SUV Car Sales Performance Report

Contents

Introduction

This report aim's to improving sales of SUV cars performance. We will do that by Finding the relation between variables, Which can help build decisions in many sectors as (Marketing & Advertising, Financial planning). As well as, Forecasting the reasons which effect in purchased process.

Giving an overview in predicted age which can buy the cars based on his salary and sex. By the end we can focus on a specific category which help us to increasing the sales . We will do that using Fisher's exact test.

In this report, We tried to put hypothesis and determine if the purchased done or not. We Analyze dataset Using R, SPSS & Power BI to summarize data and use logistic Regression To determine the main factors which effect on our output.

Research Question

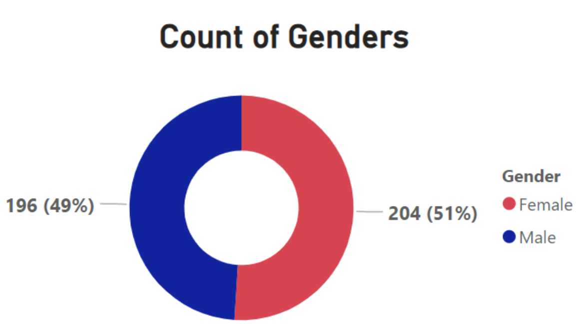
We will figure out if There is a relation between Purchased , Age , salary and Gender .

$$\frac{1}{\pi 1-\pi} = \exp\left(\beta_0 + B_1 X_1 + B_2 X_2 + B_3 X_3 + B_4 X_4 + \ldots + B_k X_k\right)$$

Purchased = C + [B1 × Age + [B2 × salary] + [B3 × Gender (Male)]

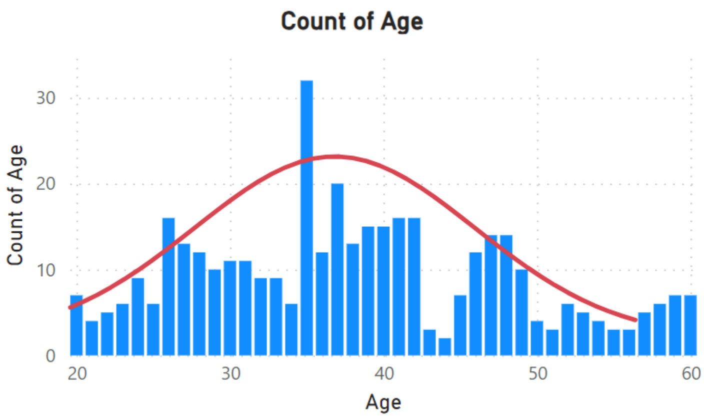
Descriptive Analysis

Gender	Count of Gender
Female	204
Male	196
Total	**400**

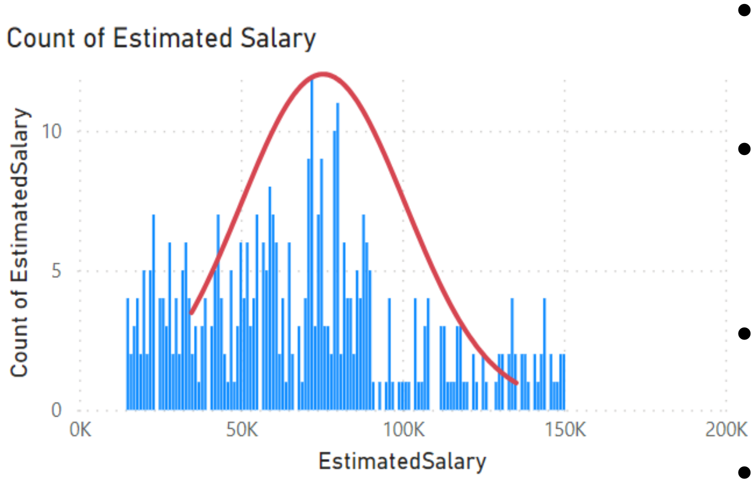


- We took a random sample from sales department which contained of 400 random and unbiased variables.
- We have 204 (51%) females and 196 (49%) males which almost close.

