Final Project Requirements
CS 4491
Due: Monday April 26th @10PM

Contents

Teams

There will be 3 teams with 3 members and 1 team with 4 members.

Team A:

- Altunian, Bram
- Buadoo, Tyra
- Xiong, Alex

Team B:

- Deberry, Damen
- Devine, Nicholas
- Halim-Rotinsulu, Jordan

Team C:

- Askew, Jakeira
- Hell Makong, Ange Armel Gatien
- Jemal, Muna
- Todd, Louis

Team D

- Sokeng, Paul Anthony Gaylord
- Stincer, Thomas
- Villanueva, Miguel

Team E:

- Hall, David
- Williams, Thomas
- Wu, Wei long

Submission Requirements

Group Submission

- All .ecl files well commented
- All WUIDs for each process
- Presentation deck
- A details document explaining all your files and work for each problem

Individual Submission

- Self evaluation
- Peer evaluation

Rubric System

- Each problem is 30 points, which includes:

    - Judges avg grade
    - Reviewing submitted materials
- Project Overall 10 points:
    - Presentation
    - Self & Peer Evaluations

Presentation Requirements

Judges grade will be based only on your presentation, so take advantage of the time and cover everything.

Presentation **MUST** include:

- Team members name
- Solution for each problem should include
    - Stating the problem
    - Code screenshots
    - Visualization/graphs
    - Process of problem solving
    - conclusions you draw for each problem
- Presentation and speech quality

Presentation Time and Order

Presentation time and orders will be announced later.

Project Description

The aviation industry is experiencing its biggest crisis on record due to the Covid-19 pandemic, which has resulted in drastically reduced passenger demand and widespread aircraft groundings. The situation has been exacerbated by the lack of harmonisation of government travel restrictions and quarantine requirements around the globe.

The industry is therefore desperate for meaningful insights into where, when and how operations are likely to be restored as demand for leisure and business travel returns. Cirium is the aviation industry's analytics engine and so its key stakeholders are looking to us to help them understand potential recovery scenarios.

Very large airlines, particularly in the US, have traditionally run on the so-called 'hub & spoke' operating model. This involves a system of connections to destinations around *hub* airports, which are generally located at cities that are the largest or most economically viable in their region. Under this model, airlines typically require travellers to make a stop at a hub to connect between two cities, creating the *spokes* of the system. This approach provides efficiencies for large carriers, while passengers only need to check in once – at their originating airport – and collect their baggage at their ultimate destination. It also offers travellers more flight frequencies due to multiple connection opportunities.

The second popular model of operation is *point-to-point*, which emphasises flying directly

between the passenger's origin and intended destination. The advantages to travellers are faster flight times and reduced overall waiting times at airports. The development of smaller, more fuel-efficient aircraft offering extended range capabilities has led to a rapid expansion of point-to-point operations, spearheaded by a new generation of 'low-cost carriers' which generally avoid routing their passengers through major connecting hubs.

With these points in mind, we would like you as Big Data analysts to work with Cirium's global flight schedules file as published on 15th June 2019. This data set is the collective statement from all airlines of which routes and aircraft they intended to fly from that day forward up to 12 months ahead (as the schedules constantly evolve, their reliability tends to decline beyond approximately 3-4 months out). You can compare this file with the equivalent schedules file published one year later, in the middle of the Covid-19 pandemic.

What we are looking for is:

1. Any evidence that the big airlines started to alter their operating models due to the pandemic. Were airlines still following the hub model, or were they adopting more of a point-to-point model (the hypothesis being that travellers would want to reduce their connection points and time at or near airports)? What conclusions can we draw when comparing datasets?
2. Looking at the overall routes (origin & destination) for specific airlines (these can be the same or different from 1.), can you identify scheduled routes that had been completely removed in the June 15th 2020 dataset? (For example, in 2019 London-Dubai was a popular route, but in 2020 there were no airlines planning to operate between these two airports.)
3. For the routes remaining, can you see a reduction in the number of airlines operating a particular route? (For example, London to New York in 2019 had five airlines competing daily, but in 2020 there were only two airlines competing.)

Outputs we would like to see are:

a. Your choices of airlines and their hubs, to analyse. Below are some ideas but we will leave the choice to you:
   - American Airlines (main hub at Dallas, Texas)
   - Delta Air Lines (Atlanta, Georgia)
   - United Airlines (Chicago O'Hare, Illinois)

These airlines have other major hubs, and it would be good to perhaps pick a few more that you can see from the data. Other airlines that operate such models outside of the USA include Emirates Airline and Singapore Airlines, as examples.

b. Your choices of countries/regions to analyse any terminated routes, for example to and from the UK or USA, or South Africa to China etc

c. Your choices of routes to analyse which airlines operated and how often in 2019, compared with 2020

Source Files

2019 and 2020 GSEC record layouts and datasets can be found in GSECFiles.ecl