**ISM 6361 - Data Visualization Spring 2020**

**Social Media Statistics using (Power BI)**

**Seattle Pacific University**

**By Satadipa Sarkar**

**1.State a business reason for selecting your tools (problem you would like to solve)**

Power BI – There are plenty of visualization tools in the market like Google Charts, Tableau, Grafana, Chartist. js, Fusion Charts, Data wrapper, Infogram, Chart Blocks etc. I always wanted to learn and work on Power BI and Tableau because most of the beautiful visualizations I came across in Reddit, New York Times about Corona Virus statistics and updates were either in Tableau or Power BI. Also, Power BI and tableau are the most sought-after skills or aptly said good to know tools for a data analyst resume.

Power BI is a business analytics service provided by Microsoft available to general public as a service embedded in Cloud with a desktop-based interface. The aim was to enable everyone create beautiful visualizations without depending on database administrators or developers to write some code and create reports and charts for making important decisions. It offers data cleaning and preparation (warehousing) activities. With Power BI, we can search and look for any public data sources (ex: Wikipedia) we want to do the analysis using search(online). I have used Microsoft excel as my data file.

The project I worked on is to understand the growth of internet and social media globally and how COVID-19 has affected the social media. My data set needed to be cleaned and joined for appropriate fields and columns. Power BI – Power Query was very helpful in building customized tables and mostly setting up the primary key and foreign key relationship was much easier than Tableau. Scheduling Data refresh was very easy too.

Power BI is very rich in visuals- In comparison with Tableau I felt the Get More Visuals had a very rich collections of visuals which could be imported into the datasheet very easily to build visualizations. The word cloud, infographic designer, play axis etc. was very helpful visuals in my project that helped in charting the data easily. I hope that readers or viewers won’t need to have much “graphicacy” to understand the content of my chart.

Page layout with various features to change the background, importing images, showcasing dynamic data, custom visuals were very interesting to work with. This was very important because I had to design the report background according to the various social media network logos etc. It was super simple to design the report in terms of color, fonts, title, backgrounds etc.

Natural Language Q&A was an add-on. Even though I didn’t end up using it much, but it is certainly a very cool feature to have in the data visualization tool.

Power BI interface is very easy to learn and does a great job in reporting.

**2. Document how/where you got your data (if it is publicly available, or internal for a work project).**

The data for this project is publicly available and it is available in the following websites:

1. <https://en.wikipedia.org/wiki/List_of_countries_by_number_of_Internet_users>
2. [www.statista.com](http://www.statista.com)
3. <https://gs.statcounter.com/social-media-stats/all/worldwide/2019>

Finding the right dataset to build a story from the rise of the internet, social media growth, usage in different countries, impact of COVID-19 on social media usage was very difficult. I had to look into number of sources to find the right data and then combine files together which I thought would be interesting to work on. Through this data set I am trying to display the social media statistics now, rankings of the different social media platforms, how broadband, mobile usage of social media have resulted in deep penetration rate for social media in various countries.

There was huge amount of data in respective websites about COVID-19, impact on social media. It will be great to study most of it, but this project mostly contains data about advent of internet, mobile, broadband by countries and the rise of social media and its usage comparison in different counties- to visualize the social-media statistics of the world. The project also contains data about the impact of COVID-19 on social media usage in U.S by age and gender, most trusted sources of news for information about the coronavirus in the United States as of March and how the people of U.S. used social medial to digitally socialize during this pandemic according to some research conducted by Ipsos, Morning Consult National Tracking Polls, Corona Virus Impact on Influencer Marketing (IZEA), PEW Research Center.

P.S -The impact of COVID-19 on social media can be a different project altogether.

**3.Document how you used your tool. Many tools are super rich in features and you probably will not be exploring all the features, explain the parts you did use.**

Power BI helps to build stunning visualizations with the help of many in-built features. For this project I have used the following feature:

**Get Data** – The most important task of building a visualization is to bring the data into the worksheet from various data sources. Thankfully, Power BI accepts to load data from number of data sources like SQL server, Access, Oracle, Postgres etc. The data source used for this project have .xlsx format. Power BI could load the data without any delay and hiccups.

