To help the car research company with the launch of its new research project which wants to analyze the fuel savings between different car manufacturers, we will use a public data set to achieve this goal, based on fuel savings for the models of the year 2015 and thus answer the different questions through the analysis that will be carried out in the Tableau tool that shows us the results in a visual way through graphics.

The data used is in a data set that has 729 records and 19 attributes, which provide the following information about each automobile model and according to the type of data in that field:

|  |  |
| --- | --- |
| 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

The data has been divided by the Tableau software itself; At the top, we have the Dimensions that normally contain categorical data such as Mfr Name and Carline, while the Measures contain numerical data such as Engine Displacement and City FE.

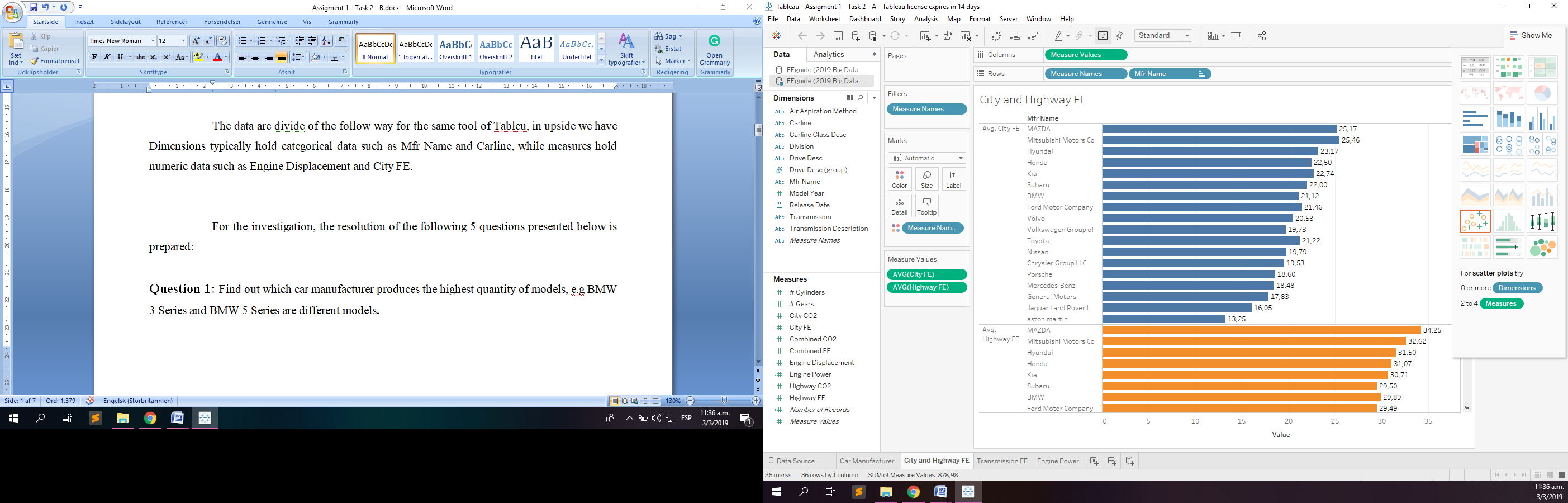


Image 1. Tableau Data Pane

For the investigation, the resolution of the following 5 questions presented below is prepared:

**Question 1:** Find out which car manufacturer produces the highest quantity of models, e.g BMW 3 Series and BMW 5 Series are different models**.**

In this question, the field Name of manufacturer that represents the name of the company that manufactured the car is used; also added the Carline field that is inside a function called Counting Distinct (CNTD) whose purpose is to count each different model within each car manufacturer. The graph chosen to represent these values ​​is packed bubbles, which allows us to show through its different sizes of shapes and labels, the comparison of the number of models that each car manufacturer owns, being the one with the largest number of models produced by BMW. a value of 86 units.

