

## Beat The Crowd: Tool / Analysis

### Functional Specification

- Background
  - Having a flight get delayed or canceled adds stress and uncertainty to traveling and with the Covid-19 pandemic, the prevalence of such occurrences has increased dramatically. By understanding what impacts the possibility of a delay occurring, routes and airlines that will provide the highest likelihood of incident free travel experience can be identified.
  - Questions:
    - What factors impact the likelihood of a delay occurring on a flight? (seasonality, route, airline, etc.)
    - What factors impact the likelihood of a denied seat booking occurring on a confirmed flight? (seasonality, route, airline, etc.)
    - What factors impact the fare of a flight? (seasonality, route, airline, etc.)
- User profile. Who uses the system? What they know about the domain and computing (e.g., can browse the web, can program in Python)
  - User will be someone who is interested in traveling within the US, and would like to know what is the expected price, delay and seat booking denial chances for their flight at their desired destination. They must have the capability to browse the web but does not require strong research capabilities and can rely on this analysis/tool to assist in determining ideal conditions for travel. They will have an idea of the source and destination city that they would like to travel to, and probably they will have a specific airport or carrier in mind but it is completely optional.
- Data sources. What data you will use and how it is structured.
  - Data Pipeline:
    - <https://www.transtats.bts.gov/ontime/departures.aspx> - Provides statistics such as Arrival time, Departure time, Cause of Delay of all flight carriers between the selected source and destination city. We will be scraping this data by building a realtime pipeline and make it easier for any user to use this data for any analysis of domestic airlines across all carriers.
  - Delay related:
    - <https://www.bts.dot.gov/chronically-delayed-flights> - This dataset contains information about the flight routes and carriers which were delayed for more than 30 minutes. We will be using this dataset along with the causes of the delay dataset.
    - [https://www.transtats.bts.gov/OT\\_Delay/OT\\_DelayCause1.asp](https://www.transtats.bts.gov/OT_Delay/OT_DelayCause1.asp) - This dataset contains information about all the flights of specific carriers and routes which were delayed for a given timeframe. This dataset will be used in conjunction with chronically delayed flights to analyze factors behind flights getting delayed.
  - Fares:
    - <https://www.transportation.gov/policy/aviation-policy/domestic-airline-consumer-airfare-report> - This dataset contains information about the range of fares

across all quarters from 1993. The dataset also has the route and airline carrier information which will help in determining the trends using the historical prices.

- Denied Boarding:
  - <https://www.bts.dot.gov/denied-confirmed-space> - This dataset contains all the instances recorded where a passenger was denied boarding for any given reason. The data is available for specific carriers and also contains the compensation paid out for the denial. This dataset will be helpful in analyzing which carrier is more likely to deny booking on a confirmed flight.
- Use cases. Describing at least two use cases. For each, describe: (a) the objective of the user interaction (e.g., withdraw money from an ATM); and (b) the expected interactions between the user and your system. Additional use cases are recommended, if any!
  - (a) A prospective passenger wants to travel from Seattle to Oklahoma city. This passenger is aware of the tempestuous climate of Oklahoma and would like to know the route that would least likely result in a delay for the passenger to board (b) Using the tool, the user will select the factor they're interested in, which is 'delay' in this case. When prompted, the user will provide the origin and destination cities for their trip, which in this case is Seattle and Oklahoma City respectively. The tool will then provide a visualization of an ordered list of flights they can take, alongside their respective flight time, available seats, carrier, pricing, and likelihood of delay
  - (a) A user wants to travel between Seattle and Chicago to visit a friend some time in the next year. They are on a tight budget so they want to know how they can make this trip as cost effective as possible (b) The user will choose the factor they are interested in using this analysis tool for, which in their case is price. They will then input their origin and destination cities i.e. Seattle and Chicago. The tool will then provide visualizations that can help the user determine the optimal flight options to minimize their cost in regards to the timing of their flight and the airline.
  - (a) A user needs to travel from Los Angeles to New York to participate in an important conference. Due to work conflicts they will need to fly out on the same day as they are presenting and thus they want to take all possible steps to avoid disruptions to their intended travel plans. They are specifically interested in which airline will provide them the most security when booking. (b) The user will select their factors of interest, being flight delay or denied boarding. The tool will then provide them information on which factors could impact their likelihood of these situations occurring so that they can make an educated decision about which flight to book.
  - (a) A user needs domestic airlines data to work on any analysis related to flights arrival, departure, and delay. (b) The user will select any or all of the parameters like source airport, destination airport, and airline carrier. (c) The tool will then fetch the data from the Bureau of Transportation Statistics and provide the data in multiple formats based on the user's chosen criteria.

**Component Specification.** The document should have sections for.

## Software components

Name: Pipeline to get Data

What it does: The pipeline component combines all the disjointed data of all the domestic flights operated by all carriers in the US. The pipeline will help us gather the data together so as to work on the statistics of each and provide a better analysis.

Inputs (with type information):

- Source city, destination city, carrier - The pipeline will fetch data from the Bureau of Transportation Statistics based on any given combination of the three inputs.