**DAX (Data Analytic Function)**- DAX formula Function have many libraries that can perform many computations on data to build powerful data models for the visualizations. I did use many formulas for calculating percentages and change in usage pattern for different social networks. Renaming, changing header, replacing the formula with updated one was very easy and simple to do. I have used calculations to understand the difference in usage pattern of the leading social media platforms like Facebook, Instagram, LinkedIn, Twitter etc between two countries (U>S and India).

1. The various calculated field are: Difference in Monthly Users (Facebook), Difference in Monthly Users (Instagram), Difference in Monthly Users (Twitter), Difference in Monthly Users (Youtube).
2. Also, used DAX formula to calculate the percentage of the population using internet till 2017. I have calculated the percentage of internet users from the population data provided.

**Power Query Editor** – Power Query Editor was helpful in building new tables from the already existing relationships between tables. For creating new parameter, I had to build a new Query table which incorporated the required columns from the ‘Percentage of Pop’ sheet.

Transform Data -> Set New Parameters🡪 Populate the parameter window🡪 Apply & close🡪 Manage Parameter🡪 Right click on the column of the table you want to set the parameter for🡪Apply filter🡪set the right parameter 🡪Click Ok

**Custom Visualizations** – Different business needs different visualizations and Power BI have ample custom visualizations of your choice. To get custom visuals, under visualizations tab click on the three dots 🡪 Get custom Visuals🡪New Window opens with vivid choices🡪Select the visual🡪Import custom Visual🡪Click Ok.

The visuals make charting more artsy and aesthetic. I have experimented with number of custom visuals as mentioned below:

**Play Axis**

A picture containing drawing

Description automatically generated

The play axis visual has been incorporated in number of reports in this project. This helped me to project the data in various charts sync together in motion.

**Infographic Designer 1.8.7**

**A screenshot of a cell phone

Description automatically generated**

The infographic visual was used to demonstrate the popular social network site by gender, to show the share of adults in US who have used digital communications to connect with others during COVID-19 by age groups and also the gender who have socialized digitally more during this pandemic.

A screenshot of a cell phone

Description automatically generated

**Word Cloud**

The world Cloud visual was used to showcase the social media platforms which stands out from others globally and the countries which end up using more social media.

A picture containing holding, people, table, woman

Description automatically generatedA screenshot of a cell phone

Description automatically generated

**Heat Map**

The heat map was used to chart active usage penetration in selected countries 2020 in percentage %.The display tells the share of internet users in selected countries visiting social networking sites (based on the number of active accounts for top social networks by population).

A picture containing computer

Description automatically generated

**Image by Cloudspace**

The image by cloudspace visual was used to chart the flags of the respective country according the slicer applied. Charting selected images in PowerBI was simple with this visual.It displays images using an Internet-accessible URL from the data table.

**A picture containing screenshot, drawing, table

Description automatically generated**

**Slicer**

Slicer in power BI serves as an alternate option to filter the dataset. It enables to sort and filter the report according to the need of the user. Slicers are used in almost every report in this project.

**Fields/Format**

The fields and format section is very helpful in viewing the columns of the various data table, which we want to chart. The Format section has three levels in Power BI – 1. Format the back-ground page of the report 2. Format the data/visual section 3. Data Formatting option. I have used custom images for the various reports which makes it look a lot better.

**Data**

The data tab on the left side of the sheet/canvas was useful for viewing the underlying the dataset, renaming columns, refreshing the dataset from the original source and calculating new column.

**Various other visuals**

**Power KPI**

The Power KPI adds labels for the current date, value, and variances. The KPI Indicator presents the status as a color indication, comparing the actual and target values. This visual helped to distinguish between various social media platforms usage in US and India.

**Donut chart –** To visualize the leading social media platforms for selected countries and their percentage of audience size.

**Stacked column, Stacked bar, maps, filled map, line chart, pie chart, forecast was also used to plot suitable data.**

**4. Explain why you chose which visualization/charts.**

**Line chart and Map** – To show growth of internet users worldwide from the year 1990-2017 and how this technology has taken over the different countries year by year was projected with a motion play axis). The data was projected with two different visuals to provide an overall understanding of the advent of the internet and the growth of internet users.

A screenshot of a cell phone

Description automatically generated

**Pie chart & Clustered column chart** – This chart shows the number of fixed broadband subscriptions per 100 people from 1998-2017. The pie chart shows the top seven countries with the world data. This refers to number of people having access to high-speed internet connection(public) per 100 people.

