

Internet Access Data Analysis Report

The Report's Objective

Use the Datasets provided to solve the following questions:

1. What are the top 5 countries with the highest internet use (by population share)?
2. How many people had internet access in those countries in 2019?
3. What are the top 5 countries with the highest internet use for each of the following regions: 'Middle East & North Africa', 'Latin America & Caribbean', 'East Asia & Pacific', 'South Asia', 'North America', 'Europe & Central Asia'?
4. Create a visualization for those five regions' internet usage over time.
5. What are the 5 countries with the most internet users?
6. What is the correlation between internet usage (population share) and broadband subscriptions for 2019?
7. Summarize your findings.

The Dataset provided are: - people.csv, broadband.csv, internet.csv, and class.csv.

Dataset Descriptions

The Dataset 1: - people.csv. The dataset contains 6379 rows and 4 columns.

The columns names are 'Entity', 'Code', 'Year', and 'Users'.

'Entity' describes the country for which the data is recorded. This column contains 223 Unique values. It is a categorical field.

'Code' describes the country code for the country provided in the Entity column. This column contains 213 unique values as the code for some of the countries are missing but does not have any effect on the dataset.

'Year' describes the year in which the data was recorded. The column values range from 1990 to 2020 (both included).

'Users' describes the total number of internet users present in the country in given year.

The Dataset 2: - internet.csv. The dataset contains 7089 rows and 4 columns.

The columns names are 'Entity', 'Code', 'Year', and 'Internet_Usage'.

'Entity' describes the country for which the data is recorded. This column contains 261 Unique values. It is a categorical field.

'Code' describes the country code for the country provided in the Entity column. This column contains 214 unique values as the code for some of the countries are missing but does not have any effect on the dataset.

'Year' describes the year in which the data was recorded. The column values range from 1990 to 2019 (both included).

'Internet_Usage' describes the total percentage of internet users present in the country in the given year.

The Dataset 3: - broadband.csv. The dataset contains 3888 rows and 4 columns.

The columns names are 'Entity', 'Code', 'Year', and 'Broadband_Subscriptions'.

'Entity' describes the country for which the data is recorded. This column contains 222 Unique values. It is a categorical field.

'Code' describes the country code for the country provided in the Entity column. This column contains 208 unique values as the code for some of the countries are missing but does not have any effect on the dataset.

'Year' describes the year in which the data was recorded. The column values range from 1998 to 2020 (both included).

'Broadband_Subscriptions' describes the percentage of broadband subscriptions present in the given year.

The code for the solution is given in the attached file named 'solution.ipynb' and 'solution.py'.

The Dataset 3: - class.xlsx. The dataset contains 2146 rows and 4 columns.

The columns names are 'GroupCode', 'GroupName', 'CountryCode' and 'CountryName'.

'GroupCode' describes the code name of the group which is named in the corresponding column named 'GroupName'.

'GroupName' describes the name of the Group in which the countries are classified.

'CountryName' describes the name of the country that are classified into different groups.

'CountryCode' describes the code of the country in the corresponding column 'CountryName'. There are 264 unique country names in the dataset.

What are the top 5 countries with the highest internet use (by population share)?

The top 5 counties with highest internet use (by population share) are: -

Country	Internet Usage
Bahrain	99.701504
Qatar	99.652804
Kuwait	99.542676
United Arab Emirates	99.149998
Denmark	98.0446435

The following bar plot shows Internet Usage by the above 5 countries.

