**[San Francisco Crime Classification](https://www.kaggle.com/c/sf-crime)**

**Project status report**

**Description of data:**

The 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 belongs to test set and week 2,4,6,8 belongs to training set.

Given time and location we have to predict the category of crime that occurred.

**Number of instances:**

*Training Data:* ***878049***

*Test Data:* **884262**

**Brief description of features:**

*Training Data: 9 features*

* **Dates** - timestamp of the crime incident
* **Category** - category of the crime incident.
* Descript - detailed description of the crime incident
* DayOfWeek - the day of the week
* PdDistrict - name of the Police Department District
* Resolution - how the crime incident was resolved
* Address - the approximate street address of the crime incident
* X - Longitude
* Y – Latitude

*Test Data: 6 features*

* Dates - timestamp of the crime incident
* **DayOfWeek** - the day of the week
* **PdDistrict** - name of the Police Department District
* **Address** - the approximate street address of the crime incident
* **X** - Longitude
* **Y** – Latitude

**Type of output variables**

Type: Class Value

The Output would be one of the category of crime from the training set.

The training set has 40 unique values for category.