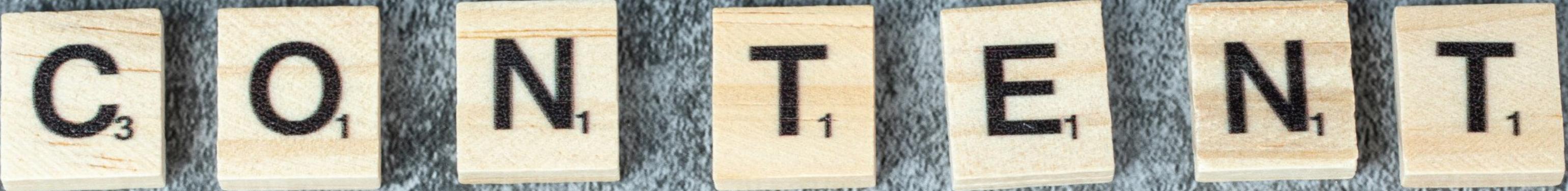




COVID VACCINES ANALYSIS

PHASE 2 :
INNOVATION



- Introduction
 - Context of phase 2
 - Advanced Machine learning techniques
 - Time forecasting series
 - Conclusion
- ○ ○
- ○ ○
- ○ ○
- ○ ○
- ○ ○



Introduction

The COVID Vaccine: A Shot of Hope in the Arm of the World. The COVID-19 pandemic has affected the world in unprecedented ways, but the development of vaccines has given us hope. This presentation will explore the science behind the COVID-19 vaccine and conduct an in-depth analysis of vaccine data, vaccine efficacy, its distribution, adverse effects and its impact on the world.

PHASE 2:

Consider exploring advanced machine learning techniques like clustering or time series forecasting to uncover hidden patterns in vaccine distribution and adverse effects data.



DATASET LINK :

<https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress>





COVID-19 World ...
kaggle.com



Search

Sign In

Register



GABRIEL PREDA · UPDATED A YEAR
AGO

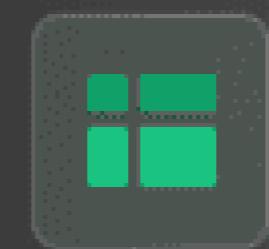
▲ 2175

New Notebook

Download (2 MB)



COVID-19 World Vaccination Progress



country_vaccinations_by_manufacturer



country_vaccinations



country_vaccinations_by_manufacturer

country_vaccinatio...

	A	B	C	D	E	F	G
1	location	date	vaccine	total_vaccinations			
2	Argentina	12/29/2020	Moderna	2			
3	Argentina	12/29/2020	Oxford/Astra	3			
4	Argentina	12/29/2020	Sinopharm/E	1			

