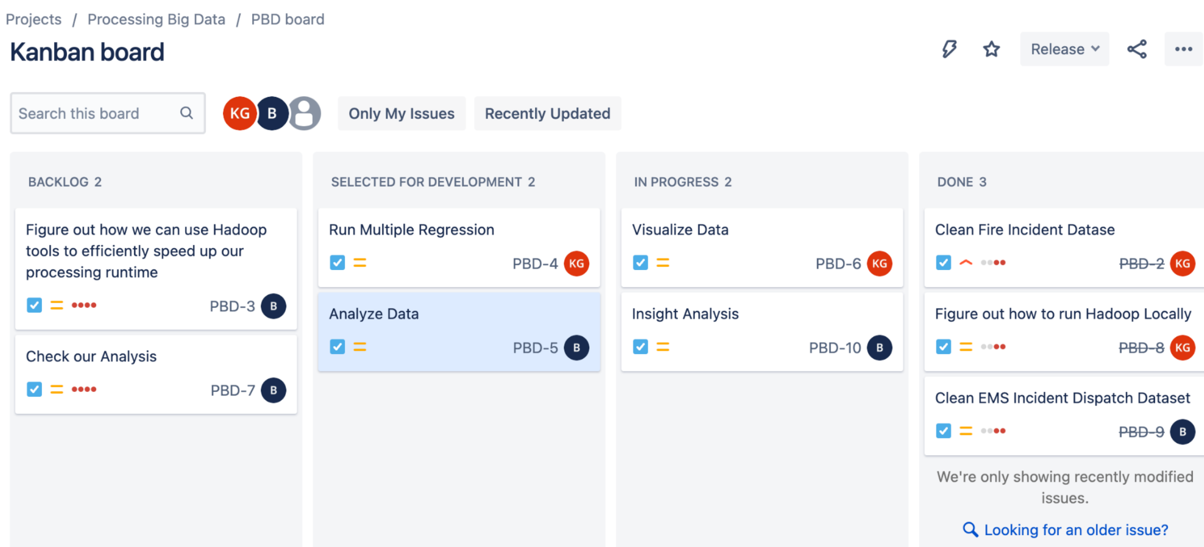
Jira Board



Code Drop

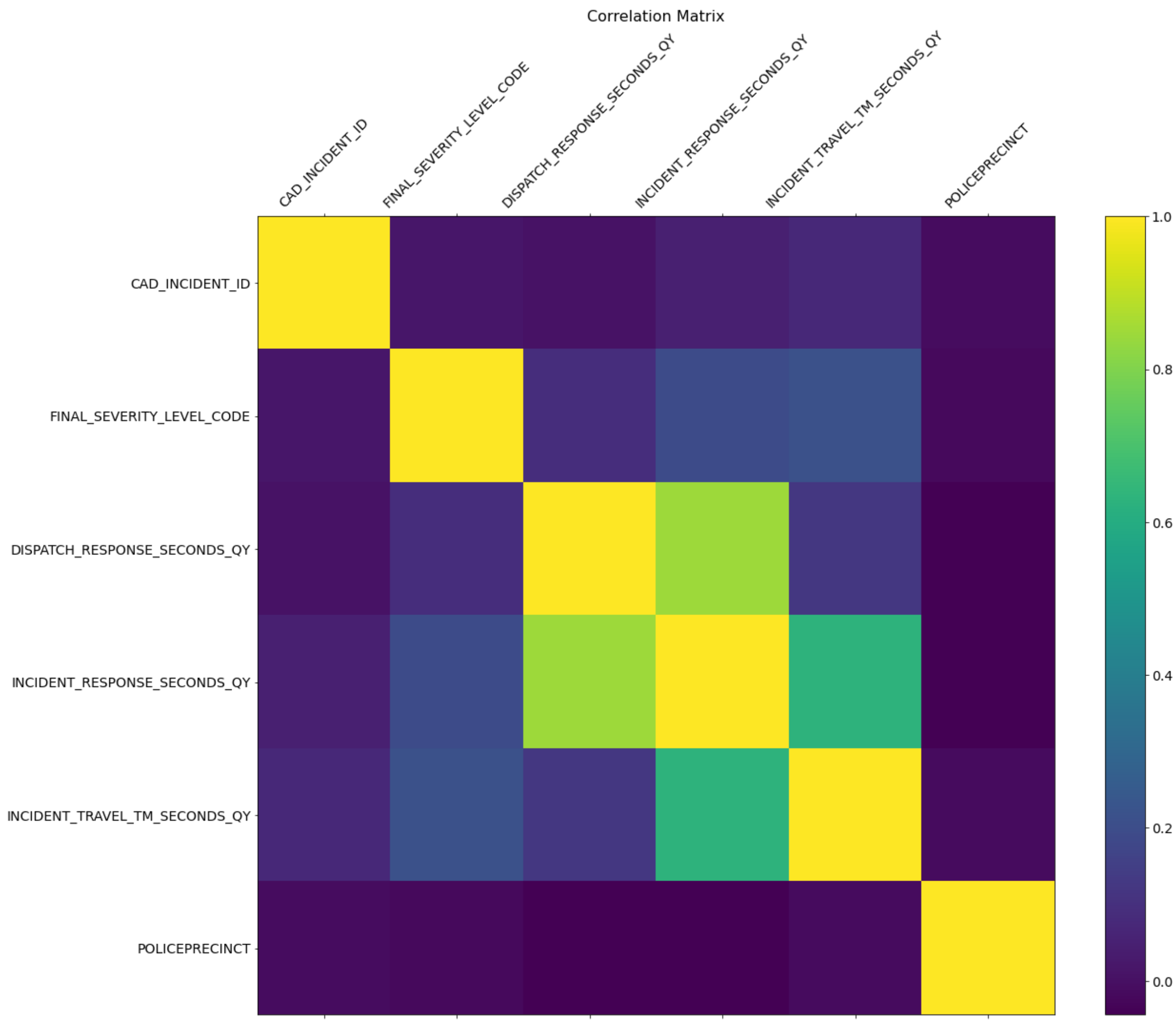
Apart from our MapReduce programs (cleaning and counting bad records), we have extended our code for visualizations and analysis.

