Crime Dataset Analysis

CS5830/6830 Data Science in Practice
Project Group: 26
Megh KC
Harichandana Kotnani

Dataset

We used the provided crime dataset. We utilized the following .csv files

- Crime dataset
- Population dataset

Motivation

- We wanted to know more about crimes happened in Austin, Texas in the year 2015
- How the different types of crime are happening? How the poverty level correlated with crime rates? Is the different crime variables are drawn from the same population?

Analysis

- Scatter plot crime per capita at the different zip code areas: how the outlier makes the analysis biased?
  Scatter plot for before and after outlier removal
- How different types of crime are committed in selected places?
  Different bar plots for crime rate comparison in five selected zip codes
- How poverty level & unemployment are associated with crime commitment?
  Pearson correlation test for two variables (crime per capita vs population below poverty level(%)
- Is the any similarities for the “theft” and “theft by person” crime distribution? T-test is drawn for comparison

Data Preparation And Analysis

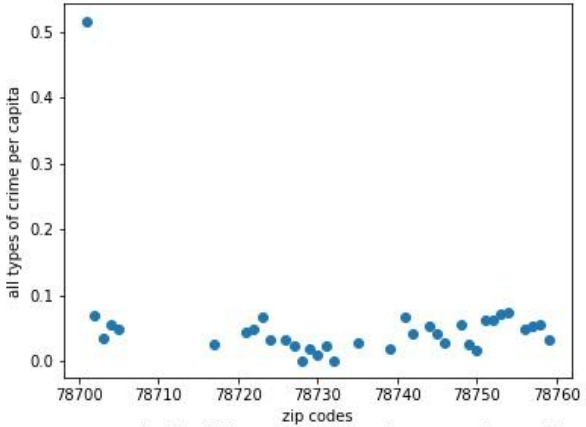
Dataset Preparation

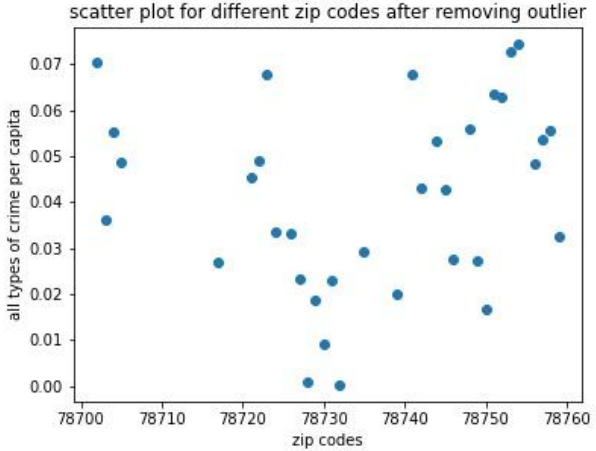
Given dataset had most raw data that we needed, once we did some preparation

- Merged the given two dataset
- Grouped by zip codes and crime_types
- Aggregated certain statistics
- Computed new columns based off existing data
- Create multiple data frames to track each crime type and their statistics
- Sorting & re-indexing of data frames
- Filtering to ensure data validity

Analysis Techniques

- Filtering data for consistency
- Scatterplot to test outliers and their removal
- Zip code selection for neighbouring places
- Comparison of crime commitment by evaluating rates and creating bar plots
- Pearson correlation test for two variables
- Interpretation of pearson test statistic and probability
- T-test for the similar attributes to know whether the variables represents the same population
- Interpreting the t-test statistics and p-value





Results- analysis:2

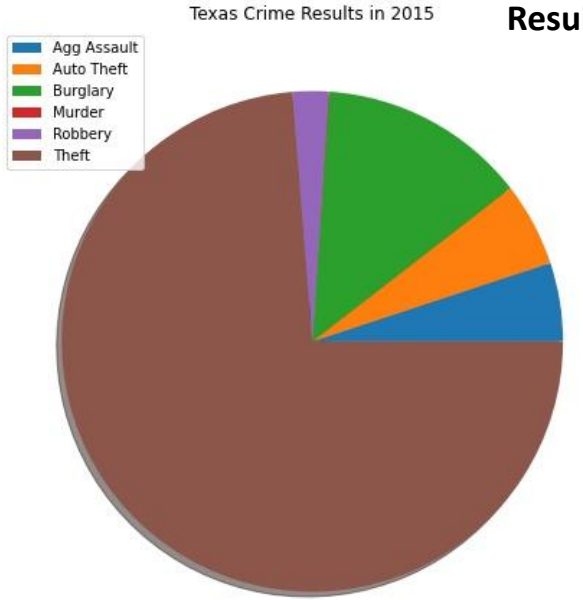
All types of crime per capita ranges from 0 to 0.51 per people

