***“A Critical Evaluation of the Big Data Approach to Car Fuel's Data Analytics”***

The Big Data analysis provides companies with the results necessary to investigate the problems presented or improve existing products and services, taking advantage of opportunities to monitor market trends. The field of Big Data is very big and covers many methods and tools that help its development and the exploitation of data with great volume, variety and growth, in all the succession of steps that entail which are the collection, elimination and Processing of all the variety of data from various sources.

The reason for all this is to generate and present graphs that allow to have a vision of the behavior of the organization in a clearer way, and to be able to assume the best route that has as objectives to improve the processes, reduce the costs and find growth opportunities.

The use of Big Data allows us to work with several types of data from different sources based on the technologies that compose it such as predictive analysis, NoSQL Database, Hadoop Ecosystem, Blockchain, among others; and it helps the research of more organizations such as manufacturing, marketing, medicine, business, whose amount of data generated grows exponentially, refining the extracted values and obtaining a true value that can be used in decision making.

For the Automobile Research Company, publicly available data that had been prepared in a document with 729 entries and 19 attributes was used, providing information on each car manufacturer and model, in which Tableau divided them for use as follows:

|  |  |
| --- | --- |
| Dimensions | Measures |
| Air Aspiration Method  Carline  Carline Class Desc  Division  Drive Desc  Mfr Name  Model Year  Release Date  Transmission  Transmission Description | # Cylinders  # Gears  City CO2  City FE  Combined CO2  Combined FE  Engine Displacement  Highway CO2  Highway FE |

Table 1. Fields of the data set

Tableau is a tool that allows us to visualize interactive data focused on business intelligence from relational databases, OLAP cubes, databases in the cloud and spreadsheets; The visualization of this data is done through different graphics suggested by the tool that depends on the selected fields. With the help of Tableau, we could answer the questions asked in assignment 1 on fuel economy between different car manufacturers and show the result for a better understanding through a graph.

Other advantages that the use of Tableau presents is to add different graphics within the same project and share them as a board or a story; Apply filters to different fields as well as the application of mathematical functions; You do not need to know any programming language to be used; It updates the data in real time and can be viewed before being used. The questions asked in the investigation of the company were as follows:

1. Find out which car manufacturer produces the highest quantity of models

e.g BMW 3 Series and BMW 5 Series are different models**.**

1. Find out the highest average fuel economy for city and highway driving from the given data set.
2. Find out high and low average fuel economy from all transmission types.
3. Find out which car manufacturers have 4WD (4 wheel drive) and 2WD (2 wheel drive) models, whose engine power is more than 3.5.

To solve each question, we used a set of fields imported from the spreadsheet about the different car models on which we want to do the research. This import was made from the same Tableau platform from the option Connect> To a file> Microsoft Excel, where it opened a window to choose the file that contains the data set.

