Syracuse University

Exploring Global Economy, Internet Utilization, and Employment Trends

Team 2:

1. Nikitha Sharma
2. Kushal Paresh Modi
3. Tejal Sanjay Palwankar
4. Bhavana Yachenahalli Channakeshava Murthy
5. Varun Sham Kumar Kannappanahalli Shamanna

I. Introduction

Global economies and employment markets have undergone a transformation due to the quick development of technology and the growing use of the internet. The relationship between them has become a crucial field of research as countries become more digitally connected.

In order to understand how internet usage affects job trends and economic growth, this research explores the interactions among these three variables. Given the growth of the digital economy, it is critical to comprehend these relationships in order to establish policies that support innovation, sustainable development, and fair employment opportunities.

Objectives:

- Analyze global trends in GDP, internet usage, and employment.
- Explore correlations between economic growth and internet penetration.
- Use data visualization to identify patterns.
- Provide insights for informed decision-making.

By leveraging data-driven insights, this report aims to provide a comprehensive understanding of how technology and connectivity are reshaping the global economic landscape and labor markets.

The findings from this report aim to provide a deeper understanding of the interplay between internet utilization, economic growth, and employment. By presenting data-driven insights, this analysis contributes to ongoing discussions about the future of work, the global digital divide, and strategies for leveraging technology for sustainable development.

This exploration sets the foundation for a deeper dive into how the internet is shaping the world's economy and workforce in an era defined by constant innovation.

II. Data Description

- **Data Sources**: Three datasets were collected from https://data.un.org. All are structured in csv format.
  - **GDP Data**: The dataset includes information about GDP per capita across different countries and years.
  - **Internet Usage Data**: This dataset includes the percentage of internet users for various countries and years.
  - **Employment Data**: This dataset provides employment information across countries, segmented by gender.
- **Data Fields**:
  - **GDP Data**: Country, Year, Value
  - **Internet Usage Data**: Country, Year, Value
  - **Employment Data**: Country, Year, Gender, Employment Rate (or count).
- **Data Shape**: The GDP dataset contains 6,729 rows and 4 columns, while the Internet dataset has 4,495 rows and 4 columns and Employment dataset contains 51039 rows and 6 columns.

```
# Here we inspect shape of GDP dataframe
gdp_df.shape
```

```
(6729, 4)
```

```
# Here we inspect shape of Internet dataframe
internet_df.shape
```

```
(4495, 4)
```

```
# Here we inspect shape of Employment dataframe
employed_df.shape
```

```
(51039, 6)
```

[64]:
```
# Here we import the GDP dataset
gdp_df = pd.read_csv('gdp.csv', nrows=6729)
gdp_df.head(50)
```

[64]:

	Country or Area	Year	Value	Value Footnotes
0	Afghanistan	2021	1517.016266	NaN
1	Afghanistan	2020	1968.341002	NaN
2	Afghanistan	2019	2079.921861	NaN
3	Afghanistan	2018	2060.698973	NaN
4	Afghanistan	2017	2096.093111	NaN
5	Afghanistan	2016	2101.422187	NaN
6	Afghanistan	2015	2108.714173	NaN
7	Afghanistan	2014	2144.449634	NaN
8	Afghanistan	2013	2165.340915	NaN
9	Afghanistan	2012	2122.830759	NaN
10	Afghanistan	2011	1961.096317	NaN
11	Afghanistan	2010	2026.163818	NaN
12	Afghanistan	2009	1823.742614	NaN
13	Afghanistan	2008	1556.844452	NaN
14	Afghanistan	2007	1528.344578	NaN
15	Afghanistan	2006	1366.993146	NaN
16	Afghanistan	2005	1352.320627	NaN
17	Afghanistan	2004	1260.060581	NaN
18	Afghanistan	2003	1292.333437	NaN
19	Afghanistan	2002	1280.463171	NaN
20	Africa Eastern and Southern	2022	3553.913370	NaN
21	Africa Eastern and Southern	2021	3519.174840	NaN
22	Africa Eastern and Southern	2020	3455.023119	NaN

[65]:
```
# Here we import the Internet dataset
internet_df = pd.read_csv('internet.csv', nrows=4495)
internet_df.head(50)
```

[65]:

	Country or Area	Year	Value	Value Footnotes
0	Afghanistan	2019	17.600000	1
1	Afghanistan	2018	16.800000	2
2	Afghanistan	2017	13.500000	3
3	Afghanistan	2016	11.000000	4
4	Afghanistan	2015	8.260000	4
5	Afghanistan	2014	7.000000	4
6	Afghanistan	2013	5.900000	4
7	Afghanistan	2012	5.454550	4
8	Afghanistan	2011	5.000000	4
9	Afghanistan	2010	4.000000	4
10	Afghanistan	2009	3.550000	NaN
11	Afghanistan	2008	1.840000	NaN
12	Afghanistan	2007	1.900000	NaN
13	Afghanistan	2006	2.107120	NaN
14	Afghanistan	2005	1.224150	NaN
15	Afghanistan	2004	0.105809	NaN
16	Afghanistan	2003	0.087891	NaN
17	Afghanistan	2002	0.004561	NaN
18	Afghanistan	2001	0.004723	4
19	Afghanistan	1990	0.000000	NaN
20	Albania	2022	82.613700	NaN
21	Albania	2021	79.323700	NaN
22	Albania	2020	72.237700	NaN

	Country or Area	Year	Sex	Age	Status in employment	Value
0	Åland Islands	2000	Both Sexes	Total	Total	12932.0
1	Åland Islands	2000	Both Sexes	Total	Own account worker	1743.0
2	Åland Islands	2000	Both Sexes	Total	Employee	9993.0
3	Åland Islands	2000	Both Sexes	Total	Unknown	1196.0
4	Åland Islands	2000	Both Sexes	0 - 14	Total	0.0
5	Åland Islands	2000	Both Sexes	0 - 14	Own account worker	0.0
6	Åland Islands	2000	Both Sexes	0 - 14	Employee	0.0
7	Åland Islands	2000	Both Sexes	0 - 14	Unknown	0.0
8	Åland Islands	2000	Both Sexes	15 - 19	Total	368.0
9	Åland Islands	2000	Both Sexes	15 - 19	Own account worker	1.0
10	Åland Islands	2000	Both Sexes	15 - 19	Employee	322.0
11	Åland Islands	2000	Both Sexes	15 - 19	Unknown	45.0
12	Åland Islands	2000	Both Sexes	20 - 24	Total	859.0
13	Åland Islands	2000	Both Sexes	20 - 24	Own account worker	22.0
14	Åland Islands	2000	Both Sexes	20 - 24	Employee	745.0
15	Åland Islands	2000	Both Sexes	20 - 24	Unknown	92.0
16	Åland Islands	2000	Both Sexes	25 - 29	Total	1242.0
17	Åland Islands	2000	Both Sexes	25 - 29	Own account worker	79.0
18	Åland Islands	2000	Both Sexes	25 - 29	Employee	1024.0
19	Åland Islands	2000	Both Sexes	25 - 29	Unknown	139.0
20	Åland Islands	2000	Both Sexes	30 - 34	Total	1572.0
21	Åland Islands	2000	Both Sexes	30 - 34	Own account worker	171.0
22	Åland Islands	2000	Both Sexes	30 - 34	Employee	1247.0
23	Åland Islands	2000	Both Sexes	30 - 34	Unknown	154.0
24	Åland Islands	2000	Both Sexes	35 - 39	Total	1676.0
25	Åland Islands	2000	Both Sexes	35 - 39	Own account worker	221.0

III. Hypothesis

1. **Countries with higher internet penetration rates experience faster GDP growth due to enhanced productivity and innovation.**
    - *Rationale:* The availability of internet access facilitates knowledge dissemination, fosters innovation, and improves business efficiency, all of which contribute to economic growth.
2. **Higher internet penetration rates reduce unemployment by enabling access to remote work opportunities and online job markets.**
    - *Rationale:* Internet accessibility opens up global job markets, allowing people to access job opportunities and skill-building platforms regardless of geographical barriers.
3. **Countries with higher GDP per capita have lower unemployment rates due to better economic stability and investment in job creation.**
    - *Rationale:* Wealthier economies are more likely to invest in infrastructure and policies that foster employment opportunities, resulting in lower unemployment rates.
4. **Increased internet usage amplifies the effect of GDP growth on employment by creating opportunities in digital and knowledge-based sectors.**
    - *Rationale:* High internet usage leads to the growth of tech and service industries, which provide employment opportunities, especially in urban areas.

