



AICP Internship Task

Spread Pattern and Behaviour Analysis of Covid-19

The year 2020 was highlighted by the appearance of the virus SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2), which causes the coronavirus disease 2019 (COVID-19), the respiratory illness responsible for the COVID-19 pandemic. This disease was first identified when WHO (World Health Organization) was reported about a cluster of viral pneumonia of an unknown cause in Central Asia. It was recognized an outbreak in January 2020, afterwards in March 2020 was named as pandemic. It has been exported to more than 200 countries. It victimized more than 25 million people all over the world till 31st August 2020 along with 846,985 deaths. It demolished the world's best health systems and their techniques seem failed to slower it down its spread.

It is necessary to conduct a study to perform analysis publically available dataset on COVID-19 in Pakistan. This study generates different reports from Feb 26, 2020 – Sep 01, 2020 to analyze the trends and patterns of COVID-19 pandemic among the population of Pakistan.

In this study we have applied different machine learning techniques to analysis COVID19 spread all over Pakistan. We have applied data manipulating techniques using pandas in machine learning to analyses spread of COVID-19 in Pakistan. We took dataset provided by John Hopkin University. We picked data from Feb 26,2020 to Sep 01, 2020 and we came to know that Pakistan have conducted a total of 26,45,397 tests to diagnose the disease to the date September 01, 2020. Daily testing at labs were being done and Pakistan have witnessed a total of 295968 infected peoples the disease till the date September 01, 2020. Daily recoveries from the disease COVID-19 were very slow at the start of march. Pakistan have seen 281389 recoveries between march 10th 2020 and September 01st, 2020. Due to the challenges and less testing capacities the confirm death count may not be accurate count of COVID-19. The total deaths were 6297 on September 01, 2020 and daily confirmed deaths were 20.

Find dataset "COVID-19 Daily.xlsx"

Here's an overview of all the columns in the dataset:

- **Date**
- **Daily Test**
- **Daily Cases**
- **Daily Recoveries**
- **Daily Deaths**



Artificial Intelligence Community of Pakistan

--Evolution to Revolution--

Q: Import data and check null values, column info, and descriptive statistics of the data.

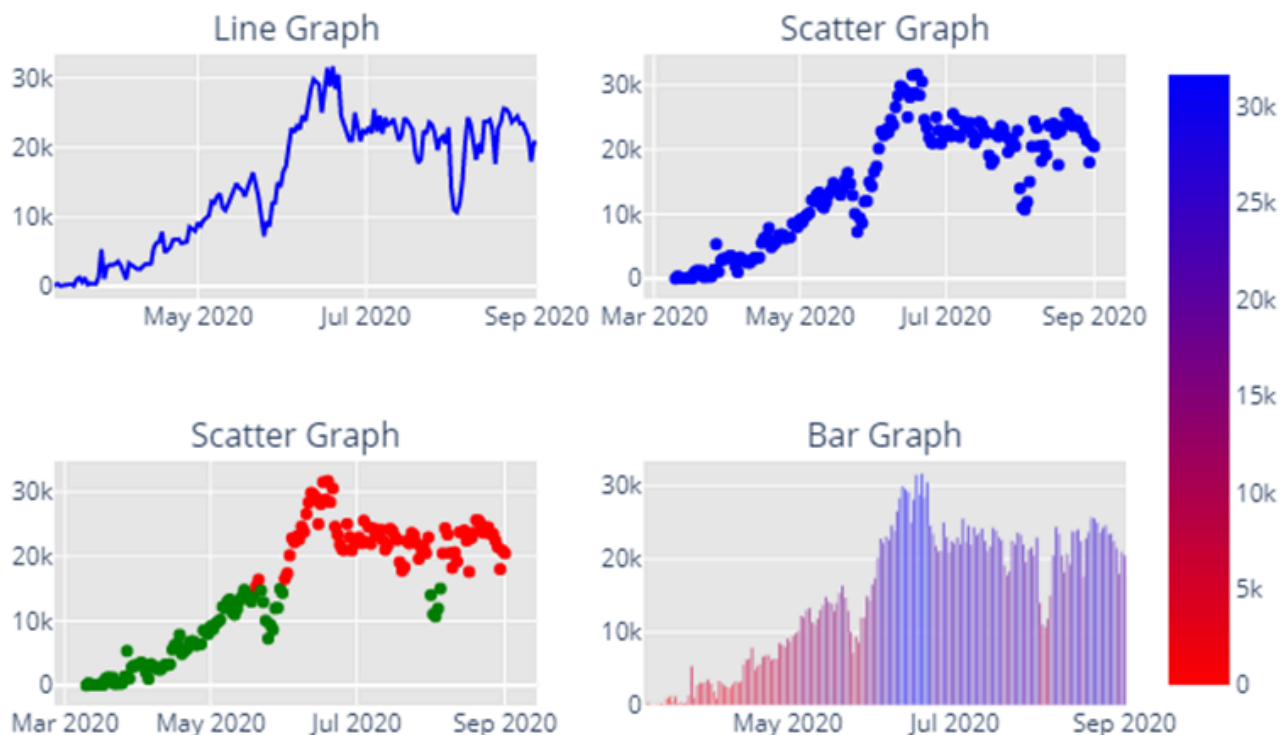
Q: Perform different type of analysis based on following diagrams on **Daily Tests, Cases, Recoveries, Death**. For sample we have mentioned tests below but implement same on others as well

