DS-SF-23 Final Project, Part 1
Potential Topics

Vanessa Grass



Idea 1: 2016 Presidential Election & Twitter

Problem: The uncertainty around the outcome of the upcoming presidential election is a major source of interest, entertainment, and worry, not only for politicians and everyday citizens, but also for news outlets and businesses.

Data: Mine Twitter for tweets regarding sentiment toward specific candidates.

Hypotheses: Sentiment on Twitter can predict the outcome of the 2016 presidential election.

Idea 2: Bay Area BikeShare

Problem: San Francisco is full of bike enthusiasts — tourists (as well as some locals) take advantage the city's small geographic size to easily get around via bike, which they rent through Bay Area BikeShare. In order to ensure that BikeShare stations have enough bikes, Bay Area BikeShare needs to understand the distribution of its inventory throughout the day to meet supply and demand.

Data: http://www.bayareabikeshare.com/open-data

Hypotheses: Given the variables of the above dataset, one is able to predict the start or end station of a particular Bay Area BikeShare user renting a bike.

Idea 3: Airbnb

Problem: Getting users to sign up for products and services is critical to any business. It's even more critical to get users to start *actually* using the product or service. For Airbnb, having a deep understanding of which actions lead to conversion (i.e. booking on the demand side of their marketplace) empowers the company to tweak variables to get the results that yield the greatest return.

Data: https://www.kaggle.com/c/airbnb-recruiting-new-user-bookings/data

Hypotheses: Given the variables of the above dataset, one is able to predict the time delta from when a user is first active on Airbnb to their first booking. Or in other words, how long it takes from first visiting Airbnb, browsing, account creation to ultimately booking an apartment or house.