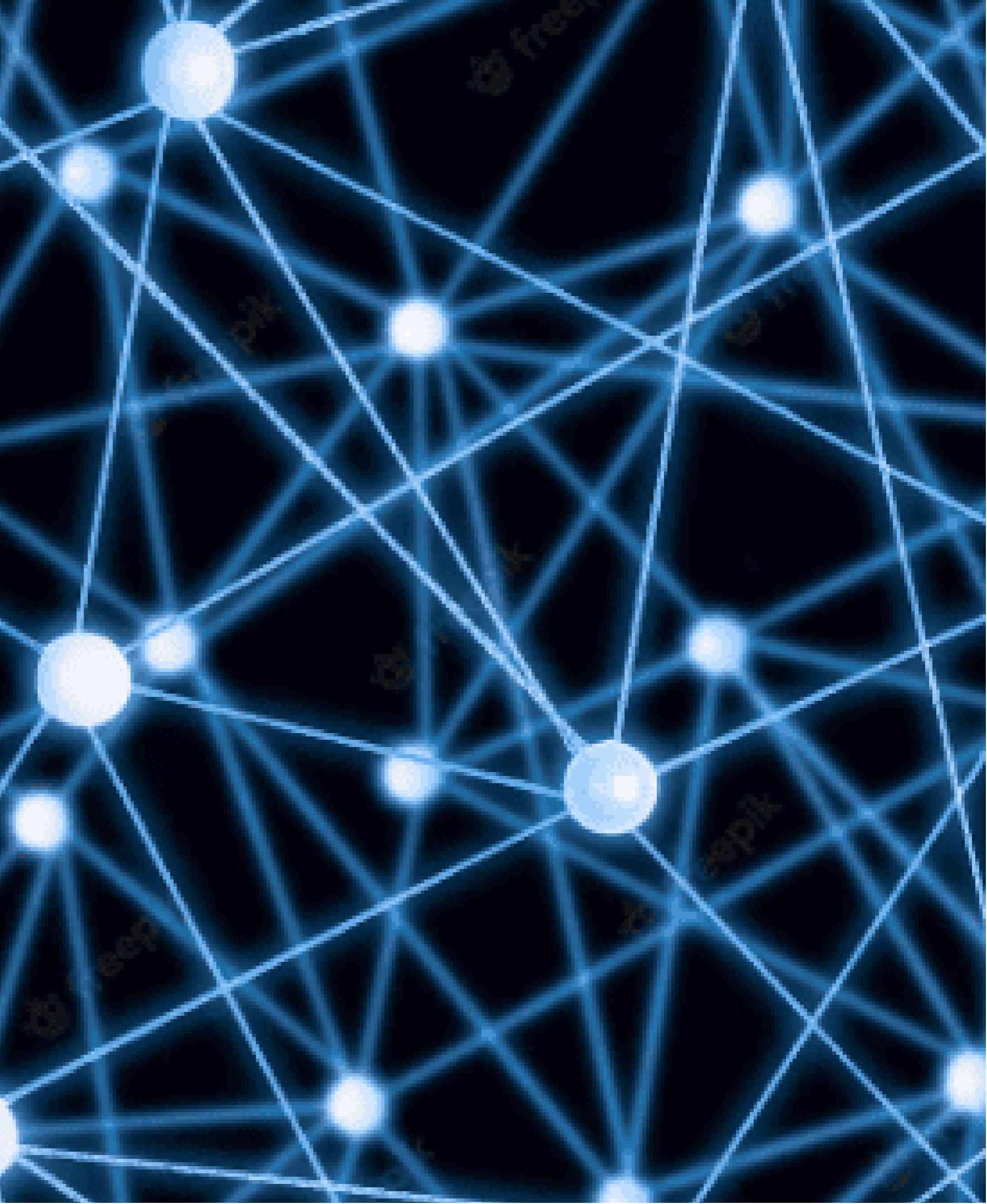
country_vaccinations

country_vaccinations

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O		
1	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations	daily_vaccinations_smoothed	total_vaccinations_smoothed	people_vaccinated_smoothed	people_fully_vaccinated_smoothed	new_vaccinations	new_vaccinations_smoothed	vaccines	source_name	source_website	
2	Afghanistan	AFG	2/22/2021	0	0				0	0				Johnson&Johnson	World Health Organization	https://covid19.who.int/WHO-COVID-19-global-data.csv	
3	Afghanistan	AFG	2/23/2021					1367						34	Johnson&Johnson	World Health Organization	https://covid19.who.int/WHO-COVID-19-global-data.csv
4	Afghanistan	AFG	2/24/2021				1367							34	Johnson&Johnson	World Health Organization	https://covid19.who.int/WHO-COVID-19-global-data.csv
5	Afghanistan	AFG	2/25/2021			1367								34	Johnson&Johnson	World Health Organization	https://covid19.who.int/WHO-COVID-19-global-data.csv
6	Afghanistan	AFG	2/26/2021			1367								34	Johnson&Johnson	World Health Organization	https://covid19.who.int/WHO-COVID-19-global-data.csv
7	Afghanistan	AFG	2/27/2021			1367								34	Johnson&Johnson	World Health Organization	https://covid19.who.int/WHO-COVID-19-global-data.csv

