

COVID-19

Analysis

&

Forecasting

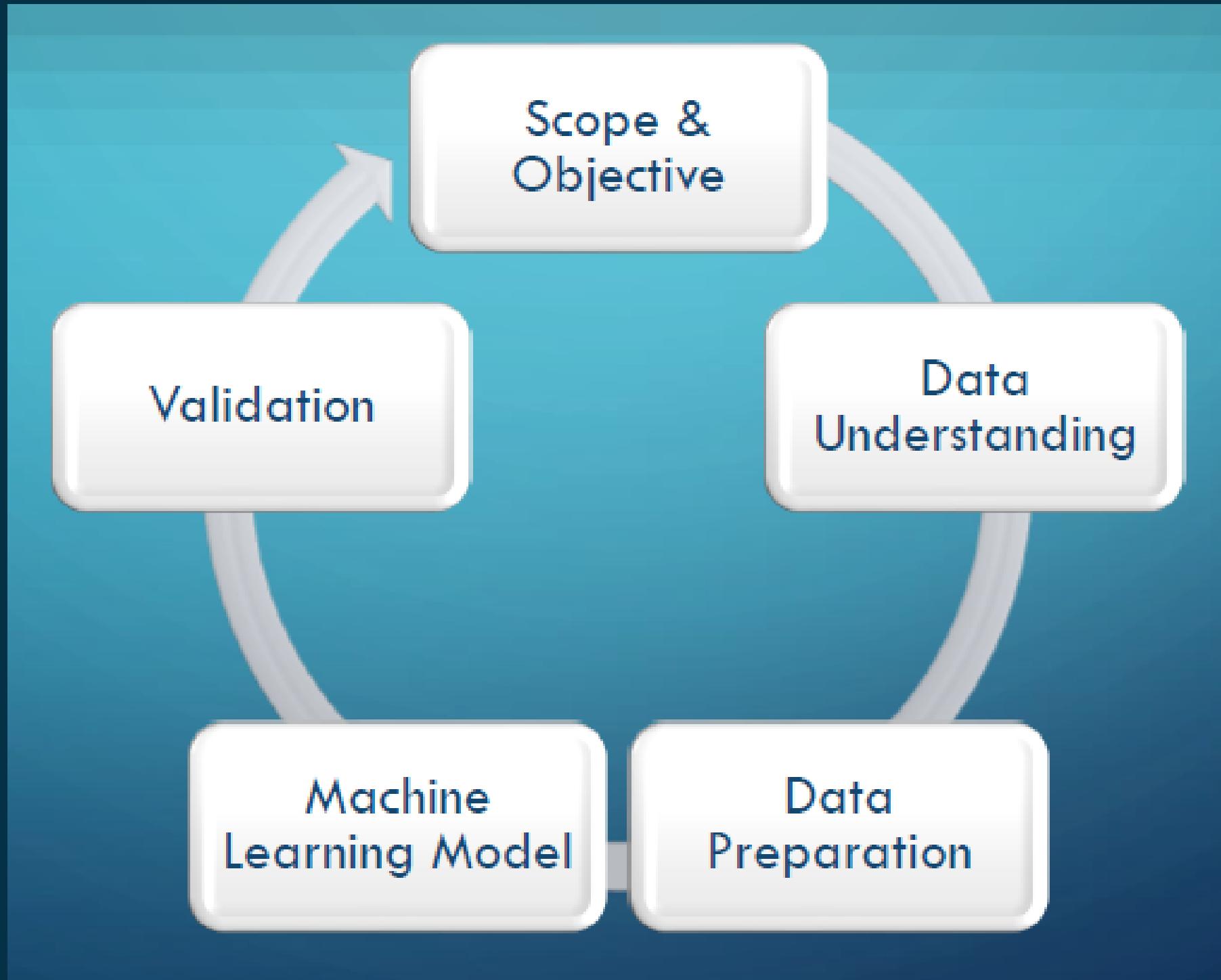
Presented By

- Amit Bagal
- Chetan Chaudhari
- Siddhesh Kanekar
- Prathamesh Ashwini
- Shubham Ugare

INTRODUCTION

- Importance of Data Analysis in Understanding COVID-19
- Data Visualization on COVID-19
- Future Prediction and Forecasting of COVID-19

