Airlines delay and passenger satisfaction

January 2023
Data visualization

Manuel Acquistapace
Miro Rava
Stefano Billeter

Syllabus

Abstract

In the highly globalized environment where we live, delays and setbacks are a real problem. These can affect travelers of all around the world, whether they are tourists or businessmen as well as the airports-airlines framework.
This data visualization project, run in an academical context, intend to verify if there is a correlation between flight delays and passengers' satisfaction.

Foreword, Preface, Acknowledgements

This report is part of the deliverables of the lab project in "data visualization" (prof. Giovanni Profeta). In this paper we go through all the aspects our work, from where we started to where it could be further developed.

Students Manuel Acquistapace, Miro Rava and Stefano Billeter are the creators of the webpage, visualizations and report. All the three are currently enrolled in the second year SUPSI's bachelor degree in "Data Science & AI".

We would like to thank prof. Giovanni Profeta for having guided us through this semester in a very important subject. We will never forget his kindness and willingness and we will treasure his teachings forever.

Introduction

The research we conducted aims to check the existence of a correlation between objective data (delays) and subjective data (satisfaction), our initial hypothesis being that it exists and it is strong and positive.

We used three datasets we found online (Kaggle, Skytraxratings, Airhelp and Airlinequality); these will be explained and analyzed in the next chapter.

In order to verify our hypothesis, we proceeded with the creation of five visualizations, according to what we have seen during the Data Visualization module.

The visualizations are shown in a specific order to provide the audience particular information along the way. Data visualization protocols and descriptions are also available for the user.

Data sources

As aforementioned we gathered our datasets from kaggle.com, skytraxratings.com, airhelp.com and airlineqality.com. These have been carefully selected in order to be able to retrieve the best possible insights, from a quality and reliability point of view. Here a glimpse of our data:

1) Air travels (year 2009)
   Because of its large number of variables and expanse, this dataset allowed us to plot the first two visualizations.

2) Airports dataset
   We needed this dataset to complete the first one. In fact, this was useful to get all the information about the airports and their coordinates.

3) Skyratings
   With the third and last dataset, we are able to come to our conclusion, that we will see in the last chapter of this report.
   It basically shows the passengers' satisfaction (rating) on an airport basis.

Data pre-processing

In this part of our project, we were supposed to use some of our python-coding and/or supervised learning knowledge, like feature processing and engineering.
Eventually, we found out that the datasets we chose were already in a very good shape and didn't require any special pre-processing phase, except for joining the airports' coordinates and symbols.
As regards the third dataset, the one about passengers' satisfaction, it was obtained through the use of web scraping tools applied on Skytraxratings, Airhelp and Airlinequality.

Interface design

HTML

The original project plan was to use an existing HTML template. Unfortunately, the template we chose did not meet the needs of the website as we had envisioned it. Therefore, we decided to modify a simple template created for a slideshow.

The visualizations were created using tools outside of our website, the procedure for creating them is described in the "Protocol Info" section and in every popup that appears by clicking respectively on the box called "Map Protocol" or "Plot Protocol" located below each visualization.

The top section of the page contains dedicated links for "Contacts" and "Protocol Info".

The description located to the right of each visualization provides a brief overview of the main visual elements of the chart or map. It highlights the key features that are displayed in the visualization and gives a clear understanding of the information being presented.
This description helps to quickly grasp the essence of the chart or map without having to study it in detail. It serves as a useful tool for quickly gaining insight into the data being displayed and is an essential component of the website.

CSS

By using CSS, all the decorative and visual aspects of the text, sections, and pop-ups of the website have been modified and made similar to create a sense of continuity that does not distract the user from the subject of the visualizations.

The CSS allows for precise control over the appearance of the website, ensuring that all elements are consistently styled and visually appealing.

This helps to create an immersive user experience and keeps the focus on the visualizations, rather than on any distractions.

The CSS makes sure that the user is able to fully engage with the content of the website and get the most out of their experience.

JavaScript

JavaScript plays a crucial role in ensuring the smooth functioning of the website. It is essential for managing all of the animations on the site, including pop-ups, cursor movements, and informative sections.

