Final Report:

- **10 Pages**
- **Same format as checkpoints**
- **Sections**
  - **Abstract**
    - 1-2 paragraphs, executive summary
  - **Introduction**
    - What problem?
    - Why this problem?
    - Potential Challenges
    - Key ideas
  - **Related Work**
    - What has been done in the literature
    - Limitations?
    - How is our work different?
  - **Methodology**
    - Datasets, tools, main tasks, analytical thinking
    - Data understanding, preprocessing, warehouse, modeling
  - **Evaluation**
    - Metrics, evaluation setup, baseline methods, results (figures, tables w/ proper labels), Interpretation
  - **Discussion**
    - Lessons learned, what worked well, what didn't, directions for future work
  - **Conclusion**
    - Summary, reiterate key tasks & findings
  - **References**
    - Check format for proper citations
  - **Appendix**
    - Honor code pledge, Individual contribution

Bayesian:

P(H|X) = (P(H and X)) / (P(X))

X - Temp

H - Type

= (# of type temp entries) / #temp entries

= (hot property crimes)/(hotcrimes)

- $$P(A|B) = \frac{|A \cap B|}{|B|} = \frac{\frac{|A \cap B|}{|S|}}{\frac{|B|}{|S|}} = \frac{P(A \cap B)}{P(B)}.$$

Things to Consider:

- To help with our predictions should we pair crimes together (EX *vehicle related crimes *Poverty crimes etc) (how many crime types)
- If we really want to we could even do a comparison between high and low-income areas or city vs suburbs
- After clean up, should we pair/ compare the cold places with cold places and hot with hot to see if there's any similarities or differences then later kinda do cold vs hot

Scope:

- How will we categorize temperatures (Ex **Cold** < 50, **Mild** 50-80 ,**Hot** > 80)
- Person, Property, society for crime categories

Global Columns:

- Crime Tables
    - Date/Time Occurred
    - Area
    - Crime Code/ type of crime
        - Check crime type with month
            - Show yearly types of crime
        - Find similar naming
            - Assault of any type
            - Theft/Burglary
            - Drugs
            - Other
        - Identify outliers with valid names
        - Check number of outliers with some assigned limit of occurrences
- Weather Tables
    - Date
    - Temp
        - Average
    - where

Create overall data data representation(IN JUPYTER NOTEBOOK)
Per city;

1. # of occurrences + date (see dips through year)
2. Counts of crime types (against, not against, other)
3. Counts of weather types (hot, mild, cold) (base on average temp/day)

Create hypothesis based off of overall/other research

1. What crimes happen on sunny, rainy, snowy, hot, cold, mild days
2. Crime group classifications, do certain crimes pop up with others?
3. Society, property, person

Outliers:

Date range
City pop
Crime count ~350,000

Meeting: 11/20:

1. Merge df's on date column
    a. Several crimes per weather day
2. Find hot, cold, mild days
    a. Find crimes occurring on those days (type)
    b. Bayesian (% probability of crime type on certain temp day)
3. Per-capita crime rates
    a. Slim down date range
    b. From date range decide how off crime entries are (# of entries)

Meeting: 11/29:

1. Possible missing % of crimes between notebooks
    a. Boston ~11%
    b. LA ~ 20
2. Add your findings (Percent of Crime types for Weather boxes)
    a. % cold crimes etc.
3. Narrow to three data frames
    a. Cold
    b. Mild
    c. Hot
4. From each "Day Temp" frame compare numbers and probabilities between cities based on these three frames
    a. Small data => big data
5. Compare cities
6. Decide what to include in final
    a. Crimes throughout the year
    b. Temp throughout year
    c. Comparison of cities

Meeting: 12/03:

1. From each "Day Temp" frame compare numbers and probabilities between cities based on these three frames
    a. Small data => big data
2. Compare cities

Meeting:

3. From each "Day Temp" frame compare numbers and probabilities between cities based on these three frames
    a. Small data => big data
4. Compare cities

Bayesian classification**

Perhaps apriori in the different weather constraints

Long term:
Decision tree

Total Crime	High	Mild	Cold	Total
Property	**15.5%**	**34.5%**	**10.4%**	**60.4%**
People	**5.6%**	**11.9%**	**4.1%**	**21.6%**
Society	**3.2%**	**7.9%**	**2.9%**	**14%**
Total	**24.3%**	**54.3%**	**17.4%**	**96%**

