Israeli Police data Analysis

Baruch Hanya (303072953), Yaniv Tordjman (204786388), Adar Assif (203181540)
baruch.hanya@post.runi.ac.il,yaniv.tordjman@post.runi.ac.il,adar.assif@post.runi.ac.il
https://github.com/yanivt16/BDP_project
Reichman University
Israel

ABSTRACT

Analysis of crime in the world in general and in Israel, in particular, is the subject of many studies that examine trends in crime, its causes, and different characteristics or features of crime. In this work, we take the database of the Israel Police which is open to the general public and make additions and corrections to the data so that they fit our work and the analysis we are interested in doing.

Specifically, we make a general data analysis that mainly refers to the segmentation of the cities in Israel and the segmentation of the types of crime characterized by the Israeli Police. In addition, we define a "crime score" according to the severity and quantity of the criminal cases for each city according to the population and look for various connections and correlations between the characteristics of the cities such as socioeconomic ranking or geographical location in the country.

Finally, we try to predict the trends in the change in crime for 2023 and define which are the cities with the biggest or the smallest change in crime to ultimately give recommendations to the Israeli Police and law enforcement agencies on how to properly invest their efforts and resources to make life safer in the state of Israel.

Throughout the analysis process, we maintained a professional and academic approach, ensuring that all steps were taken with a focus on accuracy and rigor. The resulting analysis provides valuable insights into the distribution and types of offenses occurring in various settlements in Israel.

KEYWORDS

Big data, PySpark, Crime, Israeli Police, Homeland Security, Socioeconomic Data

ACM Reference Format:

Baruch Hanya (303072953), Yaniv Tordjman (204786388), Adar Assif (203181540). 2023. Israeli Police data Analysis. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX)*. ACM, New York, NY, USA, 4 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Concerning crime trends in Israel, we feel that sometimes they are largely unknown until a few years ago. Sometimes the reports in the media or the general feeling among the public do not correspond to the real situation, which can cause populist decisions that are not prevented by a thorough analysis of the information and the real situation in the country. The lack of reliable information makes it difficult to make decisions about enforcement and crime-fighting strategies.

Given this, in 2019 the Movement for Freedom of Information petitioned the Ministry of Homeland Security to receive crime information from the Israeli Police and publish it to the general public [1]. The police responded to the petition, and since then a database regarding crime in Israel has been published on the official government information website, which is open to the public. The database has over 600,000 entrees, and is updated quarterly and includes information such as the city where the crime was committed, the police station that handled the crime, the year and quarter in which the crime was committed, the statistical type of crime (several types defined by the Israeli government) and the number of cases for each type of crime [2].

2 LITERATURE REVIEW

Analyzing crime data over the years is an academic subject that occupies many research institutes, universities, and government bodies. The results are often used to improve law enforcement mechanisms that include prevention improvements and concentrate the efforts of the various enforcement agencies for more efficient work.

Over the years, data analysis has improved as a result of enhancing the data mining process through more accurate documentation and improving the analysis methods and algorithms used by data scientists.

The article Crime Data Mining: An Overview and Case Studies (2003) [3] describes a step up in data mining and analysis in the US after the September 11 attacks for the benefit of broad learning about crime rates and predicting future criminals. Even then, the researchers mentioned the potential in data collection and analysis future capabilities.

Also, in Israel in recent years, the Israeli Police have begun to publish information about various crimes [2] which have many characteristics that include severity, crime classification, time period, and location (province/city). Our future work will focus on this given data.

The Israeli Police released the information to the general public following a request by The Movement for Free Information to the Ministry of Homeland Security [1]. The appeal was made as part of the principle of freedom of information that the association advocates, and as a result, the connection and the recruitment of the public to the work of the enforcement systems will increase to strengthen faith in them.

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

Israel Police data have been analyzed over the past few years in several works and studies that have explored the correlation between criminal offenses and the characteristics of the population. For example, in the article Crime Trends 1990-2016 [4] it was shown that the general crime rate in the USA in the last 25 years has been decreasing. First, the work presented a general analysis, then analyzed trends into normalized data of families of crime. Similar to the world's data, there is also a downward trend in crime in Israel, as shown in the article on the crime rate in Israel published by the Israel Police [5]. This review focused on property crimes and the reasons behind the drop in crime due to economic trends passing through society.

We have chosen in this work to delve into several research questions concerning the Israeli Police database [2]. Our main question is whether it is possible to identify trends in the crime rate (or, as we call it - the "crime score") according to geographic classification (police district / municipal boundary).

