# Exploratory Data Analysis Report on Tire Data

The dataset consists of 1,689 rows (observations) and 26 columns (features), 3 of which are numeric.

## Overview of Numeric Features

|  | count | avg | stddev | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| **Other Services** | 1689 | 2.5098 | 1.7491 | -1 | 1 | 2 | 4 | 10 |
| **Number of purchases made in the store** | 1689 | 3.9911 | 2.093 | 0 | 2 | 4 | 5 | 14 |
| **Number of online purchases** | 1689 | 3.7975 | 2.1125 | 0 | 2 | 4 | 5 | 13 |

## Overview of Categorical Features

|  | count | unique | top | freq | relative freq |
|---|---|---|---|---|---|
| **Cust. Neighbourhood** | 1689 | 7 | Kelowna | 705 | 41.74% |
| **Rough/Approx km per year** | 1689 | 446 | 16,400 | 16 | 0.95% |
| **Cust. Education** | 1689 | 5 | College | 608 | 36.00% |
| **Household size** | 1689 | 6 | 2 persons | 707 | 41.86% |
| **Number of children 18 years or younger in the home** | 1689 | 5 | 0 | 1137 | 67.32% |
| **#adult drivers** | 1689 | 6 | 1 | 906 | 53.64% |
| **Number of drivers in home** | 1689 | 6 | 1 | 876 | 51.87% |

| | | | | | |
|---|---|---|---|---|---|
| **Household income** | 1627 | 7 | 32000 – 55000 | 366 | 21.67% |
| **Number of vehicles in household** | 1689 | 5 | 1 | 987 | 58.44% |
| **Primary Vehicle** | 1635 | 11 | SUV | 297 | 17.58% |
| **2nd Vehicle** | 649 | 11 | Sedan | 180 | 10.66% |
| **3rd Vehicle** | 195 | 9 | Sedan | 58 | 3.43% |
| **4th Vehicle** | 23 | 4 | Van | 8 | 0.47% |
| **Wheel/rim purchase** | 1689 | 5 | 1 | 430 | 25.46% |
| **Highest quality of rim purchased in the last 5 years** | 1397 | 3 | Aluminum | 515 | 30.49% |
| **warranty extra: purchased (y/n)** | 1689 | 2 | n | 1018 | 60.27% |
| **Discount code obtained** | 1675 | 3 | Online | 888 | 52.58% |
| **Discount code** | 1686 | 4 | UQX83492b | 457 | 27.06% |
| **Responded March Promo (summer tires)** | 120 | 1 | Y | 120 | 7.10% |
| **Responded March promo (all seasons)** | 190 | 1 | Y | 190 | 11.25% |
| **Responded August Promo (all-season tires)** | 688 | 1 | Y | 688 | 40.73% |
| **Responded October Promo (winter tires)** | 237 | 1 | Y | 237 | 14.03% |
| **Responded November promo (winter tires)** | 549 | 1 | Y | 549 | 32.50% |

# 1. Univariate Analysis

## 1.1 Cust. Neighbourhood

Cust. neighbourhood is a categorical variable with 7 unique values. None of its values are missing.

### Summary Statistics

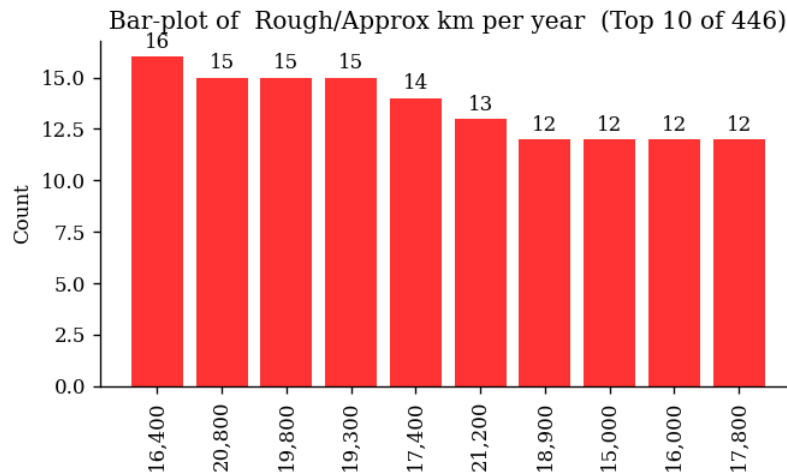| | |
|---|---|
| Mode (Most frequent) | Kelowna |
| Maximum frequency | 705 |
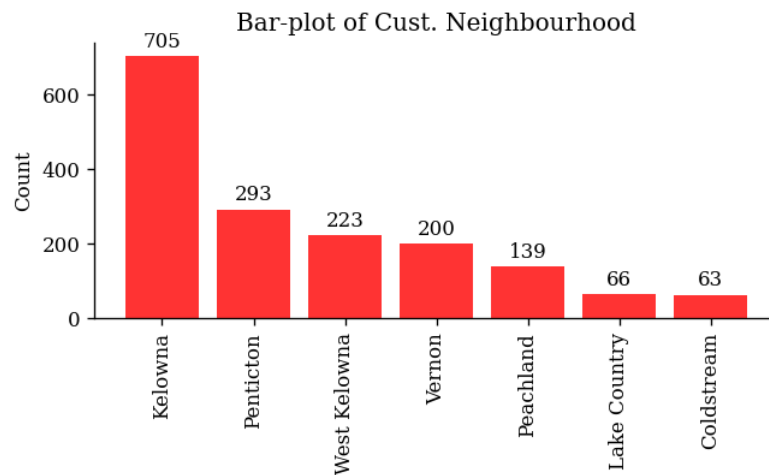


## 1.2  Rough/Approx Km Per Year

rough/approx km per year  is a categorical variable with 446 unique values. None of its values are missing.

### Summary Statistics

| | |
|---|---|
| Mode (Most frequent) | 16,400 |
| Maximum frequency | 16 |

Bar-plot of Rough/Approx km per year (Top 10 of 446)



## 1.3 Cust. Education

Cust. education is a categorical variable with 5 unique values. None of its values are missing.

*Summary Statistics*

| | |
|---|---|
| Mode (Most frequent) | College |
| Maximum frequency | 608 |
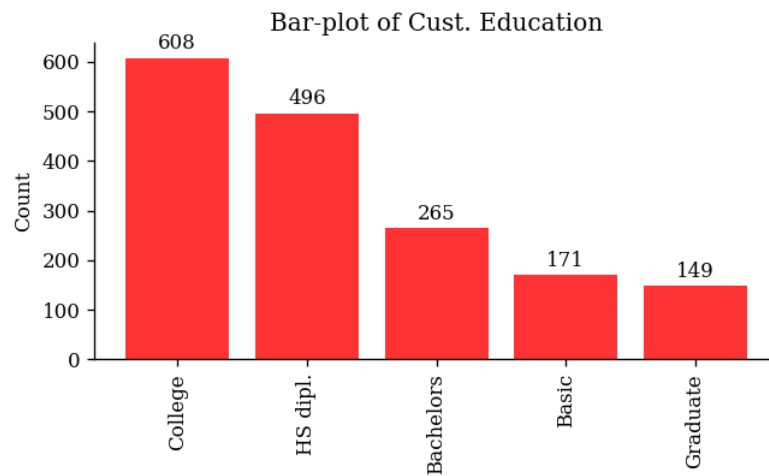


## 1.4 Household Size

Household size is a categorical variable with 6 unique values. None of its values are missing.

*Summary Statistics*

| | |
|---|---|
| Mode (Most frequent) | 2 persons |
| Maximum frequency | 707 |

Bar-plot of Household size
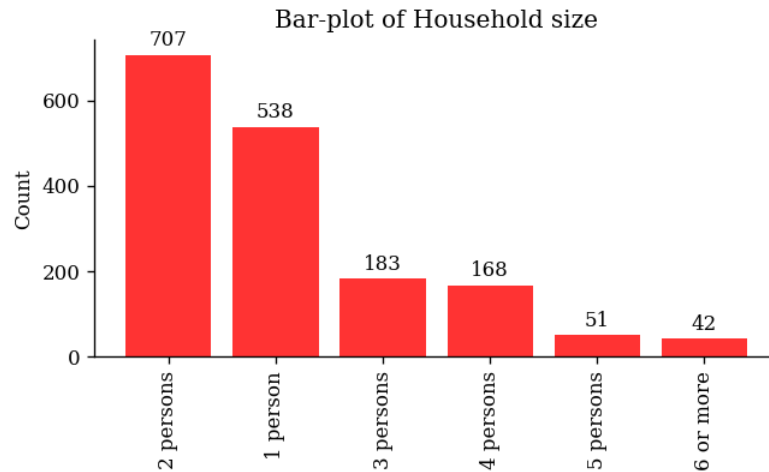
## 1.5 Number Of Children 18 Years Or Younger In The Home

Number of children 18 years or younger in the home is a numeric (<=10 levels) variable with 5 unique values. None of its values are missing.

