**DATA VISUALIZATION**

**BANA 6760E01**

**PROJECT REPORT ON**

**ANALYSIS OF GDP (1997-2020)**

A picture containing text, clipart

Description automatically generated

-Shashank Gubba

SID: 110288232

Table of Contents

**Chapter 1 - Introduction3**

**Chapter 2 – Exploratory Data Analysis (EDA)4**

2.1 – EDA using R4

2.2 – EDA using PowerBI5

2.3 – EDA using Tableau……………………………………………………………………………………………………………………….7

**Chapter 3 – Conclusion**……………………………………………………………………………………………………………………………9

**Chapter 1 - Introduction**

In the United States Gross Domestic Product is used to prepare the federal budget by the White House and Congress. GDP is basically the sum of all the final expenses or the total economic output by an economy within a specified accounting period which does not involve the output of its underground economy. In the United States, GDP 70% comprises of consumer spending of which retail and the service industries play an important role in the GDP.

For this project I’ve taken the data set from Kaggle. (ctrl+click to view the data set in browser)

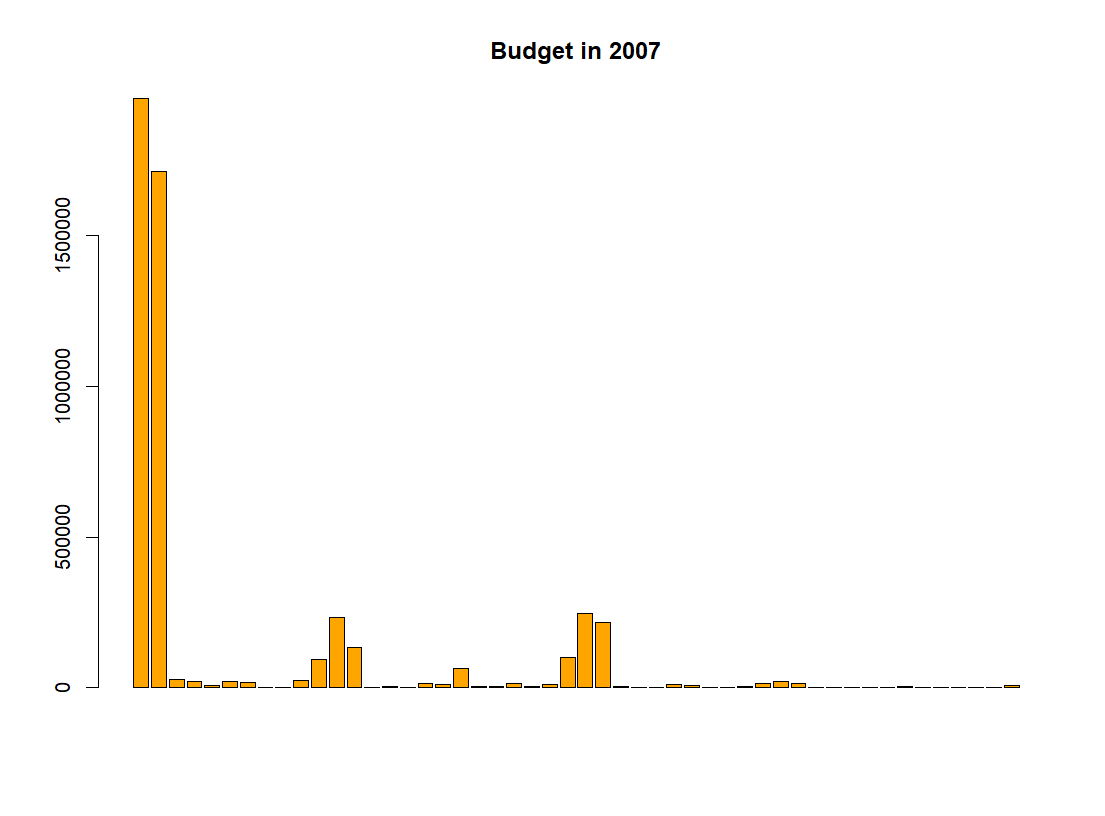
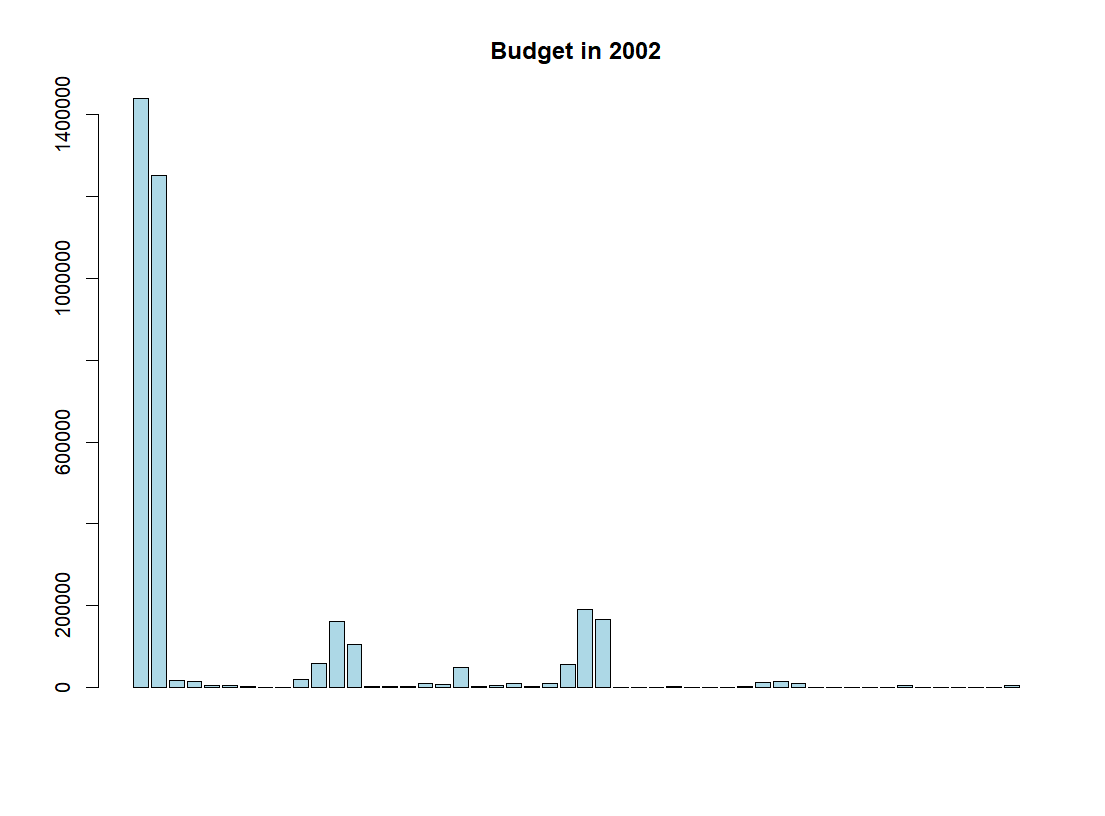
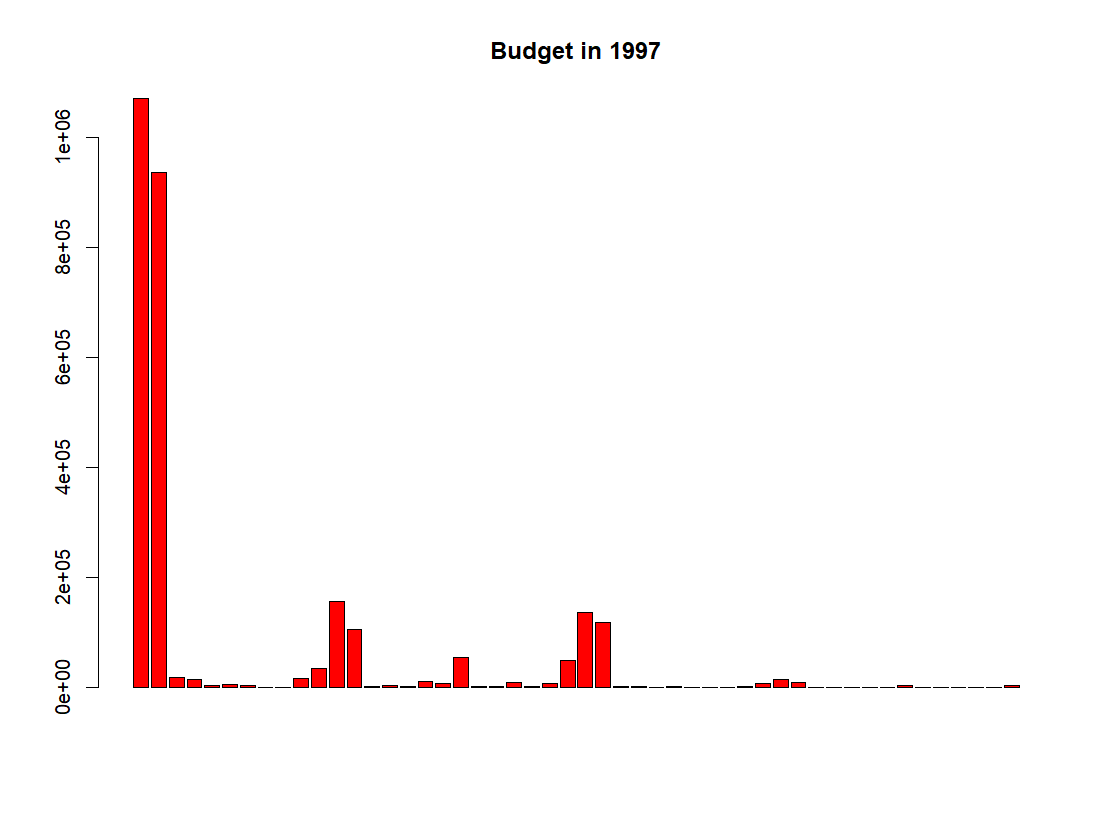
* The data set has 5528 rows and 32 columns of the GDP data representing various industries.
* The data is defined well and consists of columns that need to be removed.
* The plan is to perform EDA on the data for cleaning of the data.

