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

Big Data is a terminology used to refer to the extraction, storage, administration, and visualization of a large amount of structured and unstructured information that exists in a company from various sources. The technologies included in the Big Data are diverse and are classified according to their use or specialization, some examples of these technologies are Predictive Analytics, NoSQL Databases, Hadoop Ecosystem, Blockchain, among others.

The scenario presented for this evaluation, is an automobile research company that is launching a new research project, based on a given data set on fuel economy for the 2015 models. The company wants to analyze the economy of fuel between different car manufacturers and to achieve this goal; we use publicly available data sets that have been prepared previously. The data set contains 729 entries and 19 attributes, which provide the following information about each automobile:

|  |  |
| --- | --- |
| 1. Model Year 2. Mfr Name 3. Division 4. Carline 5. Engine Displacement 6. # Cylinders 7. Transmission 8. City FE 9. Highway FE 10. Combined FE | 1. Air Aspiration Method 2. Transmission Description 3. # Gears 4. Drive Desc 5. Carline Class Desc 6. Release Date 7. City CO2 8. Highway CO2 9. Combined CO2 |

Table 1. Fields of the Data Set

The tool used to complete this work was Tableau Desktop Professional Edition 2019.1.0 64-bit, which is a tool that allows us to analyze and visualize a large amount of data and represent them through graphs for a better understanding of the customers on the Result obtained to the automobile research company.

This research was composed by several questions that were answered with the help of the Tableau Desktop tool through the graph and the numerical results. The data provided was imported into the software as a Microsoft Excel file, and from this file was used the FEguide sheet that contains the data necessary to process them about fuel savings in automobiles. In the first instance, the Tableau software allows us to see all the records of the models and the name of the car manufacturers, so there is the possibility of creating new fields, records or making a new calculation from them.

When creating a new sheet, Tableau shows us the data organized into two different groups, which are the Dimensions (qualitative data) and the Measures (quantitative data). In addition, show us a section of shelves and a graphics section where we drag and drop each of the fields necessary to answer each question. The questions that were used to do the research of the automobile company are:

* Find out which car manufacturer produces the highest quantity of models e.g BMW 3 Series and BMW 5 Series are different models.
* Find out the highest average fuel economy for city and highway driving from the given data set-
* Find out high and low average fuel economy from all transmission types.
* Find out which car manufacturers have 4WD (4 wheel drive) and 2WD (2 wheel drive) models, whose engine power is more than 3.5.

Depending on the chosen fields, Tableau recommends us to establish a graph according to the data to be implemented, among these graphs are text tables, heat maps, highlighting tables, maps, pie charts, horizontal bars, stacked bars, side by side, treemaps, circle views, side by side circle, lines, area graphics, dual combination, scatterplots, histogram, box and whisker graphics, Gantt, bullet graphics and packed bubbles. A representation of these graphs is detailed below as an example of the first question:

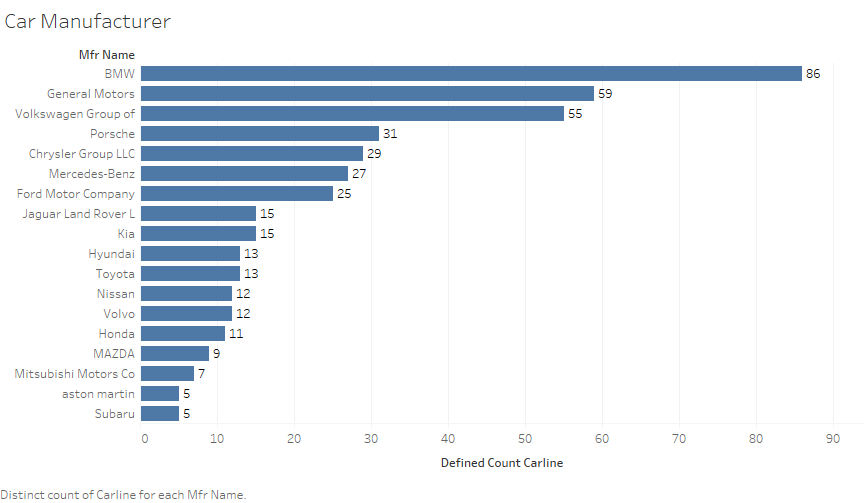


Image 1. Horizontal Bars – Graph Question 1

The representation of this data is very important to obtain a clear visualization and to understand the behaviour of the data, helping in this way the automobile research company on the fuel economy of the different models and manufacturers of automobiles. In the same way, Tableau Desktop shows us a table with the data generated for the combination of fields linked to the statistical functions. A view of this table is obtained from the fourth question:

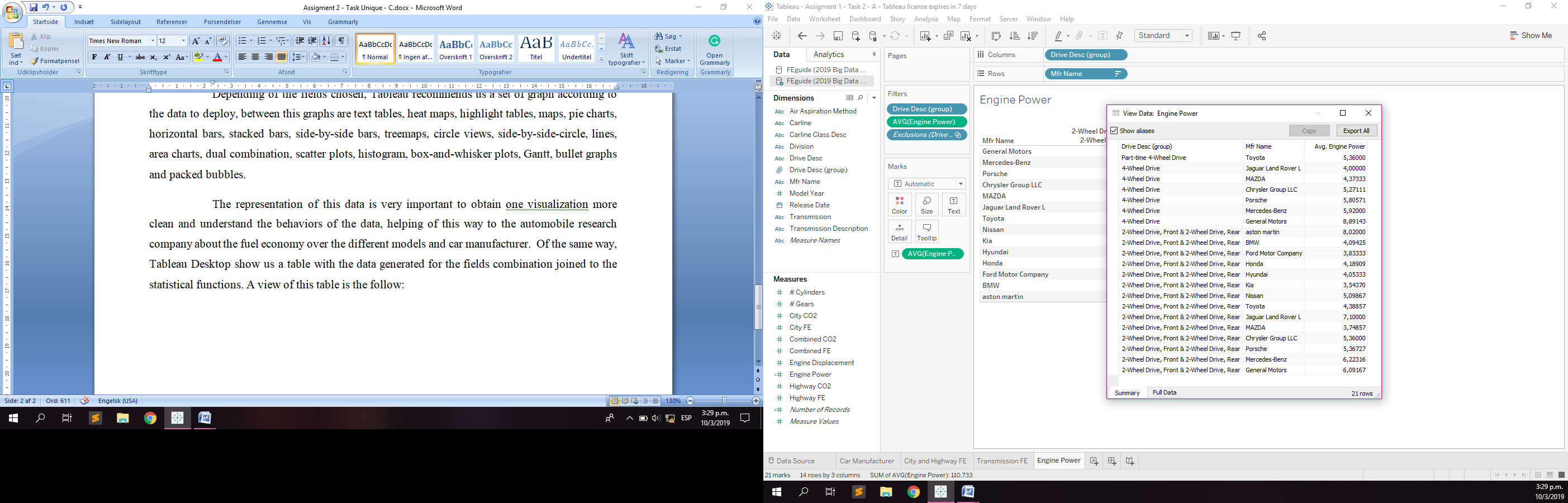


Image 2. Table of Data – Question 4

Other information that we can obtain with this tool is a summary that shows several results of the calculation, obtained from the chosen data and places them in the corresponding shelf (columns or rows), an example of this is the display in the first question shown below:

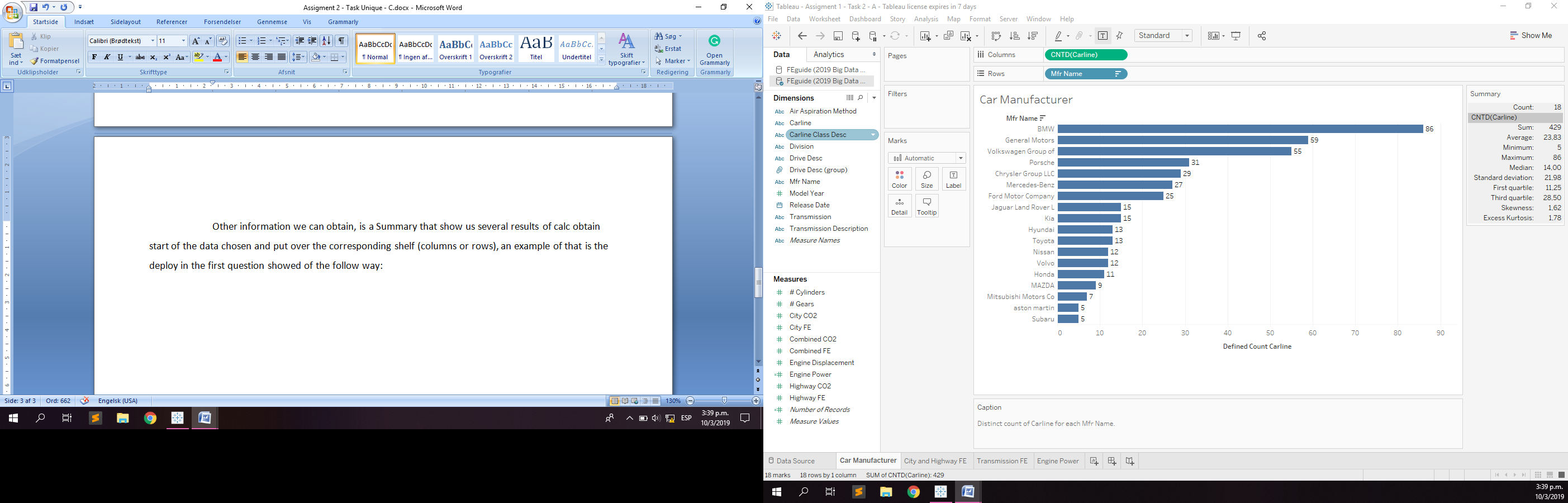


Image 3. Table of Summary Data – Question 1

Finally, Tableau Software can show us more information about the results of the chart to help us make more assertive decisions. This additional information is on the title card in text format. An example of this is the following: "Combined EF Average broken down by Transmission Description, the colour shows a Combined EF average, the marks are labelled by an Combined EF", and the graph that accompanies this text is the next belonging to the third question:

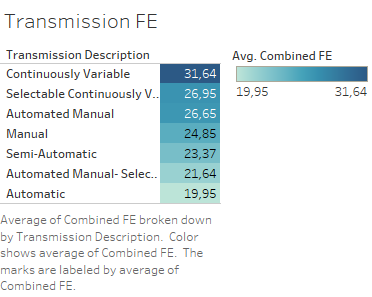


Image 4. Highlight Tables – Question 3

Of these advantages, there are many others that are related to the different technologies of this landscape, for example:

* Big Data allows us to extract information from various sources.
* We can predict future events with predictive analysis
* Supports fault tolerance through the Map / Reduce algorithm
* We can manage a large amount of information with Hadoop Ecosystem
* Storing all this unstructured information is possible thanks to the NoSQL databases.
* Make decisions at the managerial level in real time.

Currently, companies do not grant the right to Big Data technologies due to the updates in various software and the methods that integrate it to complete the different processes, since it is necessary for companies to have a data scientist in their companies so that help obtain information from the data generated from your customers and serve to improve existing products or services or create new ones that meet your needs, or to reduce the time of business process to reduce operating costs.

To conclude, we have seen that having the appropriate knowledge in Big Data technologies help us to manage important information that next it will be used to obtain the best result necessaries that cover several aspects to complete the objective of any company in the different areas such as marketing, business, medicine, governance, among others.

**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]

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]*