Exploratory Data Analysis (EDA) on Education and Crime in Chicago

AUTHOR

Penny Shi

Crime Data EDA

Data Loading and Cleaning

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2758 entries, 0 to 2757
Data columns (total 18 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Unnamed: 0            2758 non-null   int64
 1   ID                    2758 non-null   int64
 2   Case Number          2758 non-null   object
 3   Date                 2758 non-null   object
 4   Block                2758 non-null   object
 5   Primary Type         2758 non-null   object
 6   Description          2758 non-null   object
 7   Location Description 2758 non-null   object
 8   Arrest               2758 non-null   bool
 9   Domestic             2758 non-null   bool
 10  Community Area       2758 non-null   int64
 11  Year                 2758 non-null   int64
 12  Updated On           2758 non-null   object
 13  Latitude             2758 non-null   float64
 14  Longitude            2758 non-null   float64
 15  Location             2758 non-null   object
 16  Zip Codes            2758 non-null   object
 17  zip_code             2758 non-null   int64
dtypes: bool(2), float64(2), int64(5), object(9)
memory usage: 350.3+ KB
```
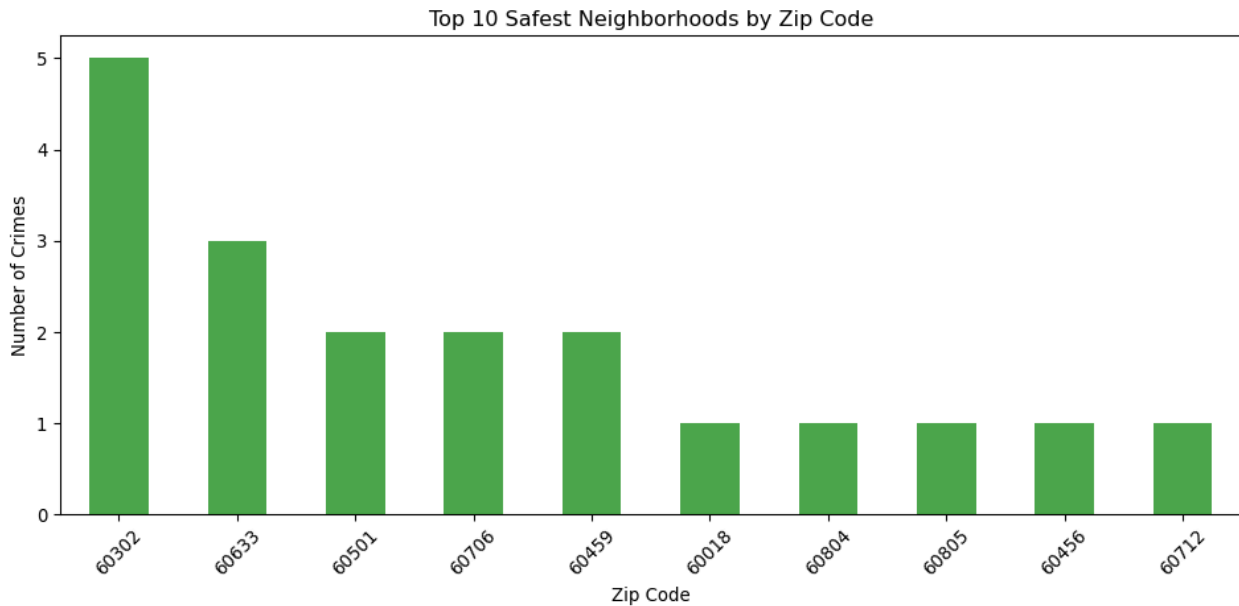
Data Visualization
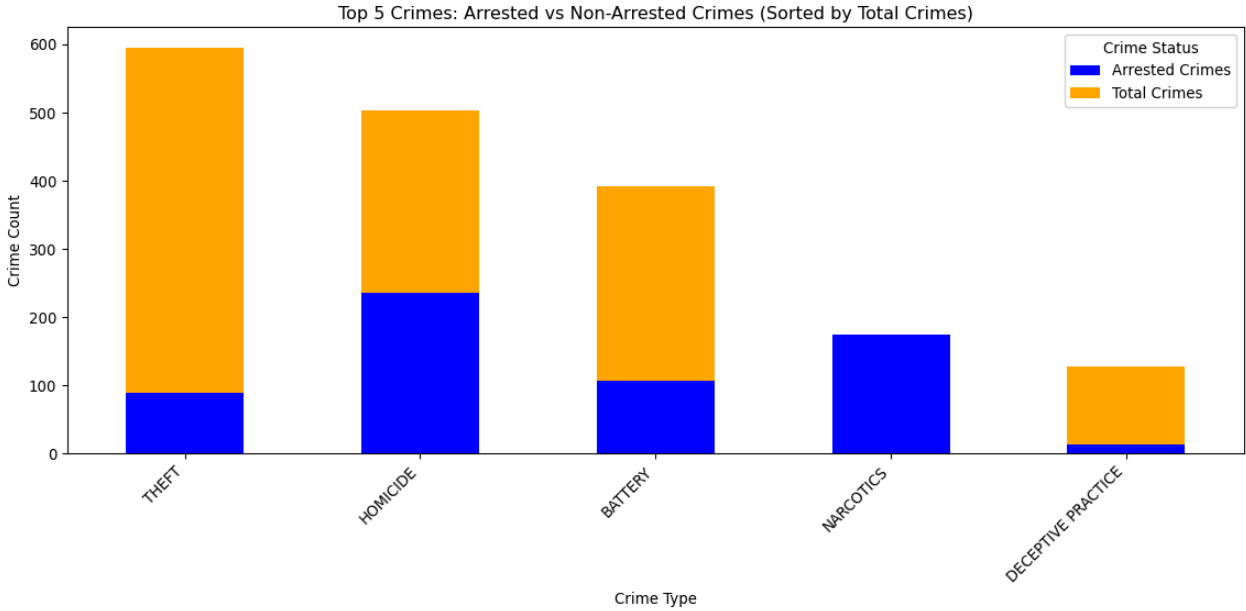
1. 10 Most Dangerous Neiborhoods (Zip Codes)
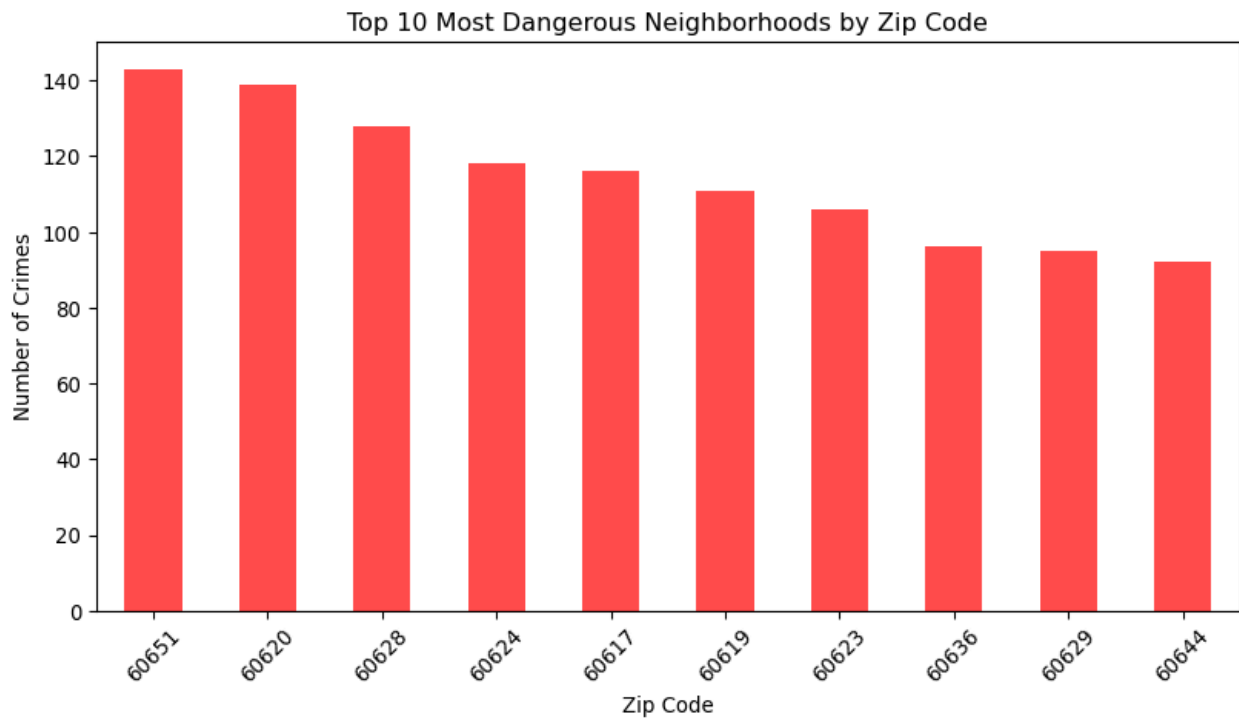
10 Most Dangerous Neiborhooods (Zip Codes)

