MODY UNIVERSITY
OF SCIENCE AND TECHNOLOGY
LAKSHMANGARH • RAJASTHAN
A Leading Women's University

PHASE-1 EVALUATION

Forecasting World GDP using Machine Learning

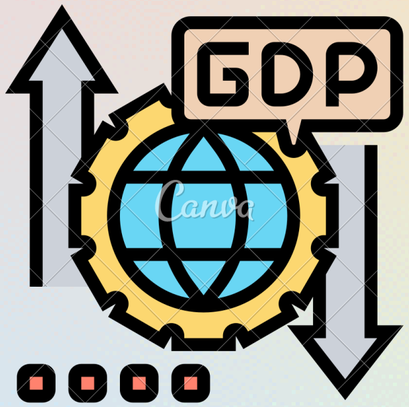
MENTOR NAME-
MR. VINAY VERMA

PROJECT BY-
ANUSHA AGARWAL (190018)
LAVINA GARG(190051)
SIDDHI BANSAL(190084)

Gross Domestic Product

Gross Domestic Product (GDP) is the main indicator used in the evaluation of the performance of an economy.

ABSTRACT



The main aim of this project is to create an ML model that would help in predicting the expected GDP of each country with a minimum margin of error for any given year in the future such as 2022 and beyond. This is very beneficial since by this data we could predict the growth of different countries and it will be able to predict how the GDP must be used in an effective way for the benefit of the state. The GDP of countries is impacted by various social, economical and cultural parameters. We are analysing those parameters from 1960 to 2021 and will predict future GDP up to 2025 of the world.

02

INTRODUCTION

Gross domestic product (GDP) is one of the most used measures of how a country's economy is performing. It is a measure of the economic growth of any given country. The economic growth of the country depends on many factors such as the social, economic, and cultural environment. GDP is usually a really large number that doesn't say much in itself, only in comparison to other countries GDP, per capita, or to the same country's GDP in different periods does it get context. The prediction of GDP has always been important with regards to identifying possible trends and future impacts on the economy. Various models have been created for predicting GDP, however in recent years there has been a growing understanding of the importance of, and use of Machine Learning based models for predicting economic performance.

Machine Learning has facilitated the identification of trends and patterns in greater depth than traditional statistical models and the power of ML algorithms and models has resulted in greater accuracy of predictions. The task for the analytics is to find which methods make the most accurate predictions and which data series to use. Because the different models give different results the analytics must make educated guesses when choosing which models to use.

Contribution to Society

- GDP is an indicator of a society’s standard of living, it gives information about the size of the economy and how an economy is performing.
- It permits to review of previous economic movements and predicts how current economic changes can amend the patterns of the previous trend; therefore, a more accurate prediction would provide a significant facilitate to the government in setting up economic development goals, ways and policies.
- Consequently, a correct Gross Domestic Product prediction presents an insightful associate with an understanding of future economics' trends.

Data Gathering

↓

Data Cleaning

↓

Data Exploration

↓

Data Modelling

↓

Model Evaluation

↓

Deployment

Project Design

04

BLUEPRINT

DATASET

- Entire _world economic database.csv(8775 rows, 56 columns)
- GDP.csv(532 rows, 66 columns)
- Factors.csv(rows, columns)

