

# SPORT vs CRIME

---



## 2 INTRODUCTION

---

Sport can attract people and help them solve a number of problems that push them into crime:

- Developing self-regulating and problem-solving abilities as a result of developing skills needed to sport activity.
- Adventurous sport can satisfy the thirst for risk.
- Sport helps people to socialize, playing sport, a person turns into a group, also can find friends and mentors, who provide positive role models.

I wanted to show that sports venues, such as sports fields, swimming pools, sports schools, can reduce crime in the city.



### 3 DATA ACQUISITION

---

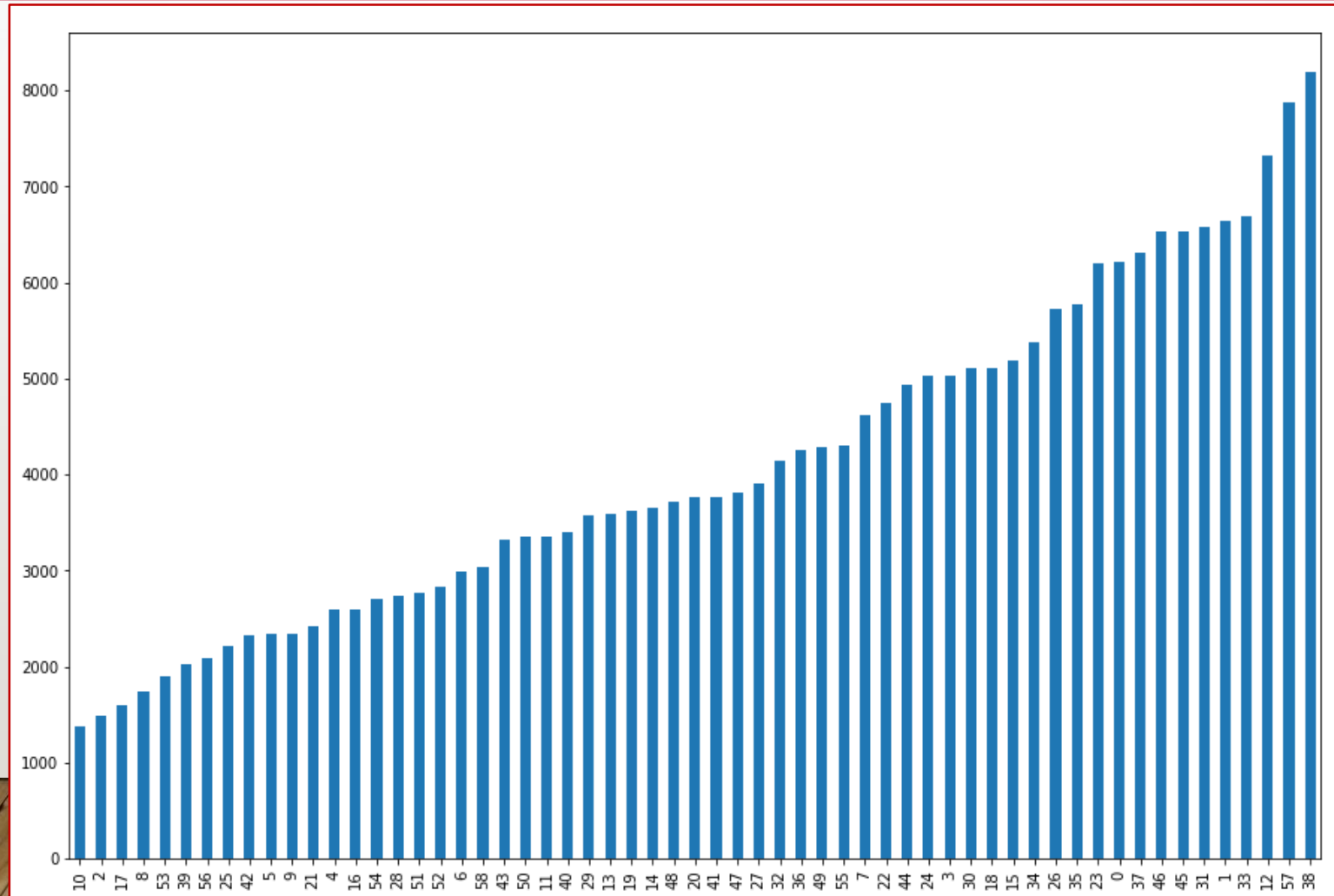
- The data about sports venues of a city from **4sq** [[foursquare.com](https://foursquare.com)]
- The data about crime rates by cities from **Wiki**  
[[en.wikipedia.org/wiki/List\\_of\\_United\\_States\\_cities\\_by\\_crime\\_rate](https://en.wikipedia.org/wiki/List_of_United_States_cities_by_crime_rate)]
- Translation the-state-name + city-name from **GeoPy** [[github.com/geopy/geopy](https://github.com/geopy/geopy)]