IV. Preprocessing

- **Data Cleansing**:
  - Dropped irrelevant or extra columns (e.g., 'Value Footnotes').
  - Converted columns to appropriate data types (e.g., 'Year' to integer).
- **Renaming Columns**: Standardized column names to make them more meaningful (e.g., renaming columns to 'Country', 'Year', 'GDP_Per_Capita').
- **Merging Datasets**:
  - Merge the GDP and Internet Usage datasets based on 'Country' and 'Year'.
  - Additional merging of the employment dataset for further analysis.
- **Subsetting Data**: Focused on specific years (2004, 2009, 2014, 2019) to analyze trends over time.
- **Handling Missing Data**: Used imputation or removal to handle missing values where appropriate.

```
gdp_df = gdp_df.rename(columns = {'Country or Area': 'Country', 'Year': 'Year', 'Value':'GDP_Per_Capita'})
gdp_df.tail(10)
#Here we changed the columns for the GDP Per Capita data frame to ‘Country’, ‘Year’, and ‘GDP_Per_Capita’
```

	Country	Year	GDP_Per_Capita
6719	Sweden	2013	48462.160003
6720	Sweden	2012	48300.842500
6721	Sweden	2011	48947.439311
6722	Sweden	2010	47791.371525
6723	Sweden	2009	45492.768175
6724	Sweden	2008	47963.493752
6725	Sweden	2007	48557.384348
6726	Sweden	2006	47292.313693
6727	Sweden	2005	45440.296749
6728	Sweden	2004	44354.391062

```
internet_df = internet_df.rename(columns = {'Country or Area': 'Country', 'Year': 'Year', 'Value':'Internet_Users_Pct'})
internet_df.head(10)
#Here we changed the columns for the Internet Users data frame to ‘Country’, ‘Year’, and ‘Internet_Users_Pct’
```

	Country	Year	Internet_Users_Pct
0	Afghanistan	2019	17.60000
1	Afghanistan	2018	16.80000
2	Afghanistan	2017	13.50000
3	Afghanistan	2016	11.00000
4	Afghanistan	2015	8.26000
5	Afghanistan	2014	7.00000
6	Afghanistan	2013	5.90000
7	Afghanistan	2012	5.45455
8	Afghanistan	2011	5.00000
9	Afghanistan	2010	4.00000

```
gdp_df['Year']=gdp_df['Year'].astype(int)

gdp_and_internet_use = pd.merge(gdp_df,internet_df,on= None, how='outer', sort=True)

#Here we merge the two DataFrames to one.
#Here we've merged all rows from each of the two DataFrames.
#Here we now call the new DataFrame gdp_and_internet_use.
```

```
# Here we view the tail of the merged dataset
gdp_and_internet_use.tail(5)
```

	Country	Year	GDP_Per_Capita	Internet_Users_Pct
7740	SÃ£o TomÃ© and Principe	2018	3987.383653	NaN
7741	SÃ£o TomÃ© and Principe	2019	4013.721897	NaN
7742	SÃ£o TomÃ© and Principe	2020	4058.674003	NaN
7743	SÃ£o TomÃ© and Principe	2021	4052.123362	NaN
7744	SÃ£o TomÃ© and Principe	2022	4012.837269	NaN

```
gdp_and_internet_use = gdp_and_internet_use[gdp_and_internet_use["Year"].isin([2004, 2009, 2014, 2019])]
gdp_and_internet_use
#Here we are subsetting the combined data frame to keep only the data for 2004, 2009, 2014 and 2019
```

	Country	Year	GDP_Per_Capita	Internet_Users_Pct
4	Afghanistan	2004	1260.060581	0.105809
9	Afghanistan	2009	1823.742614	3.550000
14	Afghanistan	2014	2144.449634	7.000000
19	Afghanistan	2019	2079.921861	17.600000
36	Africa Eastern and Southern	2004	3075.546230	NaN
...	...	...	...	...
7719	Sweden	2019	52850.569179	NaN
7726	SÃ£o TomÃ© and Principe	2004	2750.494636	NaN
7731	SÃ£o TomÃ© and Principe	2009	3187.122260	NaN
7736	SÃ£o TomÃ© and Principe	2014	3687.440493	NaN

```
gdp_df = gdp_df.drop(columns =['Value Footnotes'])
gdp_df.tail(50)
#Here we dropped the 'value footnotes' column from both datasets
```

