西南科技大学

Southwest University of Science and Technology

Project Link: Data Visualization

西南科技大学

学院名称: 计算机科学与技术

专业名称: 计算机科学与技术(本科)

学生姓名 : MD Aynul Islam (叶子)

学号：4420190030

授课教师：王松老师

Contents

1 Abstract

Tableau is a data visualization software that allows the user to not only build reports but perform exploratory analysis in order to build out a smart story. I peck some of Asian country and I tried to show some visualization respect to the dataset. Preprocessing is the process of doing a pre-analysis of data, in order to transform them into a standard and normalized format. After that, By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data. There are exceptions to the variety of output criteria, though. Some data visualization tools focus on a specific type of chart or map and do it very well. Data can be generated, captured, and stored in a dizzying variety of structures, but when it comes to analysis, not all data formats are created equal. Data preparation is the process of cleaning dirty data, restructuring ill-formed data, and combining multiple sets of data for analysis.

2 Introduction

We need data visualization because a visual summary of information makes it easier to identify patterns and trends than looking through thousands of rows on a spreadsheet. Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data. We will always wax poetically about data visualization .there are practical, real-life applications that are undeniable. And, since visualization is so prolific, it's also one of the most useful professional skills to develop. The better you can convey your points visually, whether in a dashboard or a slide deck, the better you can leverage that information. However, because both design skills and statistical and computing skills are required to visualize effectively, it is argued by some authors that it is both an Art and a Science.[1]

3 what is data visualization

Data visualization is an interdisciplinary field that deals with the graphic representation of data. From an academic point of view, this representation can be considered as a mapping between the original data. Data visualization has its roots in the field of Statistics and is therefore generally considered a branch of Descriptive Statistics. [2]

4 Data Visualization Types

Part of the strategy of visualizing data is choosing what type of data visualization to use. The trick is to select the one that will best represent your data's message and story.[3]

Bar Chart	Pie Chart	Donut Chart	Scatter Plot
Polar Graph	Tree Chart	Percentage Bar	Histogram
Venn Diagram	Network Diagram	Gantt Chart	Concentric Circles

5 Data visualization tools and techniques

Data visualization tools provide data visualization designers with an easier way to create visual representations of large data sets. When dealing with data sets that include hundreds of thousands or millions of data points, automating the process of creating a visualization, at least in part, makes a designer's job significantly easier. The best data visualization tools on the market have a few things in common. First is their ease of use. There are some incredibly complicated apps available for visualizing data. Some have excellent documentation and tutorials and are designed in ways that feel intuitive to the user. Others are lacking in those areas, eliminating them from any list of "best" tools, regardless of their other capabilities.

The best tools also can output an array of different chart, graph, and map types. Most of the tools below can output both images and interactive graphs. There are exceptions to the variety of output criteria, though. Some data visualization tools focus on a specific type of chart or map and do it very well. Those tools also have a place among the "best" tools out there.

6 Data Preprocessing

In the real world, we usually come across lots of raw data which is not fit to be readily processed by data visualization. We need to preprocess the raw data before it is fed into various data visualization. we have various techniques for preprocessing data in Python.

6.1 Data Preprocessing with Python

Preprocessing is the process of doing a pre-analysis of data, in order to transform them into a standard and normalized format.

Preprocessing involves the following aspects:

- missing values
- data standardization
- data normalization
- data binning

6.1.1 Import data

Firstly, import data using the pandas library and convert them into a dataframe. Through the head() method we print only the first 10 rows of the dataset.

```
[ ] import pandas as pd

df = pd.read_excel('/content/drive/MyDrive/Datasets/Web_Scrapped_websites.xls')
df.head()
```