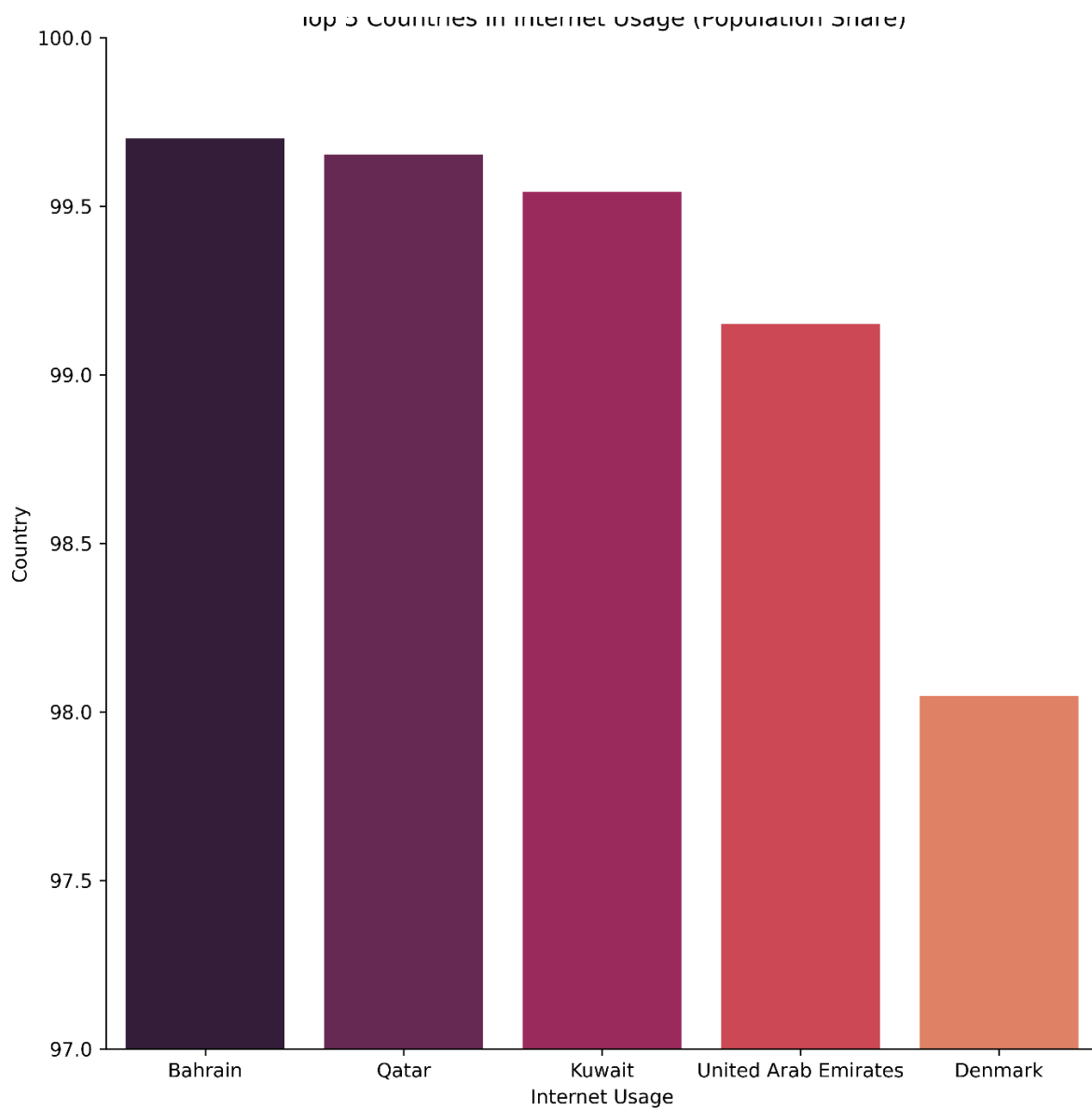


Fig.1 Top 5 Countries with most Internet Usage

How many people had internet access in those countries in 2019?

Country	Users
United Arab Emirates	9133361
Denmark	5682653
Kuwait	4420795
Qatar	2797495
Bahrain	1489735

The following bar plot shows the number of people who had internet access in above countries in 2019.

