

**STATISTICAL FOUNDATION FOR DATA ANALYTICS CSHO531DAP**

**Record of Laboratory Experiments/ Programs**



Submitted in Partial fulfillment of Laboratory Internal Assessment

for the course

STATISTICAL FOUNDATION FOR DATA ANALYTICS CSHO531DAP

in the department of Computer Science and Engineering

**October 23, 2022**

**Christopher samuel harris 2162019**

**Department of Computer Science and Engineering**

**CHRIST (Deemed to be University)**

**Kengeri Campus, Mysore Road,**

**Bangalore 560074, Karnataka, India**

**CHRISTOPHER SAMUEL HARRIS 2162019**

**Experiment 1:**

**Problem Statement:**

Introduction to IBMs SPSS (Statistical Product and Service Solutions) tool for Statistical Analysis.

**Introduction:**

SPSS is a Windows based program that can be used to perform data entry and analysis and to create tables and graphs. SPSS is capable of handling large amounts of data and can perform all of the analyses and much more.

Open SPSS on your computer, something that looks similar to the following screenshot is observed in Figure 1.

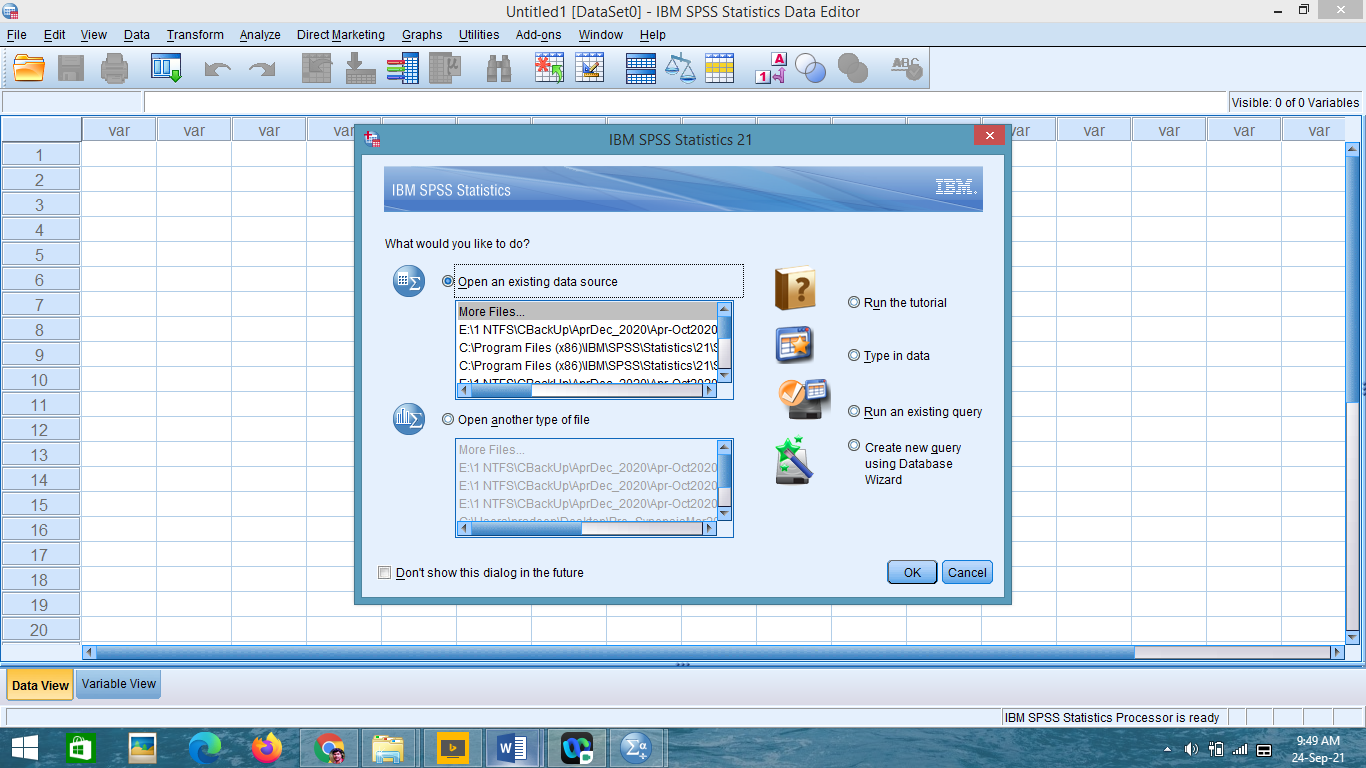


Figure 1: Basic layout of SPSS tool

SPSS automatically assumes that you want to open an existing file, and immediately opens a dialogue box to ask which file you’d like to open. It’ll make it easier to navigate the interface and windows in SPSS if we open a file.

A demo.sav file is selected to open an existing dataset which can be seen in Figure 2.

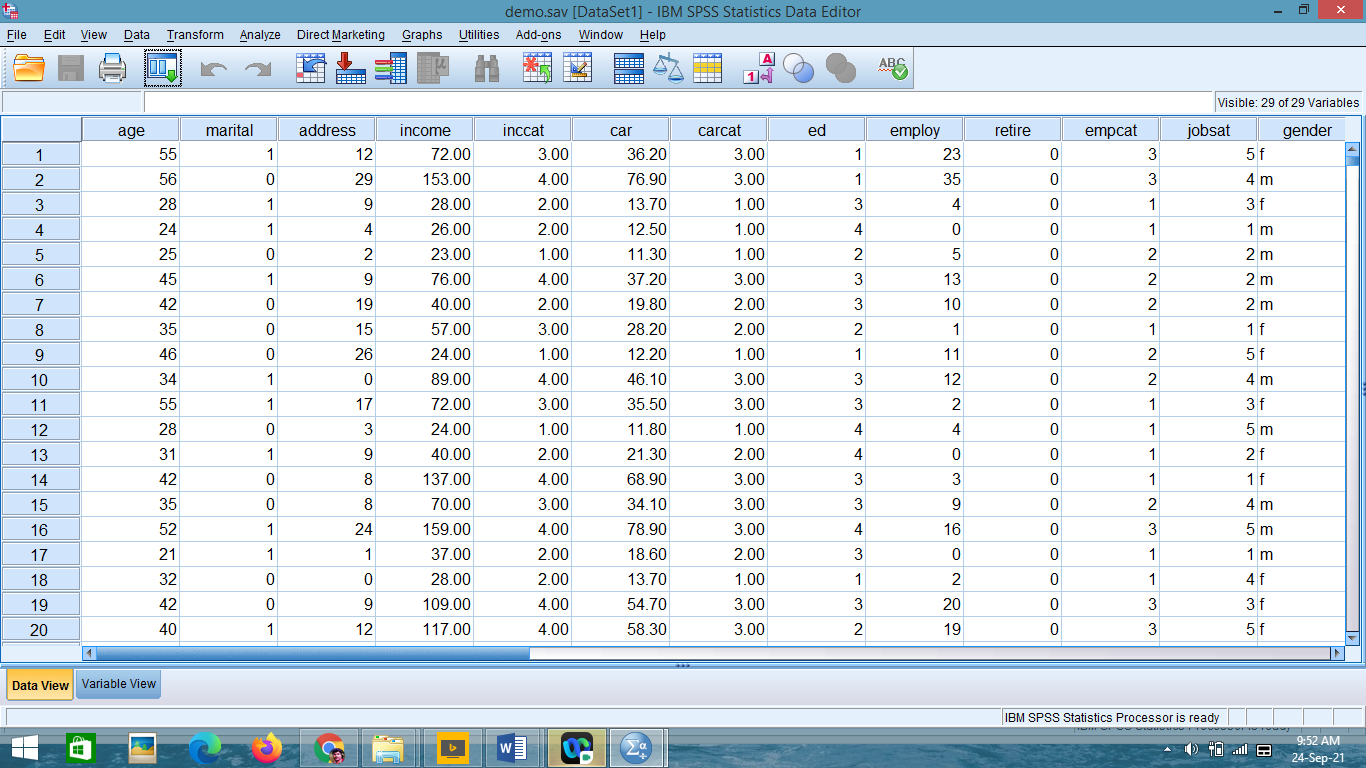


Figure 2: Demo.sav SPSS dataset/file

**SPSS Menus and Icons**

Review the options listed under each menu on the Menu Bar by clicking them one at a time. Follow along with the below descriptions.