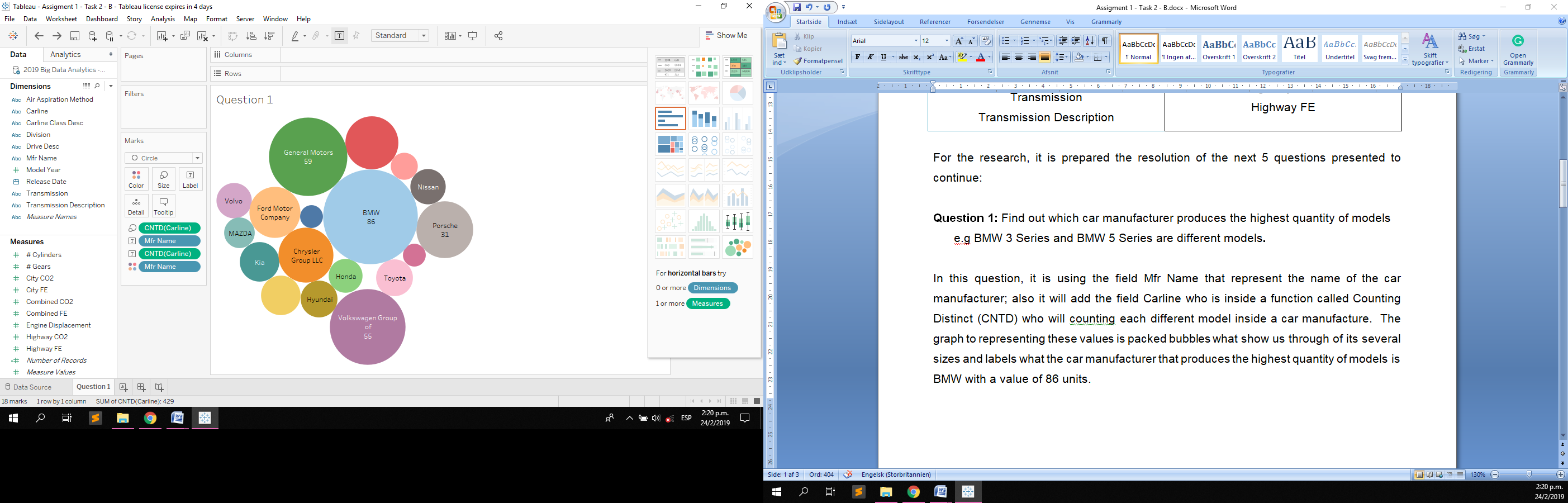


Image 2. Packed Bubbles Graph

The summaries of data for this graph can be seen in the following table sorted from largest to smallest:

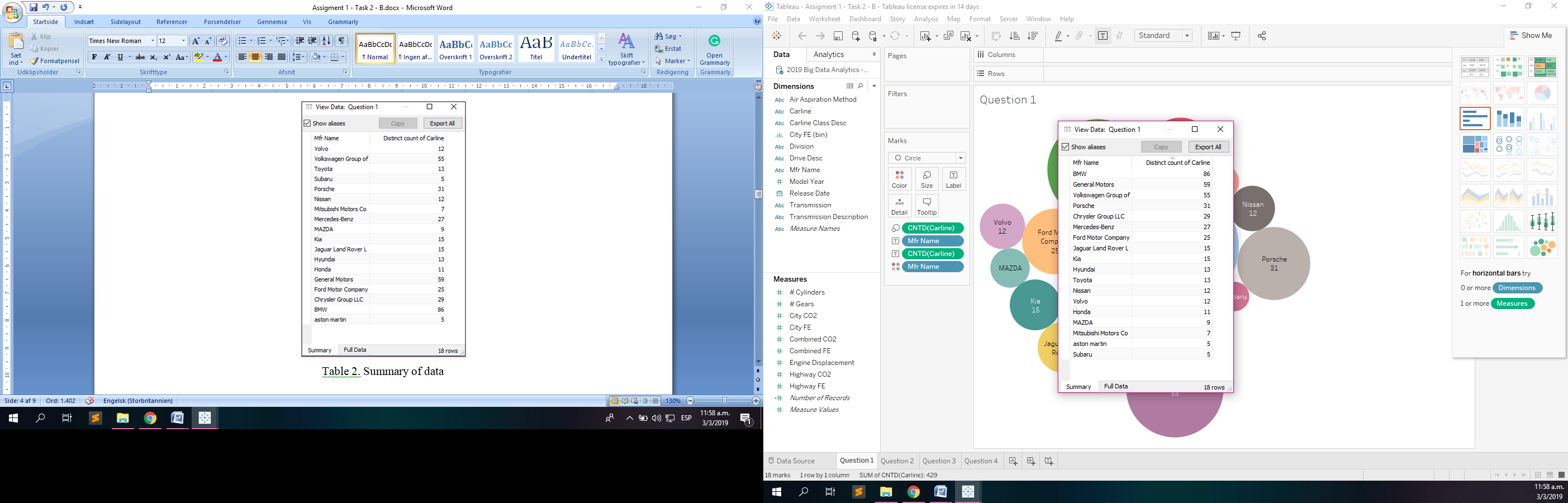


Table 2. Summary of data

**Question 2:** Find out the highest average fuel economy for city and highway driving from the given data set.

For this question we may need the fields named Mfr Name, to know which is the car manufacturer that has the highest fuel economy in the city and on the road joining the average (with the AVG function), the City FE and Highway fields FE represented with red and green respectively, and a side-by-side bar graph arranged alphabetically by the manufacturer's name.



Image 3. Side – By - Side Graph

As we see, the average highest between the car manufactures for the fuel economy in the city is Mitsubishi Motors Co with a value of 25,46 and for the fuel economy in the highway is MAZDA with a value of 34,25.

The summaries of data for this graph can be seen in the following table categorized by the values ​​of Avg. City FE and Avg. Highway FE:

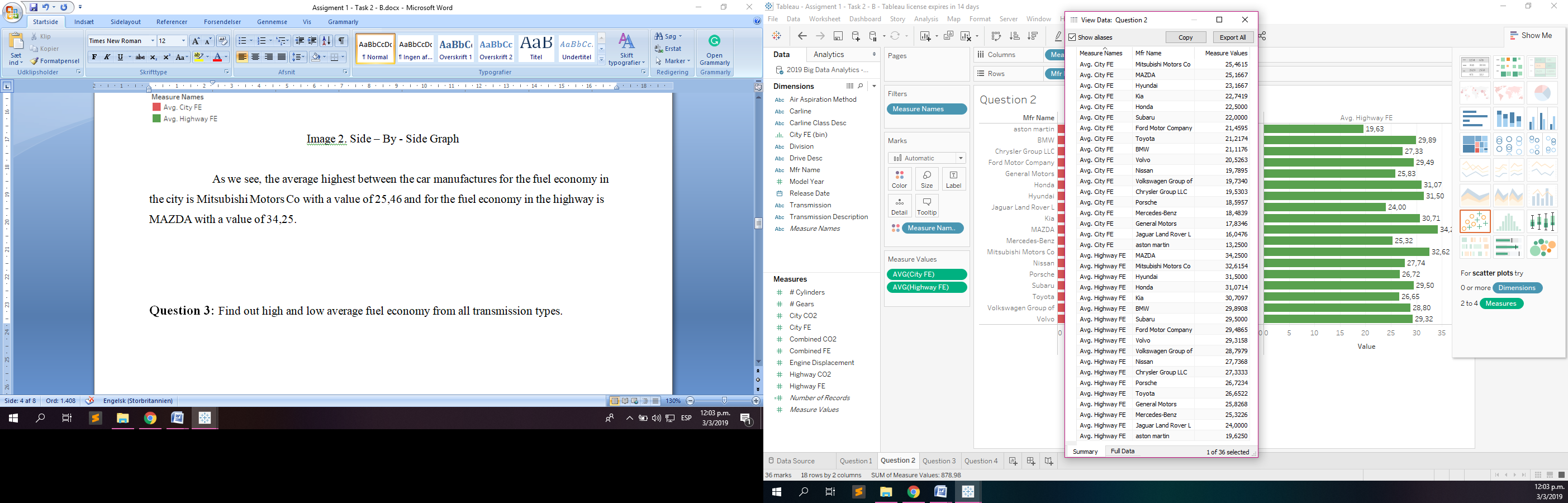


Table 3. Summary of data

**Question 3:** Find out high and low average fuel economy from all transmission types.

For this question, a horizontal bar graph was selected to represent the data. The fields used for this graph were the Description of the transmission and the average of the combined FE field within the AVG function. All transmissions were decremented according to the value obtained to better visualize the result.