Total **Temp by total crime distribution**	**Hot**	**Mild**	**Cold**
Total	**25.39**	**56.54**	**18.06**

Total Avg Bayes	**Hot**	**Mild**	**Cold**
Property	**63.2**	**62.6**	**63.2**
People	**23.2**	**22.6**	**22.9**
Society	**13.6**	**14.8**	**13.9**
Total	**100%**	**100%**	**100%**

Year Spread in Data Links from Report:
Dylan - 2008 - Dec 13 2017 (~8 full years)
Jon - 2010-2019 (9 full years)
Lauren - Jan 1 2017 - Oct 19 2020 (~3 full years)
Brady - Jun 15 2015 - Apr 1 2018 (~2 full years)

LA - Jon G
Crime dates - (Jan 1 2010 - 2019)
Weather dates - (Jan 1 1997 - Present)

Findings:

Crime type distribution

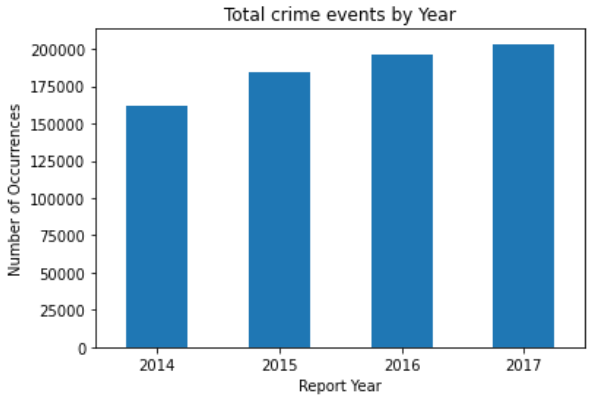
LA Crime Type Distribution	Hot	Mild	Cold	Total
Property	**30.26%**	**31.97%**	**0.56%**	**62.79%**
People	**12.14%**	**11.65%**	**0.2%**	**23.99%**
Society	**3.06%**	**4.67%**	**0.05%**	**7.78%**
Total	**45.46**	**48.29%**	**0.81%**	**94.56%**

Percent of crimes missing: 7.129

Bayesian:

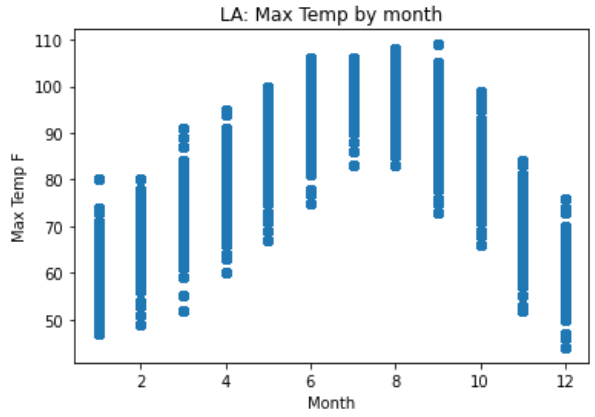
LA	Hot	Mild	Cold
Property	**66.56**	**66.21**	**69.41**
People	**26.71**	**24.11**	**24.58**
Society	**6.73**	**9.68**	**6.01**
Total	**~100%**	**~100%**	**~100%**

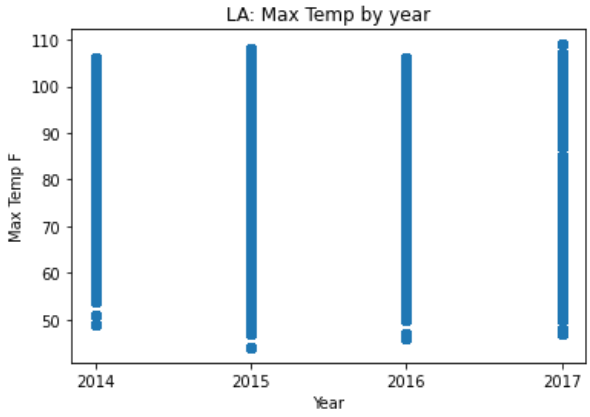
Crime by Month (LA)	Crime by Year (LA)



