DATA
VISUALIZATION

ISM 6361 - Data Visualization

Murali, Sheetal

SPU ID: 900409765

TABLE OF CONTENTS

INTRODUCTION

There is practically no limit to what you can do online. The Internet allows you quickly find information, contact with people all over the world, manage your finances, shop from home, listen to music, view videos, learn and educate, share news and do so much more. Globally, internet user numbers are growing at a very fast pace. The Internet enables computer users to quickly connect to other gadgets and data stores from any location. The history of the Internet dates back several decades, where email has been around since the 1960s, file sharing since at least the 1970s, and TCP/IP was standardized in 1982. It was the creation of the World Wide Web in 1989 that revolutionized our history of communication.

Internet usage is measuring the data you use every time one checks instant messages, notifications, or updates on your social networks. The availability of unlimited broadband subscriptions and fast download speeds is contributing to a positive trend. Mobile cellular telephone subscriptions are subscriptions to a public mobile telephone service using cellular technology. It includes postpaid and prepaid subscriptions. With billions of websites online today, there is a lot of information on the Internet. Global Internet users has already reached almost 5 billion, or about 53% of the global population as of 2021.

BUSINESS JUSTIFICATION

While many of us cannot imagine life without the Internet, when it comes to internet usage, not every country or region is created equal. In fact, some countries can have more internet users than several others combined.

In this project, I have created various visualizations to gain insights into the Internet Usage and its trends based on Market Maturity and from geographical perspective. This includes internet usage in developed countries, developing countries, Least developed countries, Landlocked Developing countries and Small Island Developing countries from 1990 to 2022.

DATA SOURCE

The data I have collected is from Kaggle, The data contains information about internet users from 1980-2020, But for this visualization project I have used Data only for 1990 – 2020. Details about the columns are as follows:

1. Entity - Contains the name of the countries and the regions.
2. Code - Information about country code and where code has the value 'Region', it denotes division by grouping various countries.
3. Year - Year from 1980-2020
4. Cellular Subscription - Mobile phone subscriptions per 100 people. This number can get over 100 when the average person has more than one subscription to a mobile service.
5. Internet Users(%) - The share of the population that is accessing the internet for all countries of the world.
6. No. of Internet Users - Number of people using the Internet in every country.
7. Broadband Subscription - The number of fixed broadband subscriptions per 100 people.

Based on further research, I found the data for Global internet access rate, by market maturity from Statista. This data consists of percentage of population accessing internet from 2005-2022 in Developed Countries, developing countries, Small Island Developing States (SIDS) Least Developed Countries (LDCs) and Landlocked Developing Counties (LLDCs).

POWER BI – FEATURES USED

In my opinion, Power BI is a highly effective tool for analyzing and visualizing business data. Its user-friendly interface is particularly beneficial for me since I have already used Tableau

I believe that my experience with Microsoft Power BI has been exceptional. I was amazed by its intuitive interface, and it enables users to create interactive reports and dashboards that allow them to visualize data from various sources. I was able to implement the lessons learnt in class as well as the textbooks while preparing visualization in this tool. Below are some of the Features I used in creating the visualization.

1. **Import Data**: Power BI supports connecting to a wide variety of data, stored in a variety of places. For this study, I imported the Data from Excel.

2. **Filter**: After importing the Data I used the filter option to remove the data from 1980 to 1990. Because the study is based only on 1990 onwards. I have also filtered (All pages) North America since the data consisted of USA which had the same numbers.

   For Motion, The auto selected filters are basic filters, But I used Top N filter to choose the top 10 Countries to determine cellular subscription by population

3. **Maps**: I used 2 different maps, 1 is the basic one which is already available in Power BI and used Bubble size based on The % share of the population that is accessing the internet for all countries of the world. Other map I used is the extension which is Globe maps where the bar heights represent Number of internet users in every country.

4. **Motion**: I have used Motion in 2 different Charts:

   First is in Change in the scatter plot, where play axis depends on the change in year to determine number of cellular subscriptions to the % of users by population. The size of the bubble determines the actual number of internet users. This motion shows us how different countries have been emerging in the digital population

   Second, I used Play Axis extension to add a motion visualization to represent the number of users in different countries based on the population in those countries. It is a great tool when you want to see trends or look for patterns in your data since you can click on "play" and just focus on how data is evolving.

5. **Calculations**: For this project, most of the charts uses Aggregation, Although Power BI has Count as a default setting, I have also used Average and Sum for the charts to represent the data.

a. **Creating Table:**

```
Table = {"Developed", " Developing", "LLDC", "LDC"}
```

b. **Creating Measure:** Firstly, I created measures for sum and average for the following:

```
MeasureBroadband = SUM(Internetusers[Broadband Subscription])
MeasureCellular = SUM(Internetusers[Cellular Subscription])
LLDC = SUM(Sheet1[Landlocked Developing Countries (LLDC)])
MeasureDeveloped = SUM(Sheet1[Developed])
```

Create Year on year Growth rate:

```
Usage Difference =

  VAR _CurrentYearUsage = [currentyearusage]
  VAR _PreviousYearUsage = [previousyearusage]
  VAR _Result = _CurrentYearUsage - _PreviousYearUsage
  RETURN _Result

Usage Growth = DIVIDE ([Usage Difference], [previousyearusage])
```

c. **Adding parameter**:

```
ParaCellular/Broadband = {
    ("MeasureBroadband", NAMEOF('Internetusers'[MeasureBroadband]), 0),
    ("MeasureCellular", NAMEOF('Internetusers'[MeasureCellular]), 1)


Parameter = {("Developed Countries", NAMEOF('Table'[MeasureDeveloped]), 0),
    ("Developing Countries", NAMEOF('Table'[MeasureDeveloping]), 1),
    ("Least Developed Countries", NAMEOF('Table'[MeasureLDC]), 2),
    ("Landlocked Developing Countries", NAMEOF('Table'[LLDC]), 3),
    ("Small Island Developing Countries", NAMEOF('Table'[MeasureSIDS]), 4)}
```

6. **Dashboard**: It is where we put everything together, In Power BI, this feature is automated. It is an interactive place where we can add an action like filter(slicer) when you select/click to perform integration between the visualizations created. Here I have created a dashboard for parameter, text, graphs etc.