	N	Range	Minimum	Maximum	Sum	Mean		Std. Deviation	Variance	Skewness	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic	Statistic	Std. Error
Age	400	42	18	60	15062	37.65	.524	10.483	109.891	.231	.122
EstimatedSalary	400	135000	15000	150000	27897000	69742.50	1704.848	34096.960	1162602701	.495	.122
	400										



- As we can see that the average of ages is 37.65 which mean that our customers center in the medium age between (25 - 44) and maximum is 60 and as we can see that the old people are less interested than the young and because of that we can see a positive skewness (0.231) but it’s very small.
- The mode is 35 years old, Which mean that the most age interesting and buying SUV cars is (35 years old)



- The average of salary is 69,742 $ and that is not mean that we just focus in people who get this salary.
- We had minimum salary which is 15,000$ and maximum is 150,000 $ .So, it means that all categories between those salary interested in this type of cars.
- Also we have positive skewness (0.495) and the mode is 72,000 $ (The most salary interested in this type of cars).
- The Standard deviation is very high that’s mean the salary is not clustered around the mean we have different types of salaries.

Logistic Regression

```
Call:
glm(formula = new_df$Purchased ~ new_df$Age + new_df$EstimatedSalary +
    new_df$gender, family = "binomial")

Coefficients:
                         Estimate Std. Error z value Pr(>|z|)
(Intercept)            -1.278e+01  1.359e+00  -9.405  < 2e-16 ***
new_df$Age              2.370e-01  2.638e-02   8.984  < 2e-16 ***
new_df$EstimatedSalary  3.644e-05  5.473e-06   6.659 2.77e-11 ***
new_df$gender           3.338e-01  3.052e-01   1.094    0.274
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 521.57  on 399  degrees of freedom
Residual deviance: 275.84  on 396  degrees of freedom
AIC: 283.84

Number of Fisher Scoring iterations: 6
```

- We can see in this model that the function will be

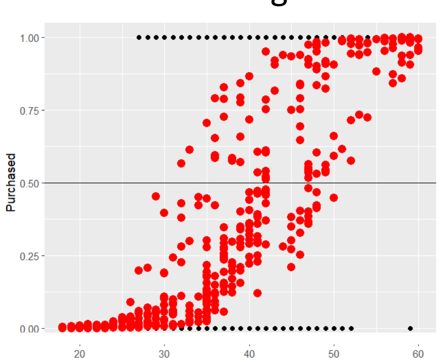
$$\mathbf{Ln\ (Purchase) = B_0 + [B_1 \times Age] + [B_2 \times Salary] + [\ B_3 \times Gender\ ]}$$

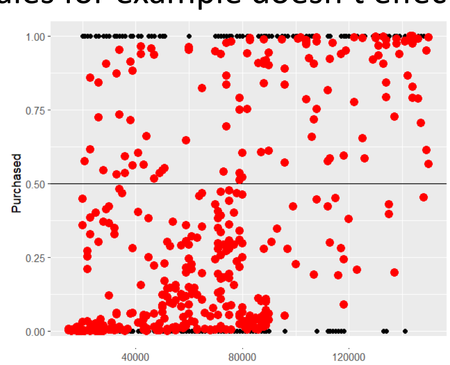
$$\frac{1}{\pi 1-\pi} = \mathbf{exp\ (\ -12.7836 + [Age \times 0.2370\ ] + [Salary \times 0.00004] + [\ Gender \times 0.3338]\ )}$$

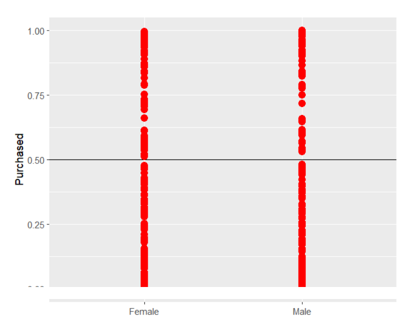
- From the output , We can see that age and salary are significant in our model . It can effect on the purchase process probability. As well , Gender is not significant and it doesn’t effect. So, We can’t focus on males and ignore females.

```
> round(exp(model$coefficients) , 5 )
           (Intercept)             new_df$Age new_df$EstimatedSalary          new_df$gender
               0.00000                1.26740                1.00004                1.39632
>
```