Testing:

- How many total tests were performed till September 01, 2020?
- Perform Daily Testing Trend Line Graph, Scattered Diagram and Bar Graph.
- Show Scatter diagram of above-average (red color) and below-average (green color) daily testing.

Q: Convert the dataset into weekly & monthly stats and perform save operations as above performed on daily tests. Also find max and min number of weekly & monthly tests, cases, recoveries and deaths with dates.

Line Graph, Scattered Graph and Bar graph for New COVID19 Tests

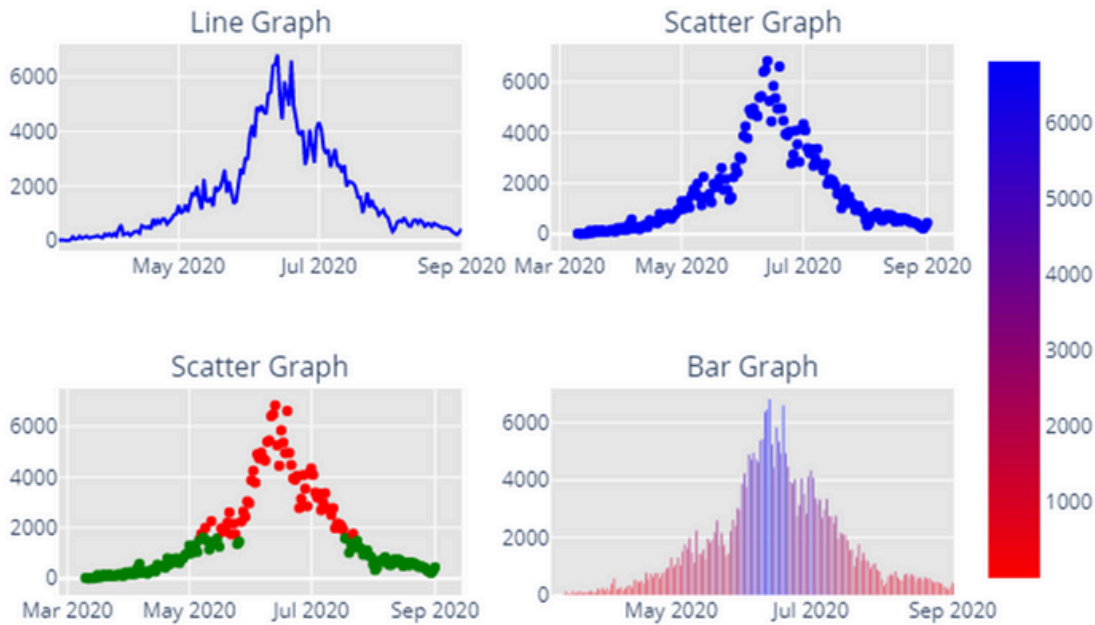




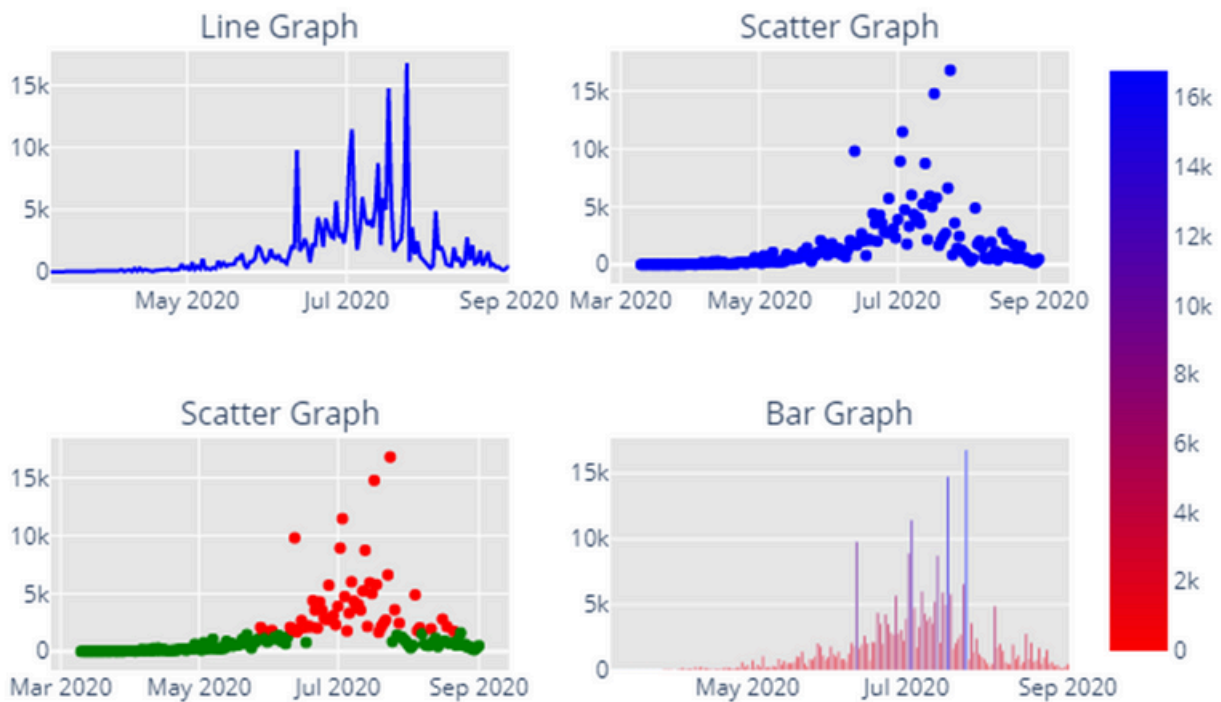
Artificial Intelligence Community of Pakistan

--Evolution to Revolution--

Line Graph, Scattered Graph and Bar graph for New COVID19 Cases



Line Graph, Scattered Graph and Bar graph for COVID19 Daily Recoveries





Artificial Intelligence Community of Pakistan

--Evolution to Revolution--

Q: Calculate **Per Thousand & Million** stats for Daily tests, cases, recoveries and deaths. Here is sample formula.

$$\text{Case Per Thousand} = \frac{\text{Total Cases}}{\text{Total Population}} \times 1000$$

$$\text{Case Per Million} = \frac{\text{Total Cases}}{\text{Total Population}} \times 1000000$$

$$\text{Positive Rate} = \frac{\text{Total Confirmed Cases}}{\text{Total Test Conducted}} \times 100$$

	Month	Tests	Cases	Cases Per Thousand	Cases Per Million	Positive Rate
0	March	23013	2038	0.010000	8.930000	8.860000
1	April	145317	14759	0.060000	64.650000	10.160000
2	May	376634	55541	0.240000	243.290000	14.750000
3	June	743969	140610	0.620000	615.930000	18.900000
4	July	690292	65676	0.290000	287.690000	9.510000
5	August	645692	16903	0.070000	74.040000	2.620000

Q: Train and implement ML model(any two) to predict the future trend of daily cases and deaths.

After implementing all above queries on national data, perform same for the different states mentioned in dataset excel sheet.