7. **Model view:** We can organize the data to help provide a clearer view into the tables to work with, and make working with complex datasets easier. With this we can add a referential integrity to the data which allows us to have relationship among the table. This helps in adding more efficiency to the interactive dashboards. In this Dataset, I have added "year" as the common property which acts as a primary key on the one side and foreign key to the many.

REASON FOR VISUALIZATION

DATA CARD

These are the number tiles, for representing numbers such as total usage, market share year over year, or total subscriptions. I have used this type of visualization to highlight the numbers and it changes based on the selection. In this project I have used data card to highlight Total cellular and total broadband subscriptions, total number of internet users and Average users for the share of the population that is accessing the internet for all countries of the world.

RIBBON CHART
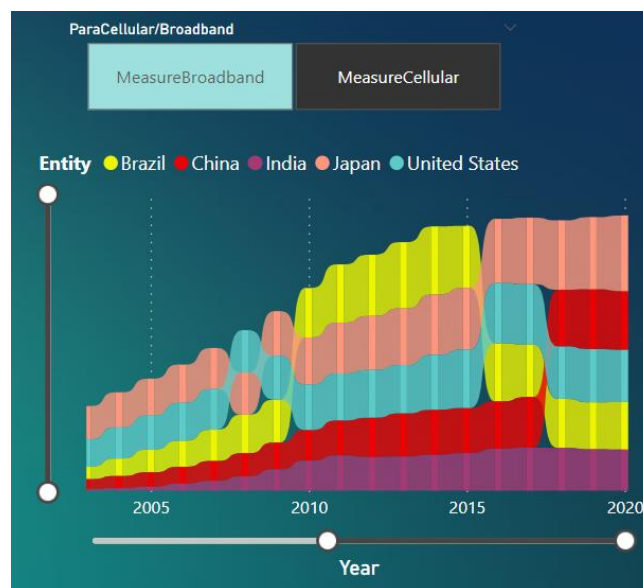
Ribbon charts is used to easily determine which category has the largest value. I have used this chart in the measure of cellular vs broadband to quickly determine the rank change over the period of time. With this chart, we can easily visualize the change in value with the changes always being displayed on top.

STACKED AREA CHARTS
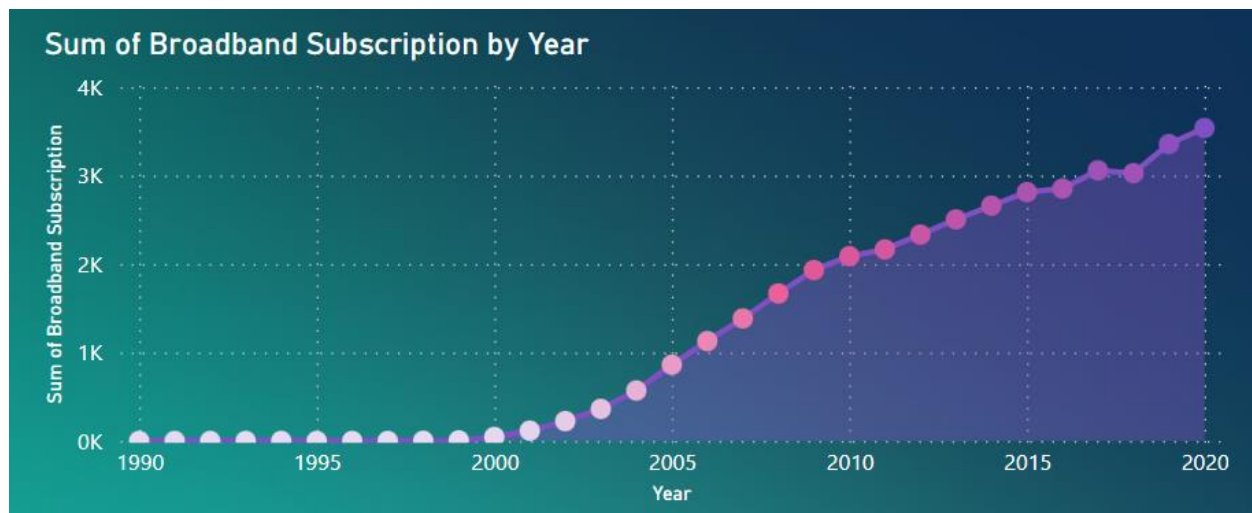
These charts emphasize the magnitude of change over time, and can be used to draw attention to the total value across a trend. For this project I have used stacked area charts to visualize The area between axis and line which is filled with color to indicate volume of Cellular/Broadband users.