The JavaScript code is integrated into the HTML template, customizing it with the styles defined in the CSS file and enabling all necessary animations to make the site user-friendly and interactive.

Additionally, the use of JavaScript is beneficial in handling the resizing of the site when viewed on different devices such as tablets or smartphones. This ensured that the user has a seamless and consistent experience regardless of the device they are using.

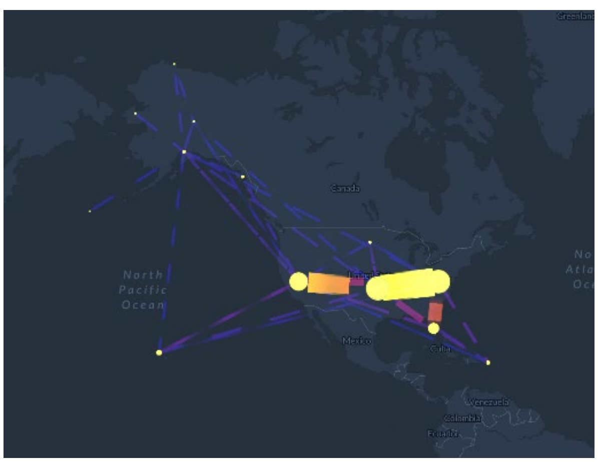
With the help of JavaScript, the website is able to provide a complete and cohesive user experience that is both visually appealing and functional.

The use of JavaScript greatly enhances the functionality of the site, making it a smooth and interactive experience for the end-user.

DATA VISUALIZATIONS

As prior mentioned, we were able to generate three visualizations. We will now go through each of them and provide the relative explanation.

1.) The visualization shown in the map called "FLIGHT ROUTES" represents the airline traffic with respect to each air route in the USA.

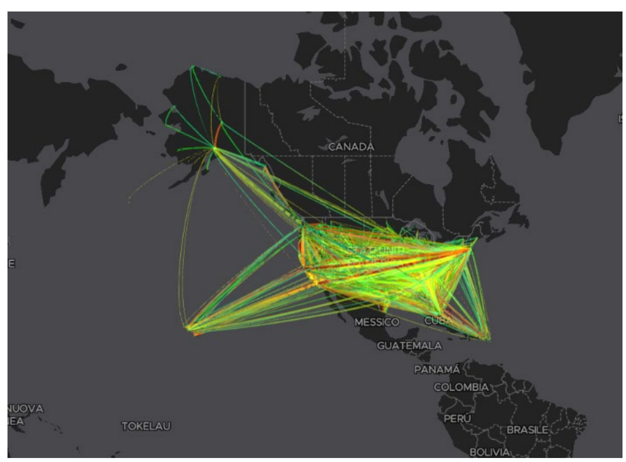


Its features and visual variables are:

- Yellow dots (shape and color): depict the airports locations. The bigger the point the higher its relevance (total number of flights).
- Thickness of segments (size, orientation): the thicker the segment, the bigger the number of flights in that direction.
- Color of segment (color): doesn't have a precise meaning. It is just a palette of colors that helps the user to understand the departure and arrival airport.

The goal of this visualization is to learn the distribution of the flights contemplated in our dataset.
There is the chance to select, by clicking on the map or writing the airport code/name in the designated field, a single airport in order to show the specific visualization.

2.) The visualization shown in the second map "ROUTES DELAY" represents the air routes in the USA, divided by airport as the first one, and the delays.

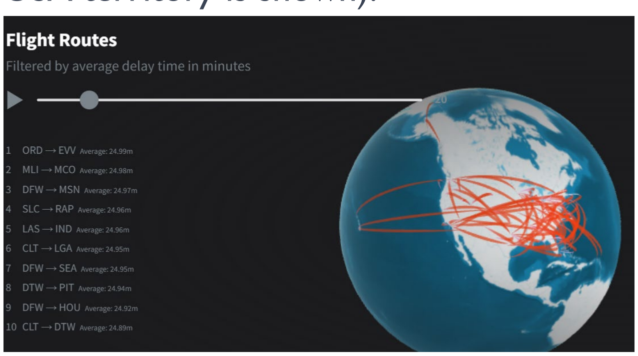


Its features and visual variables are:

- Color: the palette goes from green to red, where the first mean a low accrued delays on that route and the latter the opposite.
- Orientation: clearly visualize the route direction.

