Methods of Advanced Data Engineering

A Data-Driven Investigation of Crime and Arrest Disparities

Author: Ramisa Bhuiyan Raka (23172960)

Introduction

In urban areas, disparities between crime reports and arrests can indicate potential issues in law enforcement practices, resource allocation, or systemic biases. This report investigates which neighborhoods or regions have the highest disparities between crime reports and arrests in a specific city. By analyzing these disparities, we aim to highlight areas that may require additional attention or resources from law enforcement agencies.

Used Data

Data Description

The data used for this analysis comes from two main sources:

- **Crime Data:**
  - **URL:** Crime dataset (2020 - Present)
  - **Content:** This dataset contains records of reported crimes in Los Angeles, including descriptions, locations, dates, and areas.
  - **Data Structure:** Columns for Crm Cd Desc, Date Rptd, Vict Sex, LOCATION, AREA NAME, DR_NO
  - **License Information:** Open licensed CC0 1.0 Universal
    - **Link:** License Details (Crime Data)

- **Arrest Data:**
  - **URL:** Arrest dataset (2020 – Present)
  - **Content:** This dataset includes records of arrests made in Los Angeles, providing details like arrest types, charges, locations, and dates.
  - **Data Structure:** Columns for Charge, Arrest Date, Arrest Type Code, Charge Description, Location, Area Name, Report ID
  - **License Information:** Open licensed CC0 1.0 Universal
    - **Link:** License Details (Arrest Data)

Both datasets were processed and stored in SQLite databases (crime_data.db and arrest_data.db) using a data pipeline that ensures data cleanliness and consistency.

Analysis

Methodology

The analysis involves the following steps:

1. **Data Aggregation:** Crime and arrest data were grouped by neighborhood/region to calculate the total number of reports and arrests.
2. **Disparity Calculation:** The disparity was calculated as the difference between the number of crime reports and arrests for each neighborhood.
3. **Visualization:** Various visualizations, including bar charts, pie charts, histograms, line charts, heatmaps, and area charts, were used to present the data and highlight key patterns.

Output

Out[4]:

	Crm Cd Desc	Date Rptd	Vict Sex	LOCATION	AREA NAME	DR_NO
0	VEHICLE - STOLEN	03/01/2020 12:00:00 AM	M	1900 S LONGWOOD AV	Wilshire	190326475
1	BURGLARY FROM VEHICLE	02/09/2020 12:00:00 AM	M	1000 S FLOWER ST	Central	200106753
2	BIKE - STOLEN	11/11/2020 12:00:00 AM	X	1400 W 37TH ST	Southwest	200320258
3	SHOPLIFTING-GRAND THEFT ($950.01 & OVER)	05/10/2023 12:00:00 AM	M	14000 RIVERSIDE DR	Van Nuys	200907217
4	VEHICLE - STOLEN	09/09/2020 12:00:00 AM	None	200 E AVENUE 28	Hollenbeck	200412582

Figure-1: Crime Data (Pipeline output)

Out[5]:

	Charge	Arrest Date	Arrest Type Code	Charge Description	Location	Area Name	Report ID
0	647(B)PC	07/14/2023 12:00:00 AM	M	PROSTITUTION	POINT (-118.3091 34.0845)	Hollywood	230612568
1	10851(A)VC	11/10/2023 12:00:00 AM	M	TAKE VEHICLE W/O OWNER'S CONSENT	POINT (-118.4032 34.0475)	West LA	230801062
2	41.18ALAMC	09/08/2023 12:00:00 AM	M	LOITER/OBSTRUCT STREET/SIDEWALK/ETC	POINT (-118.5976 34.2414)	Devonshire	231715221
3	459.5PC	07/14/2023 12:00:00 AM	M	SHOPLIFTING	POINT (-118.3563 34.0736)	Wilshire	230712023
4	29800(A)1PC	07/15/2023 12:00:00 AM	F	POSS F/ARM BY CONVICTED FELON/ADDICT/ETC	POINT (-118.3866 34.1922)	N Hollywood	6641569

Figure-2: Arrest Data (Pipeline output)

Results and Interpretation

Line Chart: Trends in Crime Reports vs. Arrests Over Time

The line chart shows the trends in crime reports and arrests over time, revealing seasonal patterns or significant changes in reporting and arresting practices.