Mobile phones have been one of the most disruptive technology to change all business. The data shows how mobile adoption started rising dramatically from 21st century but it was so slow rise or almost zero in 1900’s for few countries. This was interesting to see in play axis too. This chart shows the number of mobile cellular subscriptions per 100 people from 1998-2017. Clustered column chart demonstrates the different countries with mobile cellular subscription per 100 people.

A screenshot of a cell phone

Description automatically generated

**Line chart in motion with play axis –** Social media started becoming from the early 2000’s. This chart represents the number of monthly active users by platform from the year 2004 till 2019. It shows the growth of each platform by year and how Facebook, You Tube and Reddit are present for 10 or more years now with Facebook being the highest.

A close up of a map

Description automatically generated

**Filled Map & Heat Map** – The filled map represents Global mobile social penetration rate 2019, by region where Eastern Asia is leading with 70%. Asia was ranked first with mobile cellular penetration rate of 70% followed by North America.

The heat map displays the share of internet users in selected countries visiting social networking sites (based on the number of active accounts for top social networks by population). UAE, Taiwan & South Korea topping the chart.

A picture containing indoor, monitor, photo, sitting

Description automatically generated

**Ribbon Chart** – The ribbon chart was used to demonstrate Number of monthly active Instagram, Facebook, Whatsapp, Snapchat, users from January 2013 to June 2018 (in millions) by Year. The ribbon chart demonstrates the active users-quarterly data from 2013-2020. Whatsapp have two billion monthly active users as of March 2020, snapchat 229 million daily active users.

A screenshot of a cell phone

Description automatically generated

**Donut chart** – Donut chart was used to demonstrate the Leading countries based on Snapchat, Facebook, Snapchat, Instagram, Twitter audience size as of April 2020 (in millions) by Countries. India has the largest audience for Facebook whereas United States ranks highest for Twitter, Snapchat and Instagram.

A picture containing screenshot

Description automatically generated

**Power KPI with line chart**

This visual helped to distinguish between various social media platforms usage in US and India. It represents Active Monthly Users (AMU)for Instagram, Facebook, Twitter , Youtube for both the countries and also shows the difference in number of users by month and year in motion. The chart demonstrates the difference in users of leading social media in a developed and developing country.

A screenshot of a social media post

Description automatically generated

**Gauge**

Gauge was used to chart the internet users and the population of the country along with the display of country Flag.

A screenshot of a cell phone

Description automatically generated

**5.Give an explanation/analysis of the output. What did you learn or uncover?**

* The idea behind this project was to uncover and demonstrate the social media facts and statistics of the world. The internet and mobile technology have blossomed over the social media companies to make them achieve a huge reach and penetration rate all over the world. The global social penetration rate have reached 49% with active social media users of 3.8 billion.

