In this assignment, we want to demonstrate how, through a tool, data manipulation processes can be performed to obtain relevant information and make assumptions about future events either in real time or stored from any data center. For the development of the task, we used the Tableau tools that allowed us to answer the questions asked by the car research company based on the analysis obtained from the data. Each question contains a graph obtained with the tool that shows the calculated data, along with the steps taken and the analysis or conclusion of the same.

The data dictionary shows for Tableau is the presented in the next figure:

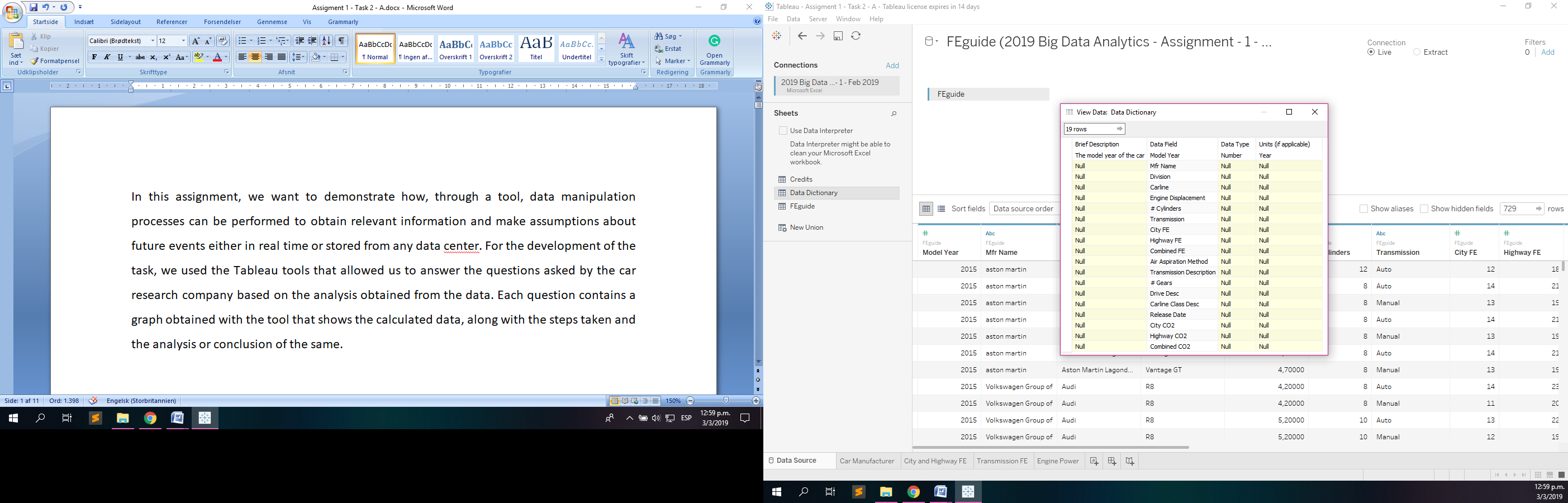


Figure 1. Data Dictionary

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

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

For this question, we need the data related to the name of the car manufacturer (Mfr Name) and the field called Carline, but the latter is configured with a measure called CNTD that allows us to count the number of relational models with each car manufacturer without those models being repeated in the count.