This chart is commonly used to represent Area to showcase the data that depicts a time-series relationship. In this project I have used this type of chart to represent total broadband and cellular subscriptions. I have also used these charts in the parameter of Percentage of global population accessing the internet from 2005 to 2022, by market maturity

SCATTER PLOT
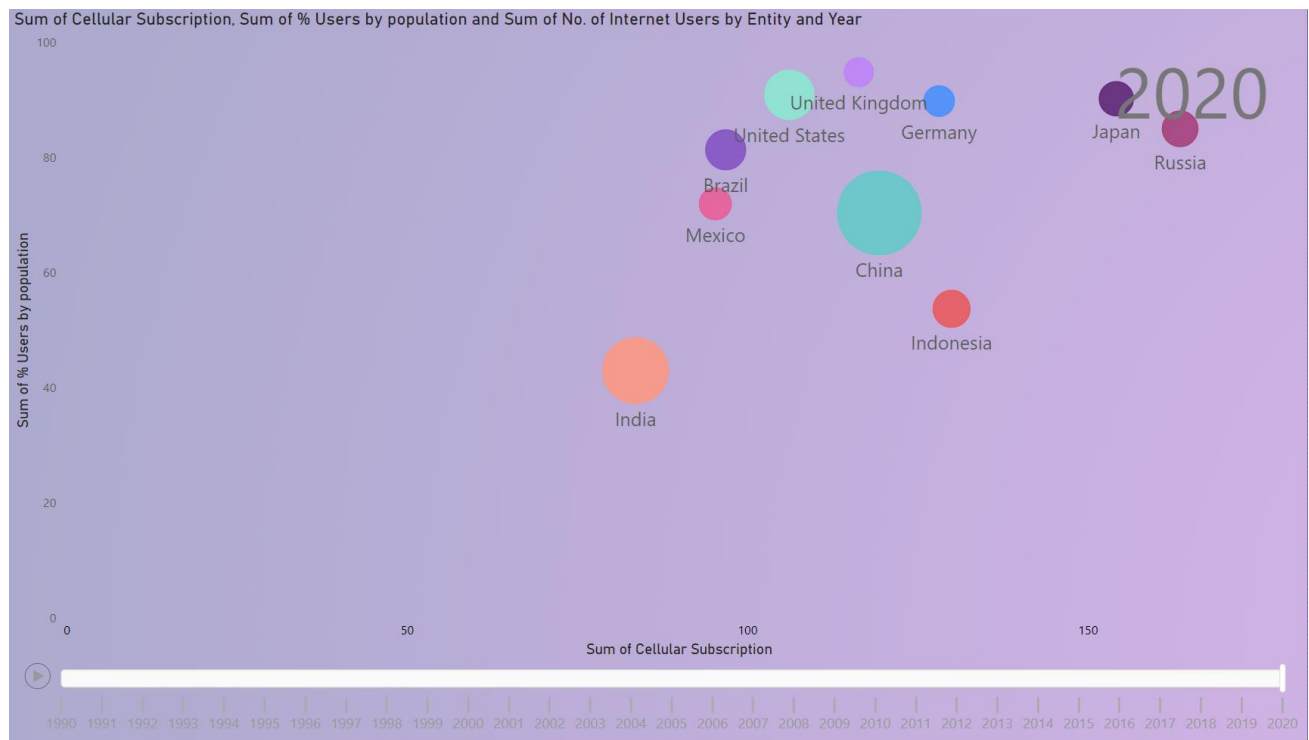
A scatter chart shows the relationship between two numerical values. In this project I have used scatter plot to determine the relationship between Number of cellular subscriptions to the % of users by population and the changes of these relationship over time. The color of the data points represent Entity(Countries) and the size represents total number of internet users in that country. I have also added a motion to monitor these changes over time from 1990 to 2022.

MAPS

Average of % Users by population by Entity
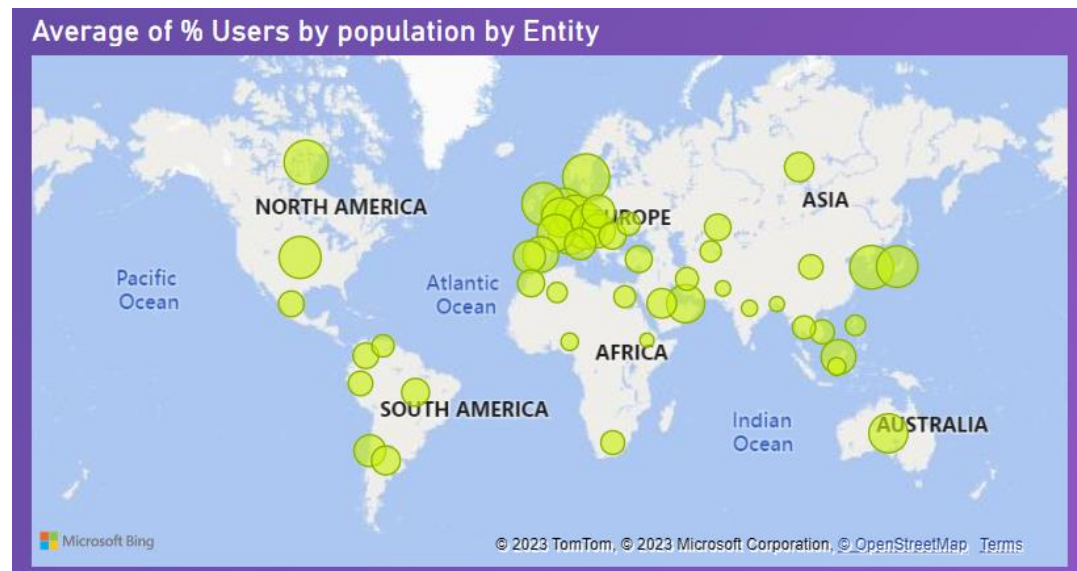
This visualization's main purpose is to compare internet users by regions on a map by sizing the bubble differently based on the average of the share of digital population. This gives a clearer picture of the regions that use Internet and the regions with lowest levels of internet adoption.