### Summary Statistics

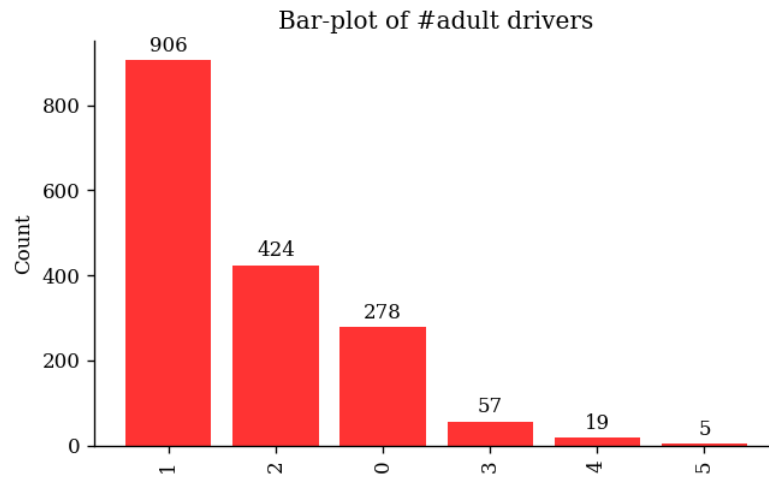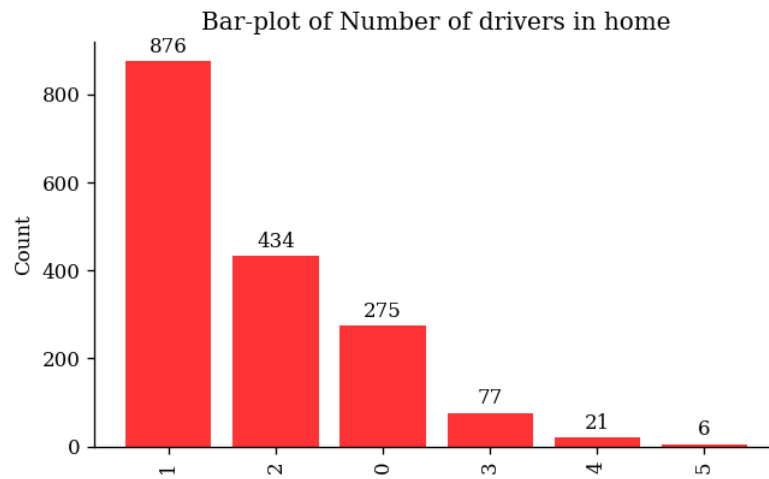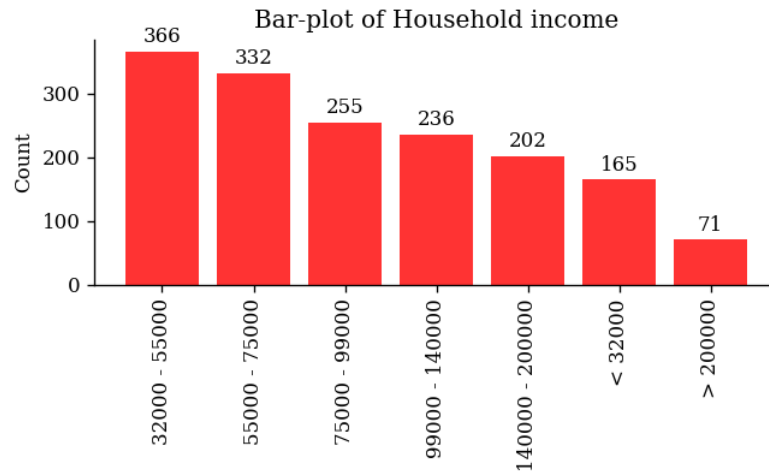| Mode (Most frequent) | 0 |
|---|---|
| Maximum frequency | 1137 |



Bar-plot of Number of children 18 years or younger in the ho...
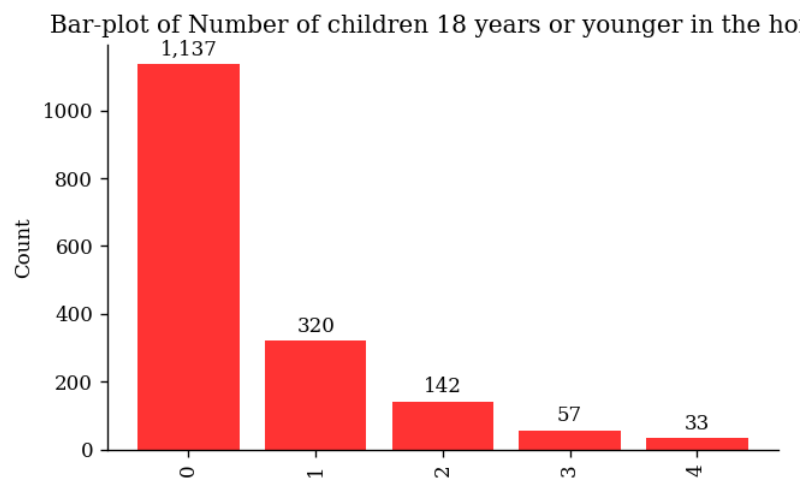
## 1.6 #Adult Drivers

#adult drivers is a numeric (<=10 levels) variable with 6 unique values. None of its values are missing.

### Summary Statistics

| Mode (Most frequent) | 1 |
|---|---|
| Maximum frequency | 906 |

Bar-plot of #adult drivers

## 1.7 Number Of Drivers In Home

Number of drivers in home is a numeric (<=10 levels) variable with 6 unique values. None of its values are missing.

*Summary Statistics*

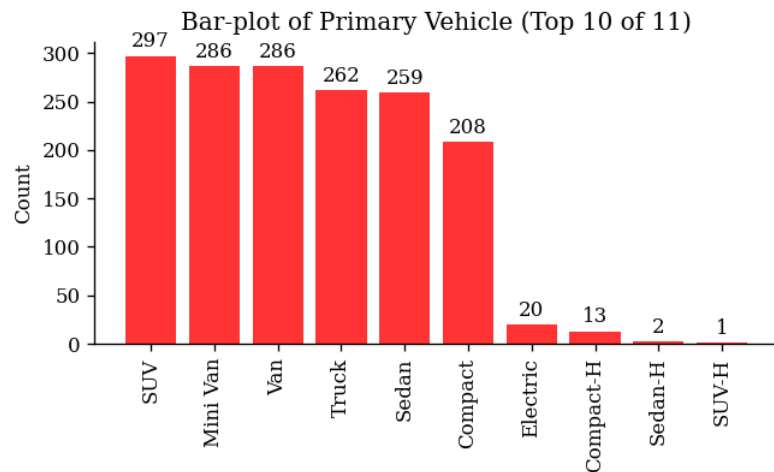| Mode (Most frequent) | 1 |
|---|---|
| Maximum frequency | 876 |



Bar-plot of Number of drivers in home

## 1.8 Household Income

Household income is a categorical variable with 7 unique values. 62 (3.67%) of its values are missing.

| Mode (Most frequent) | 32000 – 55000 |
|---|---|
| Maximum frequency | 366 |

Bar-plot of Household income



## 1.9 Number Of Vehicles In Household

Number of vehicles in household is a numeric (<=10 levels) variable with 5 unique values. None of its values are missing.

*Summary Statistics*

| Mode (Most frequent) | 1 |
|---|---|
| Maximum frequency | 987 |

Bar-plot of Number of vehicles in household
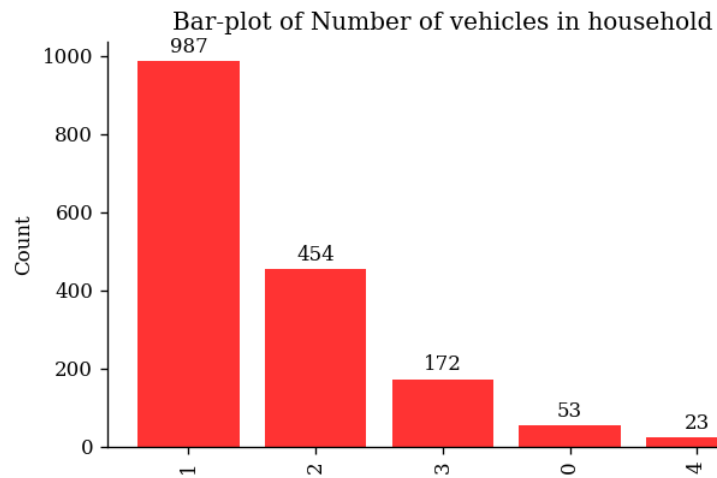
## 1.10 Primary Vehicle

Primary vehicle is a categorical variable with 11 unique values. 54 (3.20%) of its values are missing.

### *Summary Statistics*

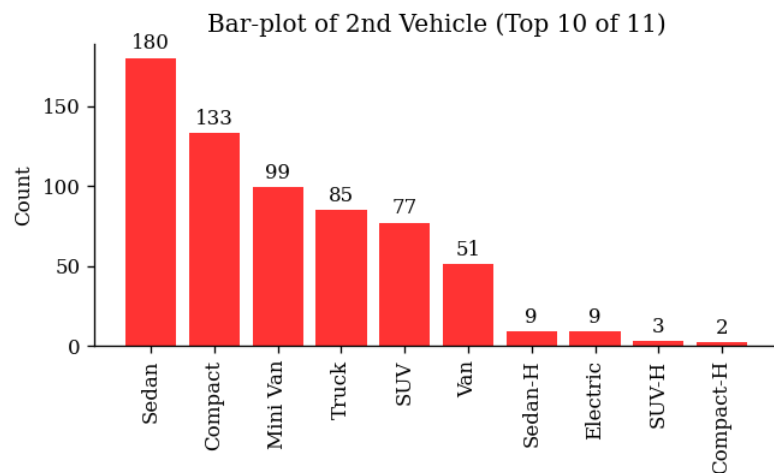| Mode (Most frequent) | SUV |
|---|---|
| Maximum frequency | 297 |

Bar-plot of Primary Vehicle (Top 10 of 11)

## 1.11 2Nd Vehicle

2nd vehicle is a categorical variable with 11 unique values. 1,040 (61.57%) of its values are missing.

### *Summary Statistics*

| Mode (Most frequent) | Sedan |
|---|---|
| Maximum frequency | 180 |

Bar-plot of 2nd Vehicle (Top 10 of 11)
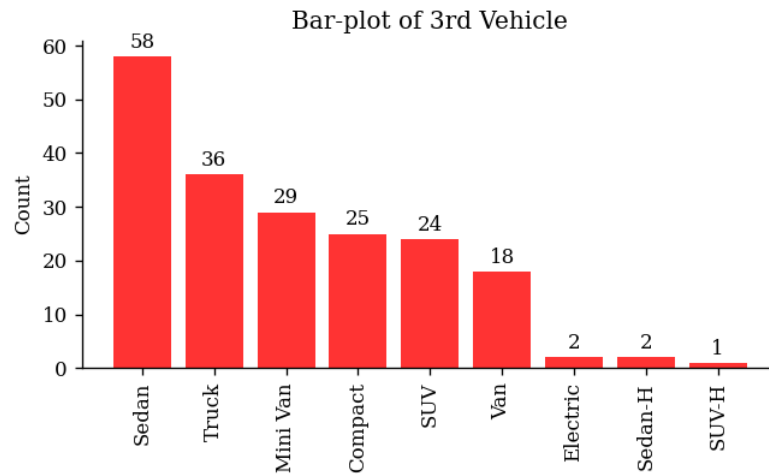
## 1.12 3Rd Vehicle

3rd vehicle is a categorical variable with 9 unique values. 1,494 (88.45%) of its values are missing.