Max Temp Range By Month

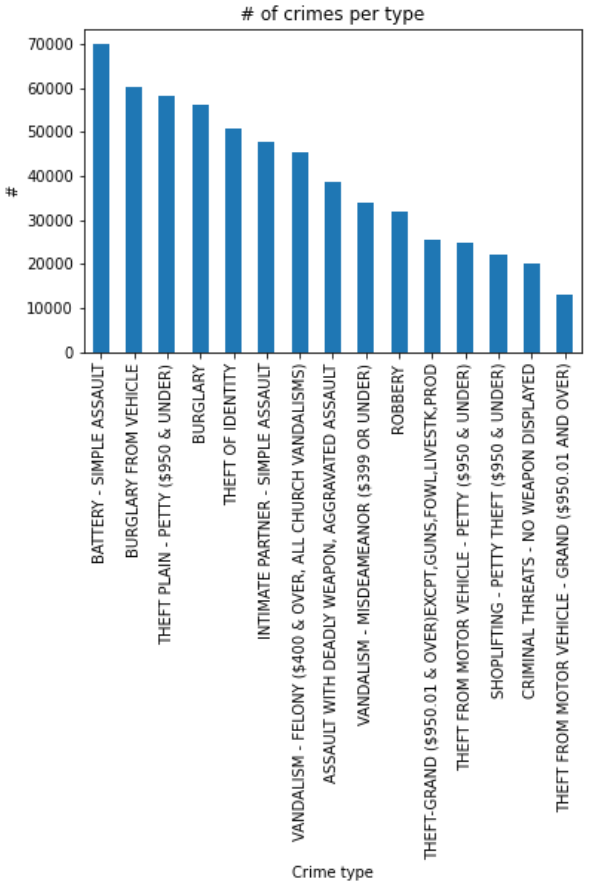
Avg Temp Range by Year & month





# of Crimes per type

# of Crimes per type



CO- Lauren

Crime dates - (Jan 1 2017 - Oct 3 2022)

Weather - (Jan 1 2002 - Oct 19 2020)

Jan 1 2017 - Oct 19 2020

Percentage of Temp distribution:

Cold -34.82

Mild - 54.12

Hot - 11.01

Percentage of crime distribution:

Crime on Property - 46.07

Crime on Person -12.99

Crime on Society -40.94

Crime type distribution Average Temp

Denver Crime	Hot	Mild	Cold	Total

Property	5.22	24.85	16.00	46.07
People	1.48	7.09	4.41	12.98
Society	4.3	22.23	14.41	40.94
Total	11	54.17	34.82	99.99