**File** includes all of the options typically used in other programs, such as open, save, and exit. Notice, that you can open or create new files of multiple types as illustrated to the right.

**Edit** includes the typical cut, copy, and paste commands, and allows you to specify various options for displaying data and output.

Clicking on **Options**, you will see the dialog box to the left. This can be used to format the data, output, charts, etc.

**View** allows you to select which toolbars you want to show, select font size, add or remove the gridlines that separate each piece of data, and to select whether or not to display your raw data or the data labels.

**Data** allows you to select several options ranging from displaying data that is sorted by a specific variable to selecting certain cases for subsequent analyses.

**Transform** includes several options to change current variables. For example, you can change continuous variables to categorical variables, change scores into rank scores, add a constant to variables, etc.

**Analyze** includes all of the commands to carry out statistical analyses and to calculate descriptive statistics.

**Graphs** includes the commands to create various types of graphs including box plots, histograms, line graphs, and bar charts.

**Utilities** allows you to list file information which is a list of all variables, their labels, values, locations in the data file, and type.

**Add-ons** are programs that can be added to the base SPSS package. Window can be used to select which window you want to view (i.e., Data Editor, Output Viewer, or Syntax).

**Help** has many useful options including a link to the SPSS homepage, a statistics coach, and a syntax guide.

Using **topics,** you can use the index option to type in any key word and get a list of options, or you can view the categories and subcategories available under contents. This is an excellent tool and can be used to troubleshoot most problems.

The **Icons** directly under the Menu bar provide shortcuts to many common commands that are available in specific menus.

**Conclusion:**

The fundamental aspects and introduction of the statistical tool SPSS are briefly explained.

**CHRISTOPHER SAMUEL HARRIS 2162019**

**Experiment 2: (Part A)**

**Problem Statement:**

Define variables with data types and related parameters for a specific data set.

**Describing Dataset (with list of variables and data types):**

The data set is the dataset of cars in the game Forza Horizon 5.

The link to the dataset is given below:

<https://www.kaggle.com/datasets/deepcontractor/froza-horizon-5-cars-dataset>

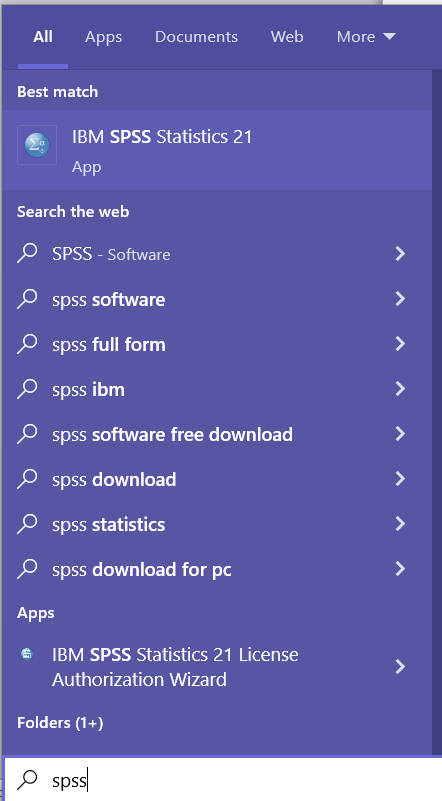
The list of variables and datatypes are as follows ( shortened):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Name** | **Type** | **Values** | **Measure** | **Role** |
| Name\_Model | String | None | Nominal | Input |
| Model\_Type | String | None | Nominal | Input |
| Price | Numeric | None | Scale | Input |
| Car\_Source | String | None | Nominal | Input |
| Stock\_Specs | String | --- | Ordinal | Input |
| PI | Numeric | None | Ordinal | Input |
| Drive\_Type | String | --- | Nominal | Input |
| Speed | Numeric | None | Scale | Input |
| Handling | Numeric | None | Scale | Input |
| Acceleration | Numeric | None | Scale | Input |
| Launch | Numeric | None | Scale | Input |
| Braking | Numeric | None | Scale | Input |
| Offroad | Numeric | None | Scale | Input |
| Horsepower | Numeric | None | Scale | Input |
| Weight\_lbs | Numeric | None | Scale | Input |

**Steps of Execution**

**Step 1:**

Open SPSS



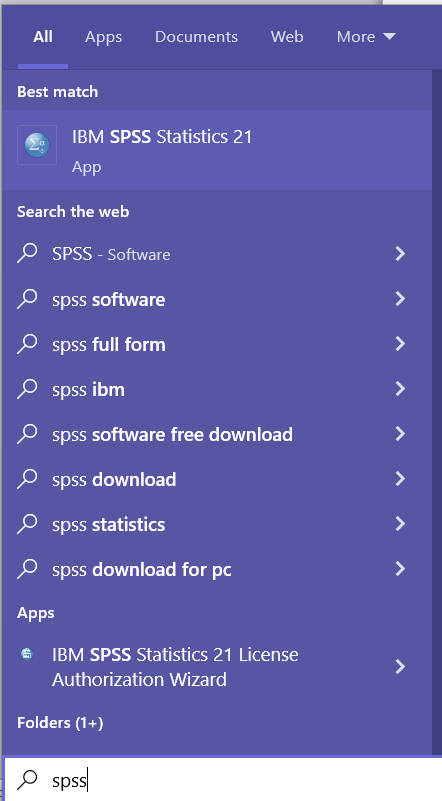
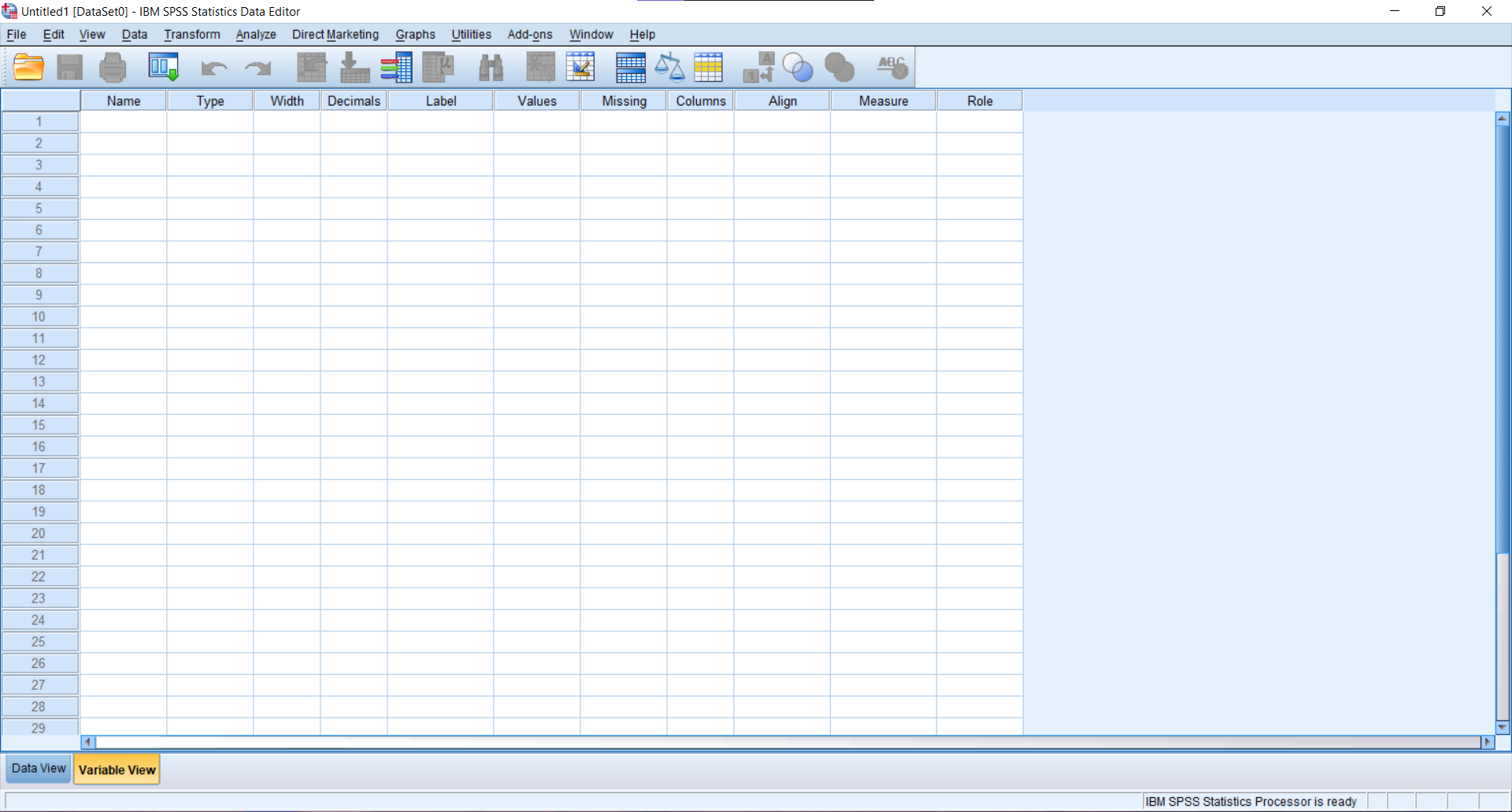