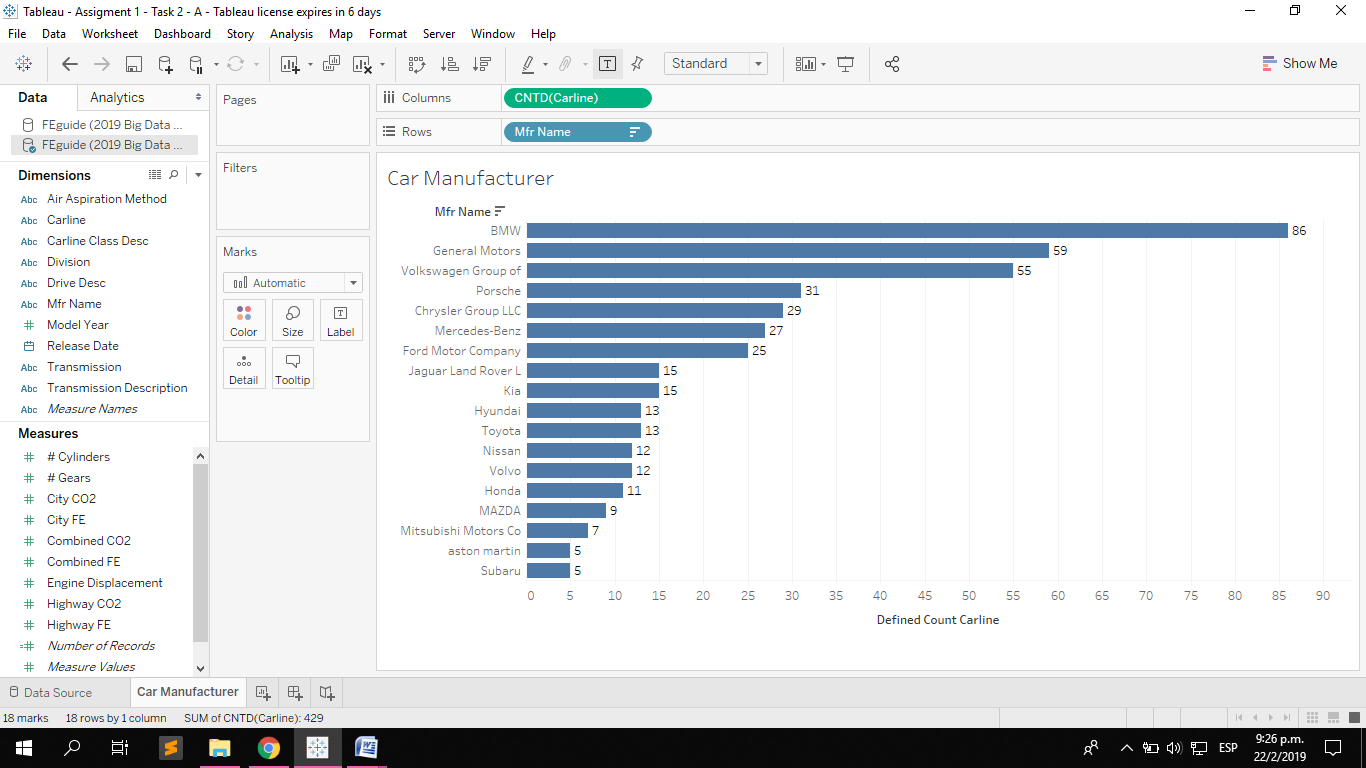


Figure 2. Horizontal Bars Graph

The layout to represent the data was a horizontal bar graph, with the name of the MFR axes for the rows and the number of different models for the columns. At the end of each bar there is a number that indicates the number of models that each car manufacturer has ordered in descending order.

As can be seen in the bar chart (Figure 1), the BMW manufacturer is the one with the largest number of models available with a total of 86.

