Nour Osama Elashry 900202786
Mariam Hassan 900232748
Jana ElMorsy 900221875

DSCI1412 - FUNDAMENTALS OF DATA SCIENCE II (SPRING 2024)

Supervised by Dr. Seif Eldawlatly

The American University in Cairo

I. Introduction

How do businesses or even entire countries know how to make informed decisions that drive increases in their prosperity and advancement? The answer lies within the data. Through in-depth analysis and predictive modeling, people can find valuable insights into the complex relations between the variables of interest(s) and, therefore, base upon it a decision. In this data analytics project, we are seeking to discover the relationship between GDP (Gross Domestic Product) and technological advancement across countries over time. These variables include GDP, broadband subscriptions, income per capita, number of cellphones, number of computers, internet users, and population. Our main goal is to investigate **whether GDP, as a measure of a country's economic prosperity, can affect its level of technological advancement.** We will analyze the data in multiple ways to determine if countries with higher GDP tend to exhibit greater technological infrastructure, such as increased broadband access, higher rates of internet usage, and greater adoption of cellphones and computers. In addition, by analyzing historical trends and patterns in the data, we can develop predictive models that estimate how the GDP influences the trajectory of technological development in different countries in the future. Using this, we can further validate our findings supporting the relationship between GDP and technological advancement. With such in-depth analysis, an accurate answer to our question is to be found.

II. Methodology

To begin with, we needed to go through all the datasets to understand it more and, if met by any challenges, to cleanse the data to fit what we are aiming for. We came upon some challenges such as the following.

- **Identifying problems within the data:**

Problem #1: Different variation in the years covered by each dataset resulted in inconsistency and made direct comparison challenging.

Problem #2: Not all the datasets contained the same countries, leading to incomplete data coverage.

Problem #3: Encountered multiple occurrences of missing data points or values that had not yet been calculated.

Problem#4: Encountered multiple occurrences of values represented as strings, such as "14K" or "8M", which required conversion into numerical values for analysis. Python couldn't read them as values, leading to missing data.

- **Solving the Problems:**

Solution for Problem #1:

- We manually examined that all the datasets start and end at a different year.

GDP 1801-2013 Internet users 1960-2020 Population 1800-2100

Broadband 1998-2020 Personal Computers 1990-2006

Income per capita 1800-2050 Cell Phones 1960-2020

- Therefore, in order for the analysis to be as accurate as it can be, we took the common years, which are from 1998 till 2006.

Solution for Problem #2:

- To make it easier for us, we developed a Python code to automate the process of identifying common countries across all datasets.

```python
import pandas as pd
filenames = [
    'broadband.xlsx',
    'cellphones.xlsx',
    'computers.xlsx',
    'gdp.xlsx',
    'internetusers.xlsx',
    'percapita.xlsx',
    'population.xlsx'
]
dfs = []

for filename in filenames:
    dfs.append(pd.read_excel(filename))

common = set(dfs[0]['country']) #the first DataFrame
for df in dfs[1:]:
    common = common.intersection(df['country'])

# Print common countries
print("Common countries:", list(common))
num_common_countries = len(common)
print(num_common_countries)
```

```
Common countries: ['Indonesia', 'Sao Tome and Principe', 'Kyrgyz Republic',
184
```

The output gave out 184 countries.

Solution for Problem #3:

- We decided to use the regression model code on python in order to get as much of an accurate answer as possible using the following code.

```python
from sklearn.linear_model import LinearRegression

years = [[2001], [2002], [2003], [2004], [2005]]
values = [1150, 3400, 4000, 4800, 5000]

model = LinearRegression()

model.fit(years, values)

year2006 = [[2006]]
value2006 = model.predict(year2006)[0]

print("Estimated value for 2006:", value2006)

```

```
Estimated value for 2006: 6400.0
```

P.s. The other values were predicted with the same code but with different values and years.

https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html
https://realpython.com/linear-regression-in-python/

The highlighted values are the ones that were predicted using the code.

	country	1998	1999	2000	2001	2002	2003	2004	2005	2006
2	Benin	7000	9000	10000	11000	15000	25800	30000	32000	50000
3	Lesotho				1150	3400	4000	4800	5000	6400
4	Madagascar	25000	30000	35000	40000	70000	80000	91000	102000	113178.5
5	Mozambique	40000	50000	60000	70000	82000	96000	112000	283000	213285.7
6	Sudan	55000	85000	100000	115000	200000	348000	606000	3250000	4240000

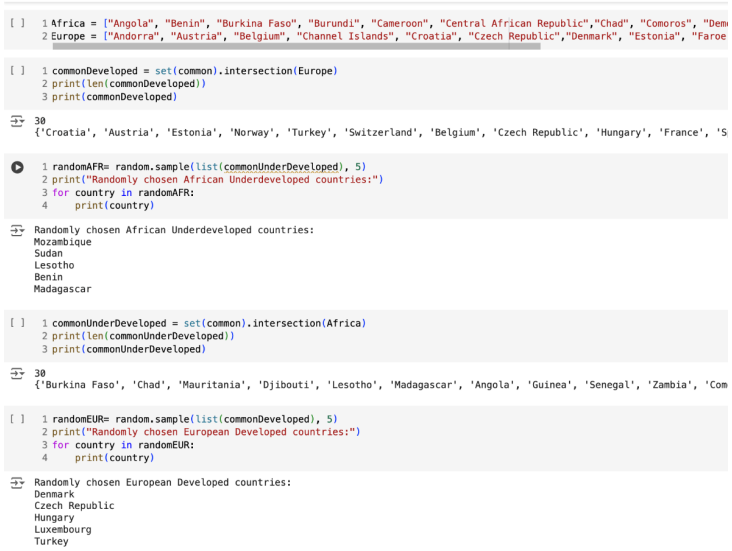
- **There was a need for us to categorize countries based on their level of development to facilitate analysis**; therefore, we decided to use some external sources to help us determine which country belongs to which category from those two: underdeveloped and developed countries. Why ?
- Categorizing countries allows us to see patterns and trends in the relationship between GDP and technology characteristics within each group. For example, high GDP may be associated with increased access to broadband and internet usage in developed countries, whereas the relationship may not be as strong in undeveloped countries.
- Furthermore, we wanted to focus on only one continent, the continent with the highest probability of having either developed or underdeveloped countries and the continent with the most developed or underdeveloped countries.

We concluded that:
Underdeveloped Countries: https://www.un.org/ohrlls/content/list-ldcs
Probability of a country being underdeveloped in African countries is **34/54**
(underdeveloped African countries /Total number of African countries).
Probability that the underdeveloped country is an African country is **34/45**
(underdeveloped African countries /Total number of underdeveloped countries).

Developed Countries: https://dcm.ffclrp.usp.br/isaac/List-countries.pdf
Probability of a country being developed in European countries is **37/50**
(developed European countries /Total number of European countries).
Probability that the developed country is an European country is **37/79**
(developed European countries /Total number of developed countries).

Clearly, **Africa and Europe** are the variables of interest here. Therefore, we will need to eliminate any countries from any other continent from our data analytics.

- Since some countries that were in the online lists weren't available in the common countries, we decided to then filter it out and only pick the common countries (also excluding any country from any other continent in the process) between the 184 countries and the online lists.
- After selecting the common countries, it got reduced to 30 countries for each of the categories. We reduce our selection of the countries we are going to examine to random 5 samples from each in order for us to be able to read the data visualization and have a more in-depth analysis of the variables and their relations.

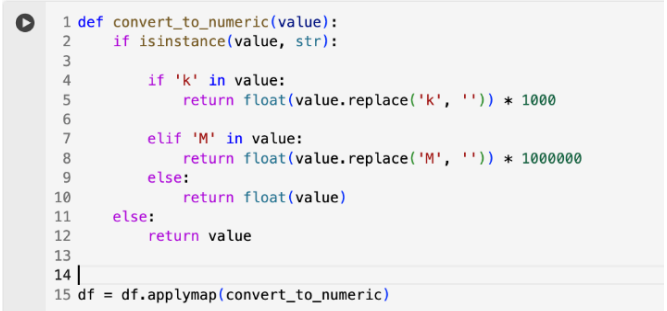


```
[ ]  1 Africa = ["Angola", "Benin", "Burkina Faso", "Burundi", "Cameroon", "Central African Republic","Chad", "Comoros", "Dem
     2 Europe = ["Andorra", "Austria", "Belgium", "Channel Islands", "Croatia", "Czech Republic","Denmark", "Estonia", "Faroe

[ ]  1 commonDeveloped = set(common).intersection(Europe)
     2 print(len(commonDeveloped))
     3 print(commonDeveloped)

30
{'Croatia', 'Austria', 'Estonia', 'Norway', 'Turkey', 'Switzerland', 'Belgium', 'Czech Republic', 'Hungary', 'France', 'S

     1 randomAFR= random.sample(list(commonUnderDeveloped), 5)
     2 print("Randomly chosen African Underdeveloped countries:")
     3 for country in randomAFR:
     4     print(country)

Randomly chosen African Underdeveloped countries:
Mozambique
Sudan
Lesotho
Benin
Madagascar

[ ]  1 commonUnderDeveloped = set(common).intersection(Africa)
     2 print(len(commonUnderDeveloped))
     3 print(commonUnderDeveloped)

30
{'Burkina Faso', 'Chad', 'Mauritania', 'Djibouti', 'Lesotho', 'Madagascar', 'Angola', 'Guinea', 'Senegal', 'Zambia', 'Com

[ ]  1 randomEUR= random.sample(list(commonDeveloped), 5)
     2 print("Randomly chosen European Developed countries:")
     3 for country in randomEUR:
     4     print(country)

Randomly chosen European Developed countries:
Denmark
Czech Republic
Hungary
Luxembourg
Turkey
```

Therefore, we have two types of datasets: underdeveloped and developed countries. Each having 5 random samples, ranging from 1998 to 2006.

