For my project I wanted to analyze a dataset by the Bureau of Transportation Statistics containing data about every domestic flight during 2023. I selected to focus on the following features: number of passengers (passengers), airline (unique_carrier_name), origin airport id (origin_airport_id), origin city (origin), origin airport code (origin_city_name), destination airport id (dest_airport_id), destination (dest), destination (dest_city_name). The goal was to find out what were the most popular routes and more specifically focus on the flight trends in the Bay Area since San Jose (SJC), San Francisco (SFO), and Oakland (OAK) airports are within a 50 mile radius of each other.

The src code is divided into two modules main.rs and lib.rs. Main.rs is very straightforward as it only calls the run function from the other module (use project::run) and verifies that it ran without any errors. The run function is inside libs.rs and is where the graph is created. Let mut graph assigns each airport as a node and the edges as flights between airports. After having multiple issues, let mut total_flights_processed was added to keep track of how flights correctly go through the run function to ensure all of the csv file entries are counted for. To answer my question of what types of flights each airport specializes in, I kept track of the frequency for each route as well as the number of total passengers. For result in rdr.records() iterates through each entry and updates the total counts. Since I did run into some issues I also printed out "Before updating, flights for {}: {}" and "After updating, flights for {}: {}" to make sure the flights were increasing. I formatted my output to display the airports (nodes) ranked by the number of originating flights (degree). To calculate the average number of passengers per route I divided the total number of passengers by total number of flights. For SJC, SFO, OAK I printed out their top three destinations. SFO: Orlando FL, Phoenix AZ, Miami FL. SJC: Dallas/Fort Worth TX, Seattle WA, Las Vegas NV. OAK: Phoenix AZ, Las Vegas NV, ST Louis

MO. It appears that SJC and OAK mainly operate shorter distance flights while SFO operates cross-country flights which is true since I always fly out of SFO when coming to Boston.

Output:

Top destination cities for SFO (San Francisco, CA):

1. Orlando, FL

2. Phoenix, AZ

3. Miami, FL

Top destination cities for SJC (San Jose, CA):

1. Dallas/Fort Worth, TX

2. Seattle, WA

3. Las Vegas, NV

Top destination cities for OAK (Oakland, CA):

1. Phoenix, AZ

2. Las Vegas, NV

3. St. Louis, MO

Top five routes

1. LAS to LAX: 388 flights

2. LAX to LAS: 369 flights

3. LAX to SFO: 366 flights

4. SFO to LAX: 318 flights

5. ANC to ORD: 299 flights

It appears that flights between Las Vegas and Los Angeles were most popular for the year 2023. The second most popular route was between San Francisco and Los Angeles which makes sense

due to California's extensive size. Unexpectedly flights between Anchorage and Chicago came in fifth.

Test.rs uses sample csv code to test the functionality of the run function. The test fails, but not because the function doesn't run correctly but rather due to the order it displays the top five routes in.

Output for test:

Airports (nodes) ranked by the number of originating flights (degree):

Node: 1, Name: SJC, City: San Jose, CA, Degree: 1, Passenger: 0, Average Passengers per Flight: 114.00, Destinations: 1

Node: 2, Name: ORD, City: Chicago, IL, Degree: 1, Passenger: 0, Average Passengers per Flight: 3414.00, Destinations: 1

Node: 3, Name: OAK, City: Oakland, CA, Degree: 1, Passenger: 0, Average Passengers per Flight: 1532.00, Destinations: 1

Node: 4, Name: SFO, City: San Francisco, CA, Degree: 2, Passenger: 0, Average Passengers per Flight: 1875.50, Destinations: 2

Top destination cities for SFO (San Francisco, CA):

1. Dallas/Fort Worth, TX

2. Denver, CO

Top destination cities for SJC (San Jose, CA):

1. Lihue, HI

Top destination cities for OAK (Oakland, CA):

1. Houston, TX

Top five routes:

1. SJC to LIH: 1 flights

2. SFO to DFW: 1 flights

3. ORD to SNA: 1 flights

4. OAK to HOU: 1 flights

5. SFO to DEN: 1 flights

Total flights in the CSV file: 5

Average passengers per flight: 1762.20

Expected output for test:

Node: 1, Name: SJC, City: San Jose, CA, Degree: 1, Average Passengers per Flight: 114.00, Destinations: 1\n\

Node: 2, Name: ORD, City: Chicago, IL, Degree: 1, Average Passengers per Flight: 3414.00, Destinations: 1\n\

Node: 3, Name: OAK, City: Oakland, CA, Degree: 1, Average Passengers per Flight: 1532.00, Destinations: 1\n\

