|  |  |
| --- | --- |
| Business analytics  Assignment | By Jonathan McDonagh  20074520 |

Table of Contents

[Introduction 3](#_Toc24829807)

[Aim 3](#_Toc24829808)

[Preparation 3](#_Toc24829809)

[Data Selection 3](#_Toc24829810)

[Data Cleaning 3](#_Toc24829811)

[Analyse 4](#_Toc24829812)

[Audi Passes/Fails 4](#_Toc24829813)

[Audi Passes 4](#_Toc24829814)

[Descriptive Statistics For Passes 5](#_Toc24829815)

[Audi Highest Failed 6](#_Toc24829816)

[Descriptive Statistics For Fails 6](#_Toc24829817)

[Cars Passed By Year 7](#_Toc24829818)

[Cars Failed By Year 7](#_Toc24829819)

[Focused Areas 8](#_Toc24829820)

[Audi Fails due to Safety Equipment 8](#_Toc24829821)

[Descriptive Statistics For Safety Equipment 8](#_Toc24829822)

[Audi Fails due to Braking Equipment 9](#_Toc24829823)

[Descriptive Statistics For Braking Equipment 9](#_Toc24829824)

[Audi Fails due to Steering & Suspension 10](#_Toc24829825)

[Descriptive Statistics For Steering & Suspension 10](#_Toc24829826)

[Focused Areas By Year 11](#_Toc24829827)

[Audi A4 12](#_Toc24829828)

[Audi A4 2002 Model 13](#_Toc24829829)

[Dashboard 13](#_Toc24829830)

[Conclusion 14](#_Toc24829831)

[References 15](#_Toc24829832)

Table of Figures

[Figure 1 Passed & Failed 4](#_Toc24829833)

[Figure 2 Highest Passed Cars 4](#_Toc24829834)

[Figure 3 Descriptive Statistics For Passes 5](#_Toc24829835)

[Figure 4 Audi Highest Failed Cars 6](#_Toc24829836)

[Figure 5 Descriptive Statistics For Fails 6](#_Toc24829837)

[Figure 6 Cars Passed By Year 7](#_Toc24829838)

[Figure 7 Cars Failed By Year 7](#_Toc24829839)

[Figure 8 Audi Fails due to Safety Equipment 8](#_Toc24829840)

[Figure 9 Descriptive Statistics For Safety Equipment 8](#_Toc24829841)

[Figure 10 Audi Fails due to Braking Equipment 9](#_Toc24829842)

[Figure 11 Descriptive Statistics For Braking Equipment 9](#_Toc24829843)

[Figure 12 Audi Fails due to Steering & Suspension 10](#_Toc24829844)

[Figure 13 Descriptive Statistics For Steering & Suspension 10](#_Toc24829845)

[Figure 14 Focused Areas By Year 11](#_Toc24829846)

[Figure 15 Audi A4 Focused Areas 12](#_Toc24829847)

[Figure 16 Audi A4 Focused Areas By Year 12](#_Toc24829848)

[Figure 17 Audi A4 2002 Model 13](#_Toc24829849)

[Figure 18 Audi Dashboard 13](#_Toc24829850)

# Introduction

For this assignment I will be investigating the NCT 2015 dataset. I will focus my attention on Audi car makes. To start of I will identify my research questions which are included in my aim below. After this I will start the data preparation and data discovery. Once I’ve collected the data needed I will present the data using suitable graphs, tables and other useful ways to represent my findings.

# Aim

In regards to my research questions the goal of this report is to help identify Audi cars most passed and failed model, to identify the model that failed most on vehicle safety equipment, braking and steering equipment. I will also identify by years and models according to the 2015 NCT dataset.

Before this I will clean the data prior to being analysed. I will familiars myself with the dataset. I will also include visual presentation of my findings which will also be in my excel workbook.

# Preparation

## Data Selection

For data selection I decided to focus on some main areas, Audi car passes and fails, fails due to Vehicle and Safety Equipment, fails due to Braking Equipment and fails due to Steering & Suspension. Audi cars with highest pass and fail rates according to the their year of birth.

## Data Cleaning

Data cleaning is an art that is very important to any data project. Good data analyst have skills to prepare the data and limit the issues such as duplicate data, extra characters and missing data.

I used the Clean() function to remove any non-printable symbols. I created a new column under the uncleaned data and used the =clean(cell reference) to clean the data. After this I created another new column under the cleaned data and used the trim() function to remove any extra spaces from the cleaned data text.

Firstly when starting the data preparation I moved all the data I will be working with to a new Excel sheet called ‘Audi Data’ here I started the cleaning process. I just used i.e. clean(D4). After this I used the trim function to trim the clean data This helped me make the text more readable before moving on.

Once all the data was cleaned I created a new excel sheet called ‘Audi Data Cleaned’, where I copied all the cleaned data over to it from the ‘Audi Data’ sheet before moving on.