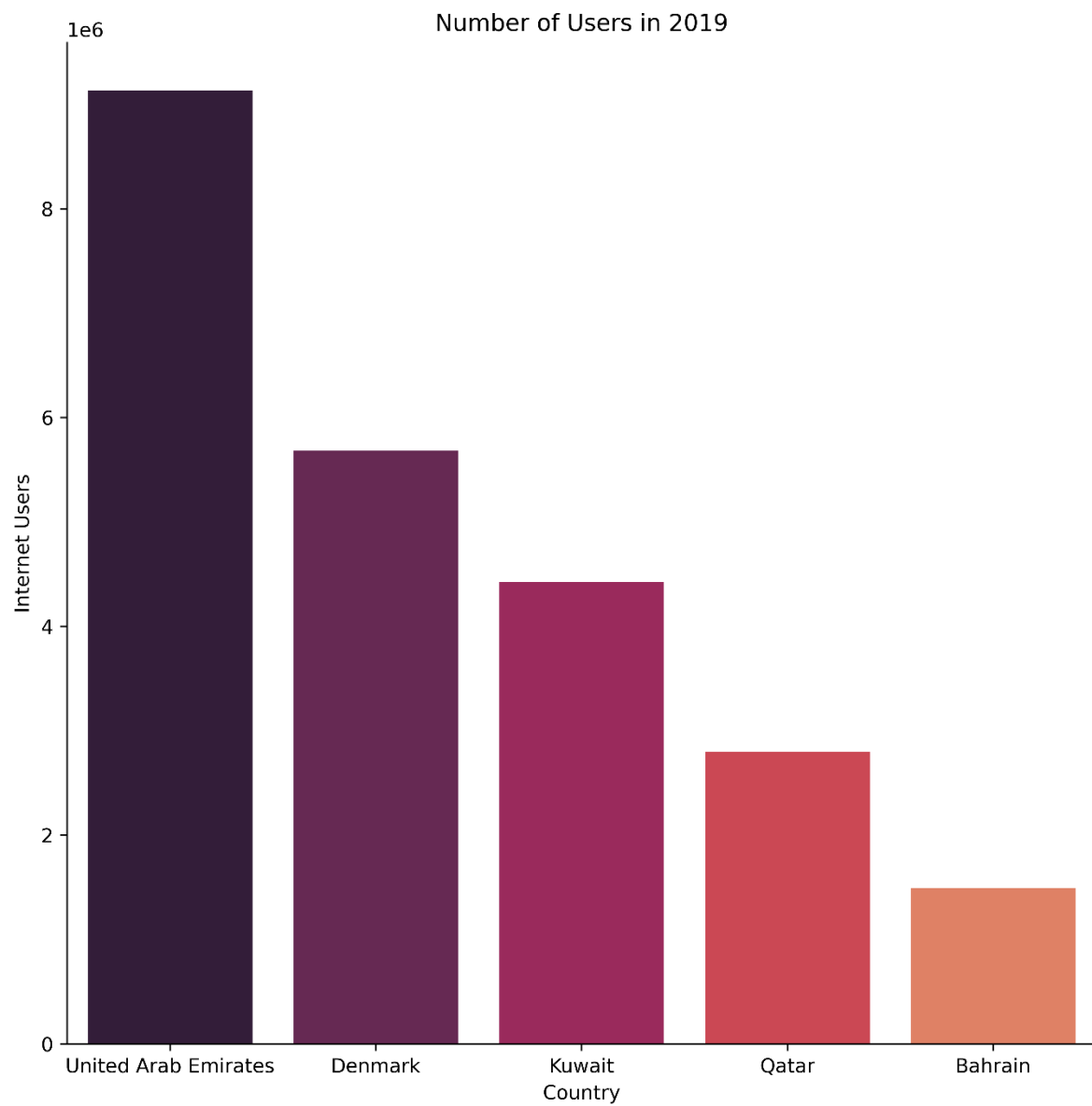


Fig 2. Numbers of Users in these above countries in 2019

What are the top 5 countries with the highest internet use for each of the following regions: 'Middle East & North Africa', 'Latin America & Caribbean', 'East Asia & Pacific', 'South Asia', 'North America', 'Europe & Central Asia'?

The six regions that are asked in the question can be found in the class.xlsx file. Most of the regions does not have the data for the recent years. Therefore, the data is matched and shown for the year 2017. We will explore each region one by one.

Region 1: East Asia & Pacific

In this region, there are 26 countries which are to be compared. There are 13 countries which have more than 50% of Internet Usage in year 2017. The name of these countries are Japan, New Zealand, Australia, Singapore, New Caledonia, Guam, Malaysia, French Polynesia, Nauru, Philippines, Vietnam, China, and Thailand in decreasing order.

The top countries among these who has the highest Internet Usage are: -

Country	Internet Usage
Japan	91.726567
New Zealand	90.811093
Australia	86.545049
Singapore	84.452268
New Caledonia	82.005841

The following line chart shows the growth of Internet Usage in above 5 countries.

The graph shows that every nation in the top 5 lists has similar growth except for New Caledonia. New Caledonia graph was below other nations in the start and started growing at a slow speed till 2000s. After 2000, New Caledonia growth grew by almost 300 percent in the next decade. After 2010, the growth continued and but slowed over the time to reach more than 82 percent in the year 2017. Countries other than North Caledonia followed a exponential growth and slowed over the time with Japan and New Zealand reaching more than 90 percent each in year 2017.

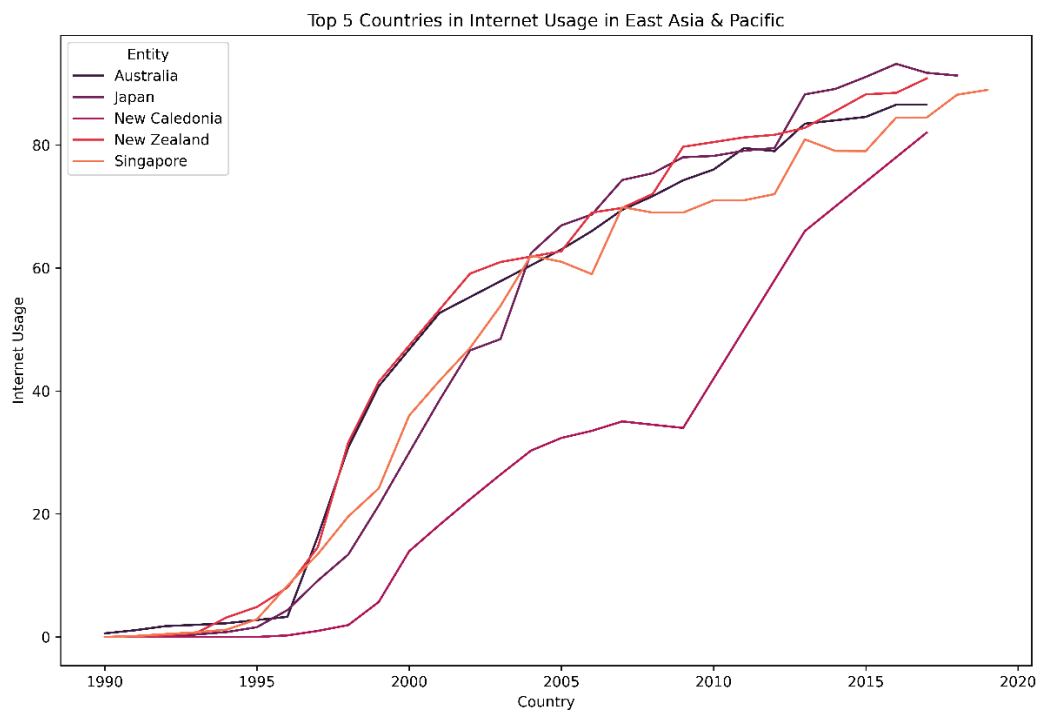


Fig. 3 Top 5 Countries with Internet Usage in East Asia and Pacific Region

Region 2: Europe and Central Asia

In this region, there are 49 countries which are to be compared. All the given countries have more than 50 percentage of Internet Usage by population in their countries except the Tajikistan and Turkmenistan. The country with most internet usage are **Liechtenstein, Iceland, Luxembourg, Denmark, and Monaco.**