This visualization acts as a mean to understand which routes are most often late, respectively on time.

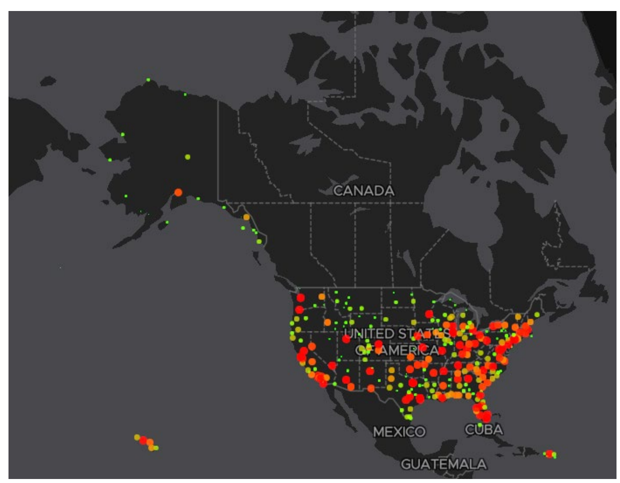
3.) The third chart, named "DELAY INSIGHT" it is just an alternative visualization of what we've already shown in the previous one.
It is an interactive and 3D map, where the user can navigate around the globe (as we know at the moment only the USA territory is shown).



Through filtering on the average delay time, this chart allows to pick a specific route, both from the left-hand side menu and the globe itself. We think that this provides a much better and less confused understanding of the previous one.
Its features and visual variables are the same as number 2.

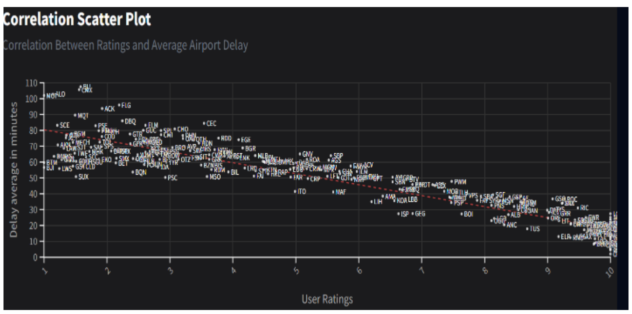
4.) With the help of the fourth plot, we approach the final conclusions of our project. "USER RATINGS" tells us the mean satisfaction ranking per each of the USA airport we took into consideration. Rankings are scaled 1 to 10, from red to green.

The usage of color and size is important to give to the user, at a glance, a full and easy understanding of the map.
As for the earlier ones, we arranged it to be interactive (by clicking on one dot some information are automatically shown).



We think that plot nr. 4 is perfect to have a 365° understanding of the ranking along the USA, but we wanted to help the user with a more classical comparison, so we decided to compare ratings against average delay in the next plot.

5.) The last, but not least, diagram is a correlation plot, which compares rankings and delays.



The data shown in this visualization also confirms all our findings.

CONCLUSIONS

The findings in our research clearly state that there is a positive and strong correlation between flights delays and passengers' satisfaction.

As stated in the abstract of the present report, in the kind of world ad environment we live today, we are pretty much influenced by delays and setbacks. It is interesting to notice how much these perturbations have an impact on the enjoyableness of our trip, probably more than the food we get on the airplane or the kindness of the flight assistants. We think we have been able, through the visualizations we created, to get this message across.

Hence, our initial hypothesis is confirmed. We can also conclude that the human being reacts linearly and objectively to such discomforts.

NEXT STEPS

These are few inputs that could be used to improve what we have done:

1) Expand analysis and visualizations outside the US;
2) Compare other means of transport, like trains or buses;
3)Add other variables to the analysis, like the reason of the delay to see if there is any change in the perception of the setback;
4) Verify if there are other variables that influence the ratings.

SOURCES

- Course slides and lectures
- Report template: templates.office.com
- Datasets: kaggle.com, skytraxratings.com, airhelp.com