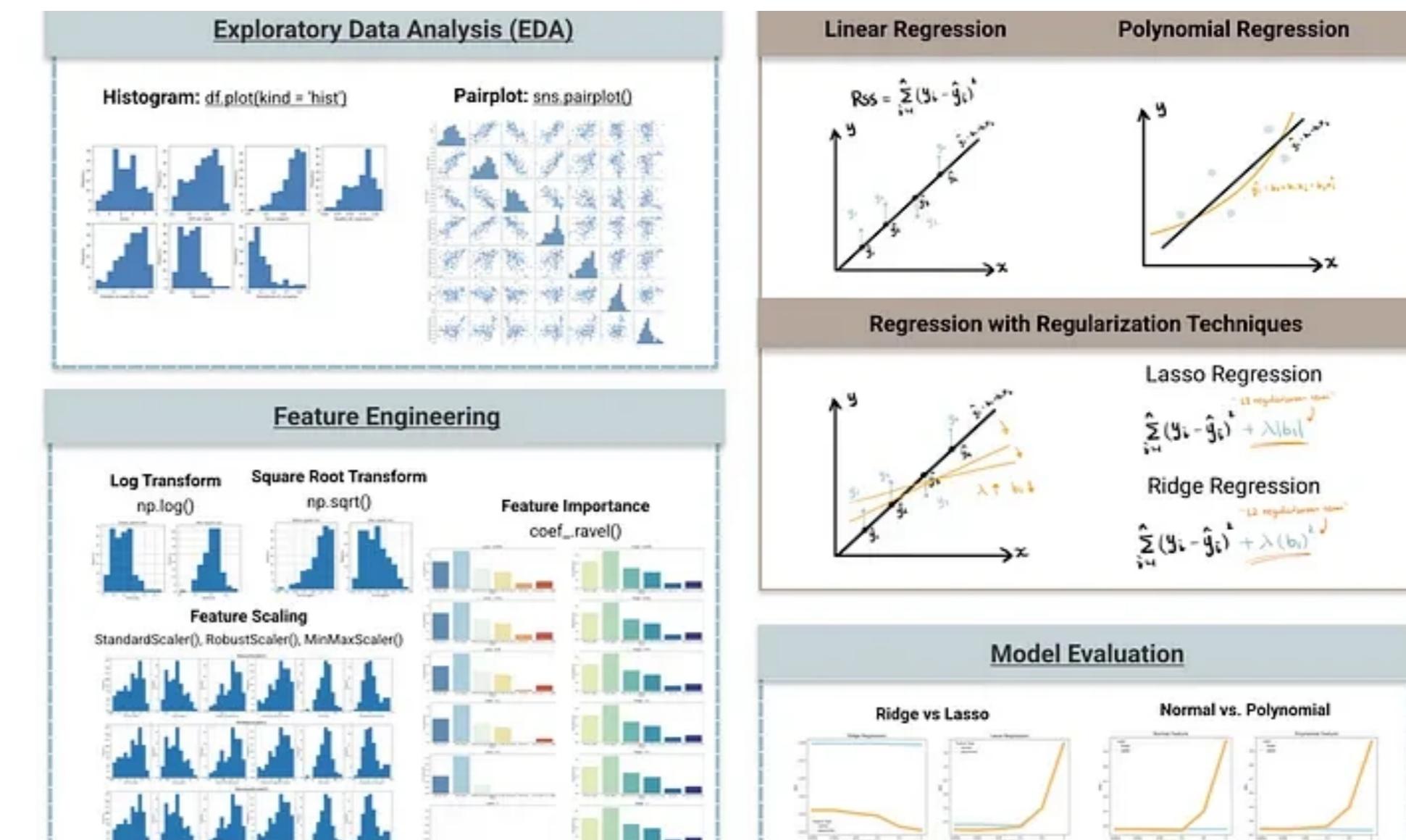
ADVANCED MACHINE LEARNING TECHNIQUES

Jupyter Notebook supports a wide range of machine learning libraries and frameworks, including scikit-learn, TensorFlow, and PyTorch. You can use these libraries to implement advanced machine learning techniques such as deep learning, natural language processing, and image recognition.