# Analyse

## Audi Passes/Fails

Firstly I created a Pivot Table to see the total passes and fails for Audi cars. The total for all the Audi cars was 59816, what surprised me was the total amount of fails was 31074 compared to 28742 this was surprising as before doing this I expected there to be more cars passing compared to failing. With failing having a percentage of 51.95% and passing having a percentage of 48.05%. As you can see in the chart below.

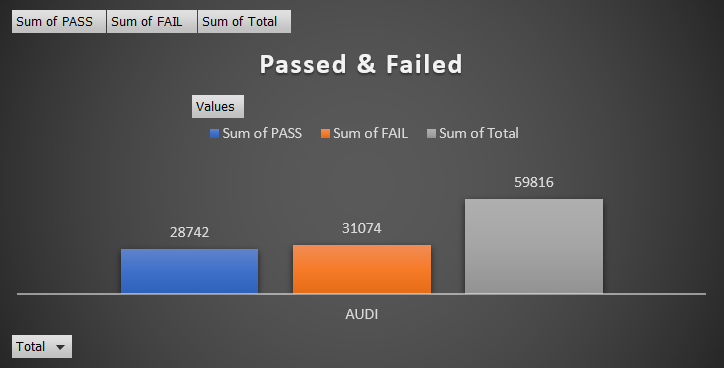


Figure Passed & Failed

### Audi Passes

As stated in my aim above I wanted to investigate the Audi car that had the highest NCT pass rate in 2015, I done this by creating a Pivot Table and filtering out cars that did not have many passes. After this I was able to see which car had the highest passes and using the max() function I could get an accurate answer to this question. It turned out that the Audi A4 2009 model had the most passes in 2015 with 2005 passes that year. With the Audi A4 2011 in second place with 1736 passes that year. As shown in my chart below.

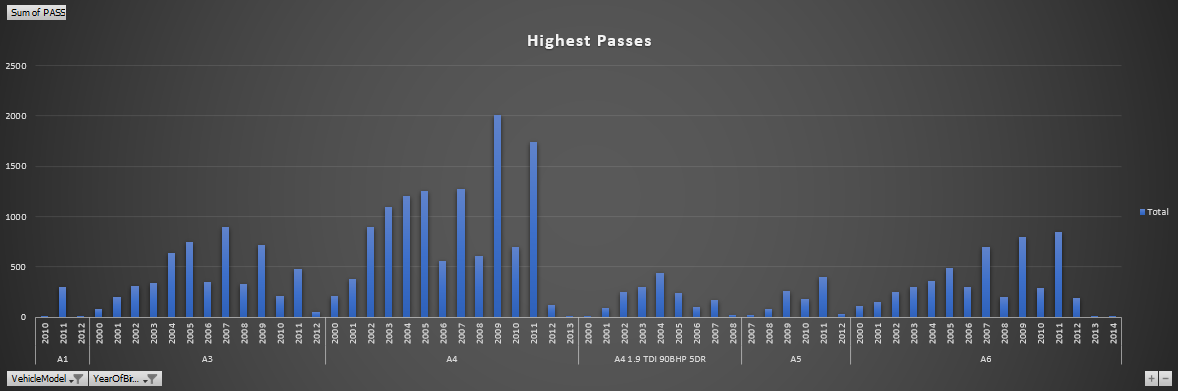


Figure 2 Highest Passed Cars

#### Descriptive Statistics For Passes

I used the Descriptive Statistics in Excel for each section to get a summary of the data I will be dealing with such as measures of Central Tendency, such as Mean, Median and Mode. For measures of variability I will be referring to Standard Deviation. I also wanted to see the Minimum and Maximum as I knew I would be seeing some of these numbers when dealing with the Pivot Tables and my charts.

The mean is the average of all values, in this case the mean is 75.83. The median is the middle observation when the data is sorted from smallest to largest, in this case the median is 2. The mode is the is value that appears most often, in this case the mode is 0.

The standard deviation is the square root of variance, which in this case is 227.82.

I am also interested in the Minimum and Maximum and Sum. The minimum is the lowest number in the dataset, in this case is 0. The maximum is the highest number in the dataset, in this case is 2005 which is the car with the highest passes as stated above in Figure 2. The sum is all the values added together in this case is 28742 telling us there was 28742 Audi cars that passed. As shown in the figure below.

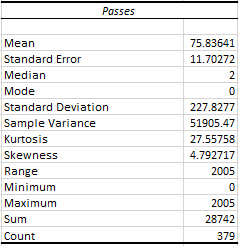


Figure Descriptive Statistics For Passes

When dealing with descriptive statistics in the rest of the report I won’t be giving the definition for each as I’ve already stated them above. I will just be focusing on the Mean, Median, Mode, Standard Deviation, Minimum, Maximum and Sum.