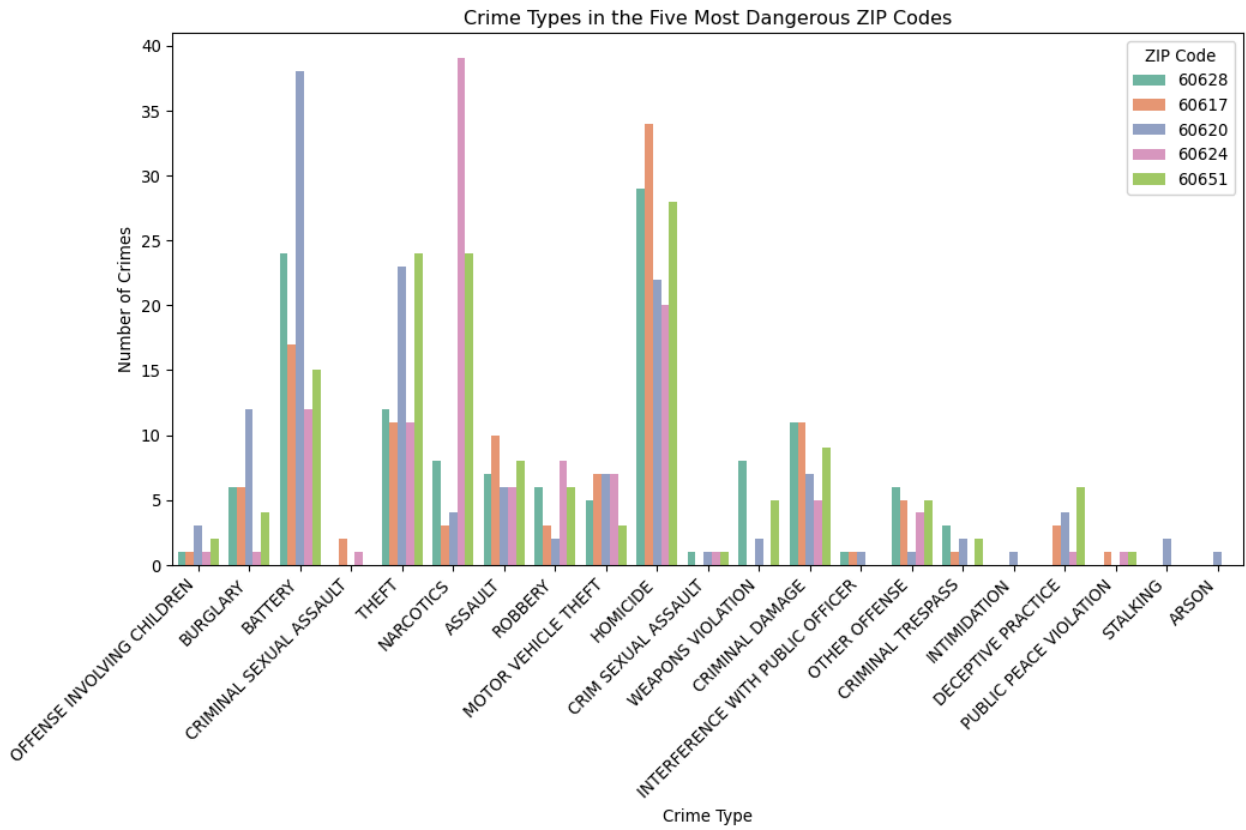## 2. 10 Safest Neiborhooods

10 Safest Neiborhooods (Zip Codes)

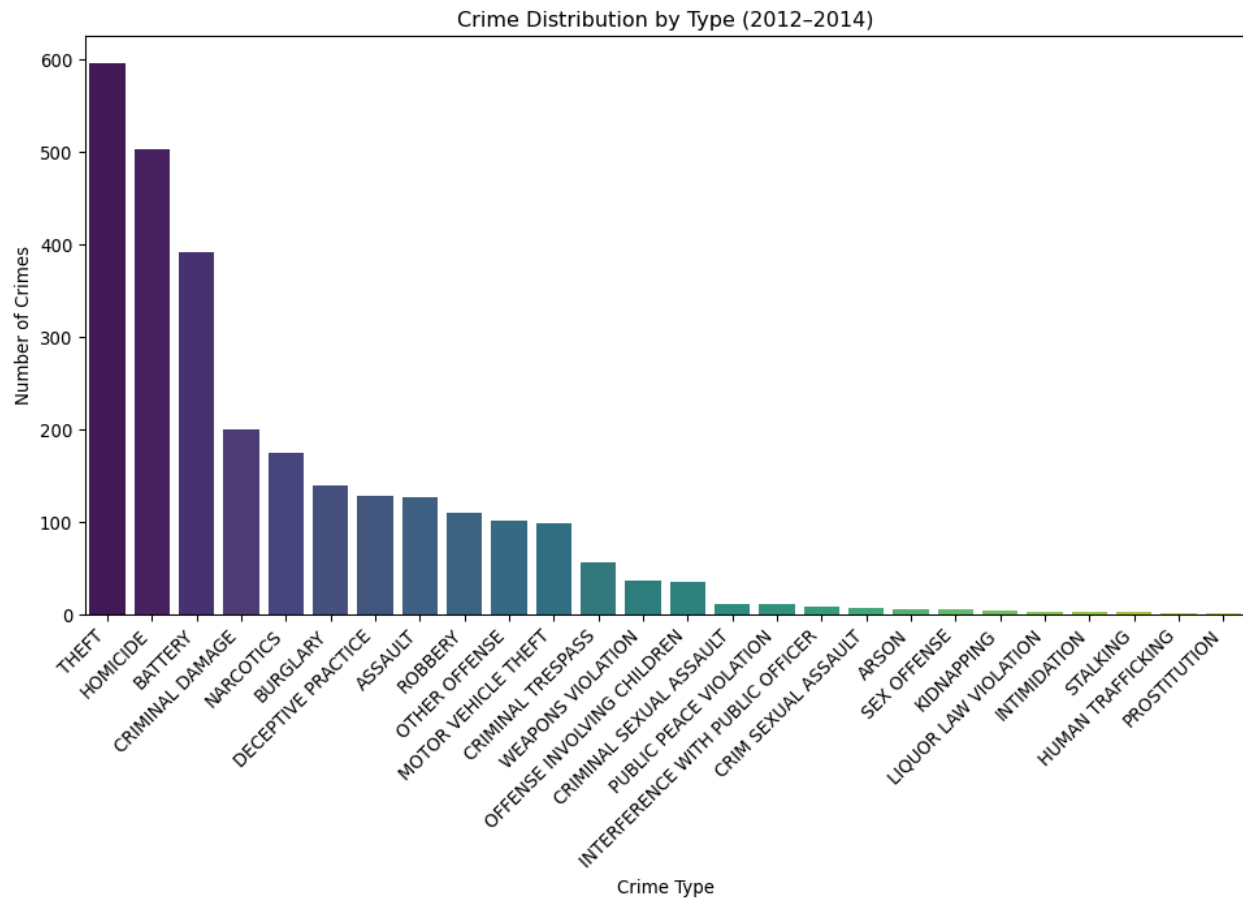## 3. Top 5 Crime Type by Crime Status

Top 5 Crime Type by Crime Status - Arrested Crimes & Non-Arrested Crimes)

4. Crime Type in the Most Dangerous Neighborhood
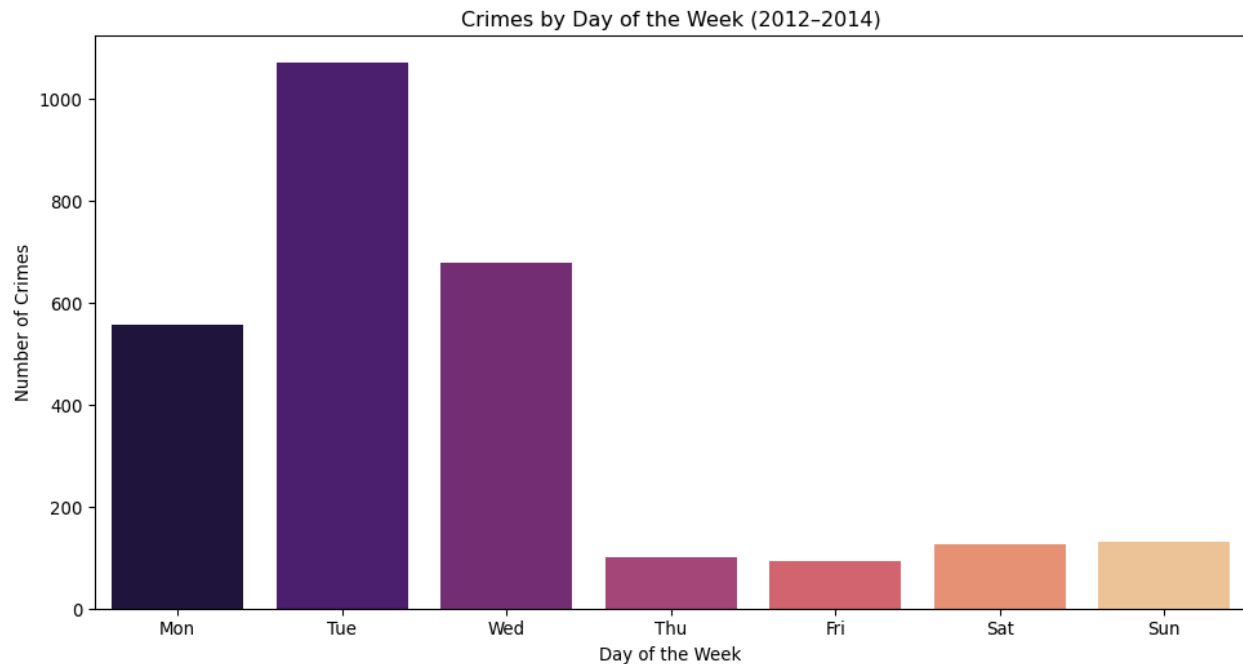

Crime Type in the Most Dangerous Neighborhood
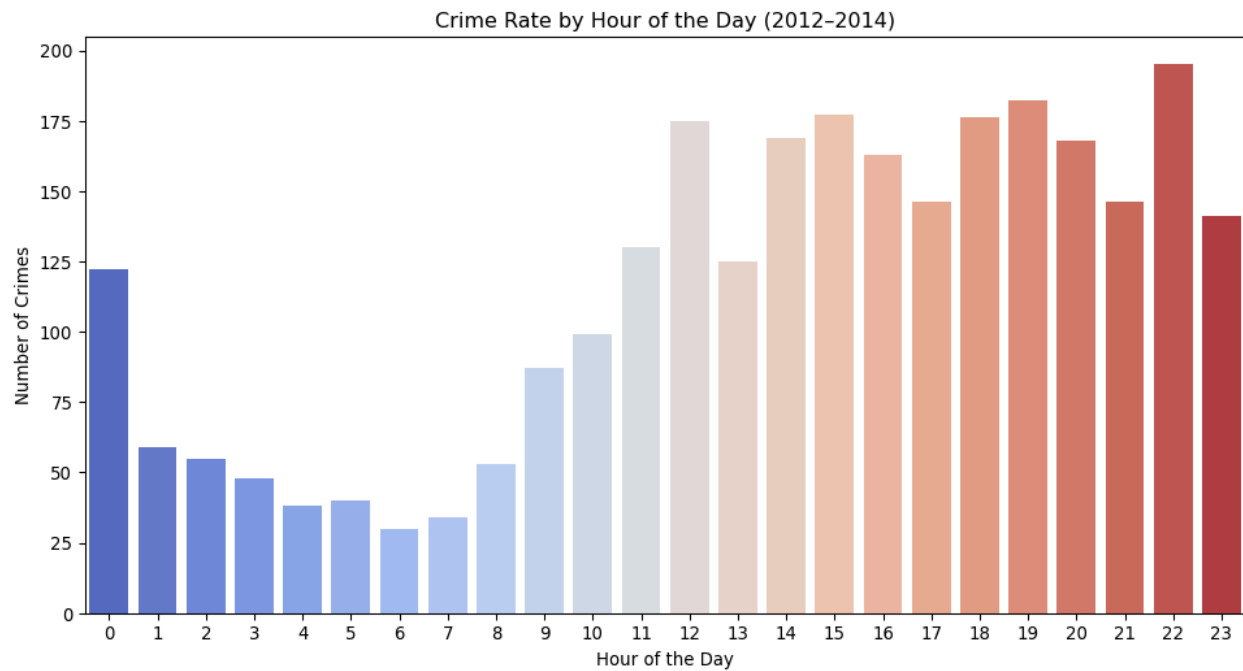
5. Crime Distribution by Type

Crime Type Distribution from 2012-2014

6. Crime Rate by Day of the Week

Crime Rate by Day of the Week

7. Crime Rate by Hour of the Day

Crime Rate by Hour of the Day (2012–2014)

Crime Rate by Hour of the Day

Education Data EDA

Data Loading and Cleaning