## 4 DATA CLEANING

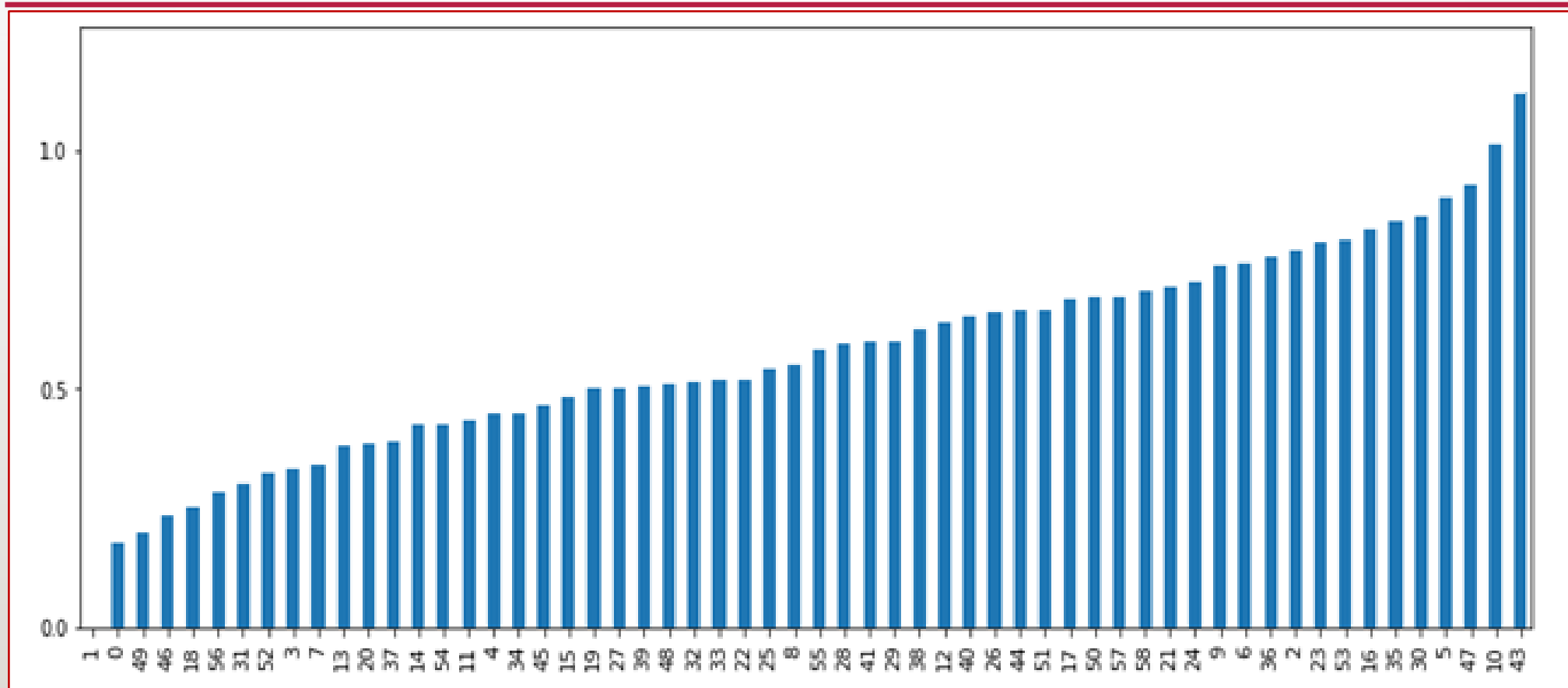
---

- Drop rows (cities) with gaps in the table of crime statistical.
- Limit the research to cities with a population of up to 500 thousand.
- Drop cities with too little data about sports venues.
- **Features:** city population and numbers of sports venues by categories per inhabitant.
- **Target:** whether the crime rate of a city is higher than the median?

