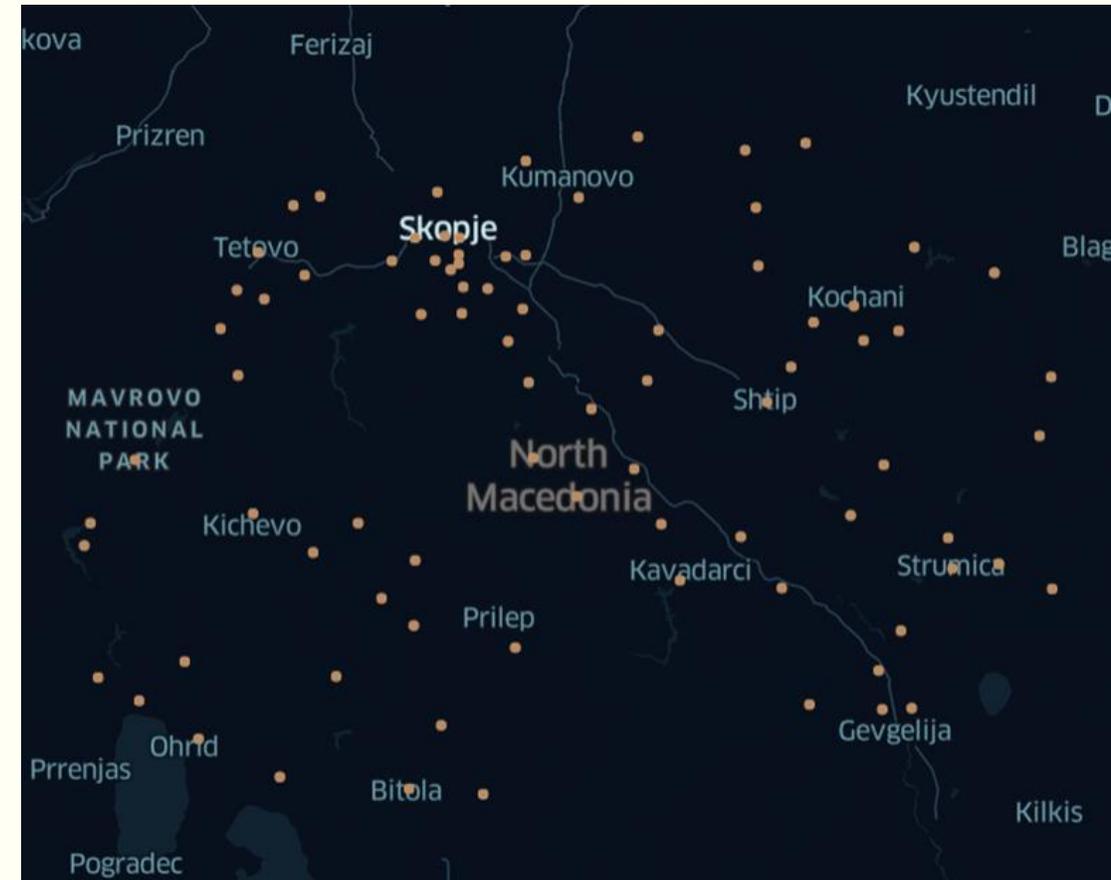


# NATIONAL CRIME ANALYSES AND FORECASTING: CASE STUDY OF N. MACEDONIA

Sonja Filiposka, Anastas Mishev and Aleksandra Dedinec,  
Faculty of Computer Sciences and Engineering, Skopje, N.  
Macedonia



# Content of the presentation

---

1

Introduction

2

Database  
description

Crime map of  
N. Macedonia

Analyses of  
historical data

3

Crime  
forecasting

4

Conclusion

Future work

# Introduction

---

- Importance of crime forecasting
- Methods used in crime forecasting
  - Statistical models
  - Machine learning based models
    - Regression problem
      - Time Series forecasting
    - Classification problem
- Scope of the paper: crime analyses and forecasting in North Macedonia

# Database

Crime records within 9 years period (2011-2020), based on data presented in Crime Map of Macedonia (<http://crimemap.finki.ukim.mk/home/en>)