Figure 3: Searching for SPSS app

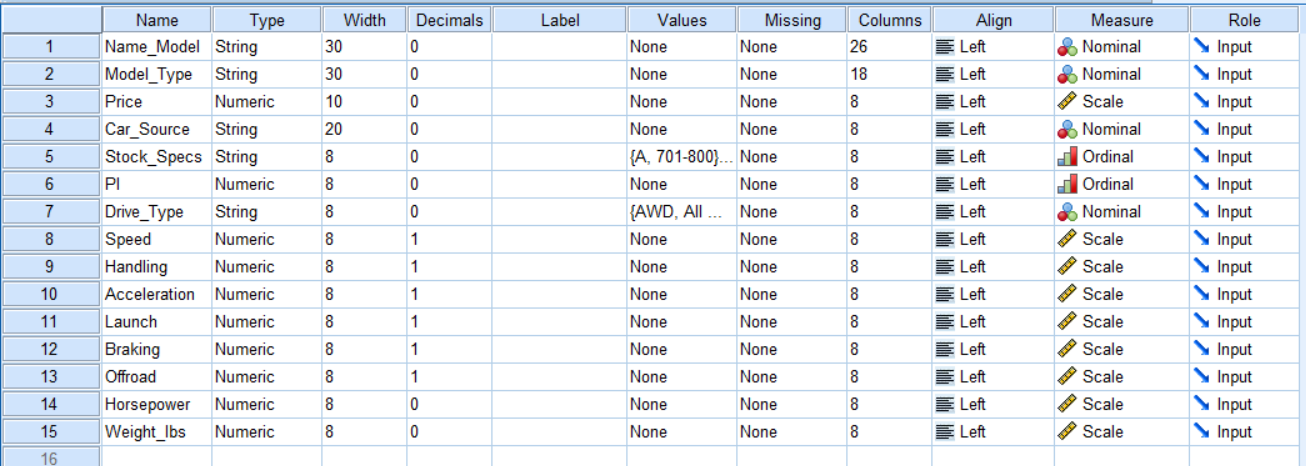
**Step 2:**

Enter variable view for the data set



**Step 3:**

Enter all the variables and their details.



**Conclusion:**

Variables and datatypes were defined and declared in the SPSS editor for a selected specific dataset.

**CHRISTOPHER SAMUEL HARRIS 2162019**

**Experiment 2: (Part B)**

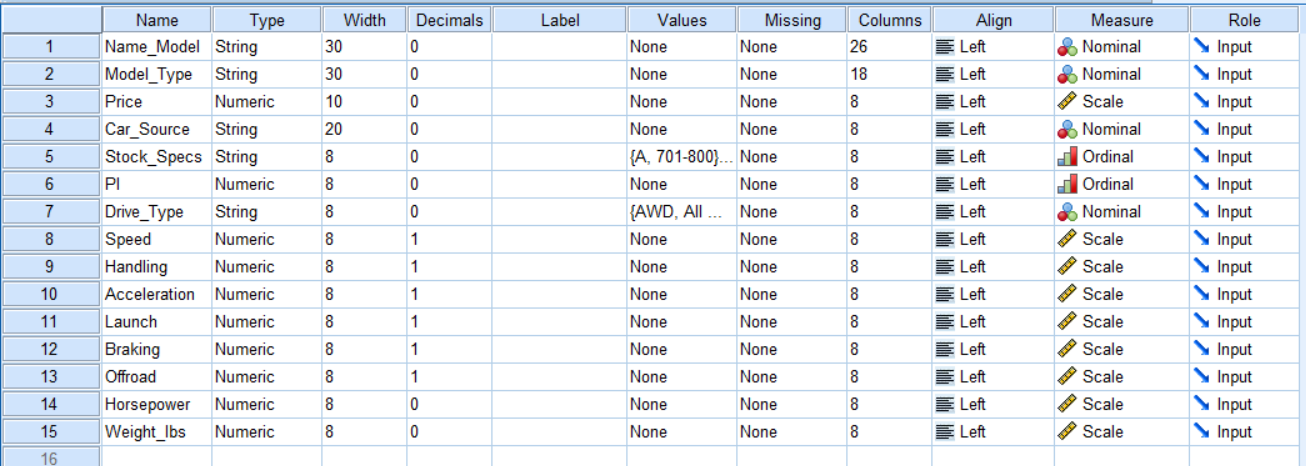
**Problem Statement:**

Data set Variable view & Data view

**Steps of Execution:**

**Step 1:**

Enter all the variables necessary



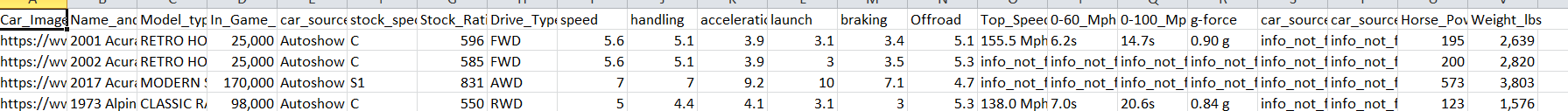
**Step 2:**

Open the dataset

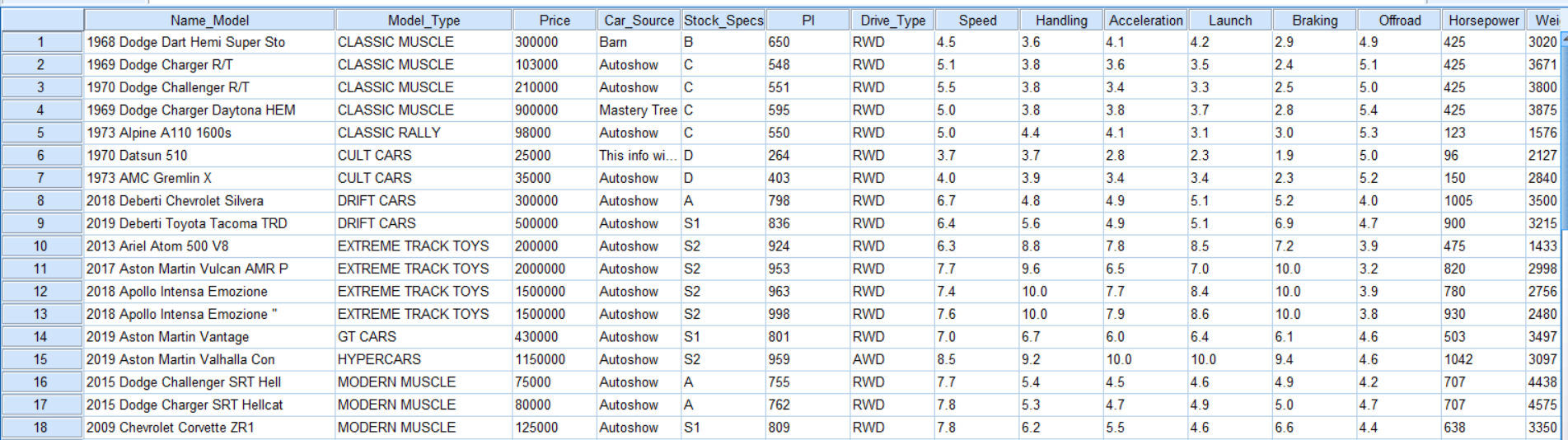


**Step 3:**

Copy and paste the relevant columns into the data view







**Conclusion:**

The variables and data have been entered into the dataset.

