Lucy Storts — 281 Project Proposal
11/27/2023

Project Proposal

Introduction

At the onset of the COVID pandemic in 2020, many opted to stop traveling and there was a noticeable decline in airport travel. By 2022, airports seemed to be back to pre-pandemic business. I want to investigate 2022 statistics for the most popular airports in the United States.

Research Questions

→ What are the most noticeable state trends for top ranked US airports in 2022?
For the top 200 ranked US airports in 2022…
- Which state listed had the greatest number of airports?
- What was the average number of airports per state listed?
- Which state experienced the lowest average security delay?
- What was the average security delay per state listed?

→ What kind of factors influence security delays for top ranked US airports in 2022?
For the top 200 ranked US airports in 2022…
- What does the relationship between flights per year and security delays look like and can it be modeled using regression?
- What does the relationship between enplaned passengers and security delays look like and can it be modeled using regression?

Data Sources

I created an extensive table for the top 200 ranked airports in the United States. Data for this ranked list was sourced from the Bureau of Transportation Statistics[1] and more details can be found in the *Notes* section near the end.

→ All data sources are linked in the *Data Sources and References* section and actual files can be found at my github repo.

Visualization Plans

→ To answer questions about the states, I will…
- Create a bar chart for the number of airports per state.
- Create a table with state airport count and an overall average for 2022.
- Create a bar chart for average security delays per state.
- Create a table with average security delays per state and an overall average for 2022.

→ To answer questions about the factors that influence security delays, I will…
- Create a scatterplot with integrated R regression and Loess regression for flights per year vs security delays. I will additionally (most likely) add in pills for airport name and 2022 rank on the tooltip.
- Create a scatterplot with integrated R regression and Loess regression for enplaned passengers vs security delays. I will probably add in pills for airport name and 2022 rank on the tooltip again.

→ To create a better overall analysis, I am going to create a map using the longitude and latitude values as well.
- I would like to use color gradient or scale for the states based on state airport count, average flights per year per state, average enplaned passengers per state, or average security delays per state.
- These options would be listed for the viewer to choose from and would change how the graph is colored based on which analysis point interested them most.

→ Finally, I will put these pieces into Dashboards and a Story in Tableau.

Notes

→ I will be using the same data for both my 386 and 281 projects.
→ My github repo for this project is: https://github.com/lucystorts/386Project
→ My 386 project is focused on data collection, cleaning, and organization. The couple visualizations I plan to submit are…
1. Scatterplot with OLS regression comparing the 2022 *Delays* proportion to the *Elevation* metric
2. Bar chart of all airports measured by their *Avg. Delay (Mins)*.

→ DataCleaning.ipynb is where the data was cleaned and organized.
- The ranked list file that was read into python needed to be cleaned.
- I cleaned it using a LOCID table[2] and a supplementary LOCID table[3].
- Next, I added originating[4] and enplaned[5] passenger data for each airport.
- Then, I used rapidAPI[6] to get ICAO codes for each row.
- Next, I used a second API[7] and the ICAO values to get long, lat, etc. columns.
- Last, I found a blog[8] that compiled an updated table of US airports for 2020 and interesting statistics. I accessed the table as a text file and merged it onto my pandas DataFrame.
- I exported my dataframe to 2022Airports.csv.

Data Sources and References

1 https://www.bts.gov/topics/airlines-and-airports/airport-rankings-2022
2 https://www.faa.gov/data_research
3 Requested unknown LOCID values for the remaining airports from chatGPT (extraLOCID.txt)
4 OriginatingPassengers - U.S. Airports ranked by 2022 Originating Domestic Passengers
Source: Bureau of Transportation Statistics, Origin & Destination Survey
DB1B Ticket, Based on 10 Percent Ticket Sample
O&D numbers are not comparable to T-100 Market Enplanement numbers
5 Bureau of Transportation Statistics DB
6 https://rapidapi.com/proground/api/aviation-reference-data
7 https://airportdb.io/#
8 https://www.stratosjets.com/blog/us-airport-rankings/

Directions

→ Proposal
- This should include your…
  - ☑ research question
  - ☑ the sources of your data
  - ☑ a brief plan for your data visualization

→ Visualization requirements
- use at least three different types of graphs to visualize your data.
- Create a responsive dashboard.
- Communicate your resulting visualization using a story with at least 4 story points which lead the viewer through your findings and interpretations.
- This should flow logically, offering clear insights, and demonstrate a deep understanding of the data.

→ Dataset
- Submit the dataset you've curated in the form of a .csv file.

→ Story Submission
- Submit the link to the Tableau story you've crafted.

→ Video Submission. Create a video walkthrough of your story. This should…
- Detail the research question you aimed to address.
- Highlight key findings.
- Demonstrate the functionality of the dashboard.

→ Final Presentation
- You will present your story during the final exam block.
- Be prepared to answer questions and explain your data collection and analysis processes.
- Classmates are welcome to attend these presentations, but the primary audience is the professor.