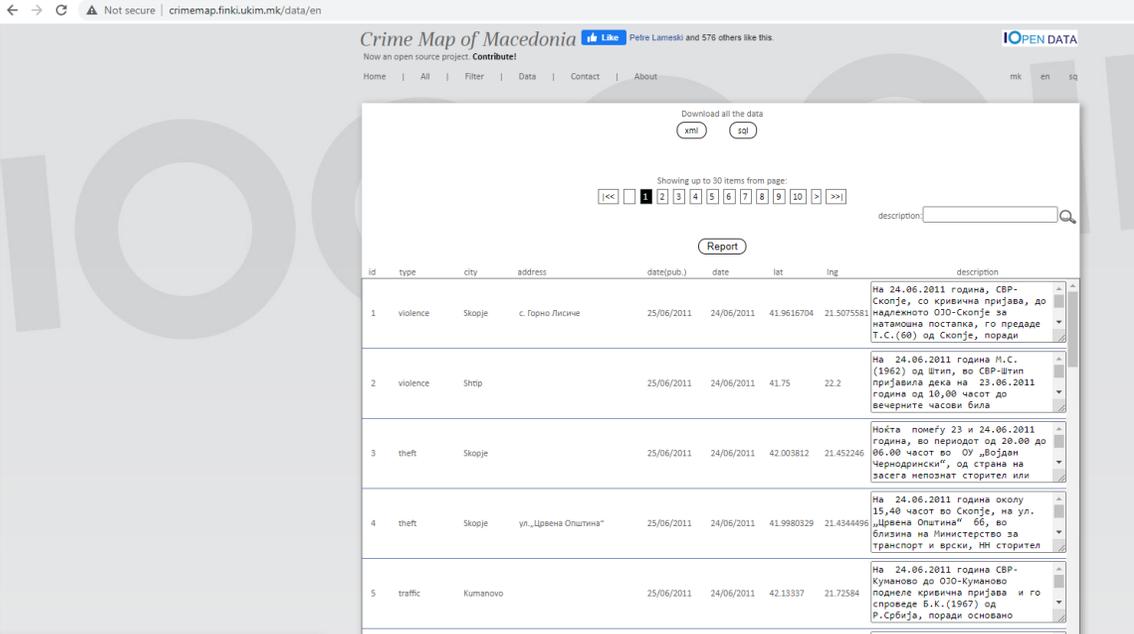
The system reads daily publications from the official website of the Ministry of Interior

The system extracts information on:

- date,
- the exact location (latitude and longitude),
- city,
- crime category and
- description

Additional data verification and cleaning process was conducted

Information on municipality was added based on the geographical location



The screenshot shows the 'Crime Map of Macedonia' website interface. At the top, there is a navigation bar with 'Home', 'All', 'Filter', 'Data', 'Contact', and 'About'. Below this, there are options to 'Download all the data' in 'xml' and 'sql' formats. A search bar is present with the label 'Showing up to 30 items from page:'. A 'Report' button is also visible. The main content is a table with the following columns: id, type, city, address, date(pub.), date, lat, lng, and description. The table contains five rows of crime records.