**CHRISTOPHER SAMUEL HARRIS 2162019**

**Experiment 2: (Part C)**

**Problem Statement:**

Sort Variables, Transpose, Aggregate, Split file & Select cases

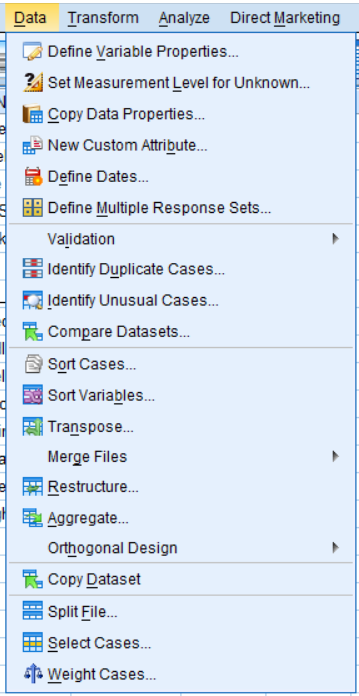
**Steps of Execution:**

**Sort Variables:**

Sort Variables is usually used to group the variables into an order to make items easier to search for.

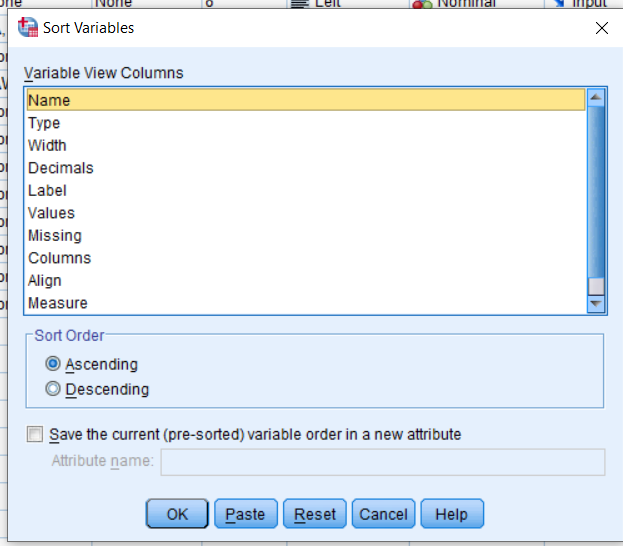
**Step 1:**

Go to the data tab, and select sort variables.



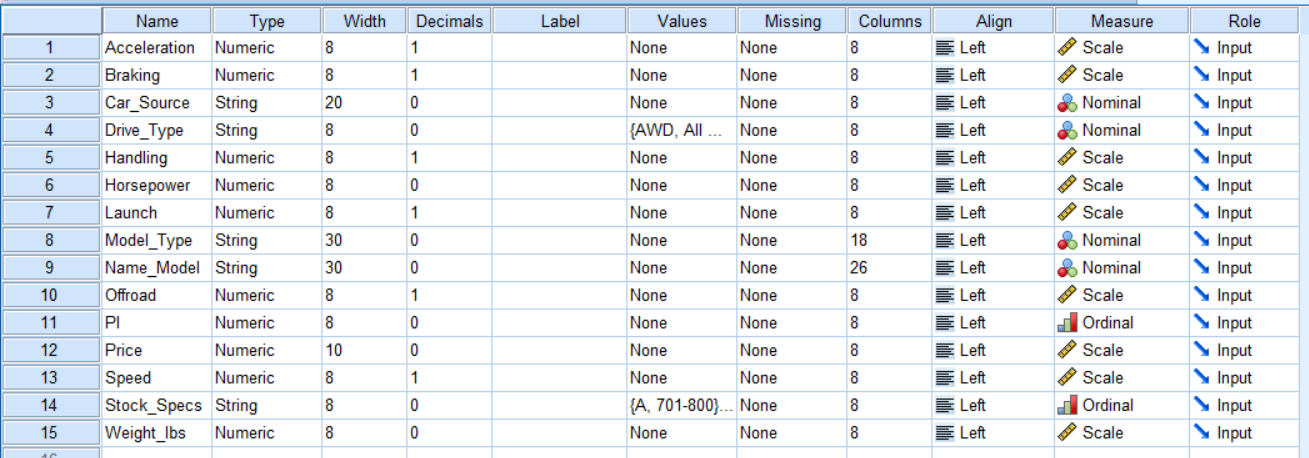
**Step 2:**

Select the sort order



**Step 3:**

Output of the sort variables.

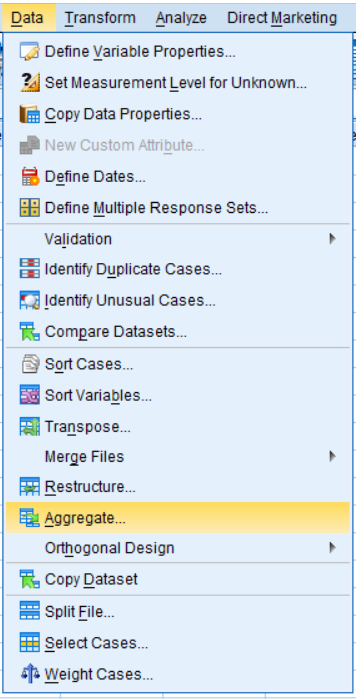


**Aggregate:**

Aggregate is generally used to get an easier idea on the averages of variables based on certain groupings.

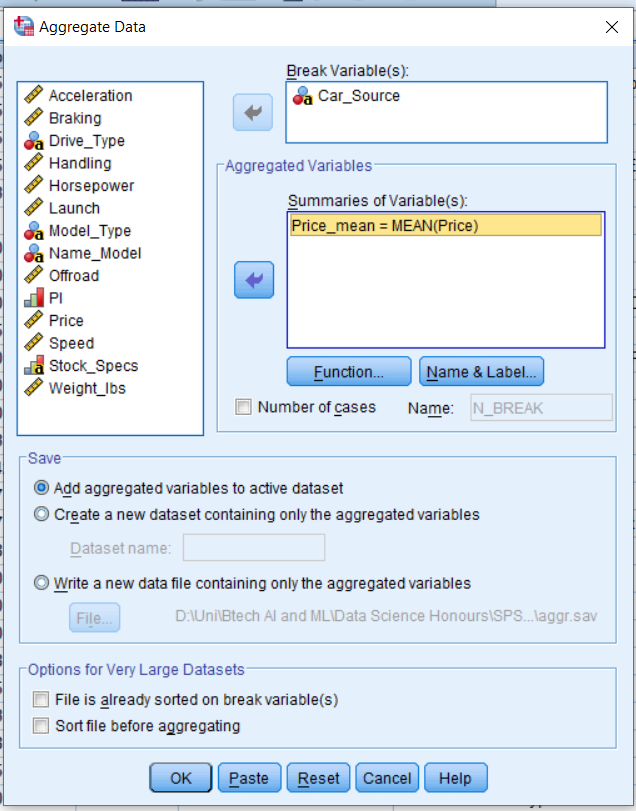
**Step 1:**

Aggregate can be found under the data tab.



