SMDM Graded Project - Coded

Great Learning DSBA program

Anirudh Sardiwal

24 December 2023

Table of Contents

Problem 1 - Austo Motor Company - Data Dictionary

Feature Name	Data Type	No. Of Null Values	Description
age	int64	0	Age of customer in years, integer
gender	object	53	Gender of customer - Male or Female
Profession	object	0	Has been divided into only 2 categories - Salaried or Business
marital_status	object	0	Married or Single
education	object	0	Graduate or Post Graduate only
no_of_dependents	int64	0	Number of people the customer supports financially - integer, values from 0 to 4
personal_loan	object	0	Whether customer has taken a personal loan - Yes or No - Binary variable
house_loan	object	0	Whether customer has taken a house loan - Yes or No - Binary variable
partner_working	object	0	Whether customer's partner or spouse works professionally - Yes or No - Binary variable
salary	int64	0	Customer's salary in $
partner_salary	int64	106	Customer's partner's salary in $
Total_salary	int64	0	Combined salary of customer and partner in $
price	int64	0	Price of vehicle in $
make	object	0	Make of vehicle - SUV, Sedan, or Hatchback

Problem 1 - Introduction

Austo Automobile is a leading car manufacturing company specializing in SUV, Sedan, and Hatchback models. Aim of the analysis is to get an idea of the demand of customers, i.e., which segment of customers are preferring which model.

1.1 - Data Overview

The data consists of 1581 rows and 14 columns. It is not very difficult to interpret. The values for each feature are less and easy to understand.

	Age	Gender	Profession	Marital_status	Education	No_of_Dependents	Personal_loan	House_loan	Partner_working	Salary	Partner_salary	Total_salary	Price	Make
0	53	Male	Business	Married	Post Graduate	4	No	No	Yes	99300	70700.0	170000	61000	SUV
1	53	Femal	Salaried	Married	Post Graduate	4	Yes	No	Yes	95500	70300.0	165800	61000	SUV
2	53	Female	Salaried	Married	Post Graduate	3	No	No	Yes	97300	60700.0	158000	57000	SUV
3	53	Female	Salaried	Married	Graduate	2	Yes	No	Yes	72500	70300.0	142800	61000	SUV
4	53	Male	Salaried	Married	Post Graduate	3	No	No	Yes	79700	60200.0	139900	57000	SUV

Fig. 1 - First 5 rows of the data

There are 6 numerical fields. Their key stats are as below.

	Age	No_of_Dependents	Salary	Partner_salary	Total_salary	Price
count	1581.00	1581.00	1581.00	1475.00	1581.00	1581.00
mean	31.92	2.46	60392.22	20225.56	79626.00	35597.72
std	8.43	0.94	14674.83	19573.15	25545.86	13633.64
min	22.00	0.00	30000.00	0.00	30000.00	18000.00
25%	25.00	2.00	51900.00	0.00	60500.00	25000.00
50%	29.00	2.00	59500.00	25600.00	78000.00	31000.00
75%	38.00	3.00	71800.00	38300.00	95900.00	47000.00
max	54.00	4.00	99300.00	80500.00	171000.00	70000.00

Fig. 2 - Statistical Description of Numerical columns

There are 53 values missing in the Gender column and 106 missing in the Partner Salary column. Spelling mistakes in the Gender column have been

treated. 53 or 3.3% of null values in the Gender column have been ignored. Null values of the Partner Salary column have been replaced by the mean of the column.

1.1.1 Data Overview - Observations and Insights

- Data is simple and easy to read.
- Number of null values is quite less.

1.2 - Univariate and Bivariate Analysis

1.2.1 Age Distribution

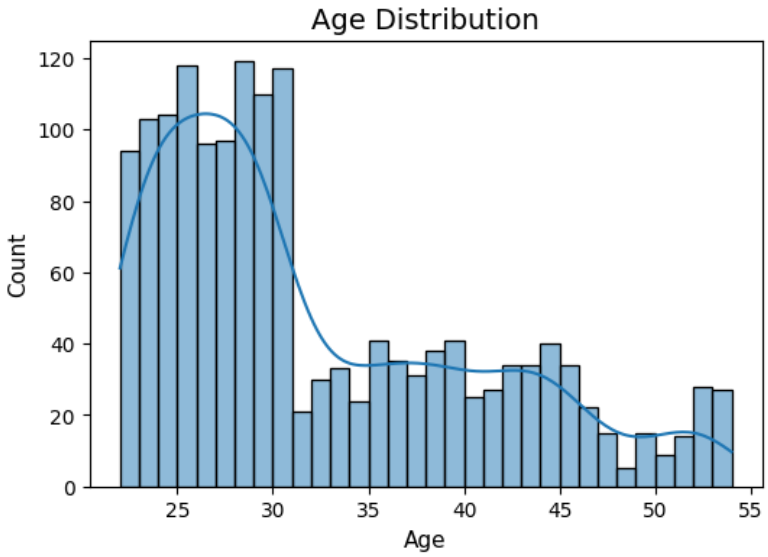


Fig. 3 - Age Distribution

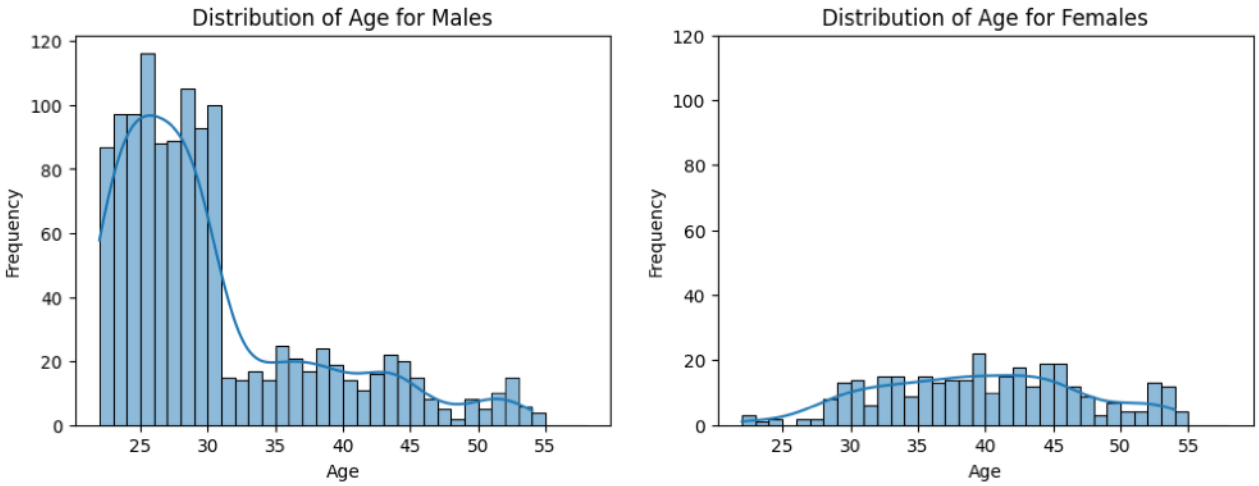


Fig. 4 - Distribution of Age by Gender

Majority of customers seem to be between the ages of 22 and 31. Men's data is right skewed with the majority being below 31 while women's data is more balanced.

1.2.2 Salaries vs Price of Vehicles

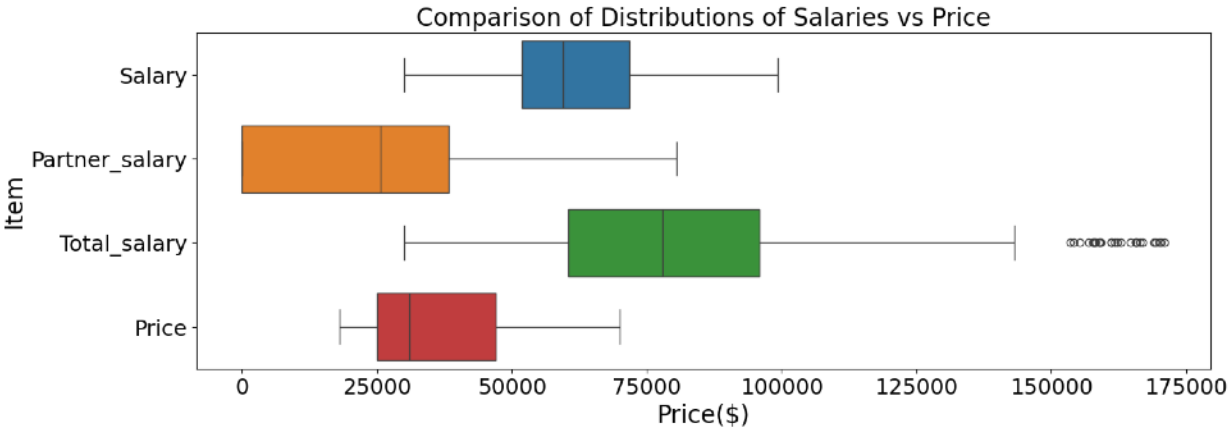


Fig. 5 - Comparison of Distributions of Salaries vs Price

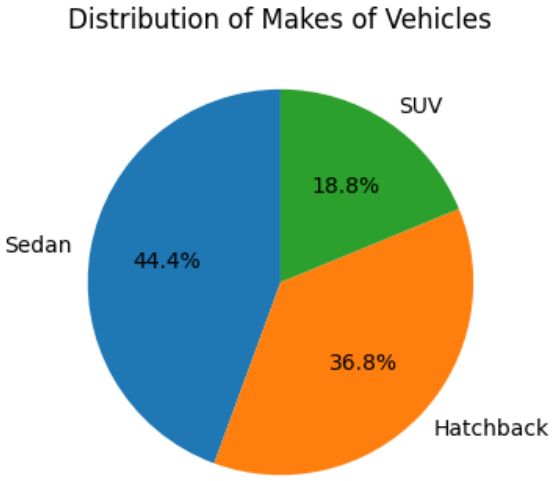


Fig. 6a - Distribution of Makes of Vehicles

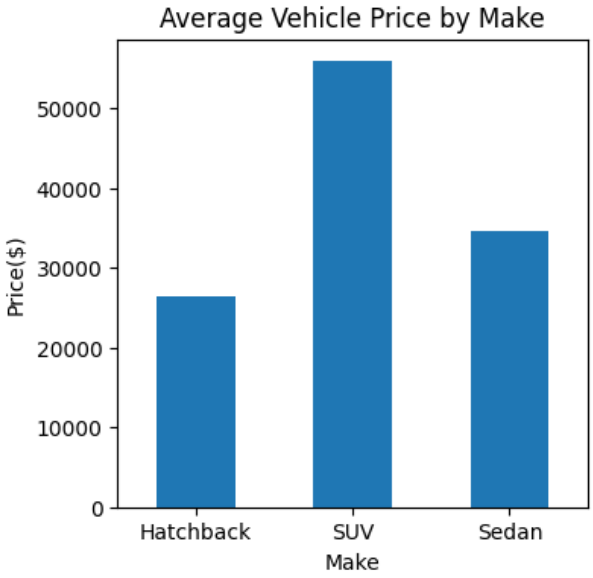


Fig. 6b - Average vehicle price by Make

Keeping Salary as the base, comparing Partner Salary and Total Salary we see that Total Salary becomes considerably more. Partner Salary is absent at many instances and the median is much less than Salary too, but whenever

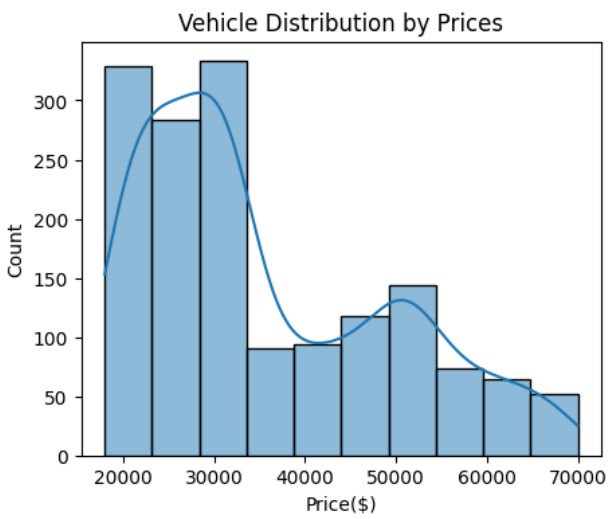


Fig. 7 - Vehicle distribution by Price

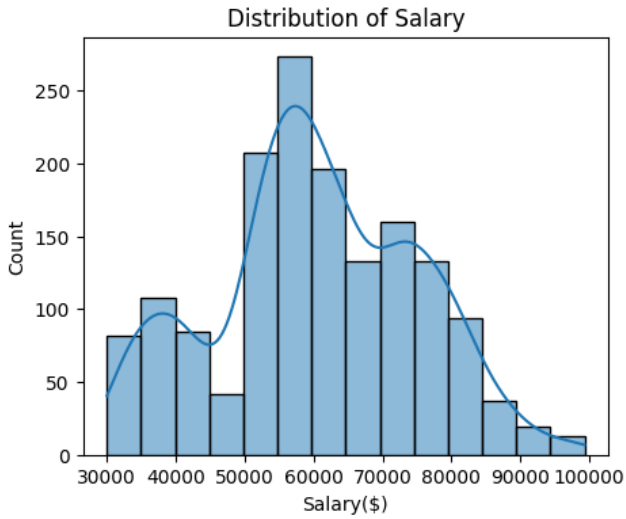


Fig. 8 - Distribution of Salary

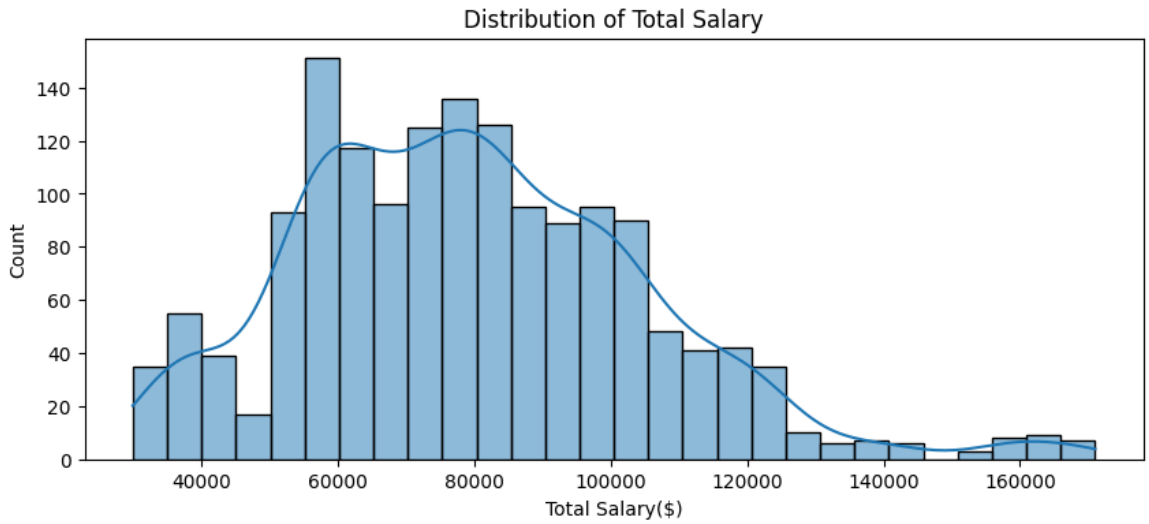


Fig. 9 - Distribution of Total Salary

there is Partner Salary present the total buying power increases with several outliers. In any case the mean Price of a vehicle, Hatchback and Sedan at <$35,000, is much less than the median Salary at ~$60,000, which means that budget would not be a serious problem while making the purchase decision.