*Summary Statistics*

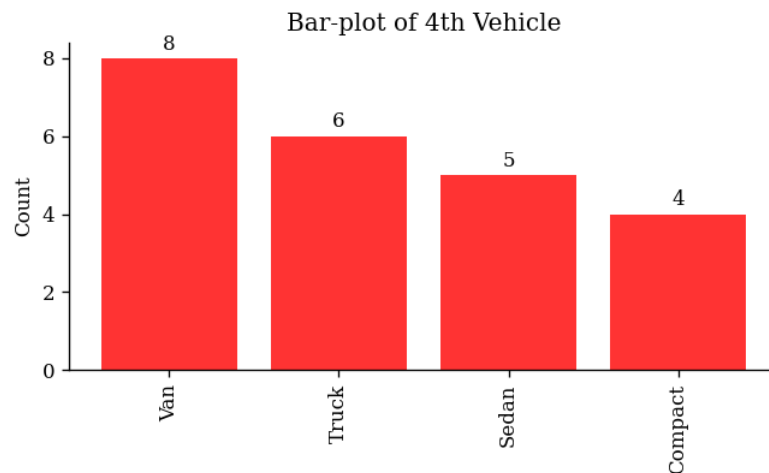| Mode (Most frequent) | Sedan |
|---|---|
| Maximum frequency | 58 |


Bar-plot of 3rd Vehicle

## 1.13 4Th Vehicle

4th vehicle is a categorical variable with 4 unique values. 1,666 (98.64%) of its values are missing.

*Summary Statistics*

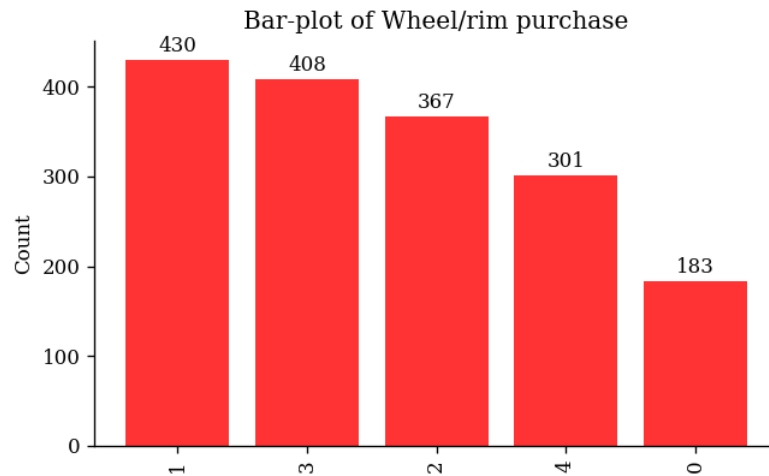| Mode (Most frequent) | Van |
|---|---|
| Maximum frequency | 8 |


Bar-plot of 4th Vehicle

## 1.14 Wheel/Rim Purchase

Wheel/rim purchase is a numeric (<=10 levels) variable with 5 unique values. None of its values are missing.

*Summary Statistics*

| Mode (Most frequent) | 1 |
|---|---|
| Maximum frequency | 430 |

Bar-plot of Wheel/rim purchase

| Value | Count |
|---|---|
| 1 | 430 |
| 3 | 408 |
| 2 | 367 |
| 4 | 301 |
| 0 | 183 |

## 1.15 Highest Quality Of Rim Purchased In The Last 5 Years

Highest quality of rim purchased in the last 5 years is a categorical variable with 3 unique values. 292 (17.29%) of its values are missing.

*Summary Statistics*

| Mode (Most frequent) | Aluminum |
|---|---|
| Maximum frequency | 515 |

Bar-plot of Highest quality of rim purchased in the last 5 year

| Value | Count |
|---|---|
| Aluminum | 515 |
| Specialised | 479 |
| Steel | 403 |
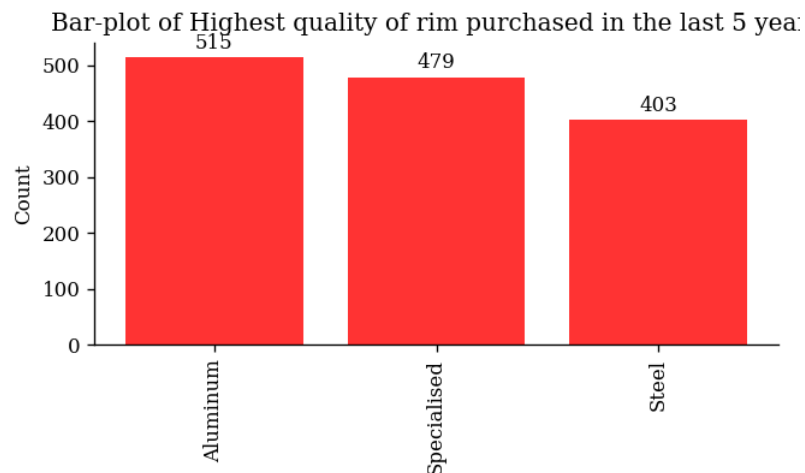
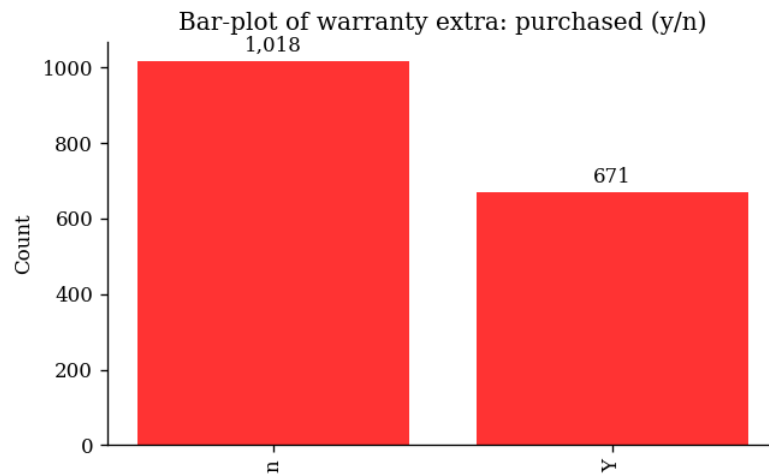## 1.16 Warranty Extra: Purchased (Y/N)

Warranty extra: purchased (y/n) is a categorical variable with 2 unique values. None of its values are missing.

### *Summary Statistics*

| Mode (Most frequent) | n |
|---|---|
| Maximum frequency | 1018 |



Bar-plot of warranty extra: purchased (y/n)

## 1.17 Other Services

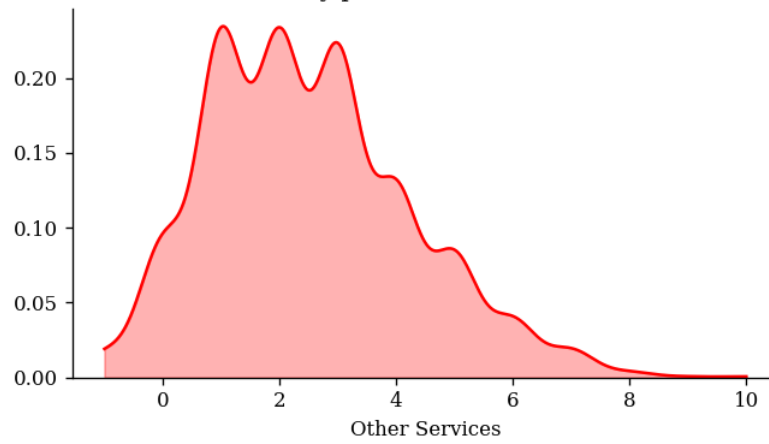Other services is a numeric variable with 12 unique values. None of its values are missing.

### *Summary Statistics*

| Average | 2.5098 |
|---|---|
| Standard Deviation | 1.7491 |
| Minimum | -1 |
| Lower Quartile | 1 |
| Median | 2 |
| Upper Quartile | 4 |
| Maximum | 10 |
| Skewness | 0.5593 |
| Kurtosis | 0.1398 |

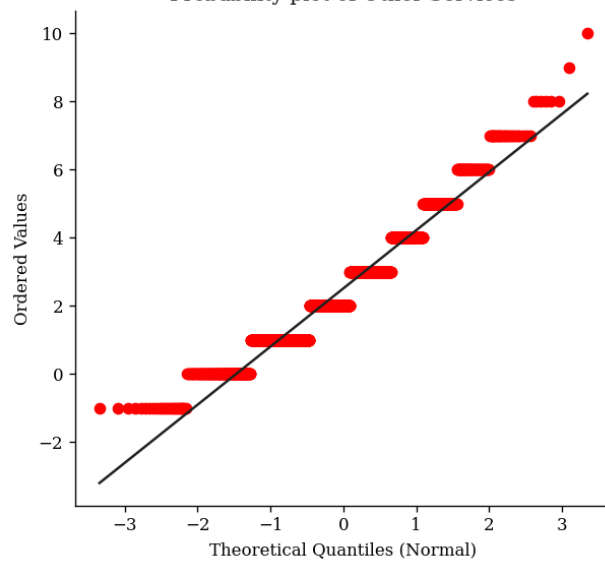## Box-plot of Other Services
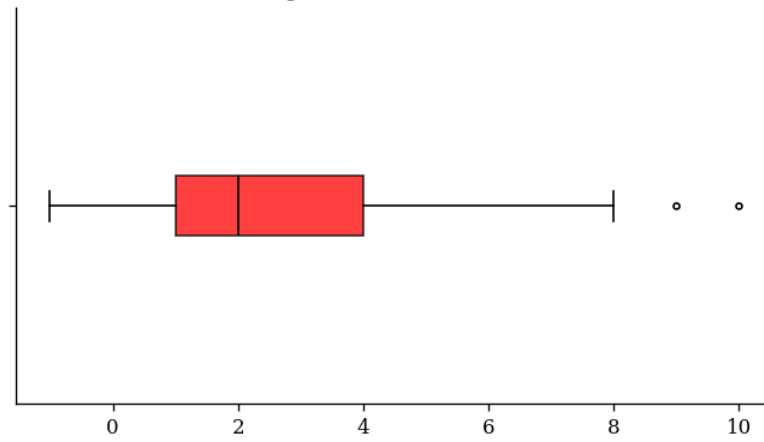
## Density plot of Other Services

