Understanding Crime: Leveraging Python in Social Data analysis

- Avinash Laddha

🌐 avinashladdha.github.io   in linkedin.com/in/avinashladdha/

Introduction

- Impressive developments in tools for data analysis and collection are happening and this can greatly benefit the contemporary social sciences.
- Integration of data science in policy making can greatly improve the standard of living with adaptive and personalized policies.
- Policy making which was earlier mostly based on survey data are now starting to analyze and consider other data points as well.
- The objective of this poster is to understand crime in India using python. Analysis of demographic segmentation and trends in subcategories* of crimes is done. Can crime rates be predicted?
- Libraries used: Selenium, pandas, NumPy, json, requests, Matplotlib

* subcategory is further segmentation of crime like economic crimes, crimes against body, crimes against women, crime against public order.

Data Collection and method

- The data has been downloaded from the Open Government Data Platform India on https://data.gov.in/

- The python script uses Selenium and requests libraries for web scrapping and data cleaning makes use of NumPy, Pandas and json libraries.

- The main URL for accessing the Crimes in India webpages on Open Government Data Platform is :
  'https://data.gov.in/catalogsv2?format=json&offset=0&limit=9&sort%5B_score%5D=desc&query=
   crime+in+india&exact_match=1'

- The full code and year wise datasets are hosted on GitHub repo:
  https://github.com/avinashladdha/CrimeDataset

- Currently **41 years** of crime datasets (year wise since 1969) are available for download.

Crime in India
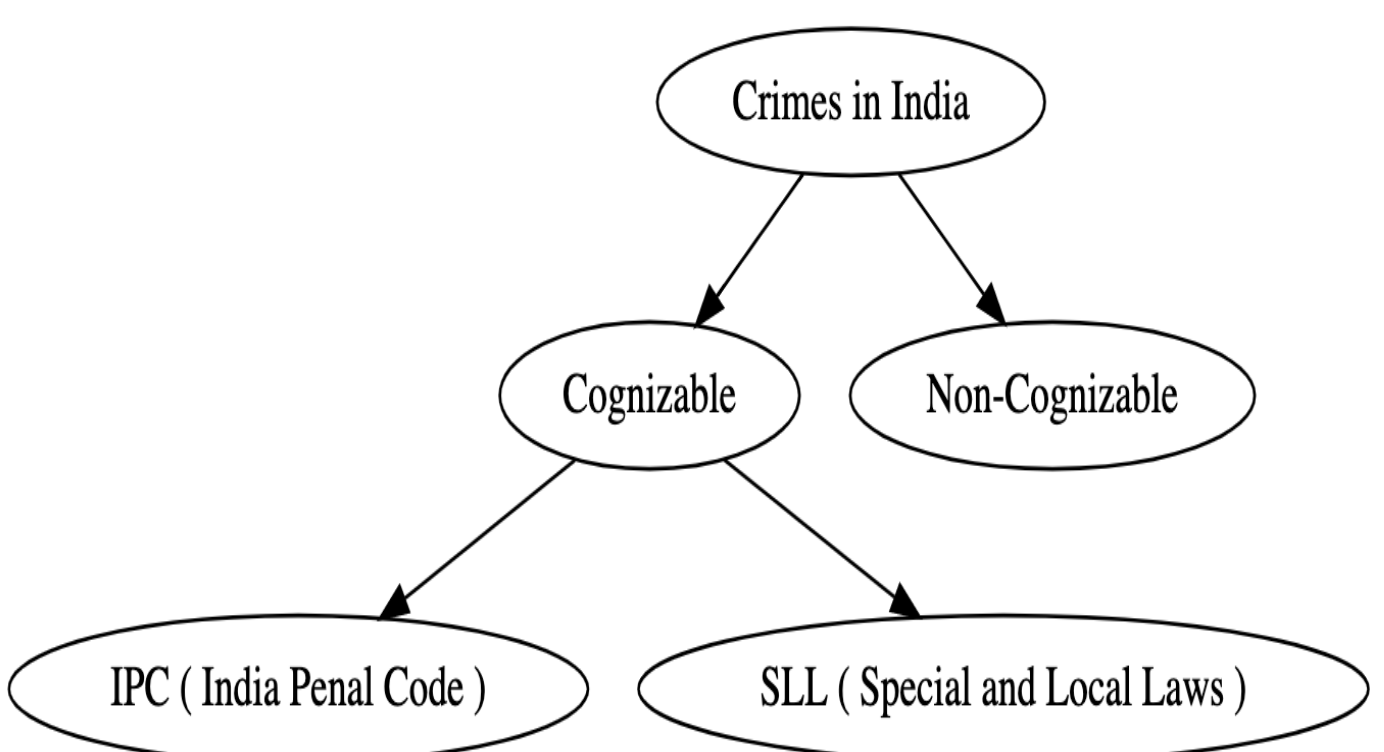
Cognizable Crime : Police have direct responsibility to take an action and affect arrest without warrant

Non-Cognizable Crime : Cannot be investigated by police without the order of magistrate