	Country_Rank	Website	Trustworthiness	Avg_Daily_Visitors	Child_Safety	Avg_Daily_Pageviews	Privacy	Facebook_likes	Twitter_mentions	G
0	1	www.google.com.af	Excellent	NaN	Excellent	NaN	Excellent	9	1	
1	2	www.google.com	Excellent	515 007 350	Excellent	4 192 159 833	Excellent	94.2K	11.2K	
2	3	www.youtube.com	Excellent	506 457 282	Excellent	2 679 159 025	Excellent	13.5K	16.5K	
3	4	www.facebook.com	Excellent	270 071 255	Good	1 082 985 733	Excellent	5.87M	64.4K	
4	5	www.yahoo.com	Excellent	99 572 035	Excellent	383 352 336	Excellent	17.2K	1.11K	

6.1.2 Missing values

We note that the dataset presents some problems. In order to check whether our dataset contains missing values, we can use the function isna(), which returns if an cell of the dataset if NaN or not. Then we can count how many missing values there are for each column.

In here checking missing value we used the function called isnull().values.any() . its return the result true value .its mean that in that data set have missing value. After that ,we used isnull().sum() for count all the missing value in dataset .the result is showing lot of missing data.

```
[ ] df.isnull().values.any()

    True

    series=df.isnull().sum()
    series

    Country_Rank                                0
    Website                                     0
    Trustworthiness                             0
    Avg_Daily_Visitors                        144
    Child_Safety                                0
    Avg_Daily_Pageviews                       144
    Privacy                                     0
    Facebook_likes                              0
    Twitter_mentions                            0
    Google_pluses                               0
    LinkedIn_mentions                           0
    Pinterest_pins                              0
    StumbleUpon_views                           0
    Status                                      0
    Traffic_Rank                                1
    Reach_Day                                  62
    Month_Average_Daily_Reach                  45
    Daily_Pageviews                            62
    Month_Average_Daily_Pageviews              46
    Daily_Pageviews_per_user                   62
    Reach_Day_percentage                       77
    Month_Average_Daily_Reach_percentage       82
    Daily_Pageviews_percentage                 77
    Month_Average_Daily_Pageviews_percentage   81
    Daily_Pageviews_per_user_percentage       437
    Location                                   73
    Hosted_by                                 124
    Subnetworks                               270
    Registrant                               1001
    Registrar                                 961
    country                                     9
    dtype: int64
```

6.1.3 Data Standardization

As we saw ,in the dataset have so many missing value. We can't drop all those values for standardization .so we are following here one techniques .we replaced our all missing value min and max value of the columns.

```
[ ] import numpy as np
    mean_values=df[['Avg Daily Visitors','Avg Daily Pageviews']].mean()
    mean_values

    Avg Daily Visitors     4.061140e+07
    Avg Daily Pageviews    2.200354e+08
    dtype: float64

[ ] df[['Avg Daily Visitors','Avg Daily Pageviews']] = df[['Avg Daily Visitors','Avg Daily Pageviews']].fillna(value=df[['Avg Daily Visitors','Avg Daily Pa
    df.head()
```

	Country Rank	Website	Trustworthiness	Avg Daily Visitors	Child Safety	Avg Daily Pageviews	Privacy	Facebook Likes	Twitter Mentions	Google Pluses	Linkedin Mentions	Pinterest Pins	Stumbleupon Views
0	1	Google.com.af	Excellent	4.061140e+07	Excellent	2.200354e+08	Excellent	9	1	37	7	0	7
1	2	Google.com	Excellent	5.150074e+08	Excellent	4.192160e+09	Excellent	94200	11200	11700000	1670	10800	246000
2	3	Youtube.com	Excellent	5.064573e+08	Excellent	2.679159e+09	Excellent	13500	16500	19300000	60000	47	329000
3	4	Facebook.com	Excellent	2.700713e+08	Good	1.082986e+09	Excellent	5870000	64400	127000	6230	4150	23100
4	5	Yahoo.com	Excellent	9.957204e+07	Excellent	3.833523e+08	Excellent	17200	1110	798000	7500	433	68900

