Name: Prayas Dixit
USC ID: 9986683071

DSCI 510 FINAL PROJECT DESCRIPTION

DESCRIPTION: For this project, I have created three datasets from three different data sources. The names of the files are 'Covid_Data.csv', 'crime_data.csv' and 'Unemp_Data.csv'. These datasets are being created when we run our python script in default mode. I have also put these three files in a folder name 'Datasets'. Using these three data sets I am doing the Exploratory Data Analysis and figuring out the relationship between covid cases, crime count per month and the Unemployment rate in LA.

MOTIVATION: I have read articles that lot of people lost their job during covid. So, here I am attempting to analyze the rate at which it increased or decreased during various phases of Covid Waves. Along with that I am also analyzing the link between Unemployment rate and Crime Rate in LA if there is any. After doing the above analysis I would like to present the actual truth whether people lost so many jobs continuously or not.

DATA SOURCES:

1. LA city Covid cases from March 2020 to Present:
(https://data.lacity.org/COVID-19/LA-County-COVID-Cases/jsff-uc6b/data)
Link to fetch json: 'https://data.lacity.org/resource/jsff-uc6b.json'

Using the request library, I am accessing the data and converting it into the dataframe by using pandas library. It consists of around 750 rows, and it is being updated on daily basis. It has various columns out of which I took the date, New_Cases, New_Deaths. After extracting the data, I have converted it to month level and added two new columns Month and Year (extracting from date column) so that it helps our further analysis. Further, I dropped the initial 'date' column and finally created 'Covid_Data.csv' file.

crime_data

Month	Year	crime_count
1	2020	18301
2	2020	17187
3	2020	16058
4	2020	15621
5	2020	17131

2. Los Angeles Unemployment Rate from 2018 to present: (https://ycharts.com/indicators/los_angeles_ca_unemployment_rate)

For this part I am scraping the datasets from the mentioned link and saving the dataset into a csv file i.e 'Unemp_Data.csv'. This dataset consists of two columns initially which are date and Unemployment Rate percentage. After extracting it I added two more column i.e Month and Year so that all our three datasets have 2 common column. Once I got the month and year column, I dropped the date column because it was unnecessary. So, the final dataset consist of 3 columns Month, Year and Unemployment Rate(In percentage). I have also converted the Unemployment rate to float type so it will be easy for me to do the future analysis and plotting.

Unemp_Data

Unemp_Rate_(In %)	Month	Year
4.7	2	2018
4.5	3	2018
4.3	4	2018
4.2	5	2018
4.8	6	2018