The "crime score" can be measured by determining a score that will consist of a formula that calculates several parameters, similar to the work carried out in the article Design and performance evaluation of crime type and crime risk score estimation technique for a fast and efficient response of severe crimes[6] in which the score is based on the severeness of the offense, the severity of the damage and the identity of the victim (gender, age). After defining the formula, the researchers tried to predict the development of that crime by the district.

Crime segmentation according to geographic classification was carried out in a study Cluster based zoning of crime info [7] on crime data in India in which the researchers used machine learning algorithms to produce heat maps of crime in the country in order to recommend a concentration of effort and enforcement strategies to the government of India.

After the analysis, we will try to hypothesize the factors that influence the difference in the "crime score" between geographical areas - are these differences in socioeconomic status (social status or economic status), or are these certain crime types that characterize districts or cities, with the help of analyzes previously carried out for the Israeli Police [8].

Furthermore, we will examine how the COVID-19 period and the lockdowns affected those trends in the crime rate in Israel after it was found that in several places in the world, the crime rate decreased during the COVID-19 period [9].

3 PROPOSED METHOD

The method we propose includes four main steps: adapting the database to our needs, initial analysis of the information, the definition of a "crime score" and a deep analysis that includes a forecast for the crime rate.

In the stage of adjusting the database - first, we translated the database into Hebrew using the Google Translate library in python. Since the translation performed is only superficial, we performed a general check and corrected several names of districts that were translated incorrectly. Regarding the names of the cities, we used the database of the Central Bureau of Statistics, which has an English transcription for each city, and thus we made an exchange in the names of the cities. After that, we added additional data to each city from databases of the Central Bureau of Statistics such as socioeconomic cluster and the amount of population of each city. In addition, we added a time field to display the offenses and the analysis performed on a timeline.

In the initial data analysis phase - during the entire analysis (as in the rest of our work) we used PySpark, except for visualization purposes where we converted the specific data to Python Pandas to create graphs. The basic analysis includes checking the number of crimes per capita in each city, ranking the 10 cities with the most crimes per capita, and checking the distribution of the number of crimes per capita among all cities in Israel.

Top 10 cities crime per capita		
City Name	Socioeconomic cluster	Crime per capita
EILAT	6	0.70159
TUBA-ZANGARIYYE	3	0.53216
ROSH PINNA	7	0.46909
TEL AVIV - YAFO	8	0.46822
BE'ER SHEVA	5	0.46677
JISR AZ-ZARQA	2	0.45256
AKKO	4	0.42884
DIMONA	5	0.39648
NOF HAGALIL	5	0.37865
MIZPE RAMON	3	0.37671

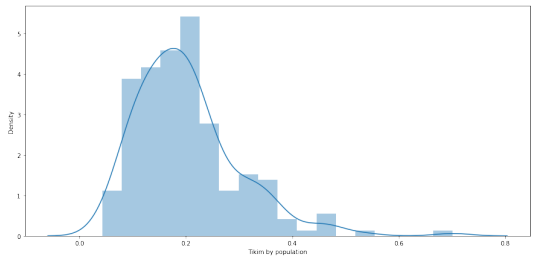

Figure 1: Crime per capita distribution

After that, we looked at the number of crimes in each district in the country from 2017 until today with an additional focus on the COVID-19 pandemic (especially in the period between the first lockdown and the end of the last lockdown in Israel).

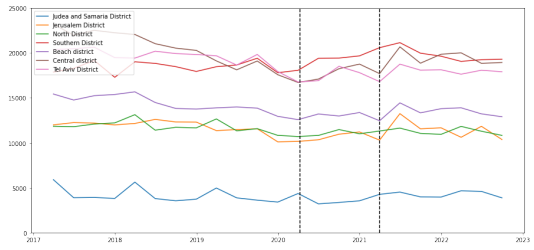

Figure 2: Number of crimes per capita by districts

Finally, we made an analysis of each district according to the amount of crime of each type, to see what the types of crime are in the district and what the trends are over the years, in the country

as a whole and in that certian district. Here too we combine an observation at the COVID-19 pandemic period so that it does not affect the general analysis.

