**Data analysis for crime in SF.**

**Background**

Crime in San Fransisco is a very interesting topic from various perspective. Personally, I visit SF every week. Although the crime is not so notorious as the other same level city in US, it has its own problem. Too much homeless make the cities unsafe even in the daytime, robbery and theft happens sometimes. In this project, I would like to investigate and analysis the crime data in San Fransisco.

**Dataset**

DataSF (<https://datasf.org/opendata/>) provides hundreds dataset of city and county of San Fransisco. In the project I will use data set of police apartment incidents 2016 and 2017 to analysis the crime in SF.

**Objectives**

Tthe relation between types of crime and specific time and specific locations.

**Tools:**

Cassandra, Spark, and any other tools needed by this project.

**Timeline:**

9/1-9/7

do more investigation.

Choose the required tools

clean the data set

9/8-9/14

store the data to DB,

inject the data to spark.

9/15-9/21

use spark to do analysis.

Learn ML in spark to predict the crime in SF.

9/22-9/30

design GUI for visualization.

Prepare for DEMO