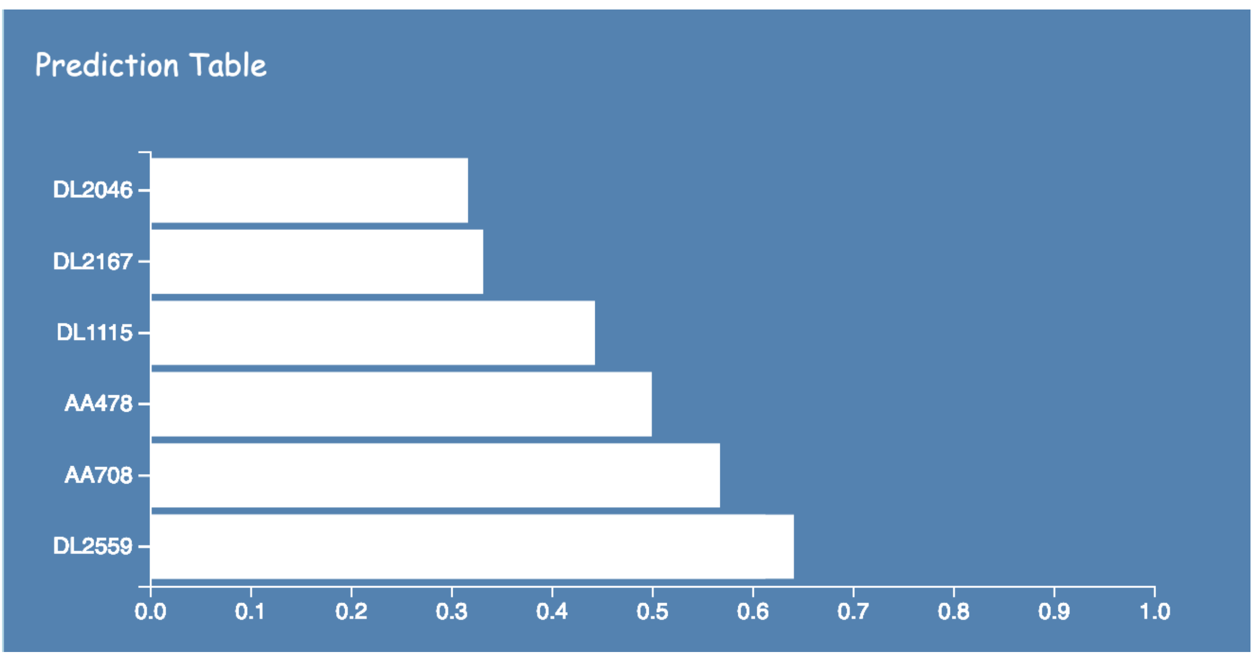

Figure 5: Interface of 'Aircraft delay rate'

5. Experiments / Evaluation

In order to evaluate our interface, our team decided to conduct a cognitive walkthrough experiment with potential user groups. As the participants work through the tasks, we, as the reviewers, would try to answer the following questions:

1. Will the user try to achieve the right effect?

2. Will the user notice if the correct action is available?
3. Will the user associate the correct action with the effect trying to be achieved?
4. If the correct action is performed, will the user see that progress is being made toward solution of their task?

After the participants finish with the tasks, we would communicate with the participants and get feedback from them to find possible issues on our interface.

Methodology

Test Scenario: You are a Tech student who is from Las Angeles and try to fly back to your hometown for Christmas. You really want to get back as soon as possible, so you want to find the flight with less delay. You want to use our interface to find out how to pick up the correct flight.

Task 1: Find the delay status for the flights from Atlanta to Las Angeles during 2018 Christmas period.

Task 2: Find the delay status for the airports of Atlanta and Las Angeles during 2018 Christmas period.

Task 3: Use the prediction functionality to find what is the possibility for future delay of this year.

Steps for each task:

Task 1:

1. Select December in the month filter
2. Select Atlanta as departure city and Las Angeles as arrival city
3. Use other filter such as day of the week, departure time and other to find more helpful data
4. Read from the information panel to find out the performance of this route

Task 2:

1. Use the mouse to hover on the Las Angeles and Atlanta icon on the map
2. Find the delay rates for the airports in those two cities

Task 3:

1. Press the predict button on the filter panel
2. Read from the bar chart to find the specific flight with best predicted performance

Results

We recruited 5 participants for the evaluation session. All of the participants are students from Georgia Tech.

Task 1:

All participants interacted with the filter panel correctly and easily. They all selected the correct month, departure city and arrival city in a short amount of time. However, some of the participants seemed to have some confusion with other filter such as the day of the week.

Future improvement: We may try to explain all the filter options better and remove the unneeded ones

Task 2:
Most of the participants knew to hover on the city icon to see the city performance under the filter condition. They reported problems such as the paths that connects the city are overlapped with each other and make the map really messy. On the other hand, some of the participants also reported that some of the city icons also stacked on top of each other, which make the hover action hard to accomplish
Future improvement: We would change the paths into curved lines and make the icons smaller.
Task 3:
All the participants knew to press the predict button to predict the future delay. However, most of them cannot find the panel that showed the prediction result. Also, they don't understand the label on the graph, which suppose to mean the possibility of flight delay.
Future improvement: We would make the prediction panel more obvious to the user. One way is to change the layout of the interface. Another way is to combine the prediction panel with the information panel. Also, we would try to make the possibility of flight delay label into percentage form.

