**Final Project**

**Object Oriented Design**

**Crime Detector**

****

**Northeastern Illinois University**

**MS in Computer Science**

**Team Members**

**Anju Shrestha**

**Latha Saradha**

**Sai Krishna**

**Contents**

[**ABSTRACT** 3](#_Toc46358028)

[**INTRODUCTION** 3](#_Toc46358029)

[**BACKGROUND** 4](#_Toc46358030)

[**Approach** 4](#_Toc46358031)

[**Prediction for the total number of crimes in the Year 2020** 4](#_Toc46358032)

[**Total arrests for top five crimes** 5](#_Toc46358033)

[**Crime rate in a district for a given year** 5](#_Toc46358034)

[**RESULTS** 6](#_Toc46358035)

[**CONCLUSION** 8](#_Toc46358036)

[**REFERENCES** 9](#_Toc46358037)

# **ABSTRACT**

For the final project, we investigated different datasets on the Chicago Data Portal. After looking at all those datasets, our team finally decided to work on the crimes dataset as our team thought we could come up with a few interesting questions using this dataset. The program follows the Object-oriented design approach and uses the techniques learned in the class.

Crime Detector is a systematic approach for identifying the patterns and trends of crimes happened in Chicago since 2001. This application can predict the total number of crimes in the oncoming year 2020 based on the crime data of the past 19 years using linear regression. A bar graph is used to display the total number of crimes for each year that makes it easier to read the data and looks clear and visually appealing. This application also finds the total arrests of the top 5 crimes types. A pie chart is used to display this information. This information is helpful in suggesting Law enforcement services to focus their attention to a particular category and reduce the crime rate in the city. The program can find the crime rate in a particular district, given the district number and year by the user. As Chicago Data Portal has a large amount of data for the crimes, analyzing those data of the crimes and crime types will be easier for the user to get information about criminal history in their neighborhood. This program will be helpful to retrieve some useful information from the crime data listed in the Chicago Data Portal.

# **INTRODUCTION**

This project is a simplified attempt to evaluate the crime statistics in Chicago, Illinois. With the help of the Chicago city Data portal, we are able to retrieve the dataset of crimes happened during the time frame 2001 to 2019. This dataset contains 22 different columns, out of which we are retrieving only the columns needed to generate the result. In this project, all the technical skills learned in the class are brought together and program is designed using object-oriented approach. Total number of crimes in the year 2020 is predicted using the data from 2001 to 2019. Top five crime which has the highest arrest rate is calculated. Total percentage of crimes in a specified district and year is calculated given the district number and year by the user.

# **BACKGROUND**

The main focus of this project is to investigate the dataset of the crimes that happened in Chicago from 2001 to 2019. The datasets are retrieved from the Chicago Data Portal (<https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2/data>).

1. First, we predicted the number of crimes in the year 2020. We are using a bar chart to display the result of the number of crimes happened in Chicago from 2001-2019. The bar chart also displays the predicted number of crimes for the year 2020. The algorithm used for prediction is Linear Regression. In our algorithm we take the Count of crimes in a year as our variable and evaluated the data from 2001 to 2019, to predict the value in 2020.
2. Second, we are finding total arrests for the top 5 crimes. With the dataset, the crimes can be divided into different categories (primary\_type in Database) such as Narcotics, Theft, Assault, etc. With the help of our application, a user can find the most significant 5 crime types where most arrests have been made.
3. Third, we are letting the user enter their district code, year and display the percentage of the crime in their district. This would help in users to determine whether a particular district is well fit for living or for their future development plans like buying a house.

### **Approach**

IntelliJ with Java 11 and JavaFX is used to develop and test the application. To access the crime data, Socrata Open Data API is used which provides programmatic access to these datasets. OkHttp3 API is used to send HTTP requests to Chicago City Portal. Received data is converted into a Reader (Java IO) and parsed using Json Parser (Json API).