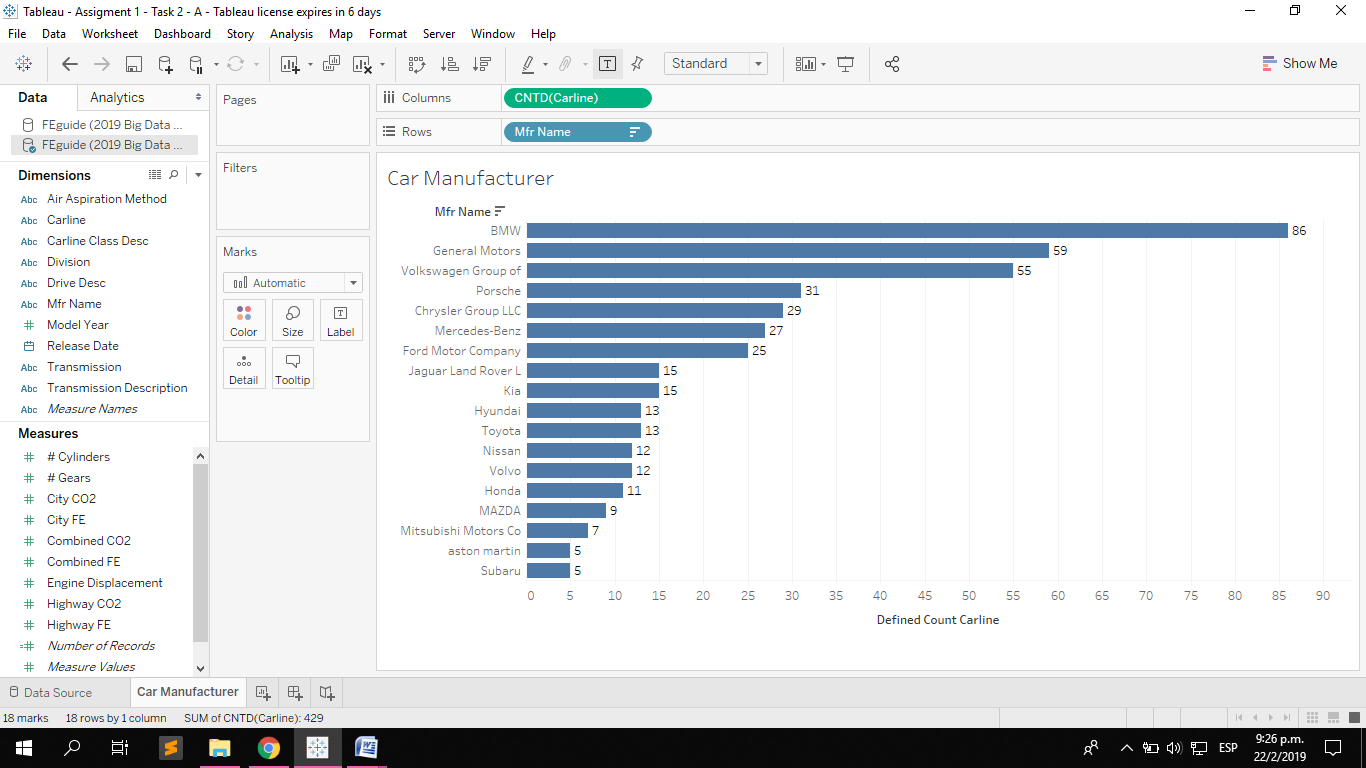


Figure 3. Columns and Rows Disposition

We can see the values of the graph represented in other graph in shape of table as the next:

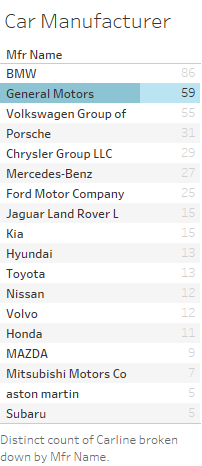


Figure 4. Text Tables Question 1

1. Find out the highest average fuel economy for city and highway driving from the given data set.

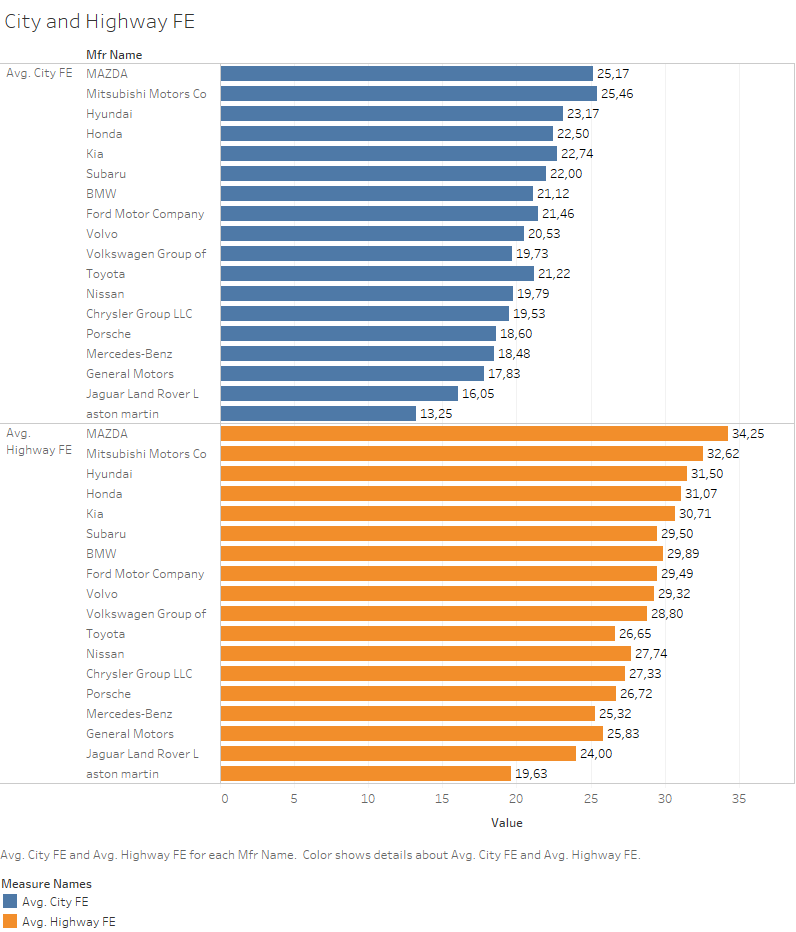
In the next bar graph called from side to side, we are representing the average fuel economy for the city (City FE) and for the highway (Highway FE) of each automaker. We can see that Mitsubishi Motors Co has the highest average with 25.46 for car manufacturers in terms of fuel savings for the city. And MAZDA has the highest average fuel economy for the highway with a value of 34.25.