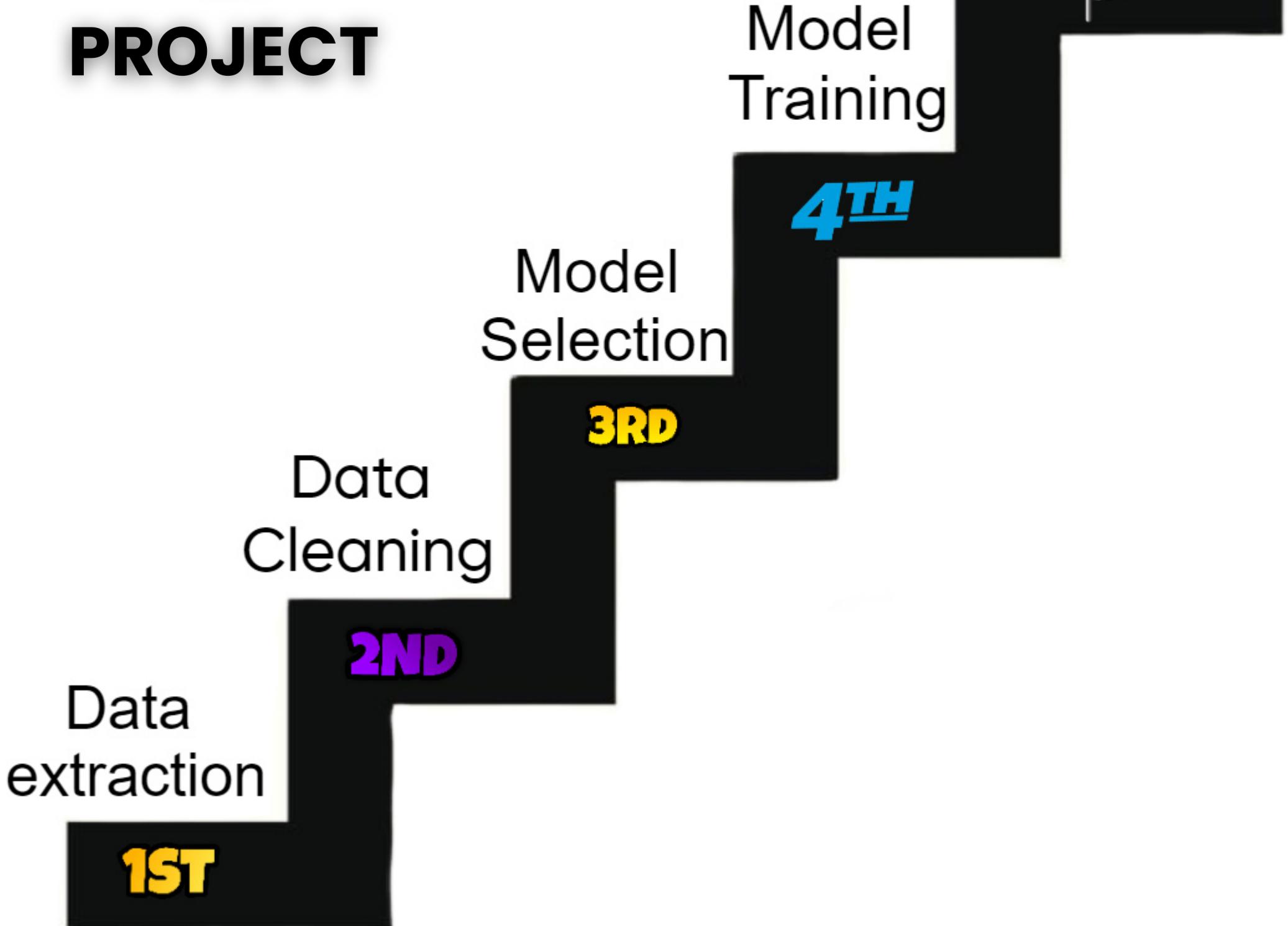
Implementation (Life Cycle)