Solution for Problem #4:

- We developed a Python code capable of transforming string representations of numerical values into their corresponding actual values.
- Then, after permanently changing the dataset itself, we use a function to save the new csv and replace this csv as the new dataset we will work on instead of the old one.



```
def convert_to_numeric(value):
    if isinstance(value, str):

        if 'k' in value:
            return float(value.replace('k', '')) * 1000

        elif 'M' in value:
            return float(value.replace('M', '')) * 1000000
        else:
            return float(value)
    else:
        return value


df = df.applymap(convert_to_numeric)
```

Folder of the final datasets:
https://drive.google.com/drive/folders/1ZAgtSWB-vdnuvHd5KfRUbaVdR-7JSgLs?usp=sharing

III. Data Description

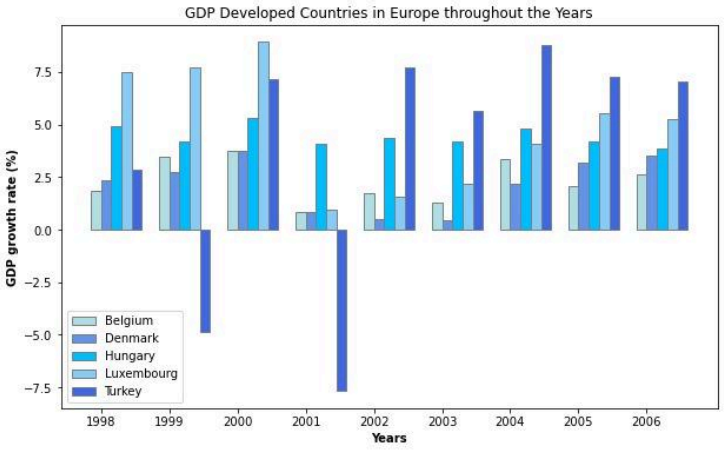
The dataset we are analyzing in this report is divided up into seven separate excel files. Each file contains numerical information about the changes incurred upon each of the seven variables over a time period that extends from the year 1960 to the year 2020. Further, the numerical data for each variable is recorded for all 215 countries. Hence, each variable is represented in a table where the years are the columns and the countries are the observations. These variables are:

1. GDP

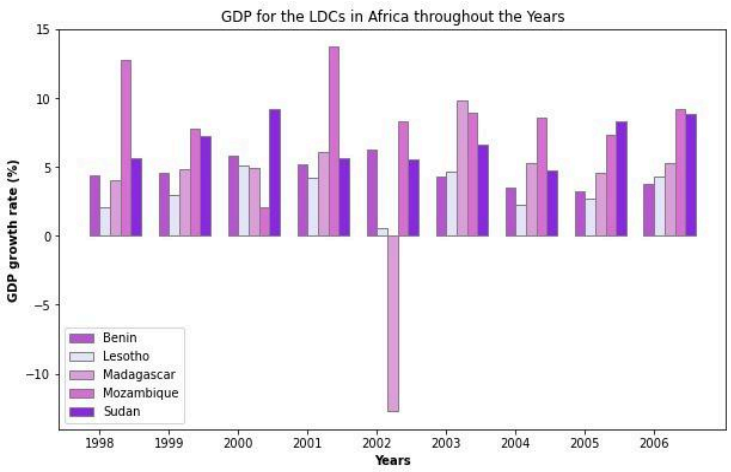
https://data.oecd.org/gdp/gross-domestic-product-gdp.htm#:~:text=Gross%20domestic%20product%20(GDP)%20is,and%20services%20(less%20imports).
https://www.britannica.com/money/gross-domestic-product#

According to the website linked above, Gross Domestic Product (i.e. GDP for short) is the monetary value of all finished goods and services produced within a country during a specific period of time. It is used as a measure of the economic status of countries. Since the final users of goods and services are households, businesses, and the government, the formula used to calculate GDP is: Consumption + Investment + Government Expenditure + Net Exports. Of those products and services, there are, of course, cellular services, broadband services, internet services, and technological gadgets like computers and cellphones.

- Developed Countries:
  The bar chart on the right shows the trends followed by each country in terms of its GDP from 1998 to 2006.
  - Turkey shows substantial fluctuations in GDP, being the only country whose economic status actually drops below 0%. The highest decrease that it reached was approximately 14% decrease in GDP between 2001 and 2002.
  - Luxembourg shows the greatest increase in GDP from 1998 to 2000, with a maximum of around 8% increase overall.



- Underdeveloped Countries:
  This bar chart shows the trends of GDP fluctuations from 1998 to 2006 for all our five chosen underdeveloped countries.

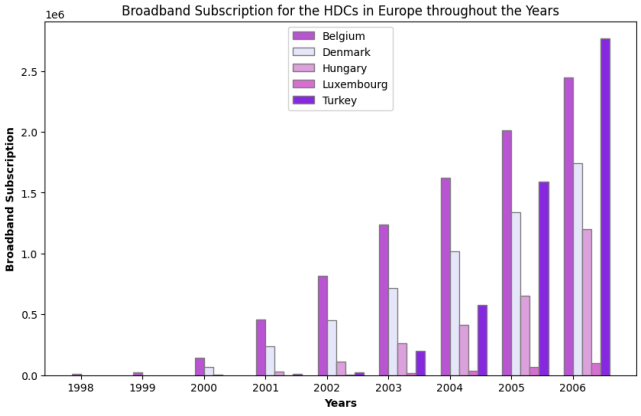


- The only country showing a decrease in GDP is Madagascar, with roughly a 20% drop in GDP from 2001 to 2002.
- Mozambique shows two peaks in the graph. The first peak was in 1998 where GDP increased by around 12% from the year before. The second peak was in 2001, where the growth in GDP reached around 14%.
- Sudan's GDP increases at an approximately steady rate throughout.

2. Broadband Subscription

- Developed Countries:

  The histogram on the right, entitled "Broadband Subscription for the HDCs in Europe Throughout the Years", showcases the trends that are followed by Broadband subscription over the time period that extends from 1998 to 2006 for all five of our chosen developed countries.



- As is evident from the visualization, Turkey reached the highest number of Broadband Subscription in 2006, with a maximum that goes to a little over 2.8 million subscribers. This is despite the fact that it remained below Belgium from 1998 to 2005.
- Luxembourg consistently and remarkably remains the lowest in terms of number of subscribers, reaching a maximum that is well beneath 0.5 million in 2005

Sample Mean Broadband (EUR)	country
Belgium	7.063333e+06

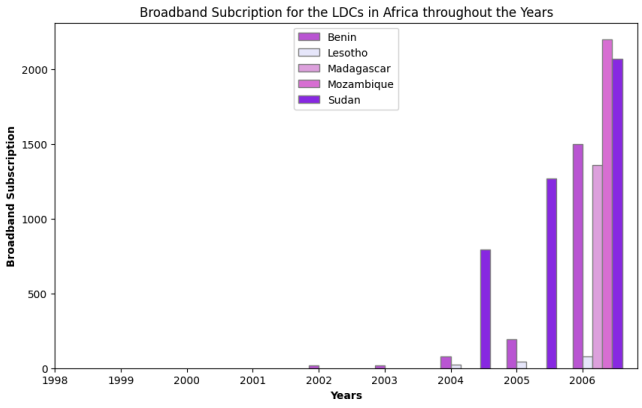
Denmark	4.175556e+06
Hungary	5.955556e+06
Luxembourg	4.174444e+05
Turkey	2.550333e+07

Standard Error Broadband (EUR)	country
Belgium	2.905297e+06
Denmark	1.320001e+06
Hungary	3.386798e+06
Luxembourg	1.786100e+05
Turkey	1.609147e+07

Sample Variance Broadband (EUR)	country
Belgium	8.440750e+12
Denmark	1.742403e+12
Hungary	1.147040e+13
Luxembourg	3.190153e+10
Turkey	2.589356e+14

- Underdeveloped Countries:

  The next histogram, entitled "Broadband Subscription for the LDCs in Africa Throughout the Years", showcases the trends that are followed by all five of our chosen underdeveloped countries in terms of broadband subscription from 1998 to 2006.



- There are barely any remarkable observations from 1998 to 2003.
- Benin begins to show a clear increase in the total number of broadband subscriptions from 2004, with a total number of roughly 100 subscriptions, ending at around 1500 subscriptions by 2006.
- The country with the highest number of broadband subscriptions by the end of the studied time period is Mozambique, reaching almost 2500 total subscriptions in 2006.

Sample Mean Broadband (AFR)	country
Benin	307117.777778
Lesotho	129825.555556
Madagascar	288966.666667
Mozambique	606892.222222
Sudan	831288.888889

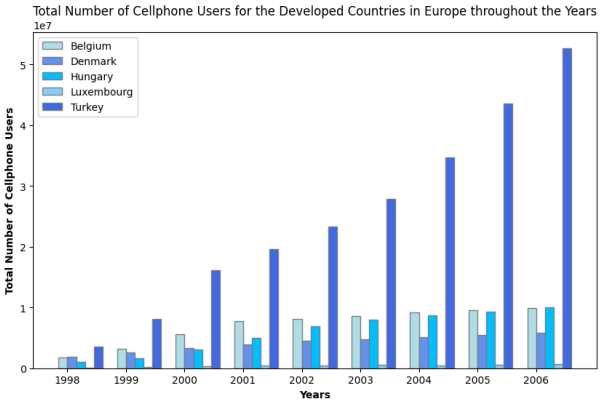
Sample Variance Broadband (AFR)	country
Benin	1.203825e+11
Lesotho	1.453301e+10
Madagascar	1.071496e+11
Mozambique	6.491670e+11
Sudan	2.421094e+12

Standard Error Broadband (AFR)	country
Benin	3.469618e+05
Lesotho	1.205530e+05
Madagascar	3.273371e+05

Mozambique	8.057090e+05
Sudan	1.555987e+06

3. Cellphone Users

- Developed Countries:



This histogram, entitled "Total Number of Cellphone Users for the Developed Countries in Europe Throughout the Years", demonstrates the changes in the number of cellphone users for each of our chosen developed countries from 1998 to 2006.

- It is clear from the graph that Turkey shows a roughly linear / uniform increase in its total number of cellphone users, starting a little below 5 million in 1998, and working up steadily to around 70 million cellphone users in 2006.
- Denmark and Belgium seem to slightly plateau around the years 2001-2003, with Belgium reaching around 10 million cellphone users by 2006, which is around the same as the number of cellphone users that Hungary has by 2006 as well. Denmark rests at a little over 5 million by 2006.
- Luxembourg shows barely any noticeable or notable increase throughout the studied time period.