### Audi Highest Failed

I also wanted to cover the area of highest failed Audi car according to the NCT 2015 dataset. I done this by creating a Pivot Table and filtering out cars that did not have many fails. After doing so I was able to see which car had the highest fails. It turned out that the Audi A4 2002 model had the most fails in 2015 with 2119 fails that year, with the Audi A4 2003 in second place with 1917 fails that year. As shown in my chart below.

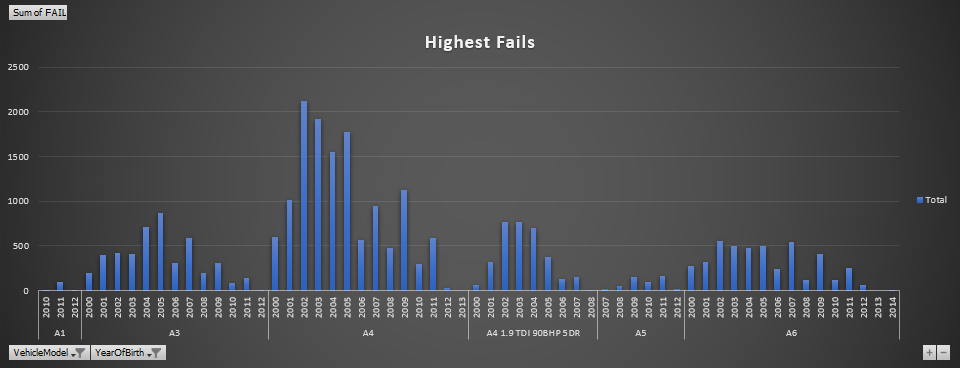


Figure Audi Highest Failed Cars

#### Descriptive Statistics For Fails

I used the Descriptive Statistics in Excel for the fails data, to get a better understand of the areas described above. I found that the mean for the fails was 81.98 which is only a little bit higher than the passes as shown in the Descriptive Statistics for Passes section. I also found that the Median for Fails was 3 and the Mode was 1. The standard deviation for fails was 245.13. The minimum for fails was 0 and the maximum was 2119. The total for the fails was 31074. This is all shown in the figure below.

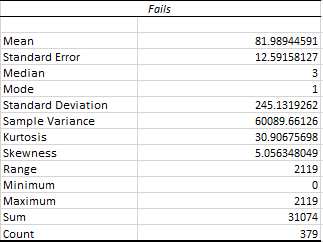


Figure Descriptive Statistics For Fails

### Cars Passed By Year

With the Audi data I am investigating I wanted to see what cars passed the most according to year of birth for all Audi cars that passed in the dataset, I done this using a Pivot table and only using the relevant data the year of birth and the sum of pass for each year. I found that Audi cars with the year of birth of 2009 had the highest passed rate with 3783 passes, in second place was 2011 with 3761 passes. Between these two years of birth there was only 22 cars that passed in the difference. As shown below.

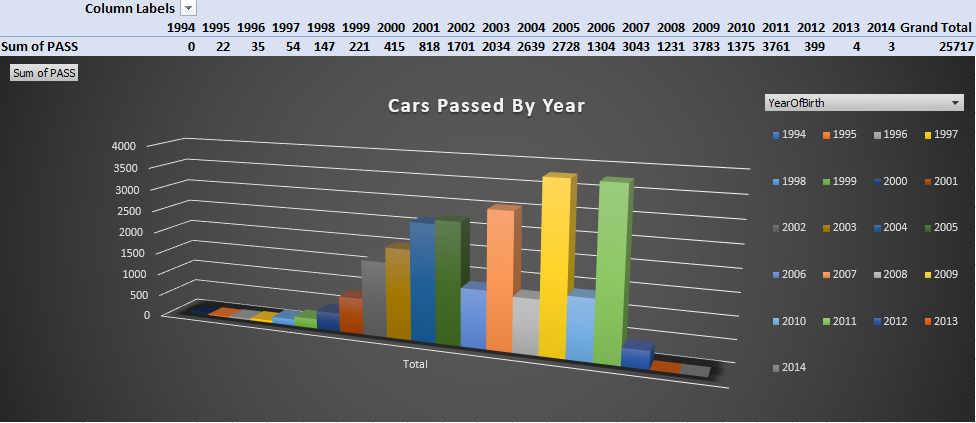


Figure Cars Passed By Year

### Cars Failed By Year

With the Audi data I am investigating I also wanted to see what cars failed the most according to year of birth for all Audi cars that failed in the dataset, I done this again by using a Pivot table and only using the relevant data the year of birth and the sum of fail for each year. I found that Audi cars with the year of birth of 2002 had the highest failed rate with 3861 fails, in second place was 2003 with 3600 fails. As shown in the chart below. I have started noticing that 2002 has the highest fails overall, with the Audi A4 2002 model having the most fails.