The top countries among these who has the highest Internet Usage are: -

Country	Internet Usage
Liechtenstein	99.546612
Iceland	98.255201
Luxembourg	97.362960
Denmark	97.099362
Monaco	97.052977

The following line graph shows the growth of the Internet Usage in top 5 countries in this region over the time.

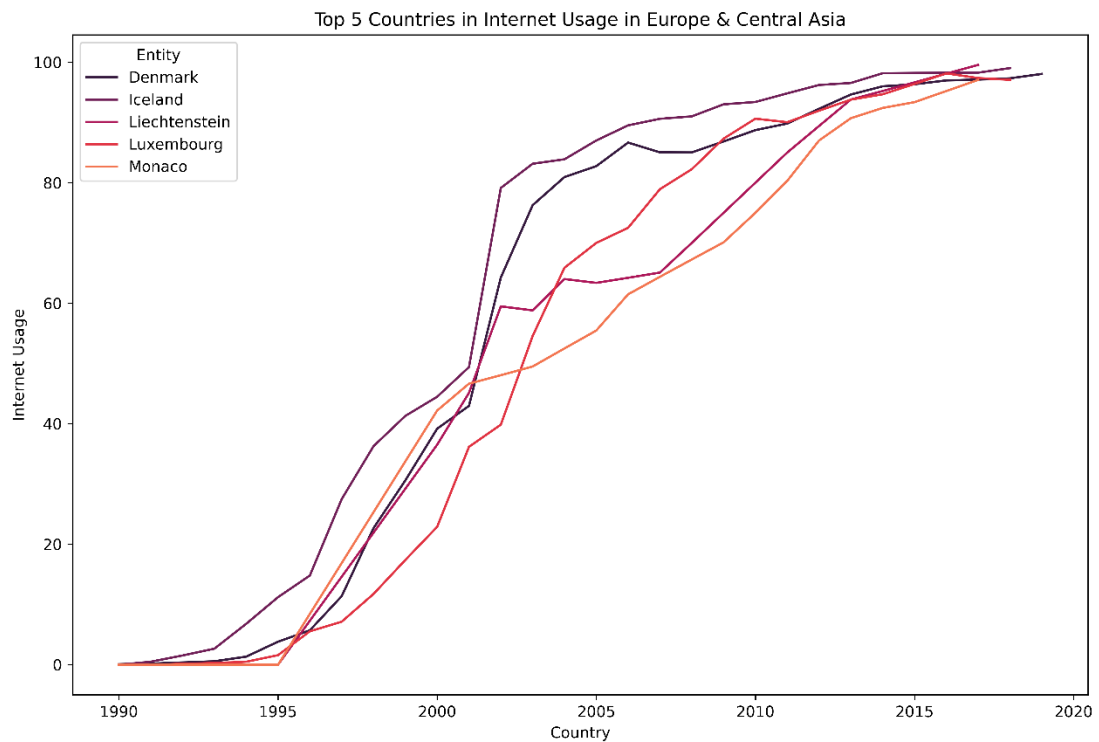


Fig.4 Top 5 Countries with Internet Usage in Europe and Central Asia Region

Region 3: Latin America and Caribbean

In this region, there are total 30 countries present to be compared with each other. 21 out of 30 countries have more than 50% usage of Internet in their countries. The countries with the most Internet Usage are **Aruba, Chile, Barbados, Cayman Islands, and British Virgin Islands**.

The top countries among these who has the highest Internet Usage are: -

Country	Internet Usage
Aruba	97.170000
Chile	82.327487
Barbados	81.760778
Cayman Islands	81.067693
British Virgin Islands	77.704268

The following line graph shows the growth of the Internet Usage in top 5 countries in this region over the time.

All the countries in the top 5 list started growing faster in late 90s. With Aruba getting a instant growth in the late 2010s making it top the list. There is a significant difference between Aruba and Chile, while Barbados, Cayman Islands, and British Virgin Islands being on-par with Chile.