## Probability plot of Other Services

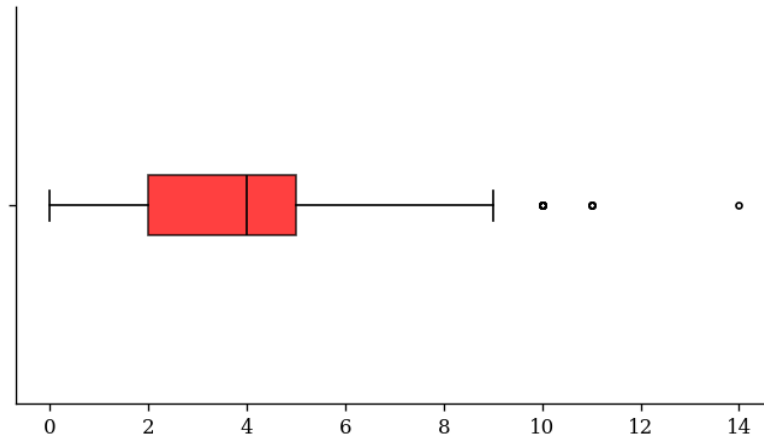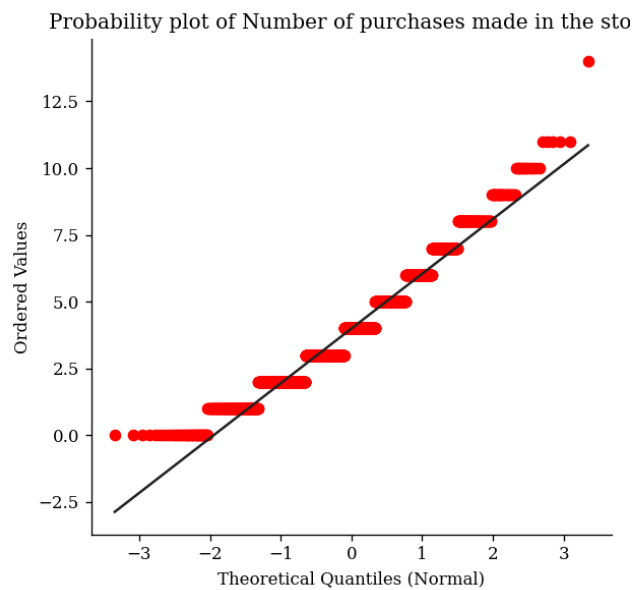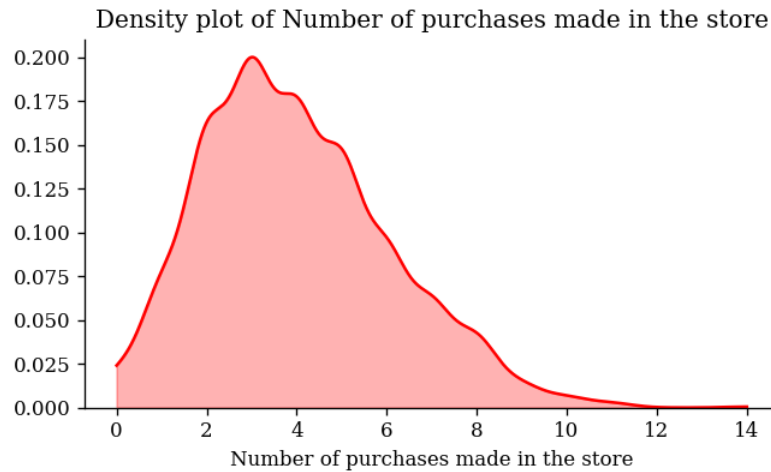|  | p-value | Conclusion at α = 0.05 |
|---|---|---|
| **D'Agostino's K-squared test** | 0.0000000 | Unlikely to be normal |
| **Kolmogorov-Smirnov test** | 0.0000000 | Unlikely to be normal |
| **Shapiro-Wilk test** | 0.0000000 | Unlikely to be normal |

## 1.18 Number Of Purchases Made In The Store

Number of purchases made in the store is a numeric variable with 13 unique values. None of its values are missing.

*Summary Statistics*

| | |
|---|---|
| Average | 3.9911 |
| Standard Deviation | 2.093 |
| Minimum | 0 |
| Lower Quartile | 2 |
| Median | 4 |
| Upper Quartile | 5 |
| Maximum | 14 |
| Skewness | 0.58 |
| Kurtosis | 0.2245 |

Box-plot of Number of purchases made in the store

## Density plot of Number of purchases made in the store



## Probability plot of Number of purchases made in the sto



### Tests for Normality

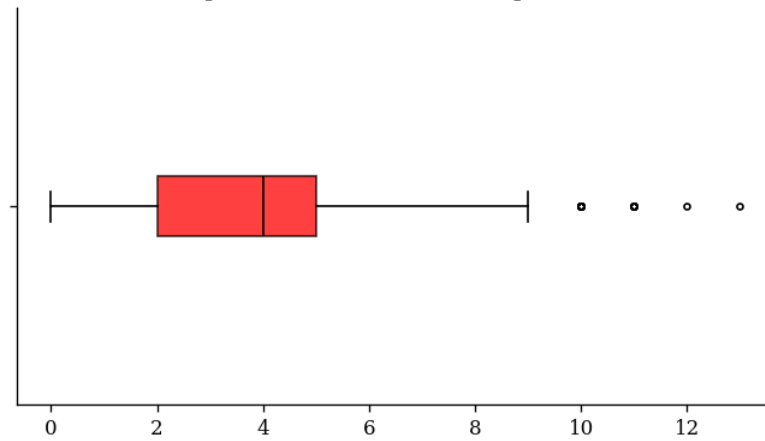| | p-value | Conclusion at α = 0.05 |
|---|---|---|
| **D'Agostino's K-squared test** | 0.0000000 | Unlikely to be normal |
| **Kolmogorov-Smirnov test** | 0.0000000 | Unlikely to be normal |
| **Shapiro-Wilk test** | 0.0000000 | Unlikely to be normal |

## 1.19 Number Of Online Purchases

Number of online purchases is a numeric variable with 14 unique values. None of its values are missing.
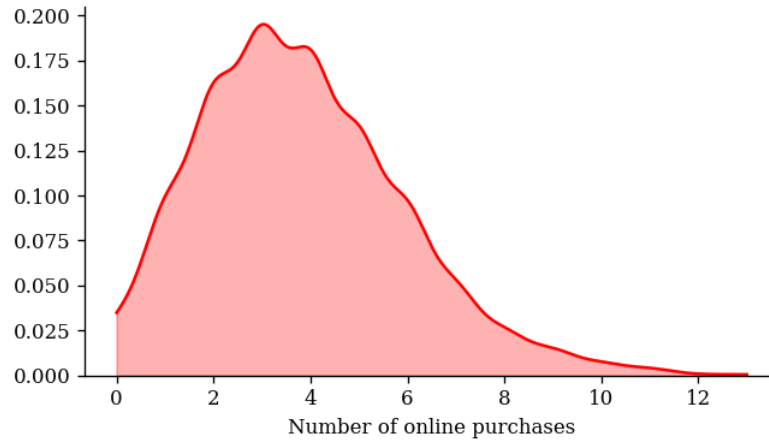
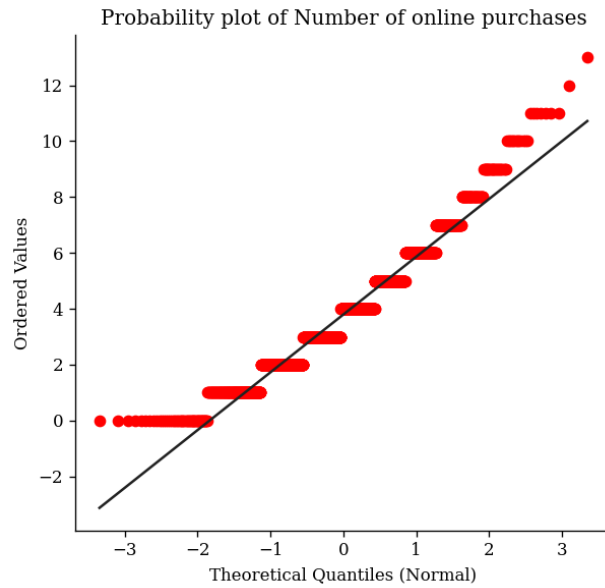| Average | 3.7975 |
|---|---|
| Standard Deviation | 2.1125 |
| Minimum | 0 |
| Lower Quartile | 2 |
| Median | 4 |
| Upper Quartile | 5 |
| Maximum | 13 |
| Skewness | 0.6387 |
| Kurtosis | 0.4833 |

Box-plot of Number of online purchases



Density plot of Number of online purchases

Probability plot of Number of online purchases

### Tests for Normality

| | p-value | Conclusion at α = 0.05 |
|---|---|---|
| D'Agostino's K-squared test | 0.0000000 | Unlikely to be normal |
| Kolmogorov-Smirnov test | 0.0000000 | Unlikely to be normal |
| Shapiro-Wilk test | 0.0000000 | Unlikely to be normal |

## 1.20 Discount Code Obtained

Discount code obtained is a categorical variable with 3 unique values. 14 (0.83%) of its values are missing.