	Country or Area	Year	Value
6679	Suriname	2020	15329.179106
6680	Suriname	2019	18449.199208
6681	Suriname	2018	18438.346668
6682	Suriname	2017	17753.060265
6683	Suriname	2016	17662.682123
6684	Suriname	2015	18768.794891
6685	Suriname	2014	19629.223247
6686	Suriname	2013	19778.306406
6687	Suriname	2012	19415.577607
6688	Suriname	2011	19111.172573
6689	Suriname	2010	18255.625820
6690	Suriname	2009	17554.265785
6691	Suriname	2008	17233.764807
6692	Suriname	2007	16735.924495
6693	Suriname	2006	16102.725520
6694	Suriname	2005	15391.999397
6695	Suriname	2004	14892.121768
6696	Suriname	2003	13910.505486
6697	Suriname	2002	13300.300155
6698	Suriname	2001	13157.618062
6699	Suriname	2000	12848.606639
6700	Suriname	1999	12812.773184
6701	Suriname	1998	13239.790702

V. Methods of Analysis

- **Questions to be Answered**:
  - Which country had the highest/lowest percentage of internet users in specific years (2004, 2009, 2014, 2019)?
  - What is the structure and content of the datasets?
  - How do you clean and prepare the datasets for analysis?
  - Are there missing or irrelevant values in the datasets?
  - What is the trend in GDP per capita over time?
  - How has internet usage evolved over time across countries?
  - What is the correlation between GDP per capita and internet usage?
  - How do GDP and internet usage vary across different regions?
  - Are there significant changes in trends over specific years (e.g., 2004, 2009, 2014, 2019)?
  - Are there any outliers in the data?
  - What patterns emerge for specific countries or regions?
  - How can key insights be visually represented?
- **Data Fields Used**:
  - For the internet usage analysis: 'Country', 'Year', 'Internet_Users_Pct'.
  - For the GDP per capita analysis: 'Country', 'Year', 'GDP_Per_Capita'.
  - For the employment analysis: 'Country', 'Year', 'Employment Rate'.
- **Methods**:
  - Sorting the datasets to find maximum and minimum values.
  - Plotting scatter plots to visualize the relationship between GDP and internet usage over time.
  - Using histograms to explore the distribution of GDP per capita and internet usage.
  - Analyzing time series trends for the top and bottom countries.
  - Correlation analysis between GDP per capita, internet usage, and employment rates.