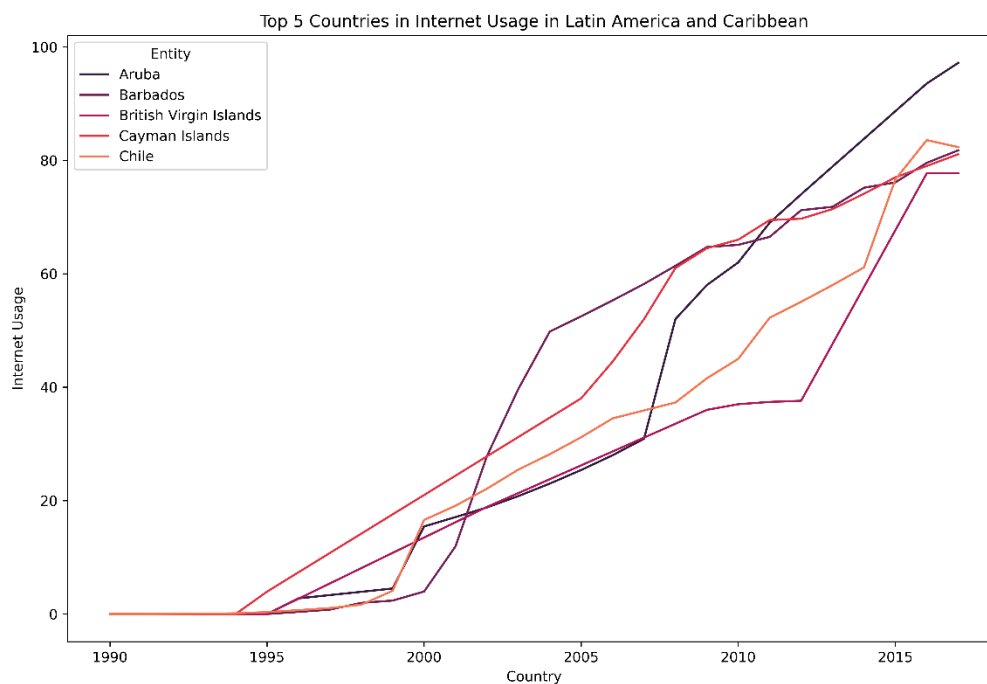


Fig 5. Top 5 Countries with Internet Usage in Latin America and Caribbean.

Region 4: Middle East and North Africa

In this region, there are a total of 16 countries. All the countries have achieved more than 50 percentage of Internet Usage Population Share except for Iraq, Algeria, and Libya in year 2017. The countries with most Internet Usage are **Kuwait, Qatar, Bahrain, United Arab Emirates, Saudi Arabia.**

The top countries among these who has the highest Internet Usage are: -

Country	Internet Usage
Kuwait	97.999989
Qatar	97.388849
Bahrain	95.878136
United Arab Emirates	94.819923
Saudi Arabia	94.175600

The following line graph shows the growth of the Internet Usage in top 5 countries in this region over the time.

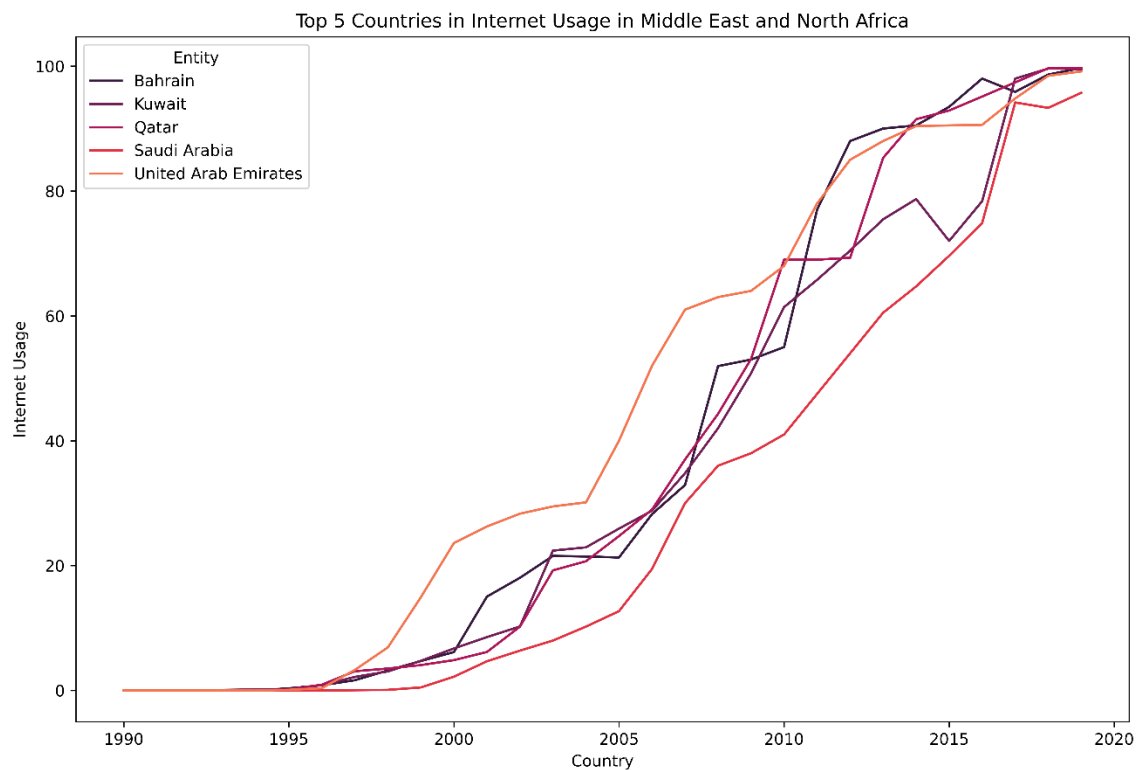


Fig 6. Top 5 Countries with Internet Usage in Middle East and Africa.