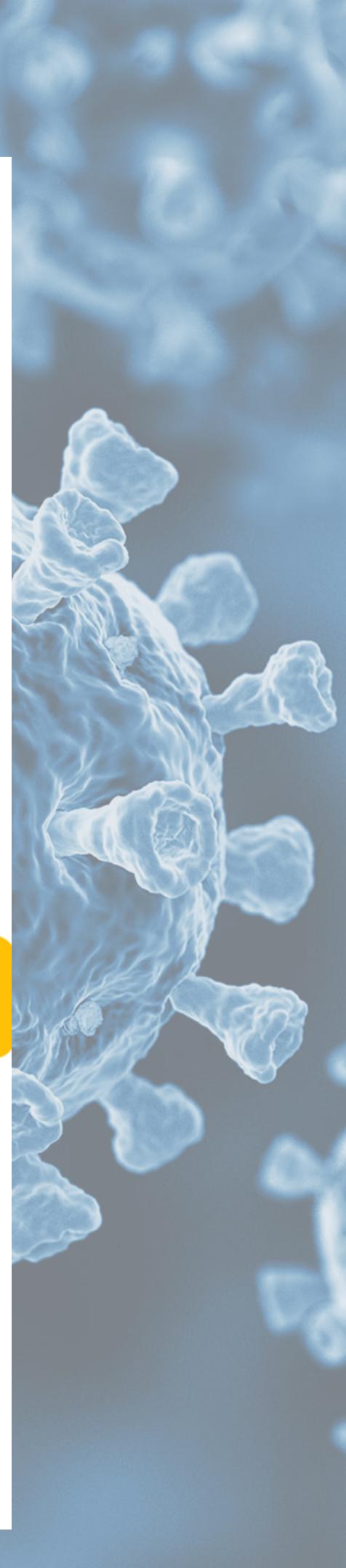
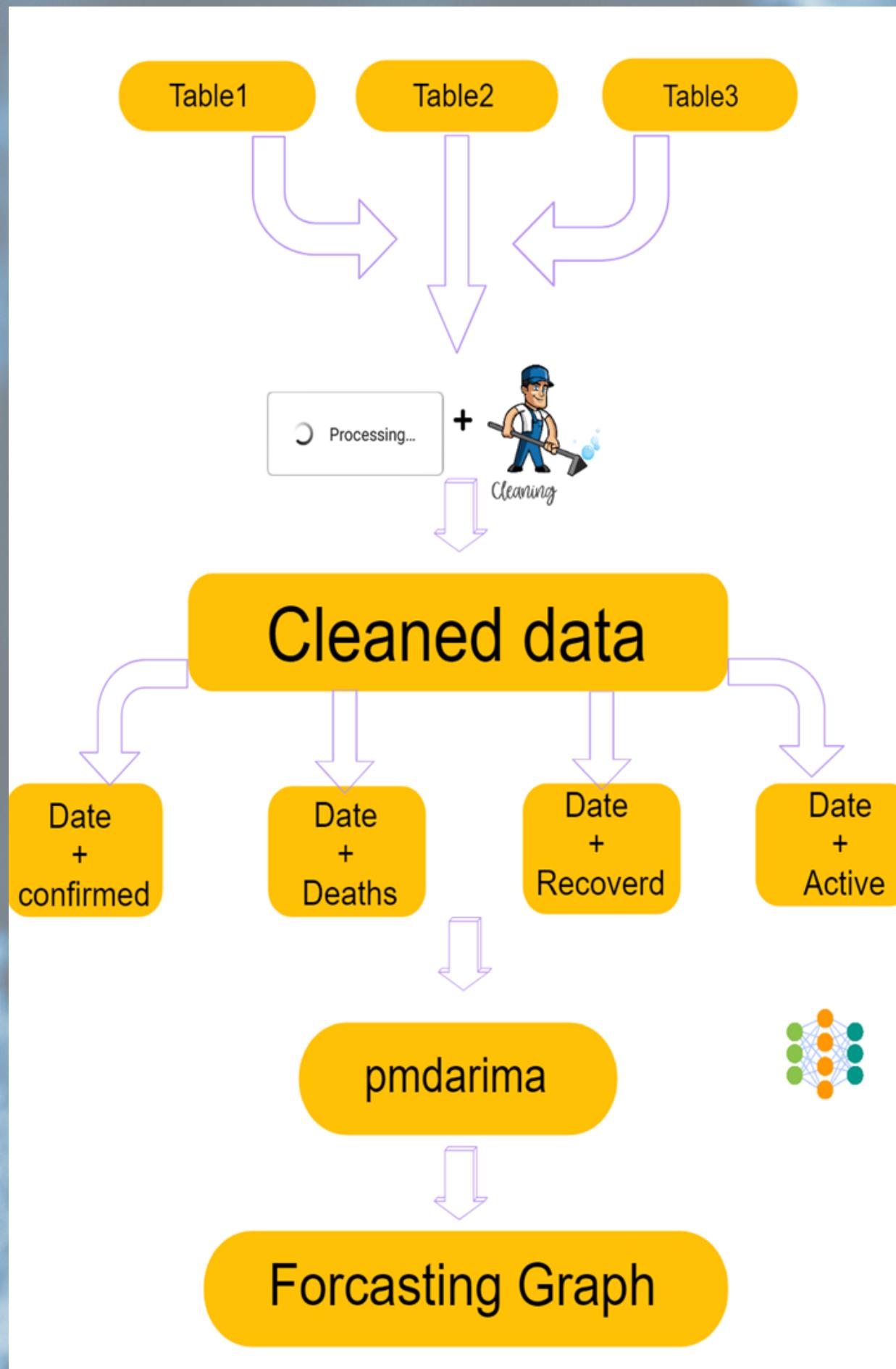
OBJECTIVE

- To forecast future trends.
- To Identify High-Risk Populations.
- To Inform Policy Decisions.

STAGES OF PROJECT



WORK FLOW



DATA EXTRACTION

Source: kaggle

- Study different tables
- Planning for Data Analysis
- Preparing dataset for Data cleaning

DATA CLEANING

- Detect useful data from dataset
- Remove redundant data
- Drop unnecessary data & join different tables using Pandas and NumPy