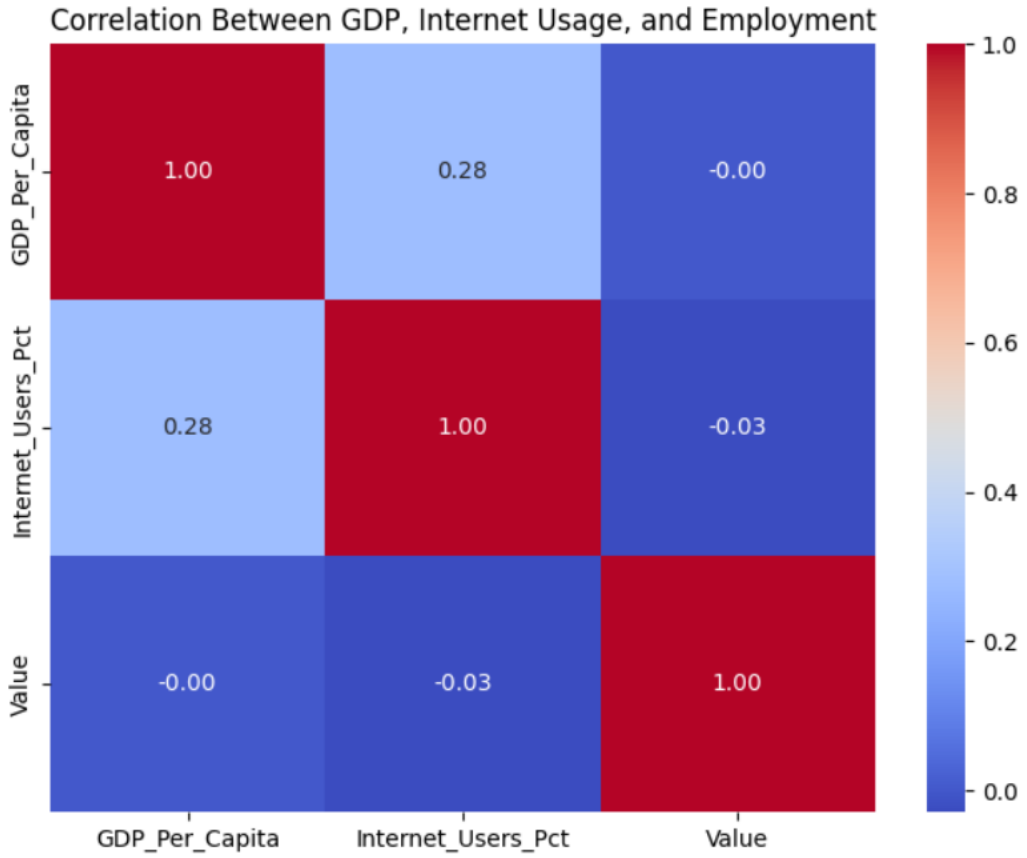
VI. Python Program Overview

- **Libraries Used**:
  - pandas: Data manipulation and merging.
  - matplotlib, seaborn: Data visualization (scatter plots, histograms, and time series plots).
  - numpy: Numerical operations.
- **Steps in the Code**:
  - Data importation and inspection.
  - Preprocessing steps (column renaming, merging, data cleansing).
  - Analyzing data (sorting, filtering, aggregating).
  - Visualizing relationships and trends using various plots.

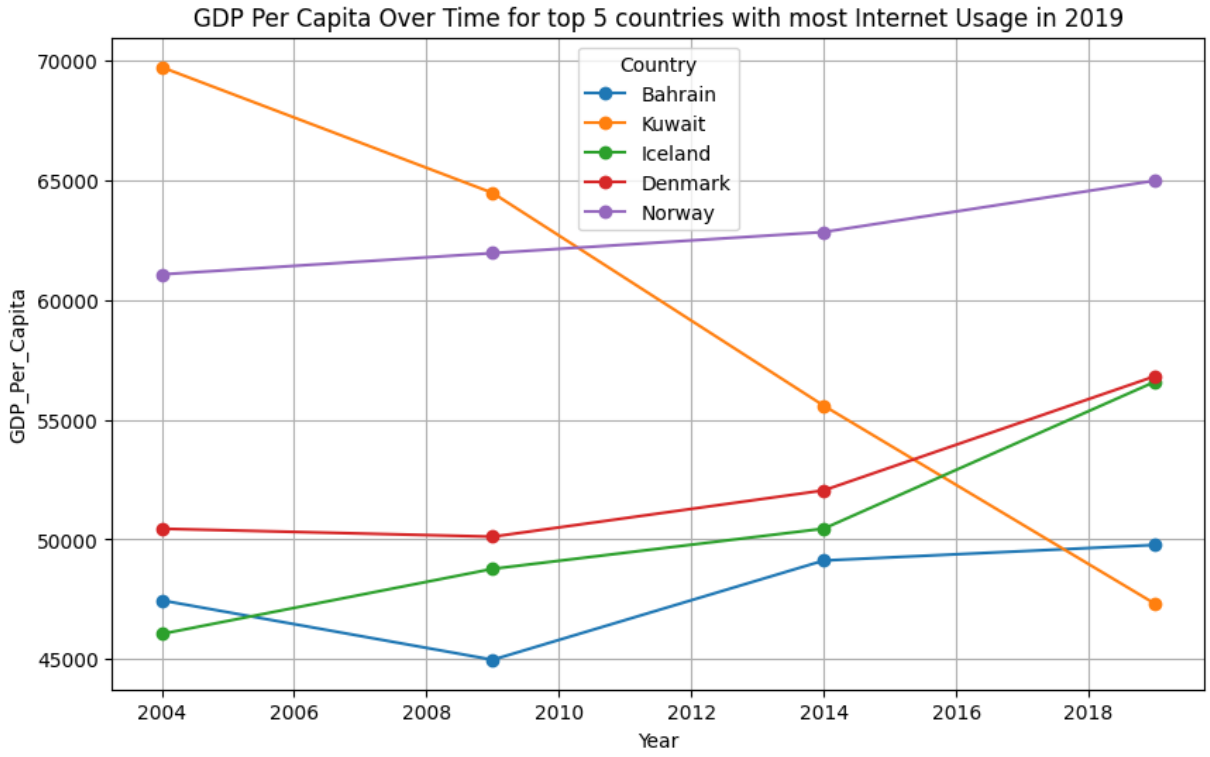
```
# Here we import the libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

%matplotlib inline
```