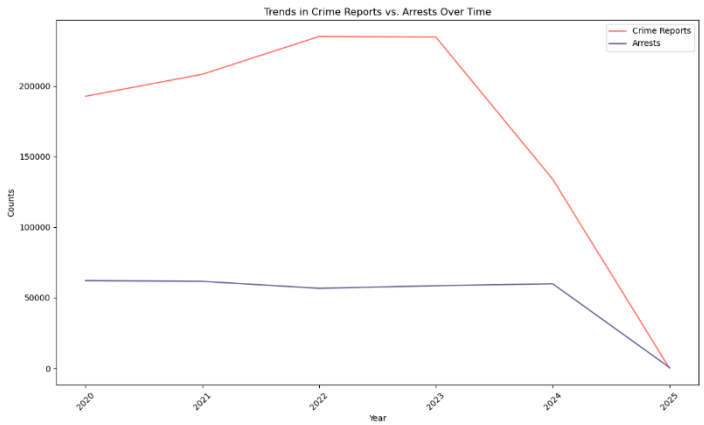

Figure-3: Line Chart

Bar Chart: Top 10 Areas with Highest Disparity

The bar chart reveals the top 10 neighborhoods with the highest disparities between crime reports and arrests. These regions show a significant gap, indicating potential under-policing or resource allocation issues.

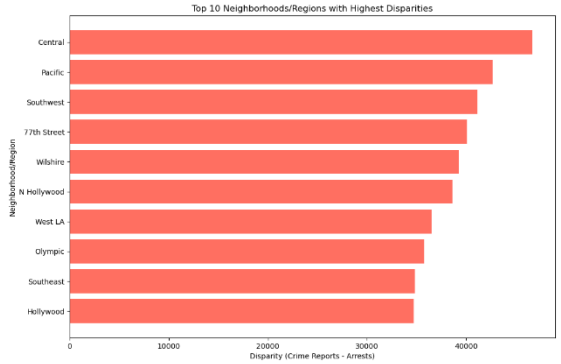

Figure-4: Bar Chart

Histogram: Distribution of Disparities

The histogram shows the frequency distribution of disparities, highlighting how many neighborhoods experience different levels of disparities. Most neighborhoods have small to moderate disparities, with a few experiencing very high disparities.

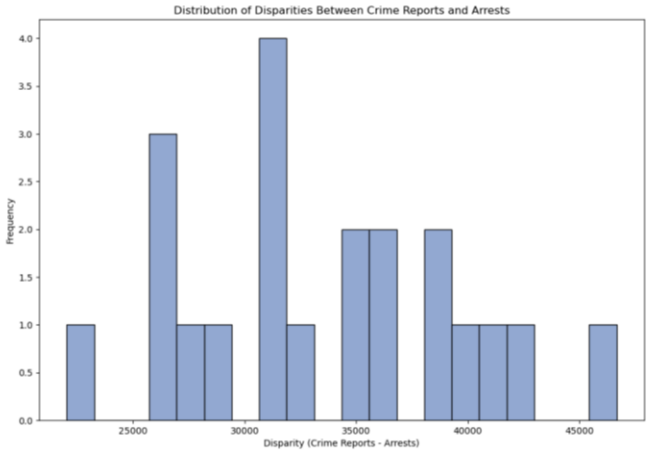

Figure-5: Histogram

Pie Chart: Distribution of Crime Reports and Arrests

The pie charts illustrate the proportion of crime reports and arrests across different neighborhoods. They provide a comparative view, showing how reports and arrests are distributed geographically.