Region 5: North America

In this region, there are only 3 countries available. The names of these countries are **Bermuda, Canada, and United States of America**. **Bermuda and United States of America** has more than 90% of Internet Usage in Population Share while Canada has 87%.

The following line graph shows the growth of the Internet Usage in these countries.

The top countries among these who has the highest Internet Usage are: -

Country	Internet Usage
Bermuda	98.370000
United States of America	92.701372
Canada	87.274889

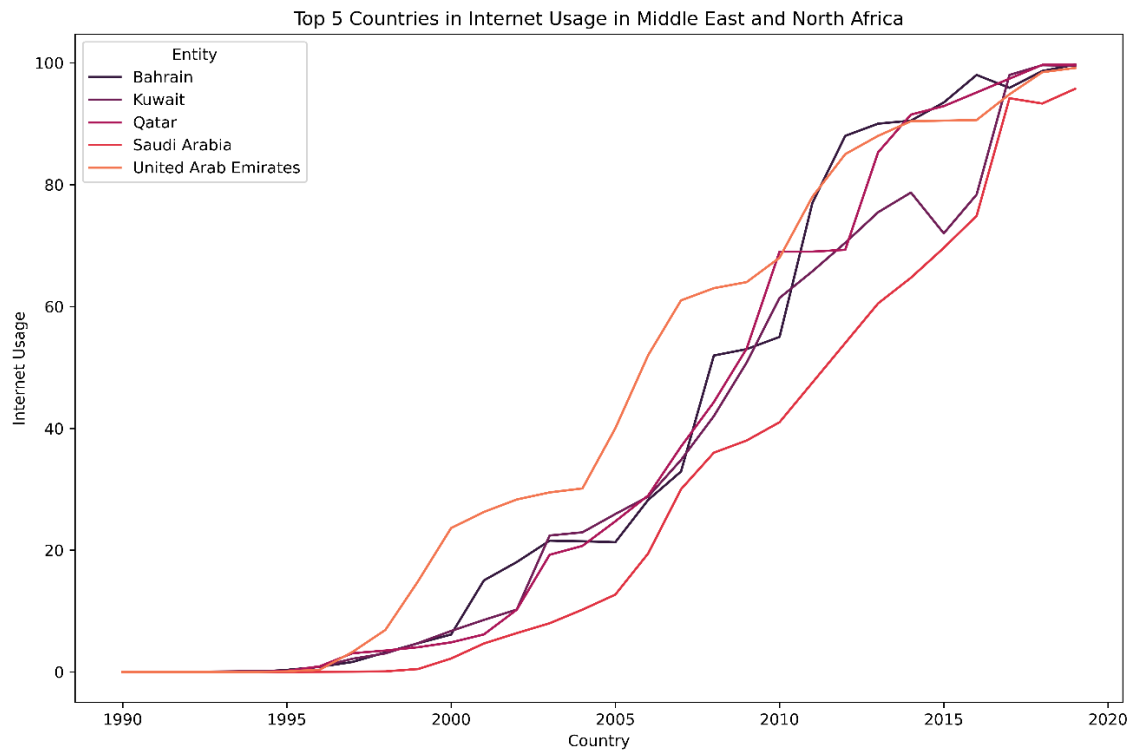


Fig 7. Top 5 Countries with Internet Usage in Latin America

Region 6: South Asia

There are 6 countries which comes under South Asia region and sufficient data till 2017. Some countries like Bhutan are missing due to insufficient data. Only Maldives among the 6 countries have Internet Usage more than 50% of the population. Maldives and India have more than 30% of internet usage in their population.

The top countries among these who has the highest Internet Usage are: -

Country	Internet Usage
Maldives	63.185666
Sri Lanka	34.113348
India	32.000000
Nepal	21.403510
Pakistan	17.105146

The following line graph shows the growth of the Internet Usage in top 5 countries in this region over the time.

Maldives had an exponential growth from late 2000s which resulted in their higher percentage among their neighbours. India and Sri Lanka had an increase in their percentage in late 2010s after decline in early 2010s. Nepal had a constant pace in their growth, while Pakistan on other hand had a horizontal line on their graph in their late 2010s.