Three functions are implemented in this program

* Prediction for the total number of crimes in the Year 2020.
* Total Arrests for top five crime types.
* Crime rate in a district for a given year

### **Prediction for the total number of crimes in the Year 2020**

Initially, to find the minimum and maximum year available, query is sent to the database using OK HTTP. After obtaining the minimum and maximum available years from the dataset, queries are sent for each year in the range of minimum and maximum to obtain the total number of crimes for that year. All the results are stored in a tree map with key as year and value as the total number of crimes in the year for further processing.

Linear Regression is used to predict the total number of crimes in Year 2020. This algorithm takes two double arrays which contain independent variables (x) and dependent variables(y). The above generated tree map is used to load values into these arrays. The algorithm calculates the best fit line and slope-intercept form (y=mx+b) which can be used for prediction. By supplying the value of x (2020 in this case) predicts the total number of crimes in 2020. This value is stored in a variable for further processing.

Separate non-JavaFX thread is utilized to fetch the necessary data for bar chart. While this thread is working, a progress indicator is displayed to the user. Once, the thread completes its work, progress indicator is closed and bar chart is displayed to the user.

### **Total arrests for top five crimes**

Initially, using Ok Http, a query is fired on the dataset to retrieve the total number of arrests for each crime type. All the obtained results are stored in a Linked HashMap. This Linked HashMap is further filtered using streams to sort the values in descending order. Further, streams are used to limit the crime types to top five and calculate the total number of arrests in these five crime types.

Separate non-JavaFX thread is utilized to fetch data from the dataset. While this thread is working in the background, a progress indicator is displayed to the user. Once the thread finishes the work, progress indicator is closed and pie chart is displayed to the user.

### **Crime rate in a district for a given year**

First, “Find crime percentage in a district and year” button is created using JavaFx. Once the button is clicked, a new window is opened with two text-fields where the user can enter district number and year. After user inputs district and year, query is sent to the database to get the total number of crimes for the user specified district and year. Additionally, another query is fired to get the total number of crimes in user specified year. Obtained values are stored in variables to calculate crime percentage in a district. Calculated percentage is displayed to the user in a dialog box. When user enters invalid inputs Alert box is displayed with a warning.

# **RESULTS**

When the application is run successfully, three buttons are displayed in a window. (as in Fig-1)

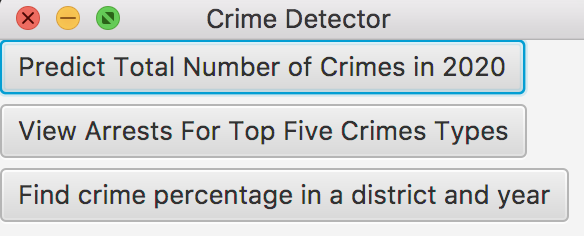


Fig-1

The first button displays the total number of crimes in Chicago from 2001-2019 and predicts total number of crimes for year 2020. (as in Fig-2)