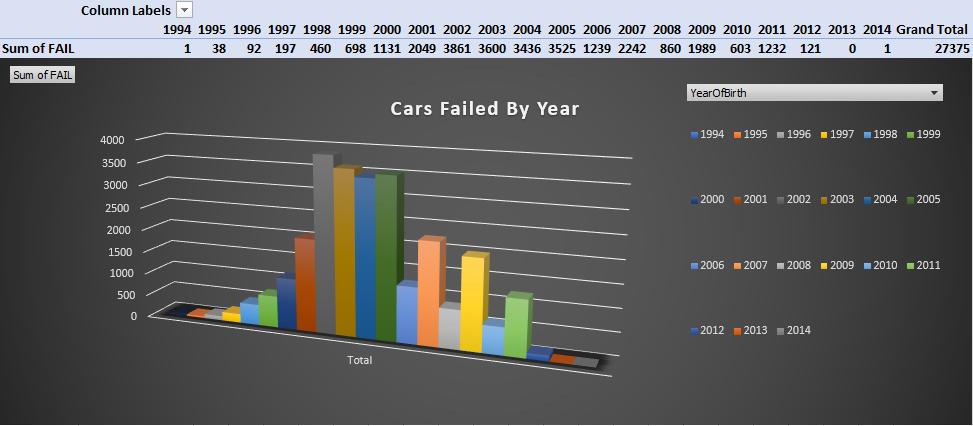


Figure Cars Failed By Year

## Focused Areas

### Audi Fails due to Safety Equipment

I thought this would be an interesting area to focus on as safety equipment is obviously important when it comes to cars, I wanted to investigate what model of car and year was failing the most in this area. I firstly created a Pivot Table I had the sum of the total fails and the total of the cars that failed due to Safety Equipment. From here I moved onto creating a bar chart I filtered out car models that didn’t have many fails in this area, I done this by filtering cars out that had under 30 fails due to safety equipment, I also filtered out cars describe as ‘Other’. This made my chart a lot more readable. The A4 model had a total of 2807 fails over all years and the second highest model with fails due to safety equipment was the A6 with 964 fails in total over all years. As you can see in the chart below Audi A4 2002 with the most fails had 483 fails.

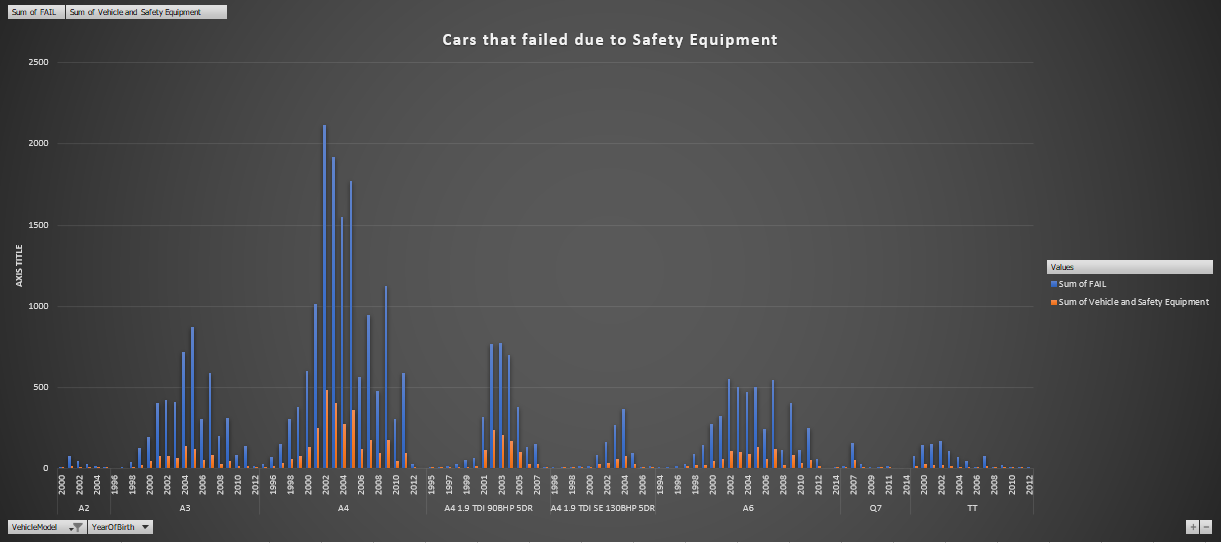


Figure Audi Fails due to Safety Equipment

#### Descriptive Statistics For Safety Equipment

I used the Descriptive Statistics in Excel for the fails due to Safety Equipment. I wanted to backup the data above as I found above that the Audi A4 2002 had the most fails with 483.

I found that the mean for the fails due to Safety Equipment was 16.74. I also found that the Median for fails due to safety Equipment was 1 and the Mode was 0. The standard deviation for fails due to safety equipment was 51.47. The minimum for fails due to safety equipment was 0 and the maximum was 483 as stated above. The total for the fails due to safety equipment was 6354. This accounted for 20.44% of all fails. This is all shown in the figure below.