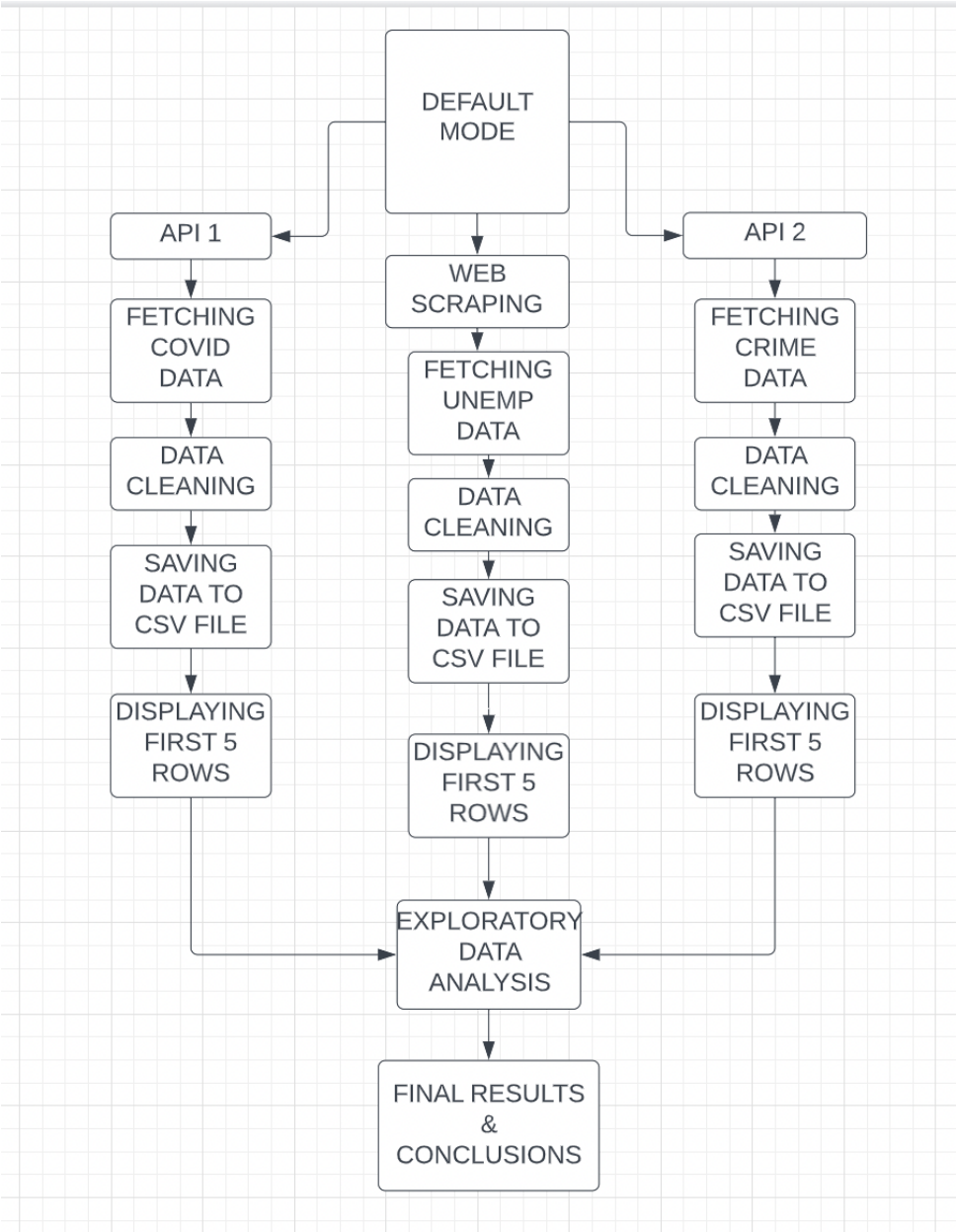
3. Crime Data from 2020 to present:
(https://data.lacity.org/Public-Safety/Crime-Data-from-2020-to-Present/2nrs-mtv8/data)
Link to fetch json: https://data.lacity.org/resource/2nrs-mtv8.json?$limit=700000

Finally for the last data, I am using the request library and accessing the data, converting it into the dataframe by using pandas library. It consists of around 50000 rows. I only wanted to take the number of crimes happening on monthly level in LA city. Therefore, I only took the 'date' data and grouped it by month and counted the number of rows which gave me the number of crimes happening in LA on monthly basis from January 2020. After that I cleaned the data according to the requirements and added two more columns i.e. Month and Year to the final dataset. I have created 'crime_data.csv' file by using this API. Regarding maintainability, this code will only take the 7,00,000 rows (because I have set 7,00,000 as limit while extracting data from API), right now the dataset has around 5,00,000 rows.

crime_data

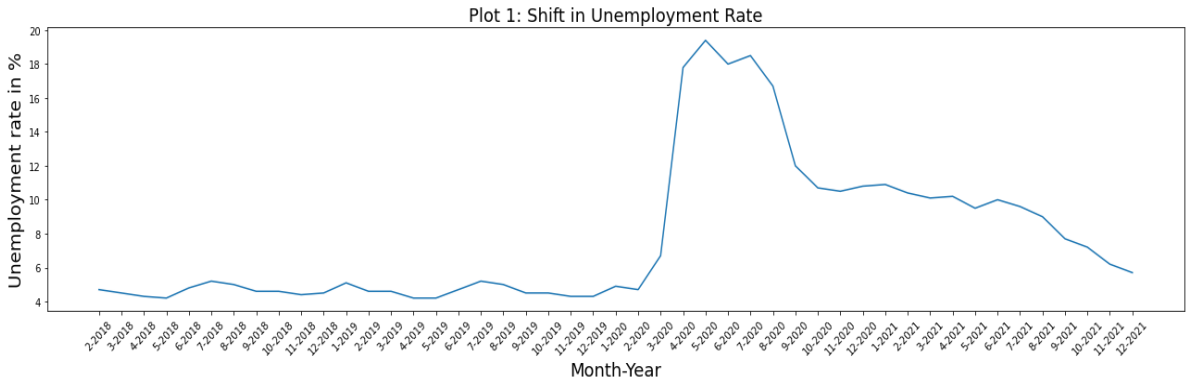
Month	Year	crime_count
1	2020	18301
2	2020	17187
3	2020	16058
4	2020	15621
5	2020	17131