The major chunk of Salary lies between $50,000 and $80,000, while that of Total Salary lies between $60,000 to $100,000. With the average price of a SUV ~$55,000, price shouldn't be a problem for customers whose partners are also working.

1.2.3 Gender Distribution

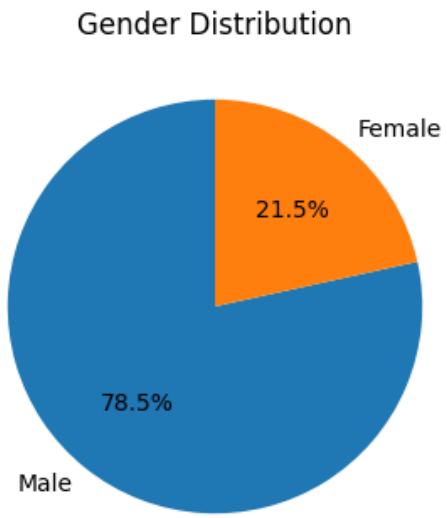


Fig. 10 - Distribution of Gender

Make Gender	Hatchback	SUV	Sedan
Female	4.6	52.6	42.9
Male	47.1	9.8	43.0
All	38.0	19.0	43.0

Table 1 - Choice of Car by Gender

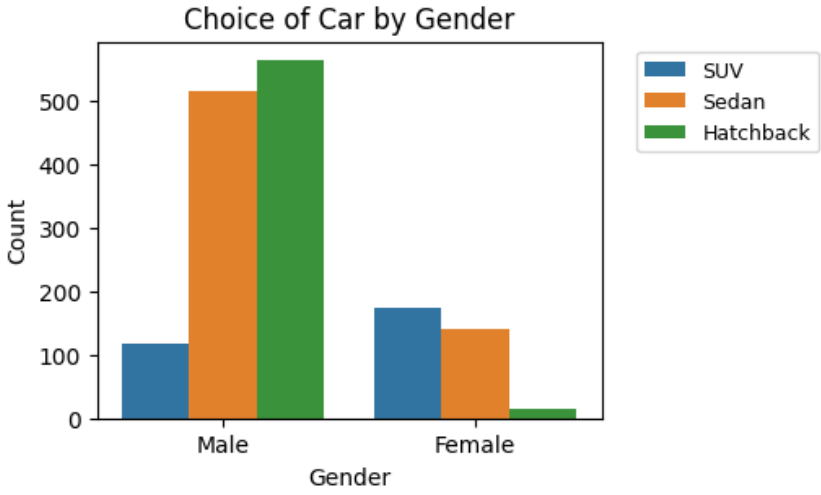


Fig. 11 - Choice of Car by Gender

Fig. 12 - Average Salary by Gender

A staggering 78.5% of customers are Male, however, females have a slightly higher mean salary. As per Table 1, 52% Females prefer SUVs and 47% males prefer Hatchbacks. Females don't like to buy Hatchbacks while Males don't like to buy SUVs. This would come as a surprise to majority of people because SUVs are seen as masculine.

1.2.4 Marital Status

An overwhelming majority, >90%, of customers are Married, which makes metrics related to Single customers quite redundant. For example, single people are choosing Hatchbacks much more than any other car, but that doesn't matter because there are very few

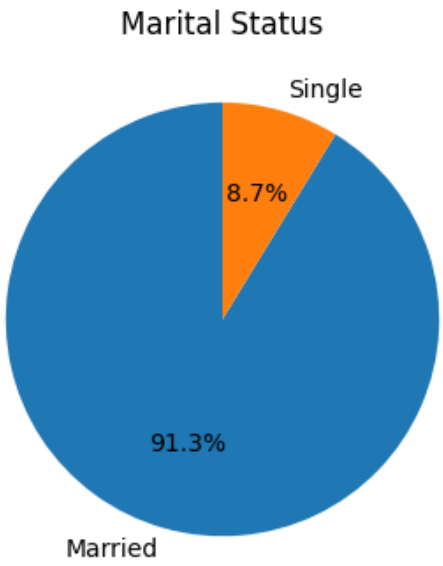


Fig. 13 - Distribution of Marital Status

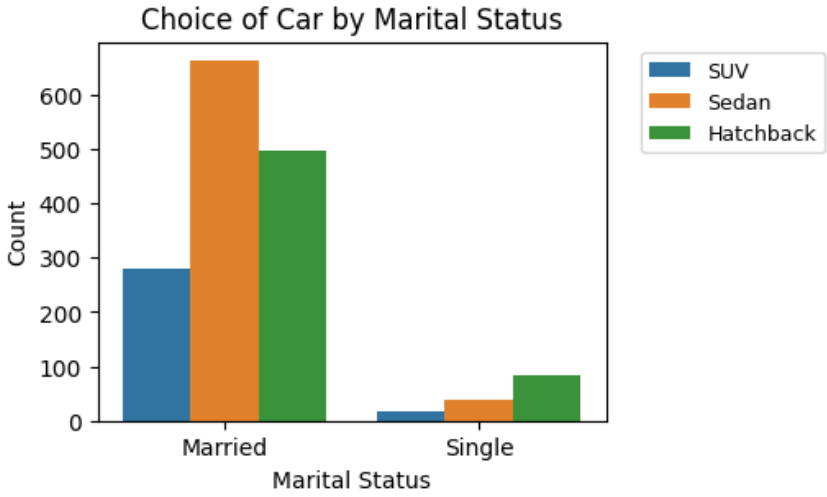


Fig. 14 - Choice of Car by Marital Status

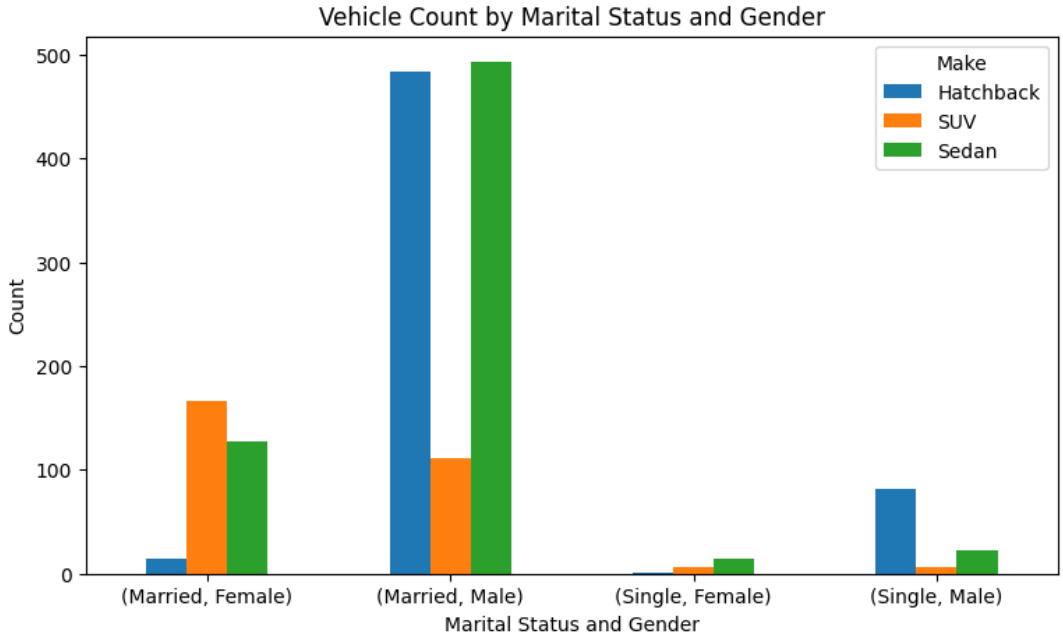


Fig. 15 - Vehicle Count by Marital Status and Gender

of them. What is making an impact is that most of the married folks are choosing Sedan followed by Hatchbacks, and only 20% are choosing SUVs.

1.2.5 Education Level

More than 60% customers are Post Graduates, and they are tending to choose Sedans followed by Hatchback. Graduates are tending to choose Sedans as well followed by Hatchback. This pattern is repeated when education level is divided by Genders.