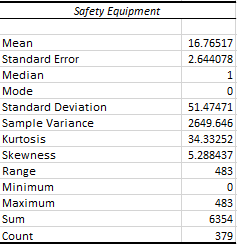


Figure Descriptive Statistics For Safety Equipment

### Audi Fails due to Braking Equipment

Seeing as I am focusing on Safety Equipment of cars I thought Braking would be another good area to focus on as braking equipment in cars is of course another important thing for any car. I wanted to focus on what car model and year is failing the most due to braking equipment. Doing this in a new sheet I created a Pivot Table which had the sum of the total fails the total of the cars that failed due to Braking Equipment. I created another bar chart and filtered out car models that didn’t have fails over 25 due to braking equipment filtering out cars described as ‘Other’ as well. The car with the most fails again was the Audi A4 2002. Audi A4 cars over all years of birth had 2509 fails due to braking equipment with the 2002 model having the highest at 547. The second highest model over all years of birth that failed in this area was the Audi A4 1.9 TDI 90BHR 5DR with 857 fails due to braking. As shown in the chart below.

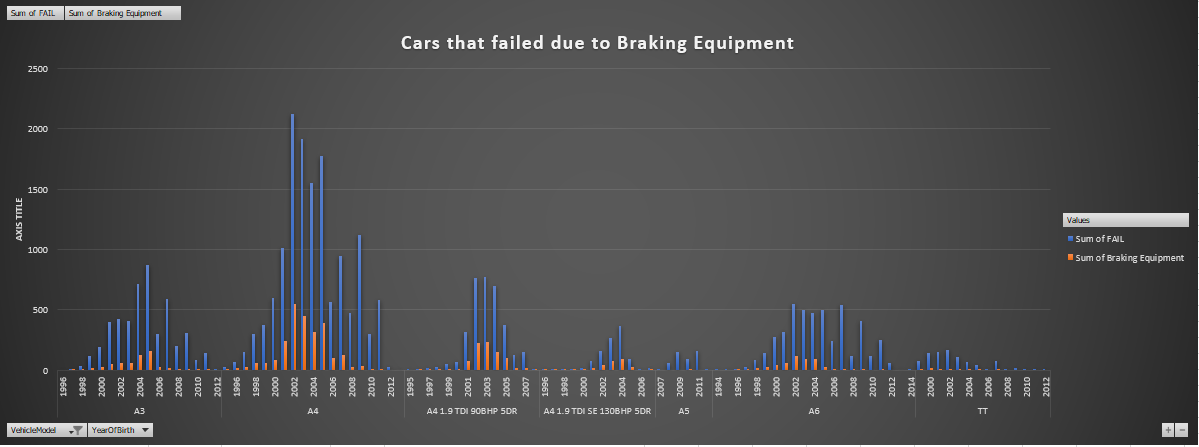


Figure Audi Fails due to Braking Equipment

#### Descriptive Statistics For Braking Equipment

I used the Descriptive Statistics in Excel for the fails due to Braking Equipment. I wanted to backup the data above as I found above that the Audi A4 2002 had the most fails with 547.

The mean for cars that failed due to braking equipment was 13.35. The median and mode both had 0. The standard deviation was 52.84. The minimum was 0 and the maximum as stated above was 547 the Audi A4 2002 model. The sum for all fails due to braking equipment was 5054. This accounted for 16.26% of all fails.

This is all shown in the figure below.

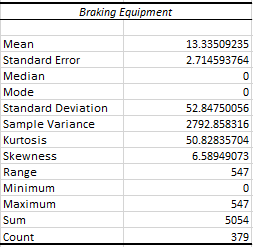


Figure Descriptive Statistics For Braking Equipment

### Audi Fails due to Steering & Suspension

Another area I wanted to focus on was Steering and Suspension. I wanted to find out what car model and year was failing the most due to steering and suspension. Doing this in a new sheet I created a Pivot Table with the total amount of fails and total amount of fails due to Steering and Suspension. I used a bar chart and filtered out car models that didn’t have over 30 fails due to steering and suspension I also filtered out cars described as ‘Other’.

I was very surprised to see that the model that failed the most was again the Audi A4 2002 model with 942 fails due to steering and suspension. Audi A4 cars over all years of birth had 4642 fails due to steering and suspension with the next highest failing due to steering and suspension was the Audi A4 1.9 TDI 90BHP 5DR over all years of birth had 1554. As you can see in the chart below.