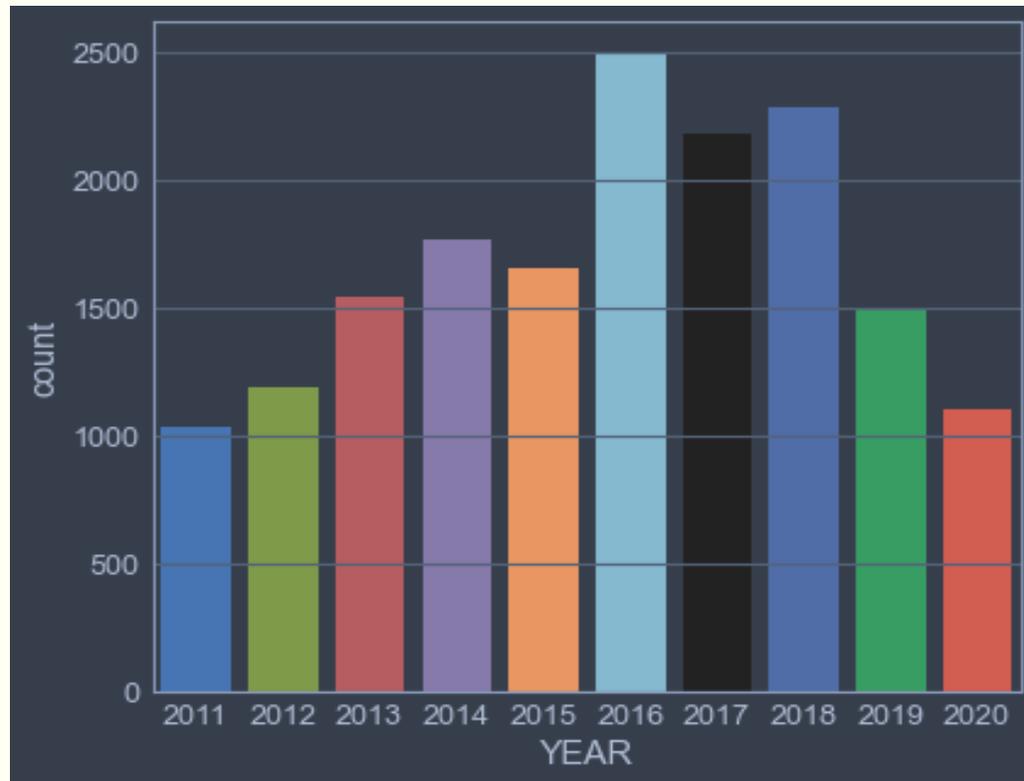
id	type	city	address	date(pub.)	date	lat	lng	description
1	violence	Skopje	с. Горно Лисине	25/06/2011	24/06/2011	41.9616704	21.5075581	На 24.06.2011 година, СВР-Скопје, со кривична пријава, до надлежното ОЈО-Скопје за натамошна постапка, го предаде Т.С. (60) од Скопје, поради
2	violence	Штип		25/06/2011	24/06/2011	41.75	22.2	На 24.06.2011 година М.С. (1962) од Штип, во СВР-Штип пријавила дека на 23.06.2011 година од 18,00 часот до вечерните часови била
3	theft	Skopje		25/06/2011	24/06/2011	42.003812	21.452246	Ноќта помеѓу 23 и 24.06.2011 година, во периодот од 20.00 до 06.00 часот во ОУ „Војдан Чернотарски“, од страна на засега непознат сторител или
4	theft	Skopje	ул. Црвена Општина	25/06/2011	24/06/2011	41.9980329	21.4344496	На 24.06.2011 година околу 15,40 часот во Скопје, на ул. „Црвена Општина“ 06, во близина на Министерство за транспорт и врски, НН сторител
5	traffic	Kumanovo		25/06/2011	24/06/2011	42.13337	21.72584	На 24.06.2011 година СВР-Куманово до ОЈО-Куманово поднеле кривична пријава и го спроведе Б.К. (1967) од Р.Србија, поради основано