- Increasing Age by 1 year correspond to increasing odds of the purchased process by 0.26 %. Also increasing Salary by 1 $ causes increasing the probability of purchasing by 0.00004 %
- Gender is not significant so males for example doesn’t effect in the purchase process .







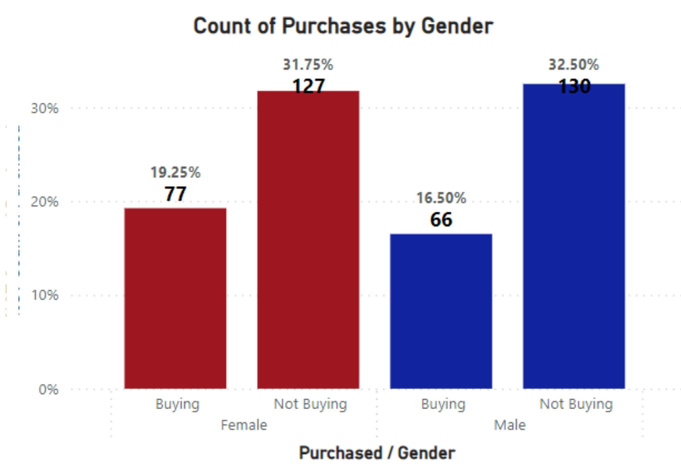
- As we can see in figure 1, the probability of purchased for ages which are higher than 30 is higher than whose age is 20 years. So, We should focus on this age.
- On the other side , all salaries has the probability to purchasing. So, We should focus on all salaries.
- For gender we can't predict by the logistic regression so we will try another method.

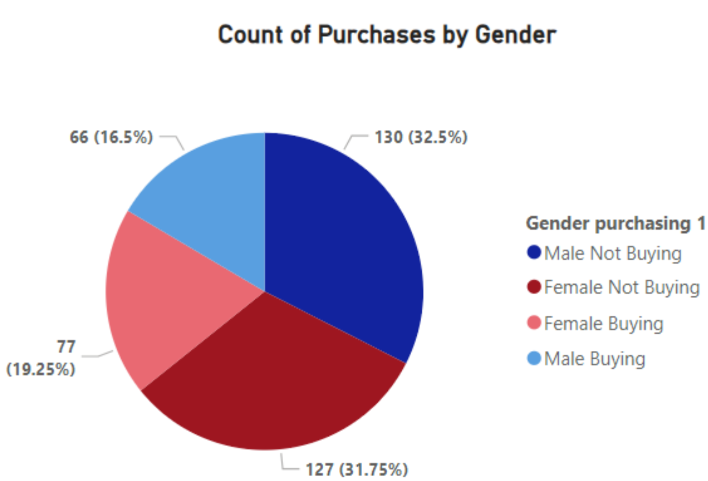
The probability of Total Purchase by Gender (Fisher's exact test)

- In this part , we want to determine whether there is a statistically significant association between Gender and Purchasing process.
- We will use Fasher's exact test to see if there is association between Gender and Purchases

Gender	0	1	**Total**
Female	127	77	**204**
Male	130	66	**196**
Total	**257**	**143**	**400**

Gender	Buying	Not Buying	**Total**
Female	19.25%	31.75%	**51.00%**
Male	16.50%	32.50%	**49.00%**
Total	**35.75%**	**64.25%**	**100.00%**





- As we can see the total number of Purchasing from male are 66 out of 196 males which represent 16.50% from total purchasing.
- Compared to females 77 out of 204 which represent 19.25% of total purchasing.
- We can see that females have higher percentage of purchasing SUV cars than males but still we can not say that there's a relation between gender and Purchasing.
- But we can say that females prefer SUV cars than males.

