2.

The data analysis started with locating the data for weather data combined with taxi data and weather data combined with passenger data. Some of the data was scraped off a website into a CSV file where it was then curated based on months. Correlations and graphs were produced based on whether there was an increased, decreased or constant amount of passengers for taxis based on the weather conditions effecting how long each ride would be. The statistics aggregated from the data were proven to produce no clear correlations between weather conditions effecting durations and amount of passengers taking yellow line taxis. However there were several clear correlations produced for green line taxis.

b. Provide a description of the choice of tools/ methods used or a description of any code or scripts written, and describe how your results were stored and managed, min. 3-4 sentences (3%)

An extension in google chrome called Data Miner was used to scrape data generated for a year off the weather underground website. Data was scraped into a CSV file in order to be edited. A Python script was used to create a heat graph with correlations on passengers, weather conditions and taxi durations. Kaggle was used to obtain New York City transportation data. The original data was managed and stored in a shared Google Drive because of the size of the file. The results of the overall project and graphs were stored and managed in folders on Github. Each new result was pushed to the master branch and reviewed by the team.

c. Perform the analysis in a form that can be validated and describe the steps and results your group took to ensure this validation, min. 3-4 sentences. (4%)

The steps taken to perform the analysis were locating the data, parsing through all the data, splitting up the data, separating the metadata from the actual data which is included in a JSON file and then conducting analysis on the data using scripts to produce resulting graphs. After the data was obtained, it had to be scrubbed and curated into a format that was free of metadata and relative to the questions the team came up with. The team wrote scripts to validate the analysis by drawing comparisons between several sources of data obtained.

3. Presentation/ Visualization (8%)

a. Prepare presentations/ visualizations of both the data (and any metadata, information) and the results of the analysis and describe them, min. 2-3 sentences. (3%)

b. Document the management of the presentation/ visualization products and any associated metadata, etc. min. 2-3 sentences (2%)

The document

c. Describe how your presentation/ visualization supports the goal of the data science investigation and highlight any value that was gained, min. 3-4 sentences (3%)