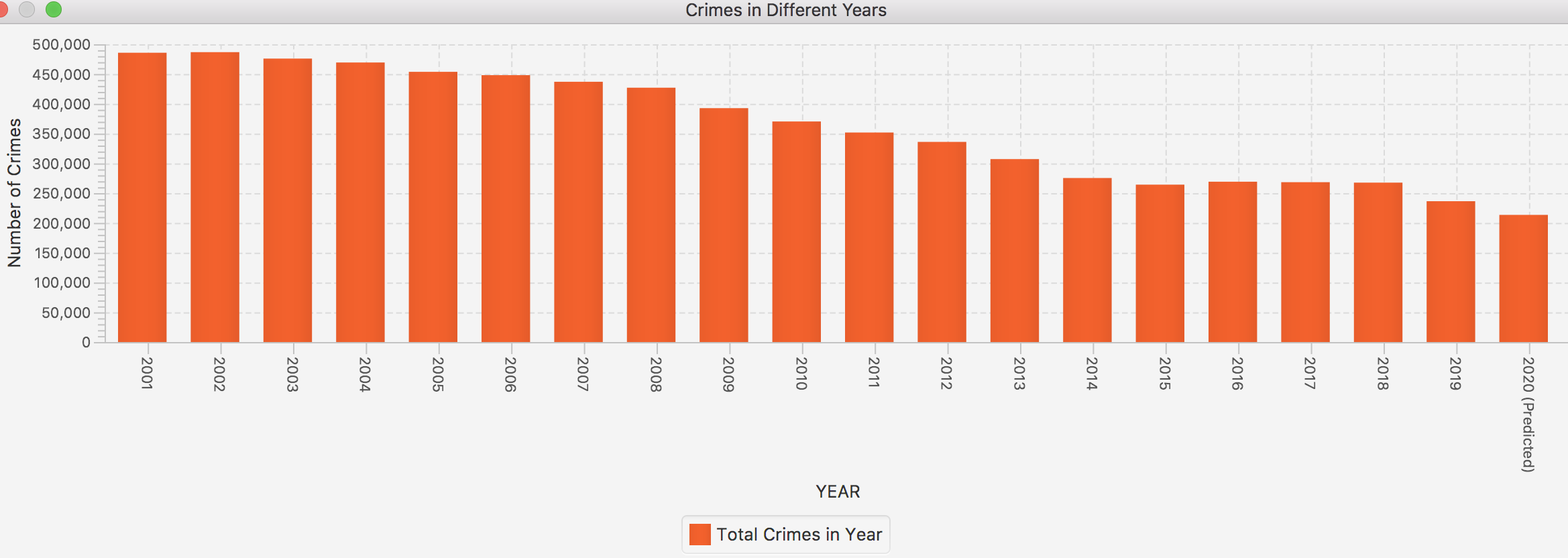


Fig-2

The second button is used to find the arrest rate of the top five crimes that happened in Chicago. Total arrested crimes are 1436036, where most of the people got arrested who committed a crime of type Narcotics. The total number of arrests for narcotics is 721993. The total number of arrests for a battery is 291663. The total number of arrests for theft is 175944. The total number of arrests for criminal trespass is 145647 and the total number of arrests for assault is 100684. (as in Fig-3)