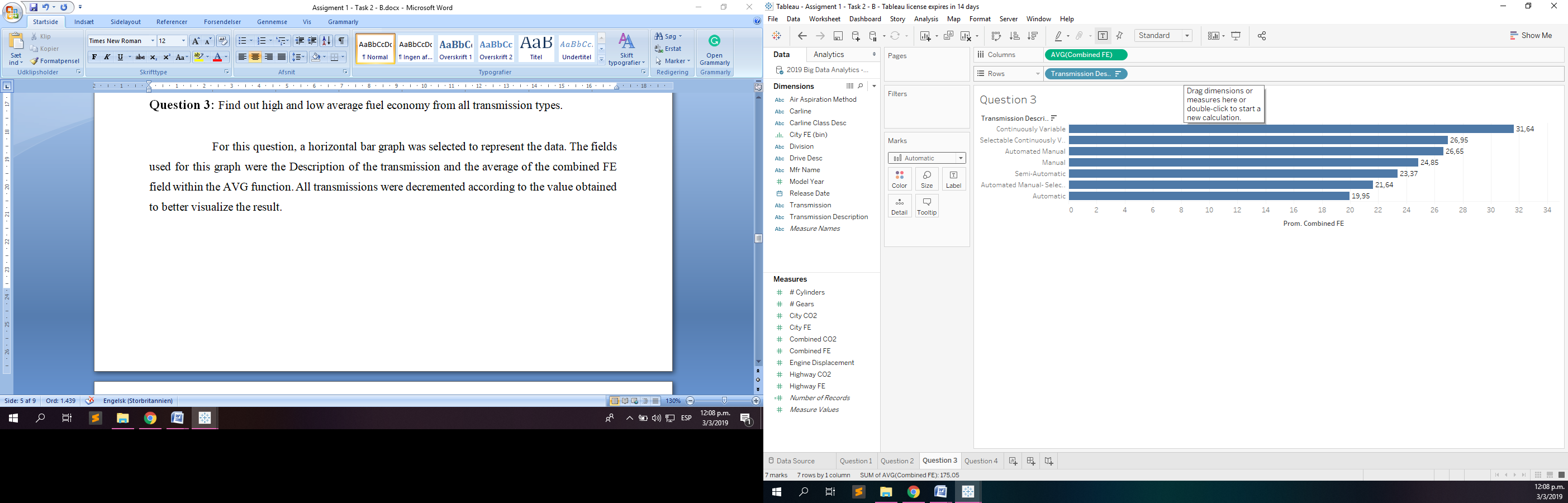


Image 4. Columns and Rows fields

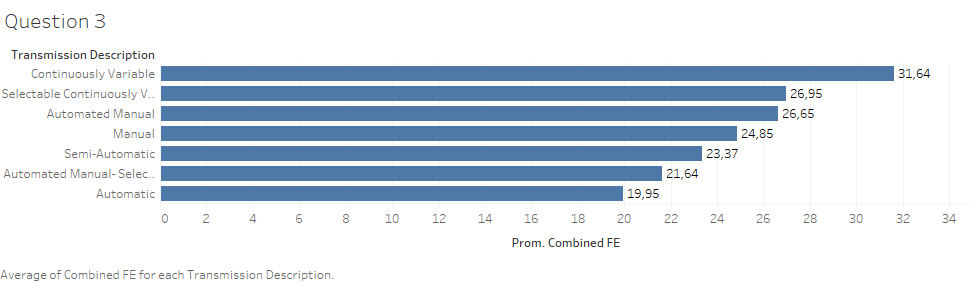


Image 5. Horizontal Bars Graph

As we can see, the transmission with the name belonging to Continuous Variable is of 31.64 becoming the highest average, while the transmission denominated Automatic has 19.95 being the lowest of all the values.

**Question 4:** Find out which car manufacturers have 4WD (4 wheel drive) and 2WD (2 wheel drive) models, whose engine power is more than 3.5.

For this question, we can use the fields Mfr Name, Drive Desc (containing the values ​​2 - Front wheel, 2 - Rear wheel, 4 - Wheel, All Wheel Drive, and Part - time 4 - Wheel Drive) to which we will apply a filter where only three values ​​will remain (2 front wheels, 2 rear wheels and 4 wheels). The field Motor displacement in the function AVG is added, whose result will be multiplied by a value of 1.6 and a filter will be made so that the manufacturers with values ​​greater than 3.5 of engine power appear.