File Format: CSV

DATA PREPROCESSING

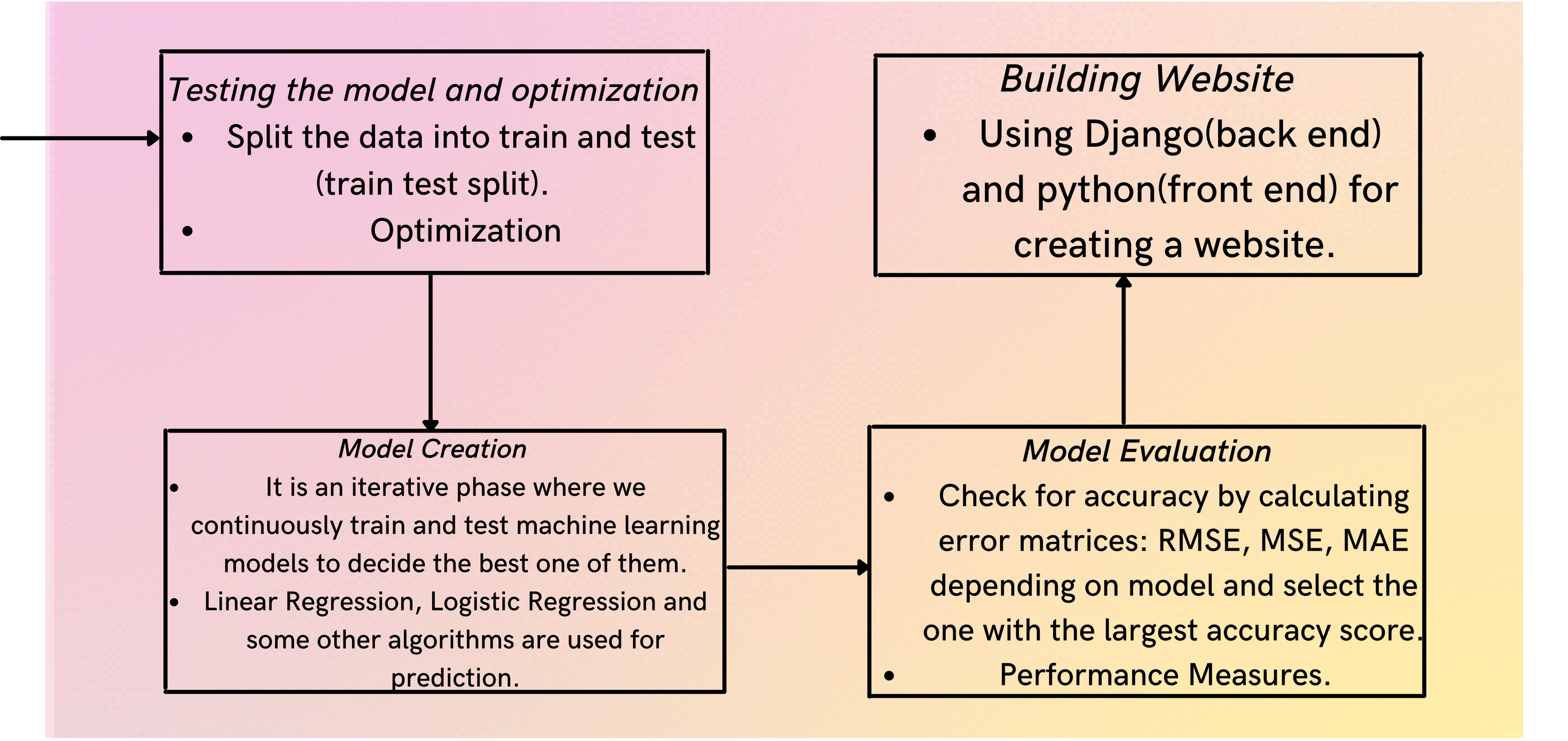
- For dataset 1: Missing values are filled and duplicate rows not.
- For datasets 2,3: Missing values are filled and duplicate rows are not present.

EDA

- Detect Outliers, anomalies and extract important variables.
- Checking correlation between variables of a dataset.
- Show datatypes, shape and other basic information of variables.
- Data visualization(Plotting the graphs like line plot barplot, heatmap etc.