```
{'Total Columns': 157,
 'Preview of Data':      School ID      School Short Name  \
 0      610340            Chgo Acad HS
 1      400018   Austin Bus & Entrp HS
 2      610524               Alcott HS
 3      609678                Jones HS
 4      609692              Simeon HS


                                   School Name      Street Address_x  \
 0                  Chicago Academy High School     3400 N Austin Ave
 1   Austin Business and Entrepreneurship Academy HS     231 N Pine Ave
 2           Alcott High School for the Humanities     2957 N Hoyne Ave
 3    William Jones College Preparatory High School       606 S State St
 4        Neal F Simeon Career Academy High School  8147 S Vincennes Ave


      City_x State_x    ZIP                         Website_x  \
 0   Chicago      IL  60634          http://www.chicagoacademy.org
 1   Chicago      IL  60644   http://www.austinbusinessacademy.net
 2   Chicago      IL  60618                http://www.alcotths.org
 3   Chicago      IL  60605          http://www.jonescollegeprep.org
 4   Chicago      IL  60620                http://www.simeonca.org

                       Managed By Network  \
 0                            AUSL Schools
```

```
1                West Side High School Network
2   North-Northwest Side High School Network
3                South Side High School Network
4                South Side High School Network


                          Geographic Network  ...  \
0  North-Northwest Side High School Network  ...
1                West Side High School Network  ...
2  North-Northwest Side High School Network  ...
3                South Side High School Network  ...
4                South Side High School Network  ...


  Teacher Attendance Percentage 2013 Teacher Attendance Percentage 2012  \
0                               95.9                               95.6
1                               95.0                               94.3
2                                NaN                                NaN
3                               95.1                               95.3
4                               93.5                               94.6


  Gr3-8 On-Track Percentage 2013  One-Year DropOut Rate Percentage 2013  \
0                            NaN                                    5.2
1                            NaN                                    1.4
2                            NaN                                    1.1
3                            NaN                                    1.0
4                            NaN                                    2.9


  One-Year DropOut Rate Percentage 2012  X Coordinate_y  Y Coordinate_y  \
0                                   4.0     1135740.091     1922002.941
1                                   6.9     1139494.763     1901274.258
2                                   1.8     1161870.556     1919857.440
3                                   1.5     1176428.317     1897119.100
4                                   4.6     1174719.074     1850730.192


   Longitude_y  Latitude_y                     Location_y
0    41.942154  -87.776506  (-87.77650575, 41.94215439)
1    41.885205  -87.763212  (-87.76321191, 41.88520477)
2    41.935761  -87.680524  (-87.68052441, 41.93576106)
3    41.873049  -87.627711  (-87.62771142, 41.87304937)
4    41.745792  -87.635370     (-87.63537, 41.74579217)