# Analysis of historical data

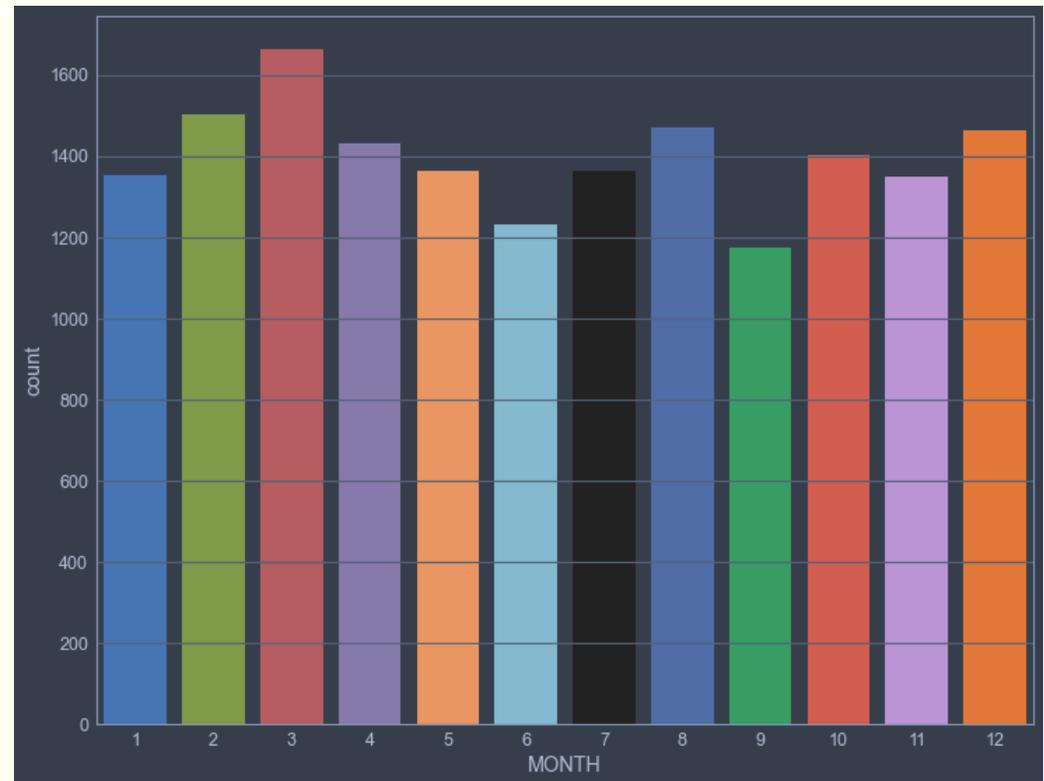
---

---

## Distribution of crimes per year

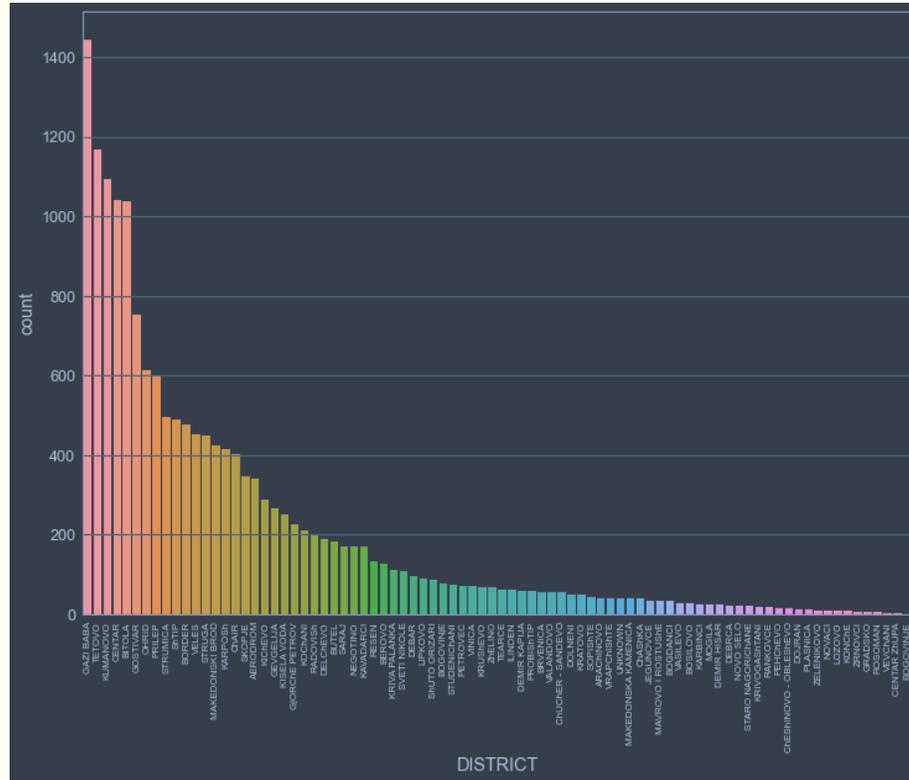


## Distribution of crimes per month

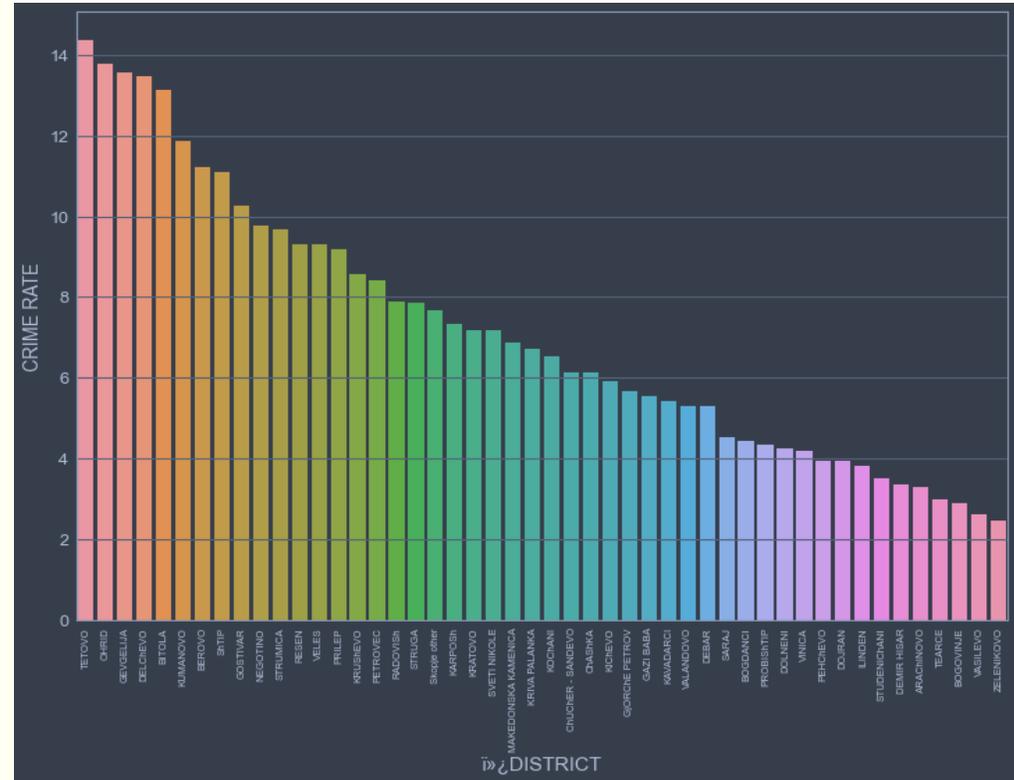


# Analysis of historical data

## Distribution of crimes per municipality

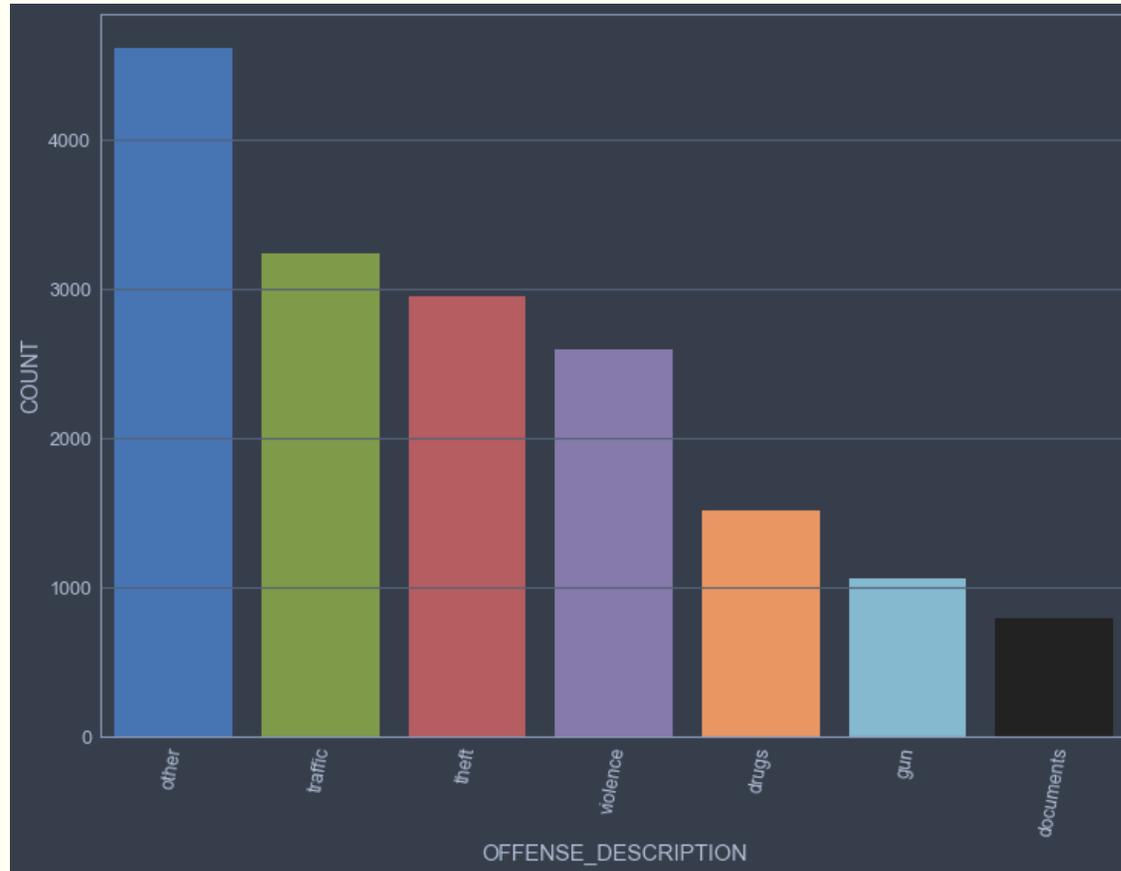


## Number of crimes/population per municipality



# Analysis of historical data

---

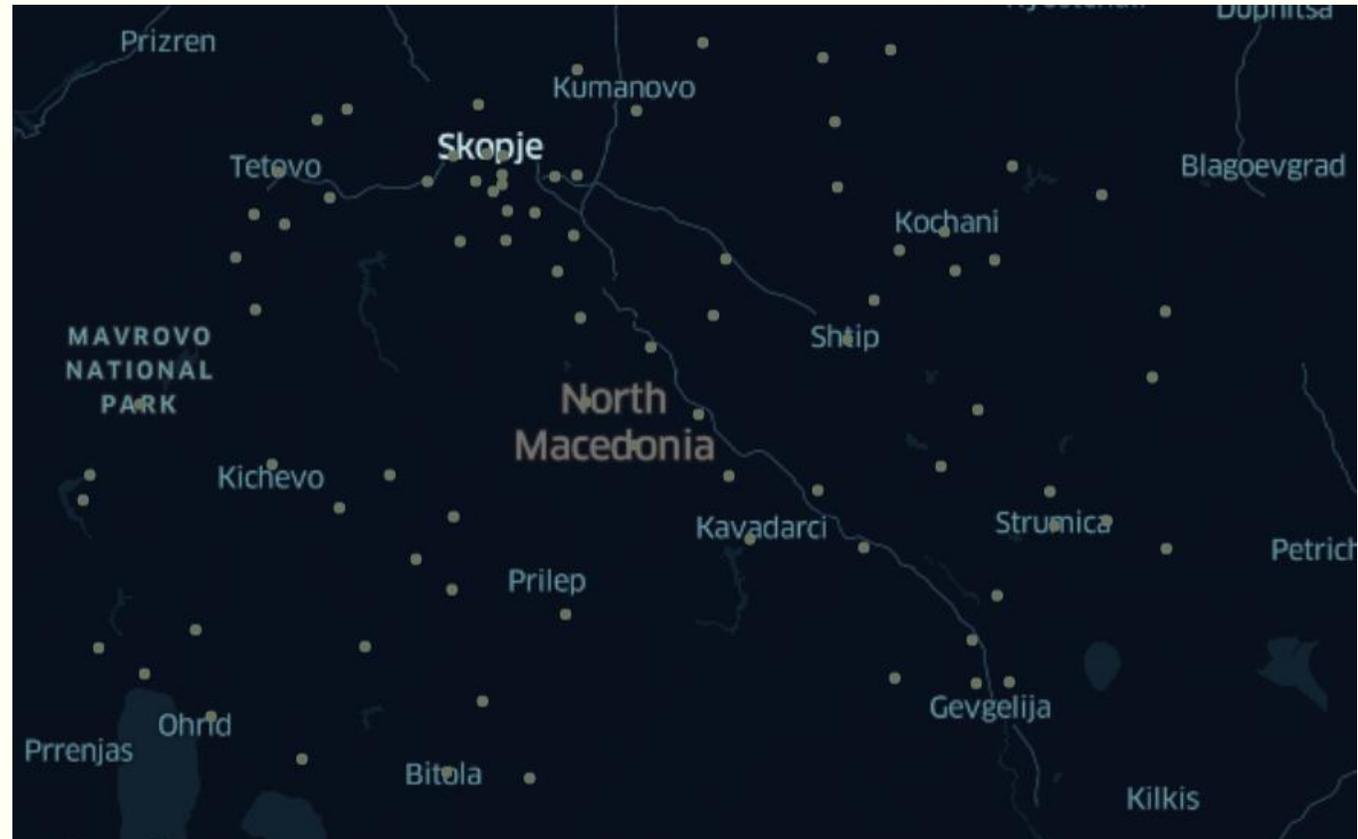


- Distribution of the crime per category
- There are seven categories in the analyzed dataset:
  - Traffic
  - Theft
  - Violence
  - Drugs
  - Gun
  - Documents
  - Other

# Geographical location of cluster centers

---

- Cluster centers of reported crimes
  - Using K-means clustering
  - 83 clusters obtained (based on number of municipalities)



# Feature calculation process

---

- Crime categories are grouped in the following types:
  - Violent
  - Property
  - Other
- Besides longitude and latitude, for each cluster center the following information is calculated:
  - Number of crimes per type (violent, property, other) in the last 120 days
  - Number of crimes per type (violent, property, other) in the last 30 days
  - Number of crimes per type (violent, property, other) in the last 7 days
  - Number of crimes per type (violent, property, other) in the last day
- Aggregated historical features created for each cluster center and day

# Forecasting model

---

- Goal: predict whether a violent crime will happen on a certain day and certain location based on the calculated historical features
- Classification problem
  - model based on gradient boosted decision trees
- Train and test sets created (70%/30%)
- Selection of threshold
  - Percent of regions labeled as positive
  - Assumption: the police can patrol only 30% of the regions (the calculated threshold in that case is calculated to be 0.38)