GLOBEMAP

Sum of No. of Internet Users by Entity
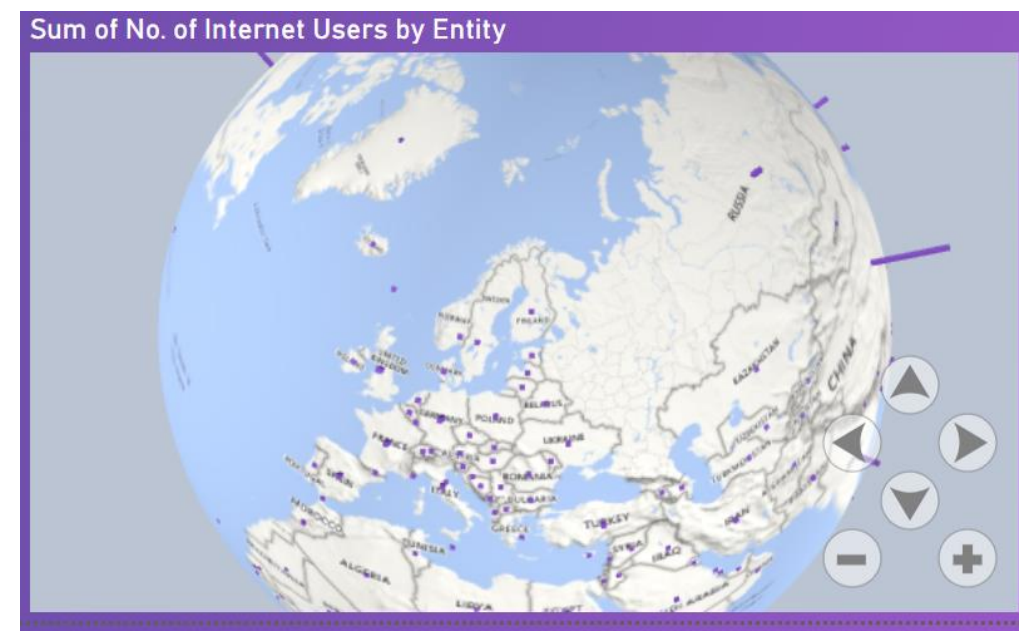
Globe Map is a 3D Map that makes the map exploration experience more interactive. It provides the sense of connection to the data with the number of internet users. This has been integrated with the play axis, Hence as the year changes the bar size on the globe changes in the respective countries.

PIE CHART
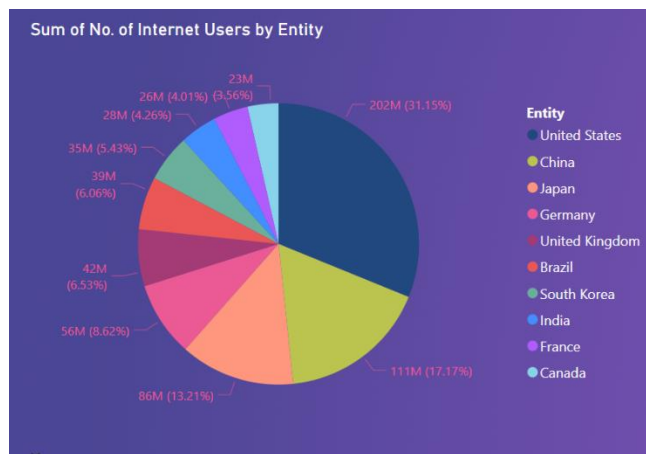
Pie charts display each data group as a separate slice on the chart. These charts are most commonly used to make comparisons between groups. I have used the Top N filters(10 entities) to demonstrate the changes of % share of population using internet over the years. I have referenced the "years" to play axis and used Edit Interactions to filter so that the slices in the chart move as the years change.
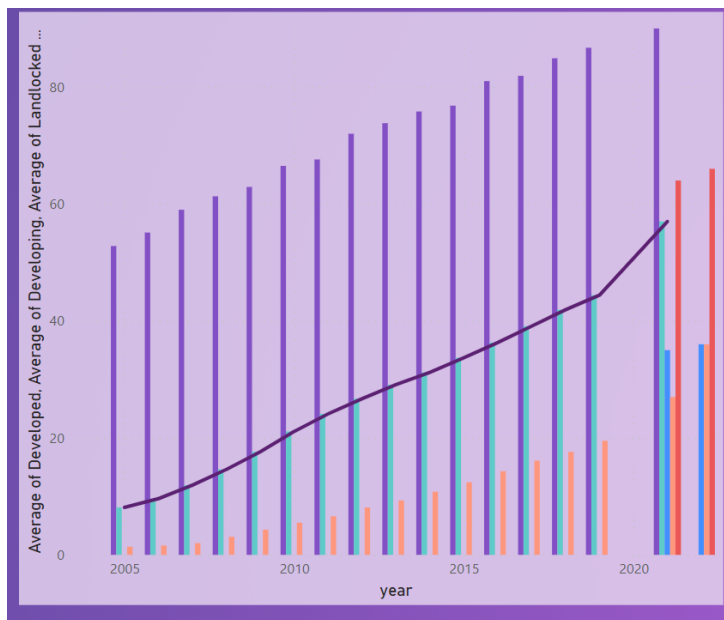
LINE CHART
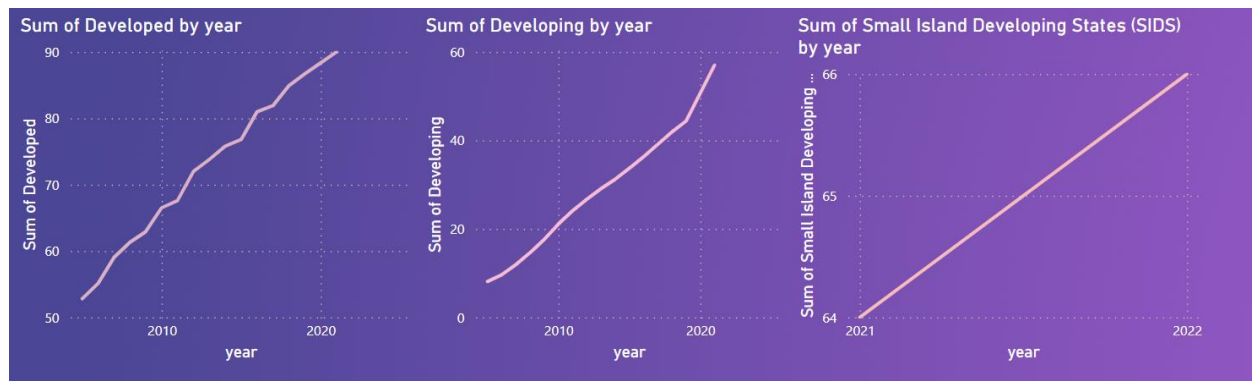
These graphs show how numbers have changed over time. They are used when you have data that are connected, and to show trends. I have used this chart to visualize the following:

1. Sum of No of users in developed countries over 3 decades
2. Sum of No of users in developing countries
3. Sum of users in SIDS

LINE WITH CLUSTERED COLUMN CHART

This is a combo chart where a single visualization combines a line chart and a column chart. Combining the two charts into one lets you make a quicker comparison of the data. Hence I have used Clustered column chart to represent different categories of countries by market maturity and line graph to represent the average % of internet users of the selected parameter.

GAUGE

I have used this chart to represent the average % of internet users in Developed, Developing, LDC , LLDC and SIDS countries. This chart is mainly used to show progress toward a goal, In this case, I have used the KPI to be the average of developed countries. This is used to represent a percentile and health of a single measure as compared to the progress towards the target value.

ANALYSIS OF THE DATA

In this project I was able to analyze how the internet continues to transform how we connect with others, arrange the flow of things, and share knowledge. The internet has become an increasingly important element of our daily lives, with its rising influence on both individual consumers and major economies. However, while the digital population is clearly expanding in many regions of the world, internet access and availability vary greatly by region.

BROADBAND AND CELLULAR USERS
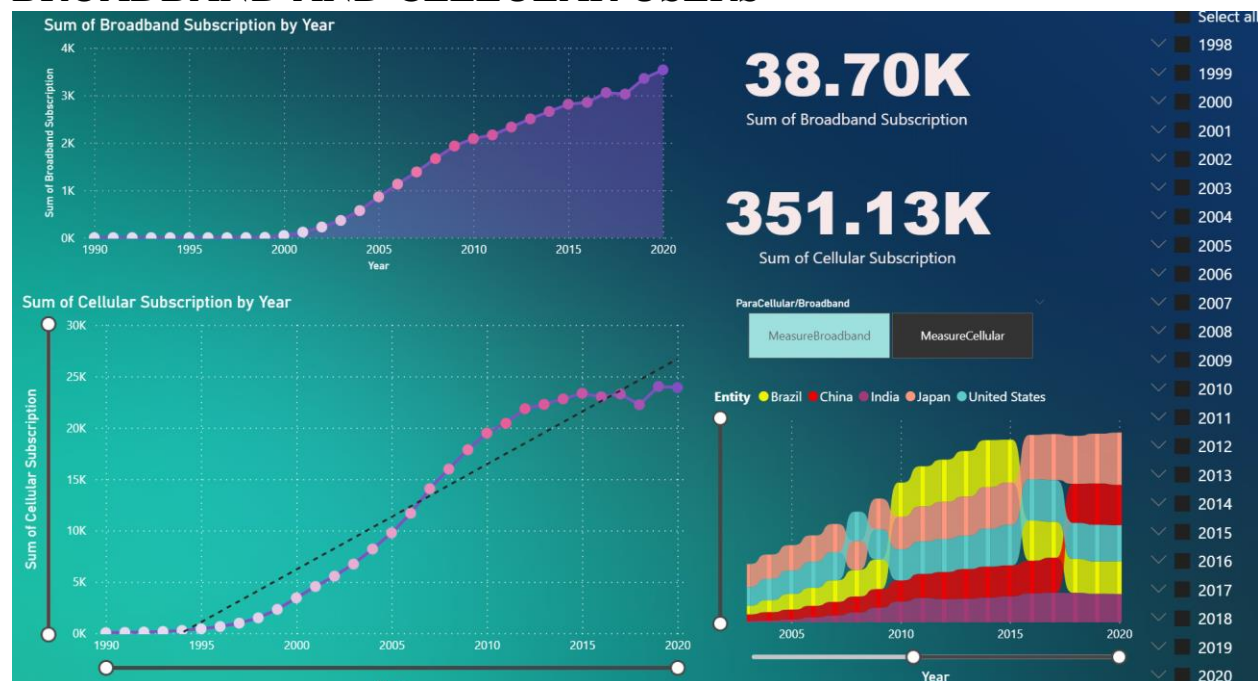
INTERPRETATION: Mobile internet has become increasingly widespread and popular over the past few years, as smartphones are more readily available and affordable than ever. As internet users are gradually switching to mobile devices to browse the web on the go, total Cellular subscription in 2020 is over 351K while broadband only had 38K which accounts for 1% of the total number of cellular subscriptions.