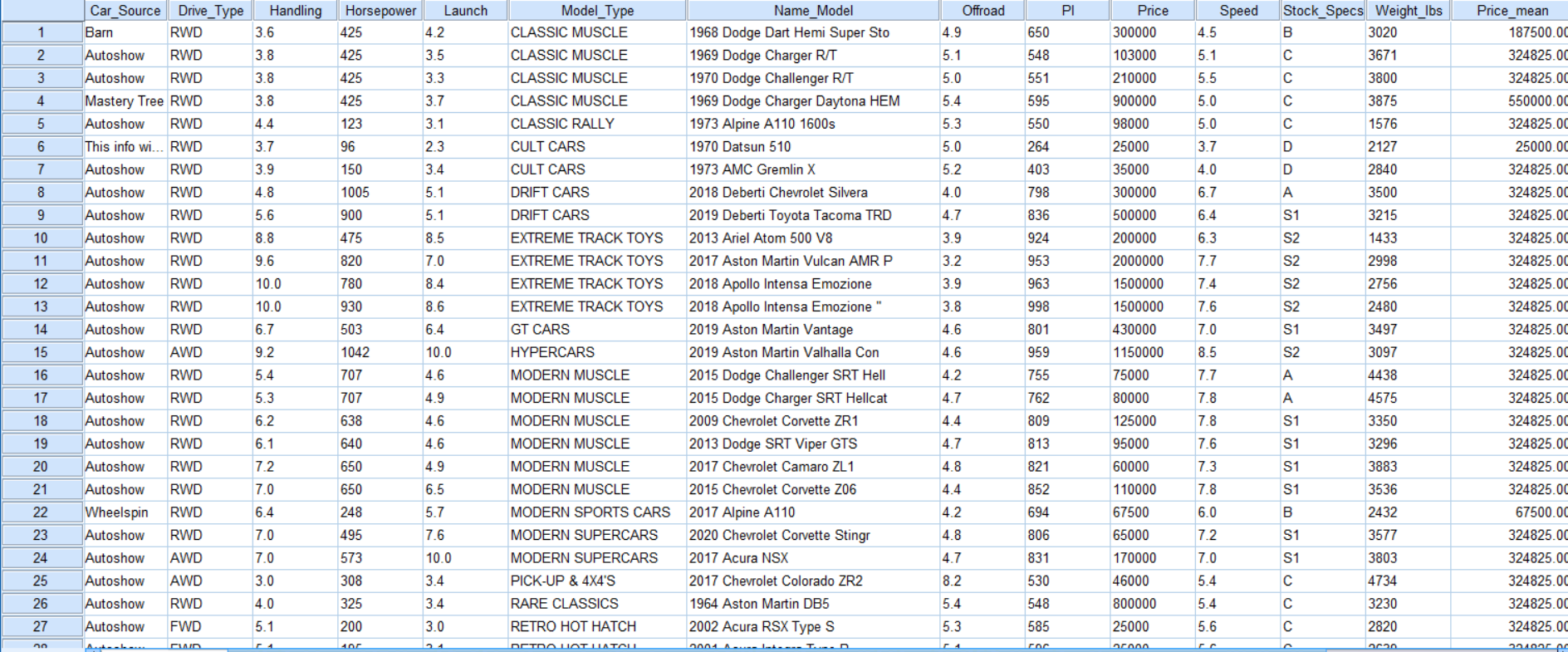
**Step 2:**

Break variables is the category to divide the aggregate on. Summaries of variables marks the variables to take the sum of.



**Step 3:**

Output.

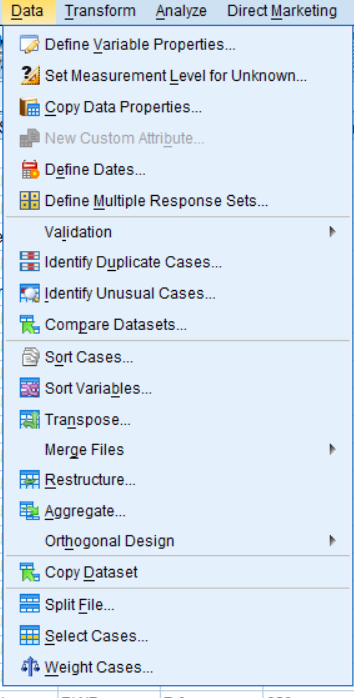


**Split File:**

Split file is another way of grouping data into different categories.

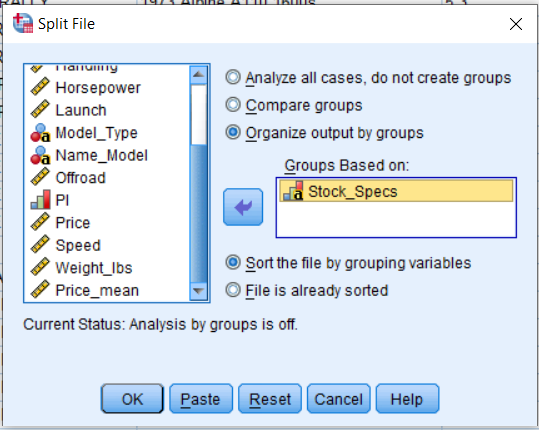
**Step 1:**

The option can be found under the data tab.



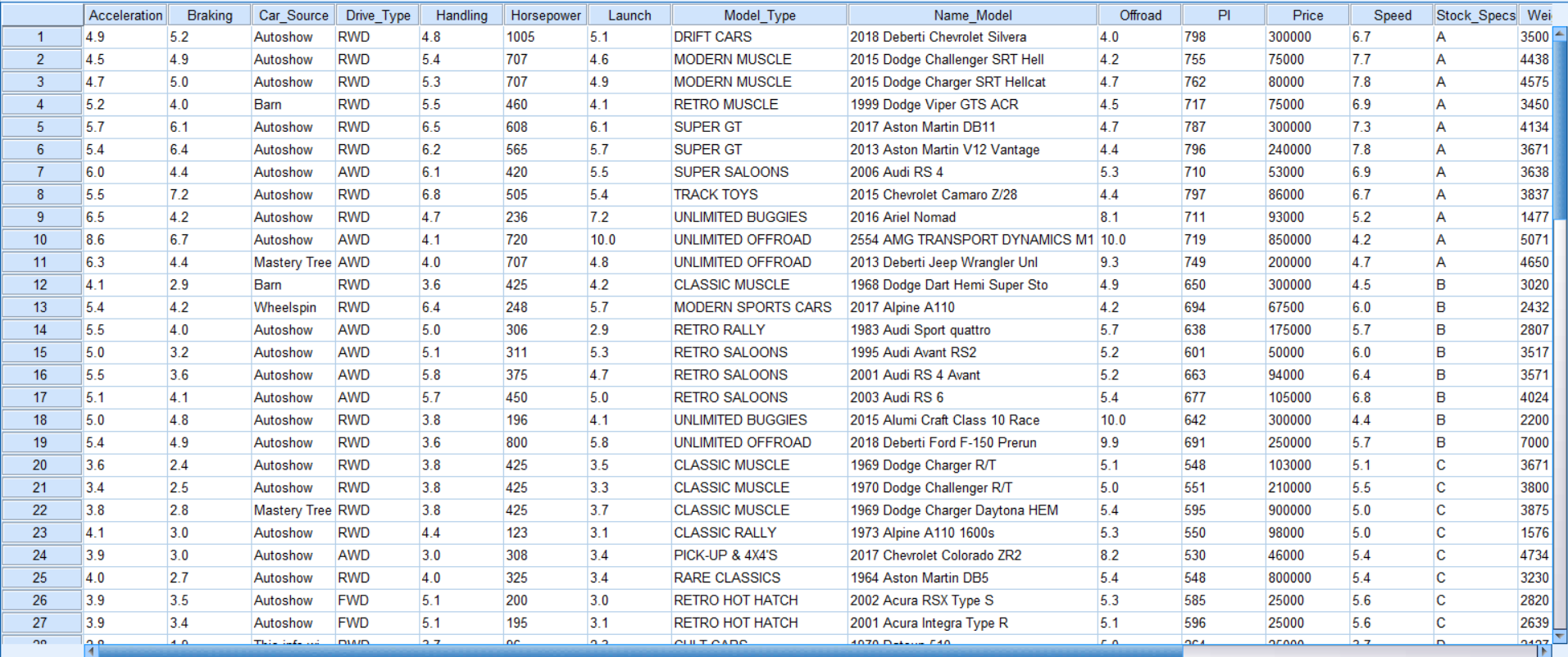
**Step 2:**

Select the variable to group the data by.