Zip 78729 and 78731 corresponds almost 0

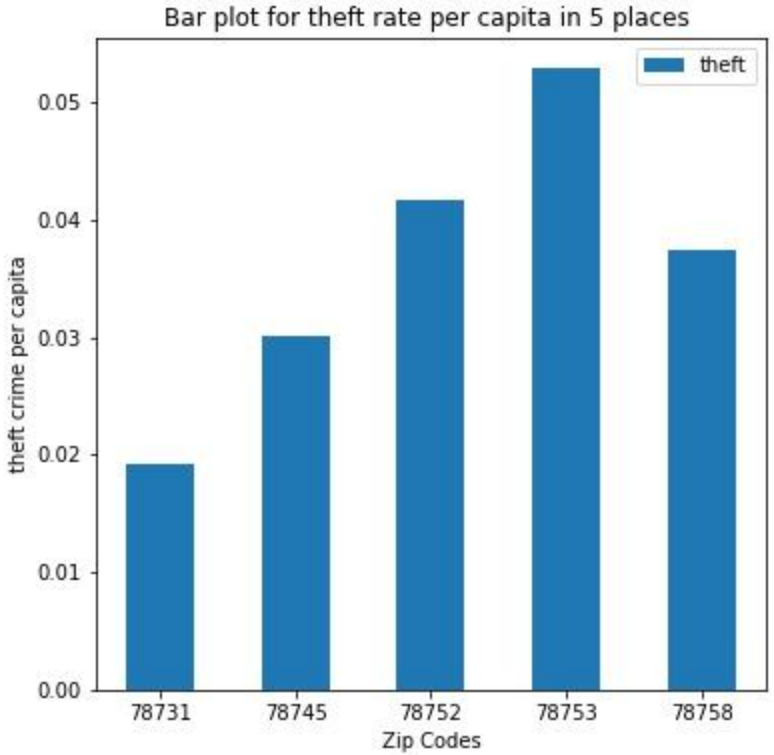
Zip 78701 has at least one crime for every two people (regarded as outlier) : because of lower population

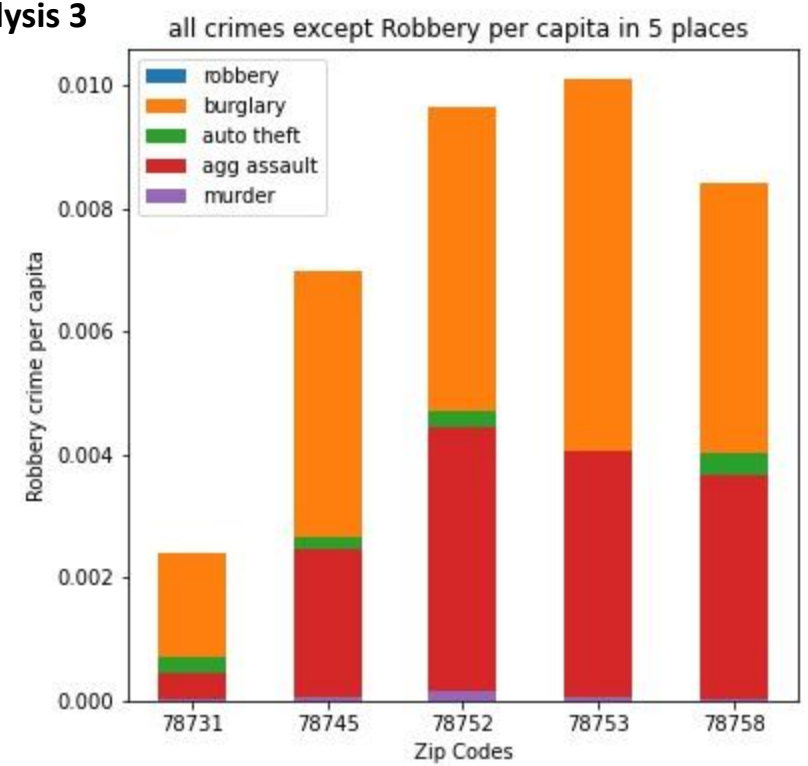
Note: zip codes are categorical, interpretation should be done with visual consideration

