US Airport Efficiency based on Flight Delays

Priyanka Dubey, 1211136222

1 INTRODUCTION

Every day about 87000 flights take off and land with about 1.73 million passengers. With so many flights and passengers, delays are bound to happen. Delays may occur due to assorted reasons such as weather, security, carrier, and late aircraft. In winter, most parts of the world experience bad weather and snowfall. Hence, runways are covered with snow or weather is not suitable for landing and low visibility adds to the causes of flight delays also leading to cancellation in worst cases. Sometimes, the delays occur because of National Airspace System (NAS) due to reasons like busy runways or lost navigation, etc. Security delays occur when there are long lines for boarding or checking the baggage due to vacation time or international students travelling to school . Carrier delays occur when the aircraft cleaning takes extra time, or the aircraft is waiting for the connecting passengers to arrive, or the aircrafts develop faults. Late aircraft delay corresponds to situations where an aircraft started late at one airport causing the delay to be propagated to the next connecting airport leading to subsequent delays. Irrespective of the type of delay, the effect is going to have a considerable impact on the passengers.

This visualization helps to establish the most efficient airports in the United States based on the arrival delay at the airport computed over a span of 2 years. The visualization targets an average traveler, who takes a flight from one place in United States to another. Air travel is the starting and the end of a journey for work or vacation. No one likes to spend more than scheduled time at the airports due to several types of delay.

If a traveler wants to book a flight from new Delhi, India to phoenix and must decide the port of entry in United States. He/she can look through this visual representation and evaluate which will be the best airport for them. Every traveler is different. Therefore, each traveler can review the airport according to their own situation. If a traveler is immigrating into the United states and has a connecting flight to another destination airport, then they can check the delays because of security checks or general delays at the airport as he/she may have little time to get all security check done and board the next flight.

Also, if there are two airports in the same city, for example Newark and JFK in New York, provided they have no preference for any of the airports, they can book a flight via a more efficient airport. Hence , it will definitely save a lot of time and the journey can be planned taking into consideration all odds.

Since the target audience of the visualisation is a variety of people such as frequent business travelers, students of all ages, and families. Some of them may or may not be aware of the several types of visualizations. There the visualizations used are common visualizations, which are easily understandable and does not take a lot of effort to draw conclusions. Simple graphs have been used.

The visualization answers frequent questions like “Which are the most efficient airports in the US for a given month and year?”, “Which type of delay is more common during a specific time of the year?” .

2. SYSTEM:

Figure 1: visualization when the page loads, the map shows the various airports in United States. On the right is a panel that shows the airport ranking according to the efficiency ratio of the selected year.

Figure 1 shows the visualization when it loads. There is a map at the center of the visualization occupying over 65 percentage of the screen.

The black dots represent the various airports. When an airport is selected, the various visualizations are updated to show results for the selected airport. Map is at the center because most of the interaction takes place through it. There is a slider at the bottom of the Map. This helps to select the year and month the user wants to identify data of. There is a bar chart ranking the various airports in United States according to the efficiency ratio.

There is a radial menu at the top left corner representing the various delay types. there also a bar chart and line chart answering other insights with respect to data.

3 RESULTS

3.1 Finding the list of efficient airports for a given month

and year

When a user wants to find out which airport had the least delays during a specific year and month, the user drags the slider to the desired position on the horizontal bar.

This would populate all the data corresponding to that month and year based onto the panel which is above it on the right.

Figure 2: Figure showing the Map with airport dots colored according to the airport efficiency .The darker shade of blue represents a more efficient airport whereas Maroon represents a less efficient airport .

3.2 Finding out the delay details of a specific airport for a given month and year

When the user wants to know the delay details specific to an airport, he would click on the airport about which he would like to know the details and re-position the slider to the desired month and year. Now, the user can see the following charts will appear giving the user a better picture of the delay details.

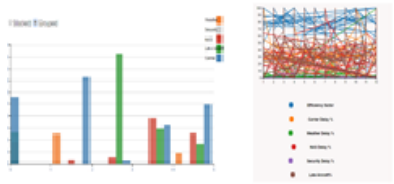

Figure 3: Figure showing the conclusion drawn from the visualization. The figure on the left shows a grouped bar chart showing the delay at an airport because of according to the airlines operating at that airport. On the right is a multi-line chart showing the delay type of all the airports.

4. MY CONTRIBUTION

We worked as a team on this project. From formulating the idea to the implementation. I oversaw collecting the dataset from the various reliable sources. the data consisted of the 4 excel files. I wrote the SQL queries for merging the various data sets and the resultant was a single data file, containing all the data needed for the visualization. This allowed faster access to our components to extract data from a single file to avoid runtime delays caused due to merging data at runtime. I also formulated the conclusions that one can draw from the data. Because the most essential information should be represented in the best feasible way. I specifically wrote the code for the Radial Menu, Stacked Bar Chart and Multiline chart. I also wrote the codes for connecting these components with the overall system so that they help in the overall story telling.

4.1 Radial Menu

The radial menu sits on the upper left corner of the visualization. The radial menu is meant for selecting the type of delay at the airport. This acts as a site wide filter which narrows the analysis down to the selected type of delay. Sump, Krystina [1] in his paper titled "the design and evaluation of graphical radial menus" proposed compact radial layout menu. This radial menu was 20-30% faster, reduced navigation errors, induced users to faster response and decreases screen consumption.

