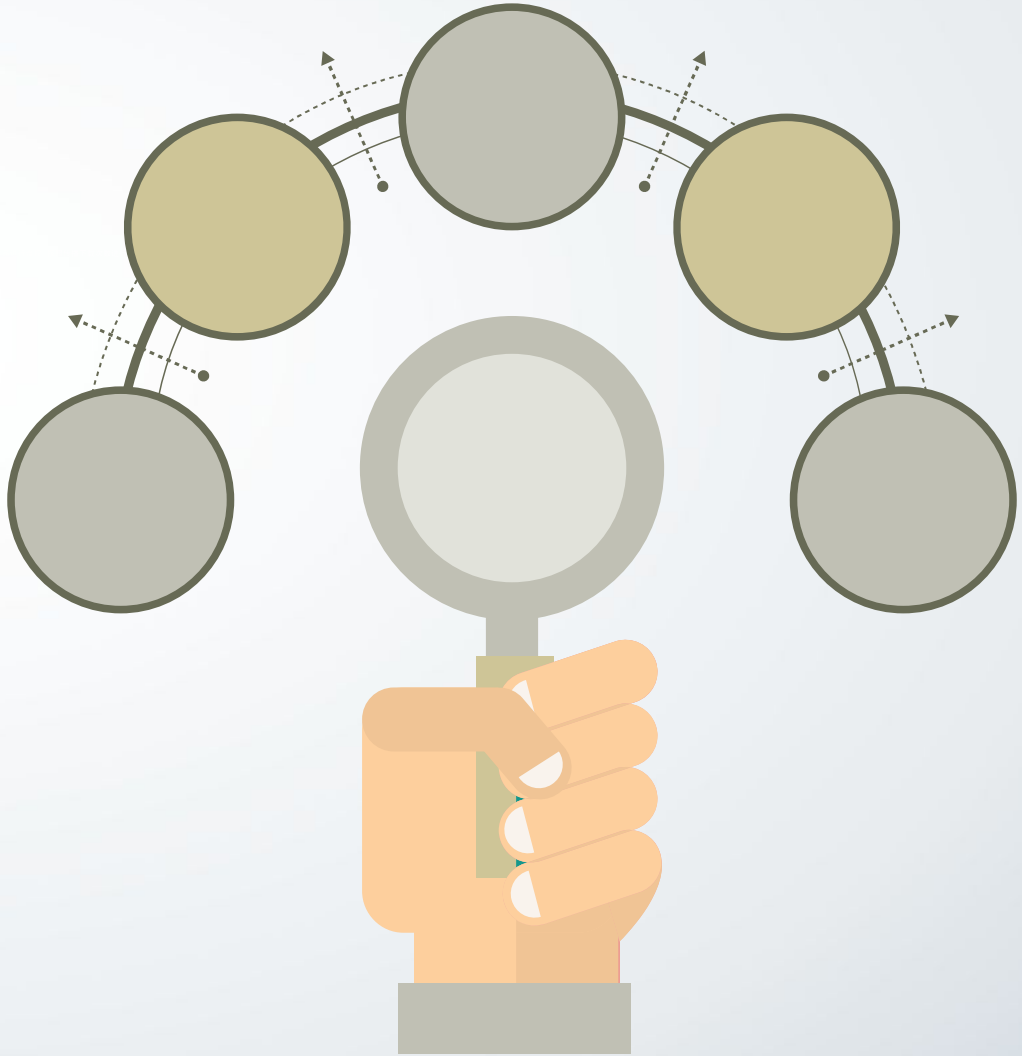
San Francisco Crime Analysis

Project Plan

Team 10
Ning Huang 001839327
Yixuan Wang 001494410

01 Overview

We'll explore the 12-year crime report for San Francisco, build models to classify the types of crime and predict the crime occurrences for specific periods and locations.



02 Use Cases

Visitors can input the area name to check the crime occurrences, system will show the probabilities of several crime types in that area

Users can search for crime records, system will get back to them the records which matched for those inputs information.

System can analyze the data to show the dangerous place in the specific time periods

System runs crime records data (Machine Learning) model and predicts the crime types of those records.

03 Methodology

01

Data Processing

Read and explore Data
Understanding the features and the problems
Clean the Data with Spark

02

KNN(K-Nearest Neighbor classification)

Use KNN algorithm to deterimine the relation between of crime catagory and location

03

Prediction Models

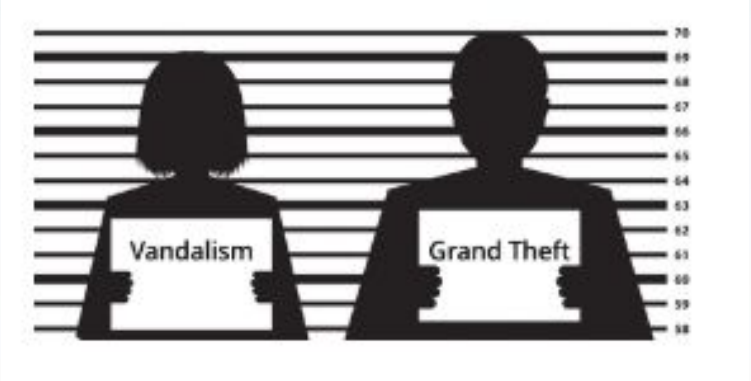
Using ARIMA or other regression model to predict the crime trend.
Regression model to predict the probability of different types of crime occurrences

04

Data Visualization

With Play framework, users can interact with system to search data and will get results to prediction

04 Data Resources



https://www.kaggle.com/c/sf-crime/overview

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.