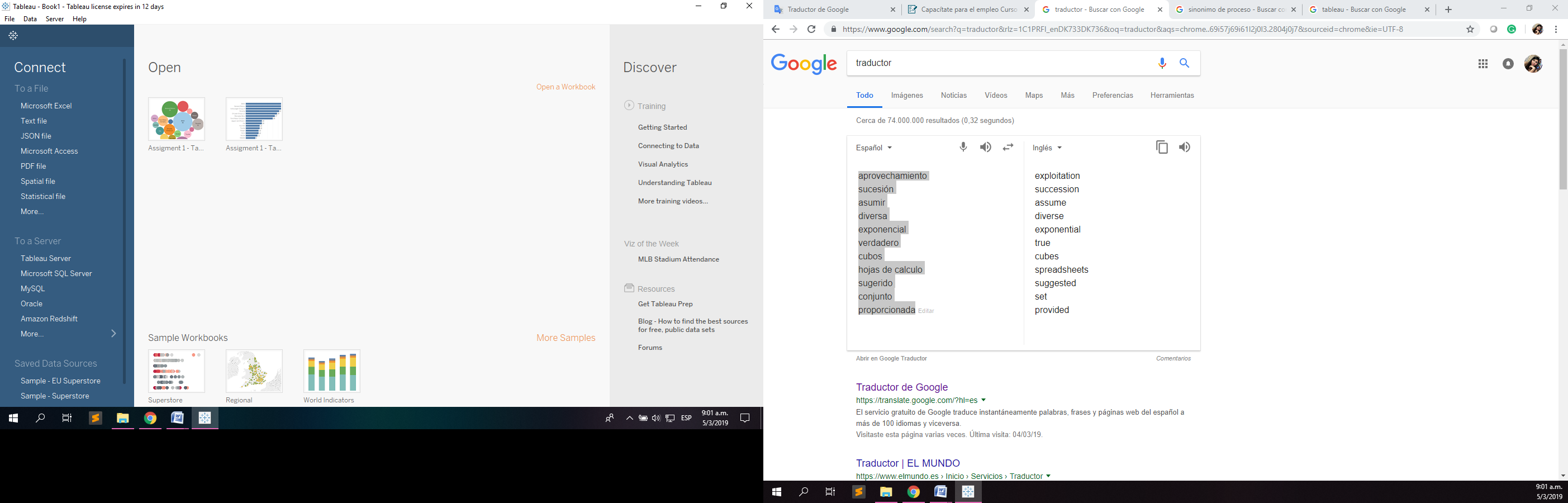


Image 1. Connect Panel

After the import, the table containing the data (FEguide) is dragged to the panel with the message "Drag the sheet here"; and then it shows us all the data contained in the file. At this point, we can also clean the data before using it or create fields calculated from it. Then, at the bottom is the option to create a new worksheet where we create the necessary chart for each of the questions.

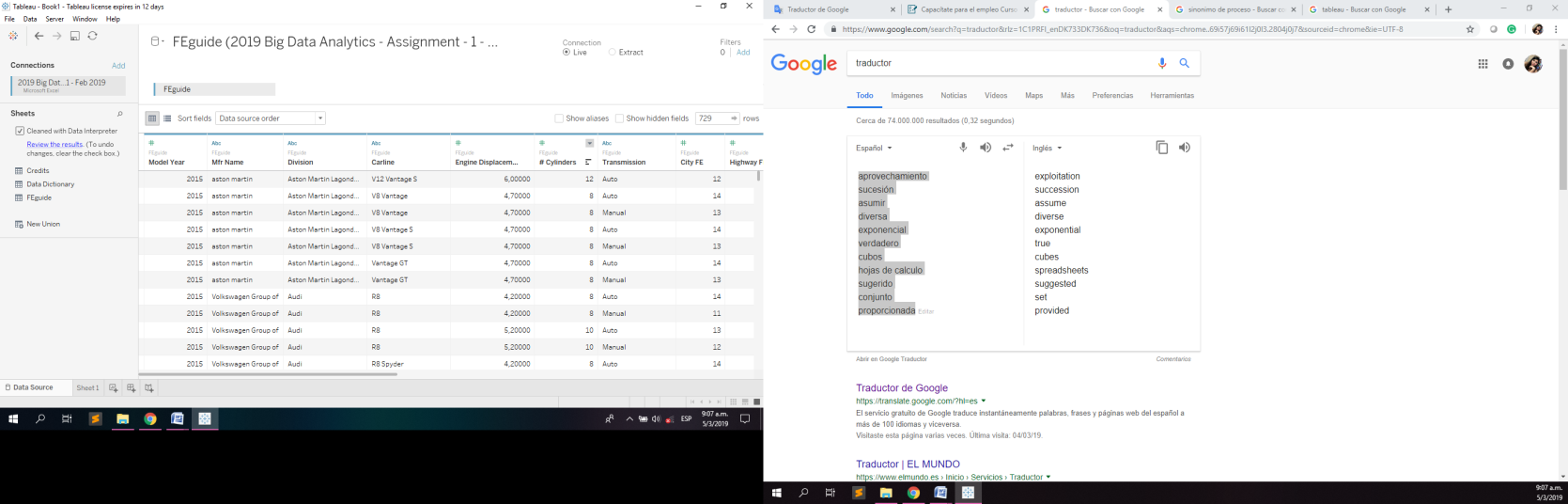


Image 2. Data Source View

On each sheet, we can configure the name of the graph, establish the fields belonging to the columns or rows, these fields are located on the left side of the screen divided into two groups, dimensions (they are the fields that contain the qualitative value as name and dates that can be used to classify the data to affect the level of detail of the view) and measure (the fields that contain numerical values which can be measured and apply calculations), we also have a toolbar, a filter section , and the most important part is the view section where our data is displayed to be displayed in different graphics.

