Data Science Capstone Project

SAN FRANCISCO CRIME DATA 2018

The project is about in-depth study of the city of San Francisco. It is mainly focused on analyzed to improving and prevent crime of the city .

This project also provide deeper understanding on how many crimes succed in the city.

Reports included are those for incidents that occurred starting January 1, 2018.

AIM OF ANALYSIS

The description of the crimes is varied and the objective of this analysis is to be able to identify if crimes with the same description have a correlation with a particular area in order to act in the future in a preventive manner.

To make this analysis and to group crimes based on different characteristics we can use different types of cluster analysis which we will analyze later.

We don't know if exist any common group of crime wich are related on particular characteristics and we can't assert which number of group is more efficient to do on future an analysis to prevents them.

DATASET

We begin the work by describing in detail the available dataset.

The dataset is open source and can be downloaded and made available on the website of the US state.Dataset has 533 rows and 22 columns.

Each crime is identified by specific ID and Case Number of reports submitted by an agent. The variables 'Description', 'Name of District' and the location (indates with two variables 'Community Area' and 'Location') are fundamentals to do a cluster analysis. The coordinates can be used to build a map.

```
ID                          int64
Case Number                object
Date                       object
Block                      object
IUCR                       object
Primary Type               object
Description                object
Location Description       object
Arrest                       bool
Domestic                     bool
Beat                        int64
District                  float64
Ward                      float64
Community Area            float64
FBI Code                   object
X Coordinate              float64
Y Coordinate              float64
Year                        int64
Updated On                 object
Latitude                  float64
Longitude                 float64
Location                   object
```

There are variable avaiable for the analysis. Each variable is fundamental to our analysis.

With the coordinate we can plot the map of the city with respective cluster of crime, based on Location and Community Area. it is also possible to frame the most skilled districts in detecting the crimes of the city.

METHODOLOGIES USED:

- CLUSTER ANALYSIS with K-MEANS