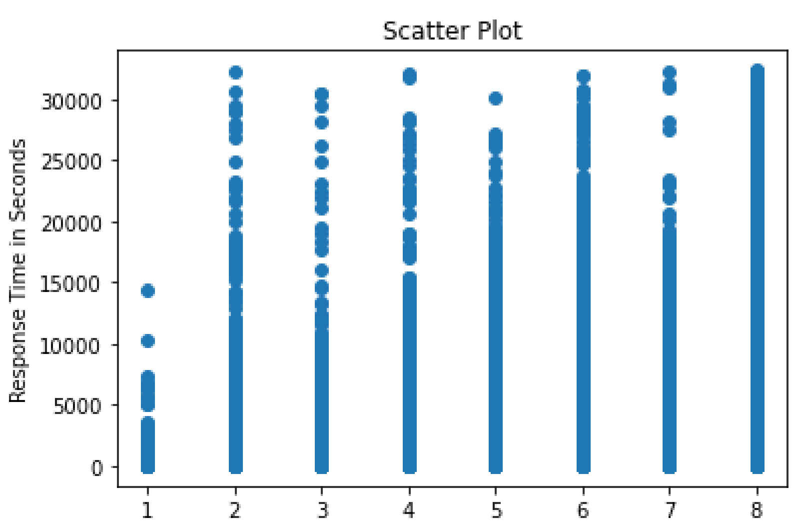
```
import numpy as np
import pandas as pd
import random

data = pd.read_csv('EMS_Incident_Dispatch_Data.csv')
df = pd.DataFrame(data)

sample_df = df.sample(frac = 0.2)
cleaned =
sample_df[['CAD_INCIDENT_ID','FINAL_CALL_TYPE','FINAL_SEVERITY_LEVEL_CODE','DIS
PATCH_RESPONSE_SECONDS_QY','INCIDENT_RESPONSE_SECONDS_QY','INCIDENT_T
RAVEL_TM_SECONDS_QY','BOROUGH','ZIPCODE','POLICEPRECINCT']]
cleaned = cleaned.dropna()

# Figure 1
import matplotlib.pyplot as plt

f = plt.figure(figsize=(19, 15))
plt.matshow(cleaned.corr(), fignum=f.number)
plt.xticks(range(cleaned.select_dtypes(['number']).shape[1]),
cleaned.select_dtypes(['number']).columns, fontsize=14, rotation=45)
plt.yticks(range(cleaned.select_dtypes(['number']).shape[1]),
cleaned.select_dtypes(['number']).columns, fontsize=14)
cb = plt.colorbar()
```