Figure 5. Side – by - Side Bars Graph

This graph was made by adding as rows the names of the measurements (Avg. City FE and Avg. Highway FE) and the field named Name Mfr that includes the names of the car manufacturers, and as columns are the measurement values ​​that contain the averages of the fuel economy of the city (City FE) represented by the colour blue and the fuel economy by highway (Highway FE) represented by the colour orange. At the end of each bar is the average measured for each car manufacturer.

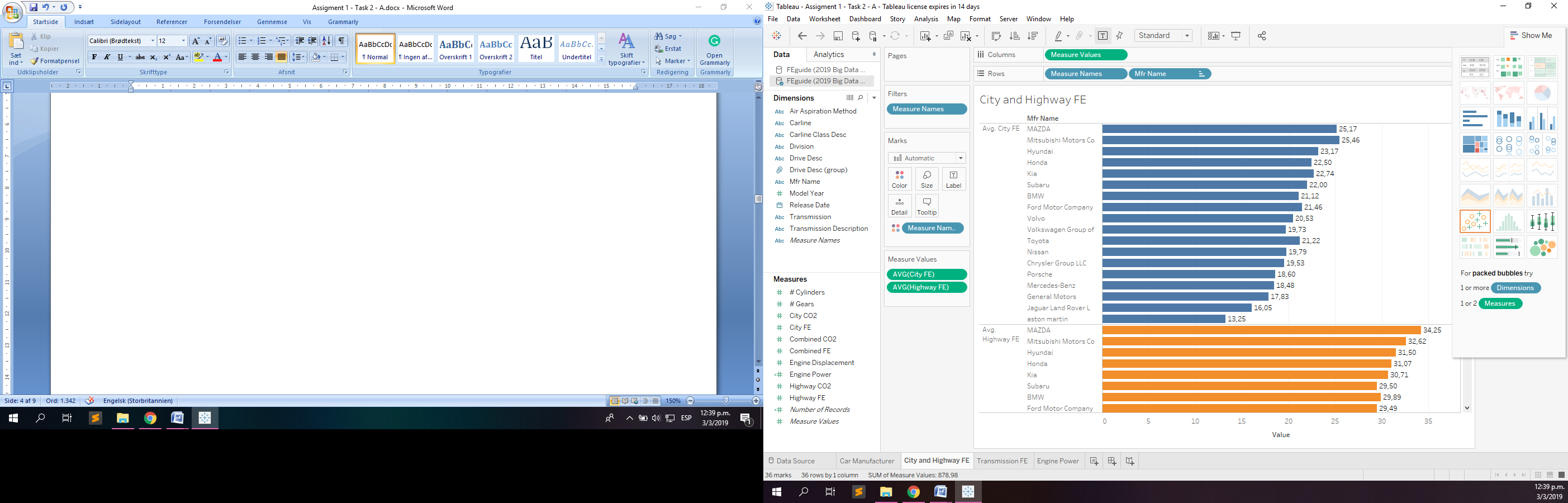


Figure 6. Shelves of Column and Row

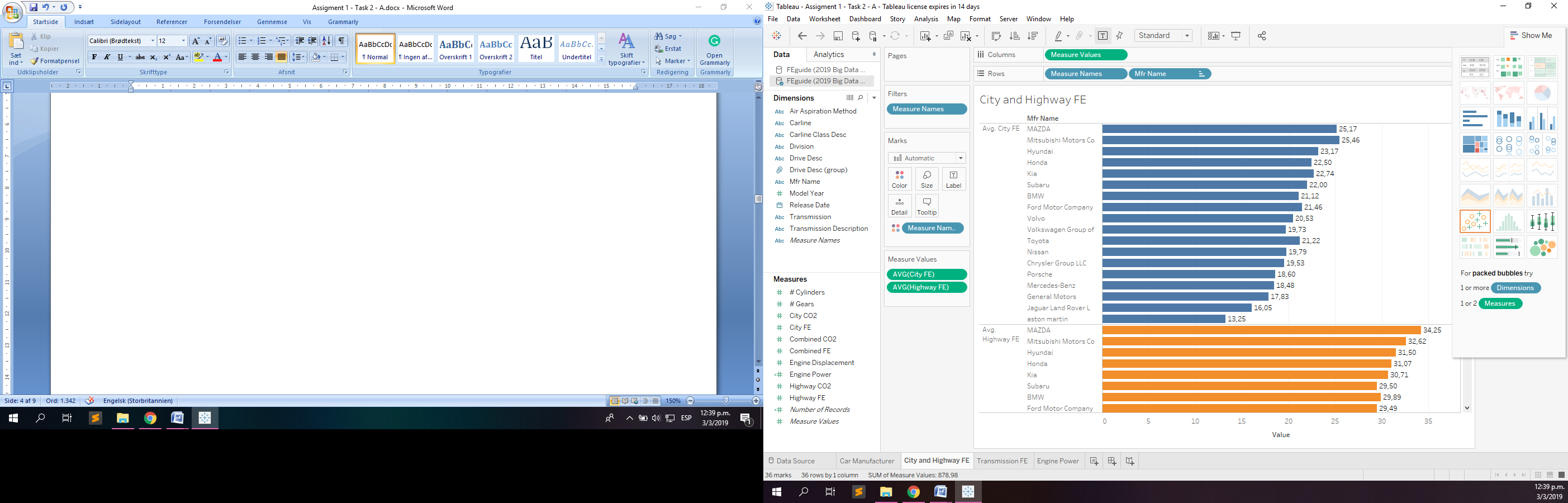


Figure 7. Fields of Measure Values

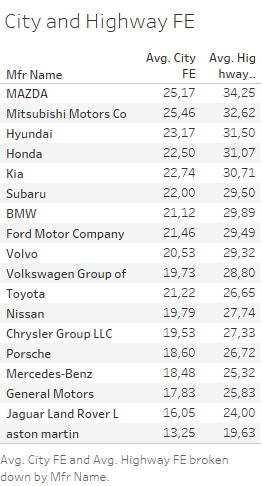


Figure 8. Text Tables Question 2

1. Find out high and low average fuel economy from all transmission types.

In this table, we can see which type of transmission is counteracted with the highest and lowest fuel economy. The fields used were the description of the transmission and the combined FE as the value of an average measurement (AVG). The table of highlights shows us that the value belonging to the continuous variable (31,64) is the highest value and the owner to Automatic (19.95) is the lowest value of all types of transmission found in the data given.