### Summary Statistics

| Mode (Most frequent) | Online |
|---|---|
| Maximum frequency | 888 |

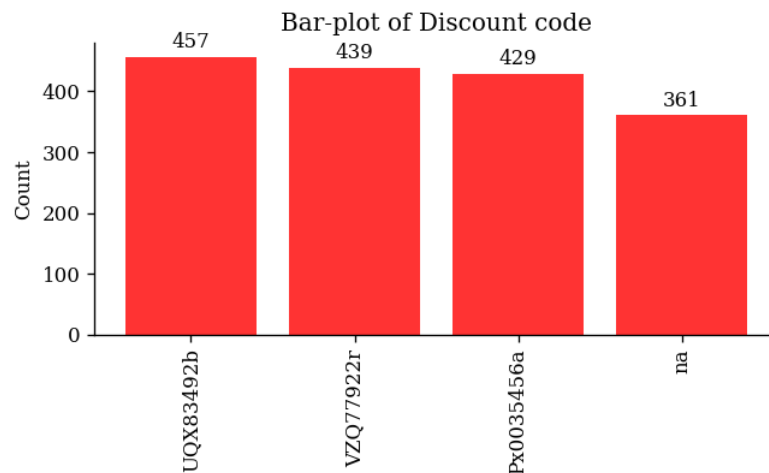Bar-plot of Discount code obtained
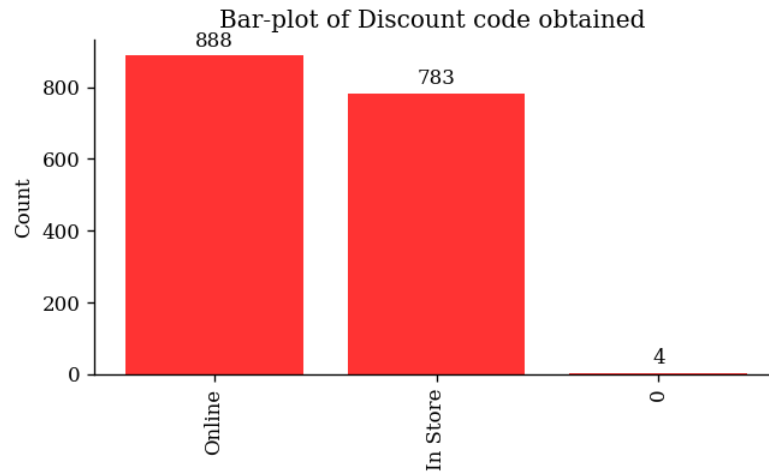
## 1.21 Discount Code

Discount code is a categorical variable with 4 unique values. 3 (0.18%) of its values are missing.

### Summary Statistics

| Mode (Most frequent) | UQX83492b |
|---|---|
| Maximum frequency | 457 |



Bar-plot of Discount code

## 1.22 Responded
## March Promo (Summer Tires)

Responded
march promo (summer tires) is a categorical variable with 1 unique value. 1,569 (92.90%) of its values are missing.

| Mode (Most frequent) | Y |
|---|---|
| Maximum frequency | 120 |

### Bar-plot of Responded
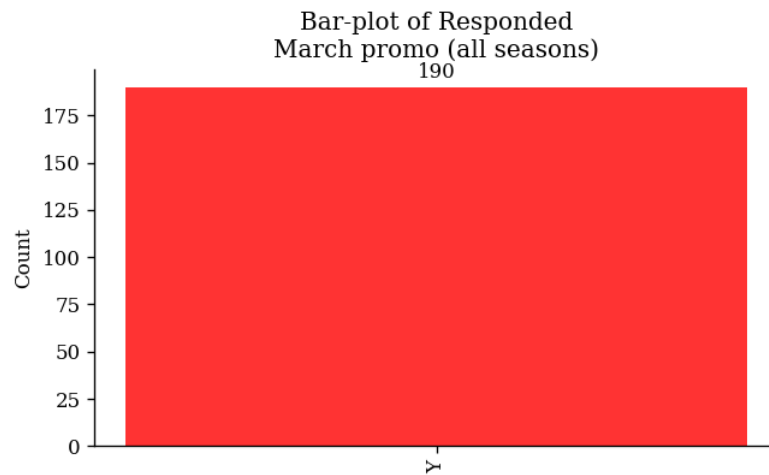### March Promo (summer tires)



## 1.23 Responded

## March Promo (All Seasons)

Responded
march promo (all seasons) is a categorical variable with 1 unique value. 1,499 (88.75%) of its
values are missing.

*Summary Statistics*

| Mode (Most frequent) | Y |
|---|---|
| Maximum frequency | 190 |

### Bar-plot of Responded
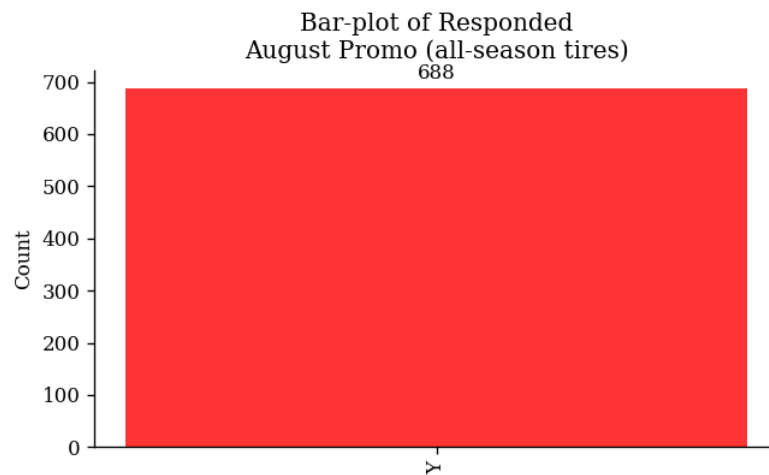### March promo (all seasons)

## 1.24 Responded

## August Promo (All-Season Tires)

Responded
august promo (all-season tires) is a categorical variable with 1 unique value. 1,001 (59.27%) of its values are missing.

### *Summary Statistics*

| Mode (Most frequent) | Y |
|---|---|
| Maximum frequency | 688 |



Bar-plot of Responded
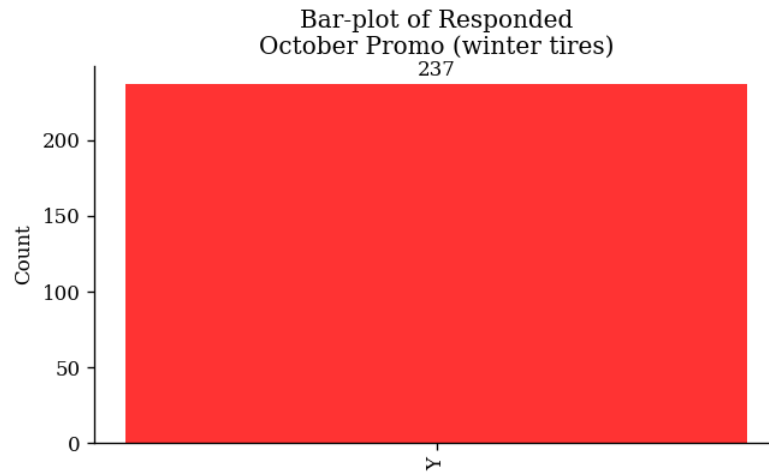August Promo (all-season tires)

## 1.25 Responded

## October Promo (Winter Tires)

Responded
october promo (winter tires) is a categorical variable with 1 unique value. 1,452 (85.97%) of its values are missing.

### *Summary Statistics*

| Mode (Most frequent) | Y |
|---|---|
| Maximum frequency | 237 |

Bar-plot of Responded
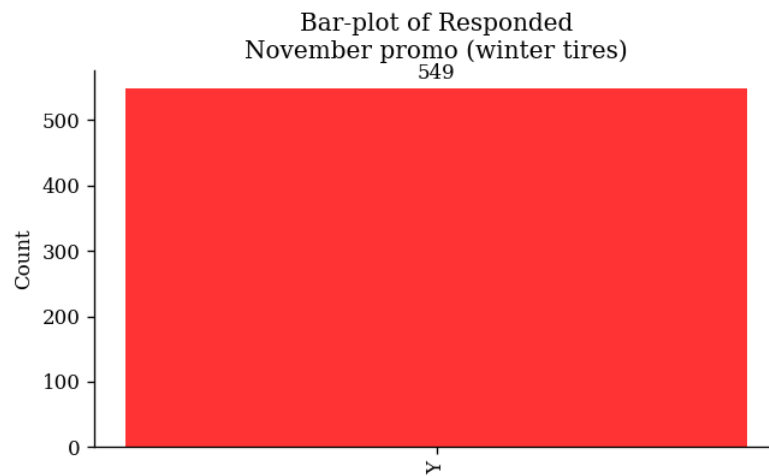October Promo (winter tires)