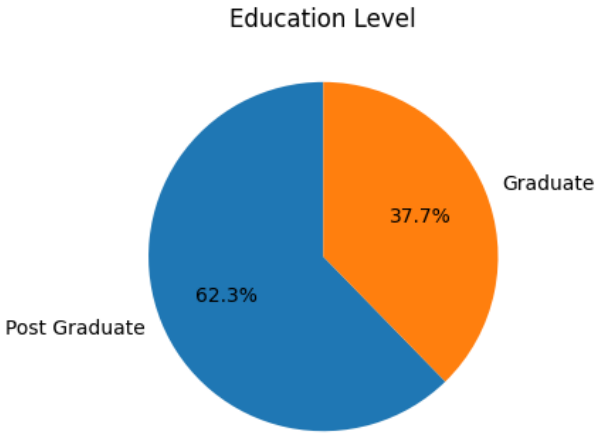


Fig. 16 - Distribution of Education Level

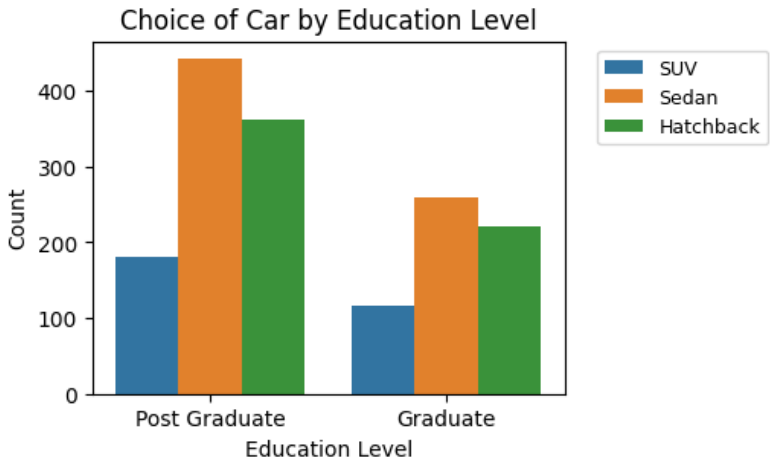


Fig. 17 - Choice of Car by Education Level

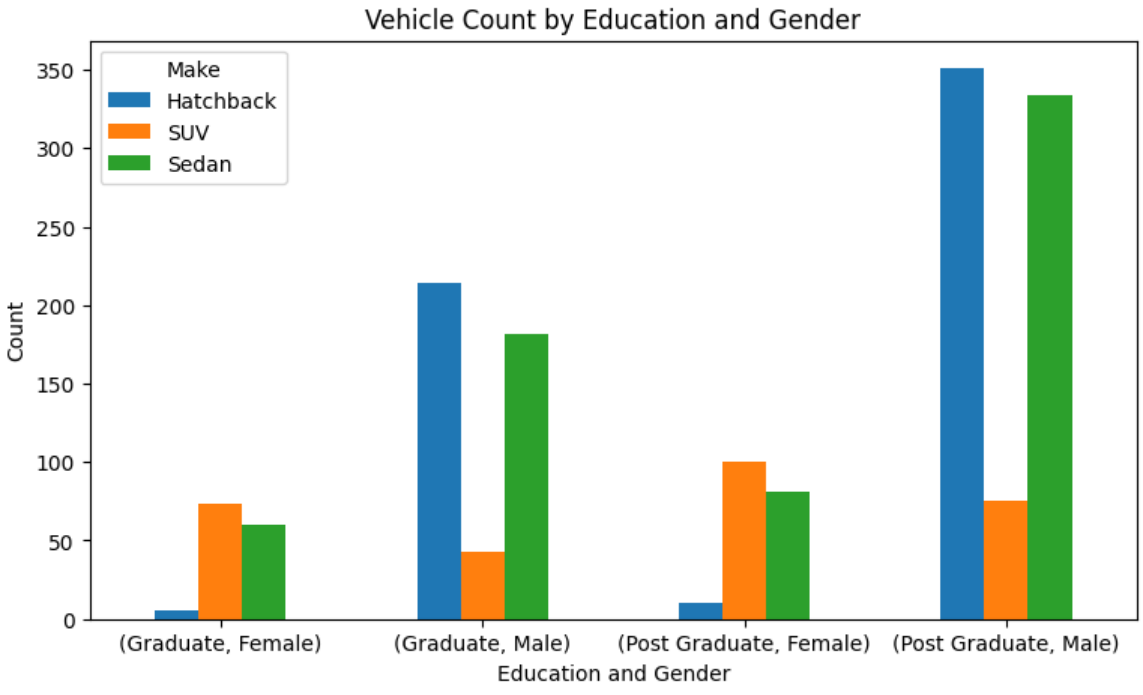


Fig. 18 - Choice of Car by Education Level and Gender

1.2.6 Number of Dependents

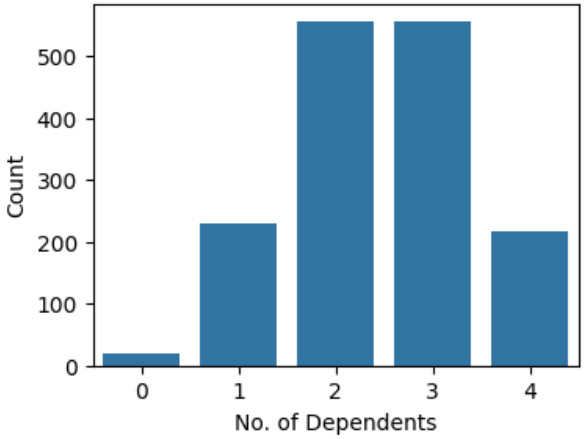


Fig. 19 - Distribution of No. Of Dependents

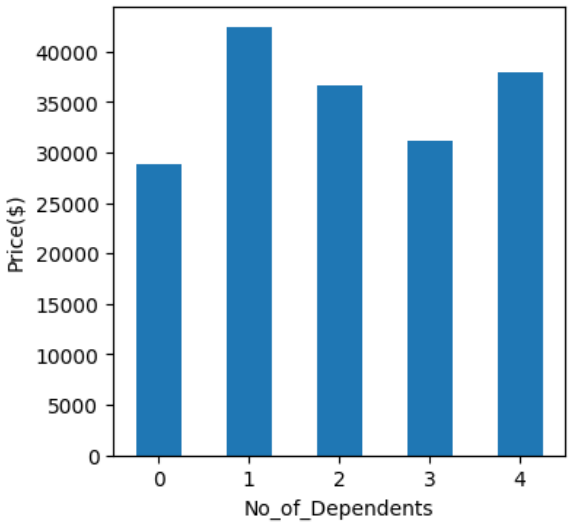


Fig. 21a - Average vehicle price by number of dependents

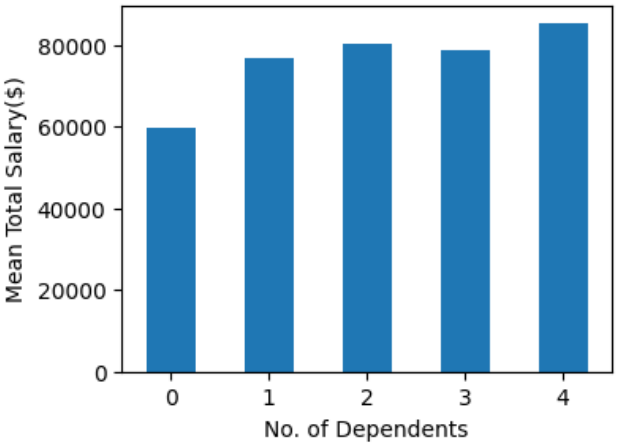


Fig. 20 - Mean Total Salary by No. Of Dependents

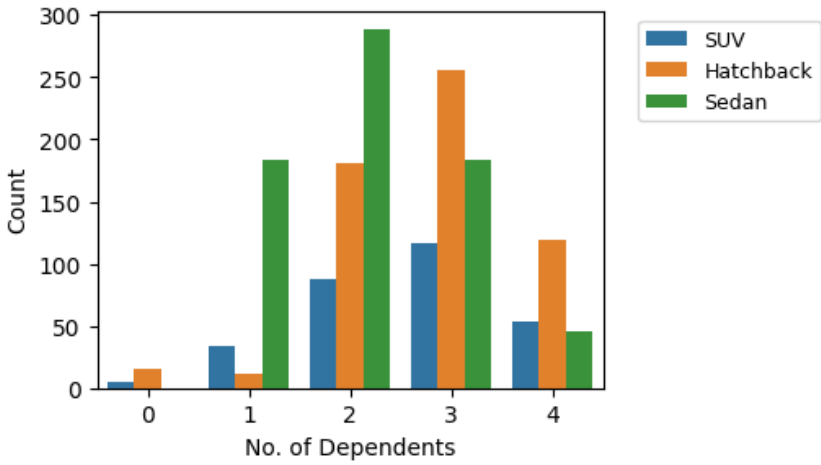


Fig. 21 - Choice of Car by No. Of Dependents

Majority customers are having 2 to 3 dependents, and there's hardly anyone with no dependants. Mean Total Salary increases from $60,000 to $80,000 for a span of 0 to 4 dependents, which means that people with dependents will have most purchasing power for more expensive vehicles. As for the choice of make we see that with 1 dependent, customers tend to choose

Sedans almost exclusively. Hatchback is the primary choice with 3 dependents. Sedans are more popular with 1,2, and 3 dependents, while Hatchbacks are more popular with 2,3, and 4 dependents. Also, we see that most money spent on a vehicle is by people with 1 dependent.

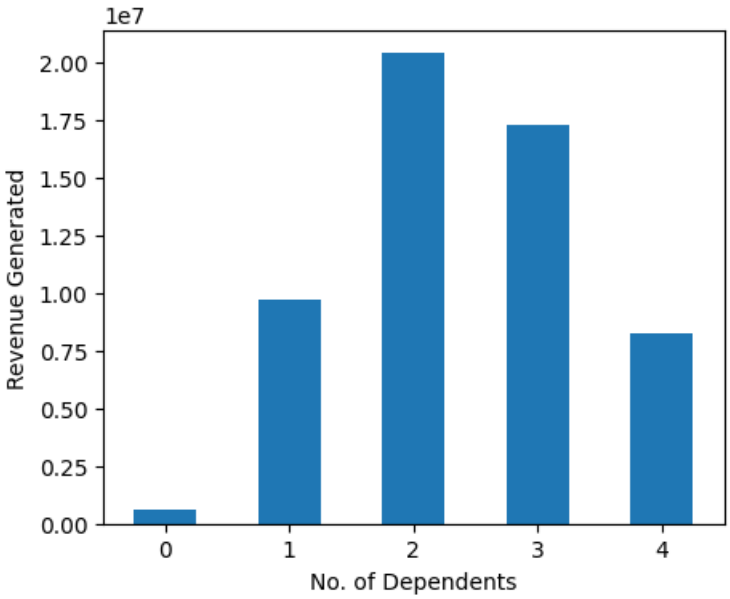


Fig. 21b - Revenue generated by No. Of Dependents

1.2.7 Personal Loan Takers

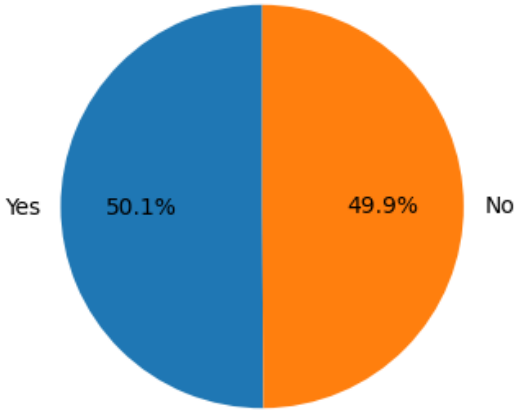


Fig. 22 - Personal Loan Taken

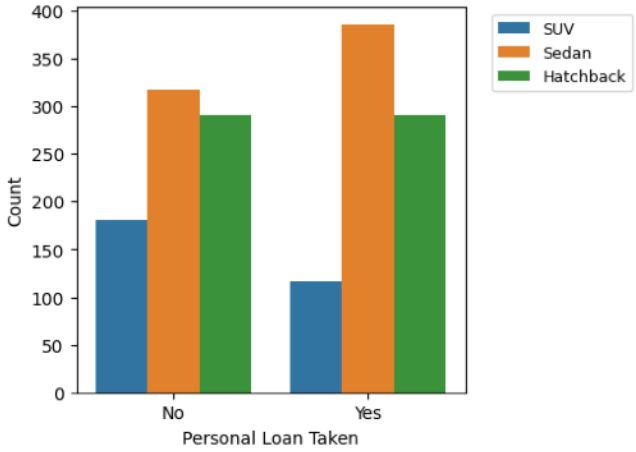


Fig. 23 - Choice of car by Personal loan taken

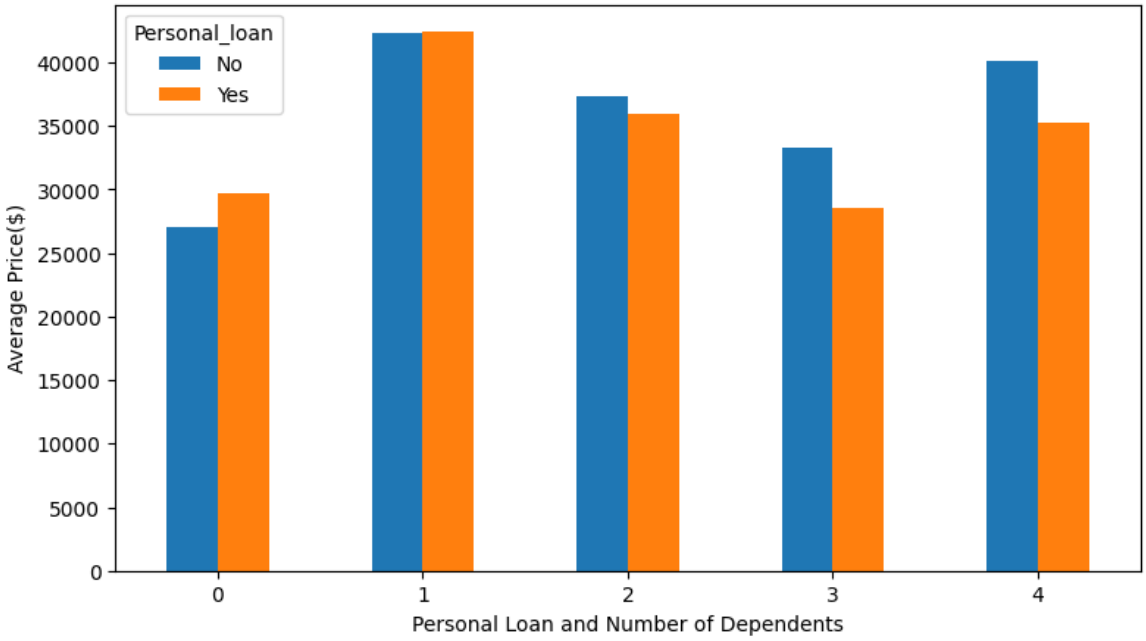


Fig. 24 - Average Car Price by Personal Loan Taken and No. Of Dependents

1 in 2 customers have taken a personal loan. Purchase of vehicles is almost in the same proportions for both groups - Sedans are the highest and SUVs the lowest. The average price of vehicles reduces slightly for those who have

taken a personal loan and have 3 or 4 dependents. This means that the loan is not having a big impact on vehicle affordability.

1.2.8 House Loan Takers

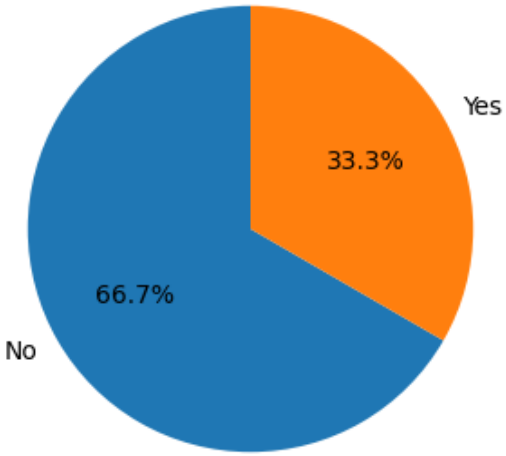


Fig. 25 - House Loan Taken

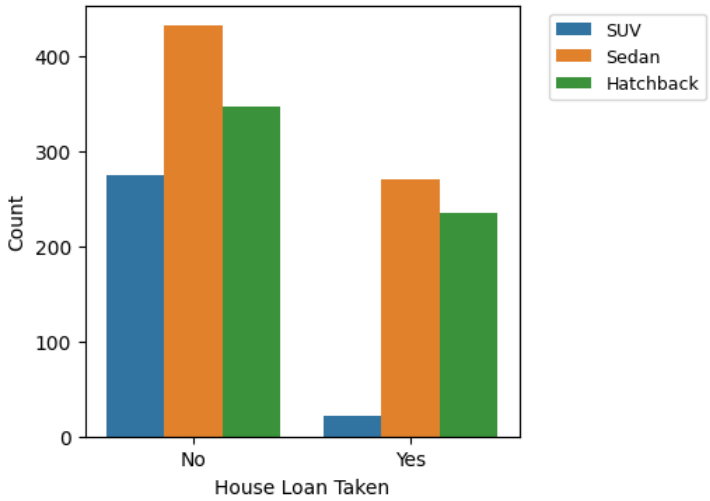


Fig. 26 - Choice of car by house loan taken

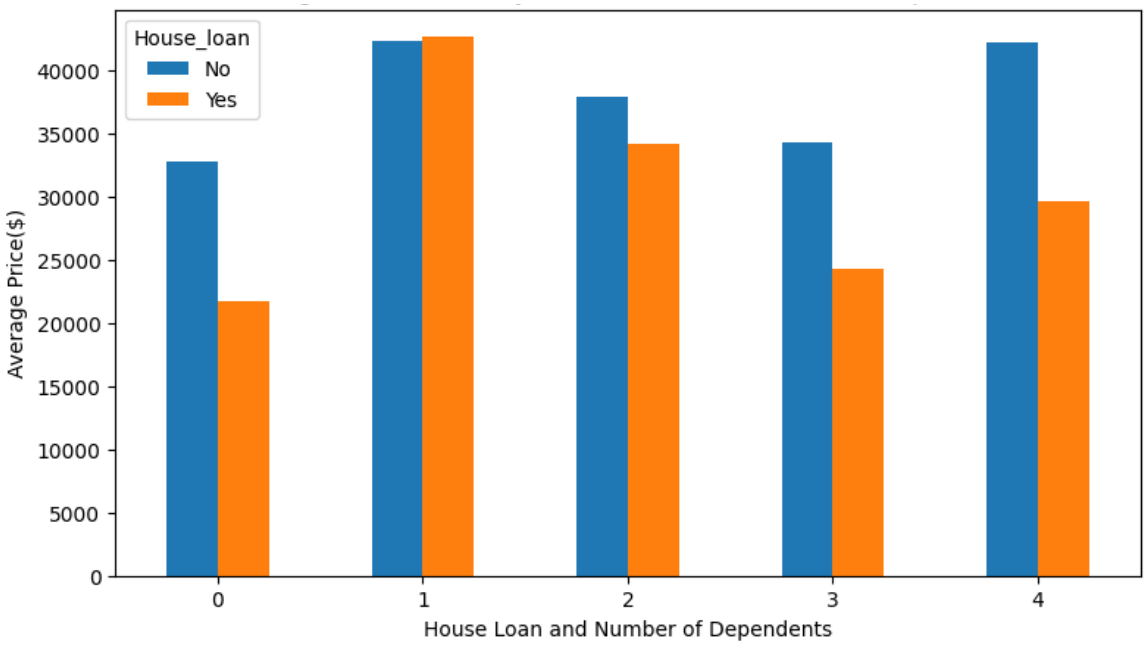


Fig. 27 - Average price by House loan taken and No. Of Dependents

Almost exactly 2/3rd people have taken house loan. Taking house loan seems to be having a major impact on SUV purchase. All car purchase is affected by

house loans but SUVs seems to be the most effected. House loan also has an effect when combined with number of dependents. Average price of car purchased reduces even without any dependents and then quite a bit with 3 or 4 dependents. Those who have taken both personal and house loan, the average price is impacted but not that much.