Sample Mean Cellphone Users (EUR)	country
Belgium	974766.666667
Denmark	796285.714286
Hungary	381828.571429
Luxembourg	37991.666667
Turkey	861516.666667

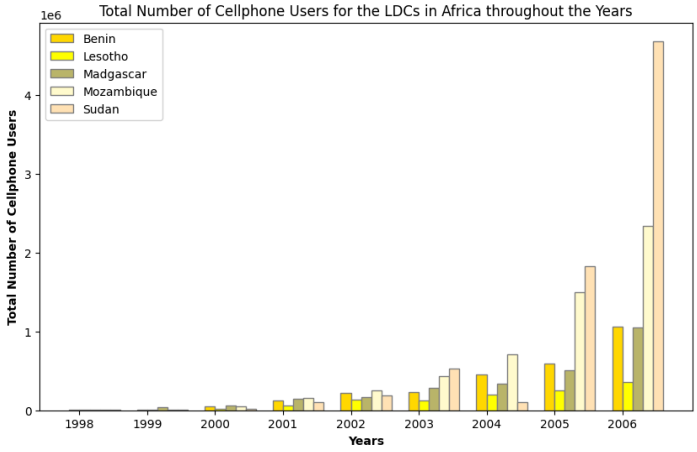
Sample Variance Cellphone Users(EUR)	Country
Belgium	8.187848e+11
Denmark	3.675269e+11
Hungary	1.827480e+11
Luxembourg	1.527981e+09
Turkey	1.224397e+12

Standard Error Cellphone Users (EUR)	Country
Belgium	9.048673e+05
Denmark	6.062400e+05
Hungary	4.274904e+05
Luxembourg	3.908940e+04
Turkey	1.106525e+06

- Underdeveloped Countries:

  The graph on the right, entitled "Total Number of Cellphone Users for Underdeveloped Countries in Africa throughout the Years", shows the changes in number of cellphone users shown by each country throughout the studied period.

- Observe that the country with the highest number of cellphone users by 2006 is Sudan, reaching a maximum of roughly 4500000 cellphone users. The increase



happens suddenly after having a total number of cellphone users less than that of Lesotho from 1998 to 2005.

- The countries showing the showest and the least increase in their total number of cellphone users are Lesotho and Madagascar.

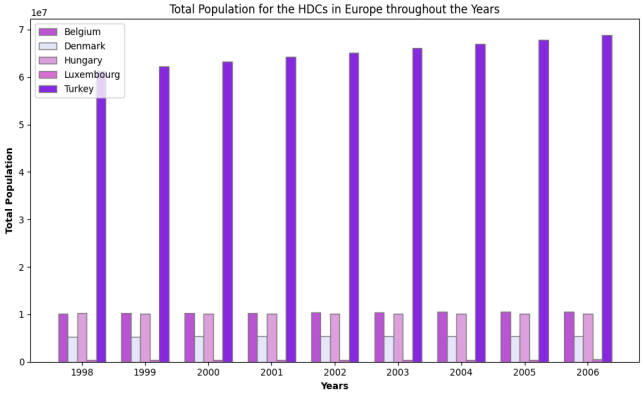
Sample Mean Cellphone Users (AFR)	Country
Benin	363.800000
Lesotho	49.333333
Madagascar	1360.000000
Mozambique	2200.000000
Sudan	1377.666667

Sample Variance Cellphone Users (AFR)	Country
Benin	408526.700000
Lesotho	884.333333
Madagascar	NaN
Mozambique	NaN
Sudan	416376.333333

Standard Error Cellphone Users (AFR)	Country
Benin	639.160934
Lesotho	29.737743

Madagascar	NaN
Mozambique	NaN
Sudan	645.272294

4. Population



- Developed Countries:
  This bar chart, entitled "Total Population for the HDCs in Europe Throughout the Years", shows the changes in population density/number incurred by each of our five chosen developed countries.
- Turkey persistently shows a relatively higher and a relatively stable population number, reaching a number that is a little below 70 billion in 2006.
- The lowest country in terms of population is Denmark, with a roughly stable population of less than 10 million throughout.

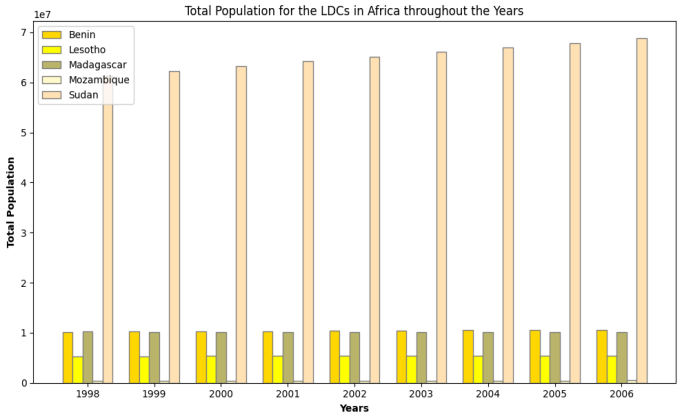
Sample Mean Population (EUR)	Country
Belgium	1.038889e+07
Denmark	5.371111e+06
Hungary	1.016667e+07
Luxembourg	4.444444e+05
Turkey	6.510000e+07

Sample Variance Population (EUR)	Country
Belgium	1.611111e+10
Denmark	2.136111e+09

Hungary	5.000000e+09
Luxembourg	1.655278e+08
Turkey	6.605000e+12

Standard Error Population (EUR)	Country
Belgium	1.269296e+05
Denmark	4.621808e+04
Hungary	7.071068e+04
Luxembourg	1.286576e+04
Turkey	2.570019e+06

- Underdeveloped Countries:
  The next bar chart, entitled “Total Population for the LDCs in Africa Throughout the Years”, shows the changes in total population number in each of our five chosen underdeveloped countries during the chosen period of study.
- The country with the highest population number all throughout is Sudan, going from roughly 60 million in 1998 to around 65 million in 2006.
- There are no remarkable changes to note for Mozambique.



Sample Mean Population (AFR)	Country
Benin	7.316667e+06
Lesotho	2.015556e+06
Madagascar	1.681111e+07

Mozambique	1.883333e+07
Sudan	2.876667e+07

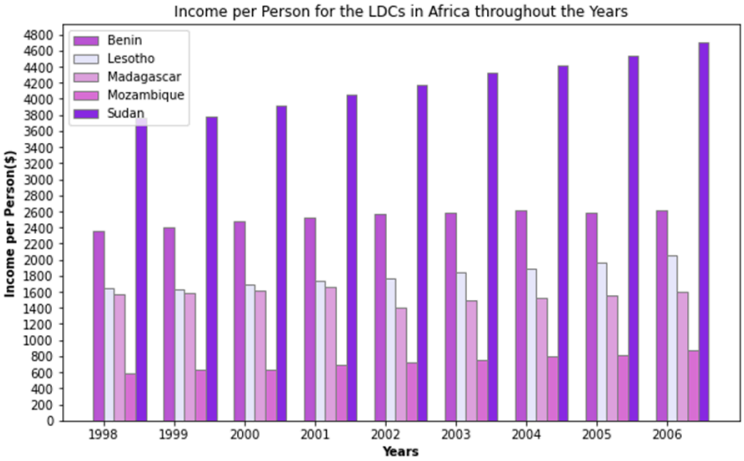
Sample Variance Population (AFR)	Country
Benin	408526.700000
Lesotho	884.333333
Madagascar	NaN
Mozambique	NaN
Sudan	416376.333333

Standard Error Population (AFR)	Country
Benin	639.160934
Lesotho	29.737743
Madagascar	NaN
Mozambique	NaN
Sudan	645.272294

5. Income per person (per capita)

- Underdeveloped Countries:

➔ The graph describes the income per person ($) of the least developed countries in Africa, Benin, Lesotho, Madagascar, Mozambique, and Sudan – from 1998 to 2006.



- ➔ Overall, the trend is increasing slightly in all of the countries at a steady rate except for Madagascar which fluctuates.
- ➔ Sudan contains the highest income per person throughout the years in comparison to the other countries. In all the years Benin is the second place, Lesotho is the third place, Madagascar is the fourth place, and finally Mozambique is always the last place.
- ➔ We can see that the income per person in Sudan increased from 3770$ in 1998 to 4700$ in 2006. While Mozambique as the least income per person increased from 588$ in 1998 to 876$ in 2006.
- ➔ Madagascar increased at a very low pace from 1570$ in 1998 to 1660$ in 2001 then it decreased to 1410$ the next year and after that it kept increasing until reaching 1600$ in 2006.

- ➔ We can observe from the bar chart that Lesotho and Madagascar are very close to each other regarding the income per person. From the table, we can see that it proves our analysis that Sudan is the highest country in the income per person with a significant difference compared to the other counties. While Mozambique is the least country in the income per person.

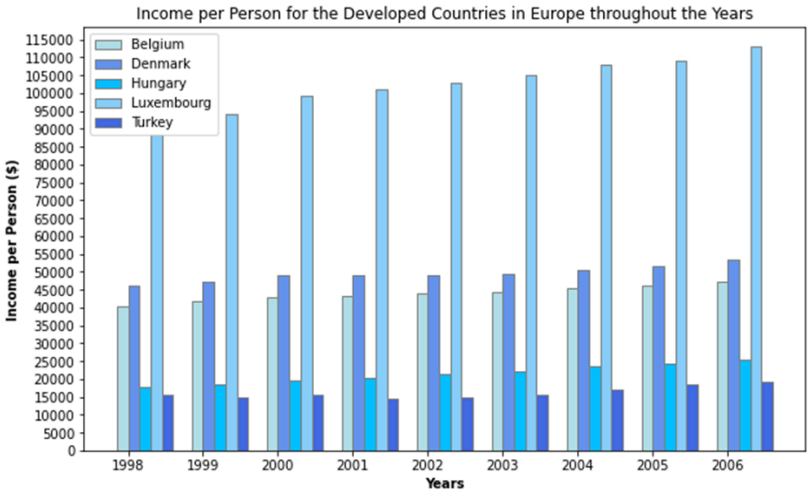
Country	**Sample Mean ($\bar{x}$)**
Benin	2525.555556
Lesotho	1803.333333
Madagascar	1558.888889
Mozambique	724.666667
Sudan	4185.555556

From the standard deviation, we can conclude that Sudan's income per person values experience huge changes among the years. Whereas Madagascar experienced the least change between its income per person values over the years.
One reason for that is during the period of 2001-2005, Madagascar experienced significant changes in its economic landscape that impacted growth, inequality, and poverty. This period was marked by political instability, economic reforms, and various international interventions that shaped the socio-economic conditions of the country.