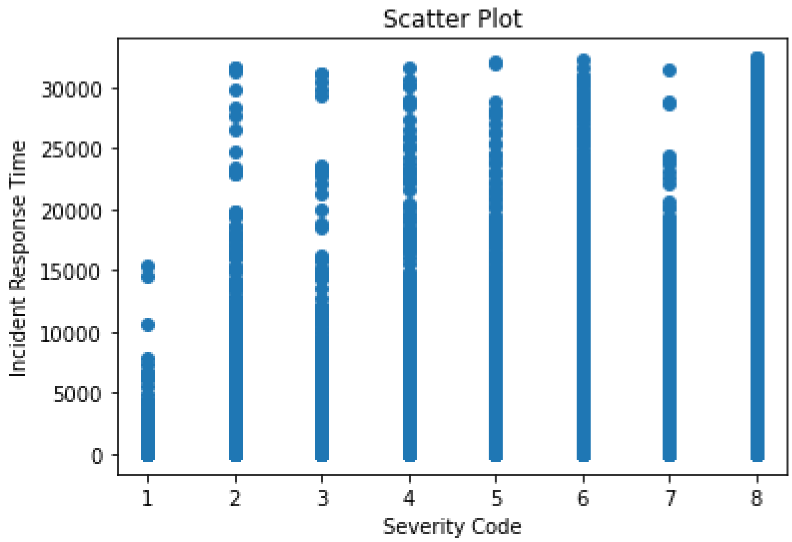
```
cb.ax.tick_params(labelsize=14)
plt.title('Correlation Matrix', fontsize=16);

response_time =
cleaned.groupby(by=['FINAL_CALL_TYPE','FINAL_SEVERITY_LEVEL_CODE']).mean().reset_i
ndex()

# Figure 2
# Scatter plot with day against tip
plt.scatter(cleaned['FINAL_SEVERITY_LEVEL_CODE'],
cleaned['DISPATCH_RESPONSE_SECONDS_QY'])

# Adding Title to the Plot
plt.title("Scatter Plot")

# Setting the X and Y labels
plt.xlabel('Severity Code')
plt.ylabel('Dispatch Response Time')

plt.show()

# Figure 3
plt.scatter(cleaned['FINAL_SEVERITY_LEVEL_CODE'],
cleaned['INCIDENT_RESPONSE_SECONDS_QY'])

# Adding Title to the Plot
plt.title("Scatter Plot")

# Setting the X and Y labels
plt.xlabel('Severity Code')
plt.ylabel('Incident Response Time')

plt.show()

# Figure 4
import seaborn as sns

sns.scatterplot(x='DISPATCH_RESPONSE_SECONDS_QY',
y='INCIDENT_RESPONSE_SECONDS_QY', data=cleaned,
          hue='FINAL_SEVERITY_LEVEL_CODE')
plt.show()

# Figure 5
```