Outputs (with type information):

- All the combined data for the selected source city and all the airline carriers operating from that city. The data also has statistics like arrival time, departure time, delay and its cause.

Assumptions: The source city is a valid US city.

How it uses other components: This component takes input from the visualization manager and then extracts the data from the Bureau of Transportation Statistics based on the input. The output data will then be fed into the data manager which will then perform the analysis.

Name: Data Manager on Delays, Denied Boarding, and Fares

What it does: This component will provide the interface to the supplemental flight data so that it can be navigated easily. This tool will allow for querying of the data subsets that are used for the tool which are flight delays, denied boardings, and fares.

Inputs (with type information):

- Datasets containing information on delays, denied boarding, and fares, as specified in the Functional Specification Document.

Outputs (with type information):

- Subsets of the flight data that serve as inputs to the visualization manager

Assumptions: information used for querying the data is valid and can be used to subset the flight data.

How it uses other components: This component is the backbone of all the analysis shown on the webapp and will take input from the pipeline manager to perform analysis and give output in the form of visual information to the Visualization manager.

Name: Visualization Manager

What it does: This component will display information contained in the data manager in the form of visualizations, based on user input.

Inputs (with type information):

- Origin and Destination City that is inputted by the user
- Factor of interest that is chosen by the user: Delay, Price, or Denied Boarding
- Subset of data from the data manager based on the origin and destination cities

Outputs (with type information):

- Plots depicting the impact of different factors such as seasonality or airline on chosen focus point

Assumptions: user input is valid and exists in the dataset

How it uses other components: This component takes in information from the user interface and uses that input to obtain a subset of data from the datamanager. This component will interact with the data pipeline and Data manager based on the visualization required.

Name: Web App

What it does: This component will orchestrate all the managers and act as an interface between user and the application.

Inputs (with type information): The component will take input from the user which could be the origin and destination city, airline carrier or could be selected all cities and all carriers (in case of pipeline).

Outputs (with type information): The output of webapp is either visualization or different file formats like excel, csv etc which will be the case of output from pipeline.

Assumptions: The user has basic knowledge of how to use the internet and access a website.

How it uses other components: The webapp will interact with the user and take input from the user. This input will then be forwarded to the data manager or pipeline manager based on the use-case. It will then take the output from the data manager or pipeline manager and display it to the user and send it as a downloadable file.

### **Interactions to accomplish use cases.**

Describe how the above software components interact to accomplish your use cases. Include at least one interaction diagram.

A user will enter input into the web app in the form of an origin city and destination and they will select what they are interested in visualizing: delays, denied boarding, or fares.

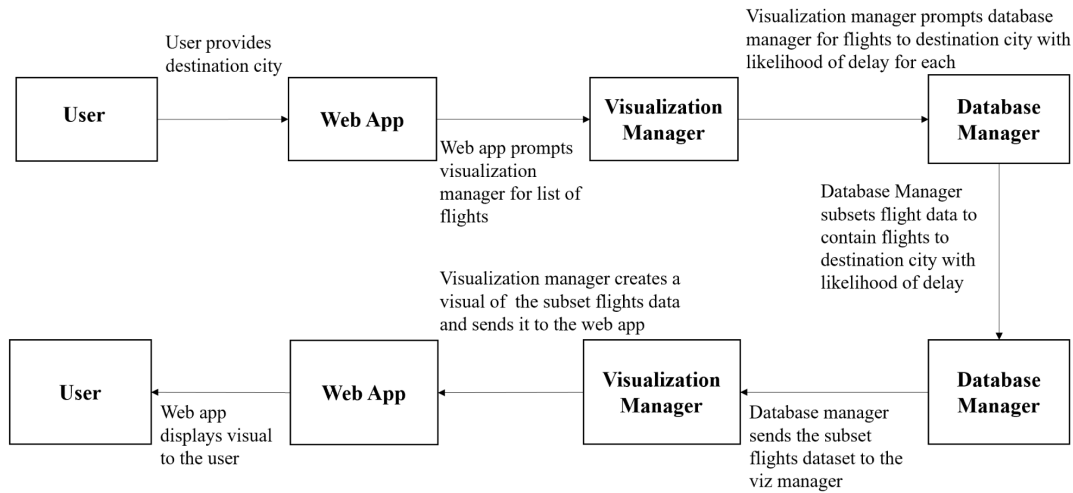
The data manager confirms that the user input exists in the dataset and outputs a subset of data for that origin and destination city.

The pipeline takes in the user input and collects the available data from the Bureau of Transportation Statistics. It outputs a dataset of flights filtered by the input.

The visualization manager takes in the datasets outputted from the Pipeline and the datamanager as input and outputs visualizations.

The Web App takes in the visualizations from the visualization manager as input and outputs them in the user interface. The Web App can also take input from the user and output the data as a downloadable file.

### Example: Likelihood of Flight Delays to a given Destination



#### Preliminary plan. A list of tasks in priority order.

1. Determine Functional and Component Specification
2. Perform Technology Review
3. Create pipeline to collect data from <https://www.transtats.bts.gov/ontime/departures.aspx>
4. Clean/Process supplemental datasets
5. Create data manager
6. Create visualization manager
7. Construct web app