Background:

United Airlines prides itself on providing a safe, dependable, and comfortable travel experience to its customers. United Airlines TechOps supports that mission by maintaining the airline's fleet. With hundreds of aircraft scattered across the world every day, each with a unique set of required maintenance activities, the task of coordinating, planning, and executing all required maintenance is an impressive challenge. As the airline grows and travel rebounds from the pandemic, the strain on the maintenance system is greater than ever. TechOps has limited resources to support an ever-expanding demand for travel. TechOps leadership is looking to invest the organization's time and money in the most efficient way to support the airline's growth.

Leadership has asked the TechOps Data Analytics team to analyze historical flight data to identify areas of strength or weakness, benchmark United's performance against other carriers, expose possible future risks, and provide recommendations for areas of focus.

Dataset:

2015 Flight Delays and Cancellations Kaggle dataset

- This collection of data contains three different datasets: airlines.csv, airports.csv, and flights.csv.
- Metadata information is available about the files on the Kaggle website: https://www.kaggle.com/datasets/usdot/flight-delays

Objectives:

With these datasets, investigate the following:

1. Are there any data quality issues? If so, how can you address them?
2. In terms of delays, which airlines perform the worst? Which the best? How does United compare to other airlines of similar size?
3. United has 5 hubs. Three of which are Newark, New Jersey (EWR), Chicago – O'Hare, Illinois (ORD), and Houston, Texas (IAH). For each of these three hubs, what are the top delayed routes leaving those airports?
4. Leadership has asked for daily reporting addressing points 2 and 3. What health checks and monitors would you set up when creating a reporting pipeline to fulfill leadership's request?

Please prepare a 10-minute presentation that covers:

- Your findings to the above questions
    - Any data prep/cleaning/code you wrote to answer the question
- Recommendations on areas where the business could focus its efforts based off your findings.

Be prepared to answer a few follow-up questions on your work.

Expectations:

- In addition to your findings and recommendations, we are interested in seeing your process for this exercise. Please explicitly note any caveats and assumptions, and provide insights into your thought process used in developing your answers.
- Python and/or R should be the primary tool you use to complete this exercise. However, you may also use any other tools/applications (Excel, Tableau, Alteryx, SQL, etc).
- You are allowed to use external resources to complete this exercise (e.g. looking something up on StackOverflow)
- **Please submit all files, including code, visualizations, and slides that you used to answer the questions by 9AM Central the day of your interview. You should email these materials to greg.jackson@united.com.**