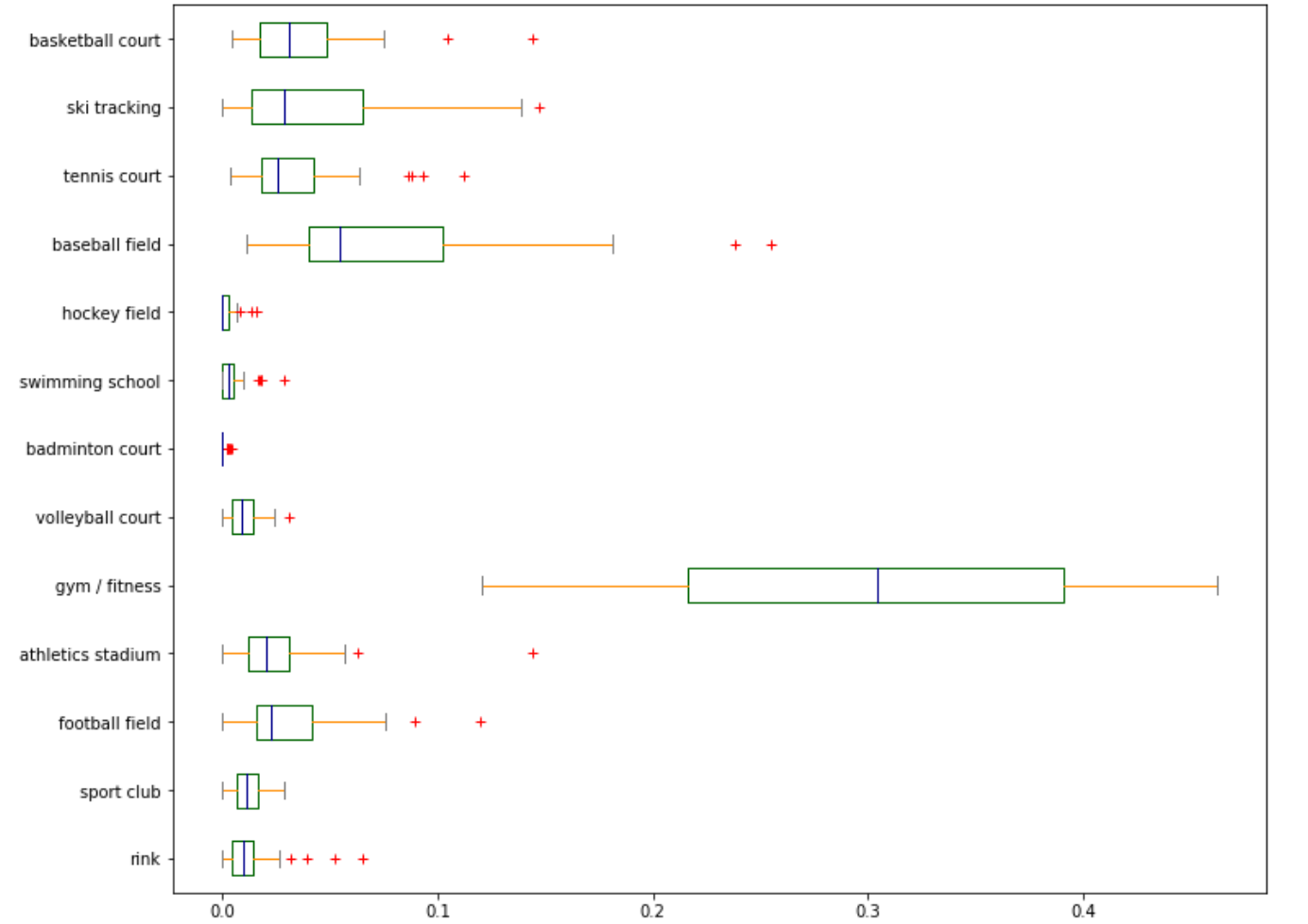
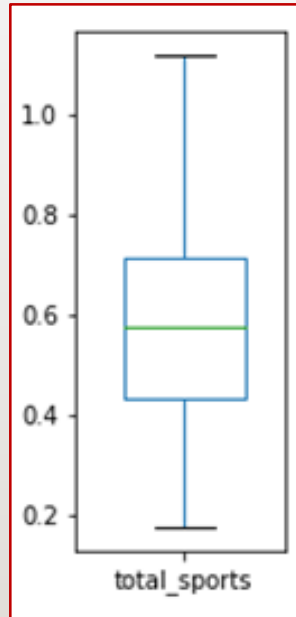
## 5 HISTOGRAM OF CRIME RATE