This is how we handle all our missing value of dataset. After fill all null value with integer we have almost clean data. Result is like

```
df.isnull().sum()

Country Rank                              0
Website                                   0
Trustworthiness                           0
Avg Daily Visitors                        0
Child Safety                              0
Avg Daily Pageviews                       0
Privacy                                   0
Facebook Likes                            0
Twitter Mentions                          0
Google Pluses                             0
Linkedin Mentions                         0
Pinterest Pins                            0
Stumbleupon Views                         0
Status                                    0
Traffic Rank                              0
Reach Day                                 0
Month Average Daily Reach                 0
Daily Pageviews                           0
Month Average Daily Pageviews             0
Daily Pageviews Per User                  0
Reach Day Percentage                      0
Month Average Daily Reach Percentage      0
Daily Pageviews Percentage                0
Month Average Daily Pageviews Percentage  0
Daily Pageviews Per User Percentage       0
Location                                 73
Hosted By                               124
Subnetworks                             270
Registrant                             1001
Registrar                               961
Country                                   9
dtype: int64
```

Then we again can see that “Location” , “Hosted By” , “ Subnetworks ” , “ Registrant” , “Country ” have still missing value .so now we replace those null value with sting. After we fill this our result is look like this.

```
    df = fill_null_value_with_string(df, columns = c
```

```
[ ] df.isnull().sum()
```

```
Country Rank                                0
Website                                     0
Trustworthiness                             0
Avg Daily Visitors                          0
Child Safety                                0
Avg Daily Pageviews                         0
Privacy                                     0
Facebook Likes                              0
Twitter Mentions                            0
Google Pluses                               0
Linkedin Mentions                           0
Pinterest Pins                              0
Stumbleupon Views                           0
Status                                      0
Traffic Rank                                0
Reach Day                                   0
Month Average Daily Reach                   0
Daily Pageviews                             0
Month Average Daily Pageviews               0
Daily Pageviews Per User                    0
Reach Day Percentage                        0
Month Average Daily Reach Percentage        0
Daily Pageviews Percentage                  0
Month Average Daily Pageviews Percentage    0
Daily Pageviews Per User Percentage         0
Location                                    0
Hosted By                                   0
Subnetworks                                 0
Registrant                                  0
Registrar                                   0
Country                                     0
dtype: int64
```

Now we successfully clean the dataset.

6.1.4 Data Binning

After successfully clean the dataset we have create a new dataset.

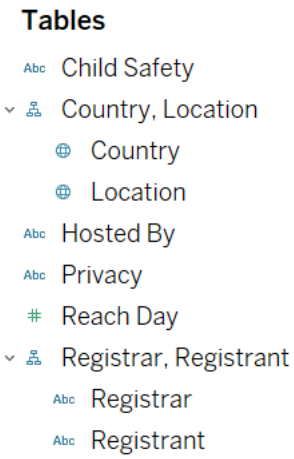
```
Registrar                                   0
Country                                     0
dtype: int64
```

```
[ ] df.to_excel("/content/drive/MyDrive/Datasets/sheet1.xls")
```

For run the df.to_excel it will create a new csv file. with that new clean csv file we can do visualization very nicely.

6.2 Data Preprocessing Tableau

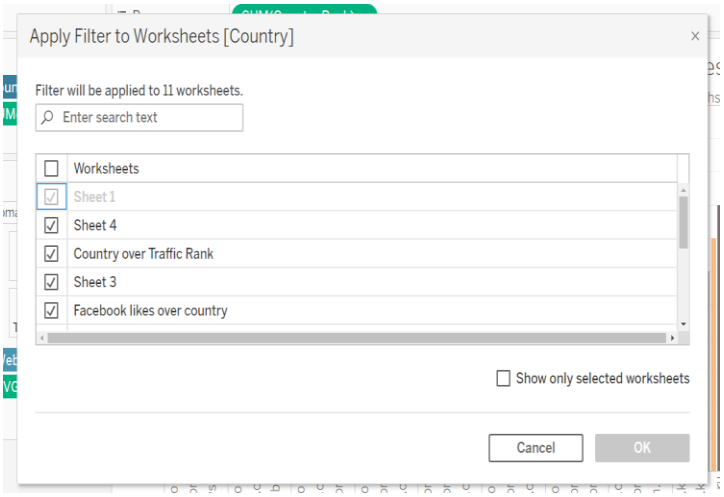
Data can be generated, captured, and stored in a dizzying variety of structures, but when it comes to analysis, not all data formats are created equal. Data preparation is the process of cleaning dirty data, restructuring ill-formed data, and combining multiple sets of data for analysis. It involves transforming the data structure, like rows and columns, and cleaning up things like data types and values. The speed and efficiency of your data prep process directly impacts the time it takes to discover insights. Understanding the scope of data you're analyzing and seeing the changes you make to the data can accelerate the entire process.