Flowchart:



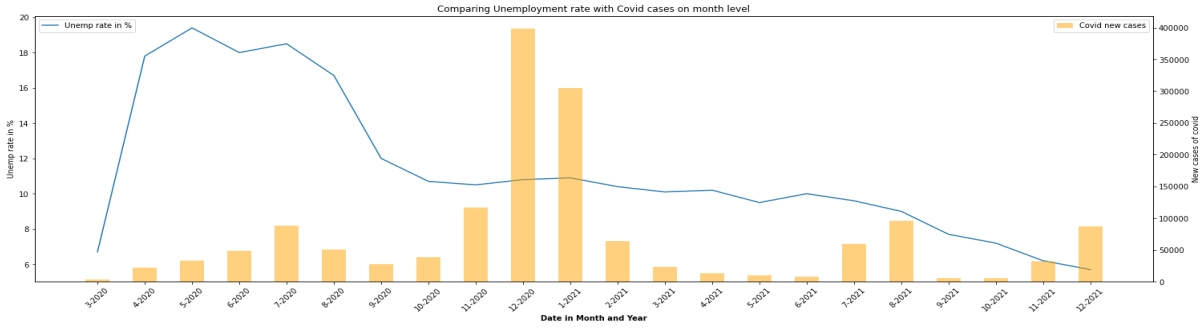
ANALYSIS PERFORMED:

Plot 1: Shift in Unemployment Rate



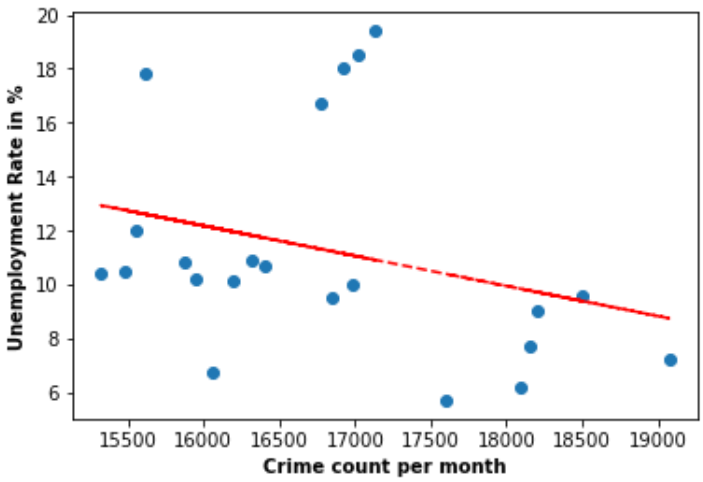
Here, I am using the Second data i.e. Unemp_Data and made a new column by extracting the month and year all together so I can represent dates on the x axis. This plot depicts the trend in Unemployment rate. Here, we can easily see how the unemployment rate spiked up when the first covid wave arrived in march 2020. After the unemployment hits its peak during the period(May 2020- Aug 2020) then it started to decrease at steady rate and now it is almost near to the same state as pre-covid times.

