Idea 1: Taxi Trip Tip Prediction
Data Source: http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml
https://www.zillow.com/howto/api/neighborhood-boundaries.htm

People providing services expect gratitude from servees and there's no exception for taxi drivers driving through the concrete jungle. With various variables available in the TLC New York Taxi Trip dataset, we can dig into the correlations between the amount of tip that passengers would be willing to pay along with various variables and build out a prediction model to predict the tip amount.

Question:
How much tip a driver could expect to receive from the passenger?

Idea 2: Airbnb Rental Price Prediction in NYC
Data Source: http://insideairbnb.com/get-the-data.html
https://www.dataquest.io/blog/machine-learning-tutorial/
https://www.zillow.com/howto/api/neighborhood-boundaries.htm

Serving as a short term rental marketplace, Airbnb allows hosts to list their properties for others to rent. In order to maximize revenue, hosts would possibly tend to set the daily rental price as high as possible. However, price over market average would result in potential renters to choose alternative listing instead. Given variables like the number of rooms / bathrooms, type of room and historical listing prices, we can build out a prediction model to predict an optimal listing price to increase occupancy.

We can also incorporate additional geographical grains categorized by Zillow to have a deeper picture of the locations of the listings.

Question:
What is the optimal listing price for certain type of room/places given the neighborhood and time NYC?

Idea 3: News Popularity Prediction
Data Sources: https://archive.ics.uci.edu/ml/datasets/online+news+popularity

Digital news article has become a prevailing content disseminating information on the internet. What category of content is more popular or what kind of title will be more likely to capture the attention of people exposed and end up being share by the readers.

Question:
How many time an article would be shared based on the popularity of the article?