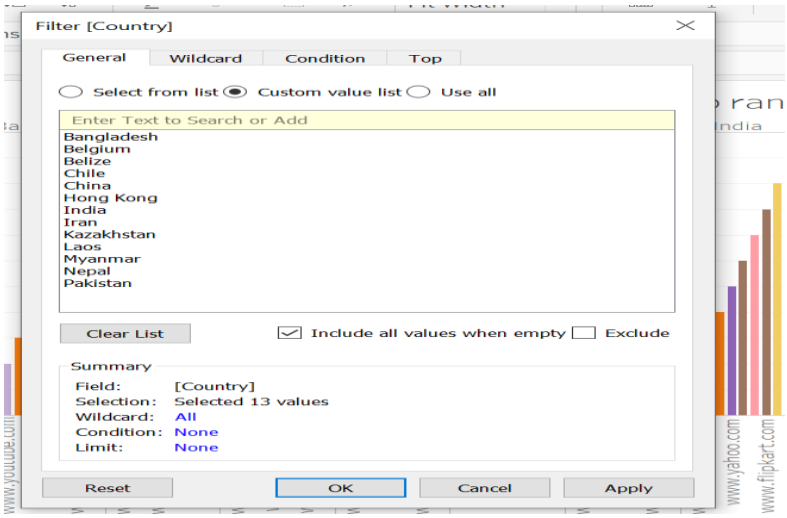


Country,location and registrar, registrant .

7 visualization project

As I said before, I peck some of Asian country and I tried to show some visualization respect to the dataset. In that visualization project I showed different countries website statuse.

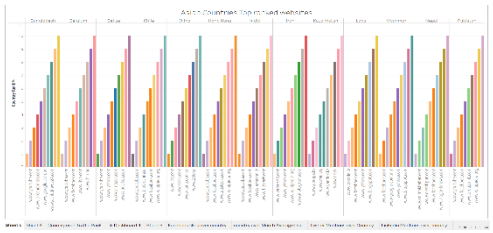


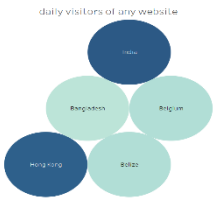


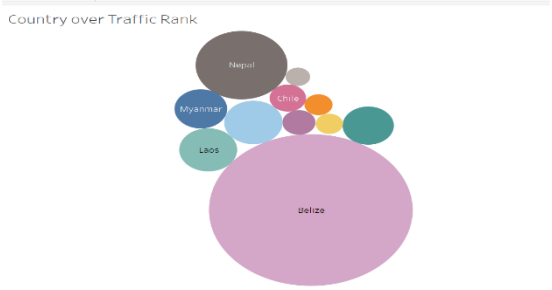
So from that picture we can see that i selected some of country from all country value .then I make it one filter .then I used this filter to all of my worksheet. all of my visualization is respect to the country.