Country	Standard Deviation (σ)
Benin	90.154189
Lesotho	149.499164
Madagascar	72.533517
Mozambique	96.147023
Sudan	334.443983

https://documents1.worldbank.org/curated/es/815371468056068786/pdf/517000NWP0wp1110Box342044B01PUBLIC1.pdf

- Developed Countries:



➔ The graph describes the income per person ($) of the developed countries in Europe – Belgium, Denmark, Hungary, Luxembourg, and Turkey – from 1998 to 2006.

➔ The overall trend of the developed countries is increasing the income per person all over the years except for Turkey in which the income per person keeps fluctuating.

➔ Luxembourg is way ahead than the other counties in the income per person as its average income per person in 1998 was 88.2K $ and in 2006 was 113K $. Compared to Turkey, as the least country, which was 15.4K in 1998 and 19.3K in 2006.

➔ For the other countries: In the years from 1998 to 2005, Belgium increased from 40.3K to 47.2k, Denmark increased from 45.1k to 53.3k, and Hungary increased from 17.9k to 25.4k.

➔ As Belgium and Denmark have similar income per person, we can conclude that their economic states are quite similar. Also, Hungary and Turkey have approximate income per person, so they may have similar economic status.

➔ The average mean of the income per person of Belgium, Denmark and Hungary, Turkey are similar which strengthen our analysis. Also, the average of Luxembourg is far more than the others which also proves our analysis.

Country	Sample Mean ($\bar{x}$)
Belgium	43922.222222
Denmark	49455.555556
Hungary	21422.222222
Luxembourg	102288.888889
Turkey	16122.222222

➔ From the standard deviation tables we can see that Belgium and Denmark have very close values of the measure of spread which indicates that they grow at approximate economic rates as we stated before. However, we can see that Hungary and Turkey have very different measure of spread, so they do not grow at the same economic rate regarding the income per person which falsify our analysis.

Country	Variance (σ^2)
Belgium	4,839,444
Denmark	4,535,278
Hungary	6,939,444
Luxembourg	59,548,610
Turkey	2,929,444

Country	Standard Deviation (σ)
Belgium	2199.873734
Denmark	2129.619163
Hungary	2634.282529
Luxembourg	7716.774657
Turkey	1711.561990

6. Internet Users

Developed Countries:

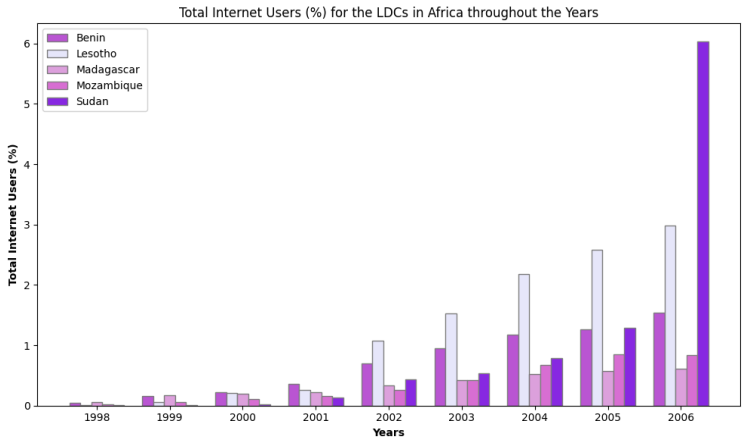
The bar chart presented on the right, entitled "Total Internet Users (%) for the HDCs in Europe throughout the Years", displays the trends of each country with respect to its Internet Consumption during the studied time period.

- Denmark shows a roughly linear and consistent percentage increase in its internet usage, with a little over 20% in 1998, ending at a little over 90% in 2006.
- The country with the smallest percentage increase relative to all other countries is Turkey, starting at less than 10% in 1998, and ending at a little less than 20% in 2006.

Underdeveloped Countries:

The next bar chart deals with the percentage of internet users in our five chosen underdeveloped African countries:

- It is clear from the graph that the percentages are much lower here than



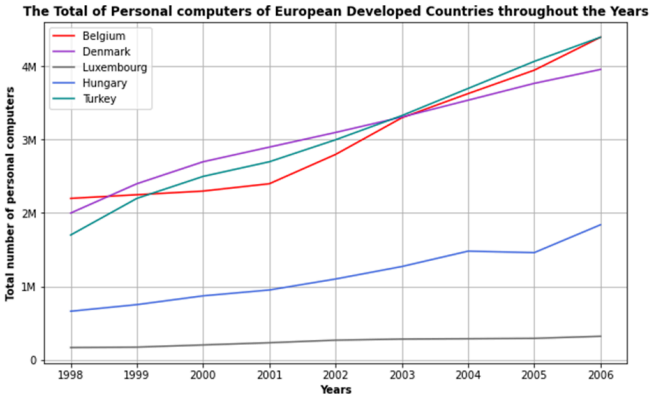
their counterparts in Europe, with the highest percentage being around 6% and scored bySudan in 2006.

- The second-ranked country from in the underdeveloped criterion is Lesotho, with a roughly linear increase that goes up to around 3%.

7. Personal Computers

Developed Countries

- This line chart represents the total number of personal computers of the developed countries in Europe over the years from 1998 to 2006.
- The overall trend is increasing in all the countries but we can observe that Luxembourg and Hungary are way lower than the other countries.
- Belgium, Denmark and Turkey have almost the same trend, they started from about 2M personal computers and doubled to about 4M in the range of 8 years (1998-2006).
- Luxembourg was almost steady in this period where it began with 165k personal computers and ended with 318k personal computers which is a very slight change compared to the other countries which increased in millions.
- Although Luxembourg has the highest income per person in all years, it has the lower number of personal computers which is very unexpected. But this is because Luxembourg has the lowest population over the years in which its population ranges between 426k and 466k while the other countries' population are in millions.

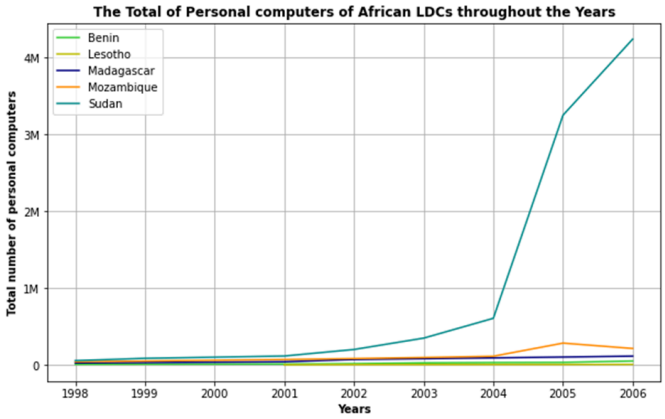


Country	Sample Mean ($\bar{x}$)
Belgium	3,025,556
Denmark	3,075,556
Hungary	1,153,333
Luxembourg	244,778
Turkey	3,066,667

Country	Standard Deviation (σ)
Belgium	824319.584735
Denmark	644284.702424
Hungary	389935.892167
Luxembourg	55737.729093
Turkey	889789.300902

Underdeveloped Countries:

- This line chart represents the total number of personal computers of the underdeveloped countries of Africa over the years from 1998 to 2006.



- The overall trend in all the countries is approximately constant except for Sudan which had a dramatic increase from 2004 to 2006 as it jumped from 606k to 4.24M in just 3 years. Although the population change in Sudan was increasing steadily over the years, thus we can conclude that Sudan is the most developed country among the underdeveloping countries.

- Benin, Lesotho, Madagascar, and Mozambique can be categorized as one group in the total number of personal computers as they somehow follow the same trend. While Sudan is far away from them, so we can put it in another group itself.

- Lesotho is the lowest country as it has the least population number where its maximum population is 2.04M while the other countries range between 6M and 32M.

Country	Sample Mean ($\bar{x}$)
Benin	21088.888889
Lesotho	4125.000000
Madagascar	65130.944444
Mozambique	111809.522222

Sudan	999888.888889

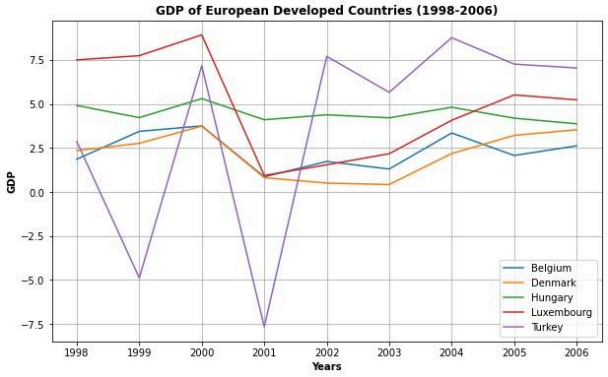
Country	Standard Deviation (σ)
Benin	14,417.74
Lesotho	1,777.006
Madagascar	33,477.63
Mozambique	82,273.53
Sudan	1,585,241

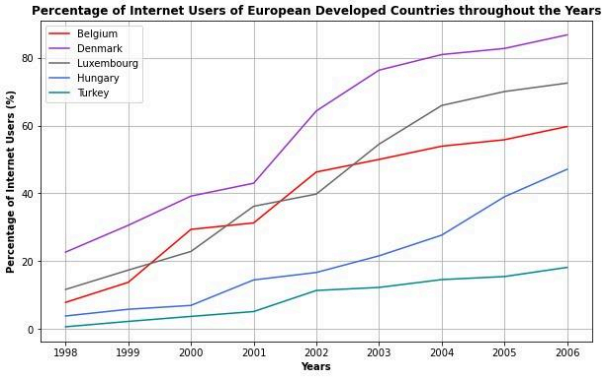
IV. Relations and Analysis

1. GDP & Internet Users

We draw upon our single variable analysis demonstrated above in order to make the below comparison between GDP and Internet Users.