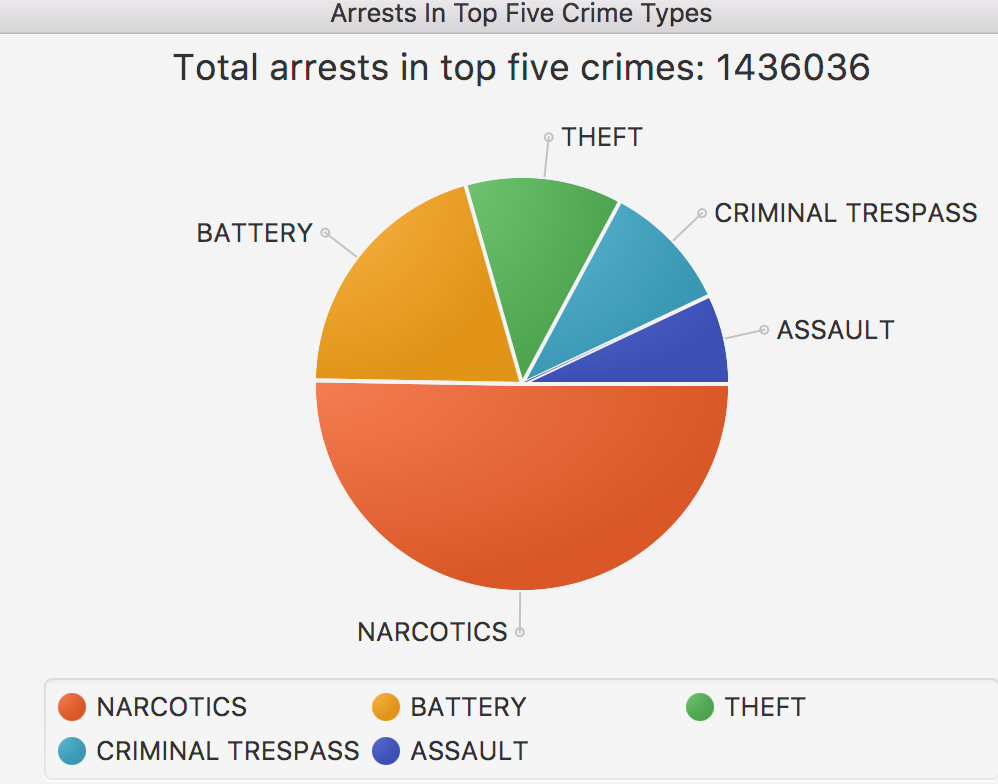


Fig-3

The third button is used to find the percentage of the crime for a particular district and year. When the “Find crime percentage in a district and year” button is clicked, it goes to another screen where the user can enter the district number and year (as shown in Fig -4). Percentage of the crime in that particular district in that year is displayed. For example, if the user entered District 017 and the year 2015, then the percentage of the crime on district 017 in the year 2015 is 2.92%(Fig -5).

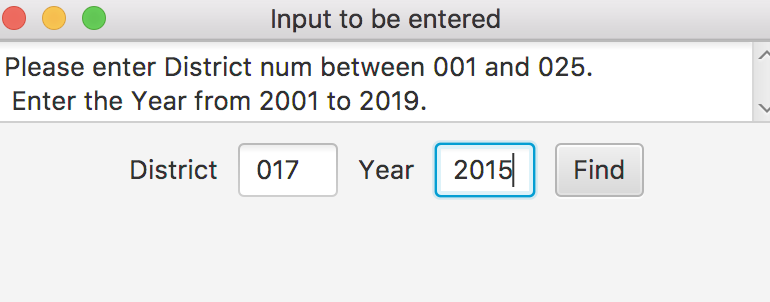


Fig-4

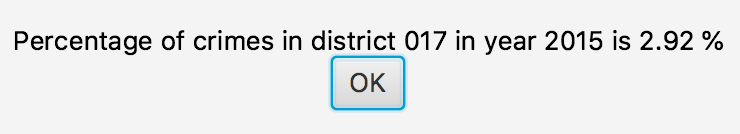


Fig-5

For invalid inputs, the warning message will be shown. (as shown in Fig-6)