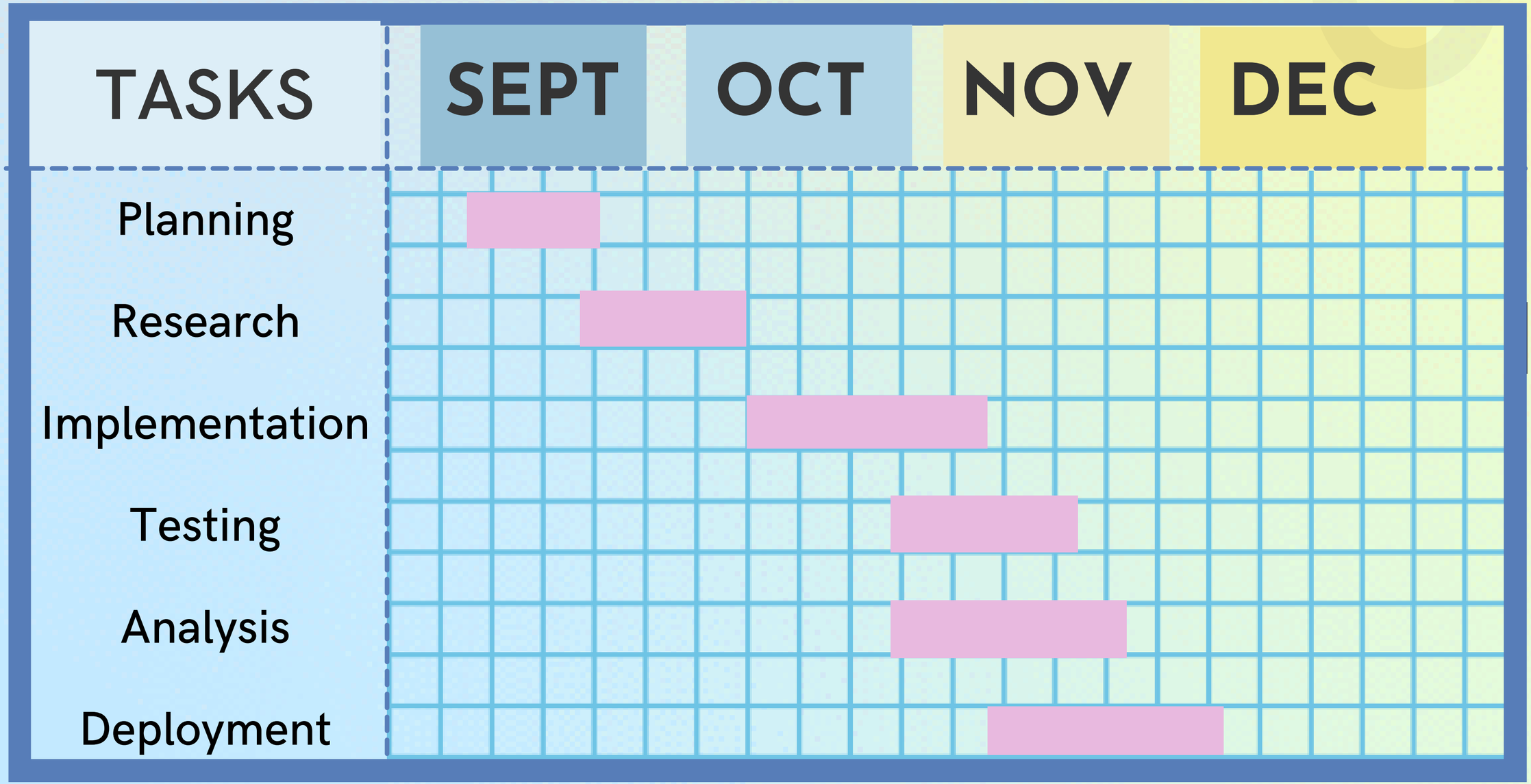
Feature Engineering

- Scaling: It is done to get the features on the same scale.
- Transformation: It is done to normalize the data(feature)


Testing the model and optimization
• Split the data into train and test (train test split).
• Optimization
Building Website
• Using Django(back end) and python(front end) for creating a website.
Model Creation
• It is an iterative phase where we continuously train and test machine learning models to decide the best one of them.
• Linear Regression, Logistic Regression and some other algorithms are used for prediction.
Model Evaluation
• Check for accuracy by calculating error matrices: RMSE, MSE, MAE depending on model and select the one with the largest accuracy score.
• Performance Measures.