A screenshot of a cell phone

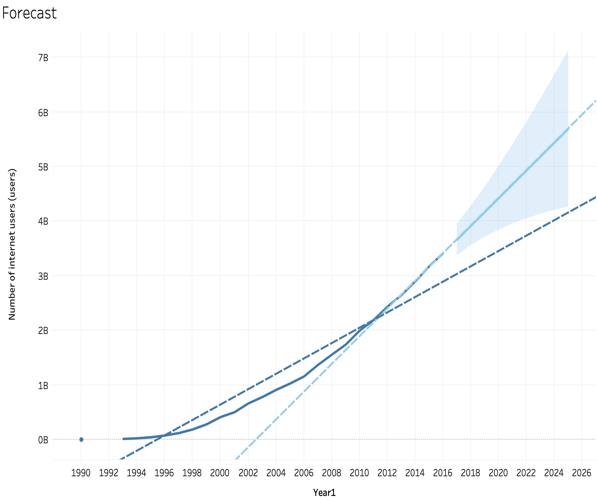
Description automatically generated

* The growth of internet users worldwide from the year 1990 till now is increasing at a huge rate worldwide. Internet users increased from 413 million in 2000 to over 4.6 billion in 2020. The top three countries are China, India and United States.
* Fixed Broadband subscription per 100 people was highest in South Korea followed by Germany and United States.
* The number of mobile cellular subscriptions per 100 people increased dramatically after 21st century and the growth were very slow in 1990’s.
* Globally Facebook is the top social media company according to the number of users and have the presence for 10 more years. YouTube, WhatsApp (Facebook company), Instagram are catching up.
* Global mobile social penetration rate is increasing. Asia was ranked first with mobile cellular penetration rate of 70% followed by North America.
* Whatsapp have two billion monthly active users as of March 2020, snapchat 229 million daily active users as of 2018.
* India has the largest audience for Facebook whereas United States ranks highest for Twitter, Snapchat and Instagram
* During COVID-19 more women interacted digitally with others than men
* It was interesting to see that even though the usage of social media is huge, but the most trusted source of Corona Virus news was CDC, WHO. Social Media Ranked last.
* People believe that the usage of Facebook, YouTube, Instagram, Twitter will increase if the pandemic lockdown continues more.
* People in the United States, age group of 18-29 years searched more about corona virus news online and shared the greatest number of posts through social media.

**6. Compare/Contrast your vizzs from Tableau to Power BI, what were the strengths/weaknesses of one product’s solutions over the other.**

**Forecasting** – Forecasting chart tried to display the number of internet users by year and its expected pattern for the next few years. Both Tableau and Power BI could plot the data properly with line chart showing the trend line at 99% confidence interval. With power BI. In power BI, the forecast editing options were under Analytics tab where we could set option for color, seasonality, interval levels etc. In Tableau if you just right click on the forecast data, under forecasting options, we can find all the necessary options. Adding the background image was easier in power BI.

A close up of a map

Description automatically generated 

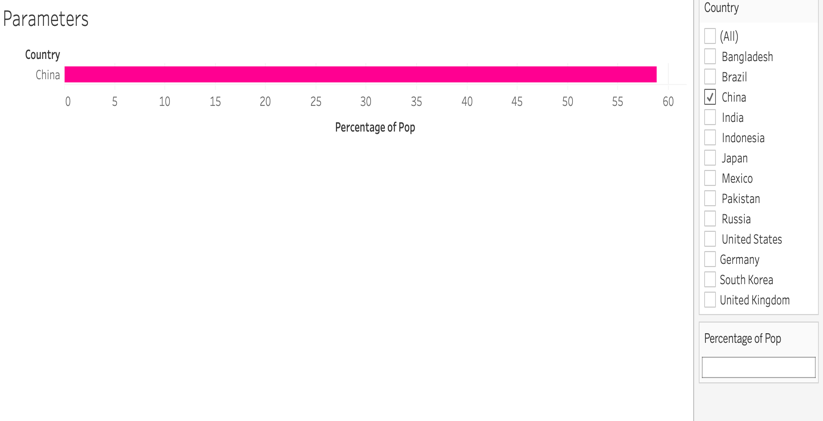
**Adding Parameters**

Tableau certainly wins in this aspect. Adding Parameters was very easy in Tableau by creating calculated fields and then setting parameters accordingly. In Power BI, adding parameters requires a few more steps than Tableau and is lot more complicated. But I could achieve the same result in both the tools. (YAYY!!)

For creating new parameter, I had to build a new Query table which incorporated the required columns from the ‘Percentage of Pop’ sheet.

**Transform Data -> Set New Parameters🡪 Populate the parameter window🡪 Apply & close🡪 Manage Parameter🡪 Right click on the column of the table you want to set the parameter for🡪Apply filter🡪set the right parameter 🡪Click Ok**

In Power BI, we need to click on Transform Data🡪Select Manage Parameters🡪Select the parameter we want to see the data of. Note: Like Tableau, it won’t be visible as parameters on the worksheet on the right panel.