**Chapter 2 – Exploratory Data Analysis (EDA)**

In Exploratory Data Analysis, there are three important stages. Data cleaning, Data modification, Data visualization. In the data cleaning stage, we must eliminate all the null values, NAs, if there are any present in the data set. Coming to the data modification stage, we must replace the missing values in the data set. Data visualization is the stage where we will visualize the modified data using numerous tools like Excel, R, Tableau, PowerBI etc.

2.1 EDA using R:

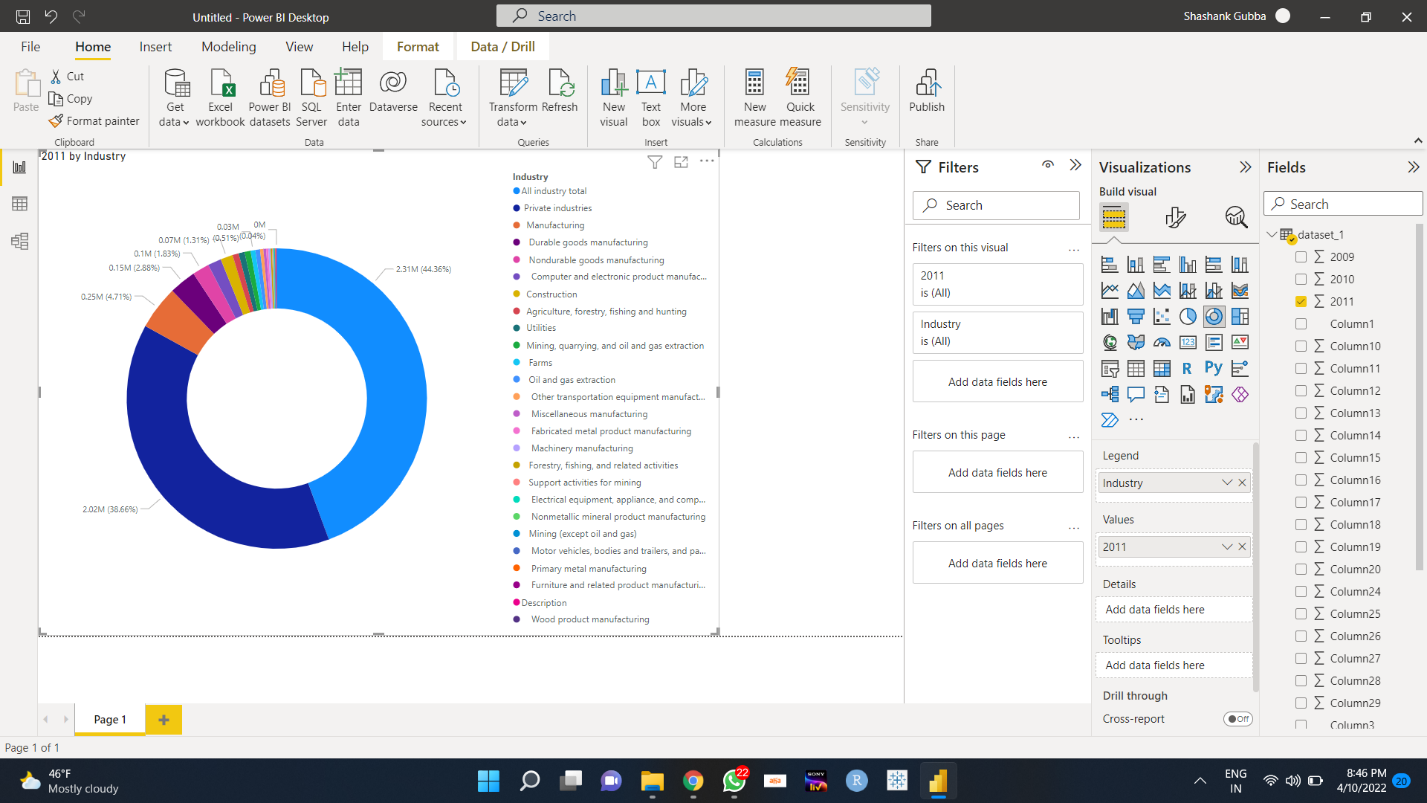
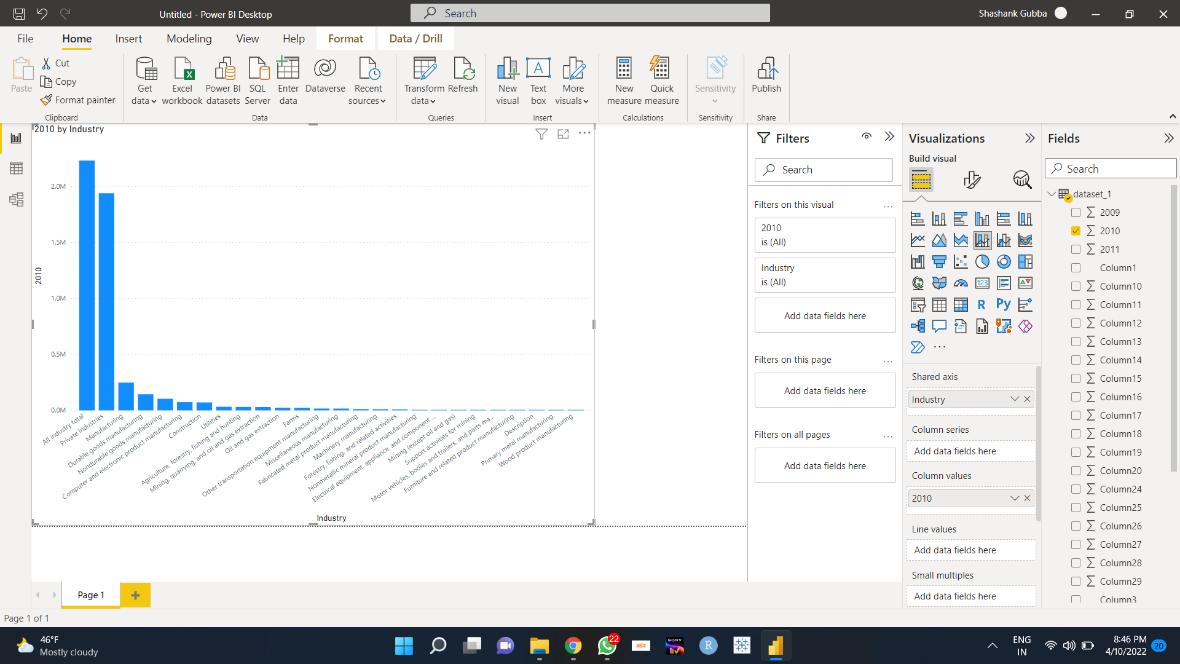
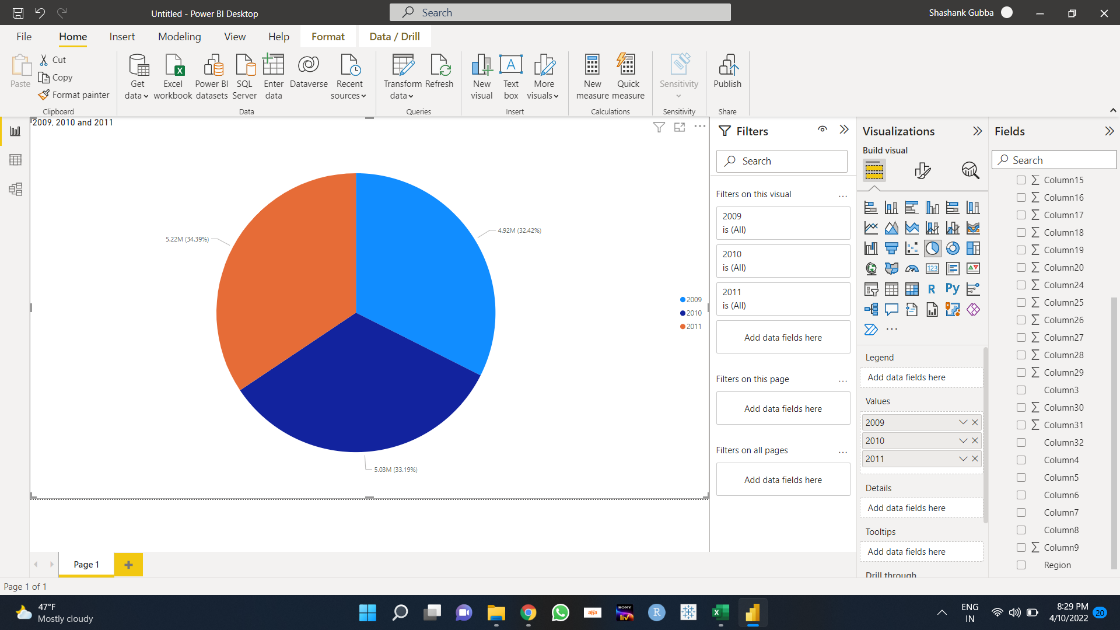
For EDA in R, I’ve used numerous libraries like “ggplot2”, “car”, “tidyverse” and some other libraries for data cleaning and visualization. Below are some visualizations using RStudio.