Results- analysis:1



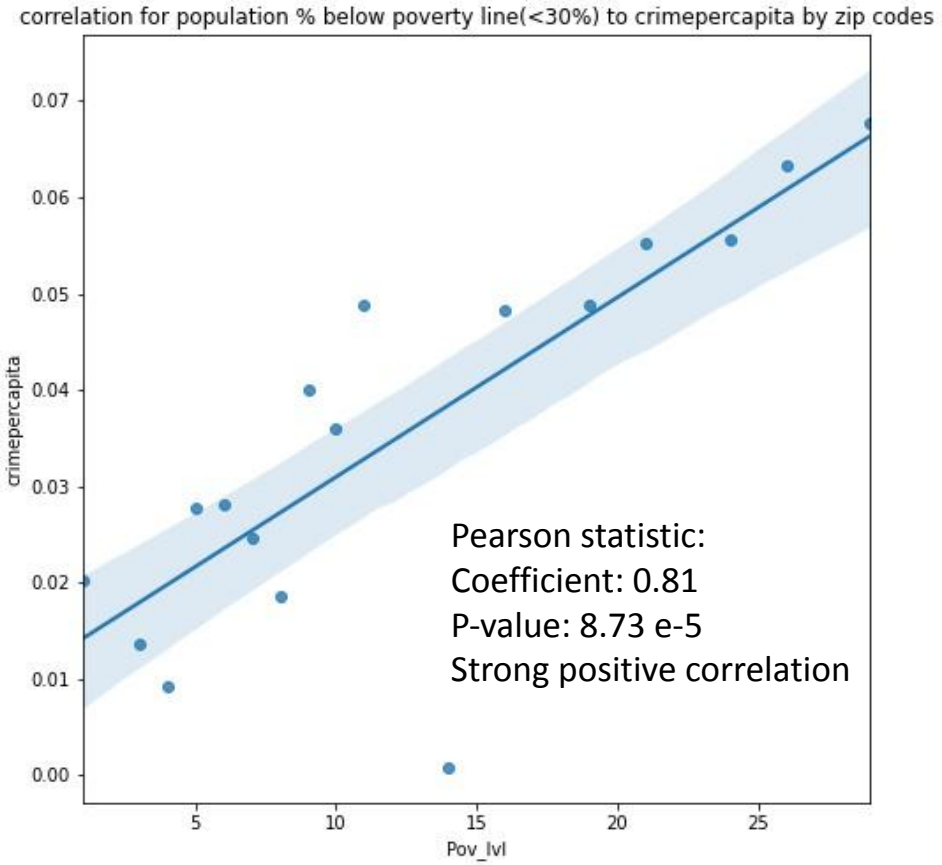
Analysis 3



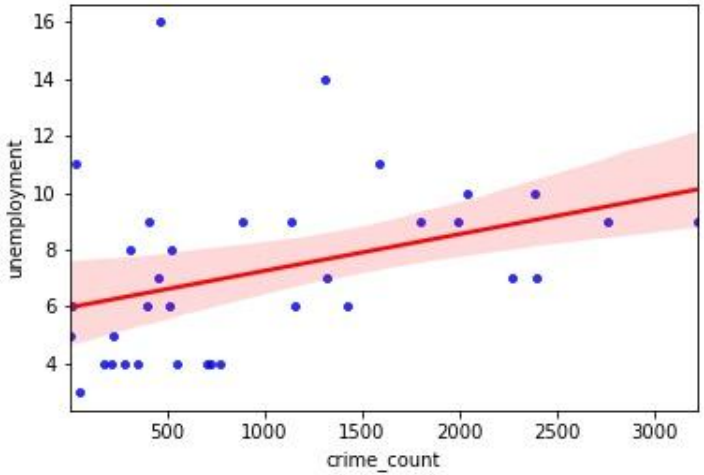


Theft, Burglary & agg assault all are highest for zip 78753 : robbery 0

Analysis 4



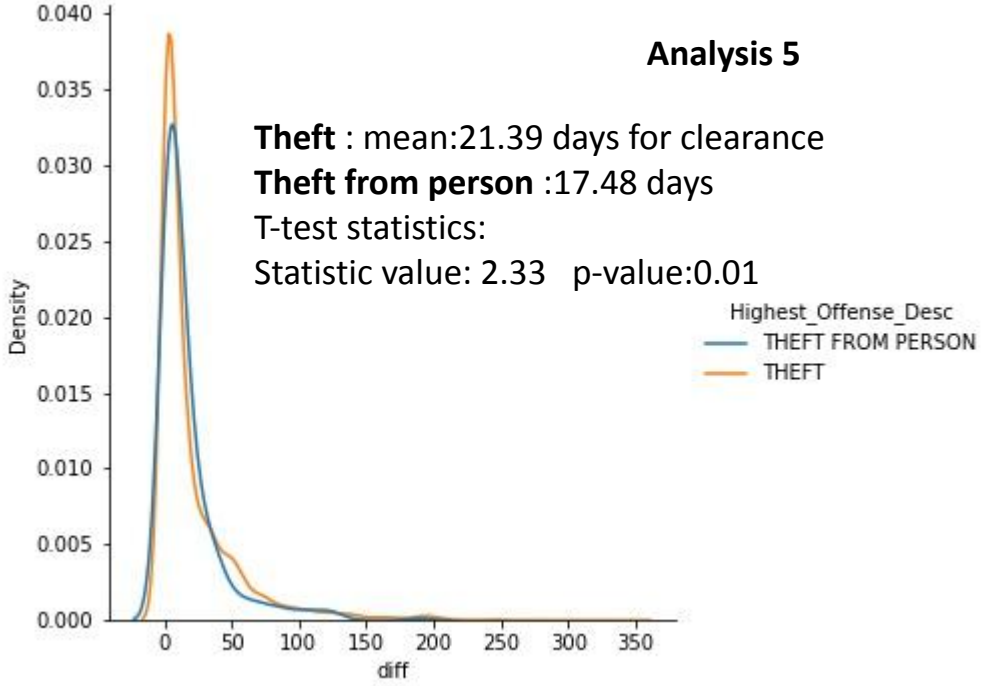
Pearson statistic:
Coefficient: 0.81
P-value: 8.73 e-5
Strong positive correlation



Pearson statistic:
Coefficient: 0.37
P-value: 0.02