Bayesian Classification Average Temp

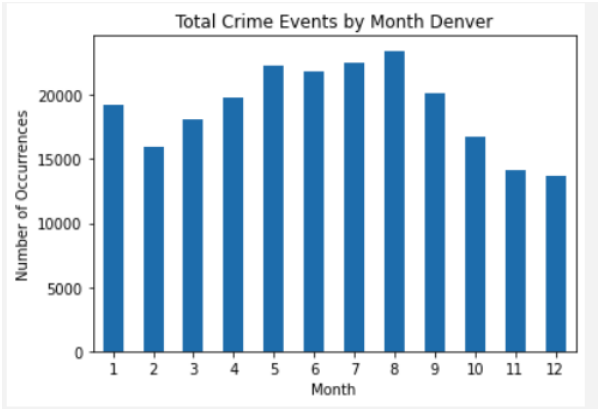
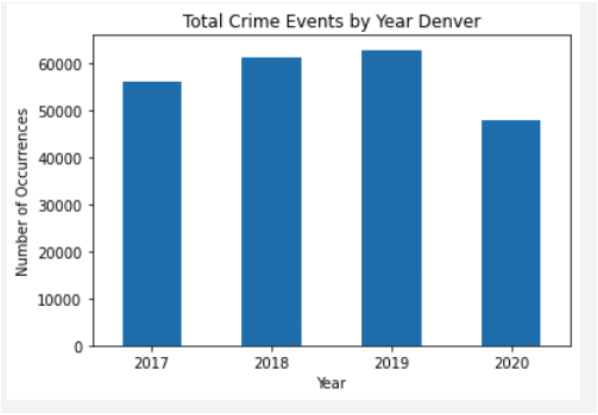
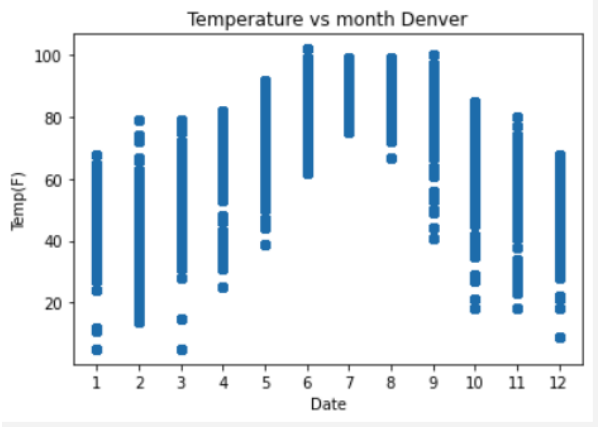
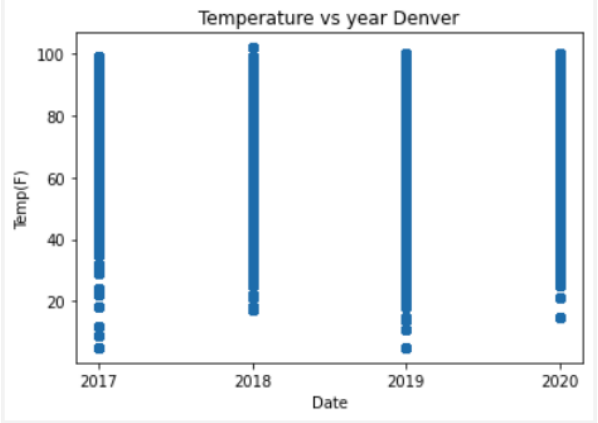
Denver Crime	**Hot**	**Mild**	**Cold**
People	13.509	12.663	13.086
Property	47.459	45.949	45.872
Society	39.032	41.387	41.042
Total	100.00	100.00	100.00

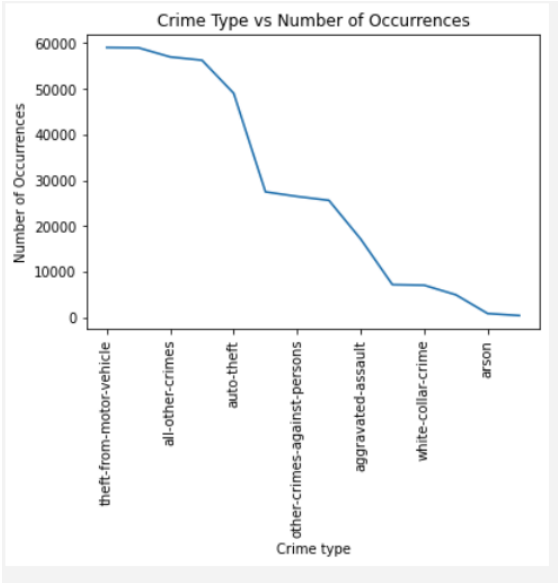
Percentage of Temp distribution:
Cold -20.68
Mild -46.299
Hot - 33.012
Percentage of crime distribution:
Crime on Property - 64.5
Crime on Person -11.81
Crime on Society -23.62

Denver Crime	**Hot**	**Mild**	**Cold**	**Total**
Property	21.598	29.742	13.223	64.563
People	3.787	5.539	2.487	11.813
Society	7.626	11.017	4.975	23.618
Total	33.011	46.298	20.685	99.994

Bayesian Classification Max Temp

Denver	**Hot**	**Mild**	**Cold**
Property	65.423	63.923	64.239
People	11.473	12.023	11.963
Society	23.102	24.052	23.796
Total	99.998	99.998	99.998

Crime by Month (Denver)	Crime by Year (Denver)
	
Max temp by Month	**Max temp by year**
	
General Lookover	




```
Offense Types in Denver and Occurrence count
theft-from-motor-vehicle        59062
public-disorder                 58994
all-other-crimes                56996
larceny                         56282
auto-theft                      49040
burglary                        27482
other-crimes-against-persons    26470
drug-alcohol                    25619
aggravated-assault              17163
robbery                          7184
white-collar-crime               7061
sexual-assault                   4956
arson                             863
murder                            417
```