Probability (Purchasing ratio) by gender “Rows Analysis”

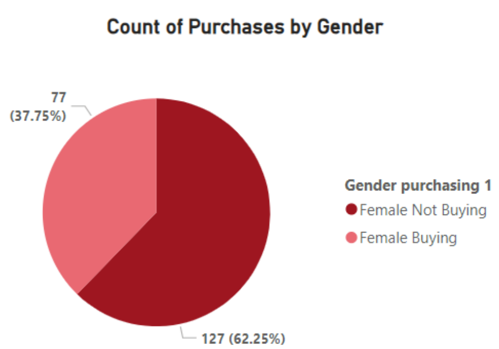
```
$data
          purchase
gender    not Buy buy Total
  Male        130  66   196
  Female      127  77   204
  Total       257 143   400

$measure
          risk ratio with 95% C.I.
gender     estimate     lower    upper
  Male     1.000000        NA       NA
  Female 1.120915 0.8608868 1.459484

$p.value
          two-sided
gender     midp.exact fisher.exact chi.square
  Male             NA           NA         NA
  Female  0.3983004    0.4057912  0.3956648

$correction
[1] FALSE

attr(,"method")
[1] "Unconditional MLE & normal approximation (Wald) CI"
```

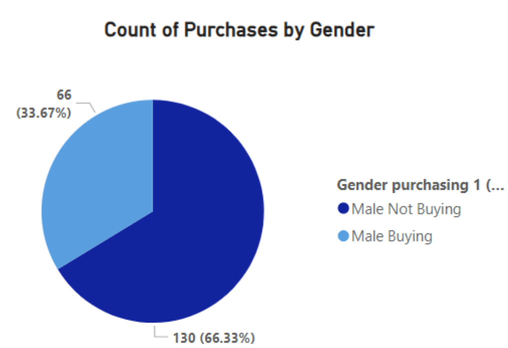
```
         purchase
gender      0        p0  1        p1 riskratio     lower    upper   p.value
  Female 127 0.6225490 77 0.3774510 1.0000000        NA       NA        NA
  Male   130 0.6632653 66 0.3367347 0.8921283 0.6851737 1.161593 0.4057912

$measure
[1] "wald"

$conf.level
[1] 0.95

$pvalue
[1] "fisher.exact"
```

Count Purchasing / Total by Gender

Gender	0	1	Total
Female	127	77	**204**
Male	130	66	**196**
Total	**257**	**143**	**400**

Gender	Buying	Not Buying	Total
Female	37.75%	62.25%	**100.00%**
Male	33.67%	66.33%	**100.00%**
Total	**35.75%**	**64.25%**	**100.00%**





- **From this data we have the probability of purchasing among Males and females**

For example:

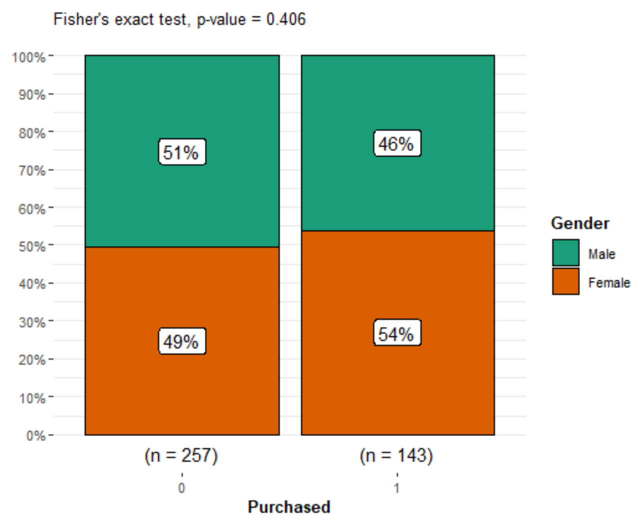
- We have 37.7% of females buying SUV cars compared to females which not buying 62.25%.
- On the other hand , 33.67% from males buying SUV cars and 66.33% not buying.