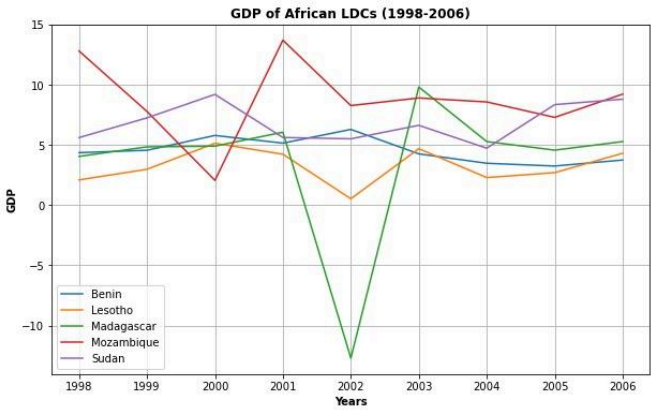
A. European Developed Countries:

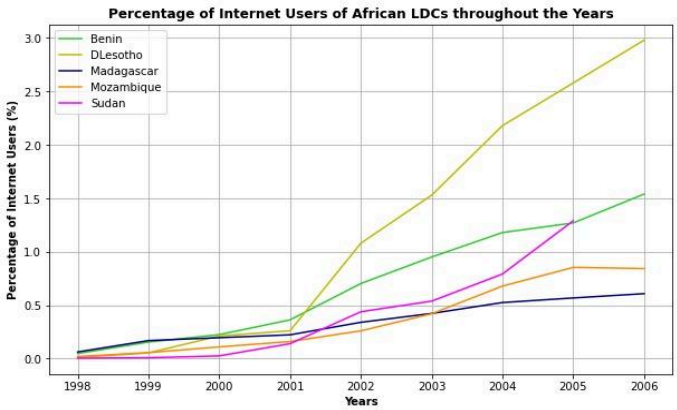




At face value, it seems that the number of internet users remains unaffected by the trends followed by GDP. In particular, we see that even though the GDP in Turkey drops by almost 14% from 2000 to 2001, the percentage of internet users increases by around 5% during the same time. This shows that there is a weak correlation between the two given variables, especially that, as per the above mentioned equation of GDP, there are other factors such as Net Exports and Investments which affect GDP. That being said, the general trend is that each European country enjoyed a steady (neither acceleration or deceleration) in GDP, which in some way or another might have accounted for the steady rise in internet users.

B. African Underdeveloped Countries

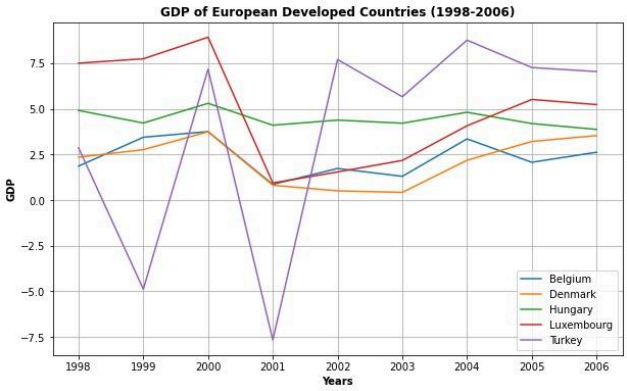


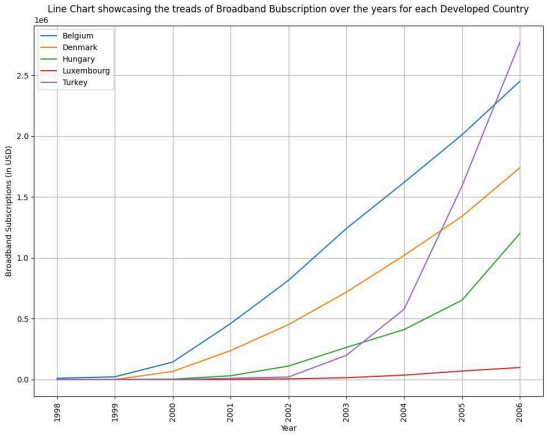


Similar to the developed countries, the trends followed by the number of internet users by underdeveloped African countries seems to be unaffected by GDP. For example, even though Madagascar experienced a sharp decrease in GDP by around 18% from 2001 to 2002, the number of internet users, the percentage of internet users remains steady (does not rise or fall unexpectedly). In particular, we see that Lesotho experienced a sharp and roughly linear increase in number of internet users from 2001 to 2006, reaching a maximum of 3% increase from only 0.2% in 1998, despite the fluctuations that happened in GDP, whence there was a 4% fall from 2001 to 2002, followed by a 4% rise from 2002 to 2003, followed by a steady GDP growth.

2. GDP & Broadband Subscriptions

   Using our single variable analysis of both the GDP & Broadband subscriptions, we can already dissect some similar trends that both graphs are following; therefore, we can conclude some relation between both of them.





A. European Developed Countries

   1. Despite experiencing fluctuations and periods of both positive and negative growth in GDP, both **Turkey** and **Luxembourg** demonstrate a steady increase in broadband subscriptions. They both have the most prominent downfall in the years of 2001 and recovery afterwards. For instance, in Luxembourg, during the period from 2000 to 2001, despite a notable decline in GDP from 3.75 to 0.859 (-77%), broadband subscriptions went from 144,000 to 460,000(+219%). One might think that there might be a negative correlation between the two variables;but

when its GDP went from 2.17 to 4.07(+87%) within 2003 and 2004, the broadband subscriptions went from 264000 to 411000 (+55%).

2. Turkey also had the most prominent recovery within the period of 2001 and 2002, where its GDP went from −7.66 to 7.7. Based on the theory made in Hungary, the broadband subscriptions should decrease. But this is not the case as it goes from 31400 to 111000. It still increases, if not more rapidly.
3. This shows that economic instability does not significantly hinder the steady growth of broadband adoption in these countries.
4. When analyzing countries like **Hungary** and **Denmark**, it becomes evident that subtle fluctuations in GDP also do not have a substantial impact on broadband subscriptions. For example, Hungary's GDP went from 5.3 to 4.1 within the period between 2000 and 2001. Within the same period, its broadband subscriptions went from 3400 to 31400.
5. Similarly, in **Denmark**, between 2000 and 2001, amidst a significant decline in GDP from 3.74 to 0.811, and until 2003, the GDP went to 0.419, broadband subscriptions increased dramatically from 238000 to 718000 regardless.

Based on the article of https://www.ecb.europa.eu/press/key/date/2003/html/sp031113.en.html#:~:text=The%20euro%20has%20already%20been,used%20during%20these%20three%20years.

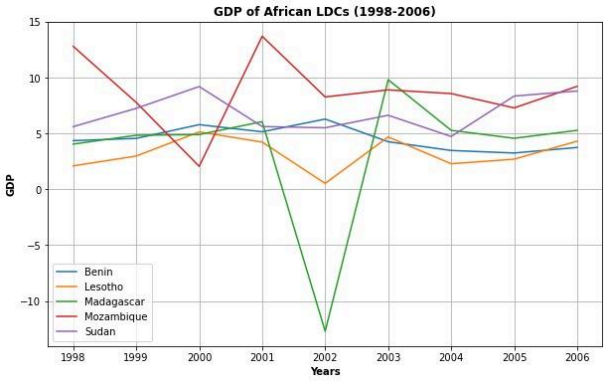
The period leading up to the full adoption of the euro, particularly from 1999 to 2002, was marked by substantial economic adjustments. Businesses, financial institutions, and consumers had to adapt to a new currency system, which might have created short-term economic uncertainties and disruptions, such as the periods within 1998 and 1999 and periods within 2000 and 2001.

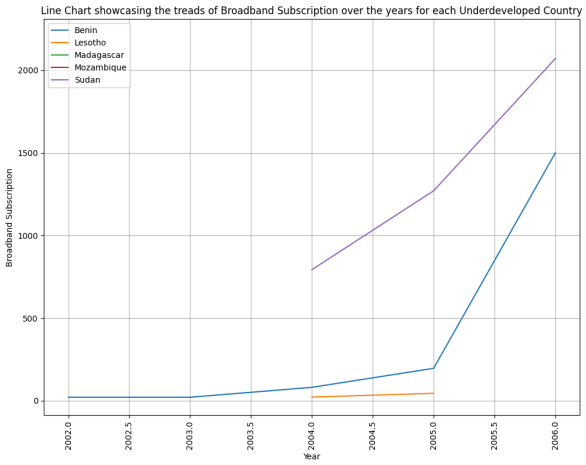
Other than that, the exchange of the old national banknotes and coins of the 12 countries of the euro area for the euro was a huge undertaking requiring enormous organisational, logistical,

technical and economic efforts. Despite any initial concerns, its success far exceeded expectations. It can be proven that within the period of 2001 and 2002 the GDP of almost all the countries increased and had a steady growth above the 0 mark.

Based on analysis on all the countries, this can further validate the fact that ever since a country's GDP starts to be stable, the country's broadband subscriptions start to increase immensely.

So what do these findings tell us? That there is no obvious systematic relationship between these two variables in these areas. These examples put light on the demand for broadband services in these areas even in the face of economic downturns, highlighting the adaptability of broadband infrastructure to economic fluctuations. However, it is not 100% accurate to say that just because there is no correlation between them, then there is no causation made by the other on one.





B. Africa Underdeveloped Countries

In our case, since the dataset only contains data for a few to no years and not all countries, and some values are missing, any analysis based on this limited data may not provide a comprehensive understanding of the relationship.

This might be because of what is stated in this article. https://www.comminit.com/global/content/twenty-african-countries-have-introduced-broadband-and-more-set-follow-says-new-report#:~:text=According%20to%20this%20report%2C%20the%20first%20broadband%20access%20was%20rolled%20out%20in%202001%2C%20the%20pace

%20picked%20up%20in%202003%2C%20and%20by%202005%20%22a%20wide%20range%20of%20both%20wireline%20and%20wireless%20broadband%20technologies%20have%20been%20deployed%20across%20Africa.%22

“According to this report, the first broadband access was rolled out in 2001, the pace picked up in 2003, and by 2005 "a wide range of both wireline and wireless broadband technologies have been deployed across Africa."”