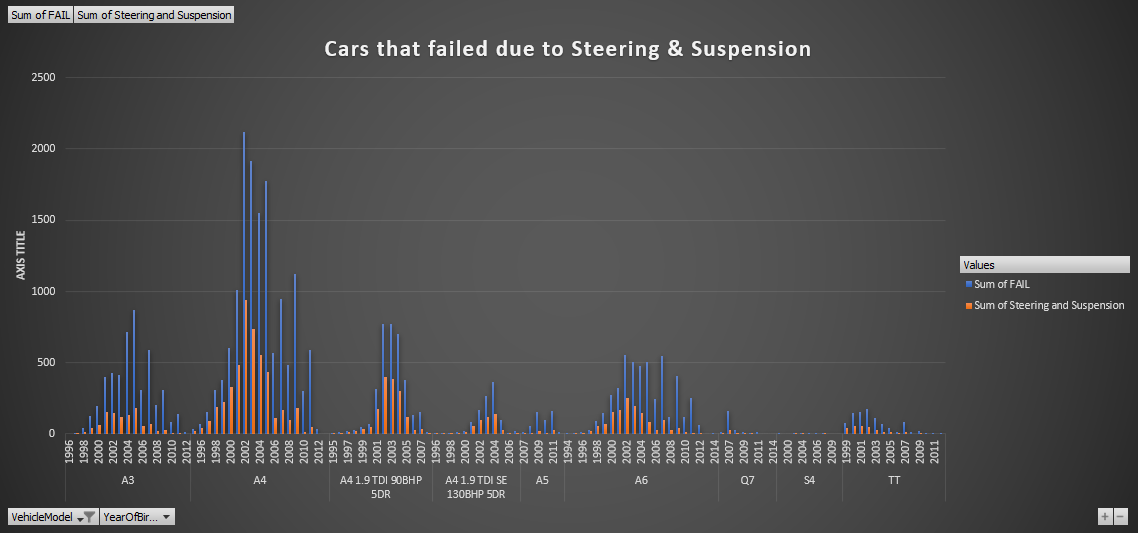


Figure Audi Fails due to Steering & Suspension

#### Descriptive Statistics For Steering & Suspension

I used the Descriptive Statistics in Excel for the fails due to Steering & Suspension. I wanted to backup the data above as I found above that the Audi A4 2002 had the most fails with 942.

The mean for fails due to steering & suspension was 25.89. The median was 1 and the mode was 0. The standard deviation for this area was 89.32. The minimum was 0 and as stated above the maximum was 942 which is the Audi A4 2002 model. The sum for all fails due to steering & suspension was 9813. This accounted for 31.57% of all fails. This is all shown in the figure below.

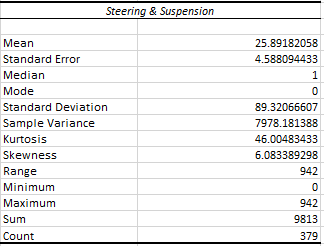


Figure Descriptive Statistics For Steering & Suspension

### Focused Areas By Year

After investigating my focused areas I thought it would be a good idea to focus on Audi cars by year of birth instead of model, I done this using a Pivot Table and a Bar Chart. I wanted to investigate the years 1993 – 2012.

From my findings I noticed that 2002 had the most total fails as well as the most fails in my focused areas of Vehicle and Safety Equipment, Steering and Suspension and Braking Equipment. With a total sum of 4321 fails. 984 fails due to Safety Equipment, 1921 fails due to Steering and Suspension and 1042 fails due to Braking Equipment. The next highest year being 2003 with a total of 4096 fails. As I was working through the data I noticed a trend happening with Audi 2002 models having the highest amount of fails, as shown in the Bar Chart below.