**Step 3:**

Output.

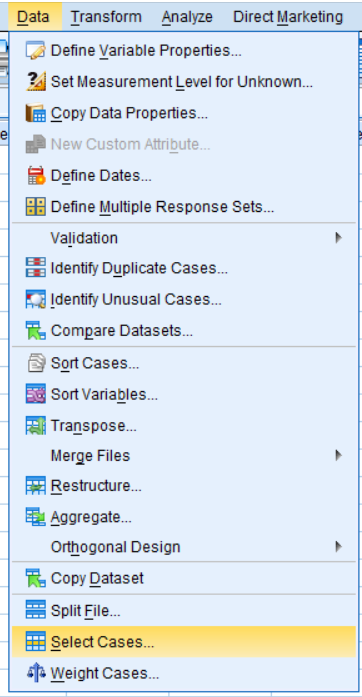


**Select Cases:**

Select cases is another strategy that is technically grouping, but it eliminates items based on pre-determined values.

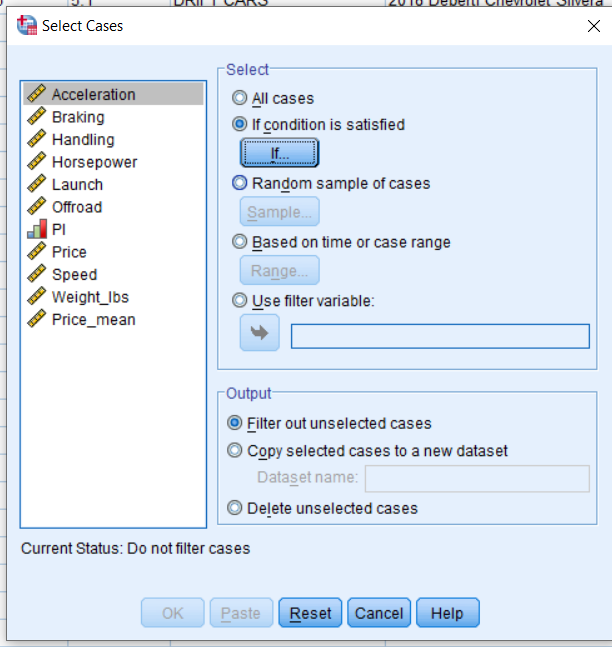
**Step 1:**

Select cases can be found under Data.



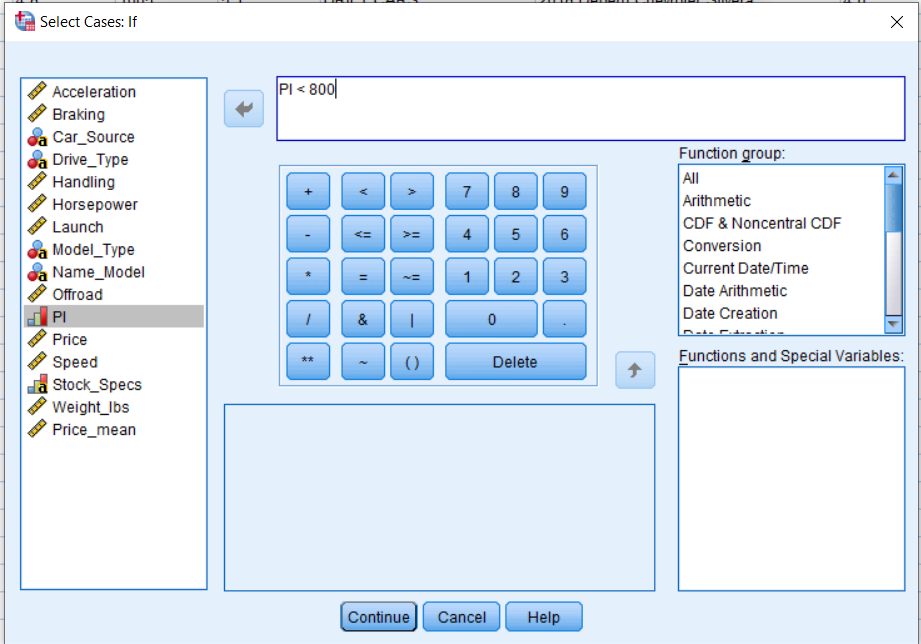
**Step 2:**

Filters can be decided depending on what is required.



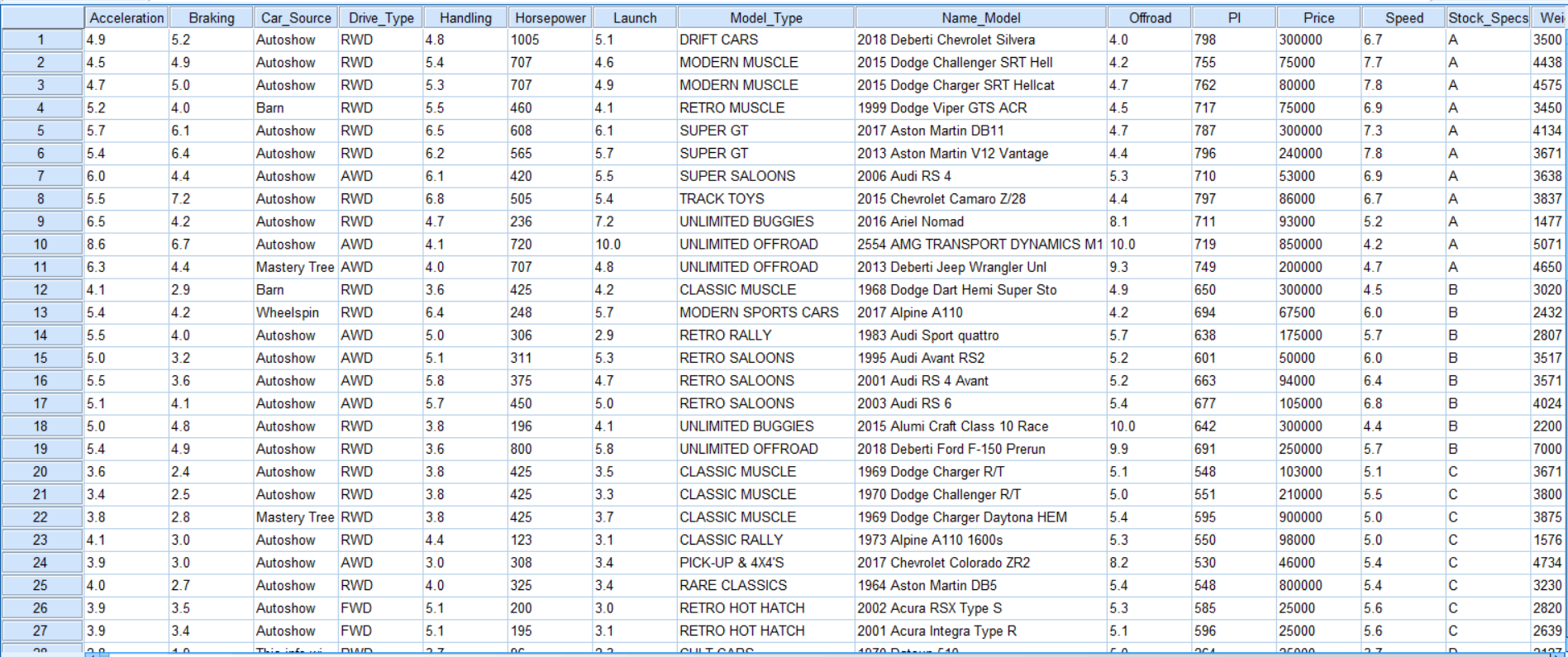
**Step 3:**

For this condition, we are filtering out cases that have PI greater than 800.