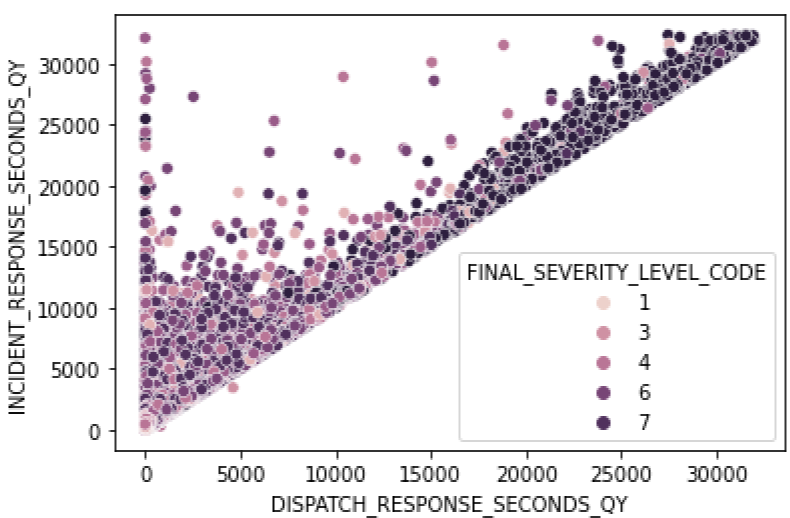
```
sns.scatterplot(x='POLICEPRECINCT', y='INCIDENT_RESPONSE_SECONDS_QY',
data=cleaned,
            hue='FINAL_SEVERITY_LEVEL_CODE')
plt.show()

boroughs = cleaned.groupby(by=['BOROUGH']).mean().reset_index()
boroughs_final =
boroughs[['BOROUGH','DISPATCH_RESPONSE_SECONDS_QY','INCIDENT_RESPONSE_S
ECONDS_QY','INCIDENT_TRAVEL_TM_SECONDS_QY']]

# Figure 6
plt.scatter(cleaned['FINAL_SEVERITY_LEVEL_CODE'],
cleaned['INCIDENT_RESPONSE_SECONDS_QY'], color='red')
plt.title('Severity Code Vs Incident Response Time', fontsize=14)
plt.xlabel('Severity Rate', fontsize=14)
plt.ylabel('Incident Response Time Price', fontsize=14)
plt.grid(True)
plt.show()

# Figure 7
plt.scatter(cleaned['DISPATCH_RESPONSE_SECONDS_QY'],
cleaned['INCIDENT_RESPONSE_SECONDS_QY'], color='red')
plt.title('Dispatch Response Time Vs Incident Response Time', fontsize=14)
plt.xlabel('Dispatch Response Time', fontsize=14)
plt.ylabel('Incident Response Time Price', fontsize=14)
plt.grid(True)
plt.show()

# OLS
from sklearn import linear_model
import statsmodels.api as sm

X = cleaned[['FINAL_SEVERITY_LEVEL_CODE','DISPATCH_RESPONSE_SECONDS_QY']]
y = cleaned['INCIDENT_RESPONSE_SECONDS_QY']

regr = linear_model.LinearRegression()
regr.fit(X, y)

print('Intercept: \n', regr.intercept_)
print('Coefficients: \n', regr.coef_)

# with statsmodels
X = sm.add_constant(X) # adding a constant
```

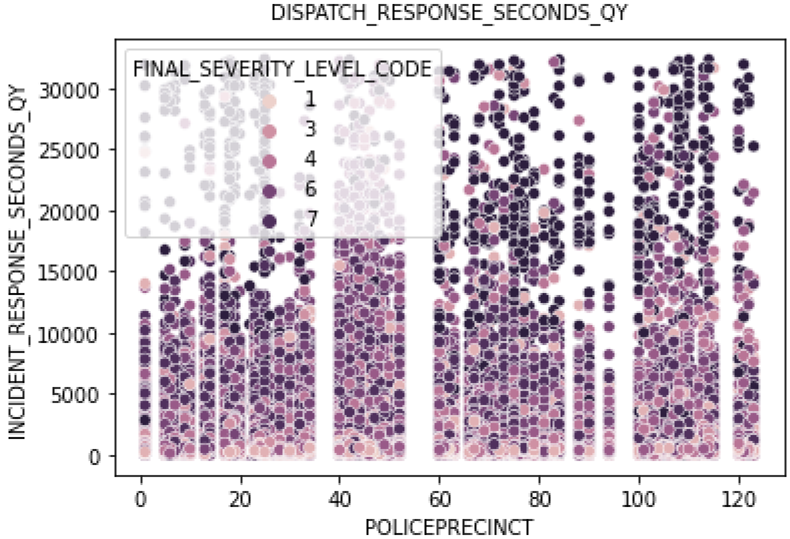
```
model = sm.OLS(y, X).fit()
predictions = model.predict(X)

print_model = model.summary()
print(print_model)

from sklearn.model_selection import train_test_split
from sklearn.metrics import r2_score
from sklearn.metrics import mean_squared_error

X = cleaned[['FINAL_SEVERITY_LEVEL_CODE','DISPATCH_RESPONSE_SECONDS_QY']]
y = cleaned['INCIDENT_RESPONSE_SECONDS_QY']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.3, random_state = 42)

regr = linear_model.LinearRegression()
regr.fit(X, y)

y_prediction = regr.predict(X_test)

score=r2_score(y_test,y_prediction)
print('R^2 score: ',score)
print('MSE: ', mean_squared_error(y_test,y_prediction))
print('RMSE: ', np.sqrt(mean_squared_error(y_test,y_prediction)))
```