Therefore, I decided to use radial menu. The radial menu was made in JavaScript and jQuery. jQuery was used because even if the number of circular menus increases or decreases it will always place itself in a position such that they are equally spaced around the central circle.

The radial menu is meant for selecting the type of delay at the airport. Initially only one circle is visible and when the user clicks on it the other circles pop out of it and the user can select any of these options. On click all the visualizations analyze the selected type of delay. Every time a menu option is selected the information on the map and the graph changes as per the type of delay selected.

4.2 Stacked Bar Chart

Figure 4 represents a 100% stacked bar chart has been used to show the composition of the delay types at a selected airport for each airline operating at that airport.

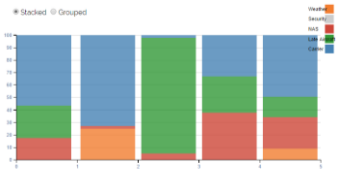

Figure 4: Stacked bar chart of the various delay types at an airport for the various airlines operating at that airport for a selected month and year on the slider

By selecting on the delay type in the legend. A delay type can be removed from the visualization. User can remove or add delay types in the stacked bar chart to have a better understanding of the composition of several types of delays per airline at an airport. He/She can make a comparison for all the airlines and tell which airline has the best track record if that delay.

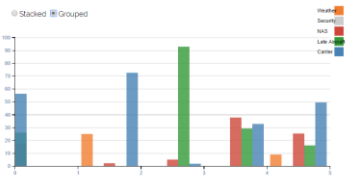

Figure 5: Grouped bar chart of the various delay types at an airport for the various airlines. All the bars begin at athe same level . Hence , help to understand better .

This visualization is dynamic in nature. The traveler can switch between a 100% stacked graph to an absolute scale grouped bar graph. This offers two perspectives: change of relationships between the parts with respect to different airlines vs. the absolute development of the delay types per airline. For this, there are radio buttons having the option to have a stacked or grouped representation of the data
In Figure 5, the bars are not stacked, but are grouped next to each other. This makes the relationships between the diverse types of delay clear. So, this is helpful when analyzing the individual delay type for every airline and comparing with the other delays and other airlines at the same time.

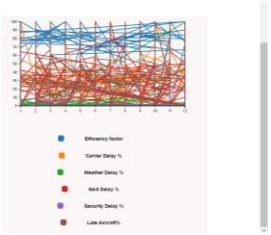

Figure 6: Figure showing the multi-line chart of the efficiency ratio and delay types for all the airports in U.S . This helps to compare the delay type to the efficiency ratio .

4.3 Multi Line Chart

I used the multiline graph to analyze the effect of several types of delay on the efficiency ratio of the airport. It is seen that not every type of delay contributes equally to the efficiency ratio. There are some delay type percentages which contributes to about 50 % of the efficiency ratio and some affect to about only 5 to 10 percentage. Also, some airports don't have a delay and hence that doesn't negatively contribute to efficiency ratio at all. The figure shows for an airport selected in the map, and the year selected in the the slider. For every month of the year, the line graph represents the efficiency ratio, Average Carrier Delay percentage, Average weather delay percentage, average NAS percentage, average security percentage, average carrier delay percentage, and average late aircraft delay percentage.
I selected his type of visualization because the advantage of using this type of graph is that we can draw relations between the different attributes in the dataset and infer the trends that may/may not occur or that you can compare multiple data sets in the same chart.
Figure 4 represents a mule line stack graph, this graph lost interaction and represents the data of all the airports hence, too cluttered. this made us learn the importance of Interaction and clarity.

5. SKILLS LEARNT:

The Project was done in a team of three. My team mates are Prannoy Chandra Pydi Medini and Amit Ranjan. I gained hands on experience in JavaScript and D3. During the project, we had a chance of applying all the visualization techniques learnt in the class to the large dataset and learn through the practical difficulties faced.
I learnt how to successfully apply Shneiderman's Information seeking mantra [2] principles such as overview, zoom and details on demand to the visualization.
Data collection was one of the most important stages of the project. It was made sure that the data was collected from reliable sources and it was large enough to represent the variety of values.
Data collection was not enough, data needed to be sanitized, so that useless features could be removed from the dataset .
Data analysis helps to realize what conclusions could be drawn from the data set. the most important attributes were more effectively visualized. Color selection skills were achieved by selecting effective and efficient colors for all the visualization components
While connecting all the components in the visualization, I understood the importance and need for interaction in a visualization and it creates another dimension for visual analysis.
The project helped me improve my data analytics skills. I also learnt and gained expertise in JavaScript and D3 during the entire course of project.

6. CONCLUSION:

This visualization will effectively help travelers to find the most efficient airports based on the flight delays that occur at the respective airports. This will ensure a hassle-free journey.

7.REFERENCES:

[1] Samp, Krystian, "The Design and Evaluation of Graphical Radial Menus", NUI Galway Theses (PhD Theses) , 2011-12-05
[2] 1 - B. Shneiderman. *The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations*. In the Proceedings of the IEEE Symposium on Visual Languages, pp. 336-343, 1996