As we can see in the above image, there is a gradual decrease in the number of broadband subscription in Developed countries like USA, the developing country getting more and more access to the digital economy.

DIGITAL POPULATION WORLDWIDE
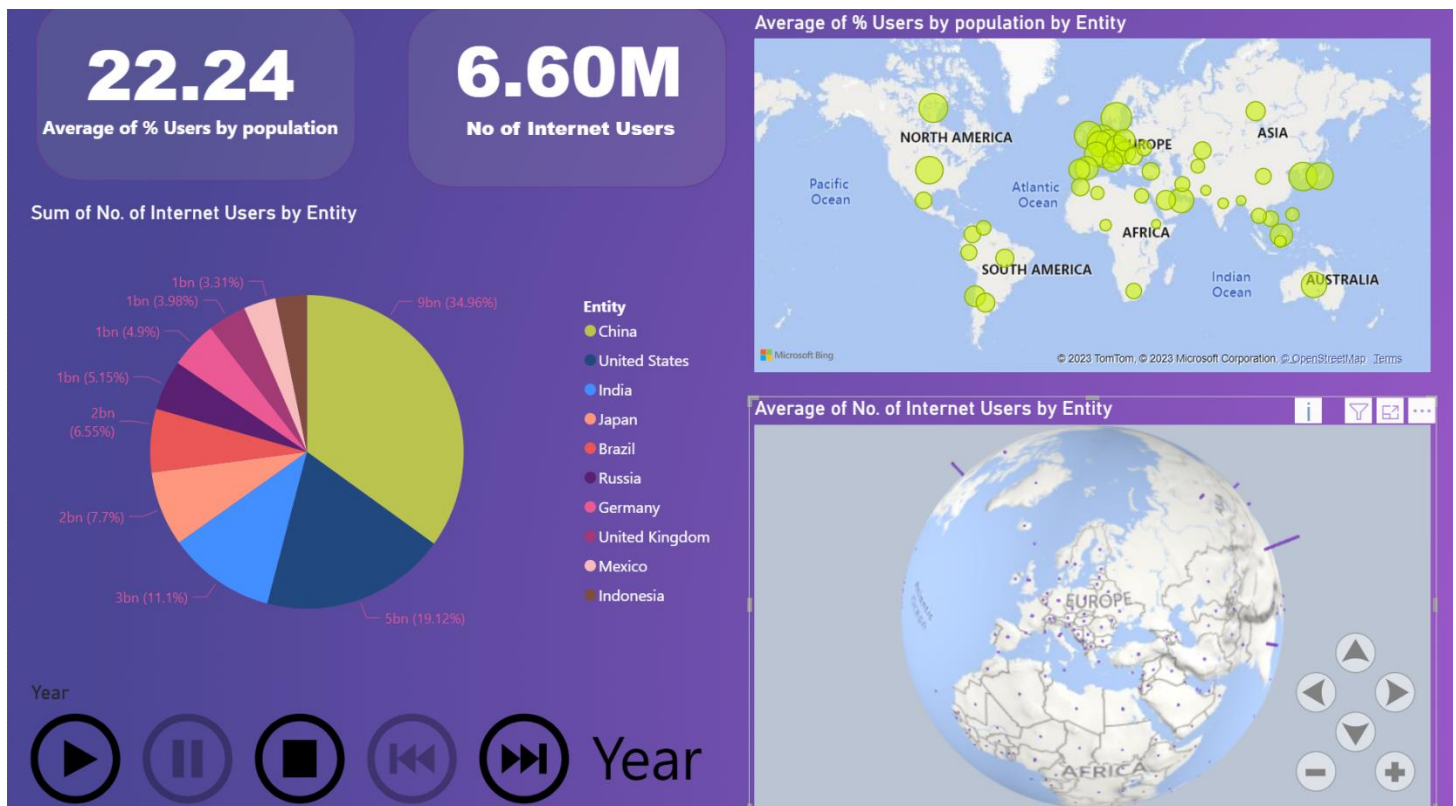
INTERPRETATION: As of 2021, China remains the leading online market in the world, followed by India and the United States. Overall, East Asia is the region with the highest number of internet users worldwide. During the 90's it was the developed countries like USA that had the majority of the online population with around 70% of the total share, but over the years we can see significant changes as the online penetration rate kept increasing.