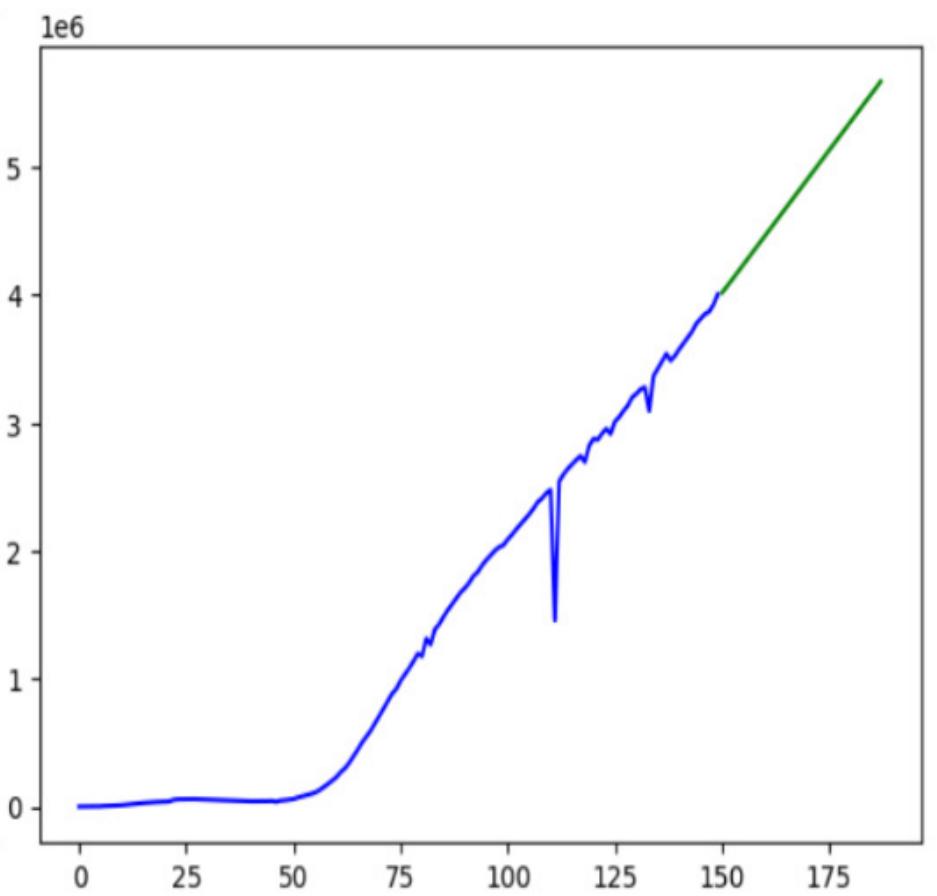
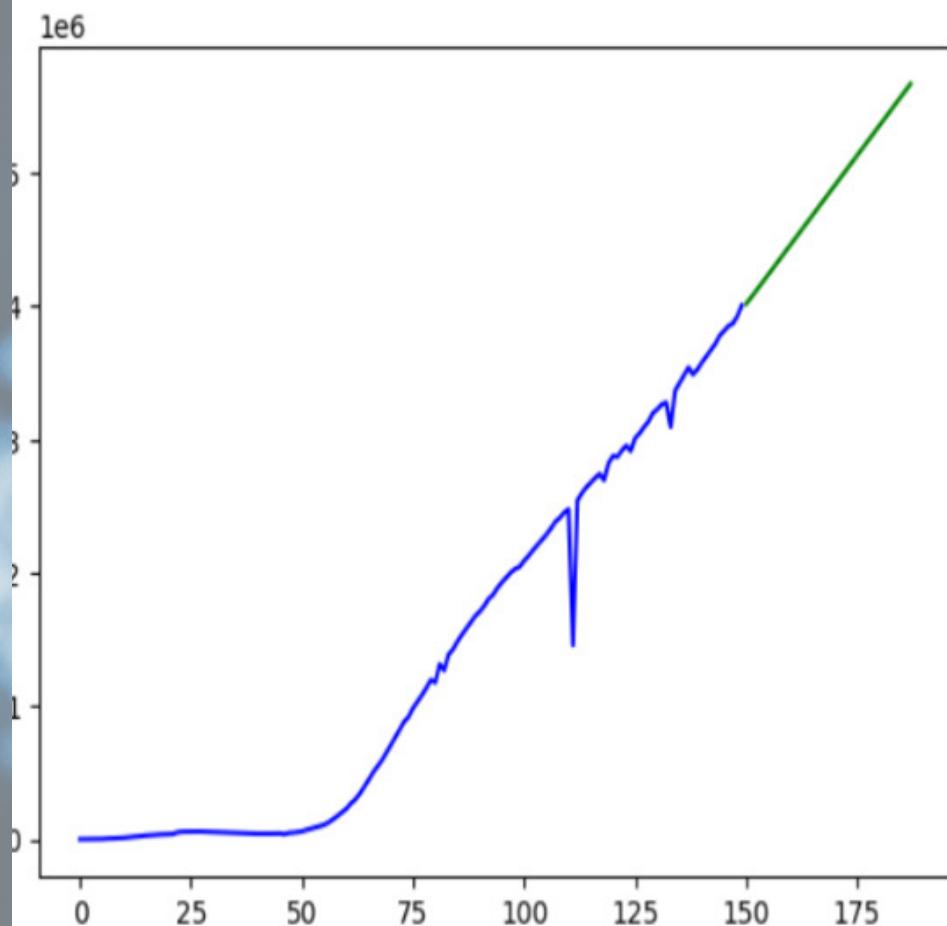