Methodology

GANTT CHART


TASKS
SEPT
OCT
NOV
DEC
Planning
Research
Implementation
Testing
Analysis
Deployment

Snippets

#importing libraries

```
In [1]: import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt
        import seaborn as sns
        import matplotlib as mpl
        import warnings
        warnings.filterwarnings("ignore")
        %matplotlib inline
```

Out[3]:

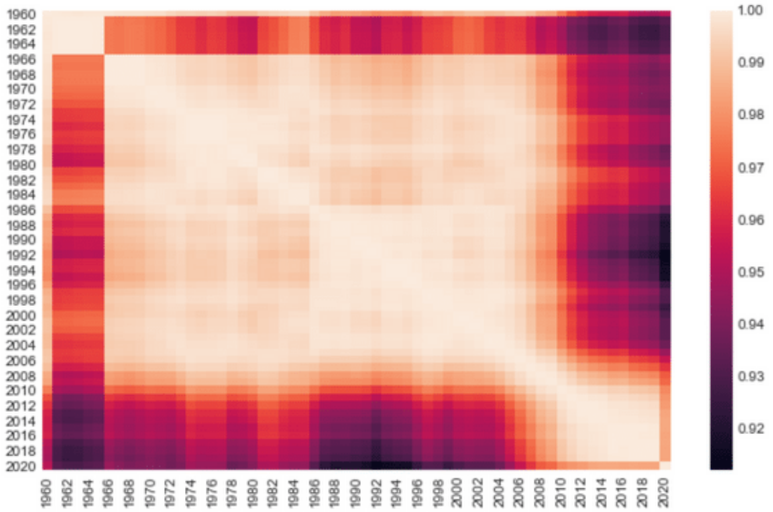
	WEO Country Code	ISO	WEO Subject Code	Country	Subject Descriptor	Subject Notes	Units	Scale	Country/Series-specific Notes	1980	...	2017	2018	201
0	512	AFG	NGAP_NPGDP	Afghanistan	Output gap in percent of potential GDP	Output gaps for advanced economies are calcula...	Percent of potential GDP	NaN	NaN	NaN	...	NaN	NaN	Na
1	914	ALB	PPPGDP	Albania	Gross domestic product, current prices	These data form the basis for the country weig...	Purchasing power parity; international dollars	Billions	See notes for: Gross domestic product, curren...	5.765	...	37.609	40.080	41.70
2	914	ALB	PCPI	Albania	Inflation, average consumer prices	Expressed in averages for the year, not end-of...	Index	NaN	Source: National Statistics Office Latest actu...	NaN	...	103.295	105.390	106.87
3	612	DZA	NGDP_R	Algeria	Gross domestic product, constant prices	Expressed in billions of national currency uni...	National currency	Billions	Source: National Statistics Office Latest actu...	2596.368	...	7364.675	7467.780	7527.52
4	612	DZA	PCPI	Algeria	Inflation, average consumer prices	Expressed in averages for the year, not end-of...	Index	NaN	Source: National Statistics Office Latest actu...	8.975	...	193.970	202.253	206.20
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
8770	582	VNM	GGXCNL_NGDP	Vietnam	General government net lending/borrowing	Net lending (+)/ borrowing (-) is calculated a...	Percent of GDP	NaN	See notes for: General government net lending...	NaN	...	-1.964	-1.023	-3.29
8771	582	VNM	BCA_NGDPD	Vietnam	Current account balance	Current account is all transactions other than...	Percent of GDP	NaN	See notes for: Gross domestic product, curren...	-1.599	...	-0.596	1.898	3.41

#dataset

#correlation between dataset

```
In [45]: plt.figure(figsize=(10,6))
         sns.heatmap(df2_gdp.corr())
```

Out[45]: <AxesSubplot:>



REFERENCES

- https://public.opendatasoft.com/explore/dataset/entire-world-economic-outlook-database/information/?
- https://www.worldometers.info/coronavirus/
- https://data.worldbank.org/
- https://knoema.com/tbocwag/gdp-forecast-by-country-statistics-from-imf-2021-2025?
- https://www.statista.com/statistics/263771/gross-domestic-product-gdp-in-india/

Thank you!

Let us know if you have suggestion for improvement.