Total number of rows: 1762313

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

05 Milestones

TimeLine

W1

3/18

Exploratary Data Analysis

- Scope the skeleton of project
- Data Wrangling and Visualization
- Feature Engineering

W2

3/25

Model Building

- Building Models and training data
- Analyzing data with Spark SQL

W3

4/1

Validation and Evaluation

- Evaluating model and tuning parameters for optimization
- Getting farmiliar with Play Framework

W4

4/10

UI Implementation

- Trying to build the UI webpage
- Finalizing the project

06 Program Details

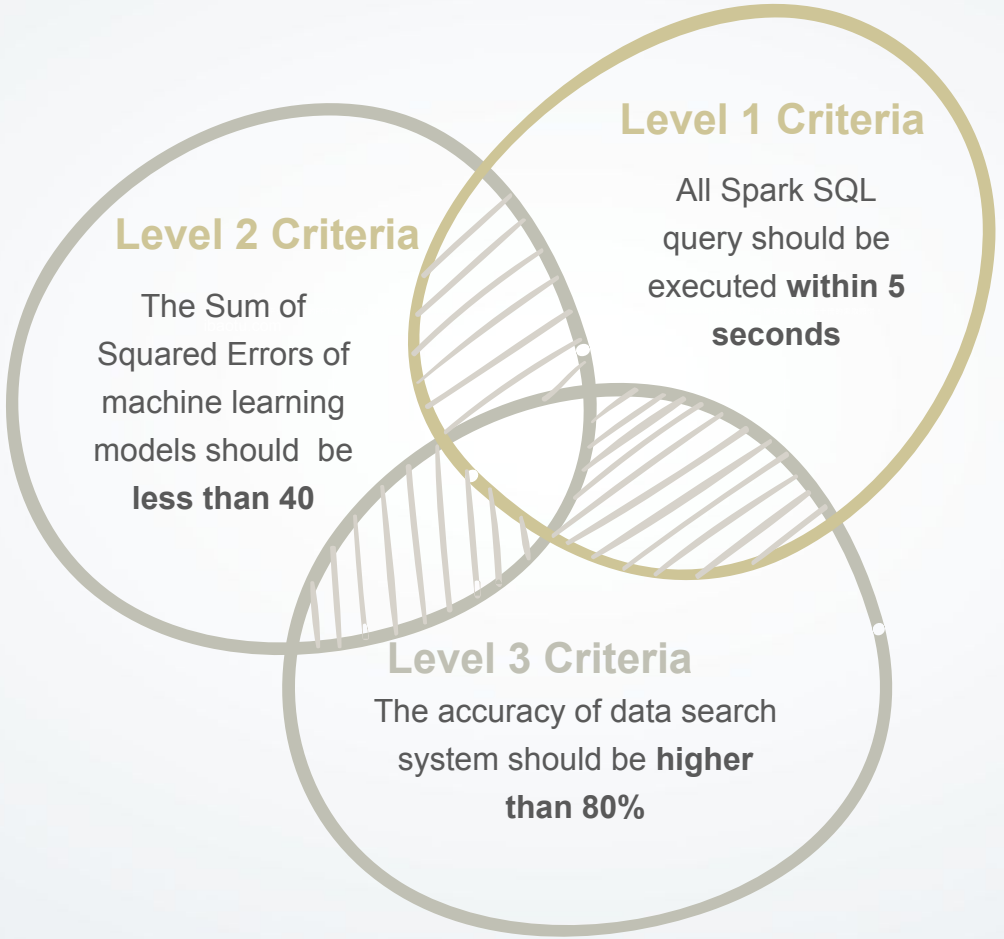
Code Language and Framework

Spark, Scala and Play framework to display visualiation and interaction

Github Repo Link:

https://github.com/HuangNing2024/CSYE7200

07 Acceptance Criteria



08 Goals

1

Analyzing the crime data to reveal the frequency of each type of crime occurrences, and explore its relationship with other factors, like time periods and locations.

2

Building model to classify and predict crime for helping police responds faster and more effective, and reduce the harm of crime incident.

3

Implementing a user-friendly system for providing people a convenient way to check the crime occurrences in specific spots, and make them keep conscious of dangers.

Thank you