Plot 2: Covid Cases v/s Unemployment Rate



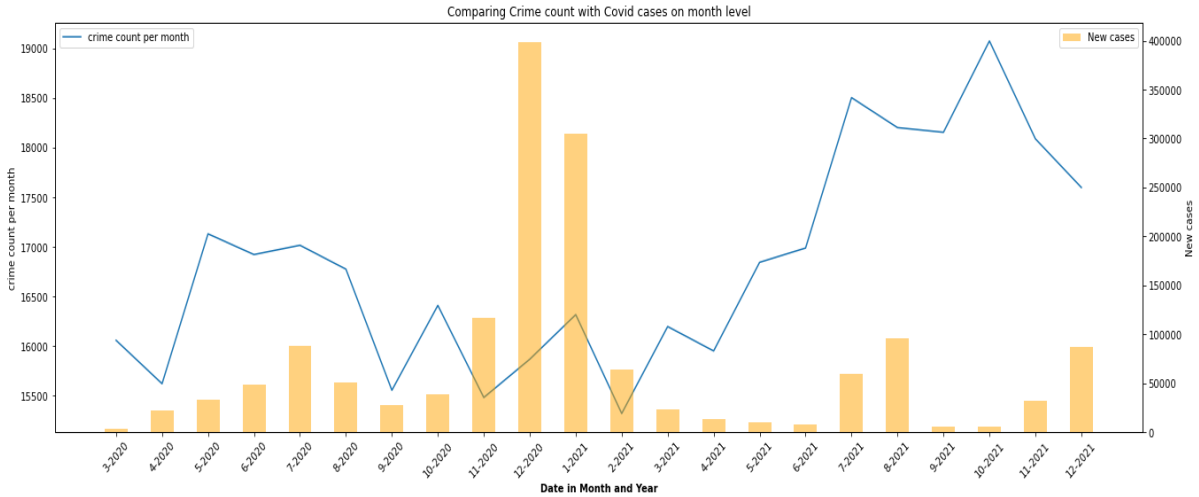
Here, I am using two datasets Covid_Data and Unemp_Data and made a single dataset so that we can compare Unemployment rate and covid cases in single graph and visualize effectively. From the plot we can easily analyze that with the increase in covid cases in March 2020 the unemployment rate got spiked up and from September 2020 it came back in control and even in the second wave that arrived in November 2020 the unemployment rate didn't spike that much due to the effective management and awareness of Covid.

Plot 3: Scatter plot between Unemployment rate and Crime count



Here, we can see that there is some correlation between Unemployment rate and Crime count per month i.e. As we have already depicted above that covid cases are managed quite well and due to that unemployment rate is decreasing. Furthermore, with less restrictions being placed for covid people started going out in late 2021 and that caused the increase in crime rate per month.

Plot 4: Plots between crime count and new covid cases



In plot 4, we can see that with the decrease in new covid cases in early 2021 the crime counts increases which can be observed as more people started going out due to less restrictions.

After this I found the Pearson's correlation coefficient between crime count and unemployment rate which was -.289 which implements there is slight negative correlation between these two variables. For example: If the Unemployment rate is Increasing then there is slight decrease in crime count.

Along with this I also found Pearson's correlation coefficient between crime count and new cases also which was -.227 which also indicates slight negative correlation.

Finally, In order to identify the significance of crime count to predict Unemployment rate, I am fitting a linear regression model with the help of statsmodel library. So, the p value is 0 as we can see in the output and when p value is less than .05 we can say that the crime count is statistically significant in predicting Unemployment rate.

```
                                 OLS Regression Results
=======================================================================================
Dep. Variable:      Unemp_Rate_(In %)   R-squared (uncentered):                   0.871
Model:                            OLS   Adj. R-squared (uncentered):              0.865
Method:                 Least Squares   F-statistic:                              141.4
Date:                Wed, 11 May 2022   Prob (F-statistic):                    8.62e-11
Time:                        13:20:34   Log-Likelihood:                         -63.312
No. Observations:                  22   AIC:                                      128.6
Df Residuals:                      21   BIC:                                      129.7
Df Model:                           1
Covariance Type:            nonrobust
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
crime_count     0.0007   5.57e-05     11.890      0.000       0.001       0.001
==============================================================================
Omnibus:                        1.859   Durbin-Watson:                   0.396
Prob(Omnibus):                  0.395   Jarque-Bera (JB):                1.550
Skew:                           0.520   Prob(JB):                        0.461
Kurtosis:                       2.218   Cond. No.                         1.00
==============================================================================

Notes:
[1] R² is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Conclusion:

Based on my analysis, Initially when the covid hit due to lots of uncertainty there was an effect on unemployment rate but later on it was managed really well by the government which led to decrees in unemployment rate irrespective of the further covid waves and by the end of 2021 it almost reached to pre covid level. So, people need not worry so much about their job security if the covid cases rises again. Also, this had one negative impact as the number of people started coming out of their houses increased there was an steady increase in crime rate in 2021.