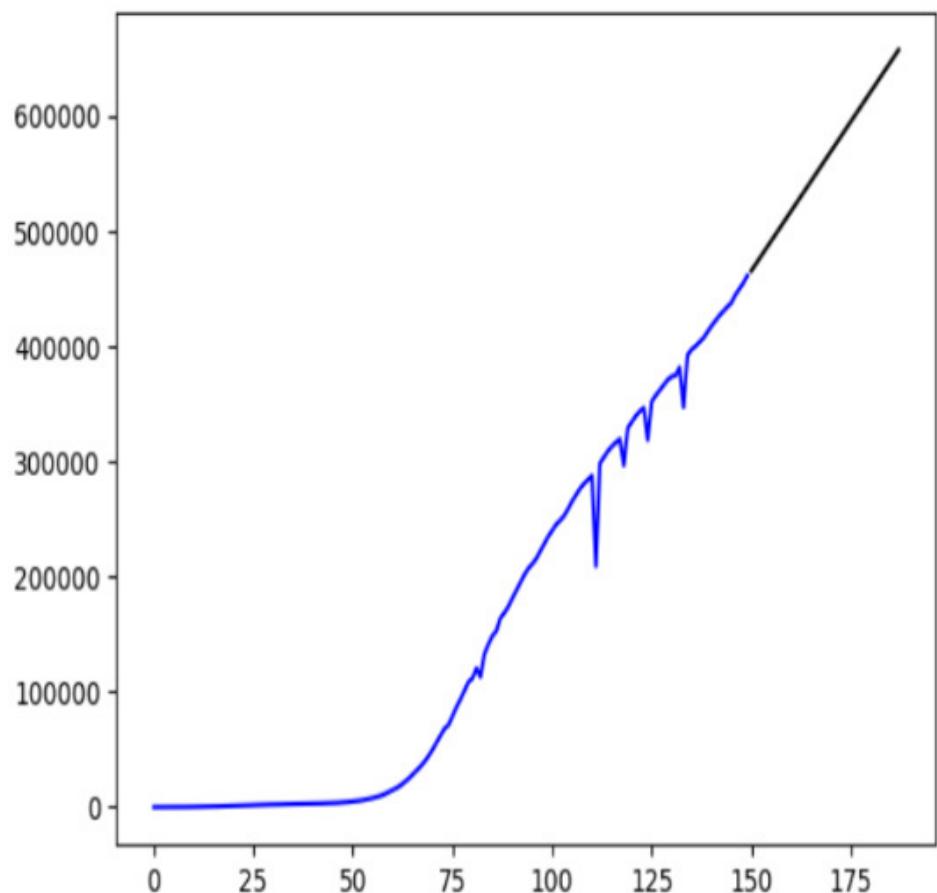
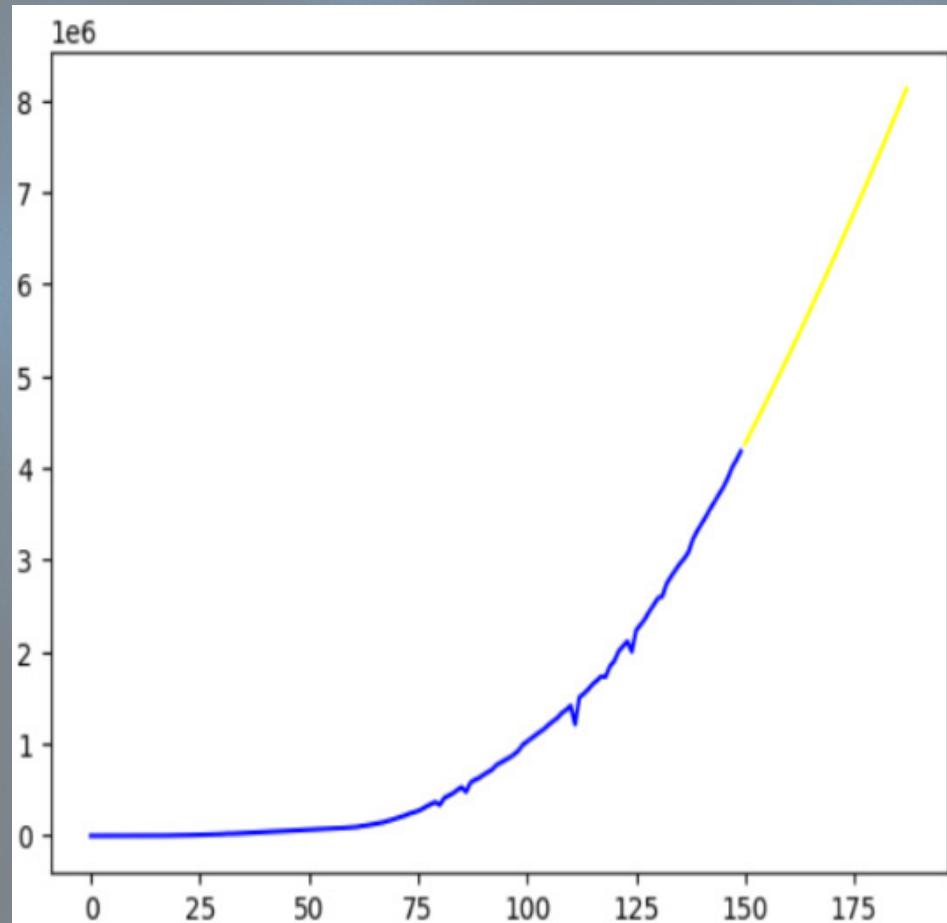
MODEL SELECTION