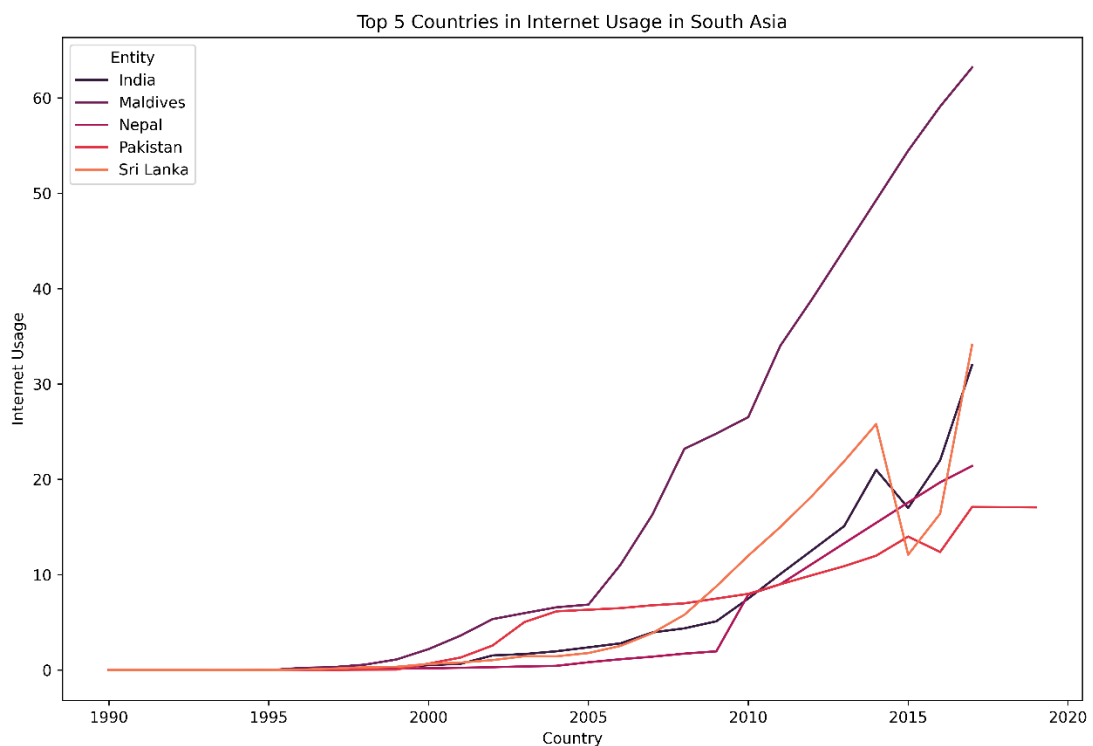


Fig 8. Top 5 Countries with Internet Usage in South Asia.

What are the 5 countries with the most internet users?

The people dataset is used here to determine which are the countries who have the most Internet Users in the world.

To start with this, we must exclude different rows of the people's dataset which have cumulative of many countries combined into one. This includes titles such as World, Asia, Upper Middle-Income Countries, High Income Countries, Lower Income Countries, Europe, North America, South America, and Africa.

This results in giving us the top 5 countries with most internet users, which are China, India, United States of America, Brazil, and Indonesia. China and India have lower population percentage as compared to others but prevail due to their massive population.