Boston - Brady

Data:

Dataframe	Dates	Entries
Crime	6/15/2015 - 10/3/2018	327,820
Weather	1/1/2008 - 4/1/2018	3,749
Overlap	6/15/2015 - 4/1/2018	277,545

Crime Type Distribution:

Crime Rates (Boston)	Hot	Mild	Cold	Total
Property	**5.12%**	**34.65%**	**4.29%**	**44.06%**
People	**3.77%**	**27.33%**	**3.52%**	**34.63%**
Society	**1.23%**	**8.68%**	**1.21%**	**11.11%**
Total	**10.12%**	**70.65%**	**9.02%**	**~89.8%**

Crime Type Distribution (High Temp (F)):

Crime Rates (Boston)(High Temp (F))	Hot	Mild	Cold	Total
Property	**8.34%**	**23.02%**	**12.7%**	**44.06%**
People	**6.16%**	**17.91%**	**10.55%**	**34.62%**
Society	**2.0%**	**5.75%**	**3.36%**	**11.11%**
Total	**16.5%**	**46.68%**	**26.61%**	**~89.79%**

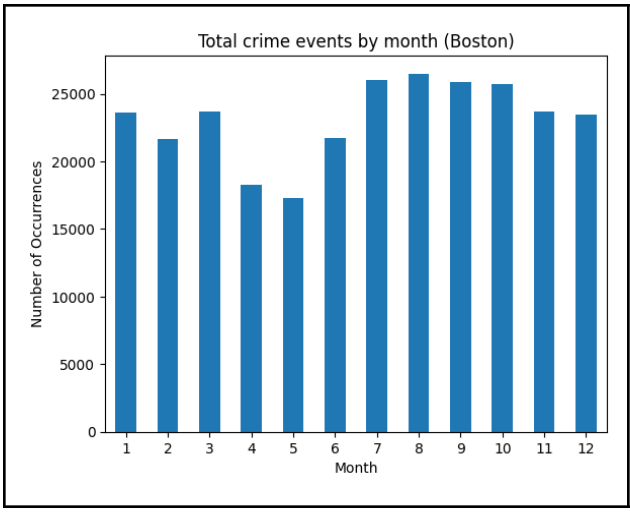
Bayesian:

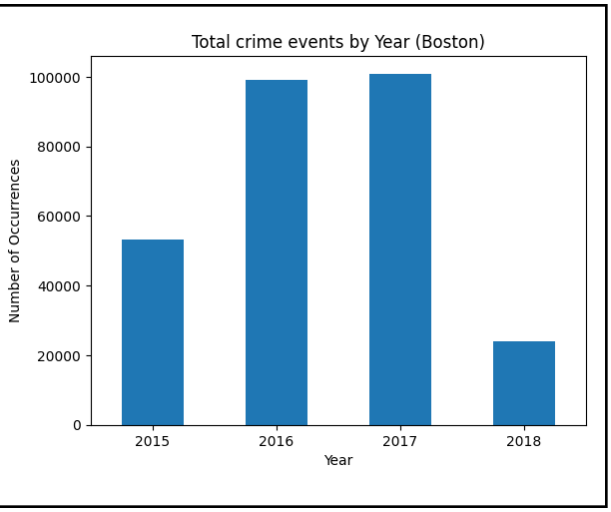
Bayesian (Boston)	Hot	Mild	Cold
Property	**50.6**	**49.5**	**47.77**
People	**37.27**	**38.26**	**39.55**
Society	**12.12**	**12.24**	**12.69**
Total	**~99.99%**	**~100%**	**~100.01%**

Bayesian (High Temp (F)):

