# Overview of the Assignment:

The term project work starts with choosing a topic and finding some data sources that will help answer some questions about that topic.

Whatever you are interested in, explore it!

**Part 1**: Are you working on your own or with a partner? If with a partner provide their name. If on your own, just state that this is the case.

Individual

**Part 2**: Determine the project scope

* In a short paragraph, describe the topic you wish to explore
* Determine three to five business questions that your data warehouse will answer.

1. In this project, I plan to use the datasets of NYC Taxi Trip Duration and NYC Airbnb Open Date. These two data are deeply relative to New York City’s tourism activity. The goal is to explore the relationship between taxi passengers, trip duration and Airbnb rentals in NYC and provide insights by using the data warehouse.
2. a. What is the average trip duration by passenger count

b. What is the occupancy rate of Airbnb rentals by neighborhood

c. How does the occupancy rate of Airbnb vary by month

d. What is the taxi taking rate of each district in New York

e. What is the relation between Airbnb density and the taxi taking rate

f. What is the average price per night of Airbnb rentals by neighborhood

**Part 3:** Data Sources

* Provide two data sources you will be using, for each data source list the number or columns and rows that are in each data source. Provide a header and first 5 rows from each source.

1. NYC Taxi Trip Duration

Train.csv: (columns: 11, rows: 1458644)

Background pattern

Description automatically generated

Table

Description automatically generated

Test.csv: (columns: 9, rows: 625134)

Table

Description automatically generated

1. New York City Airbnb Open Data

AB\_NYC\_2019.csv (columns: 16, rows: 48895)

A picture containing website

Description automatically generated

Table

Description automatically generated

* What is the URL or location of the data?

1. NYC Taxi Trip Duration: [**https://www.kaggle.com/c/nyc-taxi-trip-duration/data**](https://www.kaggle.com/c/nyc-taxi-trip-duration/data)
2. NYC Airbnb Open Data: [**https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data**](https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data)

* What information does this data provide that will help answer one or more of the above questions?

The latitude, longitude, and neighborhood in the NYC Airbnb data can let us know the location of the Airbnb. Drop-off longitude, latitude, and drop-off datetime in the NYC Taxi trip Duration can also let us know where the most common destination of the passenger is, and when do they arrive. According to the location, we might know whether there is relation between Airbnb and Taxi taking rate. We might also get to predict the level of convenience in the neighborhood.

Also,

* Do you see any issues in the data that will require transformation.

The length (amount) of rows in these two datasets has a huge difference. Therefore, we may need to sample the taxi dataset for better combination. Luckily, both datasets have latitude and longitude columns, which make the analysis more easily.

Also, in taxi dataset, train and test datasets have the different columns, which will need to be integrated.

Use the **Ask the Teaching Team Discussion Forum** if you have any questions regarding the how to approach this assignment.

Save your assignment as ***lastnameFirstname\_ProjectUpdate1.docx*** and submit it in the *Assignments* section of the course.

For help uploading files please refer to the *Technical Support* page in the syllabus.

Project scoping is graded based on the following:

1 – On track, 0-Off track, .5 – partially on track