PERCENTAGE OF POPULATION USING INTERNET BASED ON MARKET MATURITY
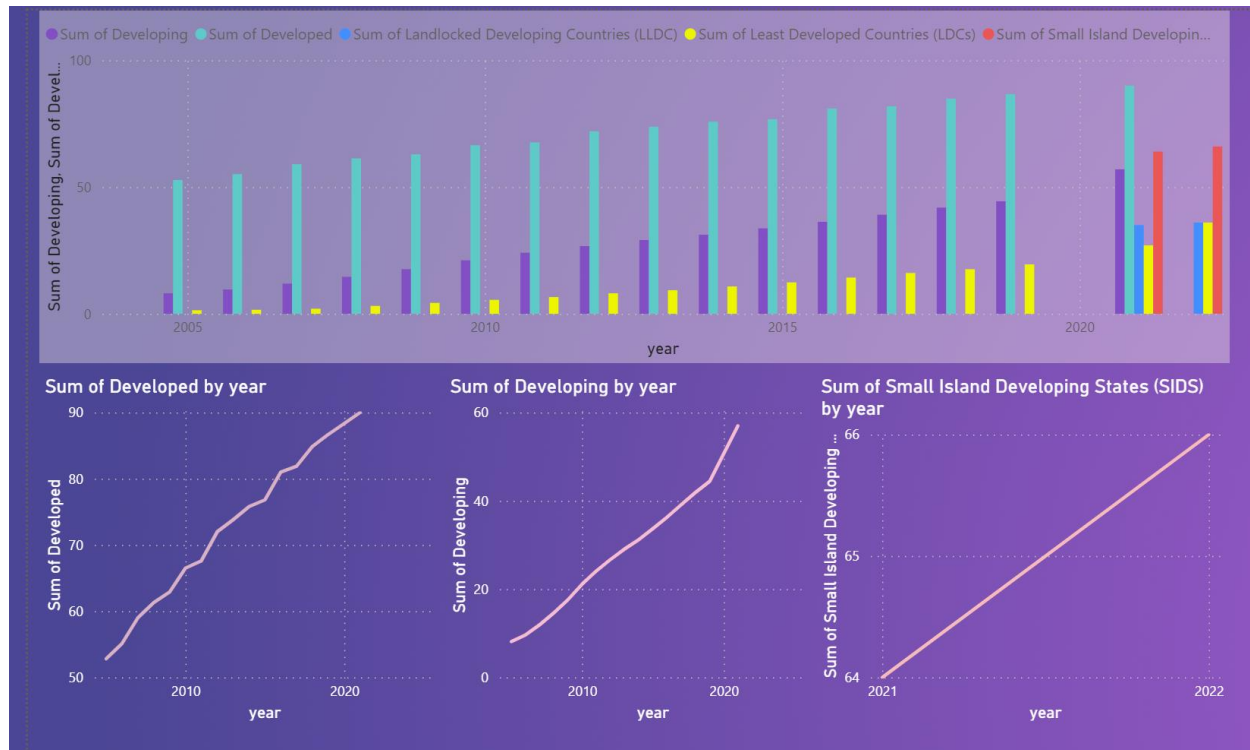
INTERPRETATION: Over the past few decades, the internet has also transformed the way products are bought and sold all over the world and how the countries have been evolving in the online population. Customer needs/desires do not appear to be evolving rapidly. A growing number of online transactions are carried out via mobile devices, making m-commerce an ever-more profitable industry. This defines the level of market maturity and we can see that the developing countries are seeing an increase of internet usage at a very high rate from 2010.

COMPARISON OF NO OF INTERNET USERS IN DIFFERENT CATEGORIES OF REGION

INTERPRETATION: As of 2020, we can see that around 89% of the population used internet, with the advancement in technology. Since most broadband services provide a continuous "always on" connection; there is no dial-in process required, the number of households in these countries with permanent internet access via broadband continues to rise. The various regions in these countries continue to have access to the internet.

PROGRESS OF DIGITAL POPULATION BY REGION TOWARDS THE TARGET WITH KPI BEING % OF INTERNET USERS IN DEVELOPED COUNTRIES.
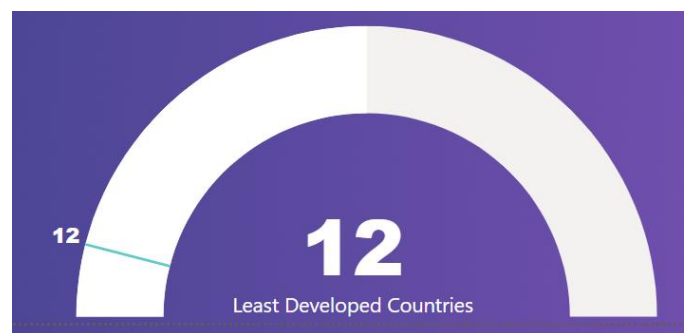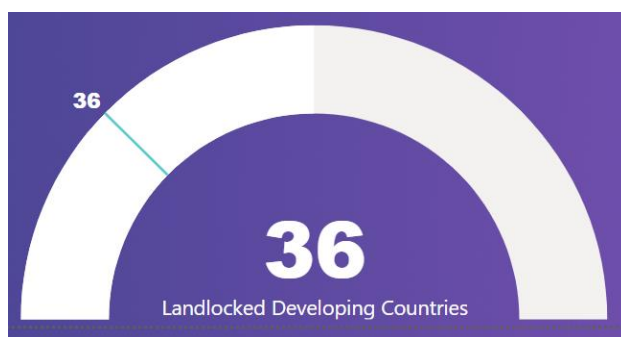
INTERPRETAION: We can observe that, on an average, 27% in the developing countries, 65% of the population used internet in 2022 in Small Island Developing States (SIDS) used the internet, compared to 12 % of individuals living in the least Developed Countries (LDCs) and a 36% of the share of those living in Landlocked Developing Counties (LLDCs). There are, however, stark differences in user distribution according to region.