Node: 4, Name: SFO, City: San Francisco, CA, Degree: 2, Average Passengers per Flight: 1875.50, Destinations: 2\n\

Top destination cities for SFO (San Francisco, CA):\n\

1. Dallas/Fort Worth, TX\n\

2. Denver, CO\n\

Top destination cities for SJC (San Jose, CA):\n\

1. Lihue, HI\n\

Top destination cities for OAK (Oakland, CA):\n\

1. Houston, TX\n\

Top five routes\n\

1. SFO to DEN: 1 flights\n\

2. OAK to HOU: 1 flights\n\

3. SJC to LIH: 1 flights\n\

4. ORD to SNA: 1 flights\n\

5. SFO to DFW: 1 flights\n\

Total flights in the CSV file: 5\n\

Average passengers per flight: 1762.20

As you can see everything matches except the order of the top destinations and routes. That is solely because the data I used only had one entry per route so it randomly sorts them. I tried to have it sorted by alphabetical order, but that resulted in more errors. So I just left it since all the other aspects of my function were correct.

This is the temporary file I used:

"PASSENGERS,UNIQUE_CARRIER_NAME,ORIGIN_AIRPORT_ID,ORIGIN,ORIGIN_CITY_NAME,DEST_AIRPORT_ID,DEST,DEST_CITY_NAME\n\

114,Southwest Airlines Co.,14831,SJC,\"San Jose, CA\",12982,LIH,\"Lihue, HI\"\n\

1990,Frontier Airlines Inc.,14771,SFO,\"San Francisco, CA\",11298,DFW,\"Dallas/Fort Worth, TX\"\n\

3414,United Air Lines Inc.,13930,ORD,\"Chicago, IL\",14908,SNA,\"Santa Ana, CA\"\n\

1532,Southwest Airlines Co.,13796,OAK,\"Oakland, CA\",12191,HOU,\"Houston, TX\"\n\

1761,United Air Lines Inc.,14771,SFO,\"San Francisco, CA\",11292,DEN,\"Denver, CO\"\n"

In total my project has 1321 nodes. Chicago is the most popular origin airport at a total of 13910 flights with over 240 destinations. The data was pulled from the United States Department of Transportation although some entries seem suspicious. The csv file is sorted by the number of passengers and the last entry is 80163.00,Hawaiian Airlines Inc.,12173,HNL,"Honolulu, HI",13830,OGG,"Kahului, HI" which implies there was 80,163 on the flight which is highly improbable as there is usually “160 seats per scheduled flight” (Grant). I revised my data to make sure the number of passengers weren’t the total number of passengers taking that flight route but just a couple entries before there is “71156.00,Hawaiian Airlines Inc.,12173,HNL,"Honolulu, HI",13830,OGG,"Kahului, HI" which is operated by the same airline. I looked on the Bureau's website and could not come to a conclusion of why this was the case.

Total flights in the CSV file: 425900

Average passengers per flight: 194658.35

I doubt this data actually reflects the average passengers per flight in the U.S but my model does run correctly. Since I could verify that some entries seem unrealistic I could confirm the issue lied on the data rather than my code. The data is directly pulled from the government website so I’m not exactly sure where there could have been a more reliable dataset.

More output:

Before updating, flights for 13830: 1059

After updating, flights for 13830: 1060

Before updating, flights for 12173: 1958

After updating, flights for 12173: 1959

Total flights processed: 425900

Top three nodes:

Node: 1319, Name: ATL, City: Atlanta, GA, Degree: 10571, Passenger: 4509471800, Average Passengers per Flight: 426588.95, Destinations: 218

Node: 1320, Name: DEN, City: Denver, CO, Degree: 13226, Passenger: 3613867700, Average Passengers per Flight: 273239.66, Destinations: 238

Node: 1321, Name: ORD, City: Chicago, IL, Degree: 13910, Passenger: 2947746600, Average Passengers per Flight: 211915.64, Destinations: 240

Works Cited

Dataset:

https://www.transtats.bts.gov/DL_SelectFields.aspx?gnoyr_VQ=GEE&QO_fu146_anzr=Nv4%20Pn44vr45

Grant, John. “Why Is Average Flight Capacity Increasing at Its Fastest Rate Ever?” OAG, OAG Aviation Worldwide Limited, 28 Mar. 2024, www.oag.com/blog/average-flight-capacity-increasing-at-fastest-rate-ever#:~:text=However%2C%20as%20the%20table%20below,160%20seats%20per%20scheduled%20flight.