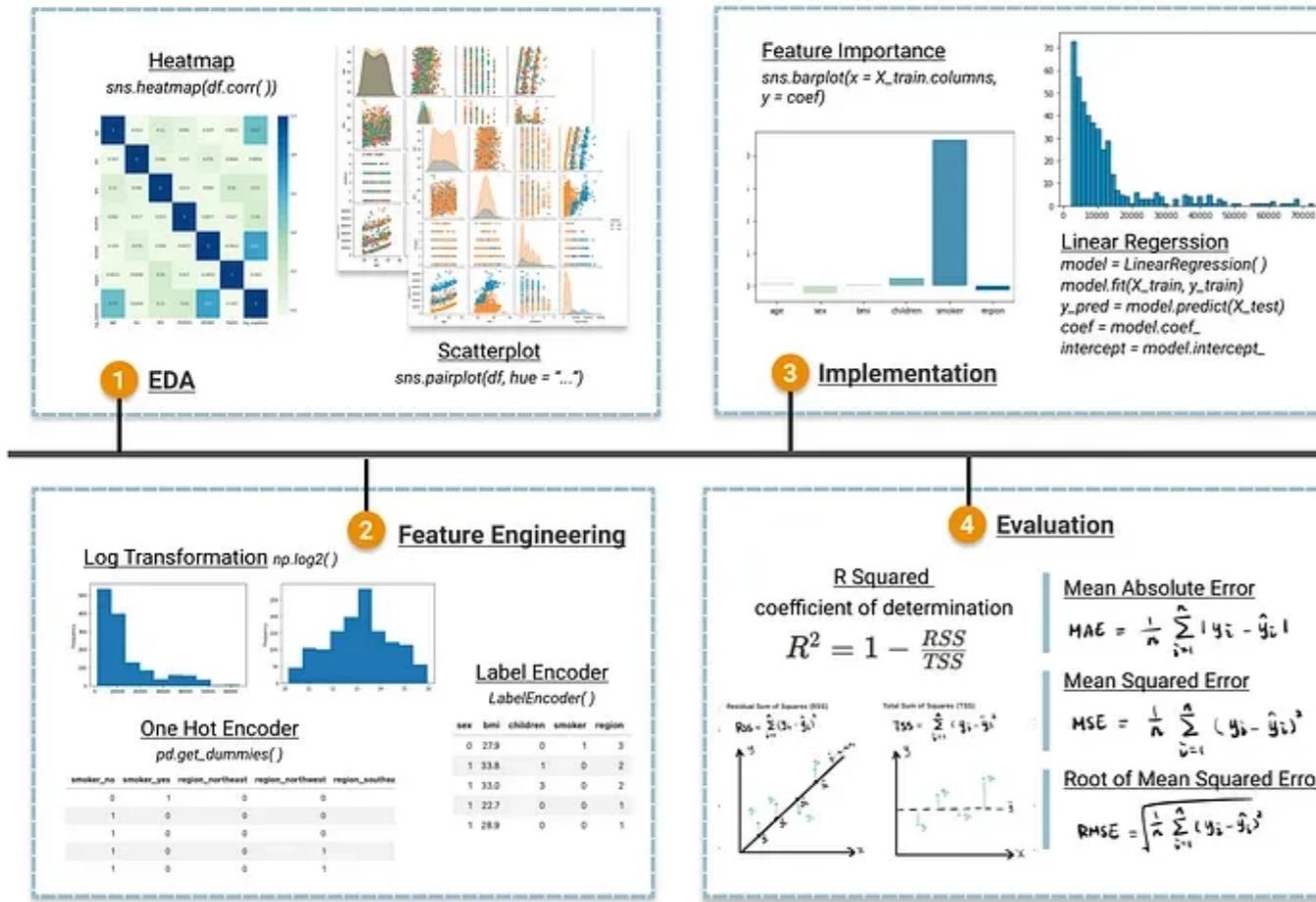
Advanced machine learning techniques play a crucial role in data analysis by enabling more accurate, sophisticated, and actionable insights from complex and large datasets. In other words advanced machine learning techniques enhance data analysis by providing powerful tools for modeling complex relationships, handling various types of data, automating tasks, and improving predictive accuracy. These techniques empower data analysts and data scientists to extract actionable insights, make informed decisions, and drive innovation across various domains. Some of the important Advanced Machine learning techniques that are most commonly used in data Analysis are listed as follows:

1. Linear Regression.
2. Logistic regression.
3. Support Vector Machine
4. Decision Tree.
5. KNN.
6. Naïve Bayes.
7. Random Forest.
8. Gradient Boosting Algorithm
9. Clustering Algorithm



LINEAR REGRESSION

Linear regression is a fundamental and widely used statistical technique in data analysis. It serves as a valuable tool for understanding and modeling the relationships between variables. Linear regression is a supervised learning algorithm used for predicting a continuous outcome variable (dependent variable) based on one or more predictor variables (independent variables). In conclusion, linear regression is a foundational technique in data analysis that helps model and quantify relationships between variables. It is a valuable tool for understanding trends, making predictions, and informing decision-making in a wide range of fields, from economics to healthcare to marketing.



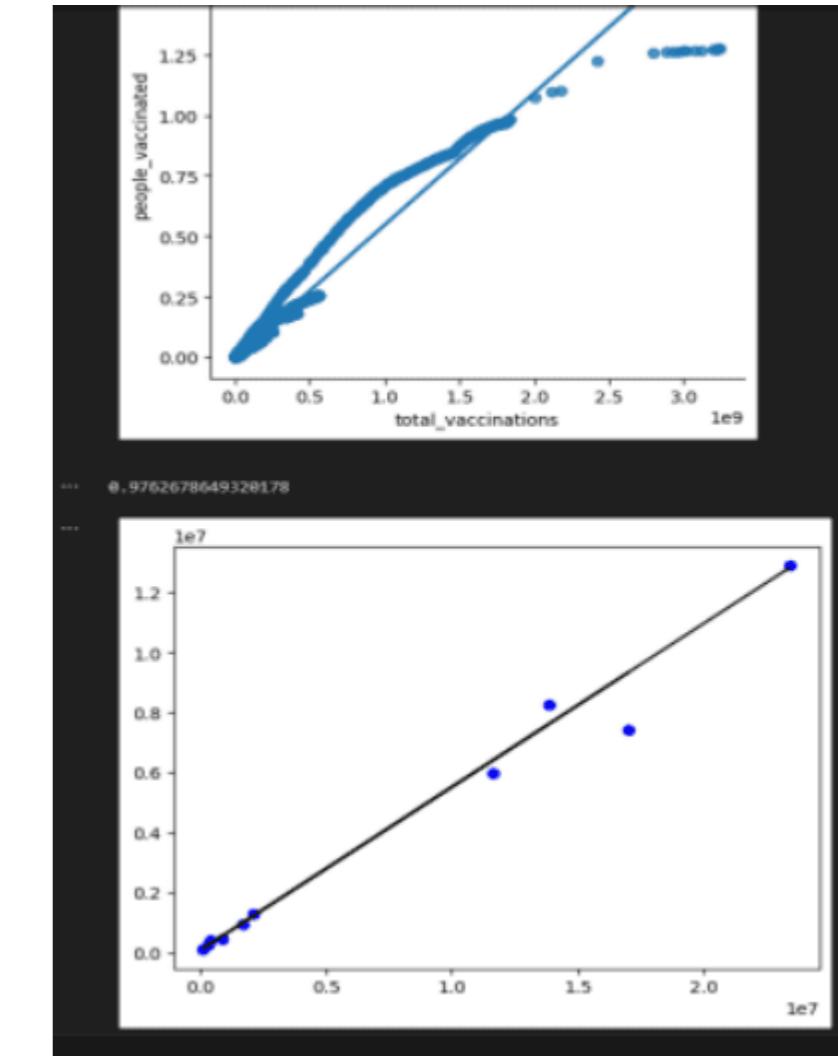
```

import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
from sklearn.model_selection import train_test_split
import sklearn.metrics as sm
data=pd.read_csv("country_vaccinations.csv")
print(data.head())
sns.lmplot(x="total_vaccinations",y="people_vaccinated",data=data,ci=None)
plt.show()
x=np.array(data["total_vaccinations"]).reshape(-1,1)
y=np.array(data["people_vaccinated"]).reshape(-1,1)
X_train,X_test,y_train,y_test=train_test_split(x,y,test_size=10)
regr=LinearRegression()
regr.fit(X_train,y_train)
print(regr.score(X_test,y_test))
y_pred=regr.predict(X_test)
plt.scatter(X_test,y_test,color="b")
plt.plot(X_test,y_pred,color="k")
plt.show()
print("R@ score = ",round(sm.r2_score(y_test,y_pred)))