TABLEAU VS POWER BI

	TABLEAU	POWER BI
Interactivity	Create interactivity by going to the dashboard tab and specifying details in the actions menu.	Create interactive dashboards by adding drill-through actions, tooltips, and drill-down hierarchies
Sheets/pages	Can add upto 1 visualization in a sheet.	Can add multiple visualization in a page
Dashboards	After creating one or more sheets, we can combine them in a dashboard and add interactivity	Automated dashboards which adds multiple visualizations in a single page and add interactivity
Maps	Able to add 3rd party maps from Other geocoding tools like mapbox etc	Able to add extensions from "Get more visualizations"
Automated Visualization	Show me automatically analyses the specified fields and provides you with a selection of view types that are acceptable for those fields.	Selected fields will be automatically updated in a tabular visualization
Query language	Uses advanced query language to perform data analysis	Able to perform analysis with Natural language
Tabs (for visualization)	Data & analytics	Data & visualization
filter	Slicer can be used for interactivity for many visualizations	Filter option available while creating a visualization. And using the filter in dashboard menu for interactivity
Motion	Can be created by applying criteria in pages shelf - Only uses shapes (Circles,square)	Can be created by applying criteria in play axis - Only scatter plot
Story	Can create story and published in Tableau public	Can create powerpoint
Q&A	Not available	available
Calculations	Perform calculations in analysis tab	By creating measures
Forecasting	Available in analytics tab	Does not have a built in forecast query
Models	Available in data source page - logical layer(relationships) - Physical layer(Joins)	Available as a separate model view to view tables, columns and relationships

CONCLUSION

While many of us cannot imagine our lives without the services that the internet provides, the key message for me from this project of the global usage of the internet is that we are still in the very early stages of the internet. It was only in 2017 that half of the world population was online; and in 2018 it is therefore still the case that close to half of the world population is not using the internet.

Two factors, in my opinion, have recently highlighted its evolution: the social web and mobile technology. These two innovations have changed the way people use the Internet. In the social web people have found a new way to communicate. Since its creation in year 2004, Facebook has grown into a worldwide network of over 2,803 million active users. Mobile technology, on the other hand, has made possible a much greater reach of the Internet, increasing the number of Internet users everywhere. According to Datareportal, The number of smartphones in use is growing at an annual rate of 4.6 percent, with more than 300 million new smartphones coming into active use over the past 12 months.

Overall, I believe that the internet has already changed the world, but the big changes that the internet will bring still lie ahead. The E-Commerce is making its mark on the digital population, The rise of e-commerce has transformed the way people shop globally. Online retail giants like Amazon and Alibaba have revolutionized the shopping experience, offering convenience, a wide range of products, and global delivery. The COVID-19 pandemic further accelerated the growth of online shopping as people sought to avoid physical stores. The demand for digital entertainment continues to grow, leading to the emergence of new streaming services and the expansion of existing ones. IoT devices, such as smart home devices, wearables, and industrial sensors, are becoming more prevalent, contributing to the growth of internet usage and generating vast amounts of data. Although there are still regions with limited access. Efforts are being made to bridge the digital divide and bring internet connectivity to underserved areas, such as Rural areas in Developing countries and least developed countries. Internet usage history has just begun.

3 W's

1. **What went Well?**

Creating the visualization: It was fun creating the visualization. I was able to understand and implement the Do's and don'ts of creating a Chart. I was able to relate the chapters in How Charts lie while creating this project.

Documenting and Digital Literacy: This project turned out to be very helpful for me as I learned how to document every step and perform the project systematically. This project helped me gain more insights towards the evolution of the internet, as it is a global phenomenon and the importance of continuous learning and digital literacy skills.

Tool: Power BI is a very interesting tool, which allowed me to unleash my creativity while preparing the charts and visually appealing dashboards. I was also able to learn new techniques and DAX formulas.

2. **What did not go well?**

Data collection: Initially, I tried to collect data from reputable sources such as the International Telecommunication Union (ITU), Internet World Stats but the data was very vast and I could not find a funnel approach to narrow it down to one single level.

Power BI forecasting: I was very curious to know the forecasting capability of Power BI to determine if 2023's usage was accurate but forecasting relies more on Power Query (a separate tool) for advanced data preparation tasks.

Extensions: I tried various visuals like Globe data bars,Sankey,ChartExpo which looked very appealing to me, but most of them were paid version. Some which were free asked to enter email address to proceed but did not actually work on the data.

3. **What would I do differently next time?**

More data: Next time I would address other aspects of internet usage like social media/ecommerce/demographics of the usage to gain more insights into the topic

Advanced Interactivity: by using advanced interactivity like defining navigation paths, interactions, and animations would help me build an immersive experience, enabling to view the report insights in a structured manner.

REFERENCES:

DATA SOURCE:

1. Kaggle: Global internet usage:

Global Internet Users: A Geographical Analysis. (n.d.). Global Internet Users: A Geographical Analysis | Kaggle. https://www.kaggle.com/code/sanjay277/global-internet-users-a-geographical-analysis

2. Statista: Internet usage by market maturity:

Global internet penetration by market maturity 2022 | Statista. (n.d.). Statista. https://www.statista.com/statistics/209096/share-of-internet-users-worldwide-by-market-maturity/

3. Other References:

Overview. (n.d.). World Bank. https://www.worldbank.org/en/topic/digitaldevelopment/overview

Digital Around the World — DataReportal – Global Digital Insights. (n.d.). DataReportal – Global Digital Insights. https://datareportal.com/global-digital-overview

Rodriguez, C. (2021, December 2). *Why A Third Of The World, Nearly Three Billion People, Have Never Used The Internet*. Forbes. https://www.forbes.com/sites/ceciliarodriguez/2021/12/02/why-a-third-of-the-world-nearly-three-billion-people-have-never-used-the-internet/

4. Power BI :

Global Internet Users - Power BI