All the above pictures give us the information of the expenditure of each industry in the years 1997, 2002, 2007. We can see that there is not much change in the expenditure trend of the industries.

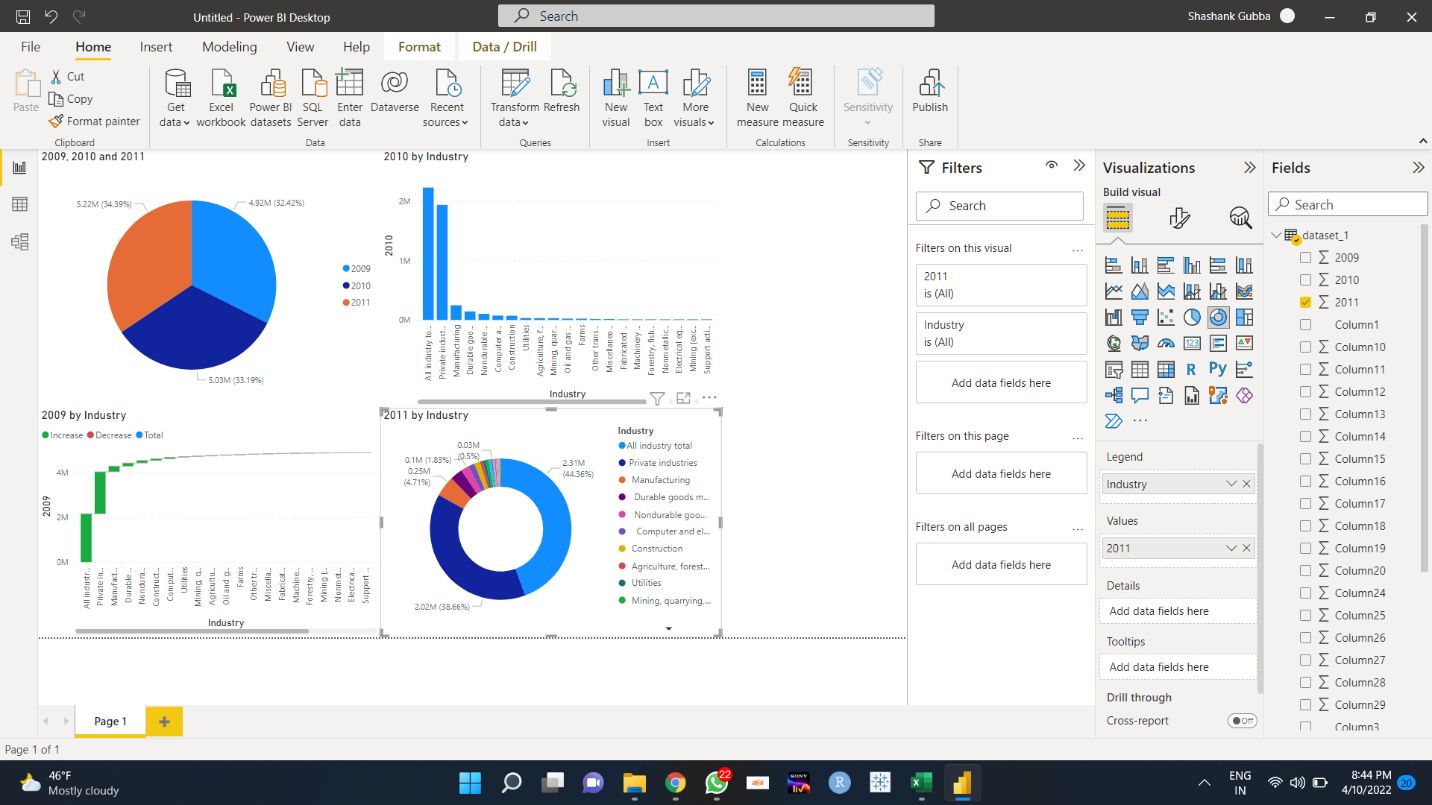
2.2 EDA Using PowerBI:

Using PowerBI, we can clean the data using the Power Query editor in PowerBI and using this tool, we can visualize the geographic data. All the visualized data in PowerBI can be shown in together in a dashboard. Following are some visualizations that are made using PowerBI.