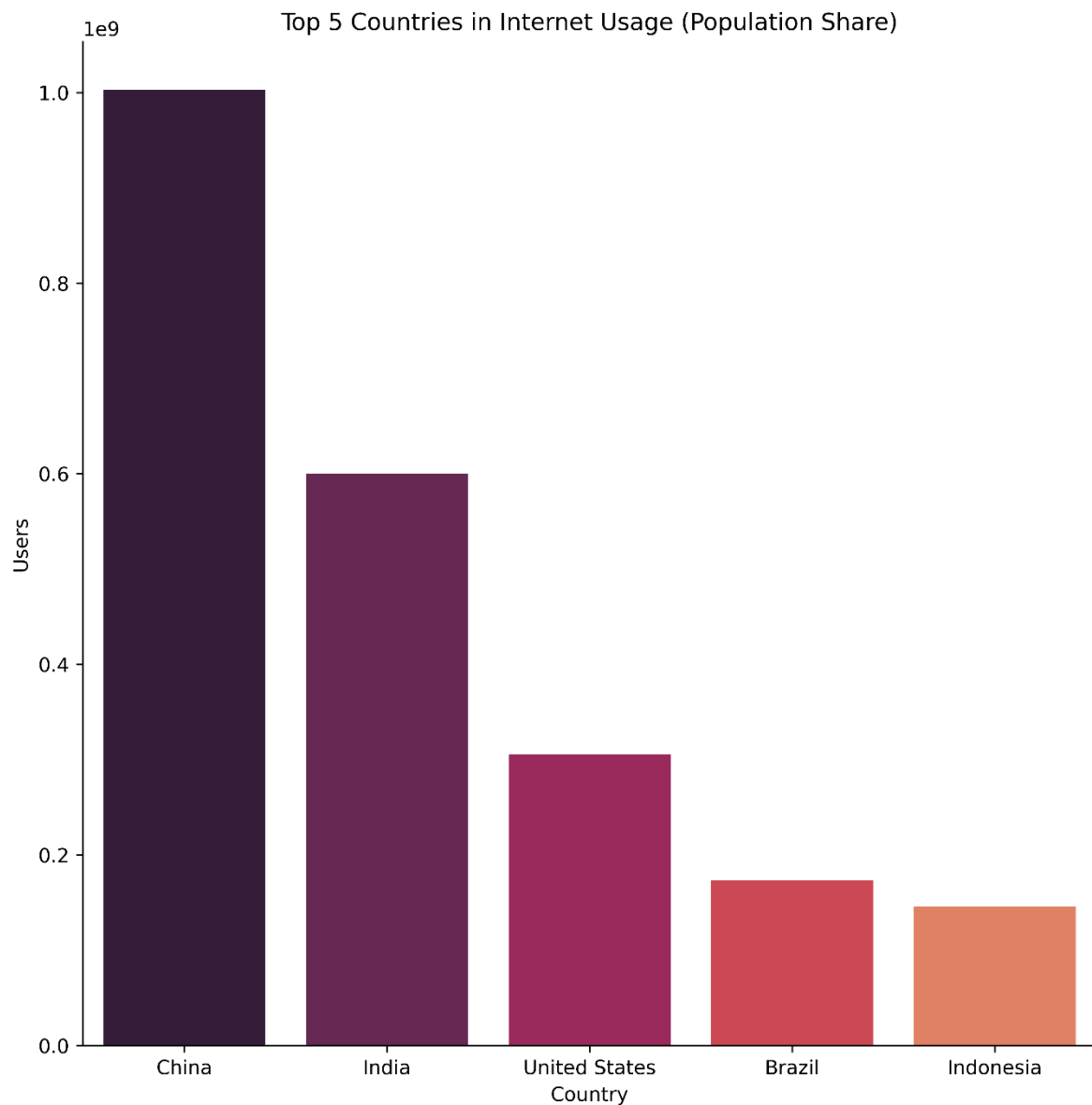


Fig 9. The countries with Most Users

What is the correlation between internet usage (population share) and broadband subscriptions for 2019?

Broadband.csv Dataset is merged with the Internet Dataset to make a merged dataset containing both Internet usage and broadband subscriptions in the same data frame. Upon merging the dataset, we can only conclude 70 rows which are available for the latest years in both datasets.

Through this dataset, we can easily calculate the correlation coefficient between Internet Usage (Population Share) and the broadband Subscriptions.

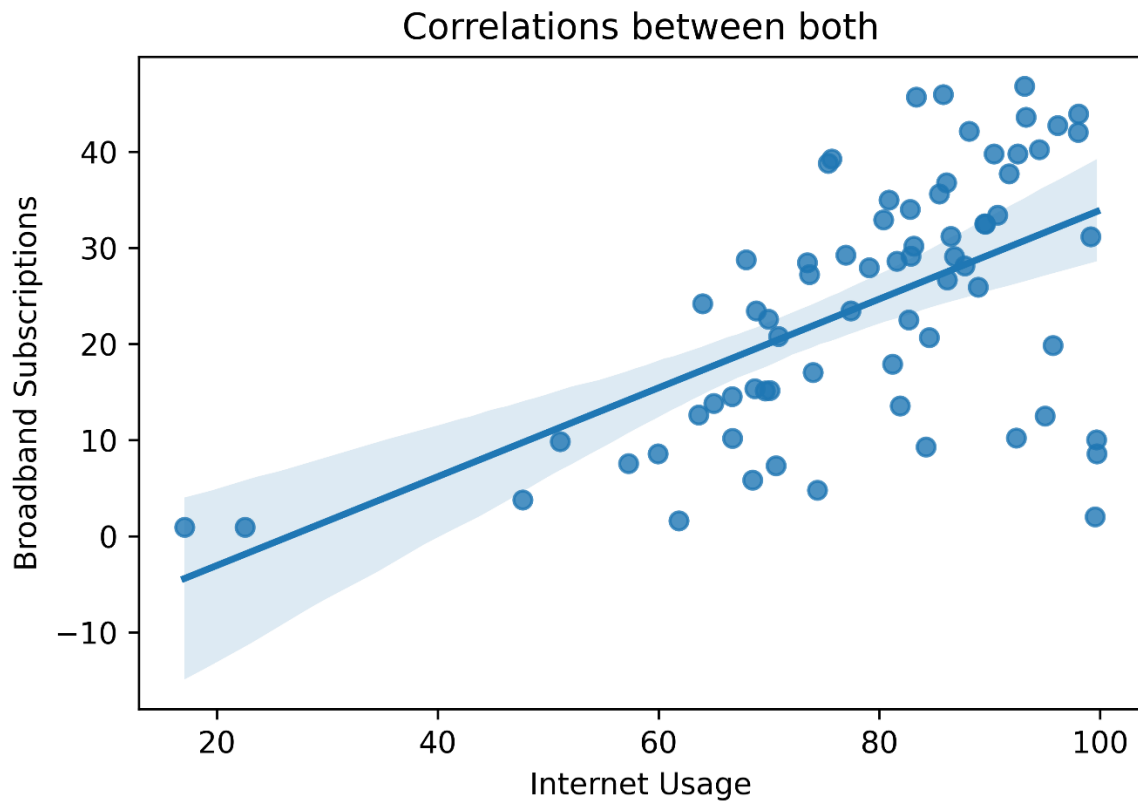


Fig 10. The scatterplot between Internet Usage and Broadband Subscription to show correlation.

The above scatterplot shows the relationship among both Internet Usage and Broadband Subscriptions. This also shows a best fit line.

Upon calculation by `corr()` function of python, we have a value of **0.56** as the correlation between the variables. This shows that both are moderately correlated to each other.

This concludes our report.