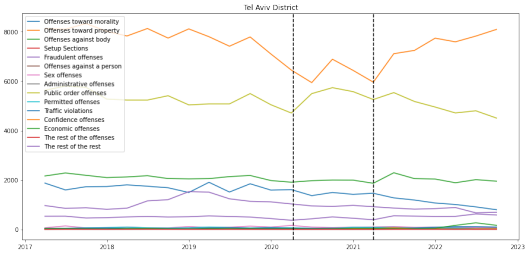

Figure 3: Tel Aviv district crimes by types

At the stage of defining the "crime score" - to learn something about the level of crime in each city, we defined several crime scores to estimate not only the amount of crime but also to normalize the amount of the population and the different types of each offense to understand better. We defined a score for each type of crime, with a score of 1 for the least severe offense (for example, public order offenses), and a score of 5 for the most severe offense (for example, physical offenses and sexual offenses). The general crime score of each city is based on two formulas that we established:

P = the population number of the city

n_i = number of specific crime group cases

S_i = the score of the specific crime group

$$CrimeScore_1 = \frac{1}{P}\sum_i n_i \cdot S_i$$

N = overall number of cases in the city

$$CrimeScore_2 = \frac{1}{N}\sum_i n_i \cdot S_i$$

After establishing these two formulas for scoring the crime rate in each city, we averaged these two formulas.

$$CrimeScore = \frac{CS_1 + CS_2}{2} = \frac{1}{2} \cdot (\frac{1}{P}\sum_i n_i \cdot S_i + \frac{1}{N}\sum_i n_i \cdot S_i)$$

Top 10 cities crime score		
City Name	Socioeconomic cluster	Crime score
EILAT	6	2.37827
JISR AZ-ZARQA	2	2.17905
TEL AVIV - YAFO	8	1.98876
AKKO	4	1.95734
TIRE	5	1.90347
BE'ER SHEVA	5	1.89396
ROSH PINNA	7	1.88766
TUBA-ZANGARIYYE	3	1.87608
TUR'AN	3	1.86332
NOF HAGALIL	5	1.85618

After defining each crime score, we checked the correlation between the crime score and the socioeconomic cluster of that city (Figure 5).

In the deep analysis phase - we are trying to predict the change in the amount of crime in each city towards the year 2023, for the benefit of establishing the recommendations for the law enforcement agencies and where and how to invest the resources and efforts. We do this with the help of a Linear-Regression model by learning the amounts of crime in each city over the years.

In the beginning, we adapted the database that will present the number of crimes in each city in Israel in each quarter from 2017 to 2022. After that, we let the model train on the changes regarding the last quarter of 2022, and finally, predict the changes in 2023. We measure the changes in percentages, by the difference between the forecast and the crime rate at the end of 2022, divided by the last crime rate in that city:

Δ_{crime} = the predicted change in the crime rate among certain city. if $\Delta_{crime} > 0$ the crime in the city is predicted to rise during 2023.

$Pred$ = the predicted crime rate in certain city from the Linear Regression model.

CR_{2022} = the last crime rate record from the Israeli Police database.

$$\Delta_{crime} = \frac{Pred - CR_{2022}}{CR_{2022}}$$

4 EVALUATION

We evaluated the crime score by looking at its distribution among the cities in Israel. The distribution is approximately normal, so it gives us confidence in the formula we established.

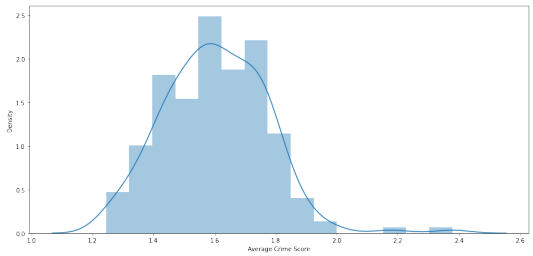

Figure 4: Crime score distribution

We evaluated the Linear-Regression model by giving a forecast for the last quarter of the year 2022 and checking the forecast on PySpark's Regression-Evaluator. The R-Squared matrix score was calculated to a score of 0.986, so the forecast model can be highly relied upon.

5 CONCLUSIONS

A general view of the amount of crime in Israel by district - tells us it's without a big change in an abnormal way, therefore considering the increase in the population (10% in 5 years) it can be said that the amount of crime in Israel has decreased in general. (Figure 2)

Crimes during COVID-19 pandemic - we see a slight decrease in most districts in crime during the first lockdown, an increase during the pandemic, a decrease towards the last lockdown, and a significant increase after the end of the pandemic. (Figure 2)

We will use the rest of this section, to recommend for the Israeli law enforcement authorities of how to make concentration of effort

in matters of resources. First, some significant trends we observed by districts:

South District - there is an increase in the number of property crimes, and a decrease in public order crimes.

Judea and Samaria district - increase in property crimes.