[5 rows x 157 columns]}
```
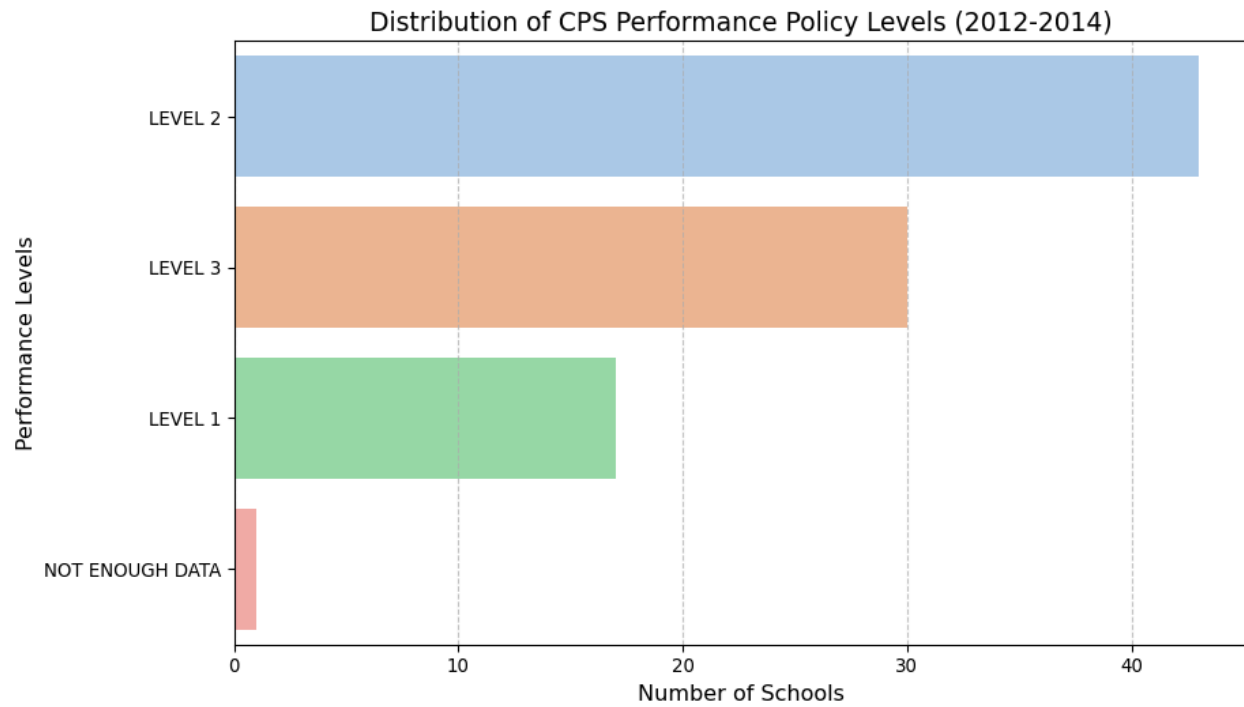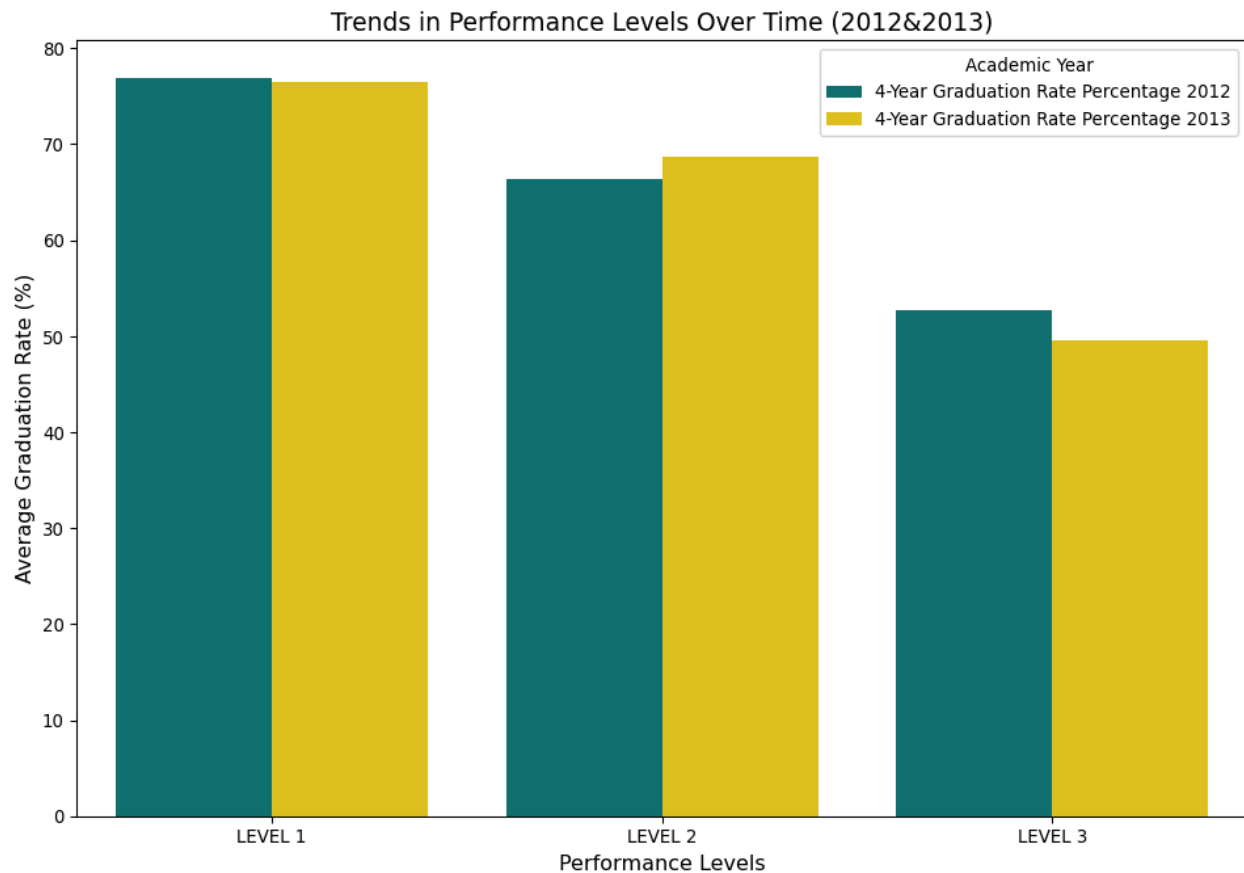
Columns: 157 Rows: 91

Data Visualization

---

1. Distribution of CPS Performance Policy Levels (Level 1-3)

Distribution of CPS Performance Policy Levels (2012-2014)

Distribution of CPS Performance Policy Levels

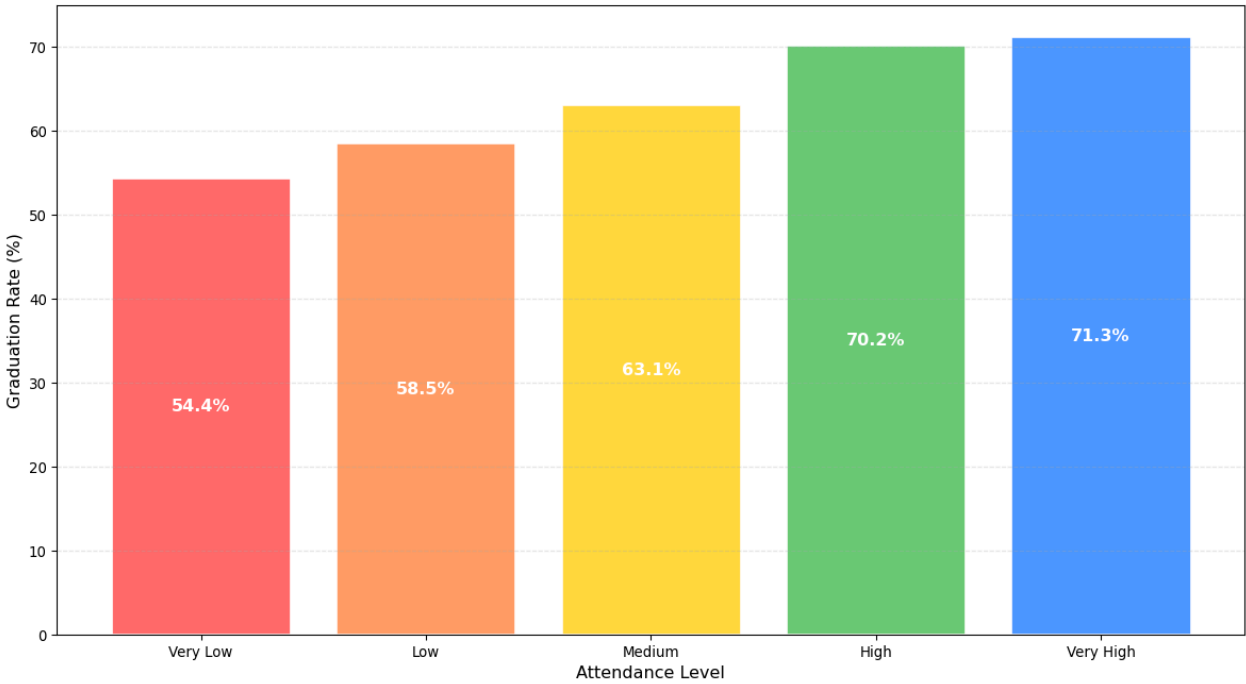## 2. Performance Levels and Graduation Rates

Trends in Performance Levels Over Time (2012&2013)

Relationship between Performance Level and Graduation Rates

3. Attendance Level and Graduation Rates

Higher Attendance → Higher Graduation Rates

Relationship between School Attendance and Graduation Rates

4. Top Performing and Bottom Performing Education Zip Codes
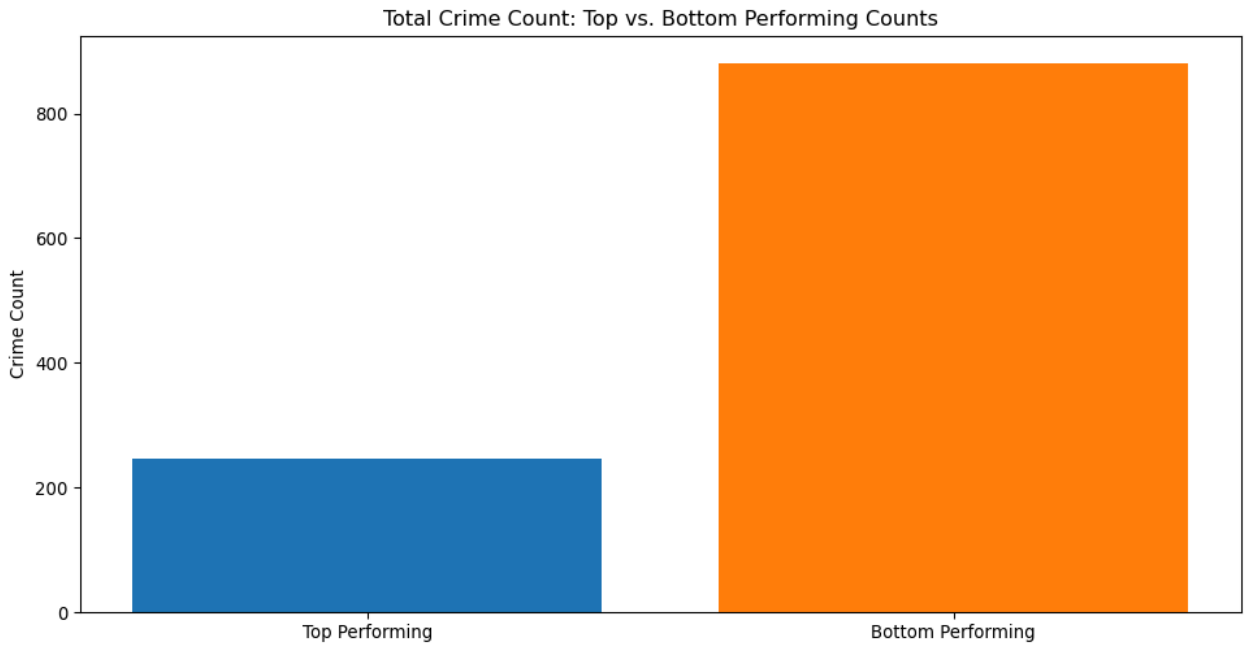
Top vs. Bottom Performing Zip Code Crime Counts (based on Graduation Rates)

Race & Ethnicity and Education

Data Loading and Cleaning

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 235 entries, 0 to 234
```

```
Data columns (total 25 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   geography_type       235 non-null    object
 1   year                 235 non-null    int64
 2   geography            235 non-null    object
 3   total                235 non-null    int64
 4   age_0-17             235 non-null    int64