237



## 1.26 Responded

## November Promo (Winter Tires)

Responded
november promo (winter tires) is a categorical variable with 1 unique value. 1,140 (67.50%) of its values are missing.
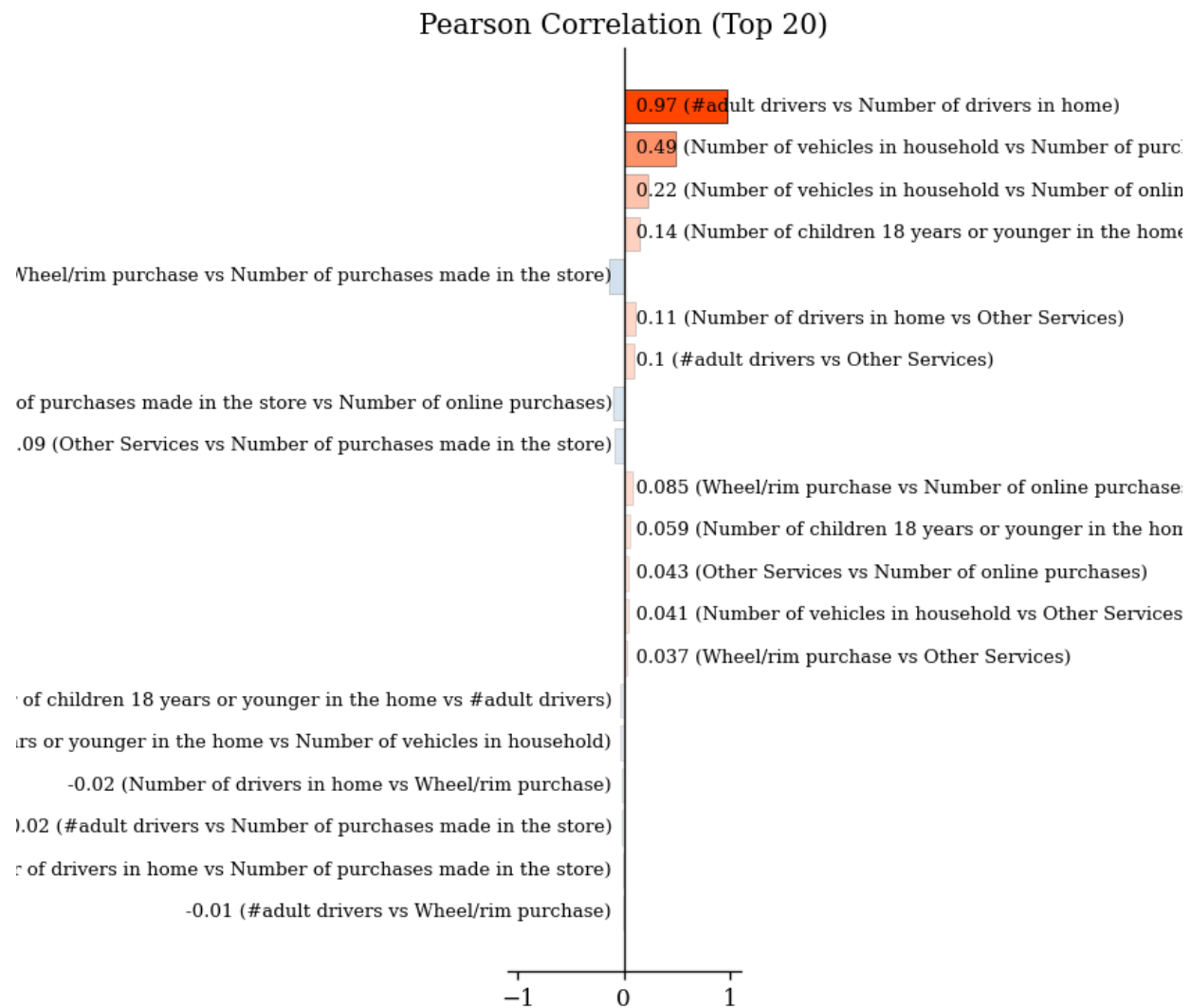
*Summary Statistics*

| Mode (Most frequent) | Y |
|---|---|
| Maximum frequency | 549 |

Bar-plot of Responded
November promo (winter tires)

549

## 2. Bivariate Analysis

### 2.1 Overview



Pearson Correlation (Top 20)

0.97 (#adult drivers vs Number of drivers in home)

0.49 (Number of vehicles in household vs Number of purc...

0.22 (Number of vehicles in household vs Number of onlin...

0.14 (Number of children 18 years or younger in the home...

Wheel/rim purchase vs Number of purchases made in the store)

0.11 (Number of drivers in home vs Other Services)

0.1 (#adult drivers vs Other Services)

of purchases made in the store vs Number of online purchases)

.09 (Other Services vs Number of purchases made in the store)

0.085 (Wheel/rim purchase vs Number of online purchase...

0.059 (Number of children 18 years or younger in the hon...

0.043 (Other Services vs Number of online purchases)

0.041 (Number of vehicles in household vs Other Services...

0.037 (Wheel/rim purchase vs Other Services)

of children 18 years or younger in the home vs #adult drivers)

rs or younger in the home vs Number of vehicles in household)

-0.02 (Number of drivers in home vs Wheel/rim purchase)

).02 (#adult drivers vs Number of purchases made in the store)

r of drivers in home vs Number of purchases made in the store)
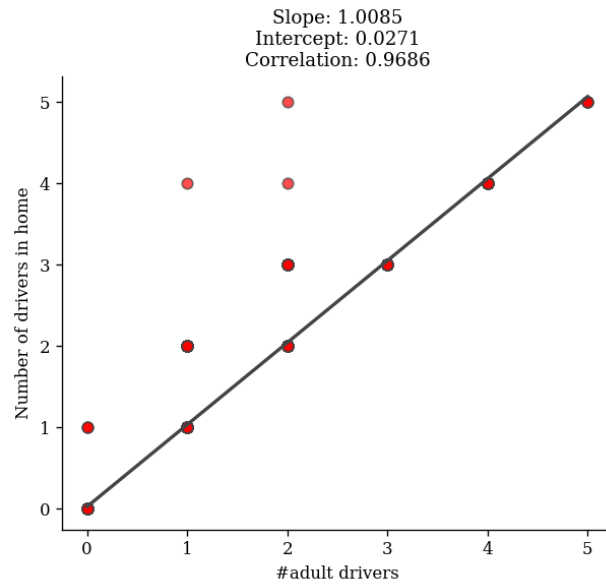
-0.01 (#adult drivers vs Wheel/rim purchase)

−1    0    1

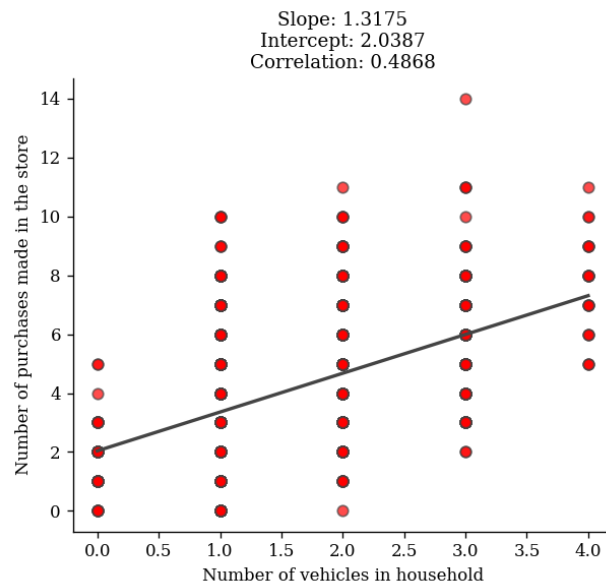## 2.2 Regression Plots (Top 20)

### 2.2.1 #Adult Drivers Vs Number Of Drivers In Home

#Adult Drivers and Number Of Drivers In Home have very strong positive correlation (0.97).



### 2.2.2 Number Of Vehicles In Household Vs Number Of Purchases Made In The Store

Number Of Vehicles In Household and Number Of Purchases Made In The Store have moderate positive correlation (0.49).

### 2.2.3 Number Of Vehicles In Household Vs Number Of Online Purchases

Number Of Vehicles In Household and Number Of Online Purchases have weak positive correlation (0.22).



Slope: 0.6040
Intercept: 2.9025
Correlation: 0.2211
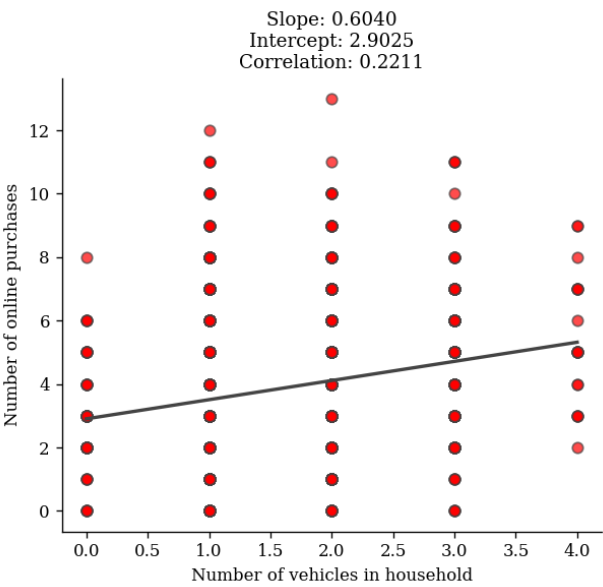
### 2.2.4 Number Of Children 18 Years Or Younger In The Home Vs Other Services

Number Of Children 18 Years Or Younger In The Home and Other Services have very weak positive correlation (0.14).
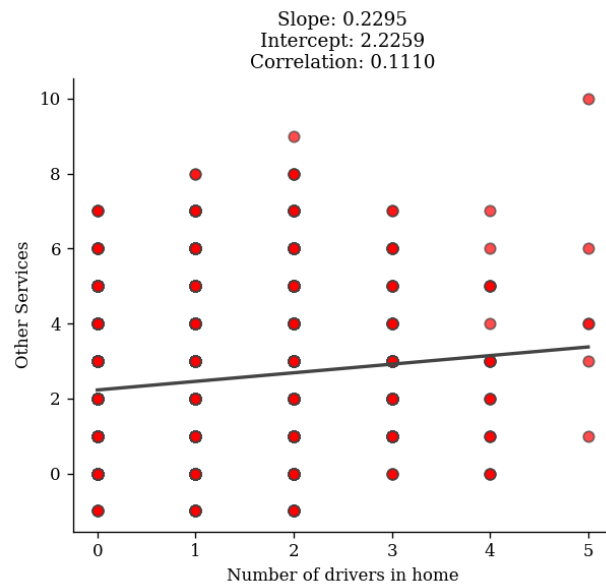


Slope: 0.2695
Intercept: 2.3650
Correlation: 0.1424
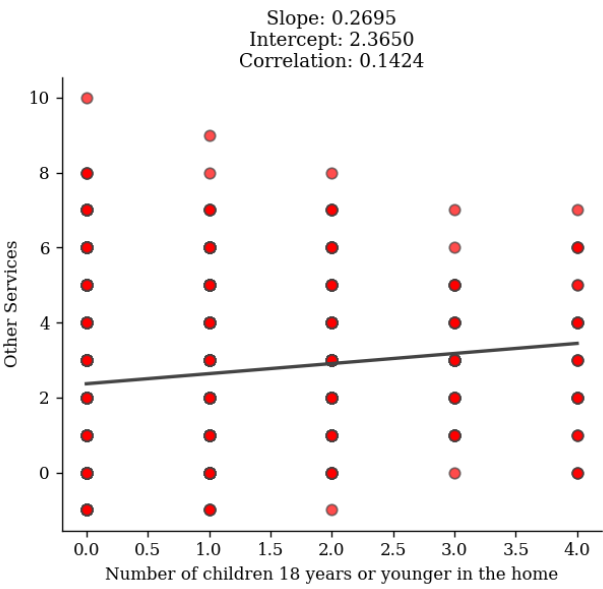
### 2.2.5 Wheel/Rim Purchase Vs Number Of Purchases Made In The Store

Wheel/Rim Purchase and Number Of Purchases Made In The Store have very weak negative correlation (-0.14).