- USE OF ARIMA
- USE OF pmdarima

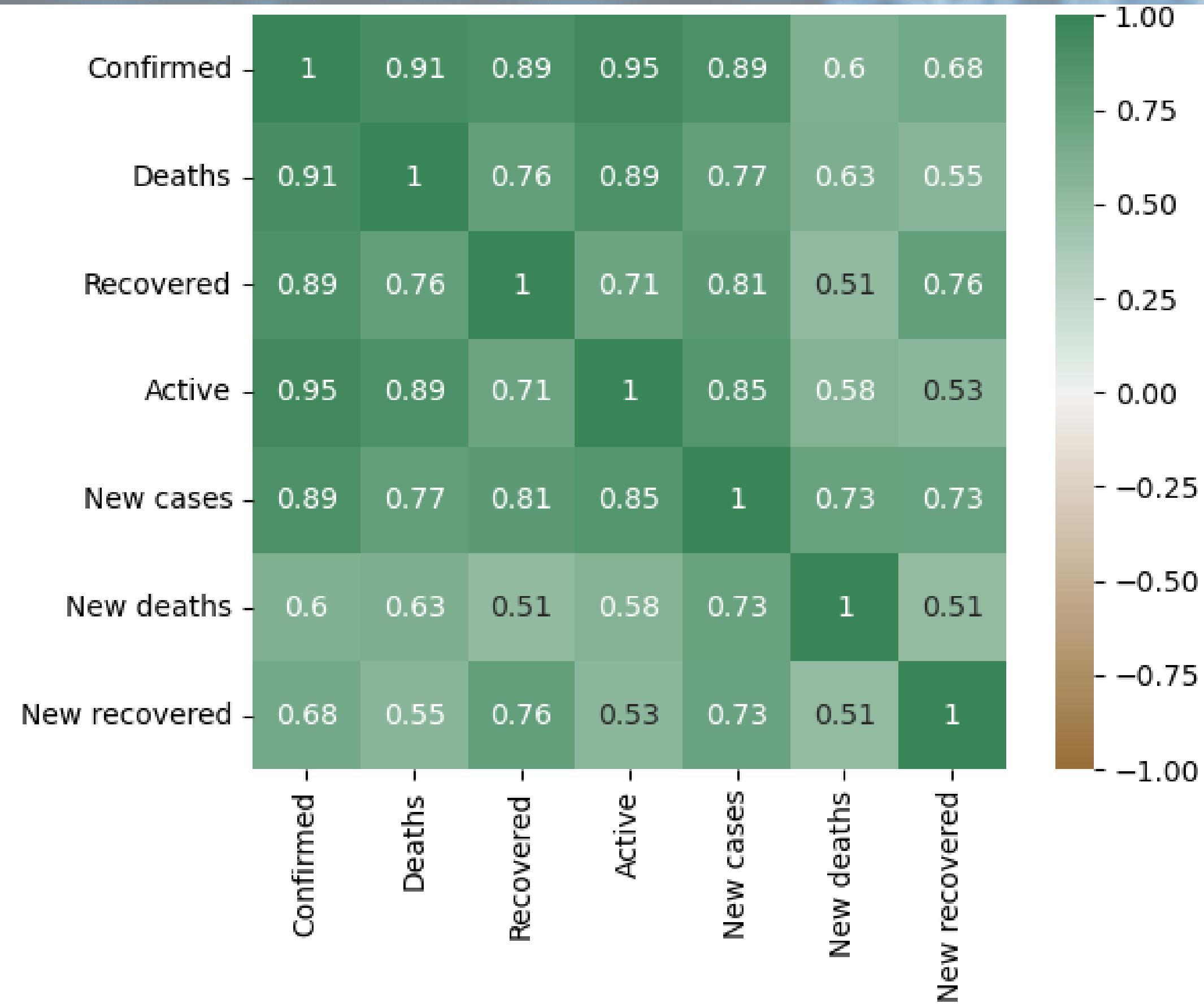
MODEL TRAINING

- Check for stationarity
- ADF test
- Fit The model

ANALYSIS & FORECASTING



CORRELATION MATRIX



CODE FOR CORRELATION MATRIX

```
In [44]: import matplotlib.pyplot as plt  
import seaborn as sns  
import pandas as pd
```

```
In [45]:  
FCG1=pd.read_excel(r"C:\Users\dbda.STUDENTS\DC\Desktop\Project\FCG1.xlsx")  
correlation = FCG1.corr()  
axis_corr = sns.heatmap(  
correlation,  
vmin=-1, vmax=1, center=0, annot=True,  
cmap=sns.diverging_palette(50, 500, n=500),  
square=True  
)  
  
plt.show()
```

C:\Users\dbda.STUDENTS\DC\AppData\Local\Temp\ipykernel_14424\4189429814.py:2: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.
correlation = FCG1.corr()

RANGE OF P-VALUE

Pearson's r value

0
0.1 to 0.3
0.3 to 0.5
0.5 to 0.7
0.7 to 1
-0.1 to -0.3
-0.3 to -0.5
-0.5 to -0.7
-0.7 to -1

Strength of relationship

No linear relationship
Weak linear relationship
Moderate linear relationship
Strong linear relationship
Very strong linear relationship
Weak negative linear relationship
Moderate negative linear relationship
Strong negative linear relationship
Very strong negative linear relationship

AVERAGE RECOVERY RATE

[19] `Average_recovery_rate=FCG1['RecoveryRate'].mean()`

[20] `Average_recovery_rate`

0.34426219002904174

✓ `Average_recovery_rate=0.34426219002904174`

AVERAGE DEATH RATE

[23] `Average_Death_rate=FCG1['DeathRate'].mean()`

[24] `Average_Death_rate`

0.04862476305829724

✓ `Average_Death_rate=0.04862476305829724`

ADF TEST

✓
0s [17] #data["Passenger_Season_Diff"]
y=pd.DataFrame(y)
adf_test(y.iloc[:,0])

ADF Statistic: 6.048136019198367

p-value: 0.01366

Critical Values:

1%: -3.48

5%: -2.88

10%: -2.58

MODEL SUMMARY

✓
0s



model.summary()



SARIMAX Results

Dep. Variable: y **No. Observations:** 150
Model: SARIMAX(1, 2, 4) **Log Likelihood:** -1963.562
Date: Fri, 16 Feb 2024 **AIC:** 3939.124
Time: 16:51:54 **BIC:** 3957.107
Sample: 01-22-2020 **HQIC:** 3946.430
- 06-19-2020