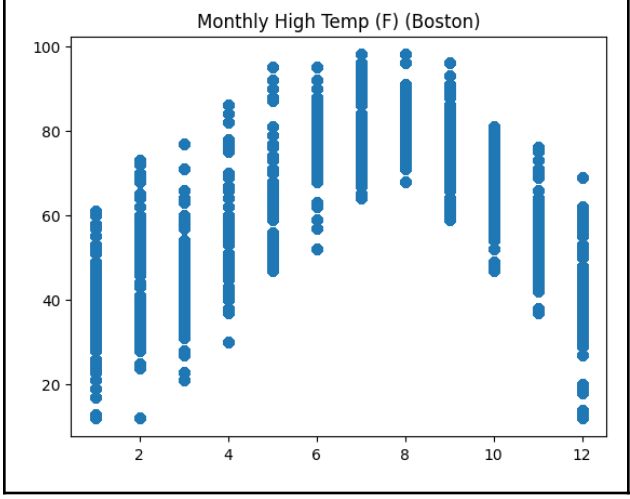
Boston	Hot	Mild	Cold
Property	**50.54**	**49.31**	**47.73**
People	**37.33**	**38.37**	**39.66**
Society	**12.13**	**12.32**	**12.61**
Total	**100%**	**100%**	**100%**

Crime per Month (Boston)	Crime per Year (Boston)

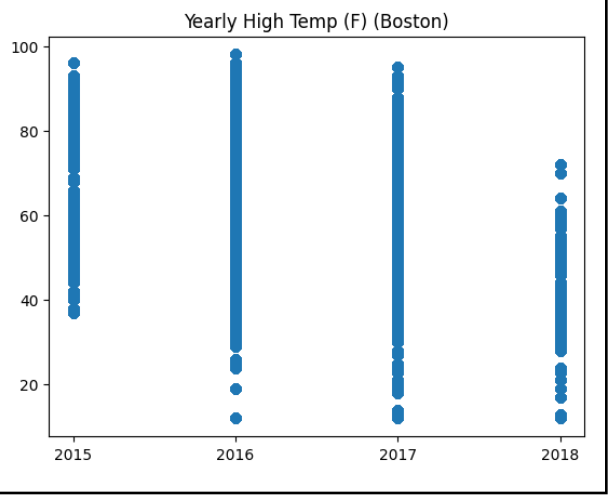




High Temp (F) Monthly Boston



High Temp (F) Yearly Boston



DESCRIPTION COLUMN:

Property: 140,120 (~43%)

"Theft" 27740
"Vandalism" 15542
"Larceny" 37790
"Burglary" 7542
"Property" 51506

Person: 114,415 (~35%)

"Assault" 24476
"Manslaughter" 173
"Warrant" 9560

“Medical” 20798
“Person” 55145
“Harassment” 4263

Society: 36,314 (~11%)
“Violation” 5996
“Threat” 9425
“Other” 20893 (here for number purposes)

Missing: 36,971 (~1%)

CODE_GROUP COLUMN:
Property: 95,180 (~29%)
“Theft” 6036
“Vandalism” 15810
“Larceny” 37790
“Burglary” 7551
“Property” 27993

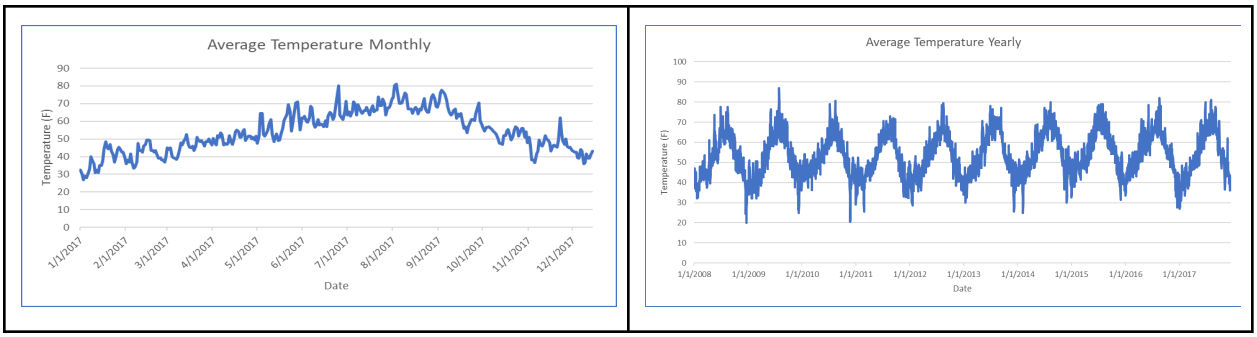
Person: 90,514 (~28%)
“Assault” 24296
“Manslaughter” 9
“Warrant” 9560 (match)
“Medical” 24226
“Person” 28160
“Harassment” 4263 (match)