Slope: -0.2261
Intercept: 4.4720
Correlation: -0.1378
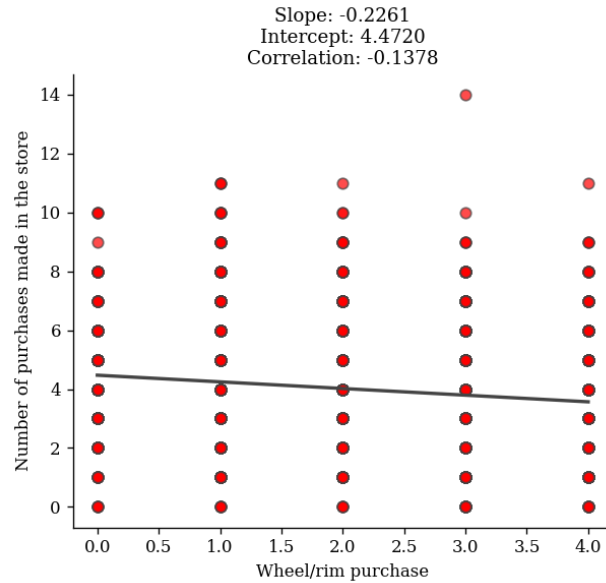
### 2.2.6 Number Of Drivers In Home Vs Other Services

Number Of Drivers In Home and Other Services have very weak positive correlation (0.11).



Slope: 0.2295
Intercept: 2.2259
Correlation: 0.1110

### 2.2.7 #Adult Drivers Vs Other Services

#Adult Drivers and Other Services have very weak positive correlation (0.10).

Slope: 0.2175
Intercept: 2.2489
Correlation: 0.1010
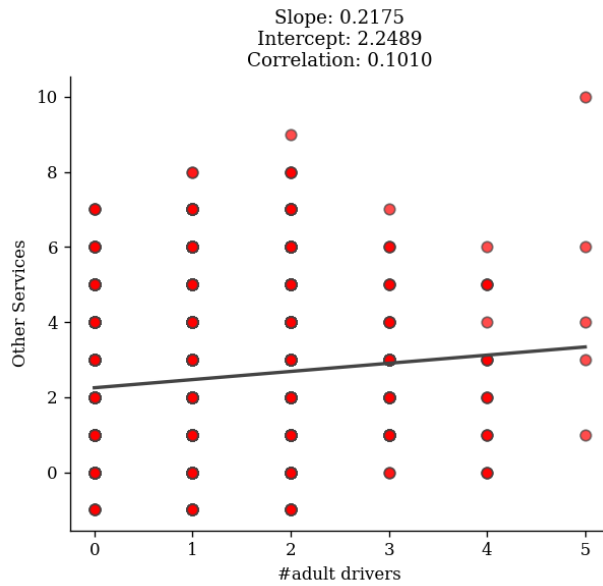
### 2.2.8 Number Of Purchases Made In The Store Vs Number Of Online Purchases

Number Of Purchases Made In The Store and Number Of Online Purchases have very weak negative correlation (-0.10).
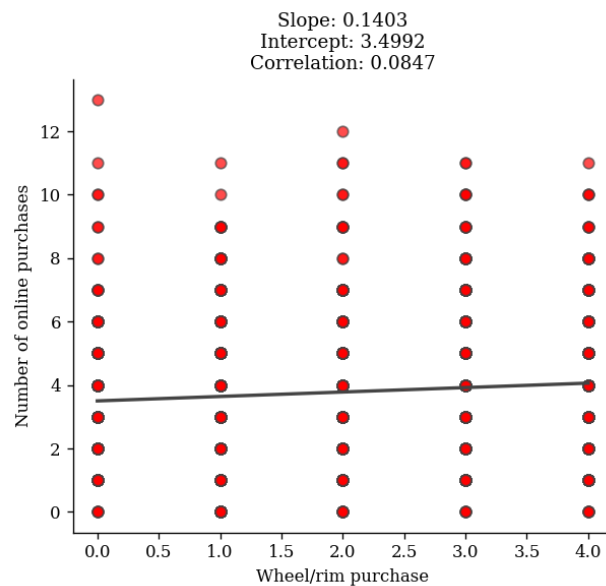


Slope: -0.0993
Intercept: 4.1937
Correlation: -0.0984
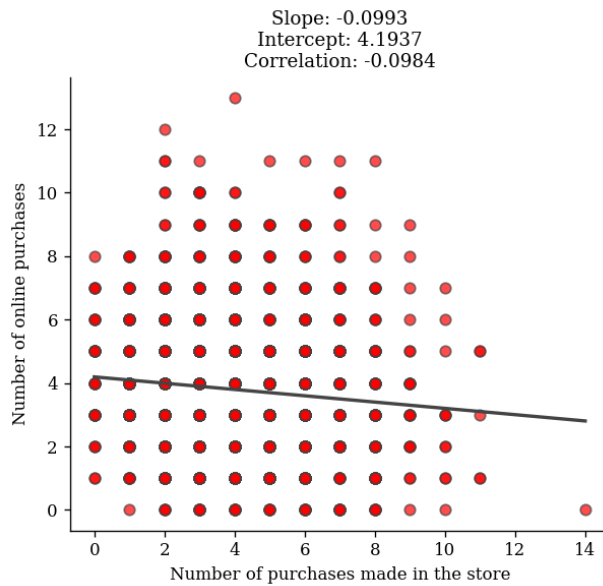
### 2.2.9 Other Services Vs Number Of Purchases Made In The Store

Other Services and Number Of Purchases Made In The Store have very weak negative correlation (-0.09).
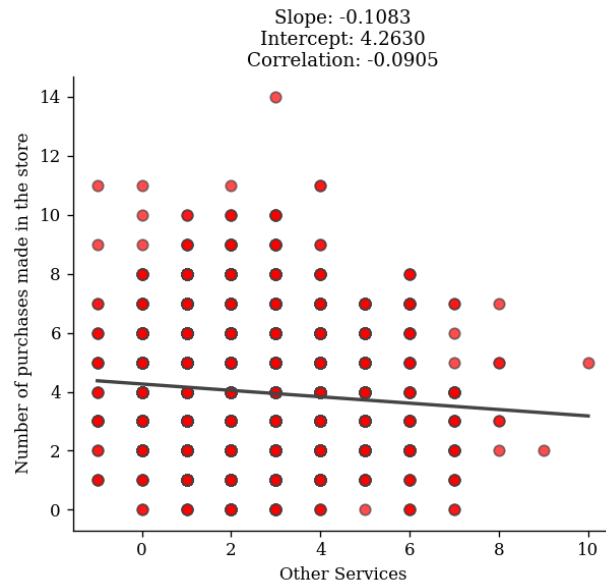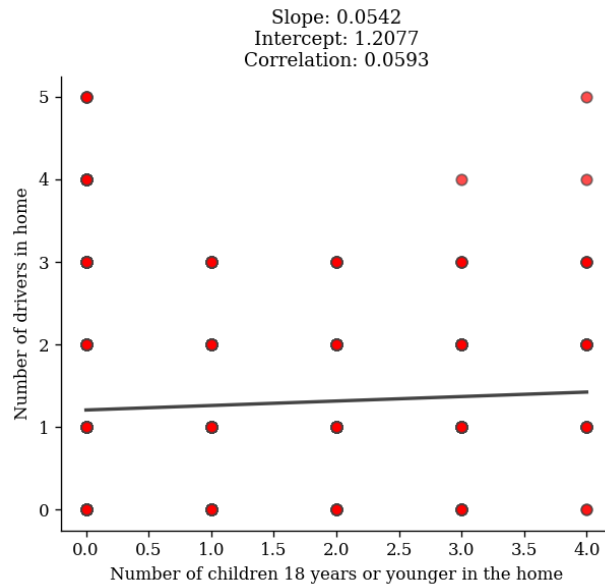
### 2.2.10 Wheel/Rim Purchase Vs Number Of Online Purchases

Wheel/Rim Purchase and Number Of Online Purchases have very weak positive correlation (0.08).
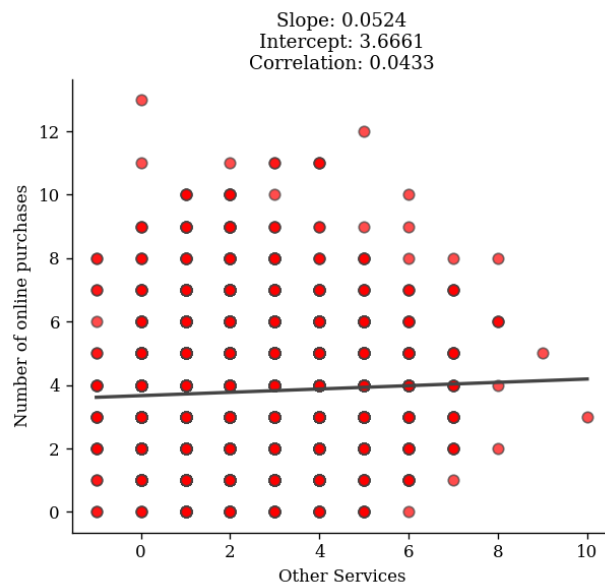