An example of the representation of the graphs will be given with the output obtained from assignment 1 shown as follows:

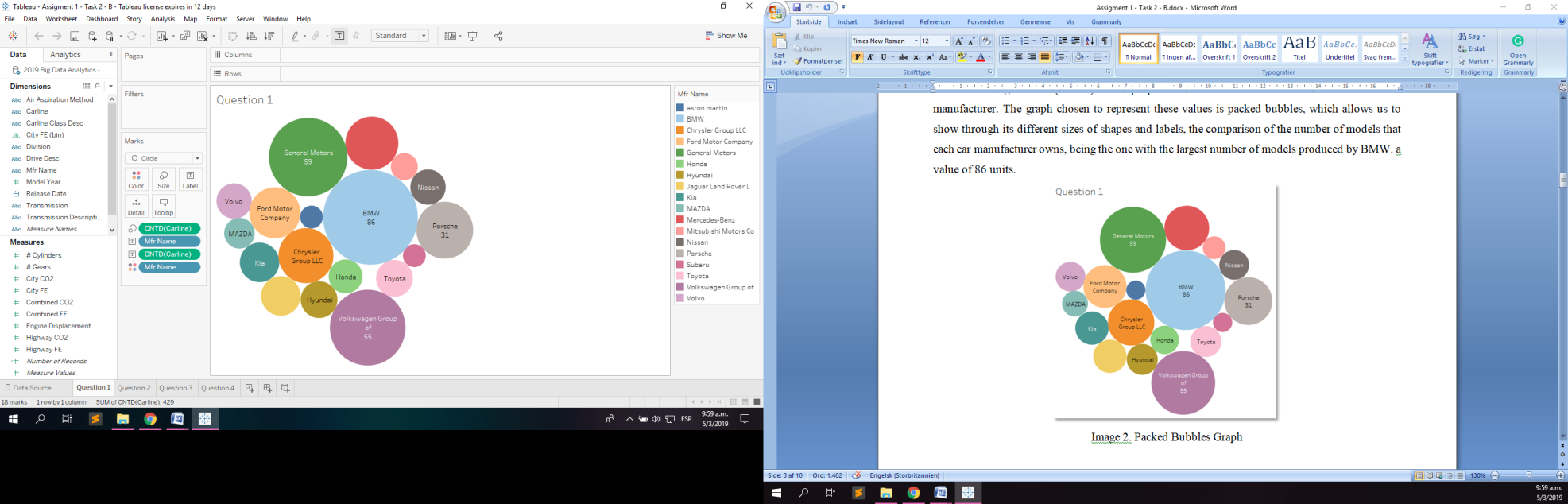


Image 3. Question 1 View

The fields used are Mfr Name and account other than Carline. The color shows details about the name field of Mfr. The size shows the count of the different models of the Carline field. The marks are labeled with the Mfr name and the distinctive Carline count. The view is filtered by Mfr name, which holds 18 of the 18 members.