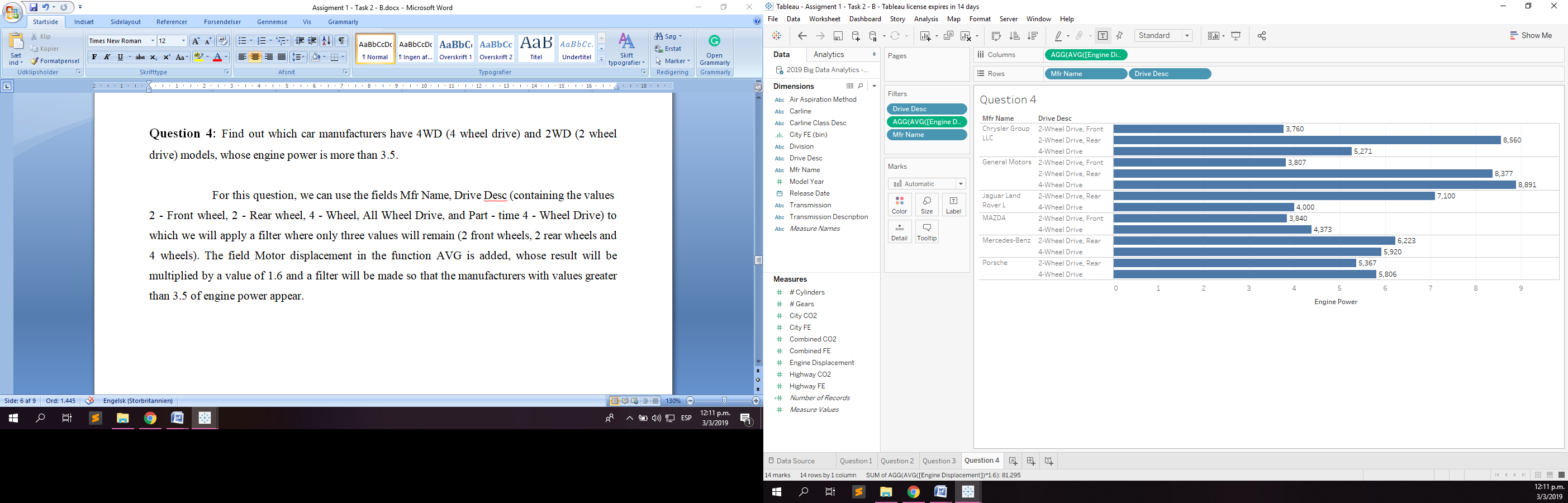


Image 6. Columns and Rows fields

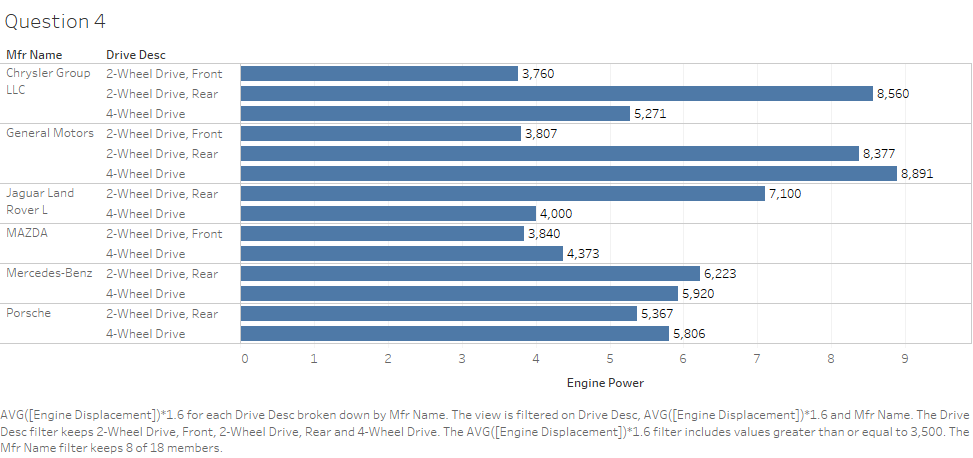


Image 7. Horizontal Bars Graph

As we can see there are several car manufacturers that meet the conditions of containing 2WD and 4WD models; and whose motor power is greater than the value of 3.5; These manufacturers are Chrysler Group LLC, General Motors, Jaguar Land Rover L, MAZDA, Mercedes - Benz and Porsche.

The summaries of data for this graph can be seen in the following table:

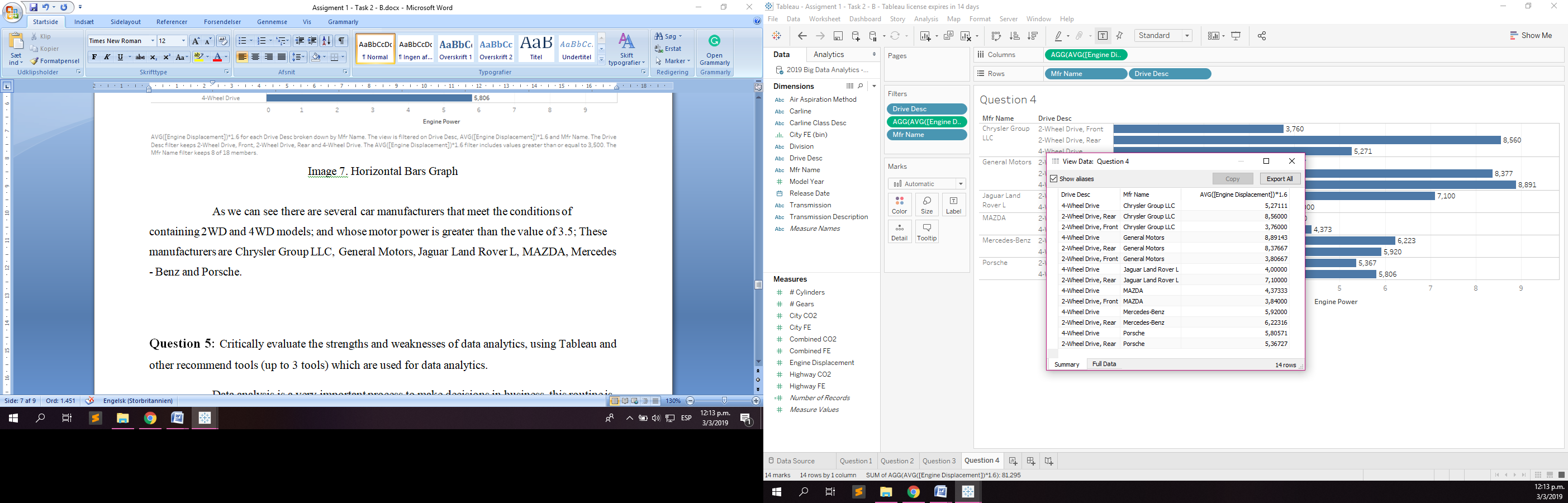


Table 4. Summary of data

**Question 5:** Critically evaluate the strengths and weaknesses of data analytics, using Tableau and other recommend tools (up to 3 tools) which are used for data analytics.