### 2.2.11 Number Of Children 18 Years Or Younger In The Home Vs Number Of Drivers In Home

Number Of Children 18 Years Or Younger In The Home and Number Of Drivers In Home have very weak positive correlation (0.06).
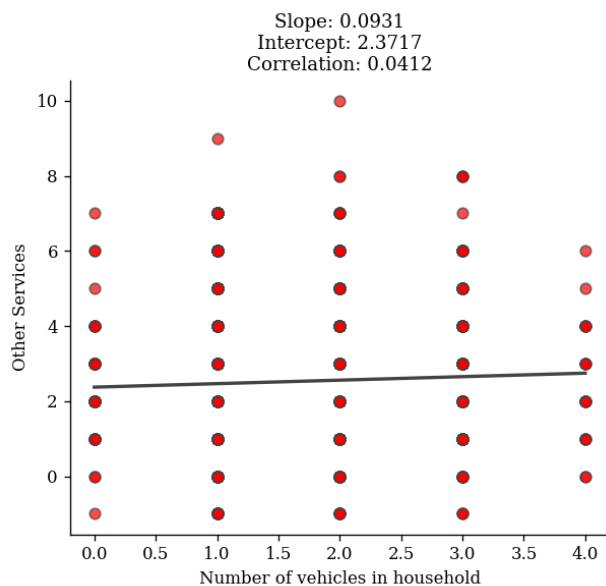
### 2.2.12 Other Services Vs Number Of Online Purchases

Other Services and Number Of Online Purchases have virtually no correlation (0.04).
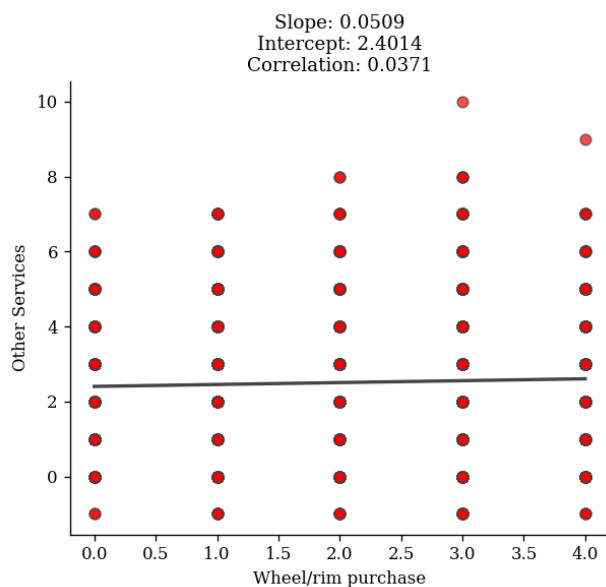


### 2.2.13 Number Of Vehicles In Household Vs Other Services

Number Of Vehicles In Household and Other Services have virtually no correlation (0.04).

Slope: 0.0931
Intercept: 2.3717
Correlation: 0.0412

### 2.2.14 Wheel/Rim Purchase Vs Other Services

Wheel/Rim Purchase and Other Services have virtually no correlation (0.04).



Slope: 0.0509
Intercept: 2.4014
Correlation: 0.0371

### 2.2.15 Number Of Children 18 Years Or Younger In The Home Vs #Adult Drivers

Number Of Children 18 Years Or Younger In The Home and #Adult Drivers have virtually no correlation (-0.03).

Slope: -0.0270
Intercept: 1.2140
Correlation: -0.0308
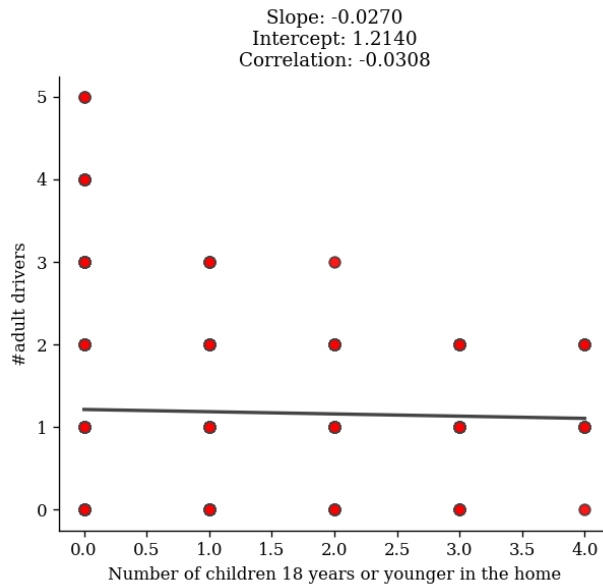
## 2.2.16 Number Of Children 18 Years Or Younger In The Home Vs Number Of Vehicles In Household

Number Of Children 18 Years Or Younger In The Home and Number Of Vehicles In Household have virtually no correlation (-0.03).
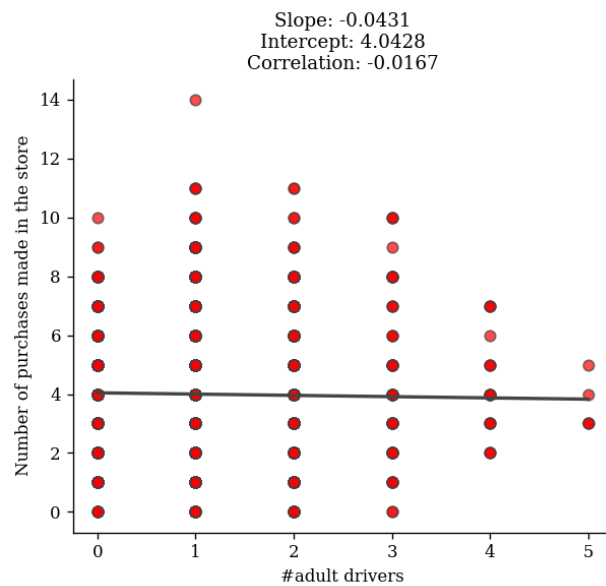


Slope: -0.0244
Intercept: 1.4950
Correlation: -0.0291
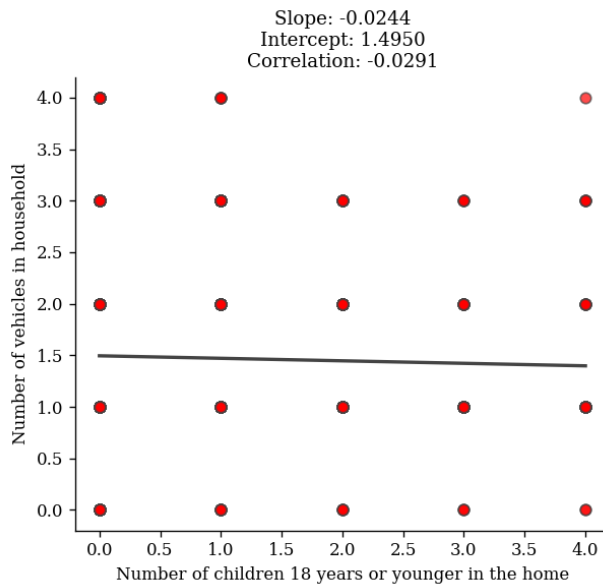
## 2.2.17 Number Of Drivers In Home Vs Wheel/Rim Purchase

Number Of Drivers In Home and Wheel/Rim Purchase have virtually no correlation (-0.02).

Slope: -0.0296
Intercept: 2.1633
Correlation: -0.0196
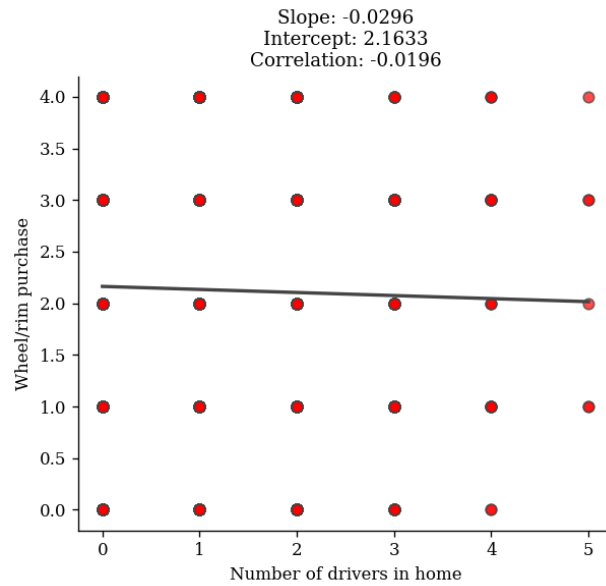
### 2.2.18 #Adult Drivers Vs Number Of Purchases Made In The Store

#Adult Drivers and Number Of Purchases Made In The Store have virtually no correlation (-0.02).



Slope: -0.0431
Intercept: 4.0428
Correlation: -0.0167

### 2.2.19 Number Of Drivers In Home Vs Number Of Purchases Made In The Store

Number Of Drivers In Home and Number Of Purchases Made In The Store have virtually no correlation (-0.01).

Slope: -0.0327
Intercept: 4.0315
Correlation: -0.0132
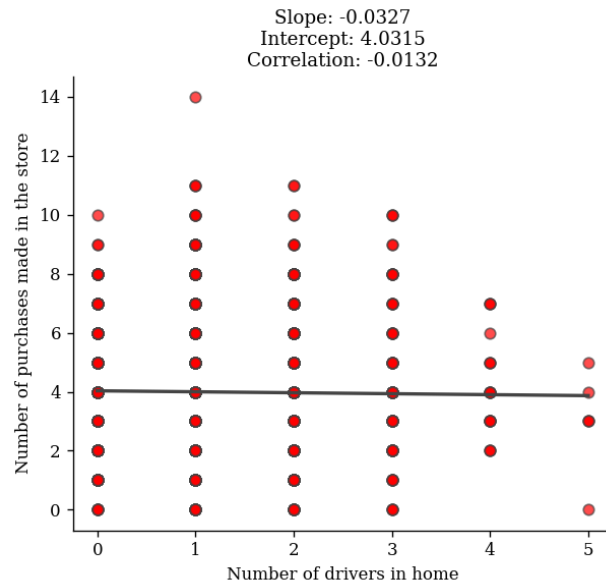
## 2.2.20 #Adult Drivers Vs Wheel/Rim Purchase

#Adult Drivers and Wheel/Rim Purchase have virtually no correlation (-0.01).



Slope: -0.0204
Intercept: 2.1511
Correlation: -0.0130