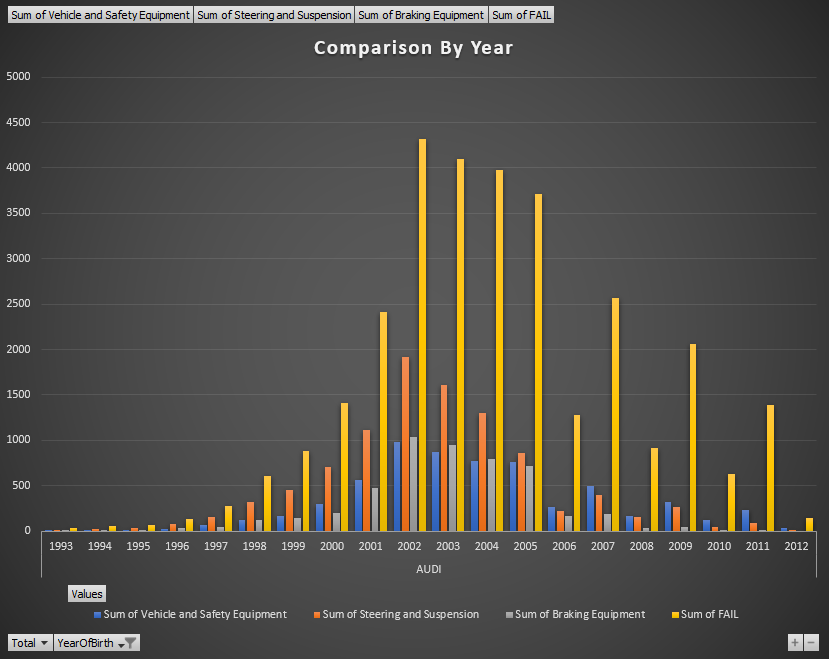


Figure Focused Areas By Year

### Audi A4

Due to Audi A4 having the most fails by far compared to other Audi cars in the areas of Vehicle and Safety Equipment, Steering and Suspension and Braking Equipment. I wanted to compare it to the total fails of Audi A4 cars. Audi A4 cars have the sum of 13937 total fails, where these three areas make up 9958 of these fails making up 71.45% of the Audi A4 fails. Making up a large chunk of the total fails for these model. As shown in the bar chart below.

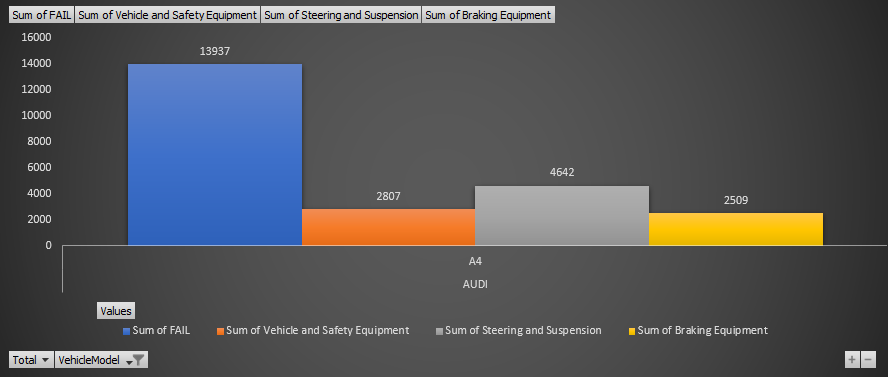


Figure Audi A4 Focused Areas

I also crated a Pivot Table for Audi A4 fails that show the fails per year in my focused areas. I focused on the years between 2001 – 2013. As shown in the chart below for Vehicle and Safety Equipment, Braking Equipment and Steering and Suspension we can clearly see that the Audi A4 2002 has the highest amount of fails in each area.

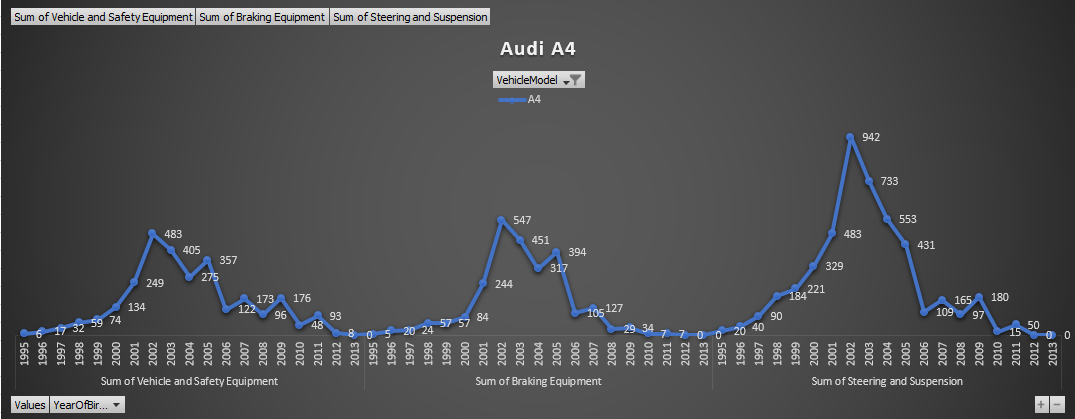


Figure Audi A4 Focused Areas By Year

#### Audi A4 2002 Model

To go a step further I wanted to compare the 2002 model as seen above this model and year had the most fails in the areas of Vehicle and Safety Equipment, Steering and Suspension and Braking Equipment. This model and year of birth in the areas above had a total of 1972 fails out of 4091. Almost half, with Steering and Suspension having the most fails out of these three areas with 942 fails. As shown in the pie chart below.