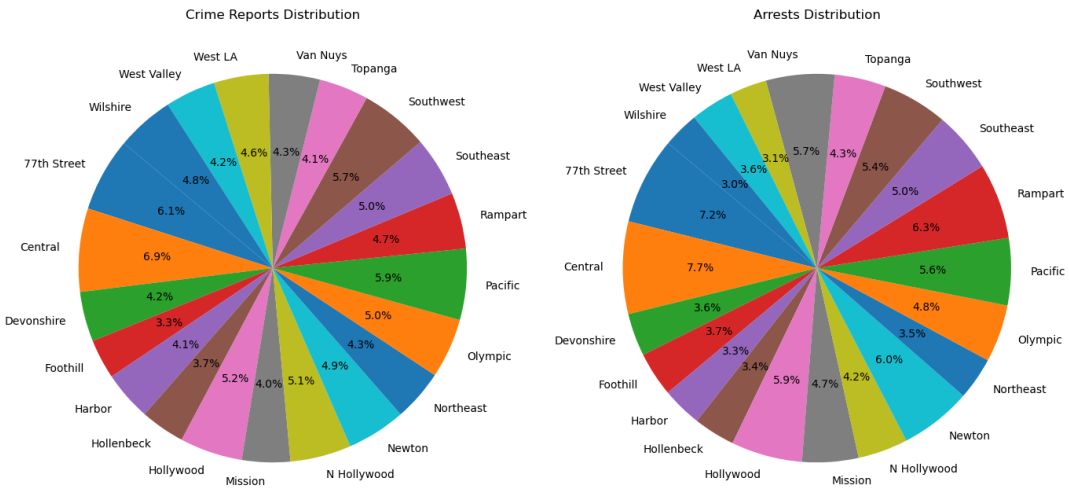

Figure-6: Pie Chart

Heatmap: Correlation between Crime Reports and Arrests

The heatmap shows the correlation between crime reports and arrests, indicating a positive relationship, but with room for improvement in achieving a higher correlation, which would suggest more consistent law enforcement practices.

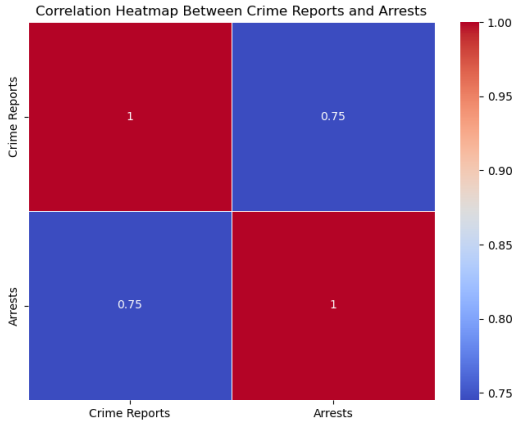

Figure-7: Heatmap

Area Chart: Cumulative Disparities Over Time

The area chart shows the cumulative disparities, providing a visual representation of how the gap between crime reports and arrests changes over time.

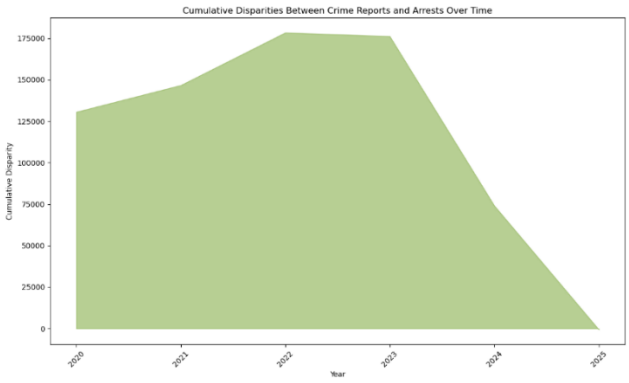

Figure-8: Area Chart

Conclusions

Analyzing crime reports and arrests across neighborhoods reveals significant disparities, highlighting areas that may need more focused law enforcement efforts. Some neighborhoods show a higher number of crime reports compared to arrests. By examining these disparities, we can better understand the dynamics of crime reporting and policing, identifying areas that may require more focused attention or resource allocation.

In summary, addressing these disparities requires a multifaceted approach, including resource reallocation, improved community engagement, and efficient crime resolution processes. Further research incorporating additional data and contextual factors can help develop targeted interventions to promote equitable law enforcement across all neighborhoods.

Limitations

Despite the valuable insights, the analysis has certain limitations. The accuracy and completeness of the data are critical factors; any missing or inaccurate records can affect the validity of the conclusions. Additionally, the analysis does not consider external factors such as socio-economic conditions, community-police relations, or policy changes that might influence crime rates and arrest patterns. These contextual factors are essential for a more comprehensive understanding of the disparities.