## 6 HISTOGRAM OF NUMBER OF SPORTS VENUES PER INHABITANT

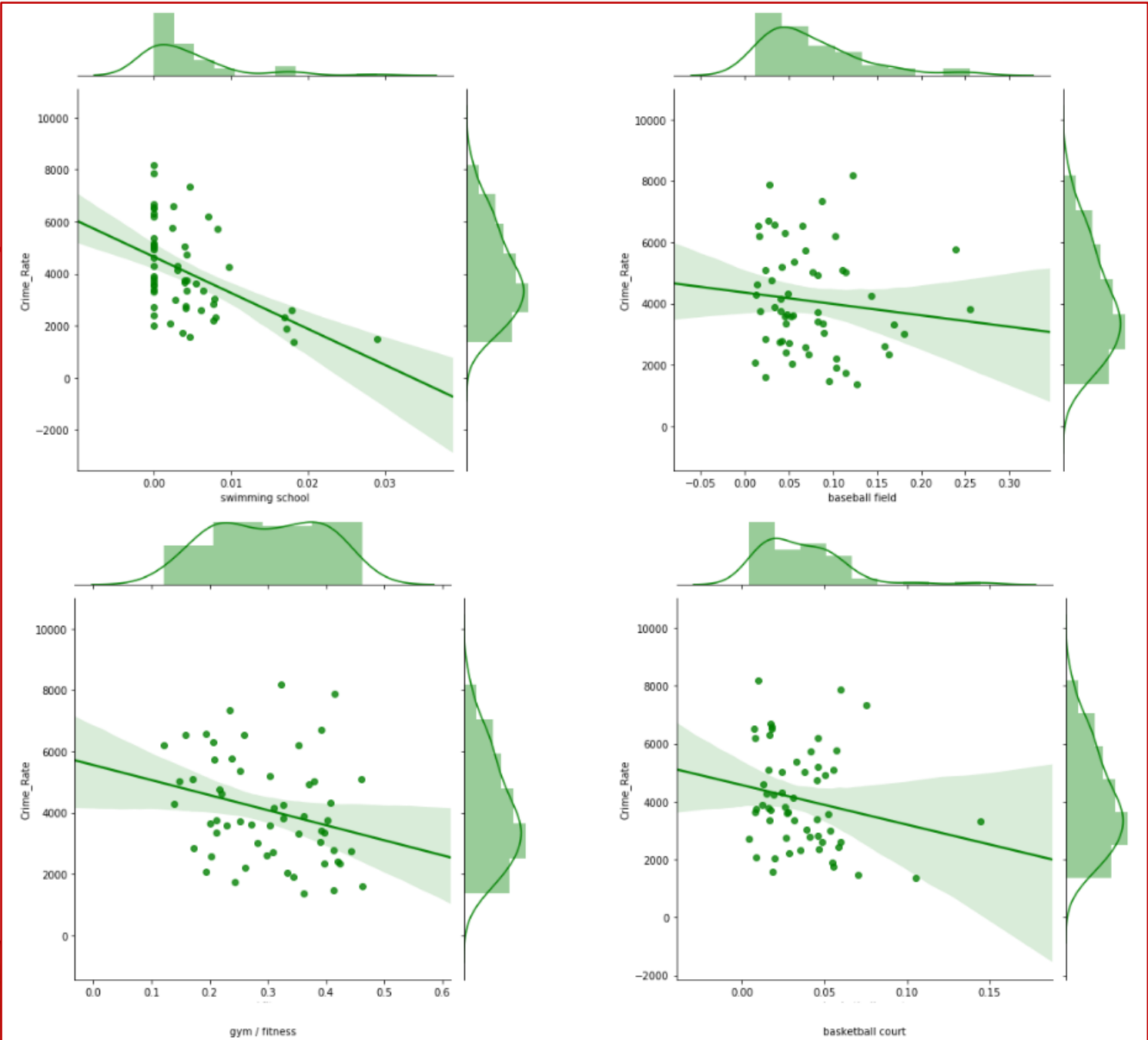
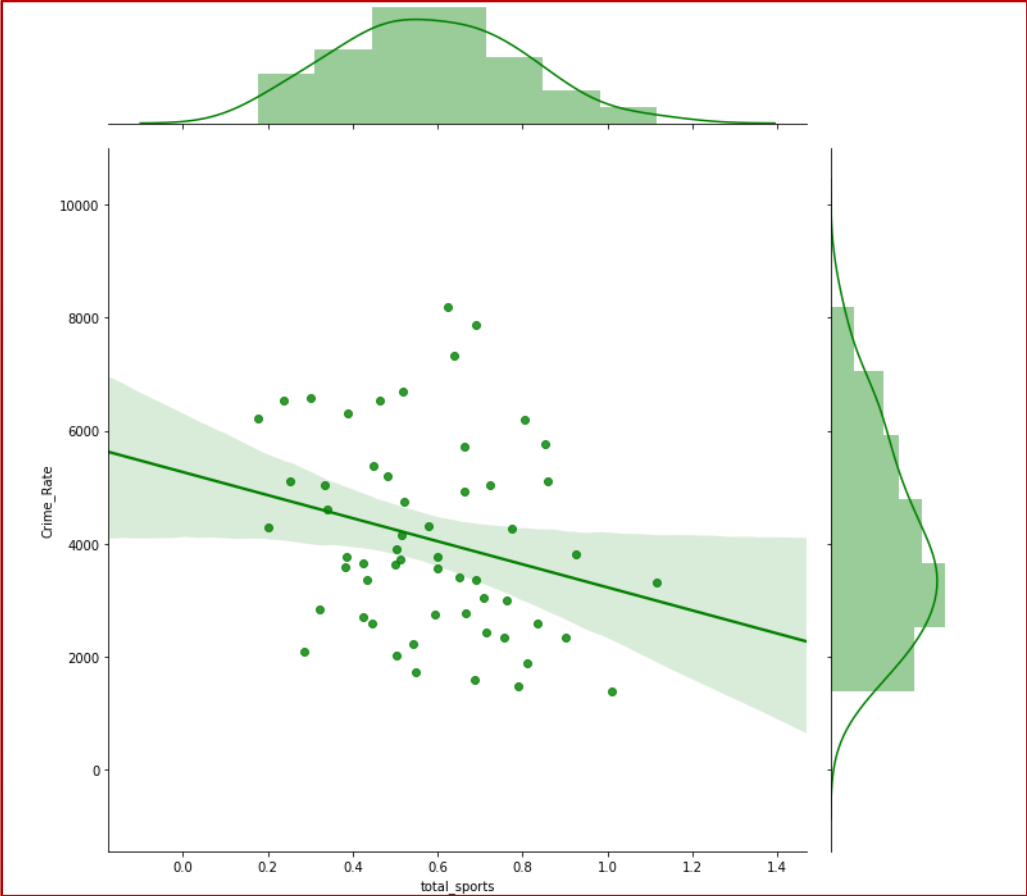