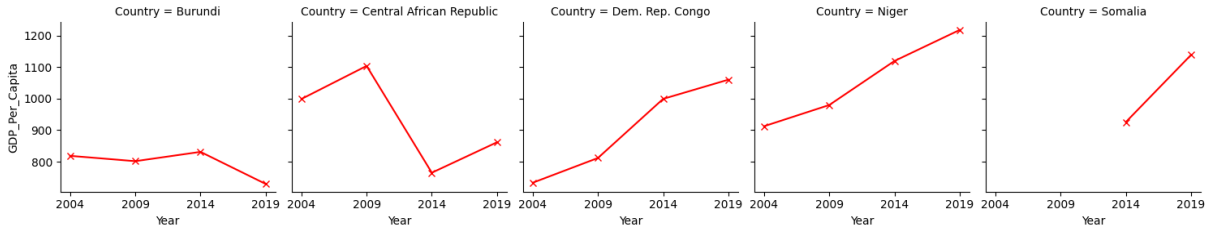
VII. Python Program Output

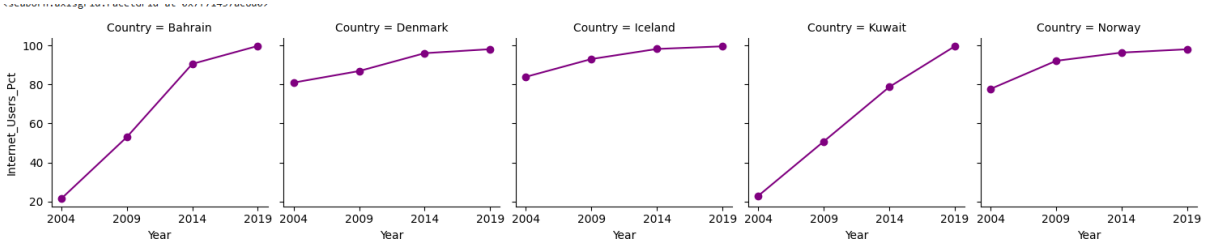
- **Descriptive Statistics**:
  - **Top Countries for Internet Usage and GDP**: Bahrain had the highest internet usage in 2019, while Malaysia had the highest GDP per capita.
  - Correlation between GDP, internet usage, and employment.