7.1 My Work Sheet

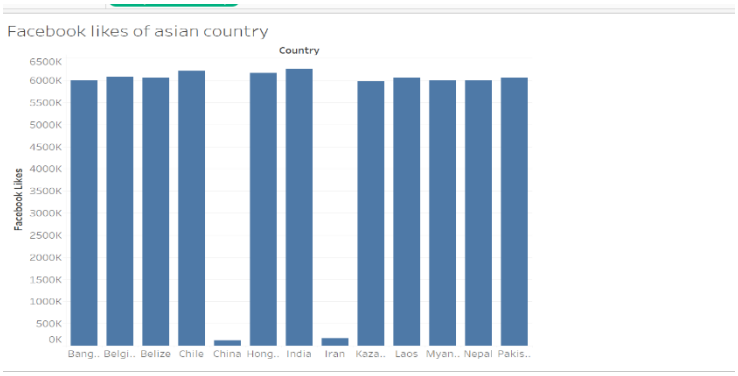
So I created total 14 work sheet respect to the dataset. In that 14 work sheet I apply my some visualization and build the worksheet. Worksheet in the Tableau screen is the area where you create the views for data analysis. By default, Tableau provides three blank worksheets when you have established a connection to data source. You can go on adding multiple worksheets to look at different data views in the same screen, one after another.

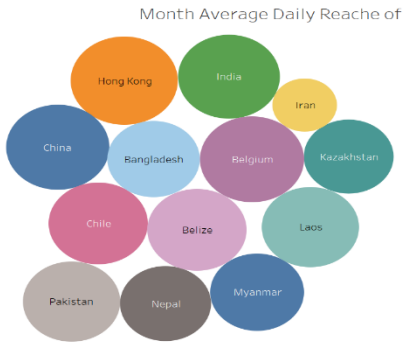


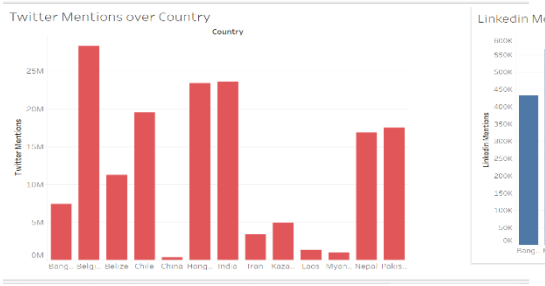

daily visitors of any website
India
Bangladesh
Belgium
Hong Kong
Belize

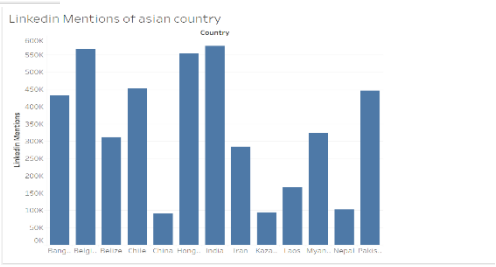

Country over Traffic Rank
Nepal
Chile
Myanmar
Laos
Belize

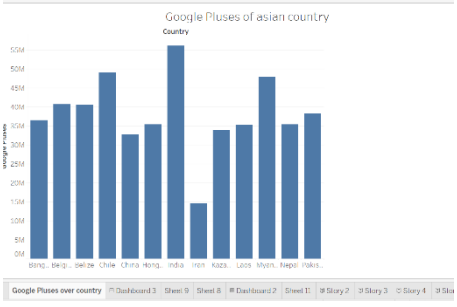

Worldwide avalabilty website map
© 2021 Mapbox © OpenStreetMap
Rank
Dashboard 4
Sheet 3
Facebook likes over country
country over Month Average Dai...
Twitter Mentions over Country

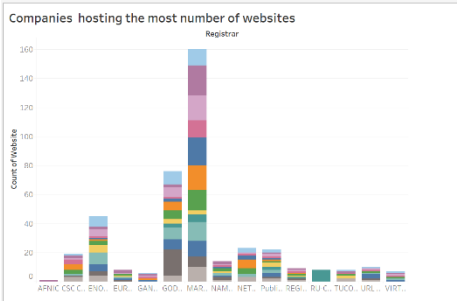

Facebook likes of asian country
Country
Facebook Likes
6500K
6000K
5500K
5000K
4500K
4000K
3500K
3000K
2500K
2000K
1500K
1000K
500K
0K
Bang..
Belgi..
Belize
Chile
China
Hong..
India
Iran
Kaza..
Laos
Myan..
Nepal
Pakis..