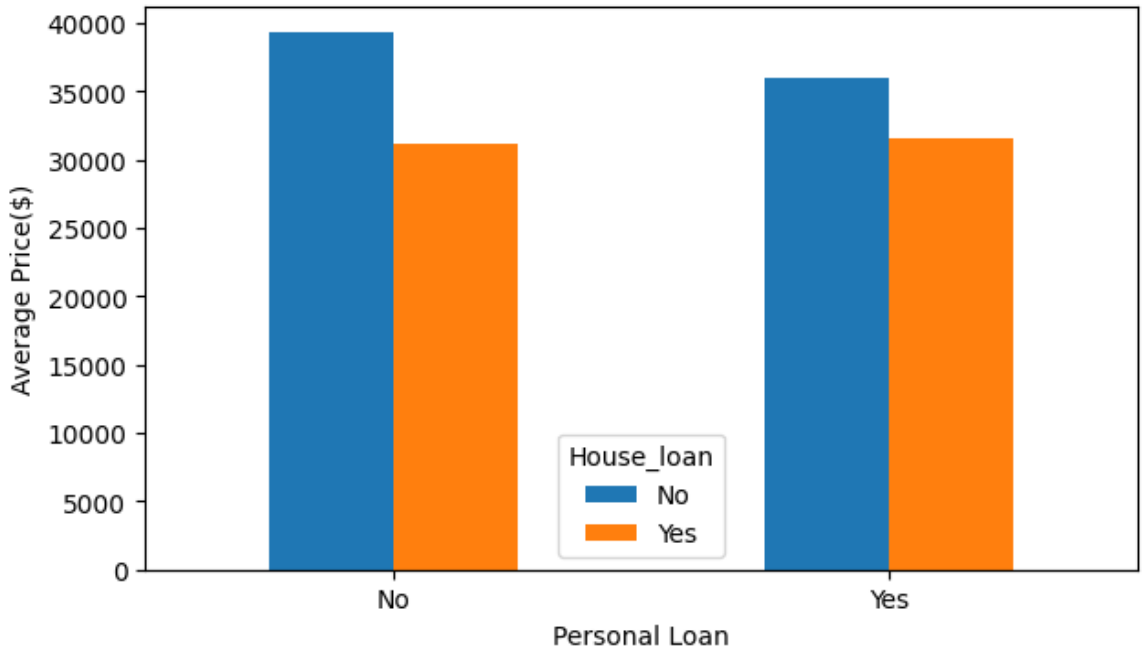


Fig. 28 - Average price by House loan and Personal Loan taken

1.2.9 Partner Working

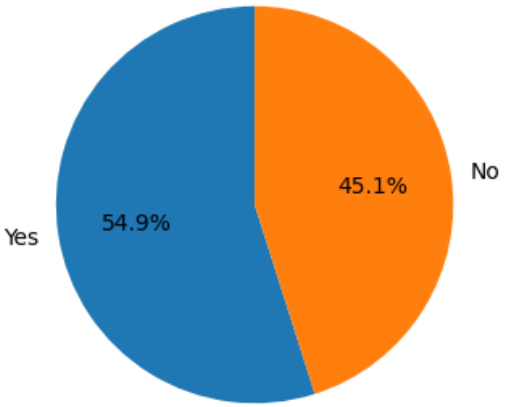


Fig. 29 - Partner Working

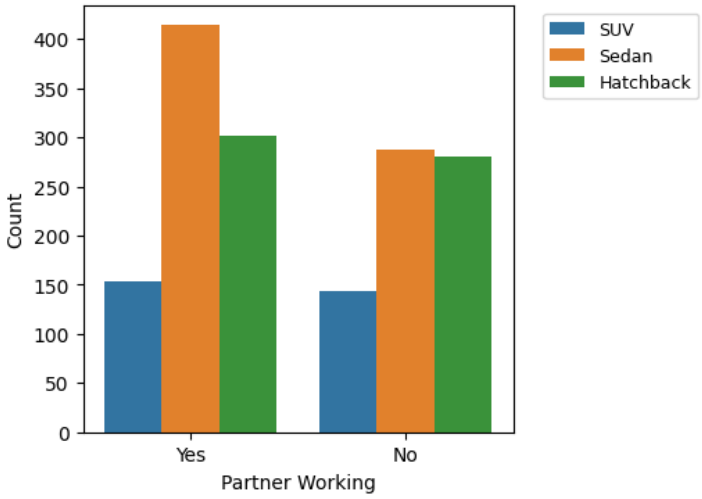


Fig. 30 - Choice of make by Partner working

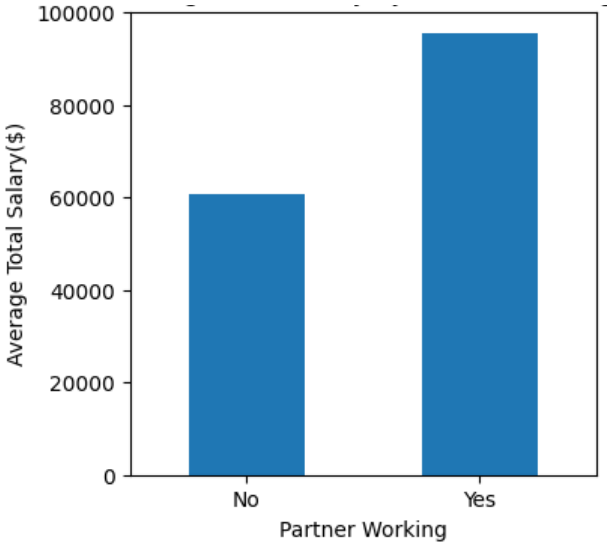


Fig. 31 - Average Total Salary by Partner Working

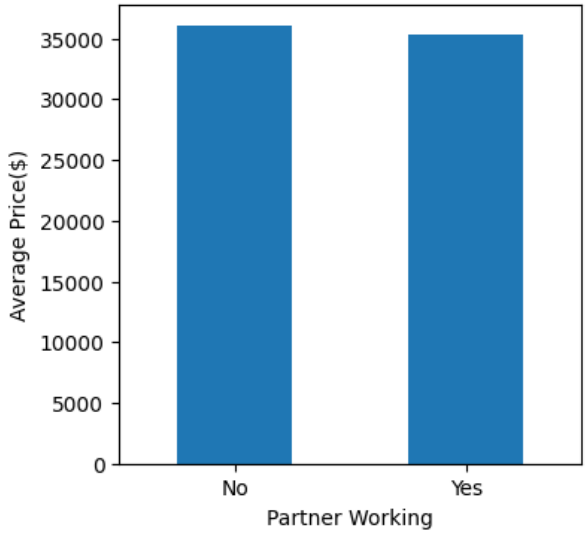


Fig. 32 - Average price of car by partner working

Over 50% of customer's partners are working. In this case customers tend to choose more Sedans. Other makes are not impacted. Also, average total salary increases quite a bit with partner working, but at the same time average car price is reducing. That means customers are not spending on expensive cars even when they have more money.

1.2.10 Average Vehicle Price

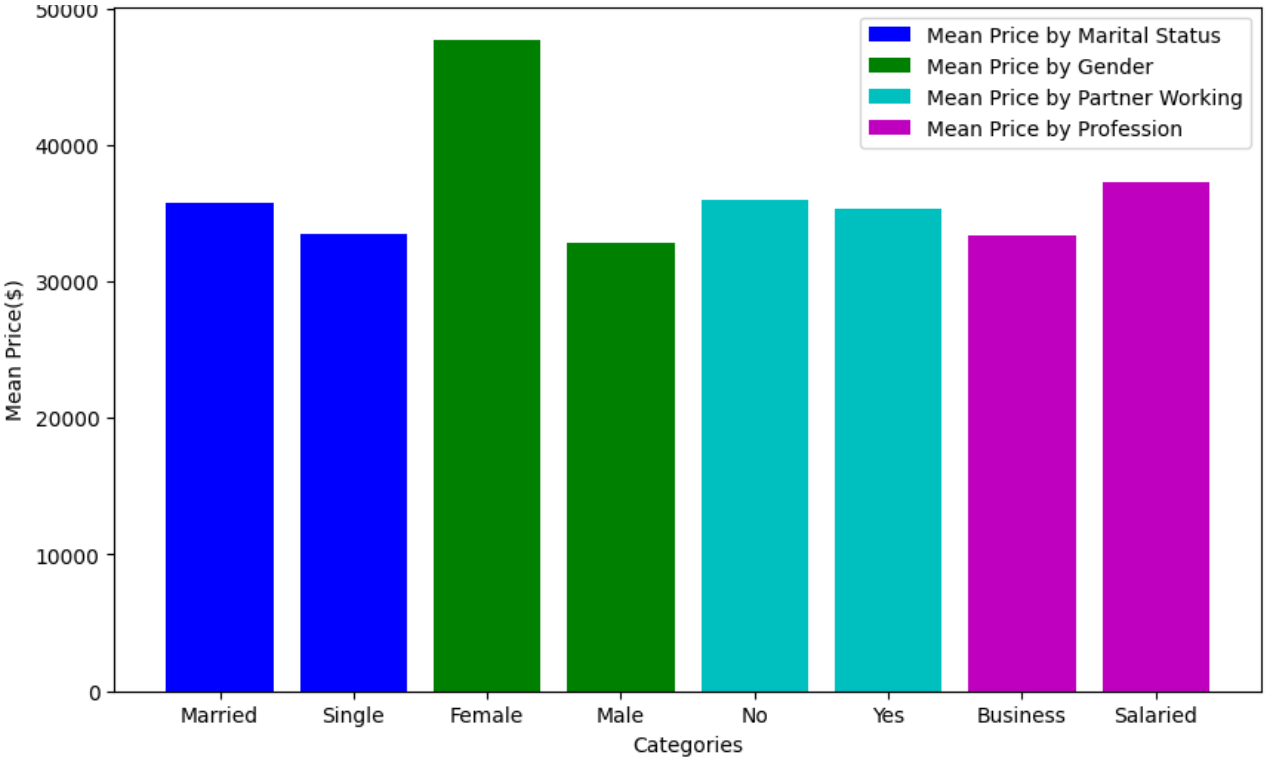


Fig. 33 - Factors effecting average purchase price

Here we take a look at what is affecting the average purchase price of a vehicle. Females seem to be spending most on vehicle purchase, while being married is not having a major impact on this statistic. Salaried people are also tending to spend a little more than businesspeople, and graduates the same over post graduates.

1.2.11 Age

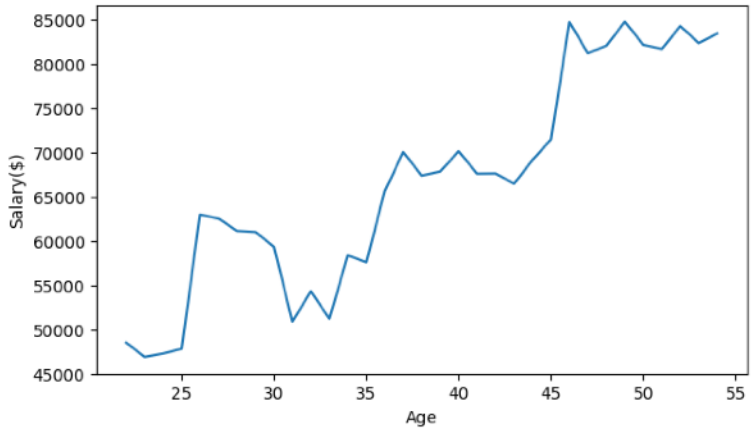


Fig. 37 - Age vs Salary

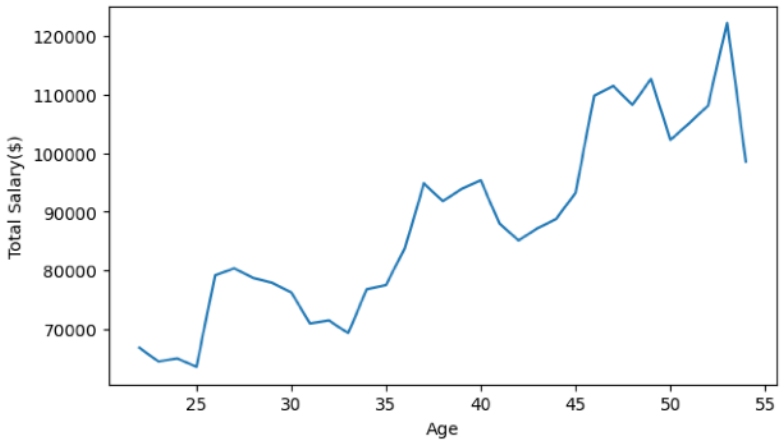


Fig. 38 - Age vs Total Salary

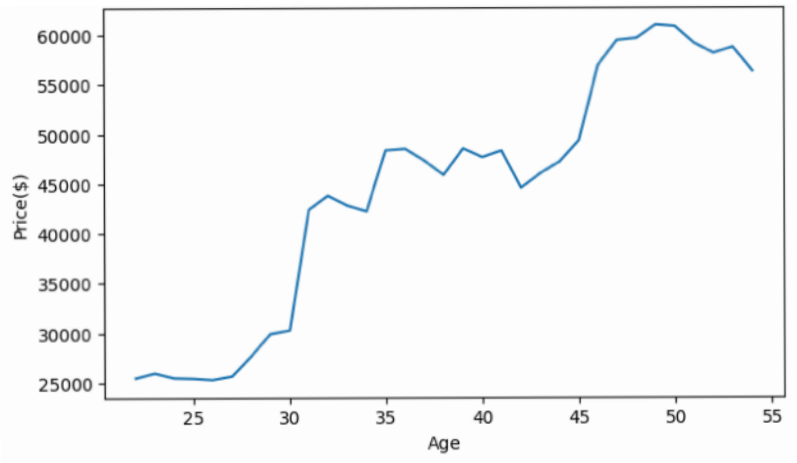


Fig. 39 - Age vs Average Price of Vehicle

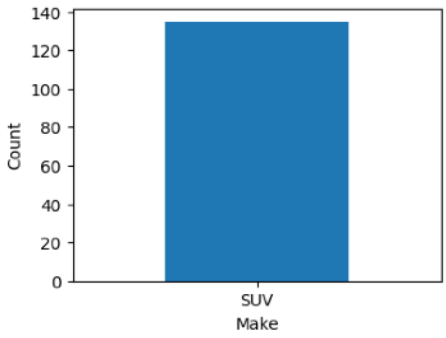


Fig. 39a - Make preferred by people over 45

Both Salary and Total Salary seem to be increasing with Age, and so does the average amount spent on vehicles. Where a 30 year old is spending $30,000 on a vehicle, a 50+ year old is spending almost $60,000!