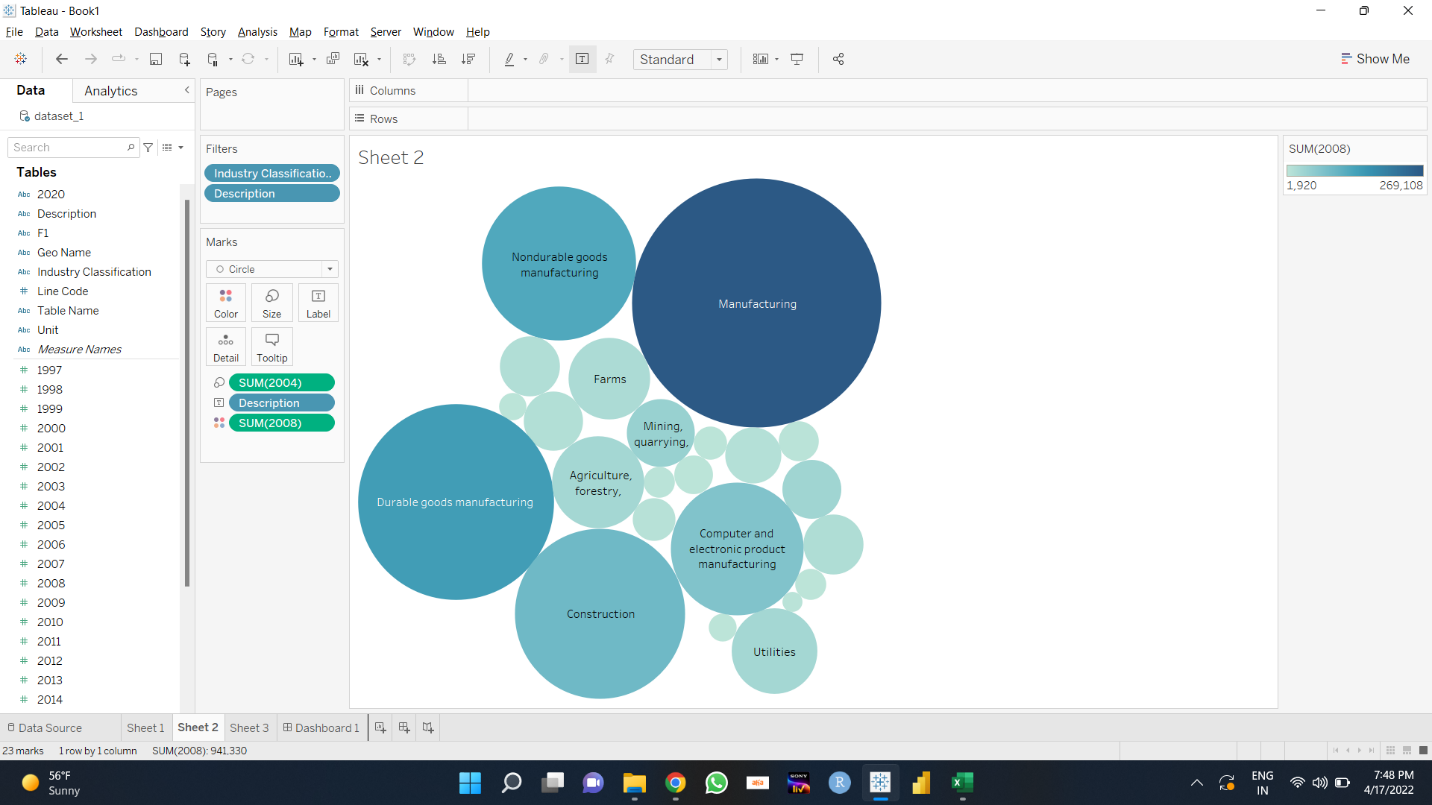
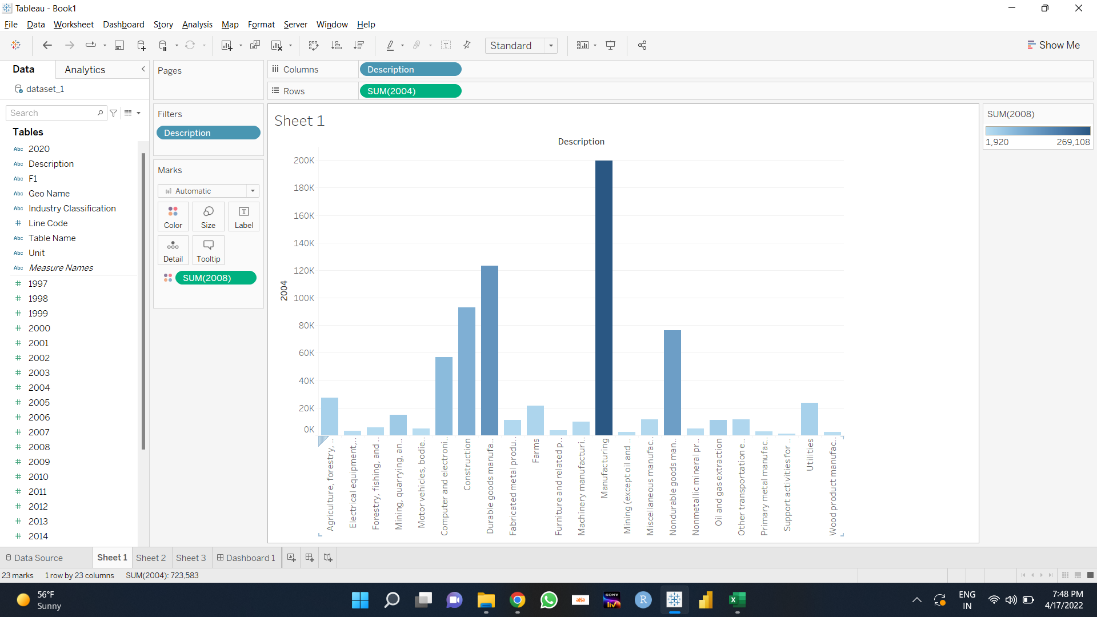
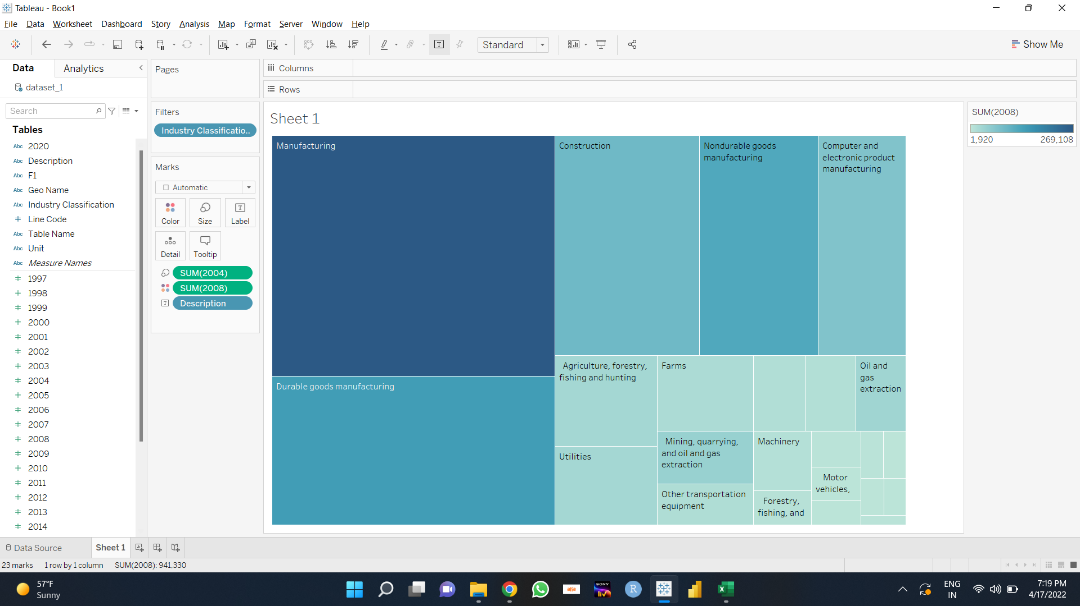
The above pictures are shows us the information of expenditure and its trend of all the industries using pie chart, bar graph and donut chart. In the pie chart, total expenditures of all the industries together in the years 2009, 2010 and 2011 are visualized. Bar graph gives us the info expenditure in the year 2010 and the donut chart gives us the info of expenditures in the year 2011.