- Crimes against women and children make the largest part of the crimes committed, closely followed by crimes against property.
- (Other IPC crimes consists of various IPC crimes like accident ,hit and run, attempt to suicide etc.
- Can the rate of crimes against women and children be predicted for the up-coming years and hence laws be put into place to deter the criminals ?
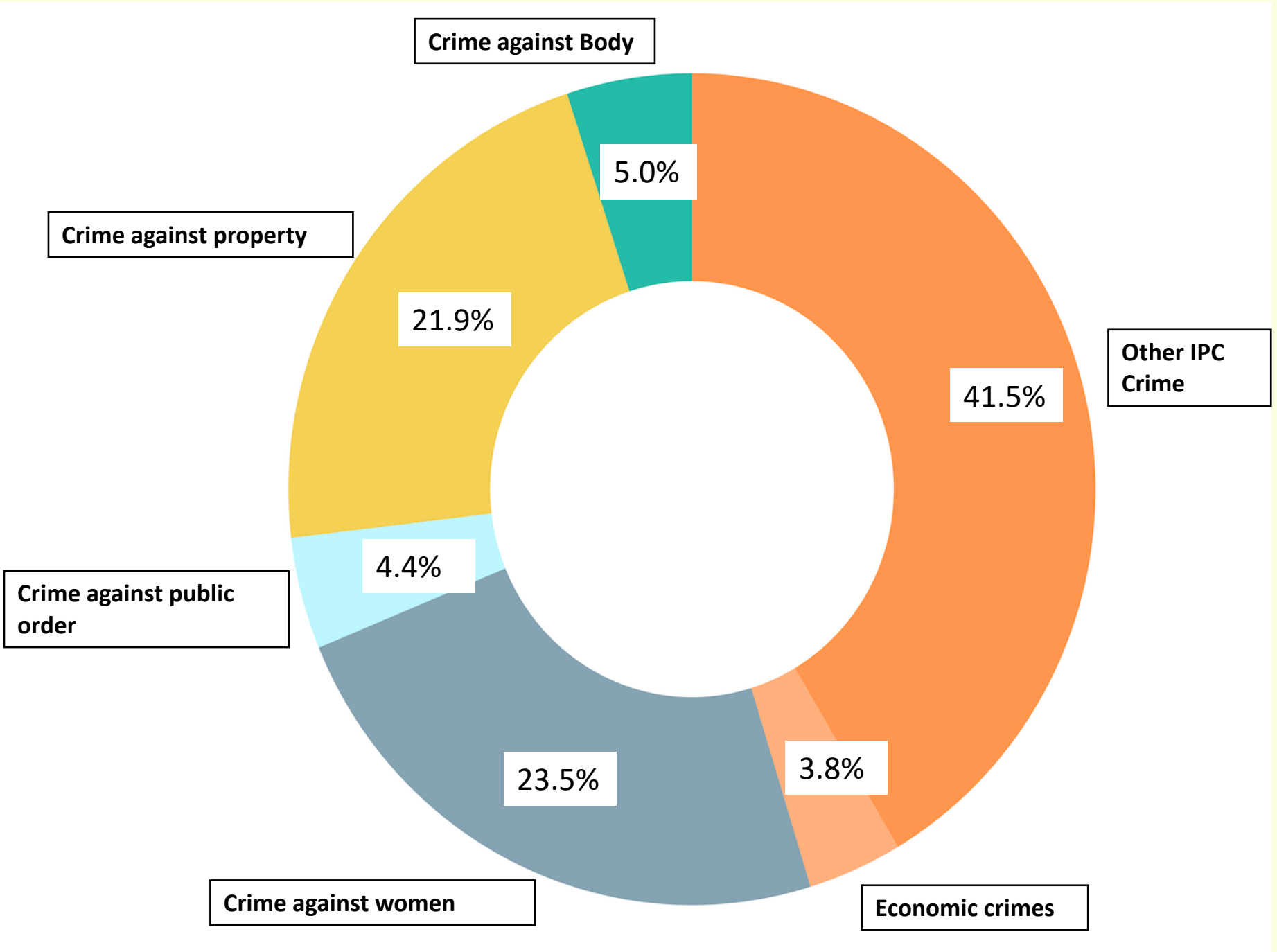
Category	Predicted Value(%)			Actual Value(%)
	Linear Regression	Time Series Analysis	Moving Average	2013
Crimes against body	4.59	5.10	5.00	5.20
Crimes against property	18.78	18.94	19.75	19.50
Crimes against public order	2.60	3.44	3.45	3.10
Crimes against women	28.69	27.29	26.10	26.00
Economic crime	4.73	4.73	4.70	4.90
Other iIPC crimes	40.60	41.48	41.00	41.30

Table 1 - Predicted values of each crime category for the year 2013 corresponding to different models

Scikit-learn and statsmodels python libraries are used for linear regression and TSA ARIMA model respectively. The exact code used to carry out analysis can be found in my GitHub repository.

It can be clearly inferred from table that predicting values using rolling average gives the best results
closely followed by TSA.
This also hints that a society can have a natural rate of crime which is intrinsically ingrained or results from the structural characteristics of a society and can persist even after implementations of stricter laws.

Juvenile Crime

- Juvenile crime serves as a classic example of Strain Theory developed by Robert Merton. The theory says that an increase in income inequality can push a person to resort to crime for profit.
- This is supported by the fact that Economic crimes making more than 40% of juvenile IPC crimes committed in India during 2001-2011.
- Juvenile thefts are usually an early indicator of other problems. These crimes are generally believed to be manifestations of unhappiness, insecurity and discontent. Studies have found that juvenile delinquents are likely to continue to offend until they are young adults.
- Are the juvenile delinquencies spread over the country evenly or is it a pressing problem in some of the states ? The below infographic aims to answers these questions.