1.2.12 Numerical Variable Correlation

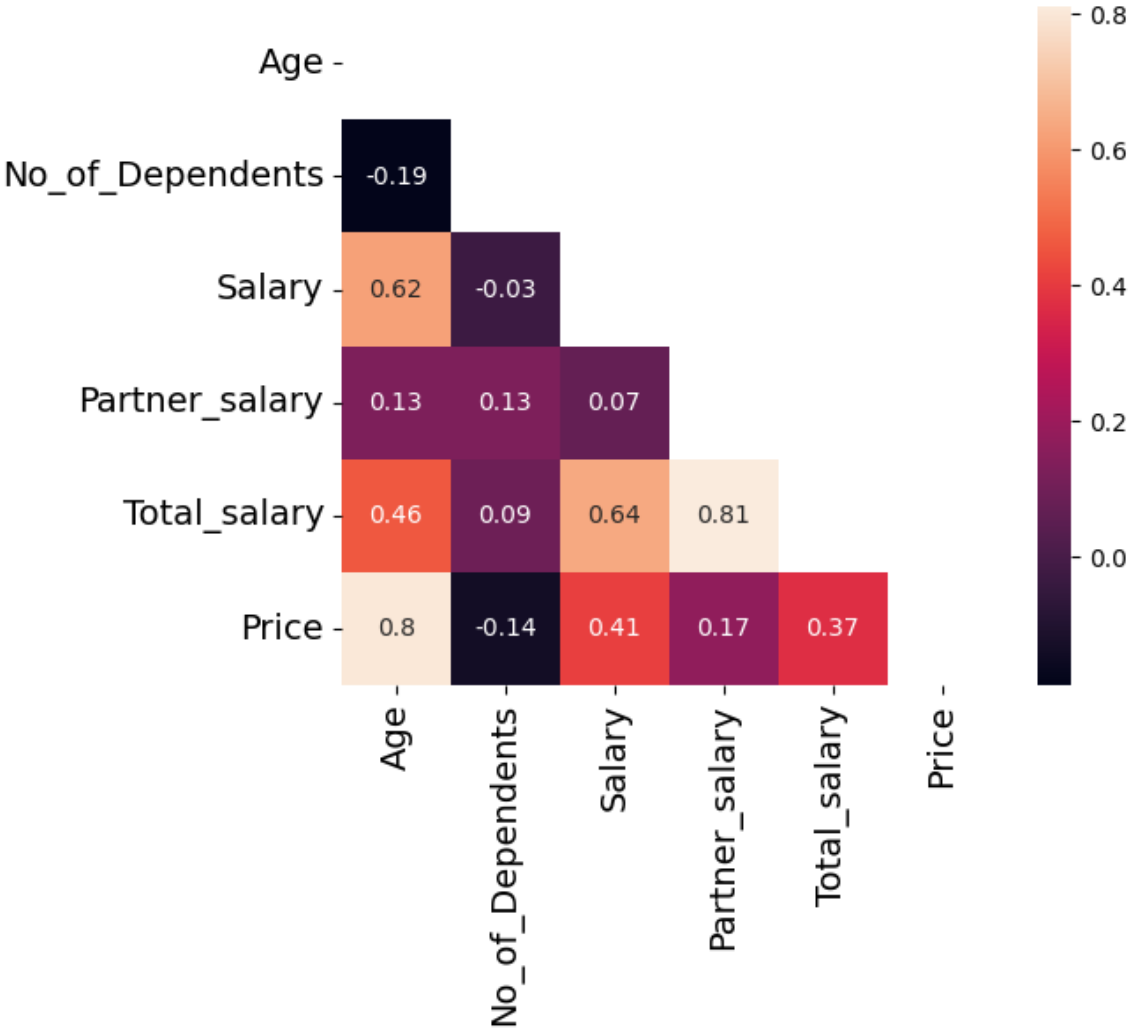


Fig. 40 - Heatmap of all Numerical variables

Two sets of variables seem to be most correlated - Age and Price, and Partner salary and total salary. Salary and age also seem to have some correlation. So spend on a vehicle increases as age increases, and salary increases with age somewhat. But total salary and price are not correlated, which means that

even though a family is earning more, they are not spending that money on vehicles.

1.2.13 Scatter Plots of Average Price

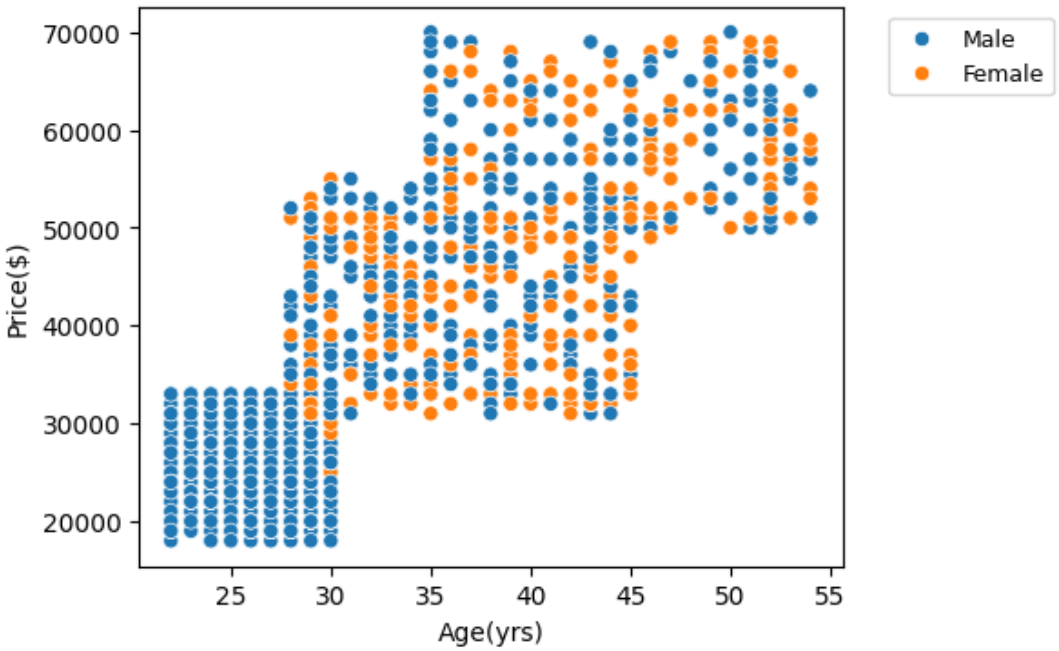


Fig. 41 - Age vs Price, differentiated by Gender

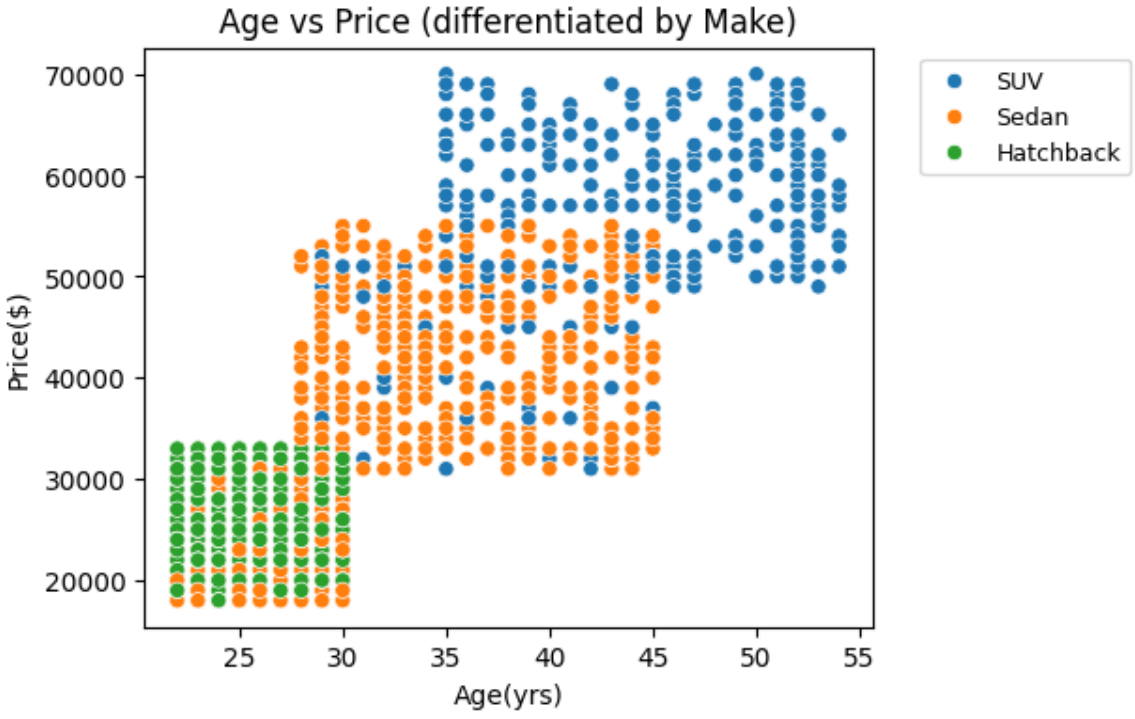


Fig. 42 - Age vs average Price, differentiated by Make

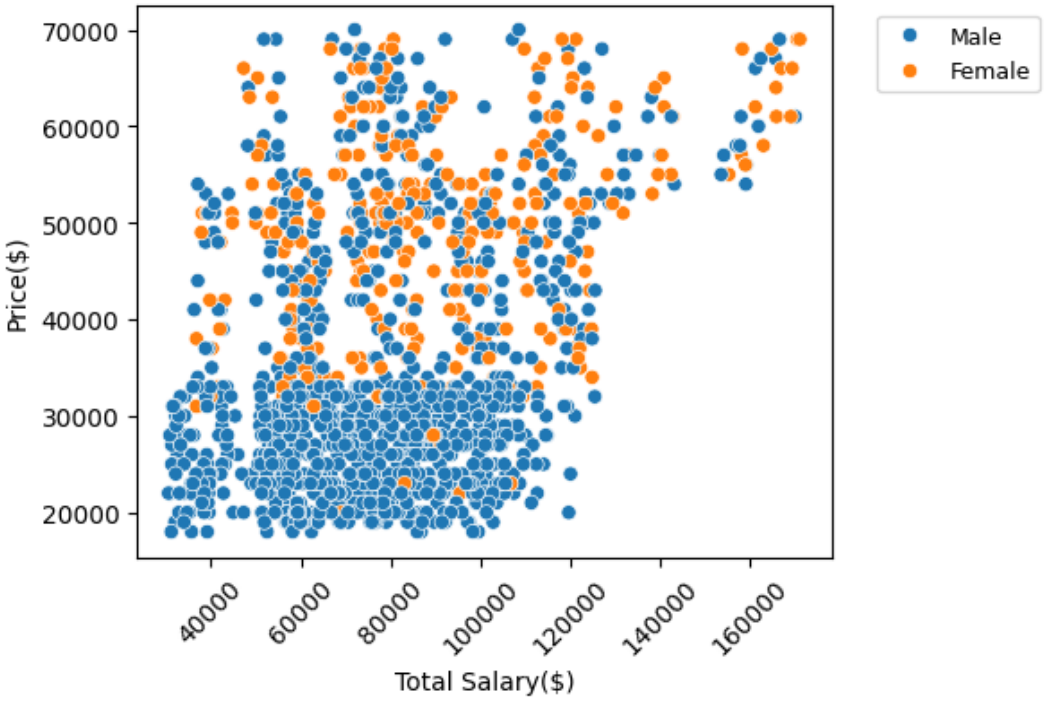


Fig. 43 - Total Salary vs Price, differentiated by Gender

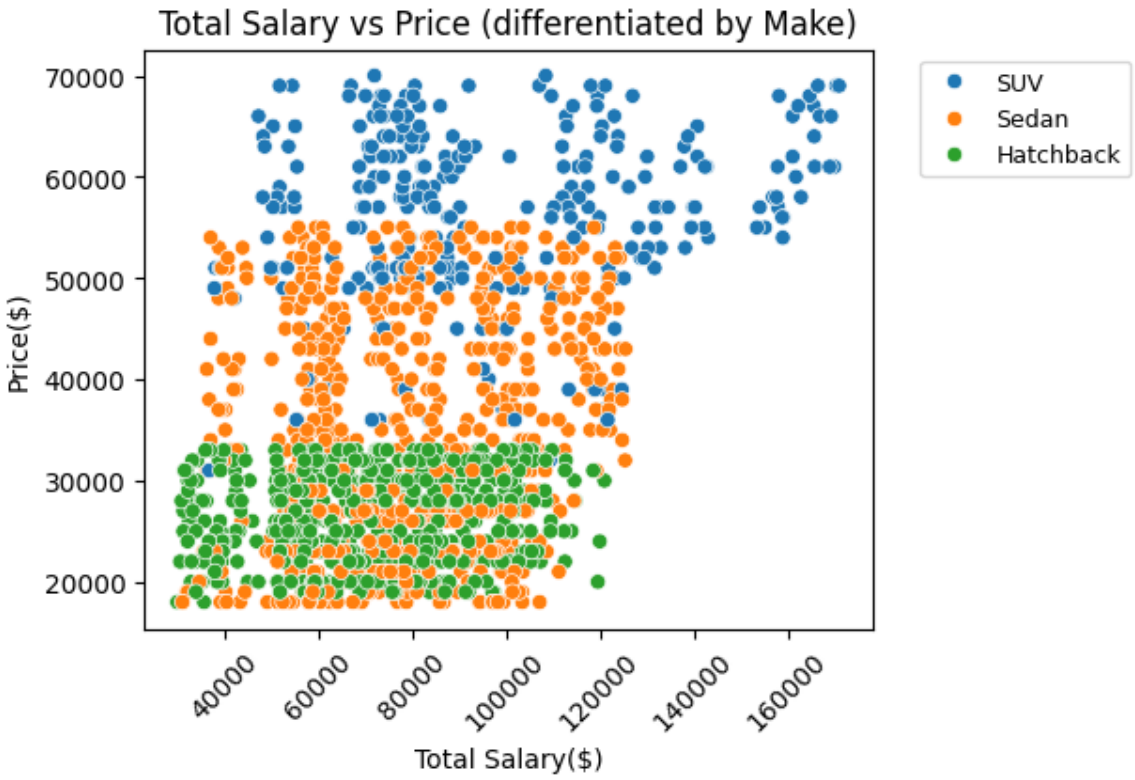


Fig. 44 - Total Salary vs Price, differentiated by Make

We see that customers below the age of 30 are all Male. Females are not buying cars below the $30,000 mark. Males till the age of 30-31 are not buying cars above $35,000.
Those Males with total salary of $100,000 and below are buying cars below $35,000. People with total salary >$120,000 are not buying vehicles below $50,000 mark. Females are mostly buying above $30,000.