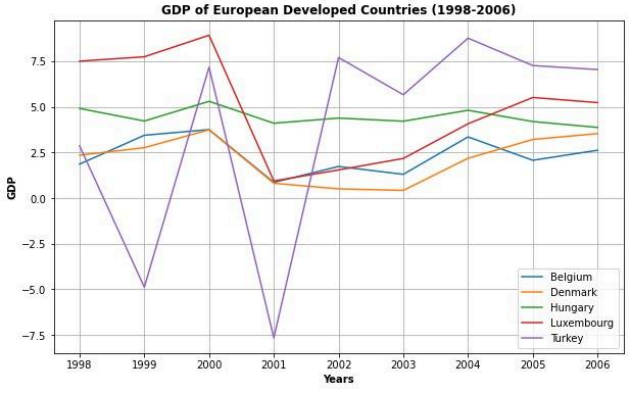
This is, in fact, proven by the data set since we have ⅗ of the samples start recording broadband subscription starting from 2005.

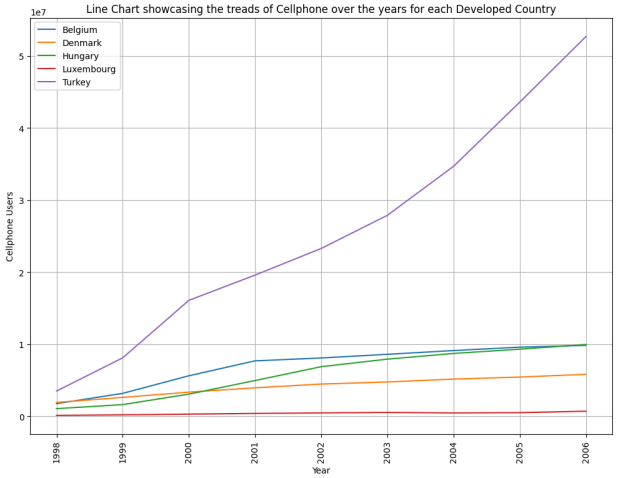
This could also prove the reasons why there were little to no data points in the african developed countries.

https://academic.oup.com/book/12417/chapter/162905356#426044008:~:text=It%20is%20within%20the%20context%20of%20these%20multiple%20political%20and%20economic%20shocks%20that%20we%20analyse%20snapshots%20of%20poverty%20and%20inequality%20in%20Madagascar%20in%20the%20first%20decade%20of%20the%202000s.

“The economic situation in Africa also would improve if the military conflicts that have plagued the continent over the past half-century stopped.”said in 2004.

https://www.nber.org/digest/jan04/economic-decline-africa





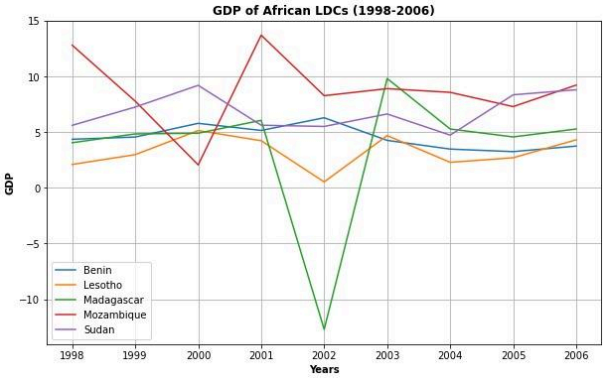
3. GDP & Cellphone Users

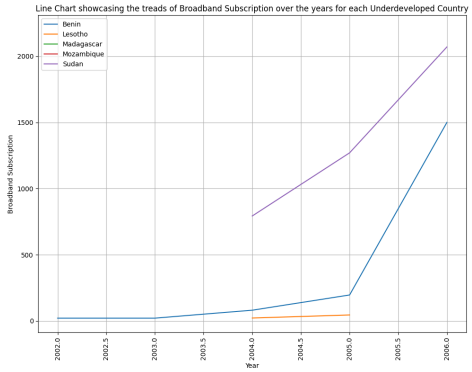
   Using our single variable analysis of both the GDP & Cellphone Users, we can already dissect some similar trends that both graphs are following; therefore, we can conclude some relation between both of them.

   A. European Developed Countries

      1. **Turkey** has the highest growth rate regarding cell phone usage and highest values compared to the other 4 countries. When analyzing the relations between the GDP and cellphone usage, one could notice that despite Turkey having the biggest plummeting in regards of GDP within the year 2000 and 2001 (the GDP going from 7.17 to−7.66) , within this time, the cell phone usages rose immensely from 16100000 to 19600000.
      2. According to the growth rate, Turkey's growth rate in cellphone users from 2000 to 2001 is approximately 21.74%, although the GDP went down by –207.12%.
      3. As mentioned in the GDP & Broadband Subscriptions segment, one might think that there might be a negative correlation. However, Turkey recovered and went from -7.66 to 7.7 within the period of 2001 and 2002. Nevertheless, the cellphone users 19600000 to 23300000.
      4. We can future investigate by analyzing countries with lesser rates of growth as **Denmark**. Denmark from 1998 to 2006 has had a steady GDP's growth rate of 0.052%. Nevertheless, the cell phone user's growth rate is about 0.14%.
      5. While analyzing **Belgium**, there was a plummeting from 3.35 to 2.07 (which is a -38.30% decrease) within the period of 2004 and 2006 in the GDP. That didn't affect the cellphone users, as it increased from 9130000 to 9600000 (which is a 5.147% increase).

In conclusion, so far, the analysis suggests that while economic conditions might not play a role in shaping technological trends, they may be one of the factors. Other underlying factors, such as consumer preferences and technological innovation also do have significant influence. However, in the context of GDP and cellphone users' analysis, there are no correlations whatsoever.





B. African Underdeveloped Countries

1. Upon examining the graph, it becomes evident that **Madagascar's** significant fluctuations in both increase and decrease rates stand out prominently, leading us to delve deeper into analyzing the factors that might be influenced due to its highest increases and decreases.
2. Within the years 2001 and 2002, Madagascar plummeted with a decline rate of -309.57% in its GDP. Comparing it to the cell phone users in Madagascar, it increased from 148000 to 163000, which is an increase of 10.135% nevertheless.
3. Now let’s analyze the period within 2002 and 2003. The GDP values for 2002 and 2003 were −12.7 and 9.81 respectively. This is a growth rate of 177.16%. Looking at the cellphones users values, we would see that it jumps from 163000 to 284000, which is 74.23% growth rate.
4. This proves that regardless of how the GDP acts, the cellphones users keep rising.
5. To further validate such findings, let’s test a country with a steady growth rate with little to no fluctuations in its GDP.
6. **Benin’s** GDP in the years 2002 and 2003 had 6.29 and 4.26, respectively. This is a decline rate of about -32.27%; meanwhile, its respective cellphone users' values went from 219000 to 236000 (which is 7.76%).

Within the years of 1999 and 2000, it had a GDP of 4.56 and 5.79, respectively. Its respective cellphone users values would be 7270 and 55500, which holds a growth rate of 663.41%.

In both Madagascar and Benin, we see that the growth rates of cellphone users tend to increase alongside higher GDP growth rates.

In **Madagascar**, despite experiencing a substantial increase in GDP from 2002 to 2003, cell phone usage was exposed to an even higher growth rate, showing that there might have been a demand for telecommunications services. Political developments during this period could have driven the need for secure and reliable communication channels for government officials and organizations.
Likewise, in **Benin**, when the economy saw stronger growth plummets, we noticed that more people tended to adopt cellphones at a faster rate. This suggests that as the country's economy grows, more people are likely to embrace cellphone usage.

In conclusion, while growth in GDP may influence cellphone use to some extent, it is not the main factor. An in-depth comprehension of telecommunications trends involves consideration of multiple factors that can contribute.

4. GDP & Income per person & Population

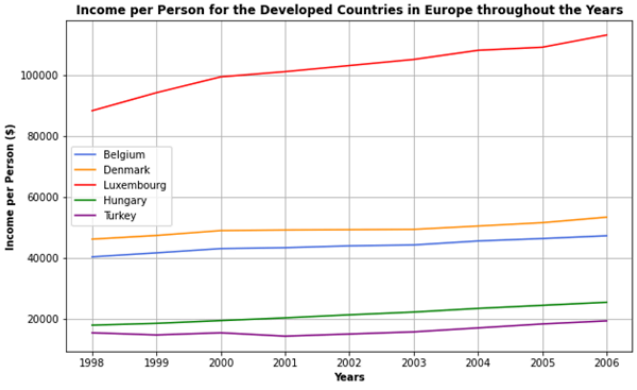
We are analyzing the three variables together as they are correlated, as will be seen in the equation.

A. The Developed countries:

The income per person (per capita) is definitely correlated with the GDP as,

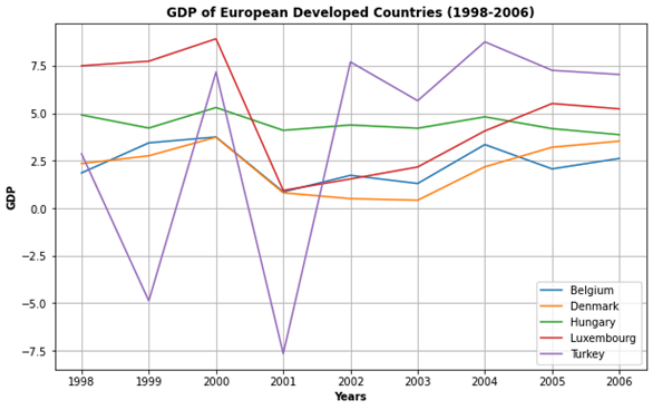
$$Income\ per\ person = \frac{GDP}{Population}$$

- However, if we compare the two graphs, we will find that, for example, in Luxembourg 2000-2001 the GDP decreased from 8.92 to 0.936 while the



income per person was still increasing from $101000 to $103000.

- This may seem to contradict with the relation that is mentioned earlier but it is not, let's explain the reasons.
- The GDP graph represents the GDP growth rate not the GDP itself, so when we see at the beginning that Luxembourg is the highest in 1998 that doesn't mean that it has the highest economy because these are not the actual values of the GDP, this just means that Luxembourg's GDP was growing at the fastest rate compares to the other countries. A GDP growth rate of 8.92% implies a considerable economic boost, which might be attributed to higher productivity, investment, or other favorable economic conditions in 2000. Then, the reduction to 0.936% growth rate indicates that the early strong expansion was not sustainable, and the economy steadied at a more modest growth rate. This is typical following a period of fast expansion.