Month Average Daily Reache of asian country
Hong Kong
India
Iran
China
Bangladesh
Belgium
Kazakhstan
Chile
Belize
Laos
Pakistan
Nepal
Myanmar


Twitter Mentions over Country
Country
Twitter Mentions
25M
20M
15M
10M
5M
0M
Bang..
Belgi..
Belize
Chile
China
Hong..
India
Iran
Kaza..
Laos
Myan..
Nepal
Pakis..


Linkedin Mentions of asian country
Country
LinkedIn Mentions

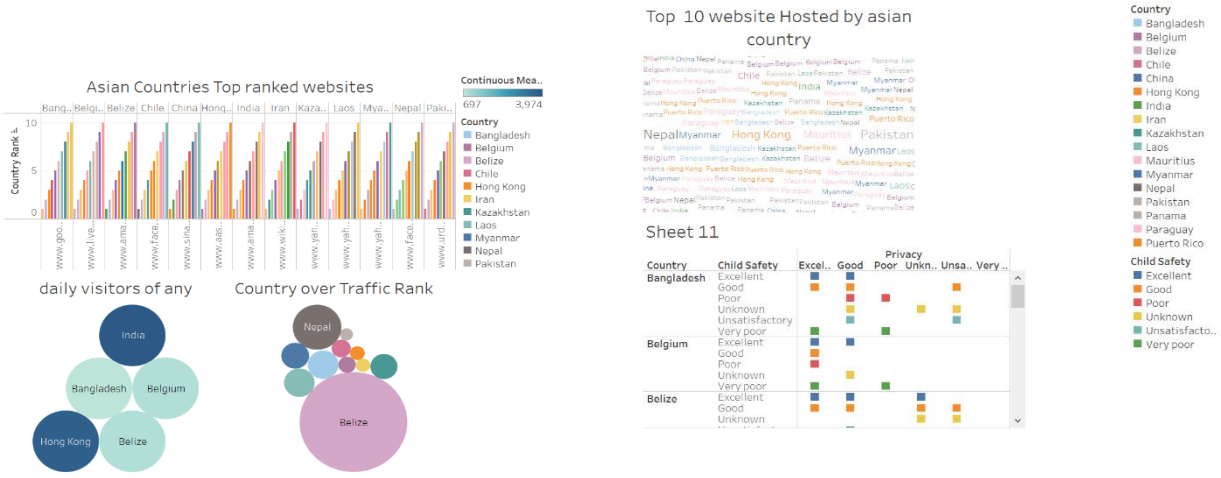

Google Pluses of asian country
Country
Google Pluses over country
Dashboard 3
Sheet 9
Sheet 8
Dashboard 2
Sheet 11
Story 2
Story 3
Story 4


Companies hosting the most number of websites
Registrar
Count of Website
160
140
120
100
80
60
40
20
0


Privacy
Country
Child Safety
Excel..
Good
Poor
Unkn..
Unsa..
Very ..
Bangladesh
Belgium
Belize
Chile
China
Hong Kong

7.2 Dashboard Data

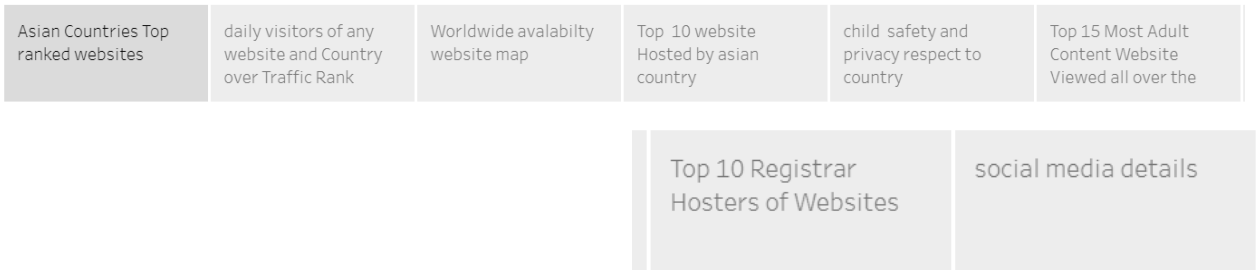
A dashboard is a collection of several views, letting you compare a variety of data simultaneously ,In Dashboard Data , I used worksheet data to build dashboard.



