# Part I

<https://public.tableau.com/app/profile/chashy.luria/viz/NLuriafinalproject/yearvsweightcylinder?publish=yes>

A graph with blue lines and orange lines

Description automatically generated

**I compared the model year to two different aspects: the acerage cylinder count and the weight. There seems to be a strong correlation between the weight and the cylinders. Cylinder counts are relatively consistent across the years, with occasional increases or decreases. I do notice a significant increase in the year 1973 and then a sharp decrease right after that. The weight trend fluctuates a bit more but remains relatively consistent with the cylinders. After 1978, it seems that the cars were made to be lighter and with less cylinders, perhaps to be more echo friendly.**

A graph showing the growth of a company

Description automatically generated with medium confidence

<https://public.tableau.com/app/profile/chashy.luria/viz/NLuriafinalproject/yearvsaccelerationhorsepower?publish=yes>

**In this model, I am depicting the model year compared to both the acceleration and the horsepower. The acceleration is slightly increasing with time, with little fluctuations. The horsepower seems to be on the decline. In 1973 there was a significant increase in the horsepower. It sharply dropped right after that and slowly plateaued.**

A close-up of a computer code

Description automatically generated

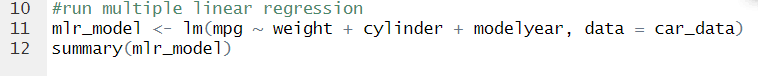
A close up of a text

Description automatically generated

A screenshot of a computer error

Description automatically generated

1. **Multiple R squared** = 0.7092
2. **Adjusted R squared** = 0.7082
3. **Complete Linear Equation** : mpg = 45.964 – 0.007(x1…xn)



A screenshot of a computer program

Description automatically generated

1. **Multiple R squared** = 0.815
2. **Adjusted R squared** = 0.813
3. **Complete Linear Equation** : mpg = -9.66 -0.006(weight) -0.243(cylinder) +0.694(modelyear)

\*Residuals Model\*

A screenshot of a computer code

Description automatically generated

A computer code with text

Description automatically generated with medium confidence

A graph with numbers and lines

Description automatically generated

\*Histogram of residuals\*

A screenshot of a computer program

Description automatically generated

A graph of a number of pink bars

Description automatically generated

# Part II

A computer screen shot of text

Description automatically generated

*Do different communication channels have a significant effect on the call duration?*

Answer:

I reject the null hypothesis that there is a channel has a significant effect on call duration. The high p-value that resulted from the ANOVA model suggests that the mean of the call durations is not very different across different communication channels.

Business Implication:

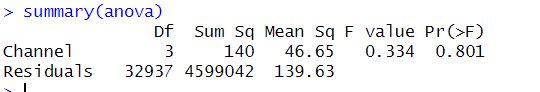
From a business perspective, the type of channel does not significantly affect the time spent on calls. This can be valuable information when the company is looking to hire more employees for different channels. It also suggests that all the departments might be doing a great job, if no department is faster than the other.

Code:

A white background with black text

Description automatically generated

Results:



* DF = 3. This means that there are 4 different channels, n-1
* P value = 0.801. This is higher than the significance level indicating no significant difference in the means of the data

Visualization:

A screenshot of a computer code

Description automatically generated

A chart with purple squares

Description automatically generated

* The median, IQR, whiskers and box are all approximately the same across the different methods of communication. The call durations are fairly consistent regardless of the communication method.
* Notice that there are no outliers, which shows that there are no extreme call durations.



Does the csat score depend on the call reason, at all?

**Answer:**

**Upon evaluation, there is no correlation between customer satisfaction score and call reason. The high p-value that resulted from the simple linear regression model shows no significance between the dependent and independent variables.**

**Business Implications:**

**Since the reason for the call doesn’t significantly affect the customer satisfaction score, the business may wish to research other factors that could positively influence customer satisfaction. These can be agent performance and professionality, and product or service quality.**

**Code:**

**A screenshot of a computer program

Description automatically generated**

**Results:**

**A screenshot of a computer

Description automatically generated**

**A close up of numbers

Description automatically generated**

* **P value = 0.2487. This is higher than the significance level of 0.05. We would fail to reject the null hypothesis because there is not enough evidence to conclude that reason has a significant effect on csat score.**

**Data Visualization:**

**A black and white text

Description automatically generated**

**A graph with purple dots

Description automatically generated with medium confidence**

* **The dots on the plot show observations of the continuous variable – csat score.**
* **The data is not skewed in any direction.**
* **There does not seem to be any outliers.**
* **The data points are very clearly uniformly arranged. This emphasizes the lack of correlation.**



Is there a correlation between the response time and the call center?

**Answer:**

**No, there is no evidence about a significant relationship between the call center and the response time. I will fail to reject the null hypothesis and assume that the 2 variables are independent of each other. This was proven from the high p-value that resulted from the chi-square test.**

**Business Implications:**

**The results indicate that that all call centers are responding to customers similarly. Response time is not a location specific issue.**

**Code:**

**A screen shot of a computer code

Description automatically generated**

**Result:**

**A screenshot of a computer code

Description automatically generated**

* **The contingency table demonstrates a breakdown of each center, and how they handled the calls they received.**
* **Above SLA means responses that were delayed. Within SLA are calls that me the service level agreement. Below SLA are calls that were responded to faster than expected.**
* **X2 = 4.0047. The chi square values is small because the observed counts are very close to the expected counts.**
* **Df = 6 because in a contingency table df = (r-1) \* (c-1)**

**-> (4-1)\*(3-1) -> 3 \* 2 = 6.**

* **P-value = 0.676. Since this is higher than the significance level, we will fail to reject the null hypothesis. The observed data does not provide enough evidence to assume an association.**

**Data Visualization:**

**A screenshot of a computer code

Description automatically generated**

**A graph with different colored bars

Description automatically generated**

* **This clustered bar chart shows the overall distribution of how each center is performing.**
* **Notice that ‘within SLA’ has the highest counts, which is a good sign for the company.**

**A computer code with text

Description automatically generated**

**A graph with different colored bars

Description automatically generated**

* **I added this stacked column chart as well because it illustrates the response time with the overall call volume in mind.**
* **Notice that Los Angles may have the highest count of above SLA calls, but keep in mind that it did receive the highest amount of total calls.**