1. Prediction model for MVP, “All-NBA Team” for 2017-18

Idea: Besides winning the MVP for the season, making the All-NBA team is also considered to be the highest award for the best players in the league (based on personal performance, culminating the entire season). But in some sense if we breaks down each position by the numbers and assess which players are considered the best playmakers in the league, for example as the analysis in this article (http://bleacherreport.com/articles/2690988-nba-metrics-101-the-best-playmakers-in-the-nba-according-to-the-numbers). I find out a majority of these players may be leading their position from a statistical standpoint, but that doesn't necessarily equate to making the All-NBA team. I want to build a model that ‘predicts’ the MVP and “All-NBA Team” for the past NBA seasons that’s more consistent with the league decisions.

Data: https://www.kaggle.com/drgilermo/nba-players-stats
This dataset is mainly consists of two files, the player data, including college background, height, weight, etc and season data of individual players in the past 50+ seasons, including points, assists, rebounds, steals, turnovers, etc. The information of MVP and All-NBA team is not included but is well accessible online.

2. Predicting flight delay.
Idea: Using past data of all the flights in the U.S, I want to predict the chance of having a flight delay of any given flight. Some variables that I plan to explore are: length of flight, age of airplane, day of week, weather condition, historical flight data. The real world clients could be companies targeting travel business, e.g. Booking.com, Google flights. Based on training on historical data, we could provide prediction of how likely a flight is about to delay and therefore better aid travelers for the trip planning. A similar application could be find in this link(http://projects.fivethirtyeight.com/flights/)

Data:
https://www.kaggle.com/fabiendaniel/predicting-flight-delays-tutorial/data
This dataset is mainly consists of two files, one for airport and another for flights, the airports file contributes the geo information of each airport, and the details about each flight in the year 2015 is included in the flights file.

3. Predicting passenger pickup location and trip duration for cabs
For cab drivers, the goal is to find the best pickup location of passengers to maximize fare amount at any given point in time. I hope to help cab drivers in NYC find the best pickup location and trip duration based on the location.

Data:
https://www.kaggle.com/c/nyc-taxi-trip-duration
The dataset(training) includes fields such as trip id, pickup time and location in longitude and latitude, and drop off time and location.