Covariance Type: opg

	coef	std err	z	P> z	[0.025	0.975]
ar.L1	-0.9629	0.132	-7.300	0.000	-1.221	-0.704
ma.L1	-0.7549	0.150	-5.023	0.000	-1.049	-0.460
ma.L2	-0.9500	0.229	-4.144	0.000	-1.399	-0.501
ma.L3	0.7728	0.089	8.703	0.000	0.599	0.947
ma.L4	0.0304	0.066	0.464	0.642	-0.098	0.159
sigma2	2.281e+10	1.31e-10	1.74e+20	0.000	2.28e+10	2.28e+10

Ljung-Box (L1) (Q): 0.08 **Jarque-Bera (JB):** 33691.26

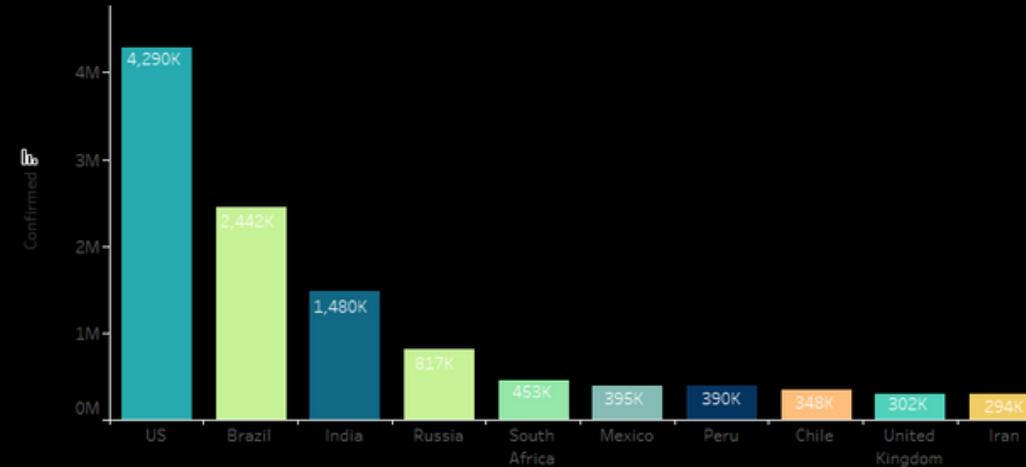
Prob(Q): 0.77 **Prob(JB):** 0.00

Heteroskedasticity (H): 1256.95 **Skew:** -7.03

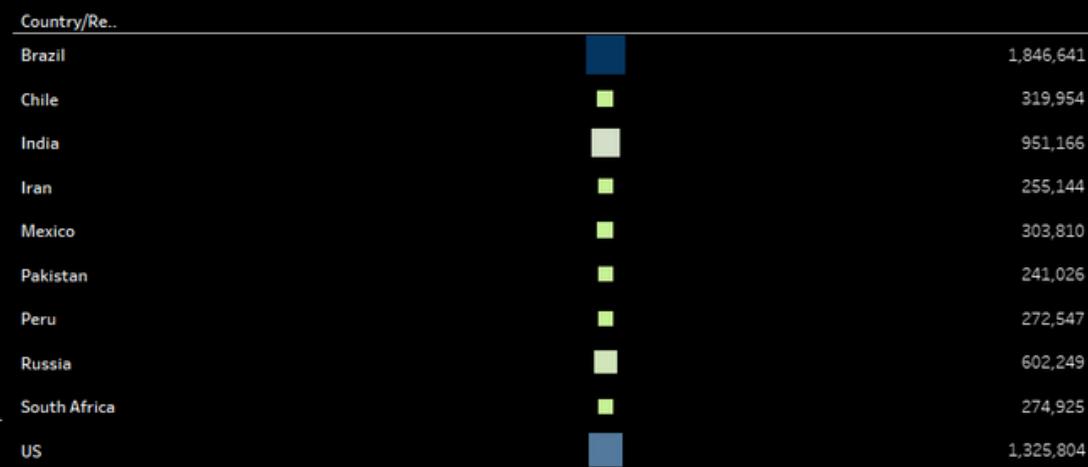
Prob(H) (two-sided): 0.00 **Kurtosis:** 75.57