# 7 BOXES OF SPORTS VENUE STATISTIC





# RELATIONSHIPS BETWEEN CRIME RATE AND SPORTS VENUES





## 9 PERSON CORRELATIONS

Feature	Correlation with the crime rate
swimming school	-0,47
gym / fitness	-0,28
total sports	-0,26
volleyball court	-0,26
basketball court	-0,20
athletics stadium	-0,18
sport club	-0,18
football field	-0,16
tennis court	-0,14
baseball field	-0,12
ski tracking	-0,06
badminton court	0,01
hockey field	0,13
rink	0,14

# 10 CLASSIFICATION MODELING

---

I used different type of ML algorithms (implementations were provided by scikit-learn):

- **k-Nearest Neighbors (kNN)**

```
knn_model = KNeighborsClassifier(n_neighbors = 4).fit(X_train,y_train)
```

- **Support Vector Machine (SVM)**

```
svm_model = svm.SVC(kernel = 'rbf', degree = 3, probability=True).fit(X_train,y_train)
```

- **Decision Tree (DT)**

```
tree_model = DecisionTreeClassifier(max_depth = 2).fit(X_train,y_train)
```

- **Logistic Regression (LR)**

```
lr_model = LogisticRegression(solver = 'lbfgs', C = 0.2).fit(X_train,y_train)
```

- **Multilayer Perceptron (MLP)**

```
mpl_model = MLPClassifier(hidden_layer_sizes = (4, 1), activation = 'relu', random_state = 0).fit(X_train,y_train)
```