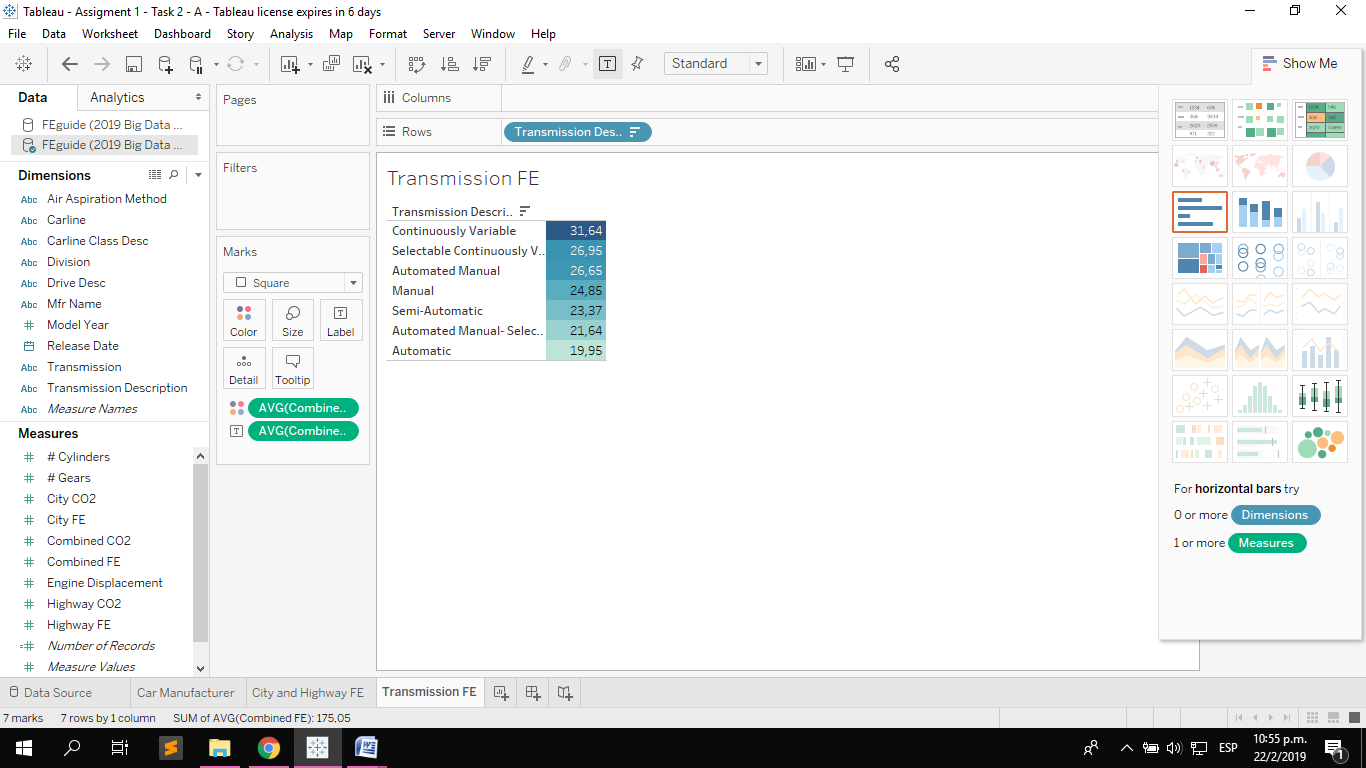


Figure 9. Highlights Table

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

With a text table, we can show the automakers that have 4WD and 2WD (in the latter are the values ​​belonging to the models with traction in the 2 wheels, front and with traction in the 2 wheels, in the part rear), and whose engine power value is greater than 3.5 (obtained with the filter to the displacement of the engine field and multiplied by 1.6 which are the liters used by that engine), called Engine Displacement calculated as an average value of each car manufacturer.