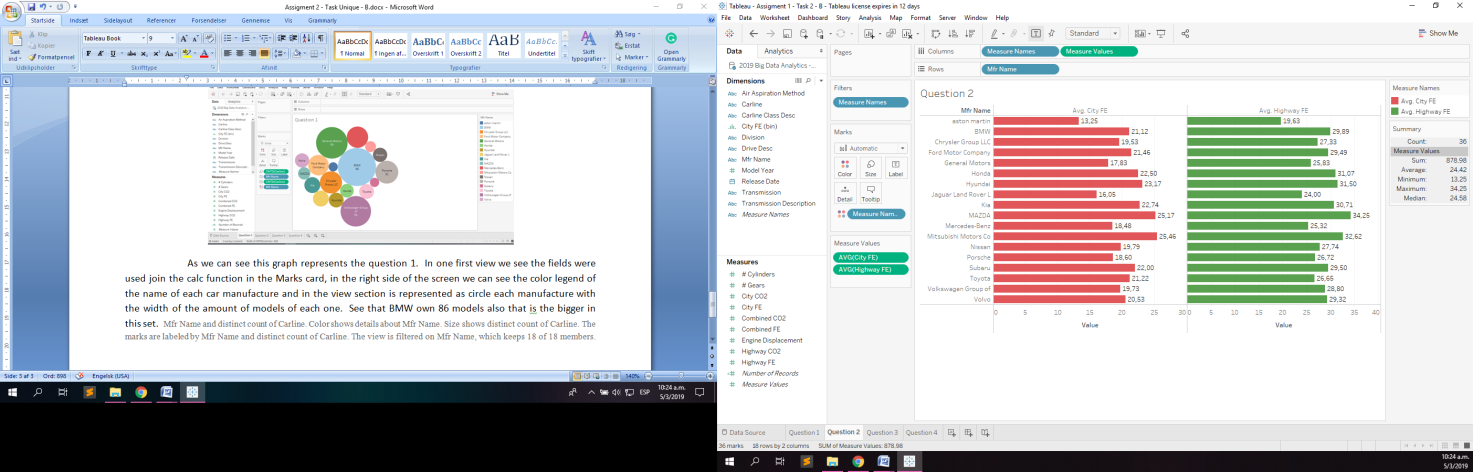


Image 4. Question 2 View

The fields used are the average of City FE and the average of Highway FE for each Mfr name. The color shows details about Avg. City FE and Avg. Highway FE.

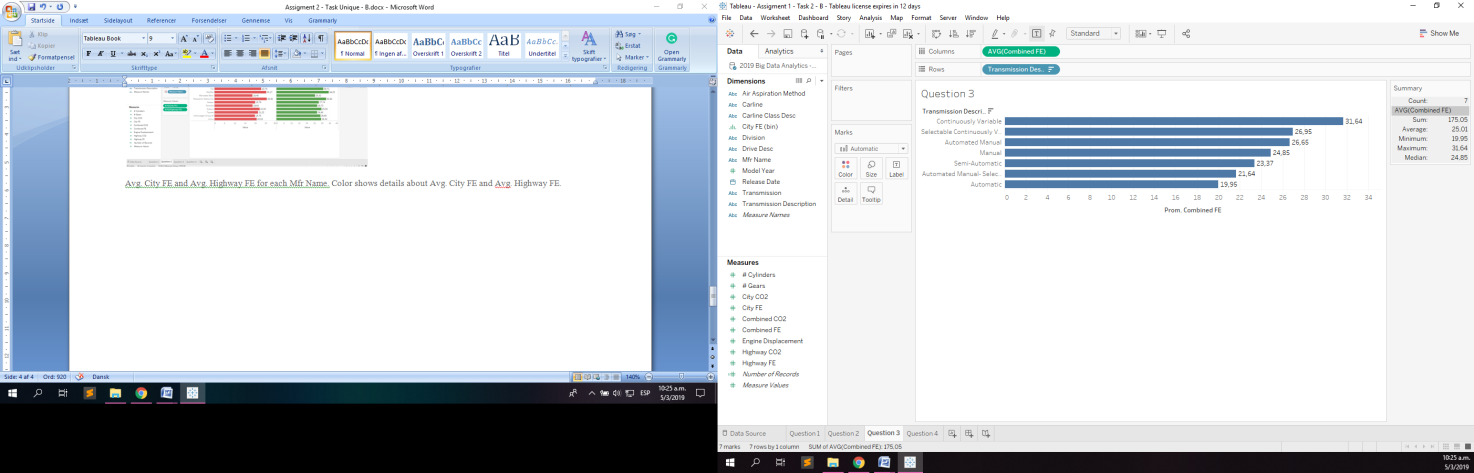


Image 5. Question 3 View

The fields are Combined FE and Transmission desc. The combined FE average for each transmission description is plotted in the image above.