Model performance

The prediction model was evaluated by the accuracy, precision, and recall of flight delay prediction on the test set and out-of-bag samples. The results are shown in Table 2. We can see that the model achieves about 80% accuracy on delay prediction. Comparing to the state of the art, although we are not doing the task on the same dataset, our accuracy is at the same level. To be more focusing on the positive delays, the precision and recall scores are also acceptable but can be improved by further optimization. The performance on hold-out test set is slightly worse than that on the training set, which implies potential overfitting.

Table 2 Prediction Model Performance

	Accuracy	Precision	Recall
Train	0.83	0.64	0.32
Test	0.79	0.45	0.21
Out-of-bag	0.79		

Random forest model automatically generated feature importances. From the results, the month of flight, day of the week, and destination airport are the most significant features for the predictive model.

Table 3 Feature Importance Generated by the Random Forest Model

	Feature Importance
Month	0.20
Day of week	0.15
Destination	0.14

6. Conclusion / Discussion

Although Flight Delay is a random event, our study shows that it is largely related to factors like date, airport, and airline carrier and thus predictable. With our integral platform for delay visualization and prediction, travelers may schedule their trips and reduce their chance of encountering flight delays in advance.

7. Contributions

All team members have contributed a similar amount of effort. JG collected, cleaned the dataset and performed data analysis on it. ZT developed and evaluated the prediction model. YL wrote the filter panel and visualization panel. ZC and YZ wrote the information panel. JL and ZT wrote the prediction panel. All team members were involved in project design, revisions of the report, and have read and approved the final version.

References:

1. Ball, M., Barnhart, C., Dresner, M., Hansen, M., Neels, K., Odoni, A. R., ... & Zou, B. (2010). Total delay impact study: a comprehensive assessment of the costs and impacts of flight delay in the United States.
2. Efthymiou, M., Njoya, E. T., Lo, P. L., Papatheodorou, A., & Randall, D. (2019). The Impact of Delays on Customers' Satisfaction: an Empirical Analysis of the British Airways On-Time Performance at Heathrow Airport. *Journal of Aerospace Technology and Management*, *11*.
3. Deshpande, V., & Arıkan, M. (2012). The impact of airline flight schedules on flight delays. *Manufacturing & Service Operations Management*, *14*(3), 423-440.
4. Cirium. “Global Survey Shows Travellers Still Struggle With Lack of Information About Delayed and Cancelled Flights, Says Cirium.” *PR Newswire: Press Release Distribution, Targeting, Monitoring and Marketing*, 8 Apr. 2019, www.prnewswire.com/in/news-releases/global-survey-shows-travellers-still-struggle-with-lack-of-information-about-delayed-and-cancelled-flights-says-cirium-868309007.html.
5. Rebollo, J. J., & Balakrishnan, H. (2014). Characterization and prediction of air traffic delays. *Transportation research part C: Emerging technologies*, *44*, 231-241.
6. Yu, B., Guo, Z., Asian, S., Wang, H., & Chen, G. (2019). Flight delay prediction for commercial air transport: A deep learning approach. *Transportation Research Part E: Logistics and Transportation Review*, *125*, 203-221.

7. Xiong, X. (2017). *Estimation and Prediction on Air Flight Delay Status in the US and Data Visualization* (Doctoral dissertation, University of Southern California).
8. Andrienko, G., & Andrienko, N. (2005, July). Visual exploration of the spatial distribution of temporal behaviors. In *Ninth International Conference on Information Visualisation (IV'05)* (pp. 799-806). IEEE.
9. Nusrat, S., Alam, M. J., Scheidegger, C., & Kobourov, S. (2018). Cartogram visualization for bivariate geo-statistical data. *IEEE transactions on visualization and computer graphics*, *24*(10), 2675-2688.
10. Teong, K. S., En Tan, Y., Shabbir, M., Chua, S. L., & Foo, L. K. (2018, December). Finding Insights through Interactive Visualization. In *Proceedings of the 6th International Conference on Information Technology: IoT and Smart City* (pp. 49-53). ACM.
11. Powell, L. J., & Varanka, D. E. (2018). *A linked GeoData map for enabling information access* (No. 2017-1150). US Geological Survey.
12. Cruz, I. F., Ganesh, V. R., Caletti, C., & Reddy, P. (2013, November). GIVA: a semantic framework for geospatial and temporal data integration, visualization, and analytics. In *Proceedings of the 21st ACM SIGSPATIAL international conference on advances in geographic information systems* (pp. 544-547). ACM.
13. Abdelghany, K. F., Abdelghany, A. F., and Raina S., (2004) A model for projecting flight delays during irregular operation conditions, Journal of Air Transport Management, Volume 10, Issue 6, Pages 385-394
14. Hansen, M., (2002) Micro-level analysis of airport delay externalities using deterministic queuing models: a case study, Journal of Air Transport Management Volume 8, Issue 2 , Pages 73-87.
15. Janic, M., Large-schale Disruption of an Airline Network: a Model for Assessment of the Economic Consequences, Presented at the 82rd Annual Meeting of the Transportation Research Board (TRB), Washington D.C., 2003.
16. Wu, C. (2005), Inherent delays and operational reliability of airline schedules, Journal of Air Transport Management Volume 11, Issue 4 , Pages 273-282
17. Rosen, A. (2002), Flight Delays on US Airlines: The Impact of Congestion Externalities in Hub and Spoke Networks, Department of Economics, Stanford University
18. Hansen, M., and D. Peterman, (2004) Throughput Impacts of Time-based Metering at Los Angeles International Airport, Presented at the 83rd Annual Meeting of the Transportation Research Board (TRB), Washington D.C., 2004.
19. Thiagarajan, Balasubramanian, et al. “A Machine Learning Approach for Prediction of on-Time Performance of Flights.” *2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*, 17 Sept. 2017.