 5   age_18-29            235 non-null    int64
 6   age_30-39            235 non-null    int64
 7   age_40-49            235 non-null    int64
 8   age_50-59            235 non-null    int64
 9   age_60-69            235 non-null    int64
 10  age_70-79            235 non-null    int64
 11  age_80+              235 non-null    int64
 12  age_0-4              235 non-null    int64
 13  age_5-11             235 non-null    float64
 14  age_12-17            235 non-null    float64
 15  age_5+               235 non-null    int64
 16  age_18+              235 non-null    int64
 17  age_65+              235 non-null    int64
 18  female               235 non-null    int64
 19  male                 235 non-null    int64
 20  latinx               235 non-null    int64
 21  asian_non-latinx     235 non-null    float64
 22  black_non-latinx     235 non-null    float64
 23  white_non-latinx     235 non-null    int64
 24  other_race_non-latinx  235 non-null  float64
dtypes: float64(5), int64(18), object(2)
memory usage: 46.0+ KB
```
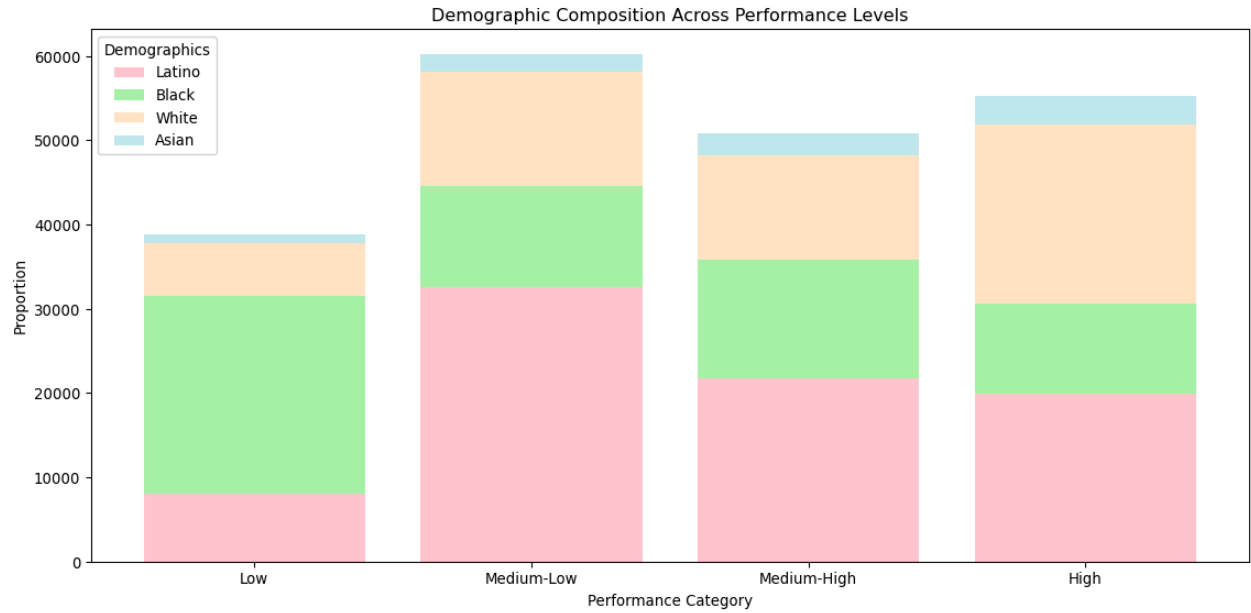
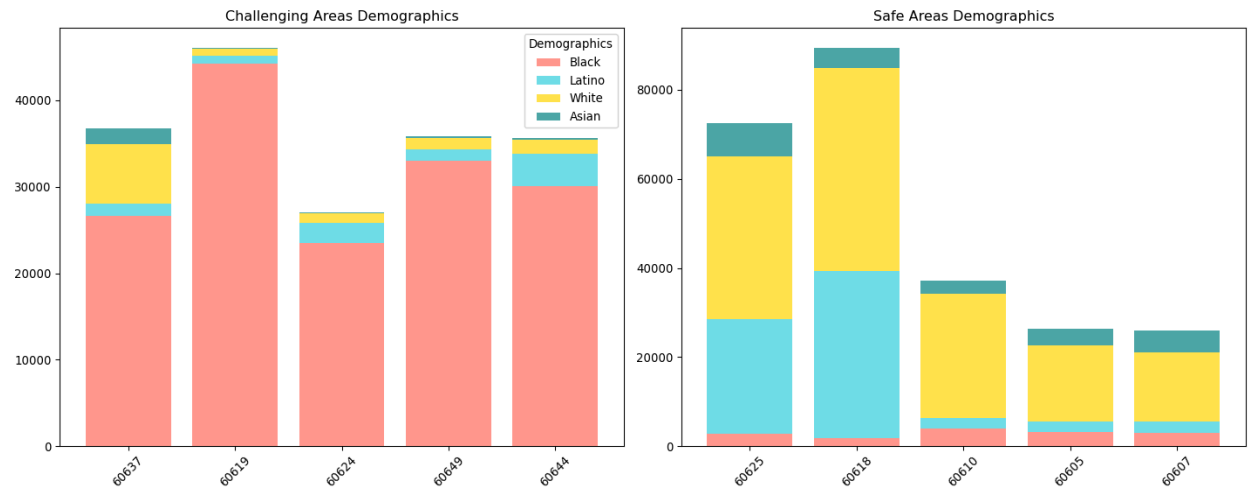## 5. Demographics in High & Low Education Performing Areas
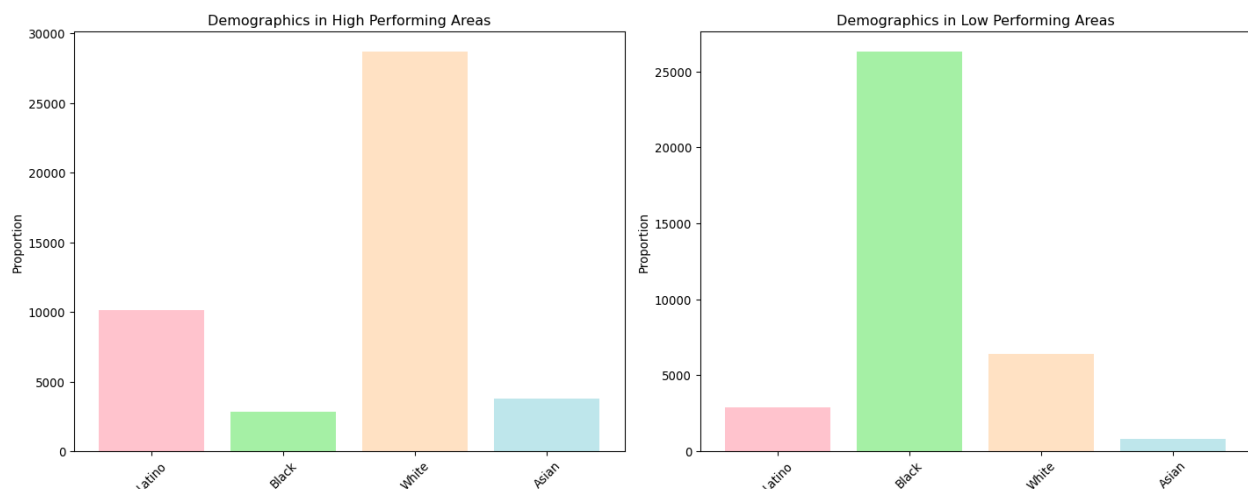
Demographics in High Performing Area and Low Performing Area