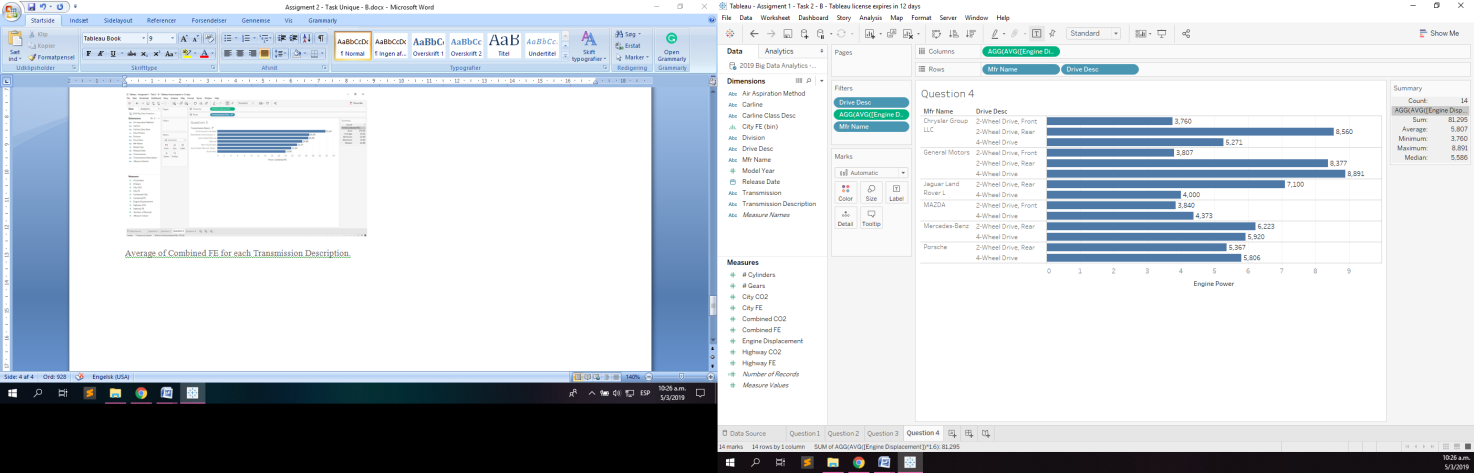


Image 6. Question 4 View

The fields used are Engine displacement, Mfr Name, Drive desc. The formula AVG ([Displacement of the motor]) \* 1.6 was applied for each Desc. Of unit broken down by name of manufacturer. The view is filtered in Drive Desc, AVG ([Motor Displacement]) \* 1.6 and Manufacturer Name. The unit's Descent filter maintains 2-wheel drive, front-wheel drive, 2-wheel drive, rear-wheel drive and 4-wheel drive. The AVG filter ([Motor Displacement]) \* 1.6 includes values greater than or equal to 3,500. The Mfr name filter is maintained with 8 registers of 18 members.

As we can see, the Tableau tool helps us to solve the questions asked by the car company's research through the analysis of data and the Big Data approach using technologies such as predictive analysis and NoSQL databases to obtain these results. The implementation of all these solutions related to Big Data helps us in decreasing response times where processes are constantly changing, as well as how data can be obtained in real time from the environment.

**REFERENCES**

Tableau.com (2019). *Tableau Desktop and Web Authoring Help*. [Online] Available at: *https://onlinehelp.tableau.com/current/pro/desktop/en-us/default.htm* [Accessed March 2019]

Tableau.com (2019). *Free Training Videos*. [Online] Available at: *https://www.tableau.com/learn/training* [Accessed March 2019]

Fueleconomy.gov (2019). *Download Fuel Economy Data.* [Online] Available at: *https://www.fueleconomy.gov/feg/download.shtml* [Accessed March 2019]

Mattturck.com (2018). *Great Power, Great Responsibility: The 2018 Big Data & AI Landscape.* [Online] Available at: [*http://mattturck.com/bigdata2018/*](http://mattturck.com/bigdata2018/) *[Accessed* March *2019]*

Whatsthebigdata.com (2017). *Big Data Ecosystem and Benefits.* [Online] Available at: [*https://whatsthebigdata.com/2017/03/28/big-data-ecosystem-and-benefits/*](https://whatsthebigdata.com/2017/03/28/big-data-ecosystem-and-benefits/) *[Accessed* March *2019]*