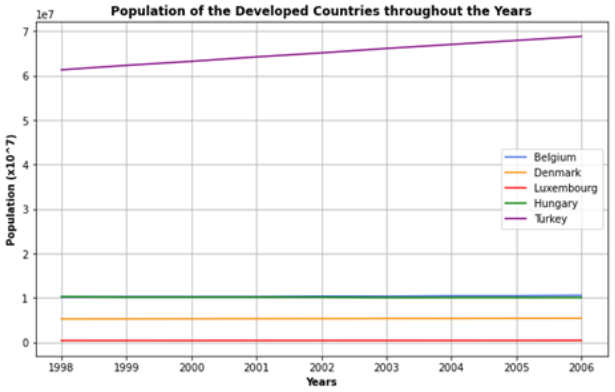
- Now, for the income per person, despite the lower GDP growth rate, income per person increased from $101,000 to $103,000, demonstrating that the economy expanded sufficiently to raise individual earnings despite a larger population. This slight gain in income per person, along with population growth, corresponds to a lower total GDP growth rate in the second year.
- Now let's explain the rest of the countries using the same understanding:

Turkey: The GDP growth rate keeps fluctuating between the positive and negative through the years, and this results in fluctuating the income per person as well. This can be seen from the two graphs although the fluctuations in the income per person is not that obvious, so let's write the numbers. It was $15400, $14700, $15400, $14300, $15000, $15700, $18300, $19300 in 1998-2006 respectively. Thus, as you can see the per capita fluctuate each year until 2002, then it keeps increasing and from the GDP graph we can see that the GDP was fluctuating between positive and negative growth rates but from 2002 and beyond the GDP growth rate was straightly positive, that explained the constant increase of the per capita after 2002.

So, we can conclude that as long as the GDP growth rate is positive the income per person will keep increasing or stabilize.

Belgium, Denmark & Hungary: they all follow the same concept that we discussed before. All of these experience a positive GDP growth rate even if it fluctuates but the rate is still in the positive part, so the income per person will increase as the time passes.

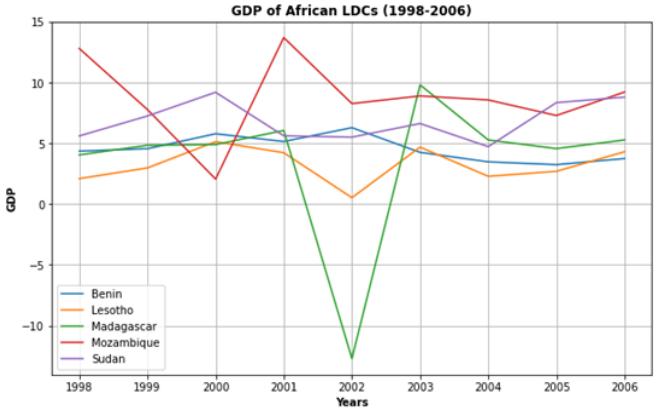
As we said earlier, $Income\ per\ person = \frac{GDP}{Population}$,which indicates a negative relationship between the population and the income per person. This is very obvious from the graphs that because Turkey surpassed the other countries in population, it has the least income per person. Because the GDP is divided for the whole population which makes the income for each person very low. On the other hand, is Luxembourg, it has the lowest population among all of the other countries which indicates that the income per person will be much higher for Luxembourg. This can be seen from the income per person graph as Luxembourg it way higher that the other countries in its per capita.

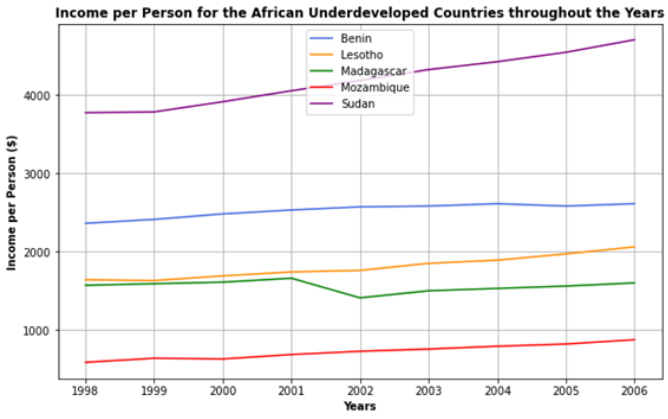


B. The Underdeveloped Countries:

Continuing on the same approach that we used in the developed countries.

Madagascar had a sudden decrease in the GDP growth rate from 2001 to 2002 and because the growth rate became -12.7% in 2002 the income per person also decreased from $1660 to $1410 in 2002. Then the GDP increased again after this period and became positive and thus the income per person started to increase.

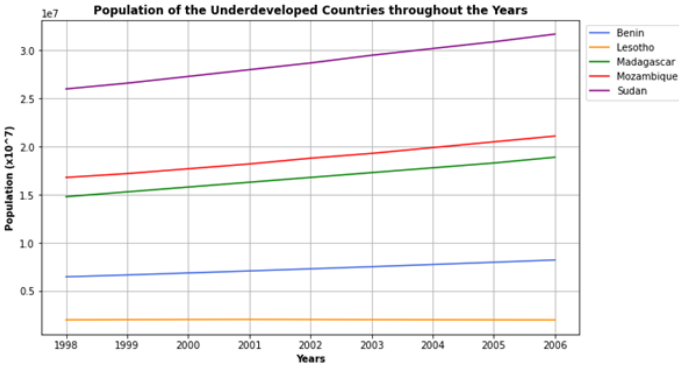




Mozambique: Although it has GDP that is positive through all the years, its population is very high which results in a very low income per person.

Sudan: it is obviously the highest economy in all of the other underdeveloped countries as its GDP growth rate is somehow stable with a little fluctuation in the positive side.

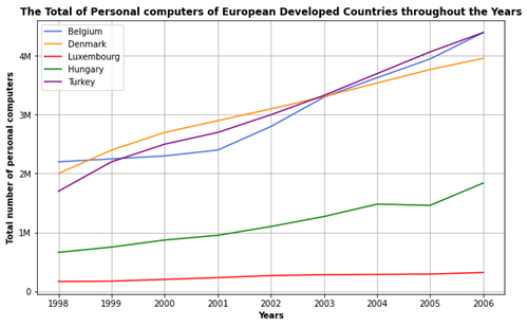
And although its population is the highest among all the countries in which its population was 26M in 1998 and increased to 31.7M in 2006, its income per person is the highest in all years. And we know that the population is inversely proportional with the GDP and the income per person; however, Sudan could manage to keep the GDP growth rate and the income per person high throughout the years.



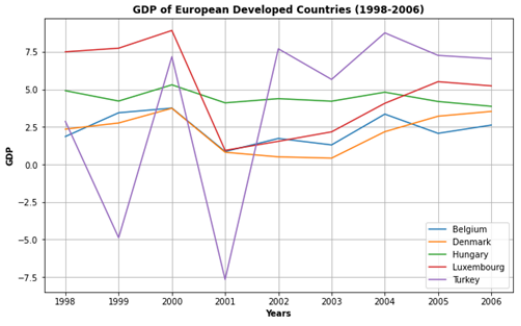
Lesotho & Benin: their statistics are very similar in which their GDP rate is moderately fluctuating on the positive side and their income per person and population are increasing in a very moderate pace.

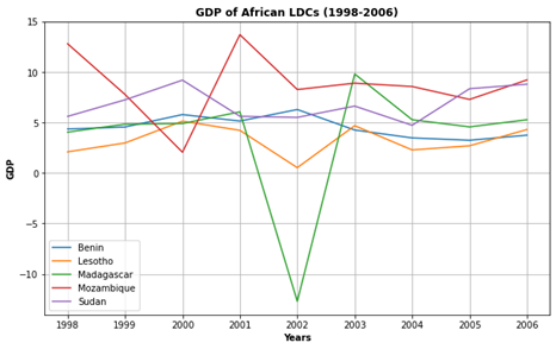
5. GDP & Personal Computers

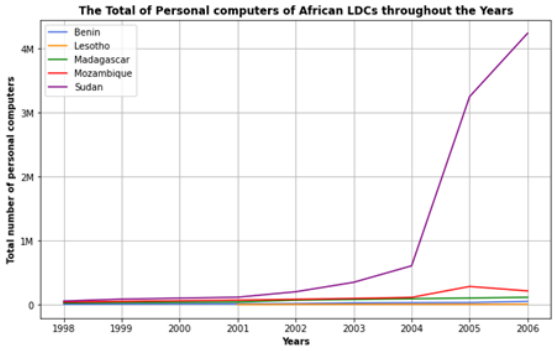
If we looked at both of the graphs of the developed and underdeveloped countries, it is hard to spot a correlation between the GDP growth rate and the total number of personal computers in a country.



The reason for that is the many factors that affect the total number of personal computers in a country such as government policies, infrastructural development, institutional quality, and global economic conditions all have a significant impact on economic growth. Thus, it is true that GDP growth rate and GDP in general plays a role in the technological development of the country but it is not the only factor.







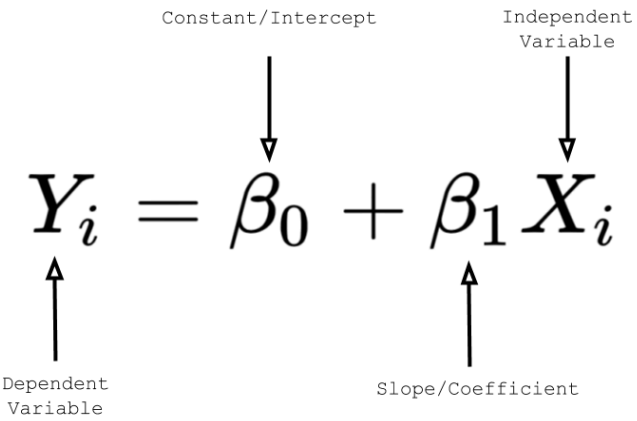
According to Moulton's paper, it highlights that technological advancement, particularly in the digital economy, has had a considerable impact on GDP and productivity; however, assessing this influence correctly is difficult due to ineffective statistical approaches. Thus, we concluded after finding a clear relationship between the variables that there are many factors that affect the total number of personal computers in a country such as government policies, infrastructural development, institutional quality, and global economy.