 A screenshot of a social media post

Description automatically generated

Tableau Power BI

A screenshot of a cell phone

Description automatically generated A screenshot of a social media post

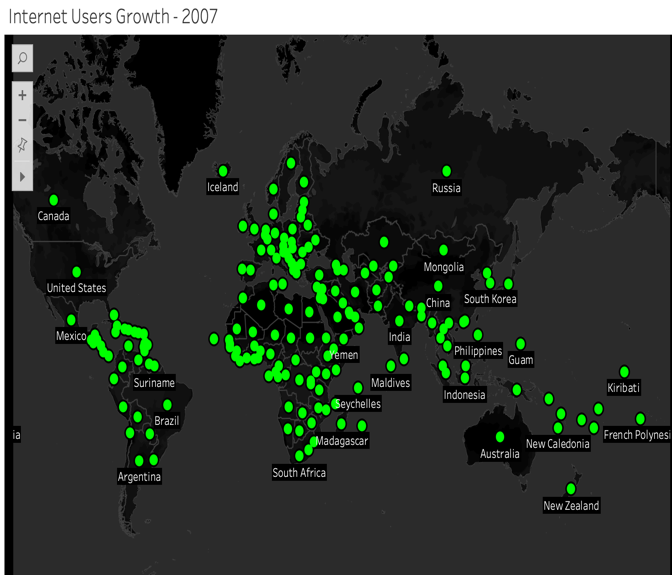
Description automatically generated

Power BI Tableau

**Maps/Other Visuals**

Creating maps and other visuals were more or less equally easy for both the tools. But I do feel the availability of custom visuals in Power BI makes the visualizations more customized to the different business problems.

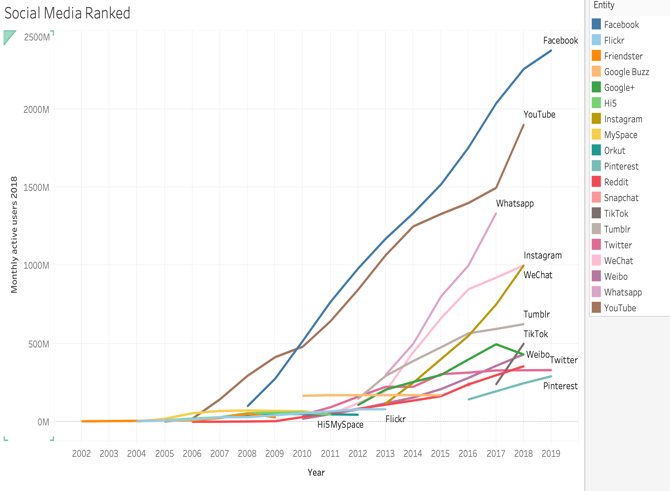
A screenshot of a cell phone

Description automatically generated 

Power BI Tableau

**Charts with motion -**was very easy (accessibility) with Play axis tool compared to Tableau where we have to add the filter on the page for the charts to be in motion and then modify the history, speed etc. In power BI the play axis tool was easy to customize with the animation settings under formatting tab.

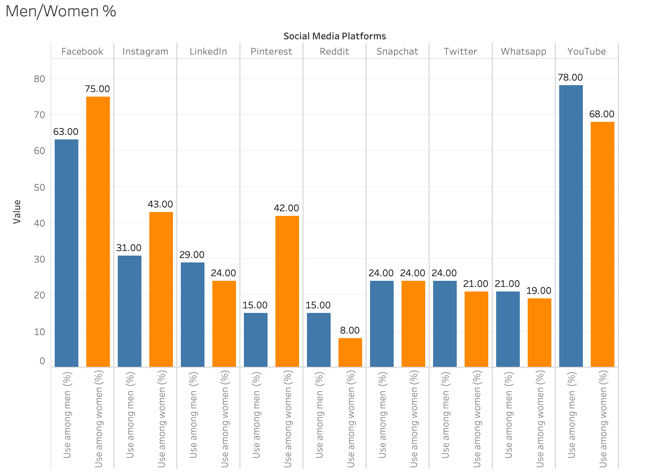
A close up of a map

Description automatically generated 

Power BI Tableau

**Word Cloud and infographics** – In Power BI it was very simple to chart by importing custom visuals in the work sheet. I haven’t explored that option for Tableau yet. This chart displays percentage use of different social media platforms by gender.

