Data Engineering Exercise

You are part of an analytics team engaged to support NYC taxis to identify opportunities to improve their revenue or cut costs. This exercise is based on the public NYC taxi data containing trip details and trip fares for 2013. Given the size of the data, we recommend you select a month or week of data for analysis.

https://archive.org/download/nycTaxiTripData2013/trip_data.7z

https://archive.org/download/nycTaxiTripData2013/trip_fare.7z

Your task is to build a data model and conduct analysis to answer the following questions.

1. How are revenue, trips and passenger volume tracking over time?
2. What are busiest locations and hours?
3. How are passengers per trip, payment type, fare and tip amounts distributed?
4. How do times and fares vary by trips?
5. Is it possible to characterise taxi drivers based on work hours and how much they make?

In addition, your data model should lend itself to the following use cases.

- Enable the development of visualisation dashboards showing performance over time
- Enable the development of models to predict fare and tip amounts

Notes for the candidate

- We are looking for the thought process and approach you take.
- We are looking for a data model that supports insights generation and predictive modelling at scale
- We are looking for structured responses and data stories communicated clearly through slides, notebooks, data visualisation or dashboards.
- You are free to work with any tools/computing platform as long as you are able to share the work (including presentations, code, comments, data visualisation etc.) for review.
- Remember that we will use this exercise to assess how you'd work in our team so it is a chance for you to impress us.