Capstone 2 Project Ideas – Anushree Srinivas

Mentor: Srdjan Santic

The three ideas that I was looking at for the capstone projects are as follows:

1. **SF Crime Classification** :

   Objective and Project Overview:

   From 1934 to 1963, San Francisco was infamous for housing some of the world's most notorious criminals on the inescapable island of Alcatraz.

   Today, the city is known more for its tech scene than its criminal past. But, with rising wealth inequality, housing shortages, and a proliferation of expensive digital toys riding BART to work, there is no scarcity of crime in the city by the bay.

   From Sunset to SOMA, and Marina to Excelsior, this competition's dataset provides nearly 12 years of crime reports from across all of San Francisco's neighborhoods. The objective of this project is to predict the category of crime that occurred at a given time and location.

   The dataset is obtained by SF OpenData, the central clearinghouse for data published by the City and County of San Francisco.

   This dataset contains incidents derived from SFPD Crime Incident Reporting system. The data ranges from 1/1/2003 to 5/13/2015. The training set and test set rotate every week, meaning week 1,3,5,7... belong to test set, week 2,4,6,8 belong to training set.

   The data fields include, dates, category, descript, DayOfWeek, PdDistrict, Resolution, Address, X(Latitude), Y(Longitude).

2. **TFI Revenue Prediction:**

Objective and Project Overview:

With over 1,200 quick service restaurants across the globe, TFI is the company behind some of the world's most well-known brands: Burger King, Sbarro, Popeyes, Usta Donerci, and Arby's. They employ over 20,000 people in Europe and Asia and make significant daily investments in developing new restaurant sites.

Right now, deciding when and where to open new restaurants is largely a subjective process based on the personal judgement and experience of development teams. This subjective data is difficult to accurately extrapolate across geographies and cultures.

The objective of this project is to predict the annual revenue sales of TFI's 100,000 regional locations using TFI's demographic, real estate and commercial data. Finding a mathematical model with these predictions to help TFI increase the effectiveness of investments in new restaurant sites.

TFI has provided a dataset with 137 restaurants in the training set, and a test set of 100000 restaurants. The data columns include the open date, location, city type, and three categories of obfuscated data: Demographic data, Real estate data, and Commercial data. The revenue column indicates a (transformed) revenue of the restaurant in a given year and is the target of predictive analysis.

3. **Airbnb New User Booking:**

Objective and Project Overview:

Airbnb travelers find themselves rising with the birds in a whimsical treehouse, having their morning coffee on the deck of a houseboat, or cooking a shared regional breakfast with their hosts.

New users on Airbnb can book a place to stay in 34,000+ cities across 190+ countries. By accurately predicting where a new user will book their first travel experience, Airbnb can share more personalized content with their community,

decrease the average time to first booking, and better forecast demand. Airbnb is now wants to predict in which country a new user will make his or her first booking

The data set will be obtained from :

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

There are 5 datasets for this project: age_gender_bkts, countries,sessions, train_users, test_users.

In the test set, you will predict all the new users with first activities after 7/1/2014 (note: this is updated on 12/5/15 when the competition restarted). In the sessions dataset, the data only dates back to 1/1/2014, while the users dataset dates back to 2010.