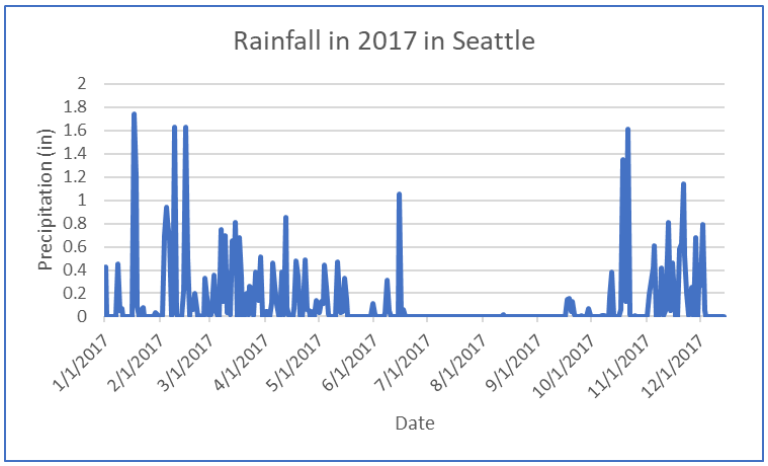
Society: 49,640 (~15%)
“Violation” 30553
“Threat” 2
“Other” 19085 (here for number purposes)

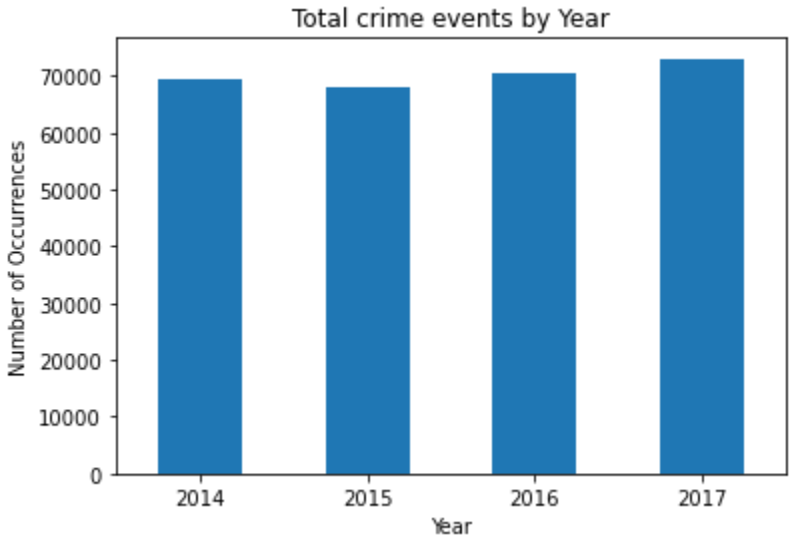
Missing: 92,486 (~28%)

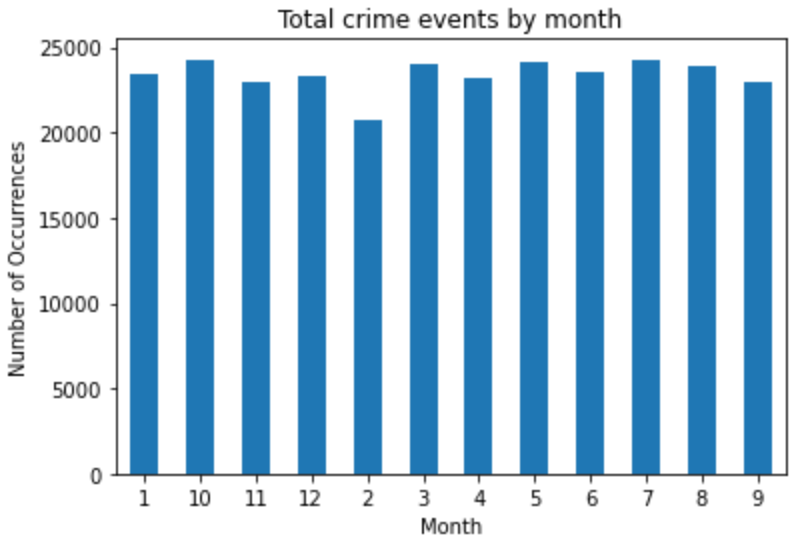
Seattle - Dylan
Crime - (2008 - Present)
Weather - (1947 - Dec 13 2017)
2008 - Dec 13 2017