6. Demographics Composition Across Performance Levels

Demographic Composition Across Performance Levels

Demographic Composition Across Different Performance Levels

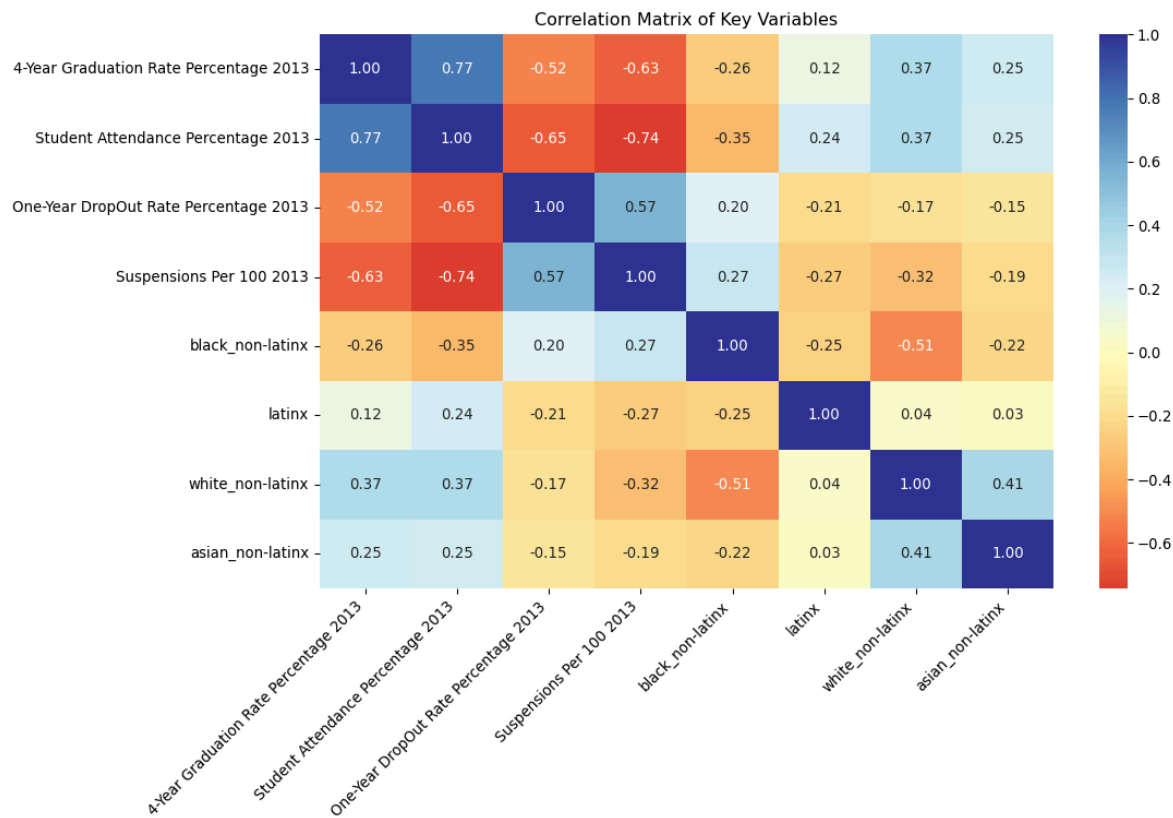# 7. Demographics in Challenging Areas and Safe Areas