**Step 4:**

Output.



**Conclusion:**

Sort Variables, Transpose, Aggregate, Split file & Select cases were executed in SPSS successfully.

**CHRISTOPHER SAMUEL HARRIS 2162019**

**Experiment 3:**

**Problem Statement:**

Introduction to Null Hypothesis

**The Null Hypothesis:**

The null hypothesis is a characteristic arithmetic theory, suggesting that no statistical relationship and significance exists in a set of given, single, observed variables between two sets of observed data and measured phenomena. Sometimes referred to as the ‘null’, it is represented as H0.

Commonly with analysis or testing, it is assumed there exists no significance or effect unless there is enough evidence to suggest otherwise, similar to how a trial presumes innocence.

When the sample contains sufficient evidence, the null can be rejected, and the effect be deemed statistically significant.

**Conclusion:**

The null hypothesis is elaborated upon, and can thus be applied in various situations, using different tests.

**CHRISTOPHER SAMUEL HARRIS 2162019**

**Experiment 4:**

**Problem Statement:**

Basic operations on Statistical parameters

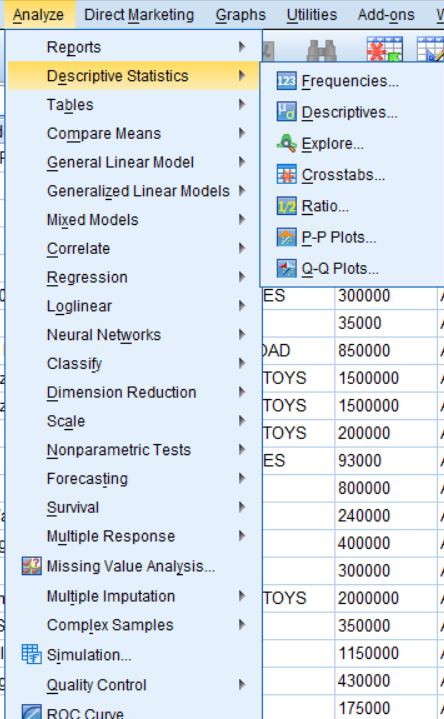
A. Mean, Median, Mode, Standard Deviation and Variance

B. Frequencies, Descriptives [Skewness & Kurtosis; Histogram with Normal curve]

**Steps of Execution:**

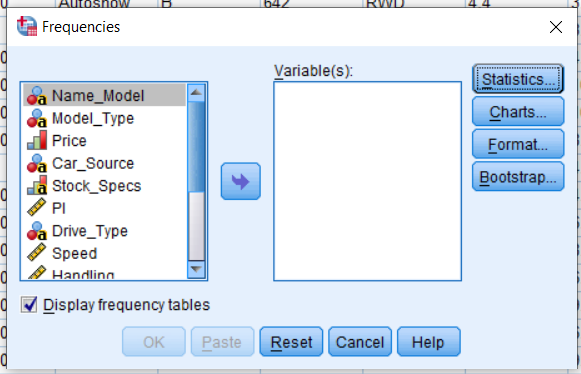
**Step 1:**

Most of the parameters can be found under Analyze->Descriptive Statistics->Frequencies



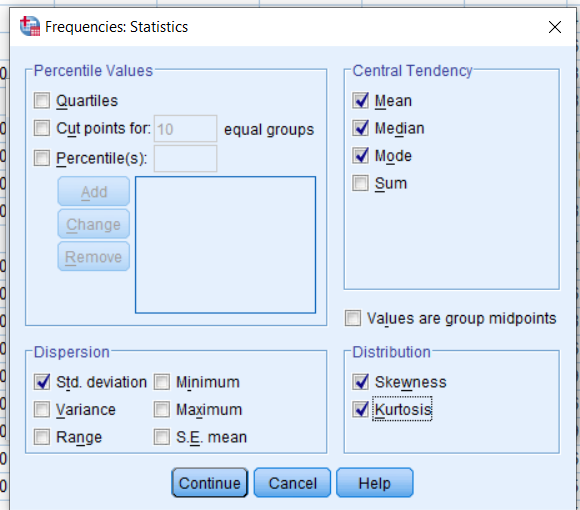
**Step 2:**

Under this, you can select which variable to check all the parameters of.



**Step 3:**

Under Statistics, you can select what items to view.



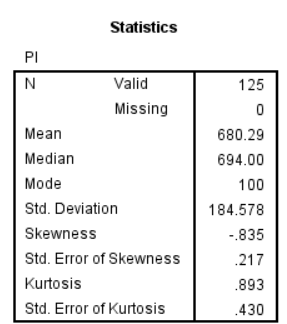
**Step 4:**

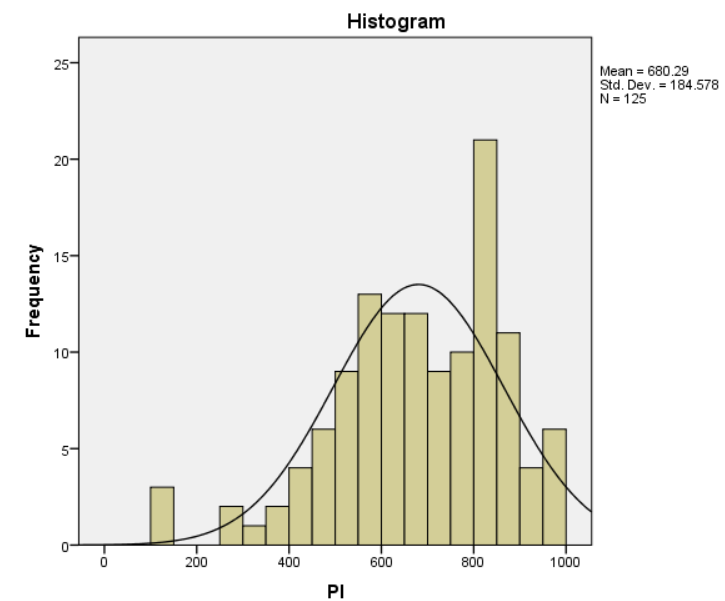
Under charts, a histogram can be applied.



**Step 5:**

Output, 2 Tables.





**Conclusion:**

Mean, Median, Mode, Variance, Standard Deviation, Skewness and Kurtosis along with a histogram is successfully displayed.

**CHRISTOPHER SAMUEL HARRIS 2162019**

**Experiment 5**

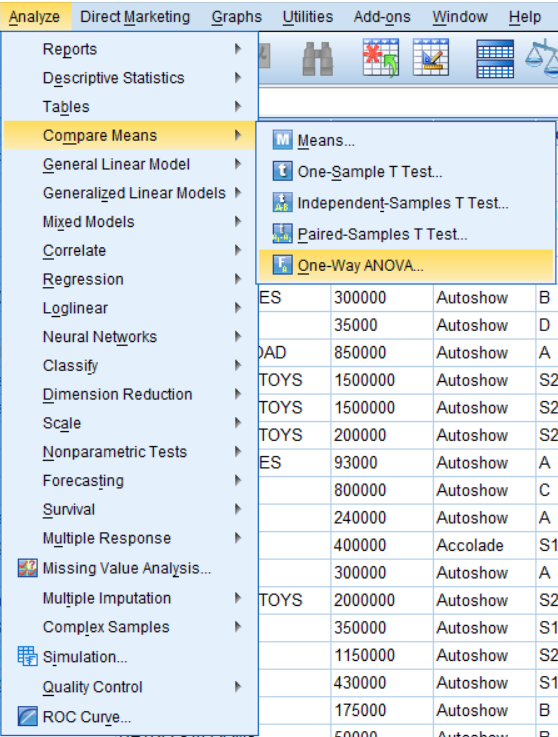
**Problem Statement:**

Statistical tests and their significance value

A. One Way ANOVA

B. Independent Sample T-Test or Chi- Square Test

One Way ANOVA and the Independent Sample T-Test can be found under Analyze->Compare Means



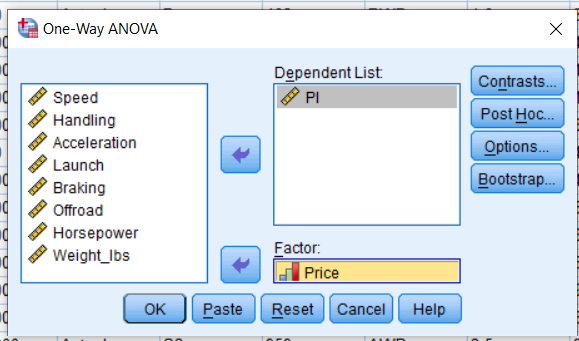
**Steps of Execution:**

**One Way ANOVA**

The One Way ANOVA test is a simple test to find the significance of one variable in comparison to another.