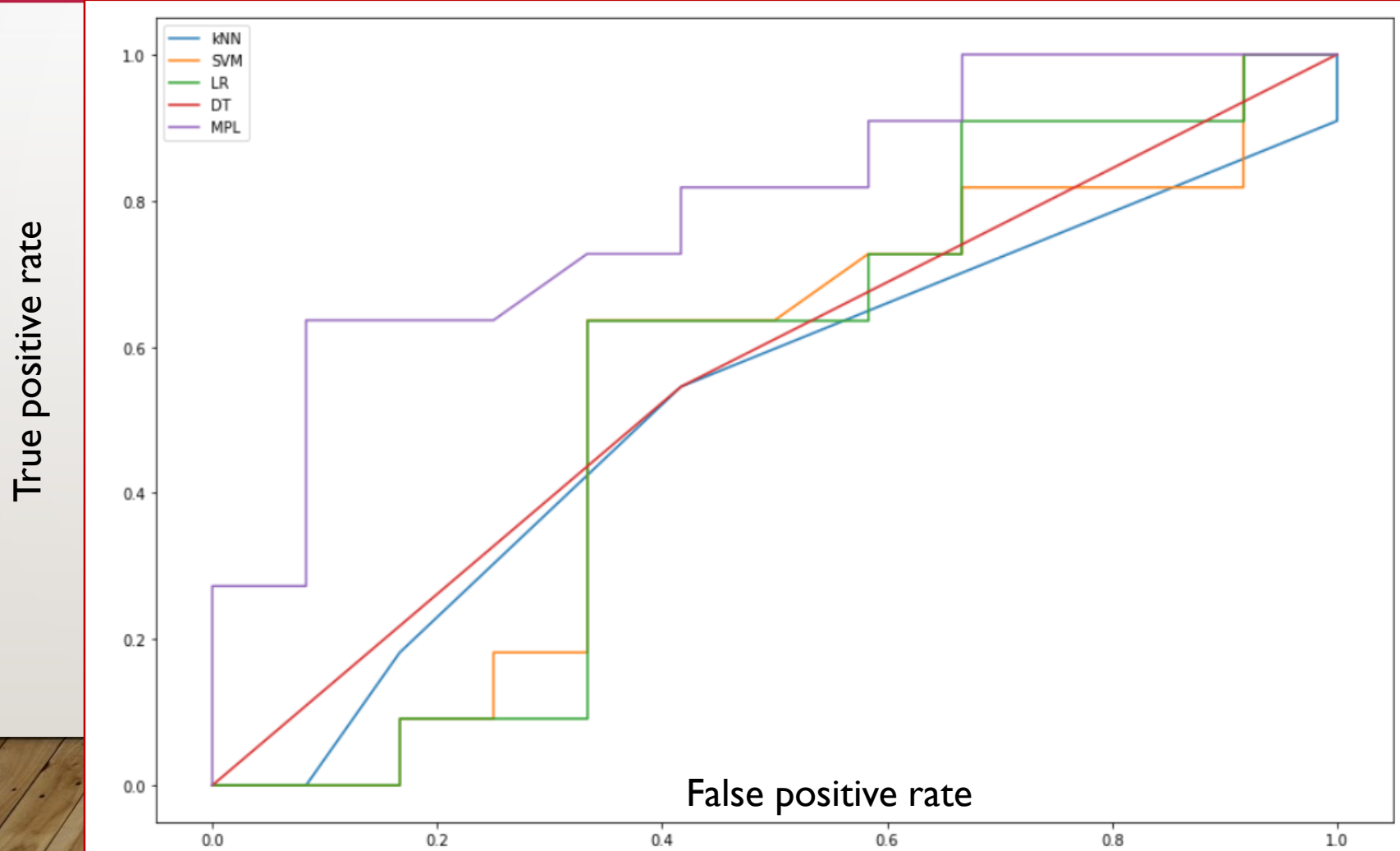
# II CLASSIFICATION RESULTS

whether the crime rate of a city is higher than the median?

---

Classifier	Accuracy	Precision	Recall	F1-score
kNN	57 %	58 %	58 %	58 %
SVM	65 %	67 %	67 %	67 %
Decision Tree	56 %	58 %	58 %	58 %
Logistic Regression	65 %	67 %	67 %	67 %
Multilayer Perceptron	73 %	71 %	83 %	77 %

## 12 ROC CURVE



## 13 CONCLUSION & FUTURE DIRECTIONS

---

Despite the fact that so far only a small exploration work has been carried out in this area of the research, I dare to formulate the following conclusion.

- Based on the results obtained, it should be recommended to invest in low-cost sports facilities for the long-term sustainable development of their city.
- I think that the initial phase of this study presented here gives showed the promise of this study. It can be continued to produce more reasonable results.
- In the future, the study can be significantly deepened by expanding the sample of output data, using not only the number of institutions but also their rating, as well as the division of crime statistics by age categories.