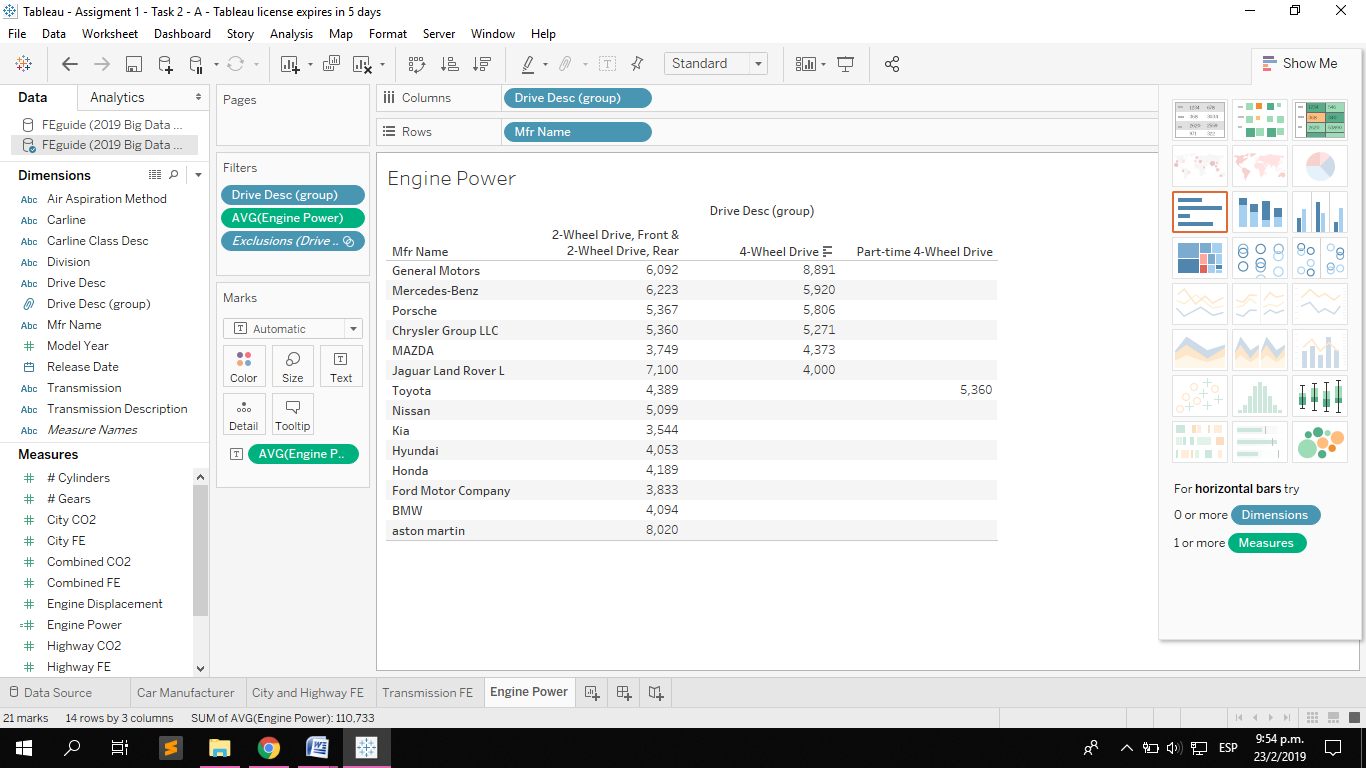


Figure 10. Text Table

As we can see, there are several car manufacturers such as General Motors, Mercedes-Benz, Porsche, Chrysler Group LLC, MAZDA, Jaguar Land Rover L and Toyota.

In the next figure, we can see the filters used for the calc in this graph:

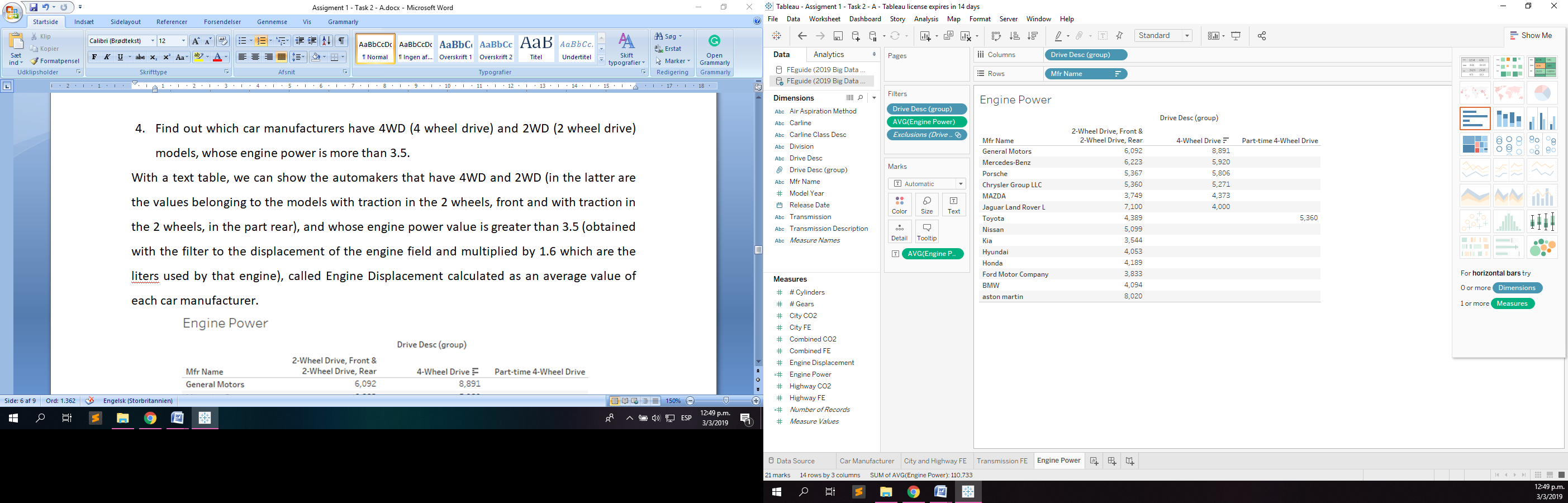


Figure 11. Filters

1. 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.

We analyze the data and we want business decisions to be made, as well as the tools of data analysis, as well as the ability to analyze the large amount of data from various sources, transactions and structured data. .

There are several tools for the analysis of big data that help us in this process of configuration and data management, cleaning and data mining, visualization, data collection, among others. Let’s list a series of tools that help us to make data analysis:

**Tableau**: It is a tool for visualization and business intelligence and has as advantages:

* To be very quick to implement and learn
* Has a very intuitive interface
* Accepts data from various sources
* Allows to see the content of the data before being used
* Has several graphics to represent the data
* Allows to export to PDF and image, as well as processing the data for excel formats
* It is a classification tool in the Business Intelligence environment
* It is not necessary to know any programming language.

For this last reason, a disadvantage is:

* You cannot create custom functions
* The software is paid and works with a license.



Figure 12. Tableau Logo

**Weka**: Advantage:

* Weka is an open source software
* It is capable of creating automatic learning algorithms
* It has tools to process, classify, model, group and visualize the data.
* The documentation is in its own web page.
* It can be executed in any platform.

Its main disadvantage is that it does not cover the algorithms for the modelling of sequences.



Figure 13. Weka interface

**R Studio**: The advantages of R Studio are:

* It is an open source software and GNU license with 4 freedoms.
* It is cross-platform, it can work on Mac, Windows and several UNIX systems.
* It has a large community of volunteers working to improve software and help with problems that arise.
* It offers all the techniques for the analysis of data, as well as the programming of new methods and statistics of routines easily.
* It has several graphics with high quality and versatility.

But the disadvantages of using R Studio is that you need to know its syntax to program the functionalities and the error messages do not specify the problem in detail.

RStudio Logo

Figure 14. R Studio Logo

**Matlab**: It is another software that helps us to integrate calculation, visualization, and programming in a fast and easy way, its advantages are:

* They contain mathematical functions of linear algebra and statistics.
* Functions for the visualization of data in 2 and 3 dimensions.
* High precision in mathematical calculation.
* Great support of resources and a large active community.

The disadvantages of this tool are:

* It is commercial software.
* You have a problem with your memory speed for a large amount of calculation you must process.

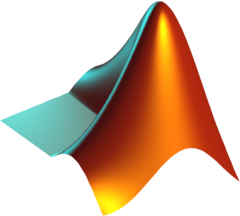


Figure 15. Matlab Logo

All this research helps us to know how was the fuel economy in the different models of cars in 2015 based on the data sets given within the excel file called 2019 Big Data Analytics - Assignment - 1 - Feb 2019, the which contains 729 records with 19 attributes that provided the necessary information to obtain the desired conclusions for each of the questions presented in this report.

**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]

Waikato.ac.nz (2017). *Weka 3: Data Minig Software in Java.* [Online] Available at: *https://www.cs.waikato.ac.nz/ml/weka/* [Accessed Feb. 2019]

Rstudio.com (2018). *RStudio: Support.* [Online] Available at: *https://support.rstudio.com/hc/en-us* [Accessed Feb. 2019]

Mathworks.com (2019). *Matlab: Support.* [Online] Available at: *https://uk.mathworks.com/support.html?s\_tid=gn\_supp* [Accessed Feb. 2019]