**Final Project**

Data Description:

**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 dataset provides nearly 12 years of crime reports from across all of San Francisco's neighborhoods.**

**For this project, I have some python scripts which analyze the data and provide informative visualizations and outputs making use of the train and test datasets.**

**The data is split between train and test data sets, hence the results slightly differ when used separately and together.**

**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.**

Data fields

* **Dates - timestamp of the crime incident**
* **Category - category of the crime incident (only in train.csv). This is the target variable you are going to predict.**
* **Descript - detailed description of the crime incident (only in train.csv)**
* **DayOfWeek - the day of the week**
* **PdDistrict - name of the Police Department District**
* **Resolution - how the crime incident was resolved (only in train.csv)**
* **Address - the approximate street address of the crime incident**
* **X - Longitude**
* **Y - Latitude**

**Packages required for the python scripts to run:**

* **Pandas**
* **Numpy**
* **Seaborn**
* **Matplotlib**

**(Using Anaconda on my laptop)**

**The submission has the following python scripts:**

* **top\_crimes\_png.py**

**Data analysis by taking column values and plots of the distribution of crimes by column values**

* **Month\_of\_year.py**

**Data analysis of the occurrence of types of crimes by month of a year and plot of the results**

* **Dist\_by\_DayOfYear.py**

**Data analysis of the number of crimes occurring distributed by days of the year and plots of the results**

* **Crimes\_on\_friday.py**

**Data analysis of the distribution of crimes on Friday after 7PM (Chose this day and time range as people generally tend to be outside on Fridays)**

* **Crimes\_by\_District\_andCategory.py**

**Data analysis of the occurrence of types of crime classified by district and then on the distribution of the crime categories in the districts**

**>>Exception handling taken care and comments in the scripts for better understanding**

**\*\*\*Please make sure to change the input file path in the script before running them…read\_csv(..path)**

**\*\*\*All the output png files attached in the submission**