Instructions to Candidate

The Data Challenge will give you an opportunity to showcase your skills and abilities in areas that align with how we challenge our Data Analysts at Capital One. We have created the Data Challenge to help us find great people to join our team as we continue to develop data products. As you work on your submission, keep in mind that we will be evaluating every applicant on the following key areas:

1. **Builder Mindset:** Utilize the right open source tools to create adaptive and innovative solutions that successfully run. Write code with effective formatting and structure, that contains sufficient comments, and is concise. Also, write code that leverages functions or other approaches that makes it reusable, as well as effectively join datasets.
2. **Data Management**: Systematically perform data quality checks, document issues, and take deliberate steps to resolve issues. In addition, create metadata for any fields that you create.
3. **Business Intelligence**: Create a variety of visualizations that tell a story and provide recommendations that address the business problem. Also, document assumptions and provide ideas for future next steps.

This challenge is your next step in showing Capital One what you can do. You have 1 week to submit a working data product, per the submission instructions.

Problem Statement

You are working for an airline company looking to enter the United States domestic market. Specifically, the company has decided to start with 5 round trip routes between medium and large US airports. An example of a round trip route is the combination of JFK to ORD and ORD to JFK. The airline company has to acquire 5 new airplanes (one per round trip route) and the upfront cost for each airplane is $90 million. The company's motto is "On time, for you", so punctuality is a big part of its brand image.

You have been tasked with analyzing 1Q2019 data to identify:
1. The 10 busiest round trip routes in terms of number of round trip flights in the quarter. Exclude canceled flights when performing the calculation.
2. The 10 most profitable round trip routes (without considering the upfront airplane cost) in the quarter. Along with the profit, show total revenue, total cost, summary values of other key components and total round trip flights in the quarter for the top 10 most profitable routes. Exclude canceled flights from these calculations.
3. The 5 round trip routes that you recommend to invest in based on any factors that you choose.

4. The number of round trip flights it will take to breakeven on the upfront airplane cost for each of the 5 round trip routes that you recommend. Print key summary components for these routes.
5. Key Performance Indicators (KPI's) that you recommend tracking in the future to measure the success of the round trip routes that you recommend.

Link to metadata and datasets:
https://github.com/CapitalOneRecruiting/DA-Airline-Data-Challenge

Here is background information on the three datasets that you will analyze:
1. Flights dataset: Contains data about available routes from origin to destination. For occupancy, use the data provided in this dataset.
2. Tickets dataset: Ticket prices data (sample data only as the data is huge). **Consider only round trips in your analysis.**
3. Airport Codes dataset: Identifies whether an airport is considered medium or large sized. Consider only medium and large airports in your analysis.

Please do not use any data other than what has been provided to you.

When joining these datasets together, use your best judgment on the join condition and document your choice.

Again, keep in mind that these are real-world datasets that come with outliers and data issues that you need to address.

You can make the following assumptions:

- Each airplane is dedicated to one round trip route between the 2 airports
- Costs:
  - Fuel, Oil, Maintenance, Crew - $8 per mile total
  - Depreciation, Insurance, Other - $1.18 per mile total
  - Airport operational costs for the right to use the airports and related services are fixed at $5,000 for medium airports and $10,000 for large airports. There is one charge for each airport where a flight lands. Thus, a round trip flight has a total of two airport charges.
  - For each individual departure, the first 15 minutes of delays are free, otherwise each minute costs the airline $75 in added operational costs.
  - For each individual arrival, the first 15 minutes of delays are free, otherwise each minute costs the airline $75 in added operational costs.

- Revenue:

- Each plane can accommodate up to 200 passengers and each flight has an associated occupancy rate provided in the Flights data set. Do not use the Tickets data set to determine occupancy.
- Baggage fee is $35 for each checked bag per flight. We expect 50% of passengers to check an average of 1 bag per flight. The fee is charged separately for each leg of a round trip flight, thus 50% of passengers will be charged a total of $70 in baggage fees for a round trip flight.
- Disregard seasonal effects on ticket prices (i.e. ticket prices are the same in April as they are on Memorial Day or in December)

Instructions

As you start the challenge, realize that this is real-world, imperfect data. Please plan to spend around 8 hours to complete, however, time may vary by candidate. The project is not timed, however, we simply ask you to return it to us within 10 days. If you find yourself uncertain of what the "right" answer is, use your best judgment, make an assumption (document the rationale), and keep going.

Overall, we first ask you to show your data skills in these areas:

1. **Quality Check** – bad data can skew results and lead to incorrect conclusions
   - Understand the data while keeping your final output in mind
   - Address any material data issues that could impact your recommendations--highlight at least 3 data quality insights
   - Create metadata for any new fields that you create to complete your analysis. This metadata can be within your code (ex. within Python docstrings) or in a separate document. Please clearly define any new fields.

2. **Data Munging** – join the data
   - Write a function that can link the data together in a scalable way

3. **Craft a visual data narrative** – visualize your insights with easy to understand charts and plots, choosing those necessary to tell the story and omitting those that do not
   - Charts and plots should be generated in your Python or R code, or can be generated in free versions of Tableau
   - Describe key trends or data issues you find using visualizations
   - Use visualizations to show the key metric drivers behind the final round trip routes you chose
   - Summarize your key insights and conclusions based on the data and your analysis

4. **Final Recommendation -** Identify both the origination airport and destination airport for each of the five round trip routes you recommend. Remember to answer the 4 other questions shown in the problem statement, as well.

You can add your conclusion and recommendations on what data to track to measure success as part of your code or in a separate write-up.

5. **What's Next –** We recognize that 8 hours is not a lot of time… and you probably came up with a number of great ideas that you did not have time to implement. Tell us (but do not do any work) what you would do next to inform a better decision or deliver a better product to your company.

Data and Tools

Solutions that require purchase of a software license or purchased access to data will not be accepted regardless of whether or not Capital One uses said software or data. Abide by all applicable laws and regulations regarding the use of software or external data sources. If you have questions about a particular software package, please contact your recruiter immediately.

How to submit

Congratulations on completing the Data Challenge! Please see the following instructions for how to submit your work.

Submission is easy – just email to dataanalysisrecruitingmailbox@capitalone.com a single ZIP file (< 10 MB) containing:
1. Working source code file(s), accompanying documentation, and code outputs
2. Documentation including metadata for data created and your data quality insights
3. Visualizations and key insights from those visualizations

Integrity and Confidentiality Disclosure

Please do not post your code or documents to any public repositories.

To maintain the integrity of the assessment process, please complete the assessment independently, keeping all content confidential. Feel free to consult resources, including people and documents on and off the internet, for advice and examples. However, all aspects of your submission delivered (excluding your packaged dependencies) must be solely your independent work. Using work sourced by others / generative AI, or failure to maintain the confidentiality of

the assessment process and materials, may disqualify you from further consideration and result in other consequences, up to and including termination of employment.

The assessment is personalized for your use only and should not be accessed by, or redistributed to anyone, as all content including instructional material is proprietary and strictly confidential. Please reach out to our Assessment Support Team at careers@capitalone.com, if you're experiencing technical issues or have any questions.

Best of luck,
Capital One Recruiting

 If you require reasonable accommodation to take the assessment, please contact Capital One Recruiting at 1-800-304-9102 or RecruitingAccomodation@capitalone.com. All information provided will be kept confidential and will be used only to the extent required to provide a reasonable accommodation.