Jerusalem district - decrease in public order offenses, a slight increase in physical offenses.

North District - decrease in property crimes, but an increase in public order.

Coastal district - decrease in property crimes.

Central district - increase in property crimes.

Tel-Aviv district - increase in property crimes. (Figure 3)

Crime score: Surprisingly, we saw that the correlation between the crime score of a city and the socioeconomic cluster is not as high as we expected, but it exists. (Figure 5) We established a score that is distributed in a normal way over all the cities in Israel. We examined all of our three formulas - in both we reached similar results.

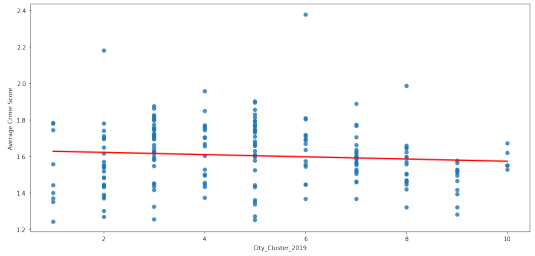

Figure 5: Correlation between socioeconomic cluster and crime score

After performing the Linear Regression model, we saw that the cities where there will be an increase in crime towards 2023, are not necessarily the cities of low socioeconomic status:

Top 10 cities crime score		
City Name	Socioeconomic cluster	Δ_{crime}
YESUD HAMA'ALA	7	285%
IMMANU'EL	2	164%
ELQANA	8	132%
MIGDAL	6	93%
MAZKERET BATYA	8	79%
ROSH PINNA	7	66%
LEHAVIM	10	58%
QIRYAT YE'ARIM	2	47%
HURFEISH	5	43%
OR YEHUDA	5	41%

Another several aspects about crime data in cities that are important to note:

Eilat has very high crime rate and crime score.

In the crime score that indicates the average crime and the score concerning the population (Formula 1) - there are only 2 arab cities out of the top 10. In the crime score that indicates the average risk according to crime (Formula 2) there are 7 Arab cities out of the top 10 - meaning that the arab cities in Israel don't necessarily have more crime, but the crimes there are more dangerous. The average score (Formula 3) is more balanced.

Our analysis shows that there are a number of cities in Israel where the forecast is for a decrease in crime cases during 2023, it can be concluded that certain strategies in the fight against crime in these communities during 2021-2022 were effective.

Top 10 cities crime score		
City Name	Socioeconomic cluster	Δ_{crime}
SAJUR	4	-127%
HAR ADAR	9	-125%
ELYAKHIN	7	-88%
MI'ELYA	7	-74%
BEIT JANN	4	-74%
KEFAR WERADIM	9	-66%
ALFE MENASHE	8	-65%
PARDESIYYA	9	-60%
TEL MOND	9	-59%
QAZRIN	5	-58%

Finally, we noticed that most of the high crime scores are found in the main socioeconomic clusters (3-8).

REFERENCES

[1] The Movement for Free of Information. Letter to the ministry of homeland security.
[2] Israeli Police. Crime records data.
[3] Hsiu-chin Chen, Wingyan Chung, Yi Qin, Michael Chau, Jennifer Xu, G. Wang, Rong Zheng, and Homa Atabakhsh. Crime data mining: An overview and case studies. 12 2003.
[4] Matthew Friedman James Cullen and Ames C. Grawert. Crime trends: 1990 - 2016. *Brennan Center for Justice*, 04 2017.
[5] Dr Yakterina Yazamski. Multi-year decrease in property crimes in the last decade - analysis of statistical trends. *Haykar Bamechkar 2017*, page 61, 2017.
[6] Myung-Sun Baek, Yong Tae Lee, Kwangho Jang, Heejo Lee, and Wonjoo Park. Design and performance evaluation of crime type and crime risk score estimation technique for fast and efficient response of severe crimes. In *2020 International Conference on Information and Communication Technology Convergence (ICTC)*, pages 1826–1828, 2020.
[7] Lalitha Saroja Thota, Mohrah Alalyan, AL-Otaibi Awatif Khalid, Fabiha Fathima, Suresh Babu Changalasetty, and Mohammad Shiblee. Cluster based zoning of crime info. In *2017 2nd International Conference on Anti-Cyber Crimes (ICACC)*, pages 87–92, 2017.
[8] Israeli Police. 2021 statistical review.
[9] Ben Stickle and Marcus Felson. Crime rates in a pandemic: the largest criminological experiment in history. In *American Journal of Criminal Justice*, volume 45, pages 525–536, 2020.