A screenshot of a cell phone

Description automatically generated 

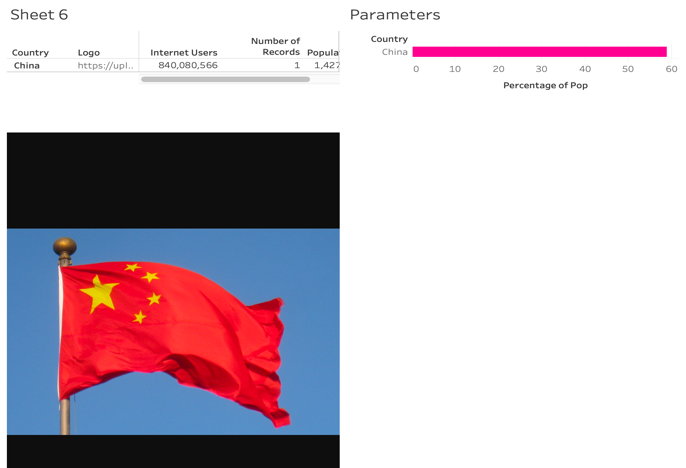
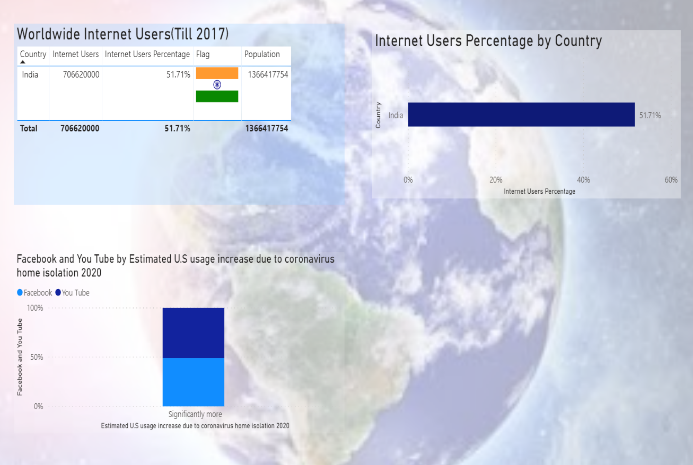
Power Bi Tableau

A screenshot of a cell phone

Description automatically generated

Power BI

**Using Web URL to display data** was very quick and easy in Power BI with Image B Cloudscape visual. In Tableau, I found it to be bit harder implementing this, I could do it in the dashboard only not in worksheets. The flag here is the Image URL.



Power BI Tableau

**7.Conclude with the 3 W’s (What Went Well, What Did NOT go Well, What Would you do Differently Next Time).**

**What Went Well**

**The whole data visualization experience was good using Power BI** starting from getting/importing the data from external sources, editing Query, creating visualizations and at the end publishing the reports and dashboards to the power BI workspace.

**Power Query Editor was very useful in formatting**, **creating new query by creating new table** by looking at relationships between tables. Once the data upload was complete, charting the required fields with plethora of visuals was handy. Creating multiple reports in one single page was easy which helped me to chart all the related content together for better story building.

**Formatting and report designing** – I could demonstrate my ideas in terms of designing and formatting without spending lot of time. Importing custom backgrounds, editing the plot area, labels, colors opacity according to a theme that the data represent was very easy and hassle-free.

**Customized visuals helped to create interesting, attractive visualizations.**

**Data Refresh was very quick from the original data source**. Changing a particular column, calculating new column with the DAX was useful.

**What Did NOT go Well**

**Creating Parameters in Power BI involves lot of steps**. There are many ways of doing it, but all are complicated as compared to Tableau. As mentioned above the steps involved to create a parameter takes bit of time but eventually, I could get the parameters up and running.

**Power BI does not have story tab like Tableau** where many worksheets can be combined to tell a story from the data. This is one of the unique and most important feature of data visualization because in the end every user wants to learn what are we trying to tell from the visualizations. Absence of this functionality (story-telling tab) in Power BI made it difficult to highlight the important data points from the chart. As an alternative I tried to put a title to all my reports and worksheets and grouped visualizations which are related in one single report and dashboards.

**Dashboards in Power BI** doesn’t have some extra-functionality like Tableau. Tableau has distinguished dashboard layout which made it easy to understand and organize the sheets, apply filters accordingly. But Power BI have a different structure of worksheets, reports and dashboards but no distinguished layout/tab for dashboard.

**Segregation of dimensions and measures** made it easier to recognize the text and numeric fields in Tableau. But Power Bi does not have that segregation, so I had to look into the data tab to understand the data type of the columns and categorize other calculated fields.

**What would you do differently Next time?**

I will try to implement the create Parameters in Power BI by another process and check if it is more convenient.

There are so many visuals available in custom visuals library, I would like to explore more and more visuals and check how it applies to the existing data set.

Also, I would like to figure out options to build the dashboard in more sophisticated way.

Import data set from two or more sources and check how Power BI performs on that aspect.