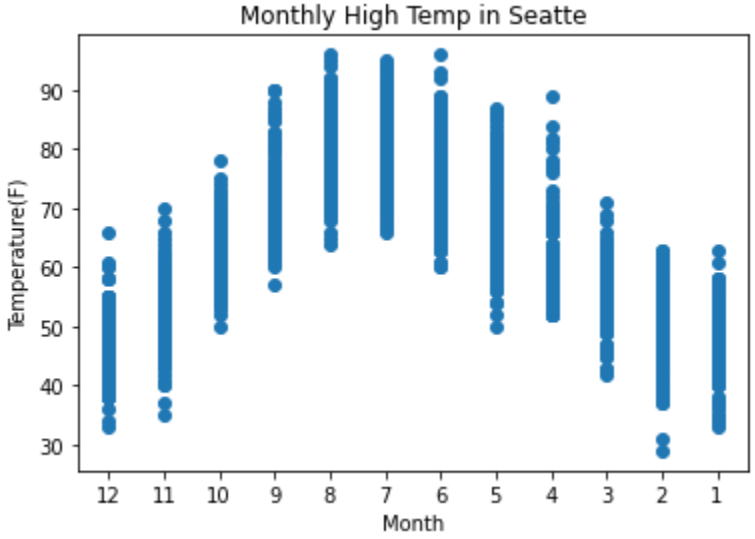
Avg Temp (F) Monthly (Seattle)	Avg Temp (F) Yearly (Seattle)

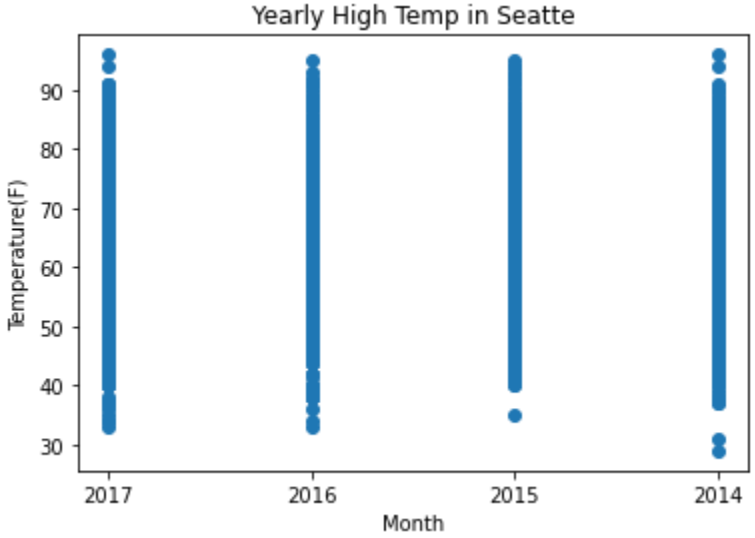

Average Temperature Monthly
Temperature (F)
Date
Average Temperature Yearly
Temperature (F)
Date


Rainfall in 2017 in Seattle
Precipitation (in)
Date


Total crime events by Year
Number of Occurrences
Year


Total crime events by month
Number of Occurrences
25000
20000
15000
10000
5000
0
1
10
11
12
2
3
4
5
6
7
8
9
Month


Monthly High Temp in Seatte
Temperature(F)
90
80
70
60
50
40
30
12
11
10
9
8
7
6
5
4
3
2
1
Month


Yearly High Temp in Seatte
Temperature(F)
90
80
70
60
50
40
30
2017
2016
2015
2014
Month

Crime Type Distribution:

Crime Rates (Seattle)	Hot	Mild	Cold	Total
Property	1.8%	53.29%	14.94%	70.03%
People	0.48%	12.56%	3.34%	16.38%
Society	0.32%	10.18%	3.02%	13.52%
Total	2.6%	76.03%	21.30%	100%

Bayesian:

Bayesian (Seattle)	Hot	Mild	Cold
Property	70.2%	71.1%	71.5%
People	17.2%	15.7%	15.3%
Society	12.6%	13.2%	13.1%
Total	100%	100%	100%