```

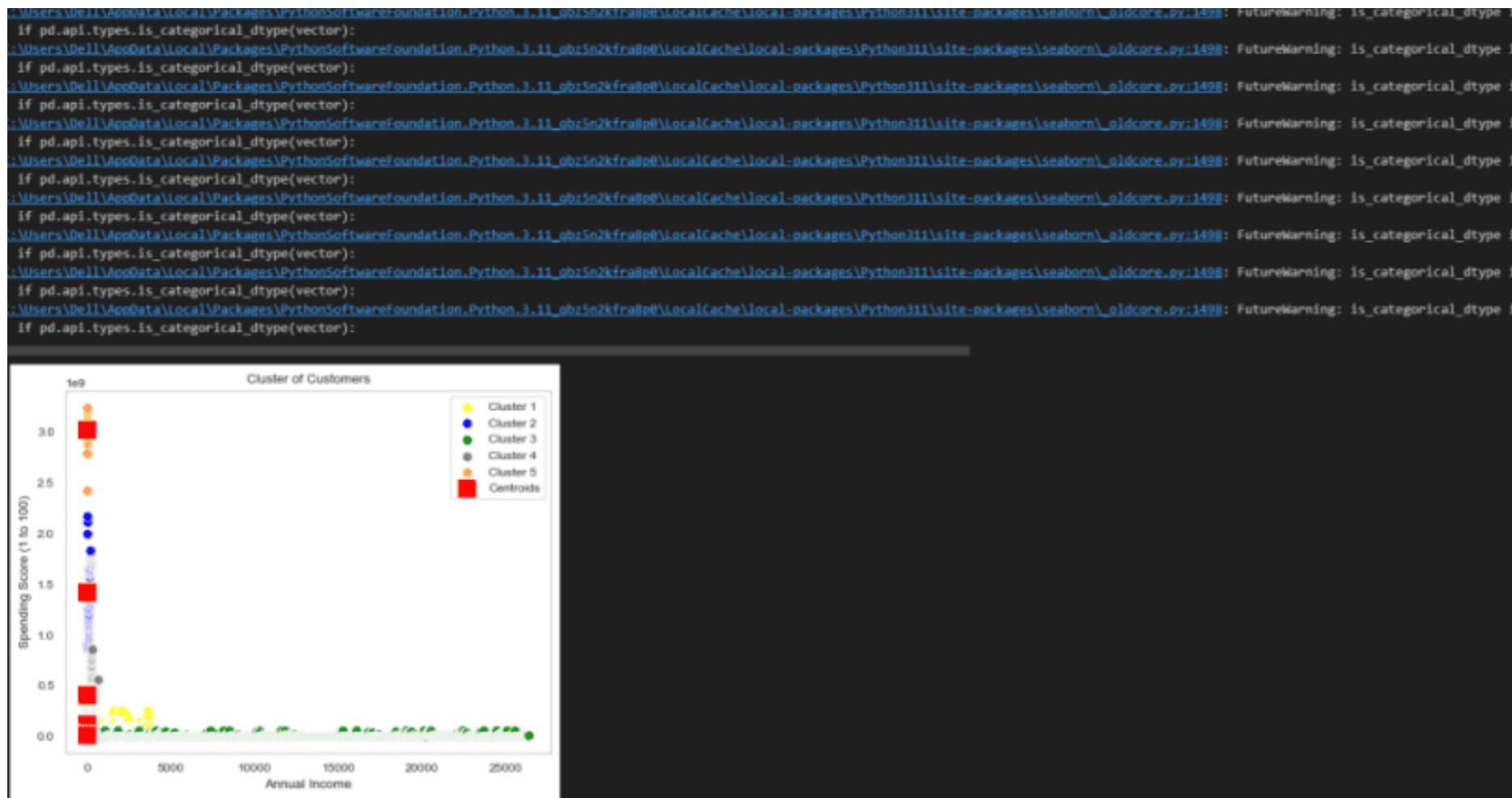
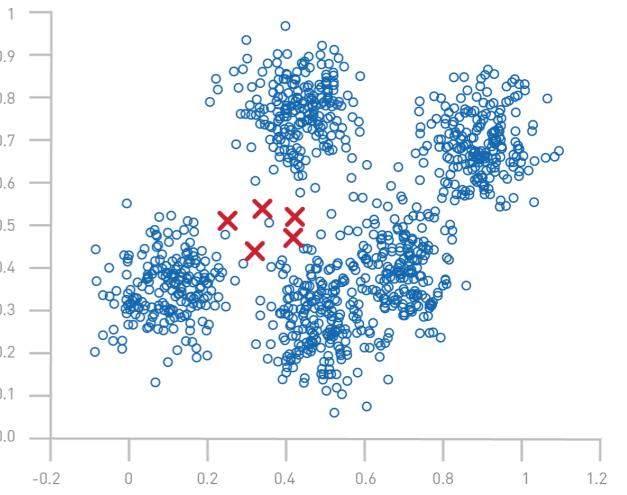
	country	iso_code	date	total_vaccinations	people_vaccinated
0	Afghanistan	AFG	2021-02-22	0.0	0.0
1	Afghanistan	AFG	2021-02-23	NaN	NaN
2	Afghanistan	AFG	2021-02-24	NaN	NaN
3	Afghanistan	AFG	2021-02-25	NaN	NaN
4	Afghanistan	AFG	2021-02-26	NaN	NaN

