Crime in Philadelphia

BY

Nimmani Vyshnavi

Data Acquisition:

- Data source is obtained from official Philadelphia city website
- https://metadata.phila.gov/index.html#home/
- Data consists columns corresponding to dispatch date, dispatch time, location, point_x, point_y, latitude, longitude.
- Data ranges from year 2010 to 2022.
- Using heat map and scatter matrix to understand correlation of data.

Performing initial data analysis to check count of crimes in every year

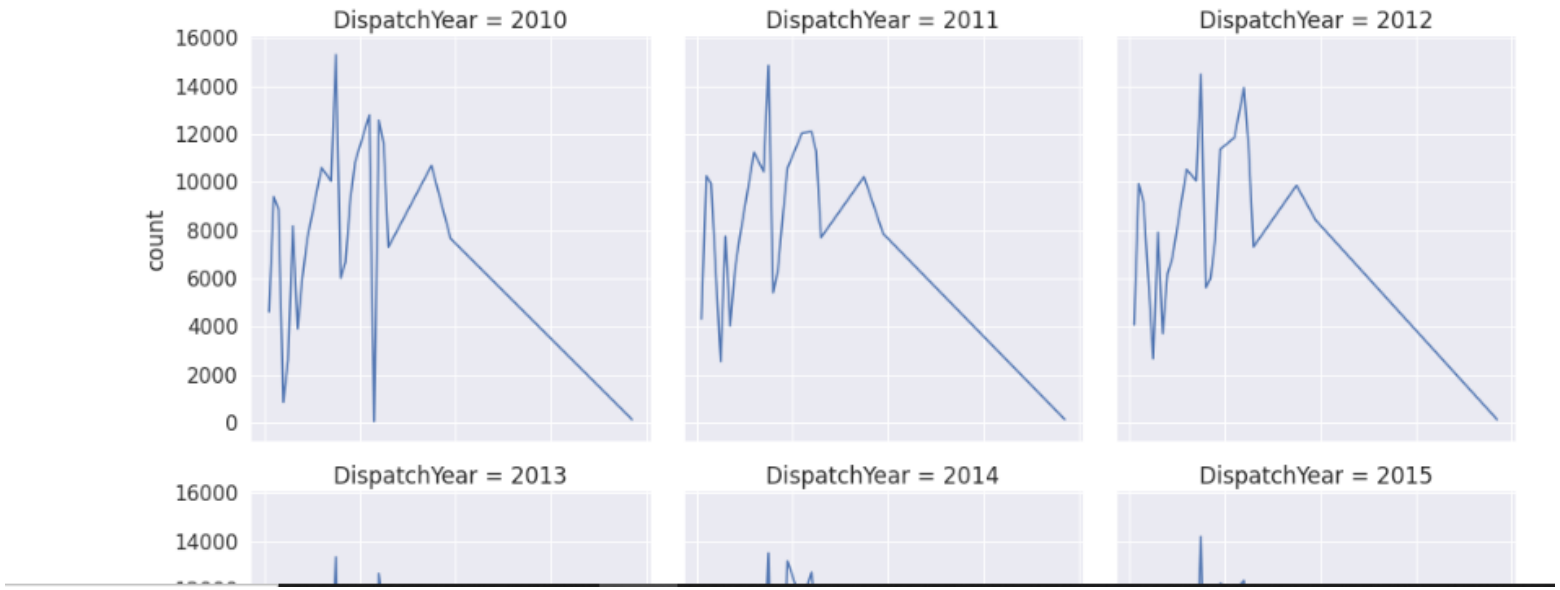
In [0]:

```
crime_dc_dist = (crime_df_data
                 .select("DCDist","DispatchYear")
                 .groupBy("DCDist","DispatchYear")
                 .count()
                 .orderBy("DCDist","DispatchYear")).toPandas()

fig = plt.figure(num=None, figsize=(50, 30), dpi=80)

sb.relplot(x= "DCDist", y = "count", col="DispatchYear",kind='line',col_wrap=3, data = crime_dc_dist)
```

Out[25]:



Problems:

- Performing EDA using pandas, seaborn, matplotlib to predict the probability of crimes by using time series analysis to see any seasonal behaviors.
- Using data to analyze district and area level we can predict the areas which are highly prone to crimes.
- Implementing binary classification methods like Logistic Regression, Gradient Boosting Classification, Random Forest Classification to generate my expected predictions.
- I hope to arrive at conclusion of predicting crimes based on seasons, areas using ML models and visualize these with respect to years.

Thank you