Demographics in Challenging Areas and Safe Areas

8. Correlation Matrix of Key Variables

Key Variables

- 4-Year Graduation Rate Percentage 2013

- Student Attendance Percentage 2013

- One-Year DropOut Rate Percentage 2013

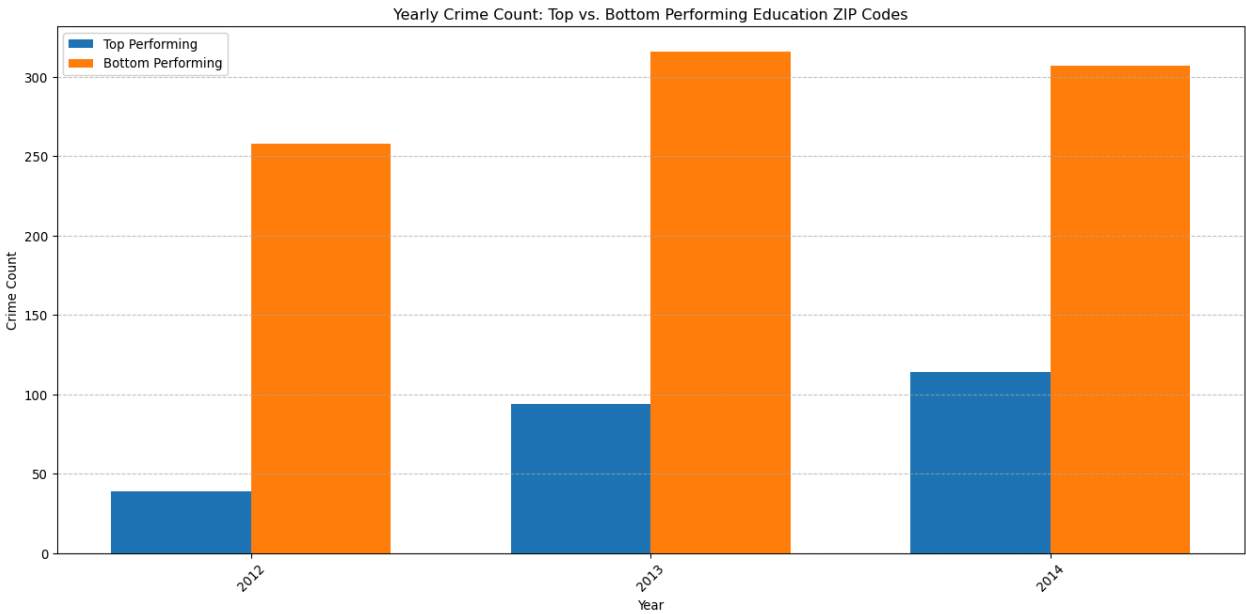- Suspensions Per 100 2013

- black_non-latinx

- latinx

- white_non-latinx

- asian_non-latinx

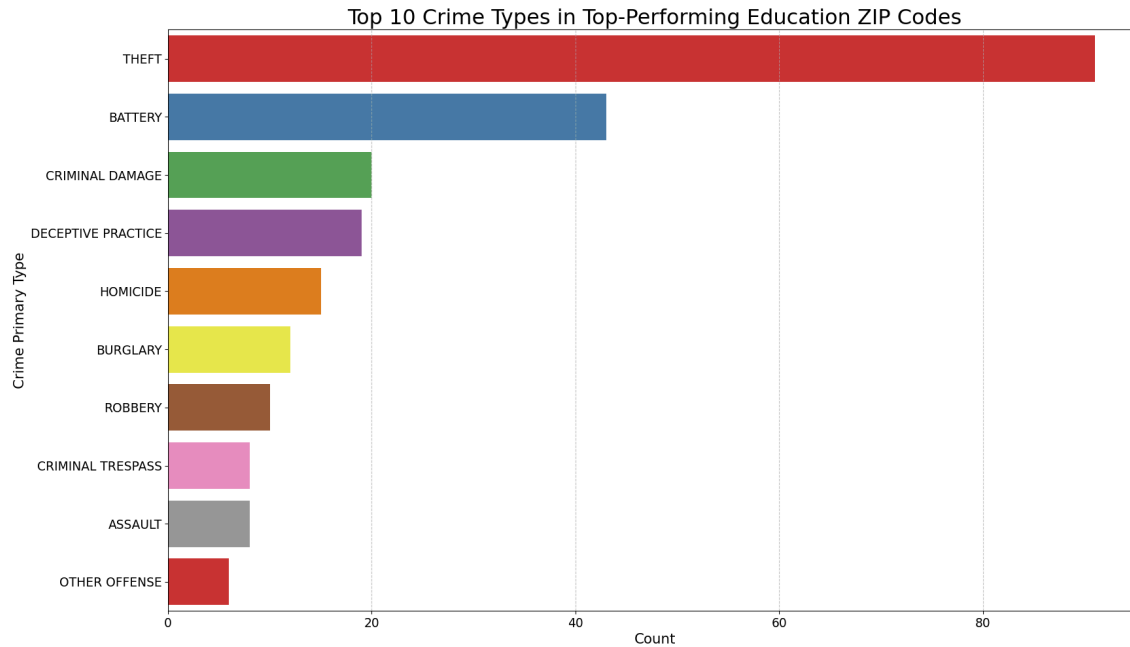**Correlation Matrix of Key Variables between Demographics & School Performance**

Crime and Education

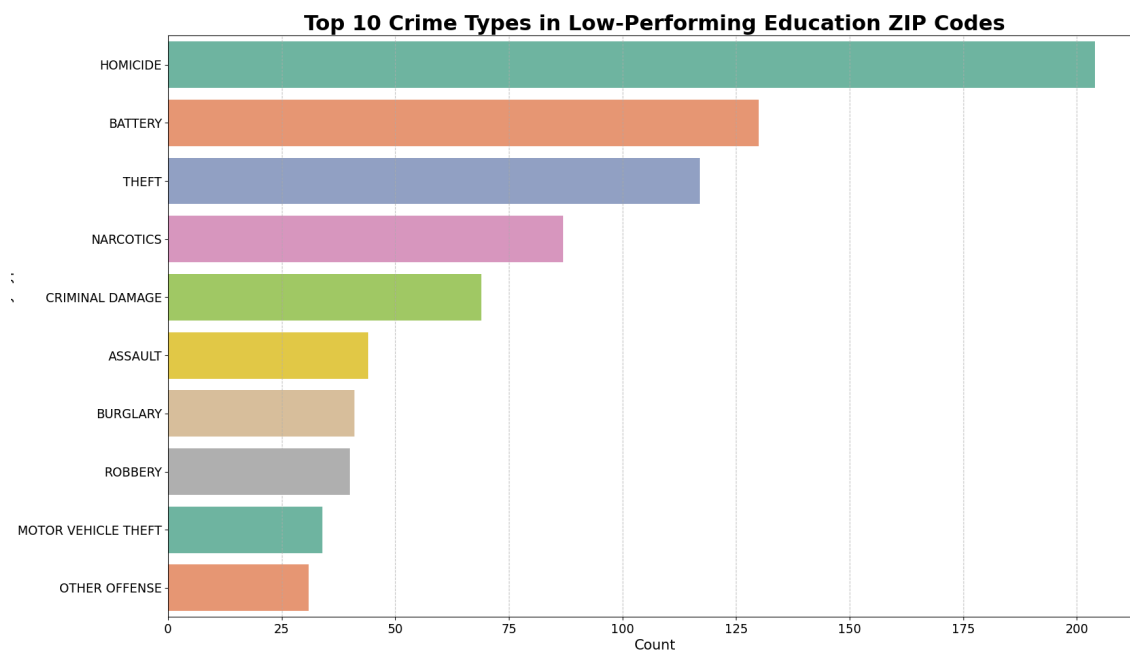## 9. Yearly Crime Counts: Top vs. Bottom Education Performing Zip Codes

Top vs. Bottom Performing Zip Code Crime Counts over Years

10. Crime Type in Top Education Performing Zip Codes

Top 10 Crime Type in Top Education Performing Zip Codes

11. Crime Type in Bottom Education Performing Zip Codes

Top 10 Crime Type in Bottom Education Performing Zip Codes

Analysis:

- First degree murder is the most committed crime in least performing neighborhoods, while it does not appear in the top performing neighborhoods.

- Narcotics-related crime - prevalent to spread through schools - also takes one of the top spots in least performing neighborhoods, and this is much lower in top performing neighborhoods.