

correlation.
one

Datathon Intro

DATATHON: CONTEXT

Based on Correlation One's datathon competition.

This is a team project



Designed for the trainees to apply and showcase their data science skills learned through this program



The details are all on the course website




DATA SETS FOR DATATHON

We give you the context of a problem and supporting data sets

Broad question: How developments in NYC for-hire transport industry relate to broader trends in the public and private transportation industries at large.

You will be analyzing NYC metropolitan area transportation data. The primary datasets are 2 files which contain information about Uber rides from 2014 and 2015.

These are supported by 7 supplemental datasets, which give information about taxi rides, subway rides, non-Uber ride-share, as well as socioeconomic and demographic data of all NYC neighborhoods



WHAT DO YOU HAVE TO DO?

We have the main dataset and several partially pre-cleaned supplementary datasets for your use.

Your Goal: You should pose your own question and answer it using the available data sets.



EXAMPLES OF GOOD VS. BAD SUBMISSIONS

GOOD

1. NYC has both green taxis and yellow medallion taxis. How have trip trends over time differed between these two classes of taxis in relation to Uber ridership?
2. Are there interesting similarities between the servicing patterns of Uber, green taxis and yellow taxis?

EXAMPLES OF GOOD VS. BAD SUBMISSIONS

BAD

1. What fraction in the change of taxi prices over time can be attributed to the existence of Uber?
-
1. Is the demand for Uber higher in the summer than winter?

SUPPORT AVAILABLE

Your team will be assigned to a TA. You will have office hours on select afternoons to meet with said TA to discuss your progress and any issues you have run into. The TA will also help guide your team in the right direction based on this.

We have allocated chunks of class time each day for you to work on this as a group

Please refer to the class schedule for the assigned time for office hours, working on the datathon projects etc.



DELIVERABLES

Your team is expected to keep a Google Doc journal of your progress on the Datathon project, and write 1 - 2 paragraphs at the end of each day in this journal.

Please check the Journals doc on the website. Each journal entry should cover:




What your team accomplished that day



Any interesting discoveries you made that day & how it will impact what you do next



Any road blocks you currently experience



DELIVERABLES (CONTINUED)

At the end of week 3, your team is expected to submit a project scoping plan to your assigned TA. The TA will review this proposal and request that your team make changes based on their judgment of feasibility and informativeness. This proposal should include:

Your team's question you set out to answer

The datasets your team is planning on using (including any externals)

How you plan on using each of these datasets/how they contribute to answering the question

SCOPE OF THE PROJECT

Scoping your project is key to success. It involves the following:



Having a clear mental map of what you want to achieve



Clearly estimating of how long each step will take you to accomplish



Having a time table so you can deliver this on time.

Important: You should get approval from your respective TA about the scope of your project before going deeper into the analysis.

SUBMISSIONS FOR DATATHON



REPORT – this should have three main sections:

A. Topic Question – What is the question that your team set out to answer?

B. Non-Technical Executive Summary – What were your key findings, and why are they important.

C. Technical Report – What was your methodology/approach towards answering the questions



CODE – please include all relevant code that was used to generate your results.

MILESTONES

WEEK 3 Project scoping plan/proposal written

WEEK 5 Datasets wrangled & cleaned, associated section written up

WEEK 7 EDA completed, associated section written up

WEEK 9 Analysis & modeling step completed, associated section written up

WEEK 10 Presentation



CERTIFICATION REQUIREMENTS

For receiving a certificate, your team must have achieved all of the milestones.

At the end of the program, we will conduct a survey whereby we ask each student to rate how each of their teammates contributed to the Datathon and final project; this information will be used in determining certification.