Correlation Matrix
CAD_INCIDENT_ID
FINAL_SEVERITY_LEVEL_CODE
DISPATCH_RESPONSE_SECONDS_QY
INCIDENT_RESPONSE_SECONDS_QY
INCIDENT_TRAVEL_TM_SECONDS_QY
POLICEPRECINCT
1.0
0.8
0.6
0.4
0.2
0.0

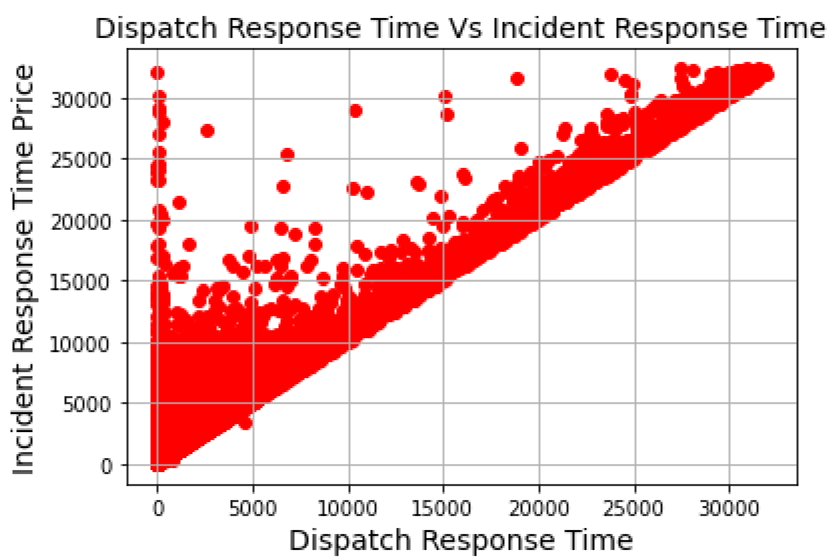

Scatter Plot
Response Time in Seconds
30000
25000
20000
15000
10000
5000
0
1
2
3
4
5
6
7
8


Scatter Plot
Incident Response Time
Severity Code
0
5000
10000
15000
20000
25000
30000
1
2
3
4
5
6
7
8


INCIDENT_RESPONSE_SECONDS_QY
DISPATCH_RESPONSE_SECONDS_QY
FINAL_SEVERITY_LEVEL_CODE
1
3
4
6
7
0
5000
10000
15000
20000
25000
30000


INCIDENT_RESPONSE_SECONDS_QY
POLICEPRECINCT
FINAL_SEVERITY_LEVEL_CODE
1
3
4
6
7
0
5000
10000
15000
20000
25000
30000
0
20
40
60
80
100
120


Severity Code Vs Incident Response Time
Incident Response Time Price
Severity Rate
0
5000
10000
15000
20000
25000
30000
1
2
3
4
5
6
7
8


Dispatch Response Time Vs Incident Response Time
Incident Response Time Price
Dispatch Response Time
0
5000
10000
15000
20000
25000
30000

```
Intercept:
 271.1995982854357
Coefficients:
 [43.49949337  1.07161985]
                                 OLS Regression Results                                
=======================================================================================
Dep. Variable:     INCIDENT_RESPONSE_SECONDS_QY   R-squared:                       0.720
Model:                                      OLS   Adj. R-squared:                  0.720
Method:                           Least Squares   F-statistic:                 5.615e+06
Date:                          Wed, 16 Nov 2022   Prob (F-statistic):               0.00
Time:                                  18:23:39   Log-Likelihood:          -3.1562e+07
No. Observations:                       4376996   AIC:                         6.312e+07
Df Residuals:                           4376993   BIC:                         6.312e+07
Df Model:                                     2                                         
Covariance Type:                      nonrobust                                         
===============================================================================================
                                  coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------------------
const                         271.1996      0.417    650.328      0.000     270.382     272.017
FINAL_SEVERITY_LEVEL_CODE      43.4995      0.094    462.990      0.000      43.315      43.684
DISPATCH_RESPONSE_SECONDS_QY    1.0716      0.000   3260.675      0.000       1.071       1.072
==============================================================================
Omnibus:                  5937314.162   Durbin-Watson:                   1.999
Prob(Omnibus):                  0.000   Jarque-Bera (JB):      10585682659.985
Skew:                           7.057   Prob(JB):                         0.00
Kurtosis:                     243.509   Cond. No.                     1.32e+03
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.32e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
```

```
R^2 score:  0.7203704041380895
MSE:  108600.12941593894
RMSE:  329.5453374210276
```