1.3 Key Questions

1.3.1 Do men tend to prefer SUVs more compared to women?

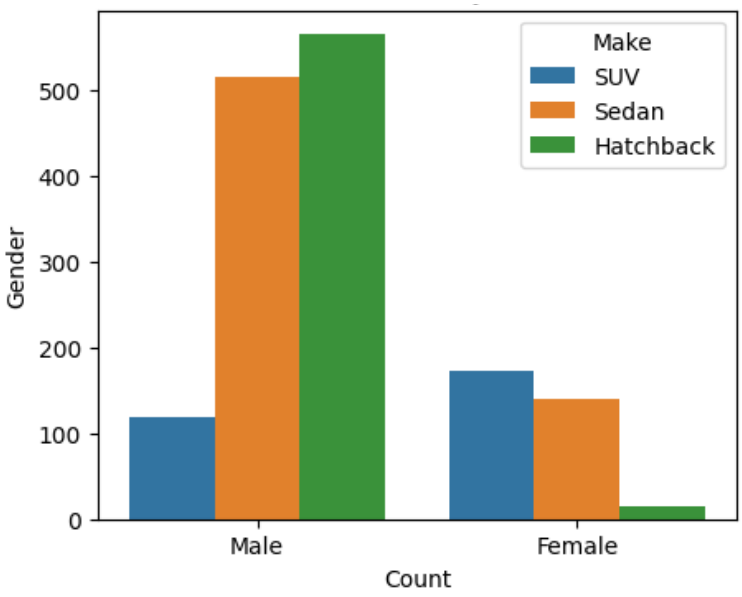


Fig. 42 - Choice of make by Gender

Make	Hatchback	SUV	Sedan
Gender			
Female	4.6	52.6	42.9
Male	47.1	9.8	43.0
All	38.0	19.0	43.0

Fig. 43 - Choice of make by Gender - Normalized

Interestingly, women are preferring SUVs slightly more than men. While men have purchased ~110 SUVs, women have purchased ~150. While 52.6% of women prefer to buy SUVs, only 9.8% men prefer the same.

1.3.2 What is the likelihood of a salaried person buying a Sedan?

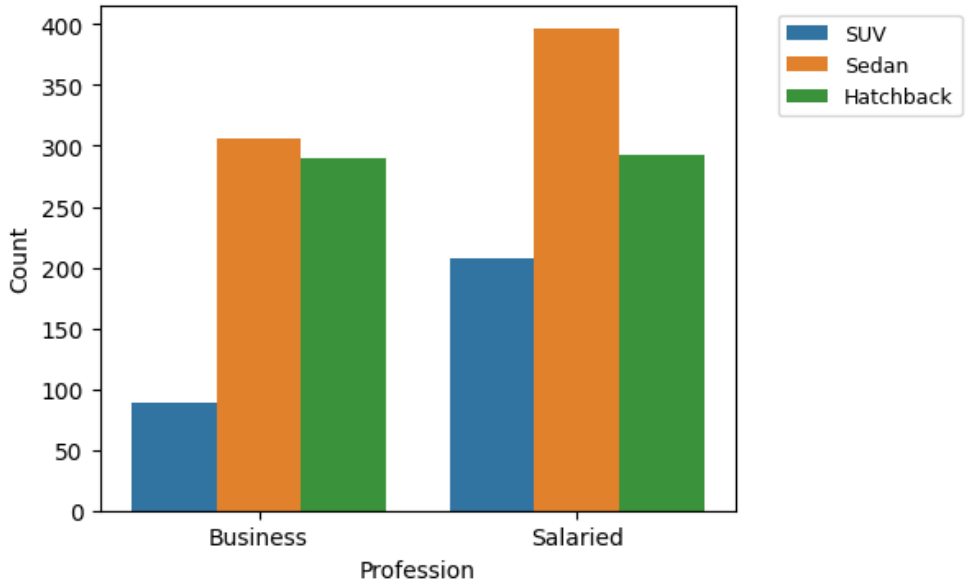


Fig. 43 - Choice of make by Profession

Number of Sedans purchased by Salaried people are 396, and total number of Sedans purchased are 702, which makes the likelihood of a Salaried person buying a Sedan 396/702 = 56.4%.

1.3.3 Sheldon Cooper's claim...

What evidence or data supports Sheldon Cooper's claim that a salaried male is an easier target for a SUV sale over a Sedan sale?

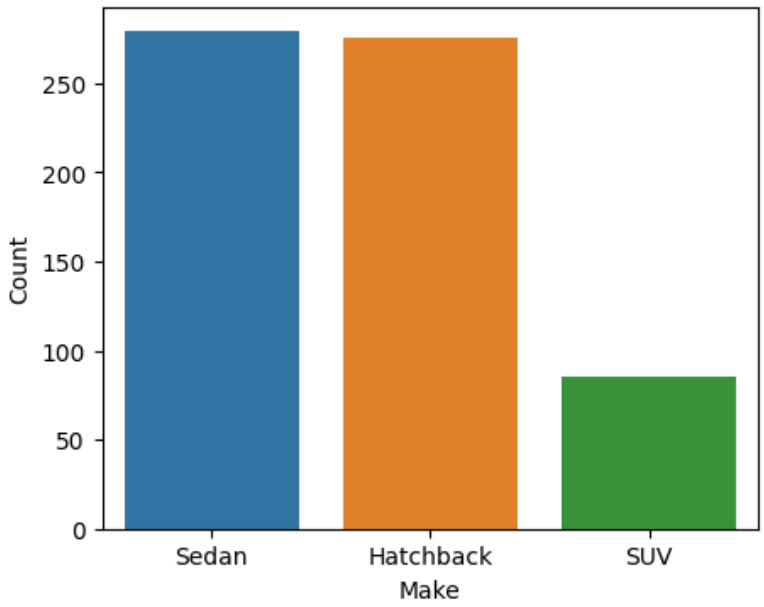


Fig. 44 - Choice of make of a Salaried Male

Clearly, salaried males are more inclined to buy Sedans, not SUVs.

1.3.4 Total amount by gender...

How does the the amount spent on purchasing automobiles vary by gender?

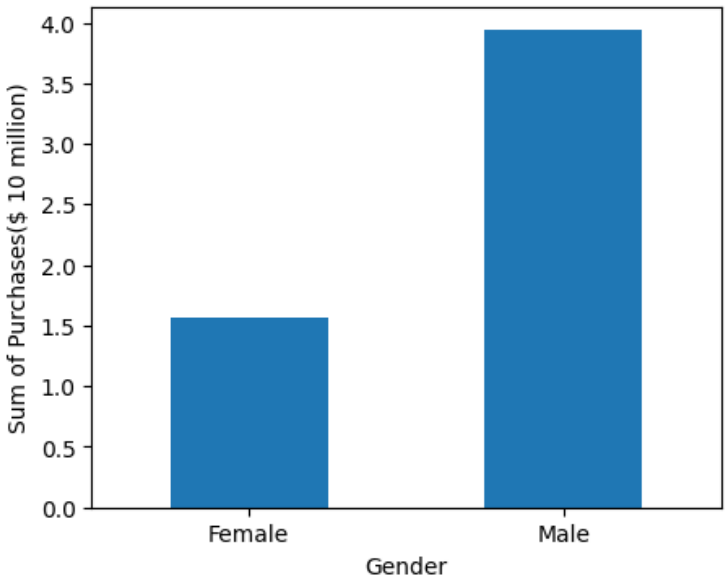
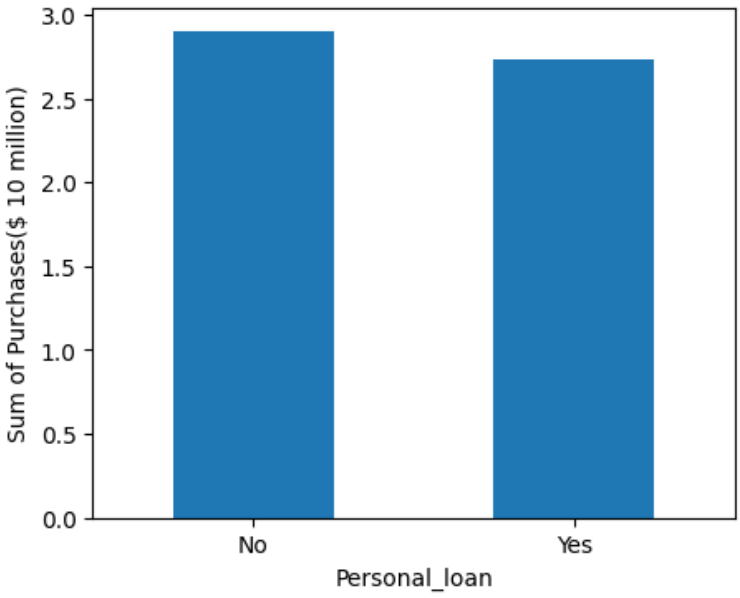


Fig. 45 - Total purchasing by Gender

Total Male spending on purchasing vehicles is more than twice than that by Females.

1.3.5 Individuals who took a personal loan...

How much money was spent on purchasing automobiles by individuals who took a personal loan?



Total money spent was close to $26 million.

Fig. 46 - Total Purchasing by Personal Loan Taken

1.3.6 Influence of Working Partner...

How does having a working partner influence the purchase of higher-priced cars?

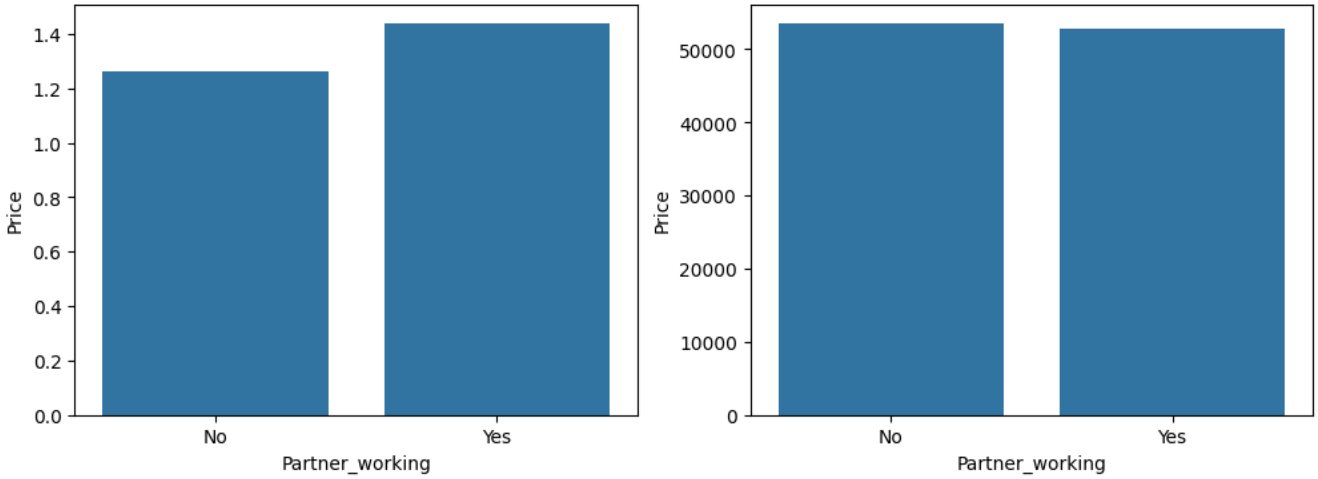


Fig. 47 & 48 - Higher priced cars Sum and Mean respectively

We are defining Higher Price as cars which cost >$40,000. The total amount spent by customers with partner working on higher priced cars is higher than not working, but the average price is slightly less. So having a working partner is actually not influencing purchase of higher cars a lot. Neither are they buying expensive cars, nor spending a large sum in total.