Moderate positive correlation

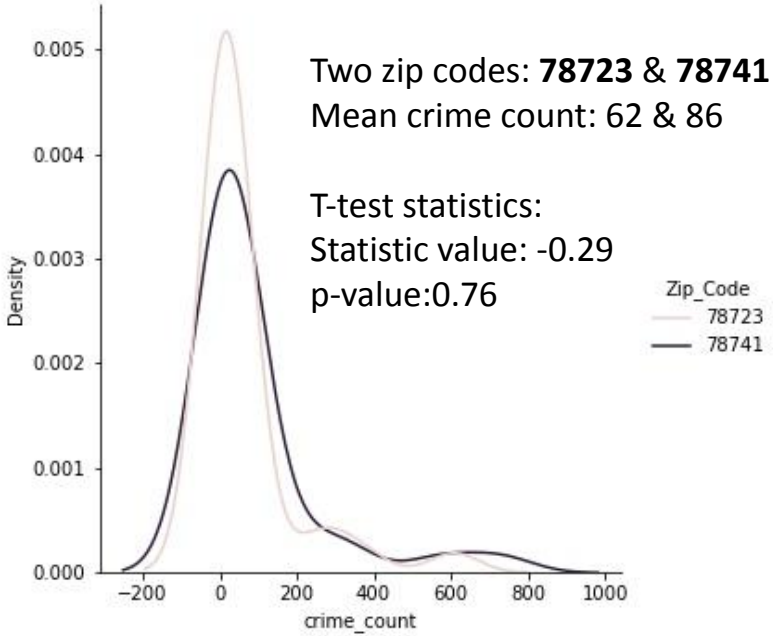
Analysis 5

Theft : mean:21.39 days for clearance
Theft from person :17.48 days
T-test statistics:
Statistic value: 2.33 p-value:0.01



Two zip codes: **78723** & **78741**
Mean crime count: 62 & 86

T-test statistics:
Statistic value: -0.29
p-value:0.76



Thank You

SELECT * FROM Audience.Questions;