Moulton's paper:
https://www.bea.gov/sites/default/files/2018-05/gdp-and-the-digital-economy.pdf

V. Regression analysis

https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-linear-regression/
https://realpython.com/linear-regression-in-python/

Linear regression is a fundamental statistical technique and a powerful tool in data science and machine learning. It is used to model the relationship between a dependent variable and one or more independent variables. In our case, GDP would be the independent variable, where it influences the other variables which we consider the dependent variable. This method is necessary for predictive analysis, allowing us to estimate future outcomes based on historical data.
The standard formula for a linear regression model is:



In our analysis, we constructed a regression model for GDP and internet users. This prediction will further validate our findings.
However, in order to deal with the regression more, we will need to reduce our scope to two countries, each from different categories. This time, we will take into account all of the years in order to give enough and fair data for the machine to study form. The countries will be **Turkey** for European Developed countries and **Mozambique** for African Underdeveloped countries.

To start with, with the large scope of dataset we have, we would need a python code to predict based on the data we feed from within 1800 till before the year we are specifically asking for its estimated value.

#Ordinary Least Squares regression (OLS) is a common technique for estimating coefficients of linear regression equations which describe the relationship between one or more independent quantitative variables and a dependent variable (simple or multiple linear regression).
https://www.xlstat.com/en/solutions/features/ordinary-least-squares-regression-ols#:~:text=Ordinary%20Least%20Squares%20regression%20(OLS)%20is%20a%20common%20technique%20for,simple%20or%20multiple%20linear%20regression.

In order to understand what the regression equation means, we referred back to the following link:
https://www.geeksforgeeks.org/interpreting-the-results-of-linear-regression-using-ols-summary/

We will focus on the following variables in the regression results:
F-statistics :

It is a measure of whether the independent variable(s) collectively have a significant effect on the dependent variable. It is greatly related to the R squares as it is in its equation. A large F-statistic value proves that the regression model is effective in its explanation of the variation in the dependent variable and vice versa. On the contrary, an F-statistic of 0 indicates that the independent variable does not explain the variation in the dependent variable.

https://www.w3schools.com/datascience/ds_linear_regression_pvalue.asp

P-values:

We test if the true value of the coefficient is equal to zero (no relationship). The statistical test for this is called Hypothesis testing: A low P-value (< 0.05) means that the coefficient is likely not to equal zero.

A high P-value (> 0.05) means that we cannot conclude that the explanatory variable affects the dependent variable (in our case: GDP and Internet Users).
A high P-value is also called an insignificant P-value.

R squared:

A statistical measure in a regression model that determines the proportion of variance in the dependent variable that can be explained by the independent variable. In other words, r-squared shows how well the data fit the regression model (the goodness of fit).

Coefficients:

A statically measure which is used to measure the average functional relationship between variables. (Beta 0- the intercept) (Beta 1- our slope)

Mozambique's GDP & internet users model

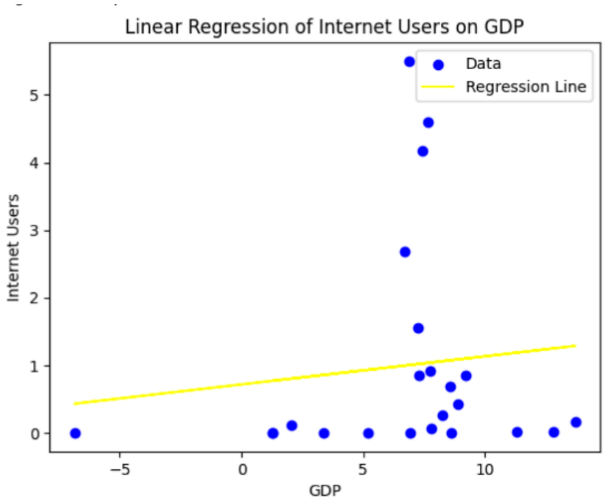
OLS Regression Results

Dep. Variable:	y	R-squared:	0.012
Model:	OLS	Adj. R-squared:	-0.035
Method:	Least Squares	F-statistic:	0.2566
Date:	Sun, 19 May 2024	Prob (F-statistic):	0.618
Time:	17:02:00	Log-Likelihood:	-43.281
No. Observations:	23	AIC:	90.56
Df Residuals:	21	BIC:	92.83
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.7155	0.648	1.104	0.282	-0.632	2.063
x1	0.0416	0.082	0.507	0.618	-0.129	0.212

Omnibus:	15.148	Durbin-Watson:	0.110
Prob(Omnibus):	0.001	Jarque-Bera (JB):	14.421
Skew:	1.743	Prob(JB):	0.000739
Kurtosis:	4.703	Cond. No.	14.9

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
Regression Equation: Internet Users = 0.72 + 0.04 * GDP



Regression Equation: Internet Users = $0.72 + 0.04 \times GDP$

Results	Values & their analysis
F-statistics	The F-statistics value is 0.2566, which is a very low value. It indicates that the regression model isn't showing the whole picture of the variability and relation between both variables.
P-values	The p-value associated with 'x1' is 0.618, which is the p-value of our independent variable. Since it is more than 0.05 with an immense difference it significantly shows that there is not enough evidence to build upon that the coefficient is in fact equal to zero, which is equivalent to saying that there is no apparent relationship between the GDP and internet users.
R squared	The value we have is 0.012. So, this means that only about 1.2% of the changes we see in Internet users can be due to changes in GDP. The values we have are extremely low indicating that the model couldn't explain much of the relation between the dependent and independent variable.
Coefficients	β_0 is our intercept, which would be 0.7155 (0.72). This tells us the value of the intercept when the GDP is equal to zero. β_1 is the slope in which the dependent variable increase per unit change in the independent variable.

Turkey's GDP & internet users model

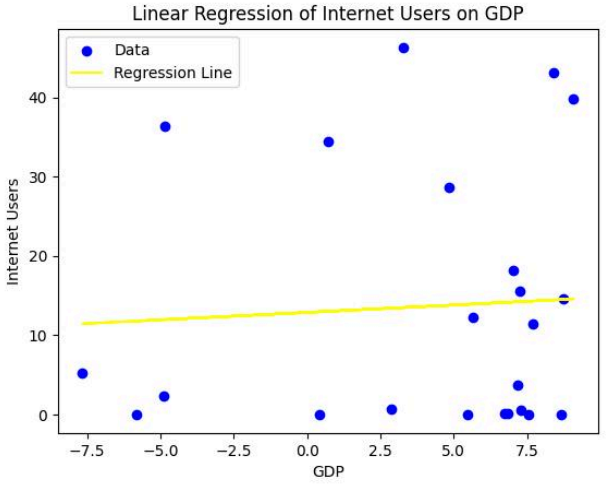
OLS Regression Results

Dep. Variable:	y	R-squared:	0.004
Model:	OLS	Adj. R-squared:	-0.044
Method:	Least Squares	F-statistic:	0.07536
Date:	Sun, 19 May 2024	Prob (F-statistic):	0.786
Time:	16:08:07	Log-Likelihood:	-96.102
No. Observations:	23	AIC:	196.2
Df Residuals:	21	BIC:	198.5
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	12.8741	4.397	2.928	0.008	3.731	22.018
x1	0.1863	0.679	0.275	0.786	-1.225	1.597

Omnibus:	4.075	Durbin-Watson:	0.047
Prob(Omnibus):	0.130	Jarque-Bera (JB):	3.398
Skew:	0.859	Prob(JB):	0.183
Kurtosis:	2.227	Cond. No.	8.34

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
Regression Equation: Internet Users = 12.87 + 0.19 * GDP



The Regression Equation:
Internet users = 12.87 + 0.19 * GDP

Results	Values & Their Analysis
F-statistic	The value is 0.07536 which suggests that the current regression model is not effective in explaining the relationship between GDP and Internet users
P-values	The P-value is 0.786. Since it is more than 0.05 with an immense difference it significantly shows that there is not enough evidence to build upon that the coefficient is in fact equal to zero, which is equivalent to saying that there is no apparent relationship between the GDP and internet users.
R squared	The value is 0.004 which is a very low value, thus the regression model may not be useful in predicting or understanding the variations in the internet users based on the GDP only. So, this indicates that the relation between the GDP and the internet users is very weak or does not exist.
Coefficients	$\beta_0 = 12.8741$ is the y-intercept which is the percentage of the internet users when the GDP was zero. While $\beta_1 = 0.1863$ is the slope of the regression line which is the rate of change of the

Conclusion of the regression models:

As we proved already in the relation and analysis segment that, in fact, the relation between the GDP and the other variables are not strong. Therefore, it's now accurate to theorize that the sample data may not be sufficient enough to draw meaningful conclusions about the relationship between GDP and Internet users. Economic downturns, political events, technological advancements, and other external factors could have influenced the observed data, making it challenging to establish a clear relationship.

VI. Conclusion

- Based on the analysis and the results that we did, we can not deduce a direct relationship between the GDP and the technological advancements as the period that we did the analysis on is not sufficient for making any correlation. However, if we increase the scope of the analysis and if we have more data, we can come up with a relation of how the GDP impacts the other factors. Despite that there are some variables that do have direct relation with the GDP such as the population and the income per person (per capita).

- Another noticeable result in this analysis is that as long as the GDP growth rate is positive, some other factors (income per person & broad) will increase or at least stabilizes because the positive growth rate means that the countries' economies are flourishing but with different rates while if the growth rate is negative, the economy is declining and they are facing financial problems.

- While the rest of the world's economy grew at an annual rate of close to 2 percent from 1960 to 2002, growth performance in Africa has been dismal which reveals the low GDP rates that the african countries experienced during this period. https://www.nber.org/digest/jan04/economic-decline-africa

- Developed countries typically show more stable growth and higher incomes per person due to smaller populations and sustained economic expansion. In contrast, underdeveloped countries face challenges with larger populations diluting GDP per capita, but consistent positive growth rates can still lead to improved incomes over time.