# Results

---

- Confusion matrix of the training data

	Negative pred.	Positive pred.
Negatives	16620	4198
Positives	18	329

- Confusion matrix of the test data

	Negative pred.	Positive pred.
Negatives	6656	2357
Positives	27	173

# Results

- Classification report for training data

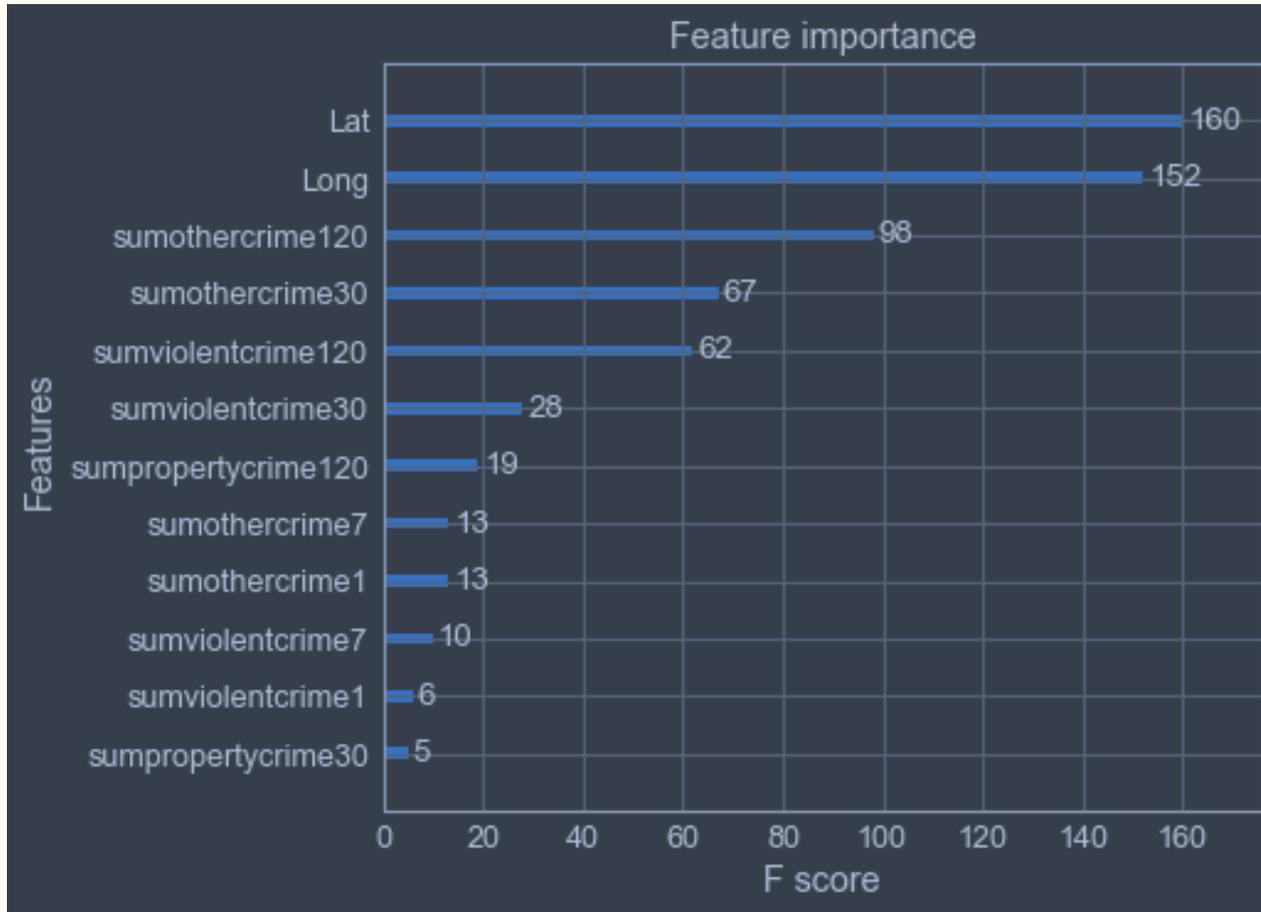
	Precision	Recall	F1-score	Support
0.0	1.00	0.80	0.89	20818
1.0	0.07	<b>0.95</b>	0.14	347
Accuracy			<b>0.80</b>	21165
Macro avg	0.54	0.87	0.51	21165
Weighted avg	0.98	0.80	0.88	21165

- Classification report for test data

	Precision	Recall	F1-score	Support
0.0	1.00	0.74	0.85	9013
1.0	0.07	<b>0.86</b>	0.13	200
Accuracy			<b>0.74</b>	9213
Macro avg	0.53	0.80	0.49	9213
Weighted avg	0.98	0.74	0.83	9213

# Results

---

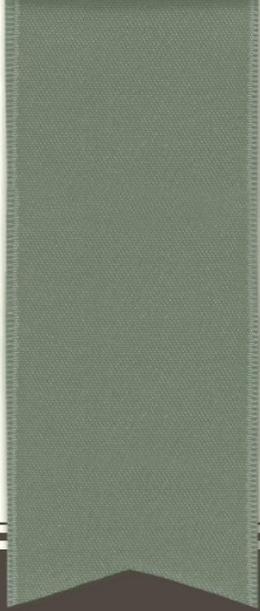


- Feature importance
  - Most important – latitude and longitude
  - Next are the statistics for other and violent crime for the period of 120 and 30 days

# Conclusion and future work

---

- Data for crime events in N. Macedonia processed and analyzed
- Features for forecasting the crimes calculated
- Gradient boosted decision trees model developed
- The results can prioritize the areas in which the probability of violent crime is the highest
- Future work:
  - Make analysis of the historical data based on GDP per municipality (or region)
  - Include other features such as weather data
  - Apply different model (including regression models) and compare the results



---

---

THANK YOU FOR YOUR ATTENTION

---

---