DATA VISUALIZATION

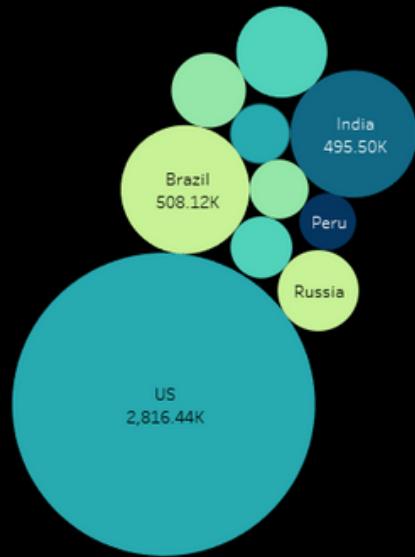
Top 10 country by total cases



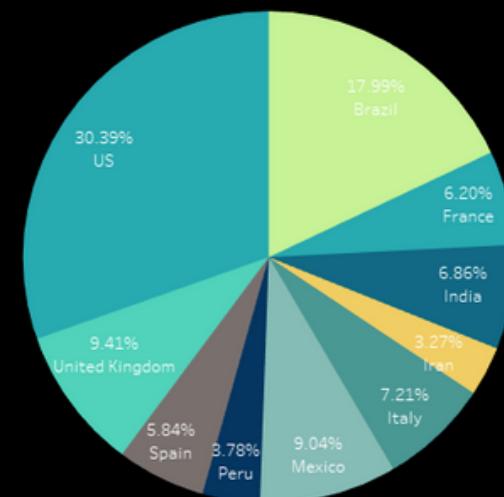
Top 10 countries by recovery cases



Top 10 countries active



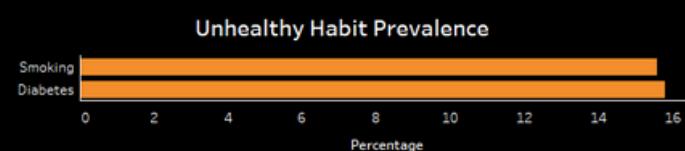
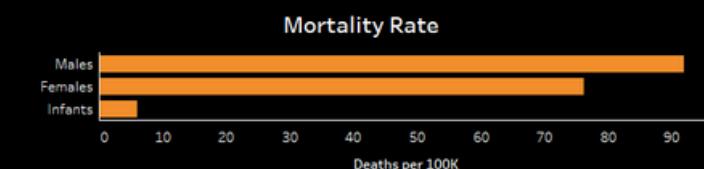
Distribution of total deaths in top 15 country



DATA VISUALIZATION



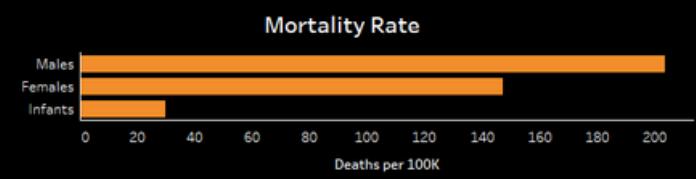
DATA VISUALIZATION



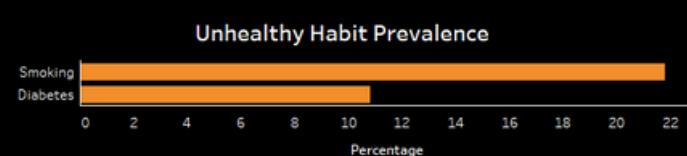
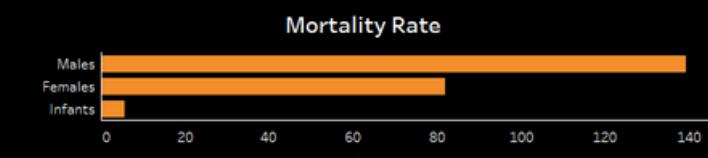
DATA VISUALIZATION



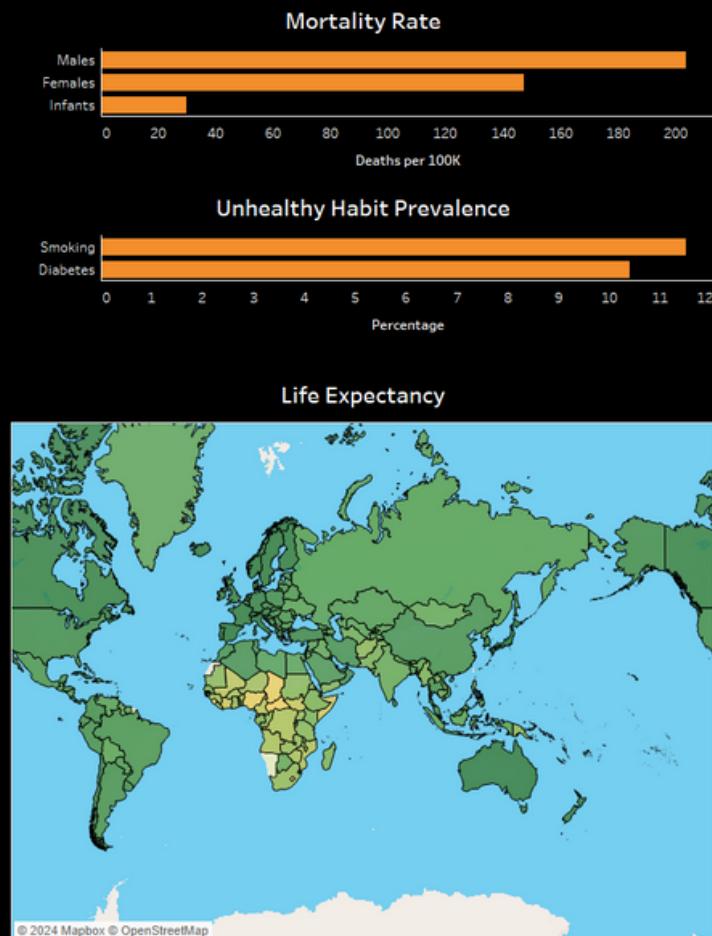
DATA VISUALIZATION



DATA VISUALIZATION



DATA VISUALIZATION



CONCLUSION

- pmdarima making it easier and more efficient to build accurate models.
- USA had highest number of cases & Death rate
- With forecasting, we can state that number of cases will increase in upcoming month.

FUTURE SCOPE

- Integration with External Data Sources
- Real-Time Forecasting
- Model Interoperability
- Scalability.

Thank You

