Project 2

DASC 5300-002

Dr Sharma

Pushpa Ghimire

Student ID 1000441575

Project Report:

Overall Status:

This project was extremely educational and I learned lots of intricacy on Python specially in these regards:

1. Moving data and taking subsets, joins and just shuffling them from datasets to datasets.
2. Inner workings of Networkx package that support the graphing functionality.

First I wanted to layout the steps that I was going to follow and made sure to verify this with Teachers Assistant Mr Enamul Karim. Here are the steps.

1. Overall identify the airline assigned to me. – AL2
   1. Process the route records.
   2. Sort the count of departures from source airport code for AL2
   3. Look at the top counts and see which airlines operates from those airports. Most likely the high frequency of departures from a specific airport indicate that is most likely their hub.
   4. Identify the airlines.
2. Next hub for AL2:
3. Graph: Based on information from Step 1 above after sorting the and identifying the airlines.
   1. Plot a graph on their flights (20) from their top two hubs. I tried plotting many more but really ran into graphs that were extremely busy. So for the sake of clarity and so I can study the characteristics of graph properly, I decided to go with 20.
4. After finding the top departures and identifying the airlines with their current hubs, it was easy to find the next hub. Researched area, locality, type of travel and its proximity to other airports.
5. The only issue I ran into was not able to fit many information in a graph. This is something I plan to research a bit more on spare time and see what are the limitations or recommended practices for drawing nodes and edges in graphs like these.

Based on the analysis, I found that ORD (Chicago O’Hare, Newark International, Denver Internation, Houston Intercontinental) were some of the top departure airport codes. It was obvious based on these information, the airline was United Airways.

File Description:

The only file used were the files provides, routes and airportmapping as inputs. I didn’t create any output files. However, I did create multiple pandas datasets within the program to handle and data lift and shifts. These are some of the files involved:

1. Attached program files
2. Airline.html – pydoc documentation.
3. Output.txt – file written out by Graph characteristics program.
4. Test.txt – file input to graph characteristics program.
5. Output.txt\_layer0\_deg\_dist.png – graph file output.

Test cases used were:

1. Isolating the airlines I was assigned. It had 2180 records. I verified this also using “excel” externally and doing several array counts with the length function.
2. Number of departures from the originating airports:
   1. Likewise with this, ran external test using “excel” also used count(), size() etc to ensure accuracy and validity.
   2. Graph: Verified that the graph had proper nodes and edges using
      1. number of nodes,
      2. number of edges
      3. list(graph.edges)
      4. list(graph.nodes) -> this showed the source to destination airport map.
   3. Verified the graph output with the pandas dataset connecting source and target airport and list(graph.nodes) function.

Division of labor:

This was an individual work by me.

Problems encountered:

I didn’t reference any other work. This was in entirety research that I had done. A problem that took up lots of my time was loading the graph edges by looping through the file. There were two issues.

1. Using graph function add\_edges\_from from:
   1. My understanding was that if you supply a list with two fields, it would automatically parse through the two fields and set the nodes and edges. First it didn’t want to take a list because of it being hashable. Second, even when I passed the list by cursoring through the records and grabbing a record at a time, it was really splitting each column into further separate letters. That I realize after an hour or so of struggle.
   2. The second issue was it was adding some numeric nodes to the graph this creating multiple graphs. One was the correct graph with the source to target airport. The other was numeric. This was an issue created upstream while processing some datasets which I had to remove.

These two were the major things that took of lot of my time. Most of the other things were fun. Networkx documentation was really good just to follow.

Analysis:

As described in the overall status part the approach that I took to handle the project helped me navigate through compiling the question/answers that I wanted to get.

Once putting the top 20 hubs (based on volume of out-flights), I quickly noticed that the first 8 (top 8) busiest airports were already their current hub.

The next airport that was not their current hub but had significant number of flights was Cleveland, OH. If they had to move their hub from existing hub, I would recommend moving from Chicago Ohare to Cleveland, OH. Chicago, IL is one of the busiest airport and a largest city in the world. I wouldn’t want to move out of there. But if I have to, I would move to Cleveland because of the location proximity it provides to Chicago. A mere less than 30 minutes flight.

Cleveland is also one of a large cities in the country and has lot of corporation and natural and tourist attraction. It has rich American history, from baseball to football.

The total out-flight currently from Cleveland is at 49.

Current Network Path:

Below is the network graph for few “United Airways” flight flight out of Houston Intercontinental airport and Chicago O’Hare International airport.

Due to the volume of data we had and the complexity of the graph, to demonstrate the graph, I, based on my discussion with My Emanul Karim picked few destination airports.

Chart, radar chart

Description automatically generated

The total number of flights in the route file provided to us had 2180 United Airways flights flying out of one or more source airport. Out of those here are the top 10 hubs:

![Table

Description automatically generated]()

Out of these the first 8 airports are United airways’s current hub. The next hub we would like UA to make is Cleveland.

Graph Characteristics:

Below are the characteristics of the graph:

The total number of nodes for United Airlines is 432 have route connection of 1115, which are the edges. Below are the stats retrieved:

number of nodes of layer 0 : 432  
number of edges of layer 0: 1115  
Number of Connected Component of layer 0 :1   
Density of layer 0 : 0.011976884076652058  
diameter of layer 0: 7Chart

Description automatically generated

Pydoc:

>pydoc3 -w Airline