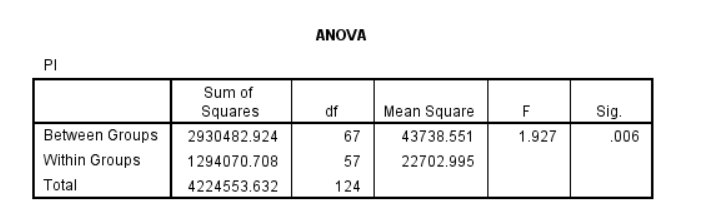
**Step 1:**

Two variables can be inputted to check significance values.



**Step 2:**

Output.



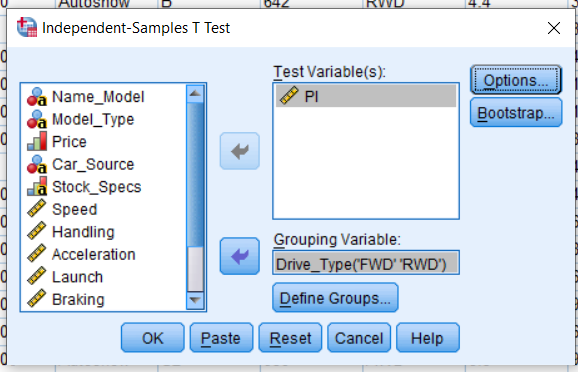
There is a significant relation between PI and Price in the dataset.

**Independent Sample T-Test:**

This test is another test that can be used to find the significance between two variables.

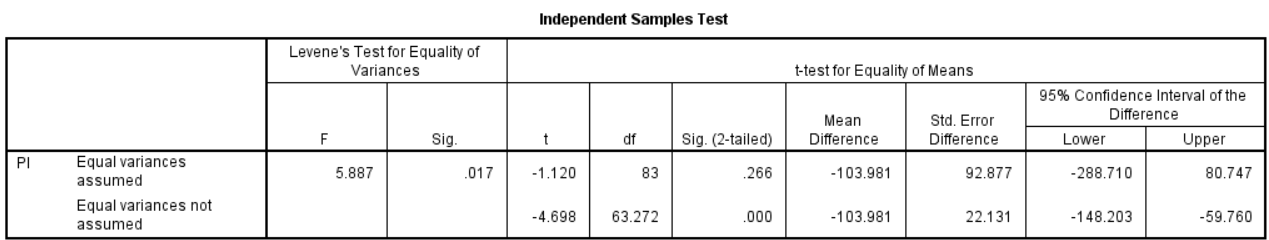
**Step 1:**

Here we have a variable that can be organized based on two values from another variable.



**Step 2:**

Output.

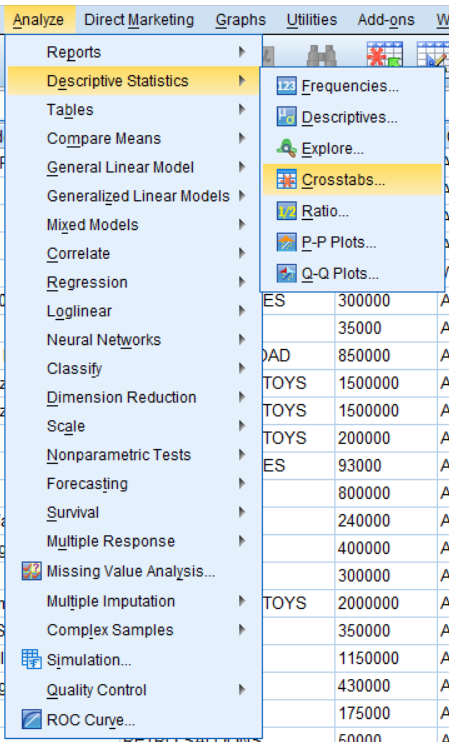


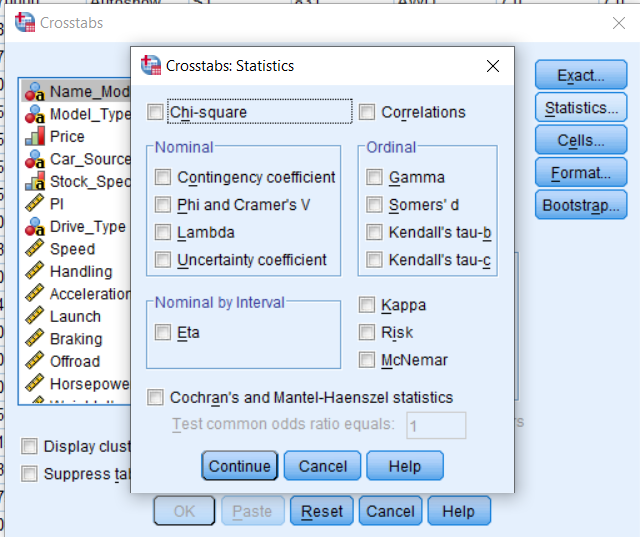
**Chi-Square Test:**

The Chi-Square Test is a simple test that can be used to test the significance between two variables

**Step 1:**

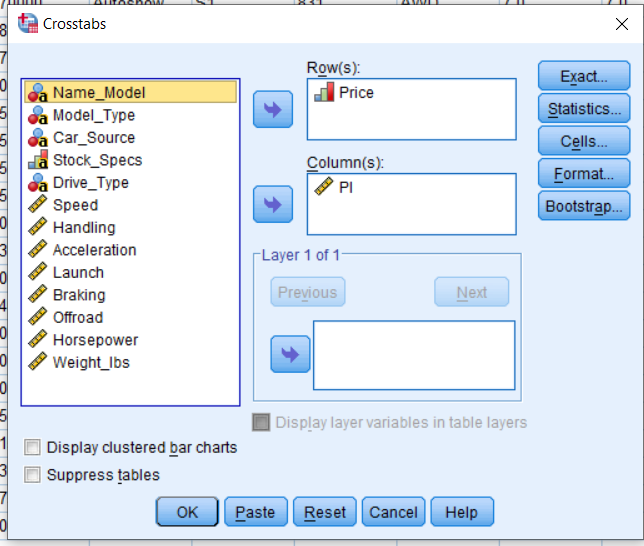
The Chi-Square Test can be found under Analyze->Descriptive Statistics->Crosstabs->Statistics





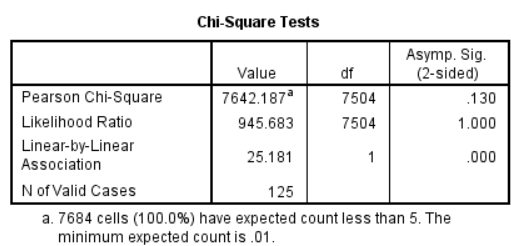
**Step 2:**

Two variables can be selected to find the significance between.



**Step 3:**

Output.



**Conclusion:**

The One Way ANOVA, Independent Sample T-Test and Chi- Square Test were executed successfully.