Analyzing Crime Data in the Charlotte Area

Brittany Marchand, Brian Peaver, Danny Barto, Elizabeth Patterson

Topic

Why topic was selected

- Team members live in the area
- Vested interest in public safety
- Area's steady growth
- More effective law enforcement

Questions we’re hoping to answer

- Overall number of calls received by CMPD
- Breakdown of calls by category and service area
- Can certain types of calls be indicative of other types of crimes that might be occurring?

Data Used

City of Charlotte Open Data Portal

- Call Data
- Offense Data
- Incident Data

Cleaning the Data

- Remove unneeded columns
- Remove nulls
- Consistency
-

```
In [83]: clean_date_4_df.CITY.unique()

Out[83]: array(['CHARLOTTE', 'MATTHEWS', 'MECKLENBURG', 'PINEVILLE',
       'HUNTERSVILLE', 'MINT HILL', 'CHARLOTTE, NC 28211', 'CHAROLETTE',
       '28277', 'CHARLOTTE NC', 'CHARLOTE', 'HUMTERSVILLE', 'CHARLOLTTE',
       'CHARLOTTE, NC 28209', 'CHARLOTTE,', 'CHARLOTT', 'MIDLAND',
       'MATTHES', 'MATHEWS', 'CORNELIUS', 'C', 'CHARLOTTTE',
       'CHARLOTTE, NC 28206', 'CHARLOOTE', 'CHARLOTTE, NC',
       'CHARLOTTE,  28211', 'DAVIDSON', 'CHAROLTTE', 'CHRALOTTE',
       'PINEVLE', 'MINT HIL', 'FORT MILL', 'CHARROLTE', 'CHARLTOTE',
       'RT SIDE', 'CHARLOTT3215E', 'MECKLENBRUG', '28273',
       'UNKNOWN/REFUSED', 'BALLANTYNE', 'HUNTERVILLE', 'CHARTLOTTE',
       '110', 'OTHER/NOT LISTED', '28205', 'CHRLOTTE', 'MOUNT HOLLY',
       '28226', 'CHARRLOTE', 'CONCORD', 'CHAARLOTTE', '1640 DEWBERRY TER',
       'CHARLOTTE/NC/28269', 'CHARLOTTLE', 'MECKLENBERG',
       'CHARLOTTEJAVASCRIPT:VOID PT_SU', 'CHARLOTTEE', 'MINTHILL', 'H',
       '28210'], dtype=object)

In [97]: clean_date_4_df.loc[clean_date_4_df['CITY'].str.contains('CH', case=False), 'CITY_NEW'] = 'Charlotte'
         clean_date_4_df.loc[clean_date_4_df['CITY'].str.contains('MAT', case=False), 'CITY_NEW'] = 'Matthews'
         clean_date_4_df.loc[clean_date_4_df['CITY'].str.contains('MECK', case=False), 'CITY_NEW'] = 'Mecklenburg'
         clean_date_4_df.loc[clean_date_4_df['CITY'].str.contains('HU', case=False), 'CITY_NEW'] = 'Huntersville'
         clean_date_4_df.loc[clean_date_4_df['CITY'].str.contains('MINT', case=False), 'CITY_NEW'] = 'Mint Hill'
         clean_date_4_df.loc[clean_date_4_df['CITY'].str.contains('PIN', case=False), 'CITY_NEW'] = 'Pineville'
         clean_date_4_df.loc[clean_date_4_df['CITY'].str.contains('CORNELIUS', case=False), 'CITY_NEW'] = 'Cornelius'
         clean_date_4_df.loc[clean_date_4_df['CITY'].str.contains('MIDLAND', case=False), 'CITY_NEW'] = 'Midland'
         clean_date_4_df.loc[clean_date_4_df['CITY'].str.contains('DAVIDSON', case=False), 'CITY_NEW'] = 'Davidson'
         clean_date_4_df.loc[clean_date_4_df['CITY'].str.contains('BALLANTYNE', case=False), 'CITY_NEW'] = 'Ballantyne'
         clean_date_4_df.loc[clean_date_4_df['CITY'].str.contains('FORT MILL', case=False), 'CITY_NEW'] = 'Fort Mill'
         clean_date_4_df.loc[clean_date_4_df['CITY'].str.contains('MOUNT HOLLY', case=False), 'CITY_NEW'] = 'Mount Holly'
         clean_date_4_df.loc[clean_date_4_df['CITY'].str.contains('CONCORD', case=False), 'CITY_NEW'] = 'Concord'
         clean_date_4_df
```

Database

- PostgreSQL- store tabular data
- SQLAlchemy- communicate databases and machine learning model

Schema

(image)

Machine Learning

Supervised Learning-Classification-Logistic Regression

Categorize violent versus non-violent crimes

Goal- map out violent crime to see where heaviest to better allocate police resources

Results

Dashboard

Summary