PowerBI dashboard is shown below,



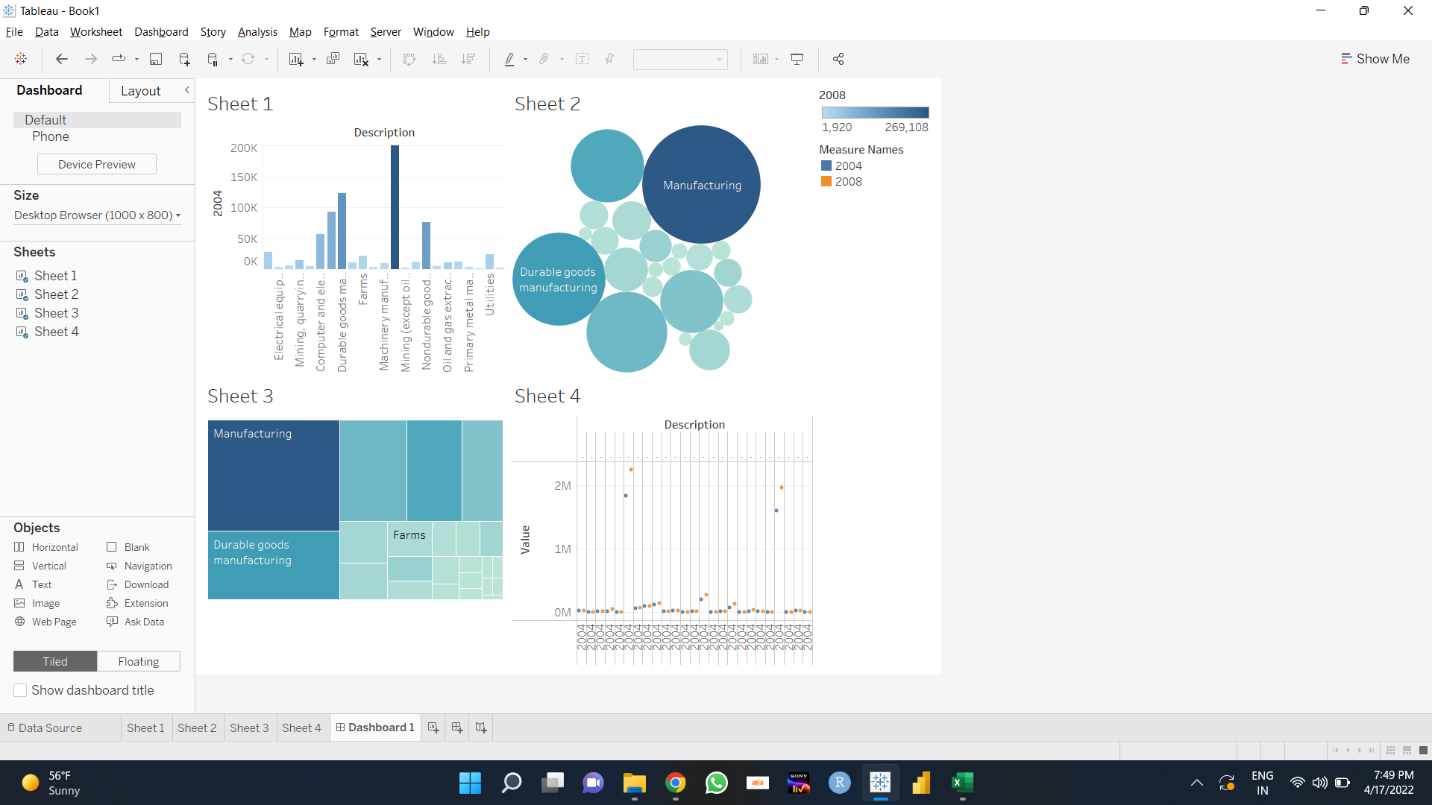
2.3 EDA using Tableau:

Similar to PowerBI, using Tableau we can perform data cleaning and visualizations too. Tableau is much better when it comes to handling the data than PowerBI and also the Tableau files can be accessed on the cloud, so it has more flexibility of doing the tasks. Below are some visualizations that are made using Tableau,



Using Tableau, data is visualized in the form of Tree map, bar graph and a bubble graph. Tree map has blocks which represent the size of expenditure, upon seeing the graph we can say that manufacturing industry has the highest expenditure. Bubble graph has the same concept, more the intensity of the units bigger is the size of the bubble.

Below is the Tableau dashboard.



**Chapter 3 – Conclusion**

After all the EDA processes using different tools and after analyzing all the visualizations, we can conclude the following,

* Manufacturing industry has the highest expenditure of all the industries.
* Retail and hospitality industries play a major role in growth of the GDP.