

flight-recommendations

A python program that recommends flights by carrier based on data given for South Carolina and North Carolina in 2018

Environment

This program uses Python 3.6.7 with tkinter as my Graphical User Interface (GUI). The pandas library is used for `.csv` intake and manipulation. The OS package is used to check for an existing file before running the `smallerDF()` function. The `calendar` package is used to call the function `month_name` to convert month numbers to their names. The Python datetime library is used to convert given month numbers back to their month names when needed.

Initial Run

The first time this project is run with `python3 main.py`, the program should read from the larger dataset and create the smaller one. Every run after that should read from the smaller data set file and not need the larger one. This is done to decrease the reading and compute time required on larger `.csv` files.

Developing the Program

This entire program was developed and run completely from my laptop. I have previous experience with Python and using the Pandas library so, I defaulted to using that for reading and manipulating the `.csv` right away. I started by narrowing down the very large dataset to something much smaller and more manageable. My program does this on first run by default now.

Once I was able to print out the data I wanted to the console, I began work on the GUI, displaying buttons for the user and tables for the results. I first made the dropdown tables for the manual mode because I knew how to do that easily. I soon switched to developing the automatic flight selection portion of my application. Given that the tkinter documentation is sparse and not very thorough, I struggled a lot at first with getting a TreeView to show all of my data.

Once the automatic section was done and I was able to list 5 flights for each month, the manual section was as simple as filtering down the dataset on the user's given input and then running that dataset through the algorithm and TreeView functions. Lots of code reuse to produce very different results!

Algorithm Explanation

The algorithm used for the auto selection of flights by month is quite simple, but effective nonetheless.

The algorithm calculates the percentage, from 0 to 1, of successfully completed flights based off of the original columns "DEPARTURES_PERFORMED" and "DEPARTURES_SCHEDULED." The average of this score is calculated for each month and is multiplied by 50, to represent the fact that it is worth half of the final score calculated.

The second part of the algorithm is made up of the percentage, from 0 to 1, of the available seats on each plane. This is based off of the original columns "SEATS" and "PASSENGERS." The average of this score is

calculated for each month and is multiplied by -50, because the higher the percentage, the less seats available which, in my opinion, should bring the overall month rank down.

These two total scores are added together to make a final "Algorithm Rank" and then sorted by highest score. I decided to make the final score "out of 20" because the highest score was ~16 and 20 could be close enough, but far enough away for the "perfect" score.

Intermediate and Supporting Files

`ProjectData.csv` is filtered down into `filtered_data.csv`. `filtered_data.csv` only contains the necessary columns for this project. I saw this necessary in order to decrease computation time of the project. This would also help if the project was ever scaled up and given data from the entire United States (instead of just South Carolina and North Carolina).

Explanation of Automatic Mode

Explanation of Manual Mode

Validation