- **From this output, We can say that females are 1.12 times higher interested and preferred purchasing SUV cars than males.**
- **Males are 0.107 times lower interested in SUV Cars than females.**
- **The percentage of males who buys SUV cars (33.67%) is lower than females (37.75 %). Beside that, Male has higher percentage for not buying (66.33%) compared to females (62.25%) .**

Probability (odds ratio for gender) by Purchasing "Column Analysis"

```
$data
        purchase
gender   not Buy buy Total
  Male       130  66   196
  Female     127  77   204
  Total      257 143   400

$measure
        odds ratio with 95% C.I.
gender   estimate     lower   upper
  Male   1.000000        NA      NA
  Female 1.194226 0.7927509 1.79902

$p.value
        two-sided
gender   midp.exact fisher.exact chi.square
  Male           NA           NA         NA
  Female  0.3983004    0.4057912  0.3956648

$correction
[1] FALSE

attr(,"method")
[1] "Unconditional MLE & normal approximation (Wald) CI"
>
```

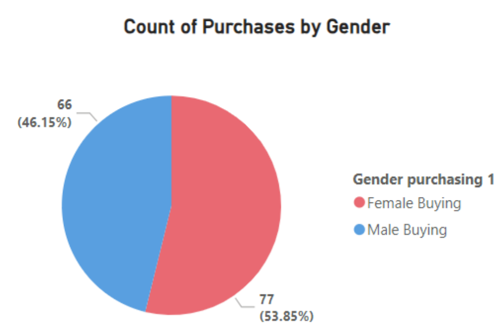
```
$tab
       purchase
gender   buy        p0 not Buy        p1 oddsratio     lower   upper   p.value
  Female  77 0.5384615     127 0.4941634  1.000000        NA      NA        NA
  Male    66 0.4615385     130 0.5058366  1.194226 0.7927509 1.79902 0.4057912

$measure
[1] "wald"

$conf.level
[1] 0.95

$pvalue
[1] "fisher.exact"
```

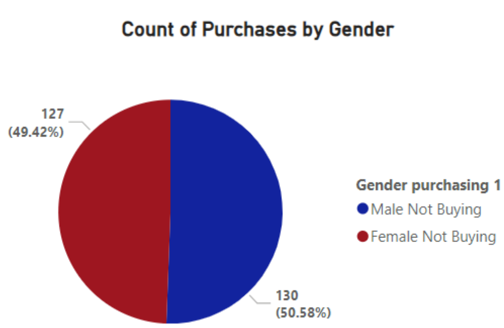


Gender	0	1	Total
Female	127	77	**204**
Male	130	66	**196**
Total	**257**	**143**	**400**

Gender	Buying	Not Buying	Total
Female	53.85%	49.42%	**51.00%**
Male	46.15%	50.58%	**49.00%**
Total	**100.00%**	**100.00%**	**100.00%**

Count Purchasing / Total by Purchasing





- **There is no association between gender and Purchasing.**
- **As we see the distribution in Purchasing with Gender is almost closed.**
- **By column analysis we can see that the percentage of buying from females is 53.85% compared to males 46.15 %.**
- **On the other hand, The percentage of not buying from females is 49.42% compared to males which is 50.58%.**
- **From this output we can say that females are 1.19 times more odds to buying SUV cars than males.**

The decision

- Finally, we can take the decision from outputs by focusing on females gender on our advertisements. we can increase the investment in advertisement for females by 20% than males.
- We should also focus on people whose age between 25 and 40 years specifically 35 years.
- The most interested are whose salary are medium between 23,000$ and 88,000$ but we should try to attract more whose salary are higher than 88,000 $.
- There is no association between Gender and Purchasing.

Conclusion

According to the result of out tests and summary, We should be more care about some variables which by that can help in improving SUV cars sales. Also we have to be careful for our advertisements for gender, salaries category and age.

References & dataset

1- "SUV data for logistic algorithm" , Retrieved from the link https://www.kaggle.com/datasets/kadamsuraj/suv-data-for-logistic-algorithm/data **.**