1.4 Actionable Insights and Business Recommendations

1.4.1 Points

- 60.6% customers are below 31 years of age but produce only 45.7% of total revenue. Out of these 91% are males and are choosing Sedans and Hatchbacks but not SUVs. Therefore **do not advertise SUVs to young males.**
- Majority total customers are Male at 78.5% and produce 69.9% of revenue. Therefore marketing needs to be geared more toward males.
- SUV sales are less as compared to Sedan and Hatchback but their total revenue generated is equal to Hatchbacks, so they can't be ignored. Number of SUVs sold are 297, which equals 18.8% of sales but they have generated $16.58 million in revenue which is 29.5% of revenue.
- Premium segment (price > $40,000) sold 508 units which is 32% of vehicles but generated $27.0 million revenue, which is 48% of revenue. The composition of premium segment is 50% SUVs and 50% Sedans, but no Hatchbacks.
- Males prefer Sedans and Hatchbacks and women prefer SUVs and Sedans.
- Most customers are married (>90%) with 2-3 dependents. Married males are the majority of customers and they prefer Sedans and Hatchbacks, not SUVs. So **do not market SUVs to married males.**
- Those with 2 dependents have generated the most revenue for the company and prefer Sedans, followed by customers with 3 dependents who prefer Hatchbacks. Customers with 4 dependents

also prefer Hatchbacks, which means that **Hatchbacks can be marketed to larger families with 4-5 people.**

- Customers with 2 dependents are choosing Sedans in a wide range from $18-$55,000, therefore **for Sedans smaller families need to be targeted both for the base customer and the premium segment.**
- Those with 1 dependent spend most on a single vehicle on an average and clearly prefer Sedans with the average price of $41,000. Therefore **Sedans in the range $30-$50,000 can be marketed to couples and single child families.**
- As number of dependents increases, total salary is also increasing, and since customers with 2-4 dependents are liking hatchbacks, **more expensive Hatchbacks can be marketed to larger families, or the company can try to create more premium Hatchbacks.**
- When partner is working, the customer has higher income, but the average price of a car purchased is low. Also, with a partner working, customers are tending to choose Sedans. Therefore, **higher priced Sedans can be marketed so as to appeal to working couples.**
- Females have a bit higher Salary and are spending more than Males for a single car on an average, and they prefer SUVs. Therefore, **one portion of marketing SUVs can be exclusively for females.**
- Salary and Total Salary are increasing with Age, and so is the average spend on a car. People aged 45 and above are exclusively preferring SUVs and they are a mix of males and females. Therefore, **SUVs can be targeted towards higher aged folks.**
- Females are buying only when they are above 31-32 years of age and above $35,000 for a car on an average, and not buying hatchbacks.
- Till $35,000, only males are buying, and since most of them (68.8%) are married, **the entry level vehicle segment <$35,000 can be**

targeted towards married males, which consists of Sedans and Hatchbacks but no SUVs.

- Above $120,000 of total salary only SUVs are being bought. Therefore, **SUVs can be marketed exclusively to the high-end segment.**

1.4.2 In Sum

- **Sedans**
  - Smaller families need to be targeted both for the entry level and the premium segment.
  - Sedans in the range $30-$50,000 can be marketed to couples and single child families.
  - Higher priced Sedans can be marketed so as to appeal to working couples.
  - The entry level vehicle segment <$35,000 can be targeted towards married males, which consists of Sedans and Hatchbacks but no SUVs.
- **Hatchbacks**
  - Can be marketed to larger families with 4-5 people.
  - More expensive Hatchbacks can be marketed to larger families, or the company can try to create more premium Hatchbacks.
- **SUVs**
  - One portion of marketing SUVs can be exclusively for females.
  - Can be targeted towards higher aged people.
  - Can be marketed exclusively to the high-end segment.

2.0 Problem 2 - Godigt Bank CC

Analyse the dataset and list down the top 5 important variables, along with the business justifications.

2.1 Networth (high_networth)

Networth categories tell us the distinctions between the customers. People from A category are much less likely to default than people in E category, and that would affect the attrition.

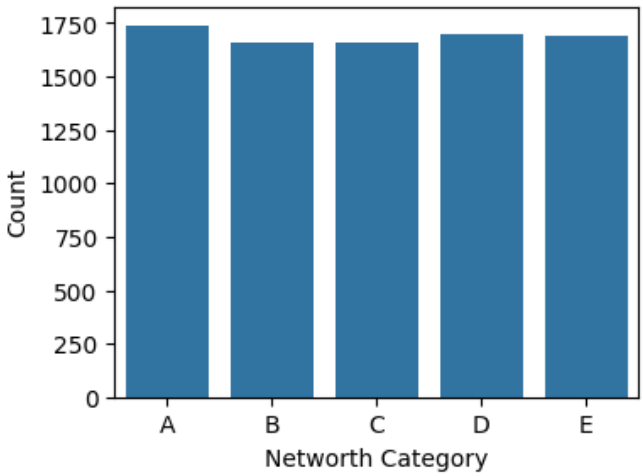


Fig. 49 - Count of customers by Networth category

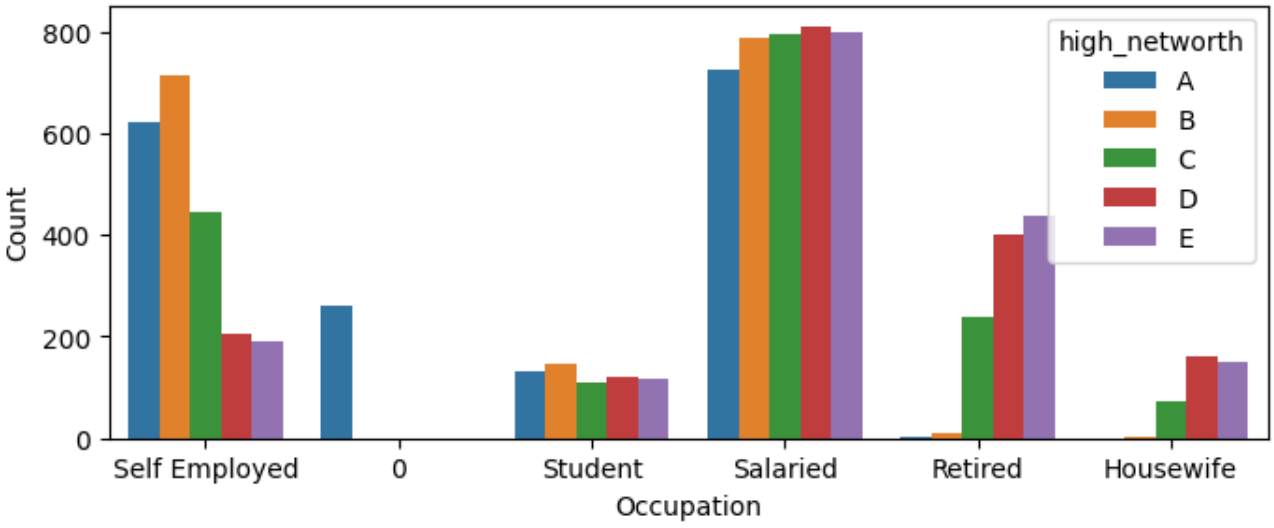


Fig. 50 - Count of Occupation by Networth category

Fig. 49 tells us that number of customers in each networth category are more or less equal. In fig. 50 we can see that Self Employed occupation has predominantly A and B categories, while Retired subgroup has mostly the lower categories of D and E. Salaried subgroup is well spread out.

2.2 Annual Income (annual_income_at_source)

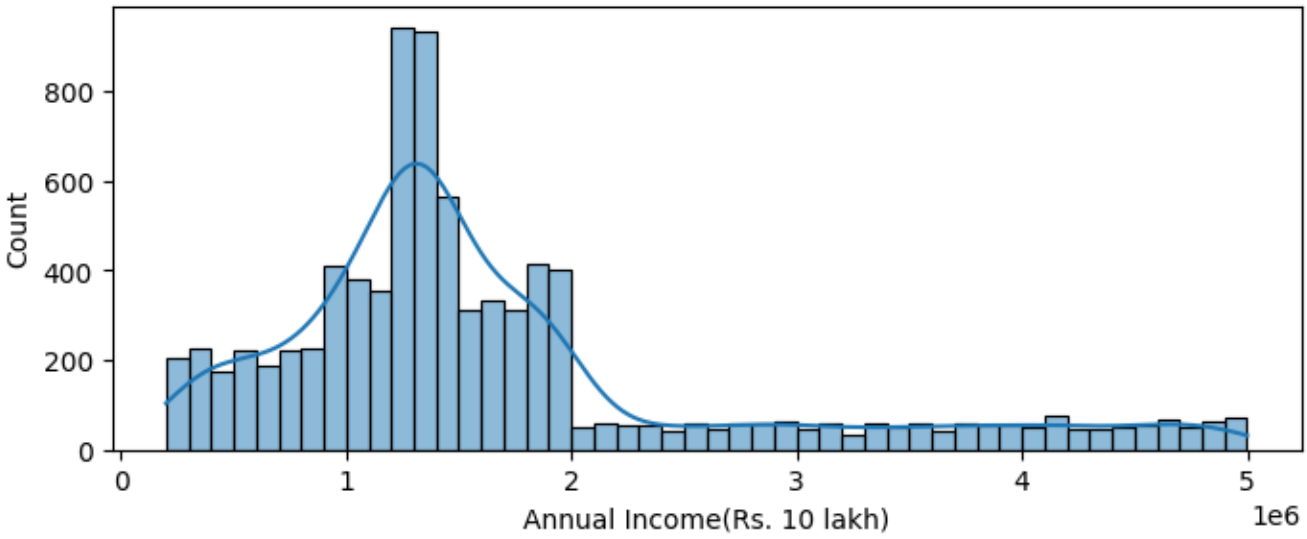


Fig. 51 - Histogram of Annual Income

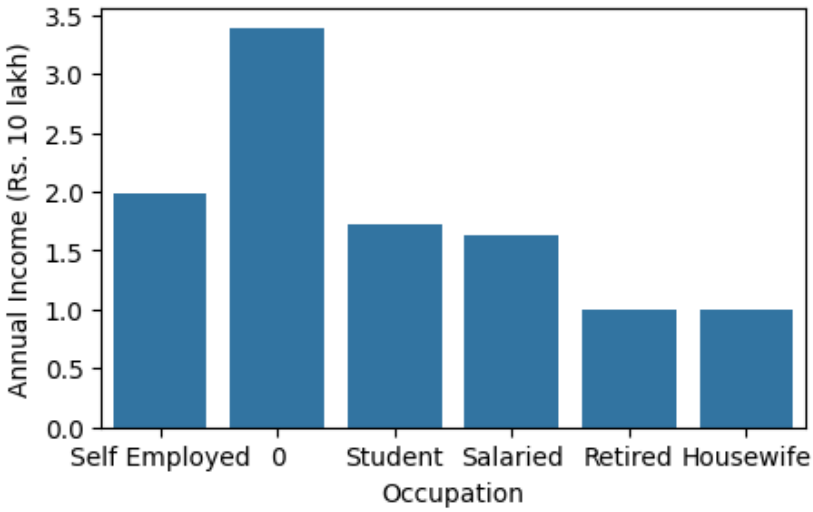


Fig. 52 - Annual Income by Occupation (0 is an anomaly)

Annual income would be critical while analysing any data related to credit cards. Here we see that while the annual income mean is at Rs. 16.7 lakh and median at 13.7 lakh, around 1/5th of customers are above Rs. 20 lakh. As for how it's divided by occupation, Self Employed was expected to have a high income but the Student subgroup surprisingly has an annual income more than Salaried people! In Fig. 53 we see that while categories C,D, and E have been given a narrow range of credit card limits, for categories A and B the range is quite large, esp category A. Category B at an annual income of 15-20L has been given the range1.5-5L, which category A with annual income ranging from 20-50L has been given the range 2-10L. In the upper reaches of these ranges, the probability of default increases and thus the attrition.

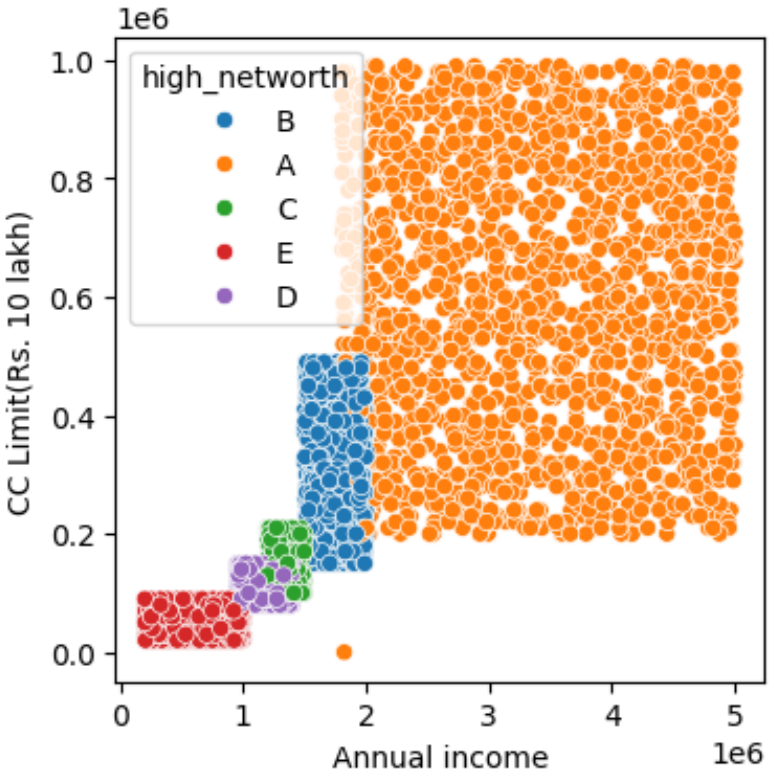


Fig. 53 - Annual income vs CC Limit, by Networth

2.3 Other Bank CC Holding (other_bank_cc_holding)

Fig. 54 - Bank Vintage by Annual Income, differentiated by Other bank CC holding

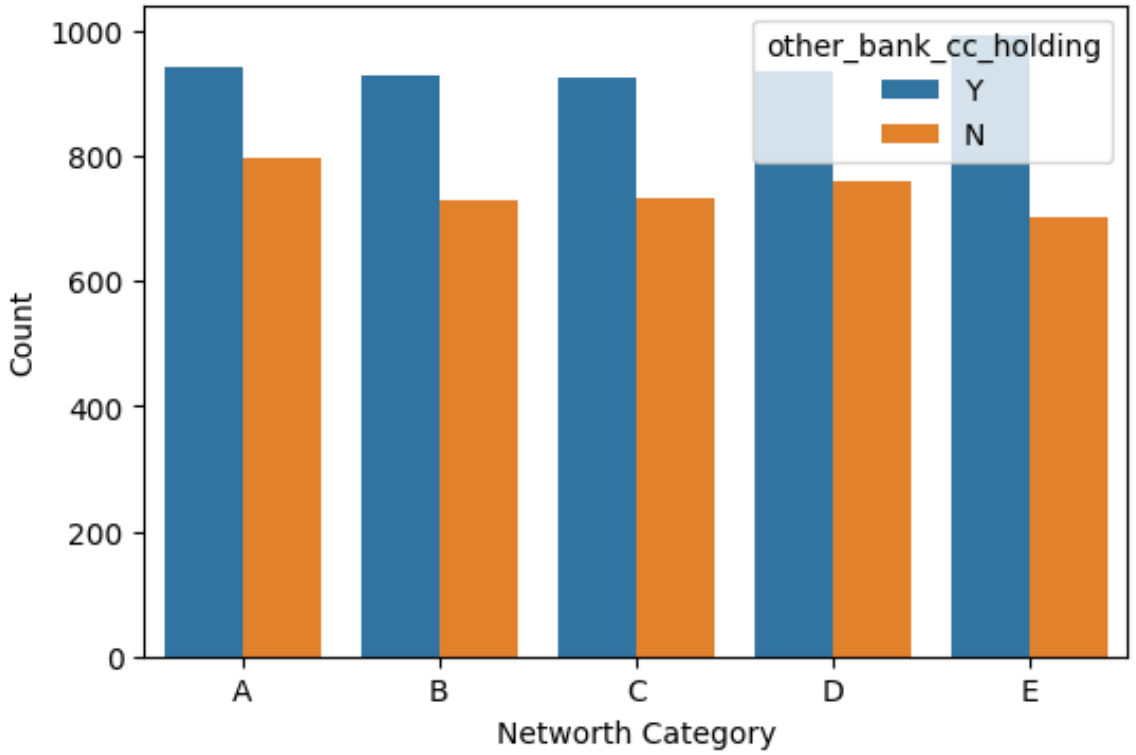


Fig. 54 - Count of Networth Category, divided by Other bank CC holding

Other bank CC holding is important because if customers are holding other bank's credit cards, they are more likely to leave our bank in case they start liking the policies of other bank more. Fig. 54 compares bank vintage with annual income, and we see that other bank CC holding doesn't have an effect on this. Also, as per Fig. 54, other bank CC holding is not effected by Networth category either.