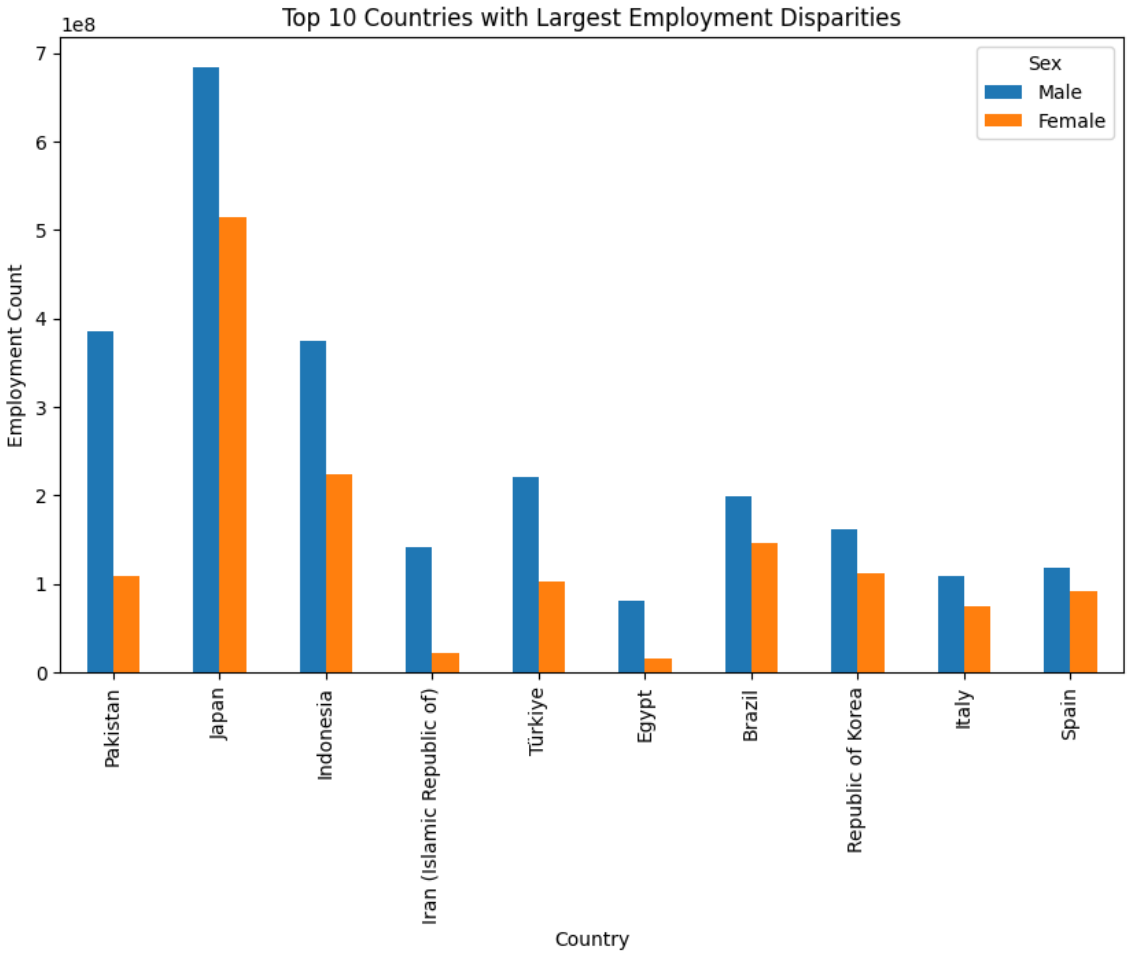


- **Visualizations**:
  - Scatter plots showing the relationship between internet usage and GDP per capita for each year (2004, 2009, 2014, 2019).
  - Histograms of GDP per capita and internet usage for 2019.
  - Time series plots of internet usage and GDP per capita for the top and bottom 5 countries in 2019.
  -


GDP Per Capita Over Time for top 5 countries with most Internet Usage in 2019
Country
Bahrain
Kuwait
Iceland
Denmark
Norway
GDP_Per_Capita
Year
70000
65000
60000
55000
50000
45000
2004
2006
2008
2010
2012
2014
2016
2018


Country = Burundi
Country = Central African Republic
Country = Dem. Rep. Congo
Country = Niger
Country = Somalia
GDP_Per_Capita
Year


Country = Bahrain
Country = Denmark
Country = Iceland
Country = Kuwait
Country = Norway
Internet_Users_Pct
Year



VIII. Team Contributions:

Regular brainstorming sessions ensured efficient task distribution and alignment of objectives. Continuous communication and teamwork helped merge different components seamlessly.

Each member brought unique strengths to the project, ensuring well-rounded and insightful analysis outcomes. Below is the task distribution for each member.

Team Member	Contribution
Bhavana	• Identifying the appropriate dataset • Data cleaning and merging and analysis • Collaborated for analysis outcomes
Varun	• Identifying the appropriate dataset • Creating data visualizations to support insights • Collaborated for analysis outcomes
Kushal	• Conducted in-depth data exploration • Created data visualizations to support insights • Collaborated for analysis outcomes
Nikita	• Conducted quality checks on datasets • Generated insights and validated results • Collaborated for analysis outcomes
Tejal	• Conducted in-depth data exploration • Generated metrics and relationships • Collaborated for analysis outcomes

IX. Conclusions

1. **Key Findings**:
   - **Internet Usage and Economic Growth**: Higher internet usage is positively correlated with increased GDP per capita, suggesting that digital connectivity contributes to economic growth.
   - **Employment and GDP Link**: Stronger job markets are often seen in countries with higher GDP per capita, highlighting the role of economic stability in job creation.
   - **Feedback Loop of Growth Factors**: There is a cyclical relationship: higher GDP enables investment in digital infrastructure, which enhances internet accessibility, leading to more job opportunities and economic stability.
   - **Digital Divide Impact**: Regions with limited internet access tend to have lower GDP per capita and higher unemployment, pointing to the need for targeted efforts to address the digital divide.
2. **Implications**:
   - **Digital Connectivity as a Growth Catalyst**: The internet plays a critical role in driving economic growth by improving access to information, enabling remote work, and fostering innovation.
   - **Interlinked Factors for Stability**: Economic stability, employment, and digital infrastructure are interconnected, suggesting that policy improvements in one area can positively influence the others.
   - **Targeted Policy Action Needed**: Bridging the digital divide is essential for ensuring equitable access to opportunities, particularly in economically disadvantaged regions.