Data analysis is a very important process to make decisions in business, this routine is accompanied by other processes, such as the collection, storage and management of millions of data, all with the aim of understanding the existing knowledge between them. These processes are used for Big Data and many companies worldwide handle all this information to improve the products and services that are offered to their customers and improve the processes of their company.

Tableau is a tool that allows us to perform this process to analyze a large amount of data through graphics to represent them. The advantages of using Tableau are the following: unite different sources, such as databases, excel, among others; suggests various types of graphics depending on the data or dimensions to be used; we can have many graphics in the same project and share them as a board or a story; we can apply filters to different fields and apply mathematical functions; it is not necessary to know how to program to use it; You can update your data in real time; We can see the data before using them. Its disadvantage is that you need a license to use it.

As Tableau there are several tools capable of processing a large amount of data to be analyzed. 3 more tools will be named with the same objective.

1. DataMelt: It is software for the analysis of statistical data and scientific visualization. This program is more frequent in natural sciences, engineering and financial markets. The software supports several programming languages ​​such as Python, BeeanShell, Groovy, Ruby, Java and others. The organization can access large libraries through dynamic scripts. It runs on several platforms such as Windows, Linux, MacOS and Android. It has two disadvantages, one is that you need a license to access the advanced functionalities and two, to use these functions you need developers.

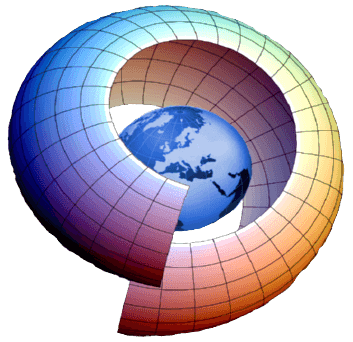


Image 8. DataMelt Logo

1. Orange: It is an open source data analysis and visualization tool developed by the University of Ljubljana (Slovenia). Users can extract data through visual programming or Python scripts in a shell window; exploration in statistical distributions, box diagrams or other techniques, and delves into data with decision trees, dynamic maps and linear projections. Its graphical interface allows the user to analyze exploratory data against coding. This tool has components for automatic learning, the extraction of data from external sources for natural language processing. It is also capable of running on platforms such as Windows, MacOS and Linux.



Image 9. Orange Logo

1. Trifacta Wrangler: This application is designed to help the data analyst to clean and prepare the data from various sources. Once the data is ready, it is imported into Trifacta Wrangler, who will automatically organize and structure the data. Its learning algorithm helps prepare the data for a more detailed analysis suggesting common transformations and additions. Trifacta Wrangles can import Microsoft Excel data, JSON files and CSV files without being processed. The tool also profiles the data to indicate the percentage of rows that have failed or mismatched values, and classifies the data depending on their type as dates or time. It is compatible with several platforms, but is limited to 100 MB of data.



Image 10. Trifacta Logo

**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 Feb. 2019]

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

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

Jwork.org (2019). *Welcome to DataMelt.* [Online] Available at: *https://jwork.org/dmelt/* [Accessed Feb. 2019]

Orange.biolab.si (2019). *Data Minig Fruitful and Fun.* [Online] Available at: *https://orange.biolab.si/* [Accessed Feb. 2019]

Trifacta.com/ (2019). *Why Trifacta.* [Online] Available at: *https://www.trifacta.com/products/why-trifacta/* [Accessed Feb. 2019]