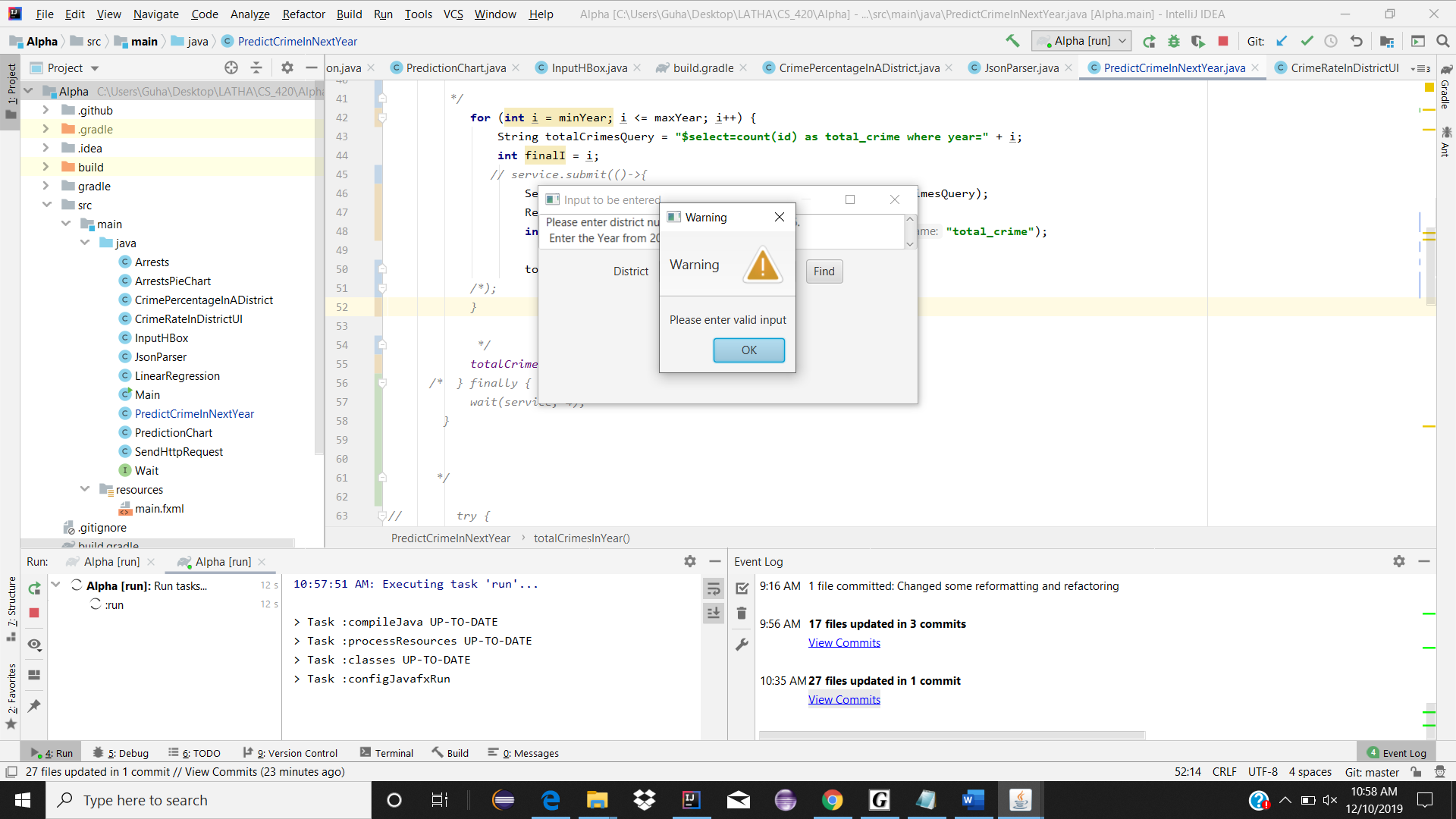


Fig-6

# **CONCLUSION**

It was a great experience for all of us to work on investigating Crime data from the Chicago Data Portal. The crime data from the portal can be broken and use those data individually. Out of 22 columns, we narrowed our search to only the columns needed to find a result. From the data, it can be inferred that the number of crimes is decreasing year by year. The top five crimes in Chicago are narcotics, battery, theft, criminal trespass, and assault. The most arrested rate out of these crimes is Narcotics. With the application, the rate of crimes on the particular district in Chicago in a particular Year can also be identified. It will be useful for people to determine whether their district is more prone to crimes and it will be helpful to find a safer district to live within Chicago. Working on this project and using the Object-Oriented Design approach to develop the program was interesting. During the design phase of the project, the team had to think the practical aspects and benefits the application can achieve to different classes of people like commoners, Law Enforcement services etc. The team came up with the questions accordingly and implementing those to find the answers was very challenging and inspiring to learn more about the techniques in Java.

# **REFERENCES**

1. **Robert Sedgewick, Kevin Wayne, Linear Regression,**

<https://algs4.cs.princeton.edu/code/edu/princeton/cs/algs4/LinearRegression.java.html> (credits)

1. **SODA Developers, Crimes - 2001 to Present,** <https://dev.socrata.com/foundry/data.cityofchicago.org/ijzp-q8t2>
2. **SODA Developers, API Endpoints,**

<https://dev.socrata.com/docs/endpoints.html>

1. **Chicago Data Portal,**

<https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2/data>

1. **Mkyong, GSon - Read and Write Json as a Stream, August 11th, 2011,**

<https://www.mkyong.com/java/gson-streaming-to-read-and-write-json/>

1. **OkHttp,**

<https://square.github.io/okhttp/>

1. **GeeksforGeeks, JavaFX Archives,**

<https://www.geeksforgeeks.org/tag/javafx/>