	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations
0	NaN	NaN	NaN
1	NaN	NaN	1367.0
2	NaN	NaN	1367.0



K-MEANS CLUSTERING

K-Means clustering is a popular unsupervised machine learning technique used for partitioning data into groups or clusters based on similarity. It's a fundamental algorithm in data analysis and has numerous applications in various fields.



TIME SERIES FORECASTING

Time series forecasting is a statistical technique used to make predictions about future values based on historical data points collected over time. It plays a critical role in various domains, including finance, economics, weather forecasting, and demand planning

AUTO REGRESSION MODEL

AutoRegressive (AR) models, often denoted as AR(p), are a class of time series forecasting models used to predict future values in a timeordered sequence based on their own past values. These models are part of the Autoregressive Integrated Moving Average (ARIMA) modeling framework. In this part of the image we have simply checked for an prevailing null Values by plotting the heatmap. This particular dataset does'nt have any null values thus an uniform color appears on the output page.

The auto Regressive model is build up for the given dataset by making use of the package pmdarima in which the module arima is installed and then the graph is plotted for it correspondingly.

```
[1]: import pandas as pd
from fbprophet import Prophet
import matplotlib.pyplot as plt

# Load the historical vaccination data into a DataFrame
data = pd.read_csv('country_vaccinations.csv')
data['date'] = pd.to_datetime(data['date'])
data.rename(columns={'date': 'ds', 'vaccinated_people': 'y'}, inplace=True)

# Initialize and fit the Prophet model
model = Prophet()
model.fit(data)

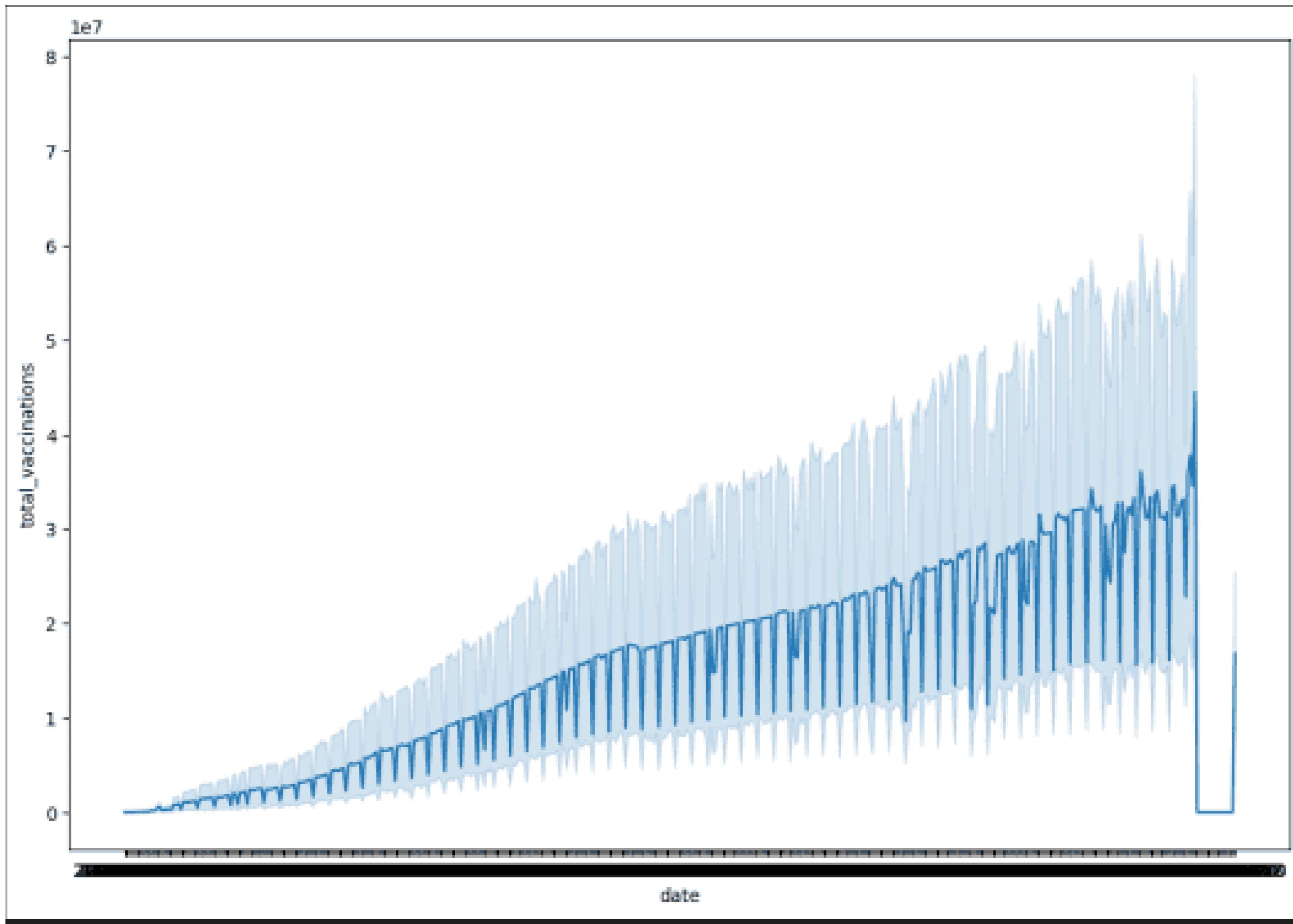
# Create a dataframe for future prediction
future = model.make_future_dataframe(periods=30) # 30 days into the future
forecast = model.predict(future)

# Plot the forecast
fig = model.plot(forecast)
plt.title('COVID-19 Vaccination Progress Forecast')
plt.xlabel('Date')
plt.ylabel('Number of Vaccinated People')
plt.show()

# Plot the forecast components (trend, weekly, yearly)
fig = model.plot_components(forecast)
plt.show()
```

TIME SERIES FORECASTING

OUTPUT :



DESIGNING REPORTS

In COVID VACCINE ANALYSIS, report design is vital. Utilizing IBM Cognos tools, data insights are transformed into visually appealing presentations, employing appropriate visualizations for KPIs and metrics. Effective design ensures clarity and relevance, enhancing user understanding. Tailored reports provide a dynamic platform for data-driven decision-making in public health strategies.



CONCLUSION



The covid -19 vaccines analysis data exploration using advanced machine learning techniques like clustering or time series forecasting to uncover hidden patterns in vaccine distribution and adverse effects data were implemented.

Thank you!