Building dashboards with Tableau allows even non-technical users to create interactive, real-time visualizations in minutes.

7.3 storytelling with data

A Tableau story is a sequence of visualizations that work together to convey information. Stories are a powerful way to tell a data narrative, provide context, demonstrate how decisions relate to outcome or simply making a compelling case.in my project a storyline based on my all work sheet and Dashboard.



A story is a sheet, so the methods you use to create, name, and manage worksheets and dashboards also apply to stories. At the same time, a story is also a collection of sheets, arranged in a sequence. Each individual sheet in a story is called a story point.

8 Importance and Benefits of Data Visualization

Analyzing reports helps business stakeholders focus on the areas that require attention. The visual mediums help analysts understand the key points needed for their business. Whether it is a sales report or a marketing strategy, a visual representation of data helps companies increase their profits through better analysis and better business decisions.

Nowadays, companies across domains are willing to record user interactions with their products or services to identify trends, patterns, anything to gain insight into their target market and make appropriate business decisions that will propel them toward success and improve their overall performance. In business intelligence, exploring this invaluable user interaction data takes the form of data visualization. Data visualization means presenting raw data through graphical representations that allow viewers—business analysts and executives—to explore the data and uncover deep insights. This visual format enables one to make quick and effective decisions since it is much easier for people to comprehend information through visuals rather than the raw reports.[4]

9 personal opinion about the project

This project provides a global hot website data set, which includes data from the top 50 websites in 191 countries around the world. Each website's data covers 31 dimensions. So we have to make a global hot website data visualization work. It is my personal opinion that such a project may possibly be worthy of consideration in the distant future. Data visualization helps to tell stories by curating data into a form easier to understand, highlighting the trends and outliers. A good visualization tells a story, removing the noise from data and highlighting the useful information .while doing project I have use python and tableau. I learn some of tool work during work on my project. Tableau is the continued growth and development of a product that is driven by small, medium and large enterprise needs, and a thriving community that only want success for the product. There is so much to learn and understand to push Tableau to its limits, and with great features being released with each version, the software is going from strength to strength.

10 conclusion

In this report, we have explored data visualization in detail, its importance and benefits, and quick tips for an effective data visualization. There are practical, real-life applications that are undeniable. And, since visualization is so prolific, it’s also one of the most useful professional skills to develop. Analyzing reports helps business stakeholders focus on the areas that require attention. The visual mediums help analysts understand the key points needed for their business. The best tools also can output an array of

different chart, graph, and map types. Most of the tools below can output both images and interactive graphs. So, This is such a project that may possibly be worthy of consideration in the distant future.

11 Reference

[1] "What Is Data Visualization? Definition & Examples | Tableau." https://www.tableau.com/learn/articles/data-visualization (accessed Dec. 6, 2021).

[2] M. Mallon, "Data Visualization," *Public Services Quarterly*, 2015. https://blog.1moby.com/ds-ep-4-data-visualization-□□□□□□□□□□□□□□□□-□□□□□□□□□□□□□□□□□□-e65515635b6e (accessed Dec. 8, 2021).

[3] "32 Data Visualization Types: Choose the One You Need." https://visme.co/blog/data-visualization-types/#venn-diagram (accessed Dec. 5, 2021).

[4] S. K, "Data Visualization: Importance and Benefits | Bold BI," *BoldBI*, 2019. https://www.boldbi.com/blog/data-visualization-importance-and-benefits (accessed Dec. 6, 2021).