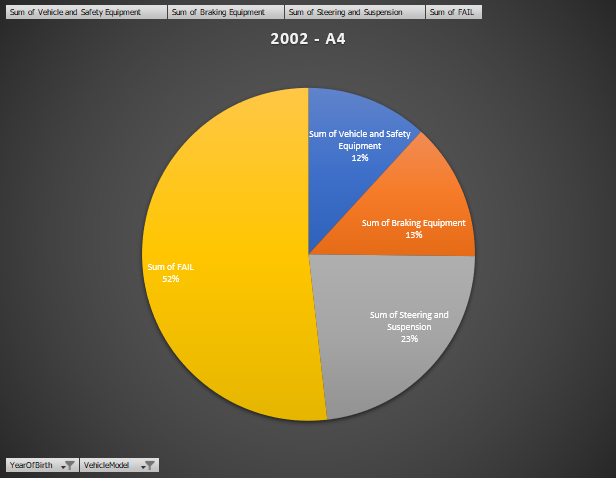


Figure Audi A4 2002 Model

# Dashboard

I created a dashboard to provide focused information in a visually appealing way, allowing the user to click between year of birth, vehicle model, number of fails and number of fails due to safety equipment, braking equipment, steering & suspension. I done this all in a new sheet called Dashboard in excel using some formatting options and report connections.

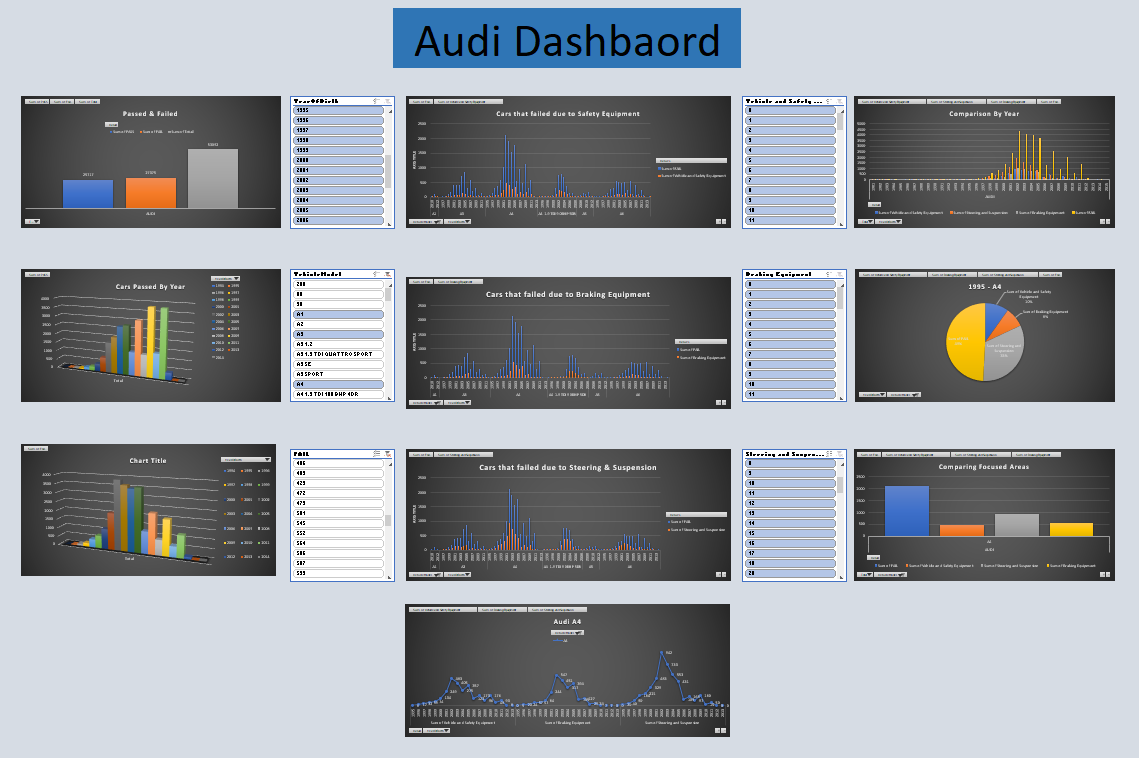


Figure Audi Dashboard

# Conclusion

I really enjoyed this assignment and feel like I have definitely improved my interpretation of data and how to better deal with data. I got to learn the importance of cleaning data, using pivot tables and charts to give a better visualization of the data collected.

I feel like I met my research questions in the aim of my report. I wanted to identity what Audi car was the most passed which was the Audi A4 2009 model with 2005 passes in 2015. I wanted to identify what Audi car was the most failed which was the Audi A4 2002 with 2119 fails.

I also identified the model and year that failed the most on vehicle safety equipment, braking equipment and steering & suspension surprisingly this was the same model for all three of my focused areas being the Audi A4 2002. With 483 fails due to vehicle and safety equipment, 547 fails due to braking equipment and 942 fails due to steering and suspension.

My three focused areas combined made up 68.27% of the total fails for all Audi cars.

Overall I really enjoyed the process of this assignment some areas are included in my excel file that was not covered in my report such as histogram for cars fails, cumulative curve for car fails and some extra pivot tables.

# References

Tutors-design.netlify.com. (2019). *Tutors*. [online] Available at: https://tutors-design.netlify.com/course/business-analytics-1.netlify.com [Accessed 12 Nov. 2019].