2.4 Transactor Revolver (Transactor_revovler)

Transactor Revolver is important because transactors will be more prone to attrition since they are paying their dues in full just the next month.

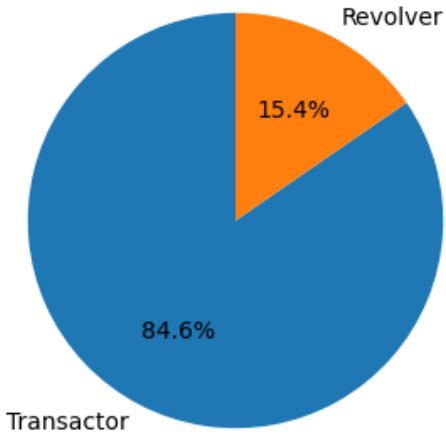


Fig. 55 - Distribution of Transactors and Revolvers

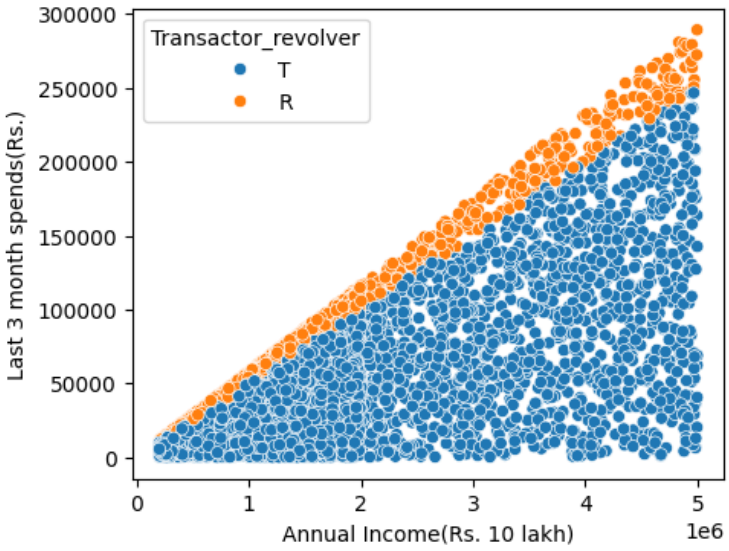


Fig. 56 - Last 3 month spends by Annual Income, divided by T/R

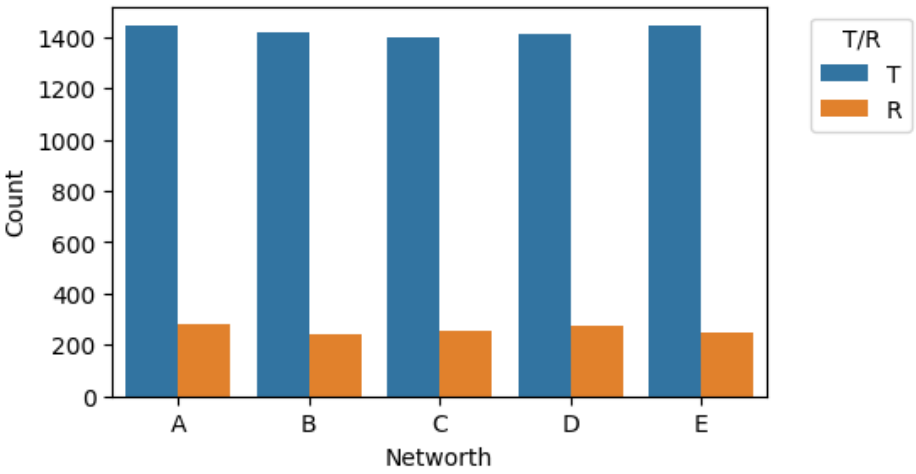


Fig. 57 - Count of Networth divided by Transactors and Revolvers

We see that Transactors are the majority at 84.6%. This means people are paying off their credit card loans quickly, which is not good for the bank. When count customers by networth, it's seen that transactors and revolvers are distributed in equal proportion. Fig. 56 displays that Revolvers are along the top edge of the pyramid which means that for revolvers, the annual income and last 3 months spend are highly correlated, but that's not the case with Transactors. Fig. 58 tells us that most Revolvers have spent more than Rs. 1L in the last 3 months while most Transactors have spent less than Rs. 1L.

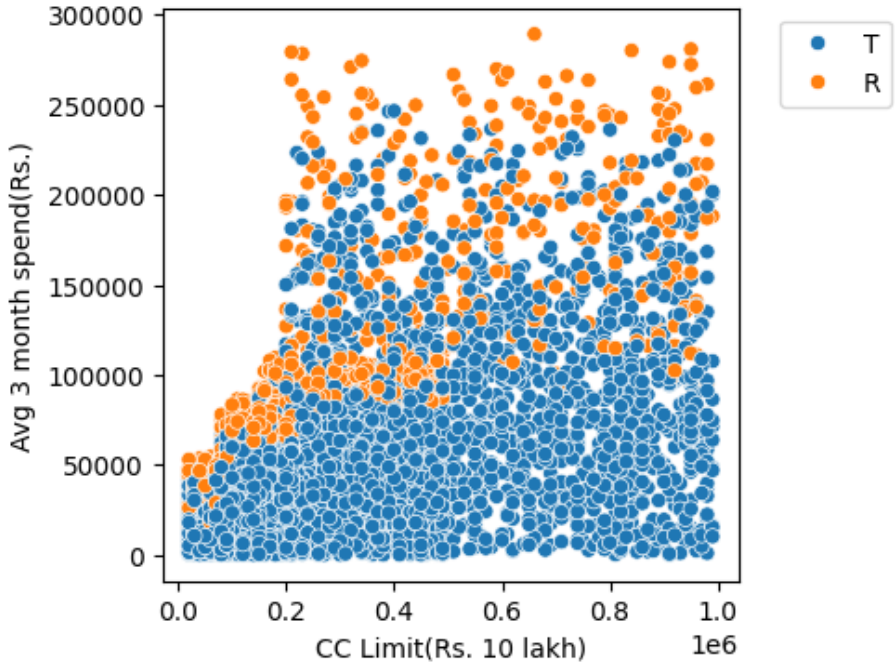


Fig. 58 - Last 3 month spends by CC Limit, divided by T/R

2.5 Credit Card Limit (cc_limit)

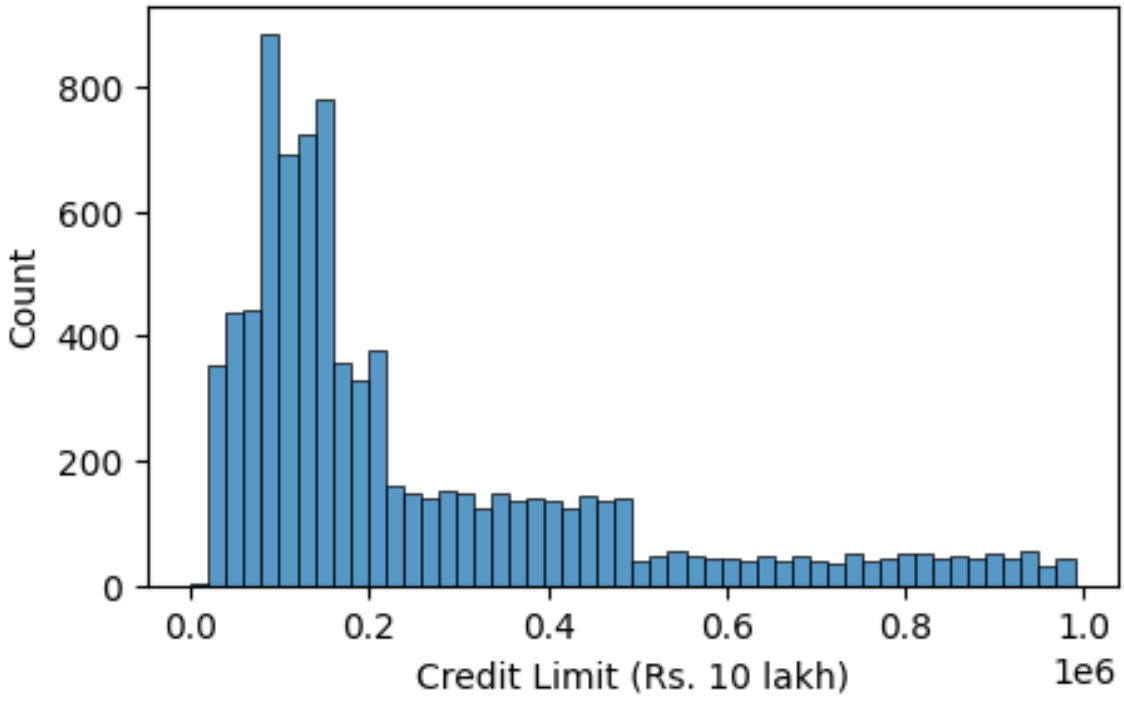


Fig. 60 - Distribution of Credit limit

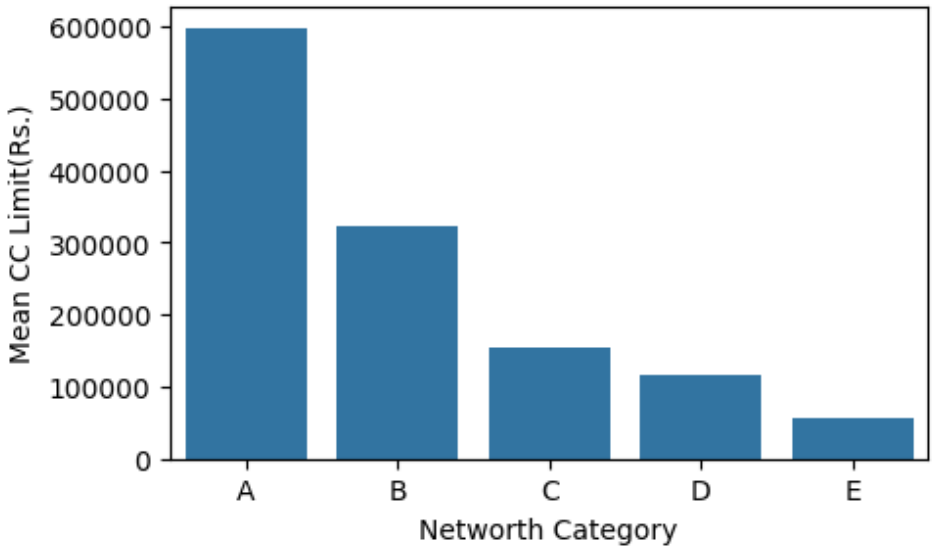


Fig. 60 - Mean CC Limit by Networth

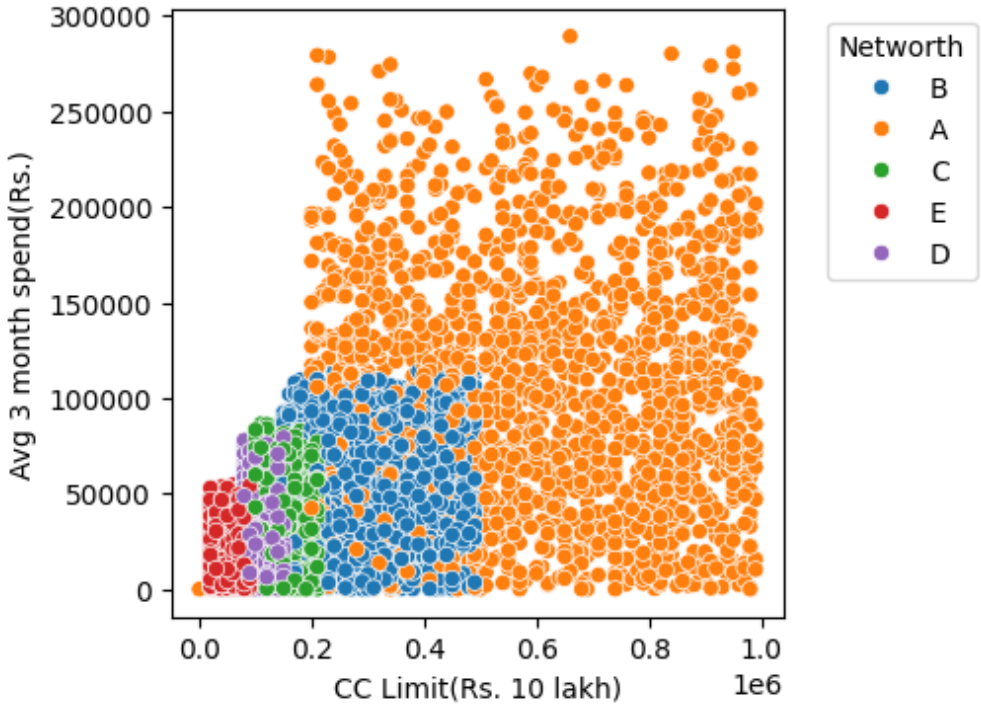


Fig. 59 - Last 3 month spends by CC Limit, divided by Networth

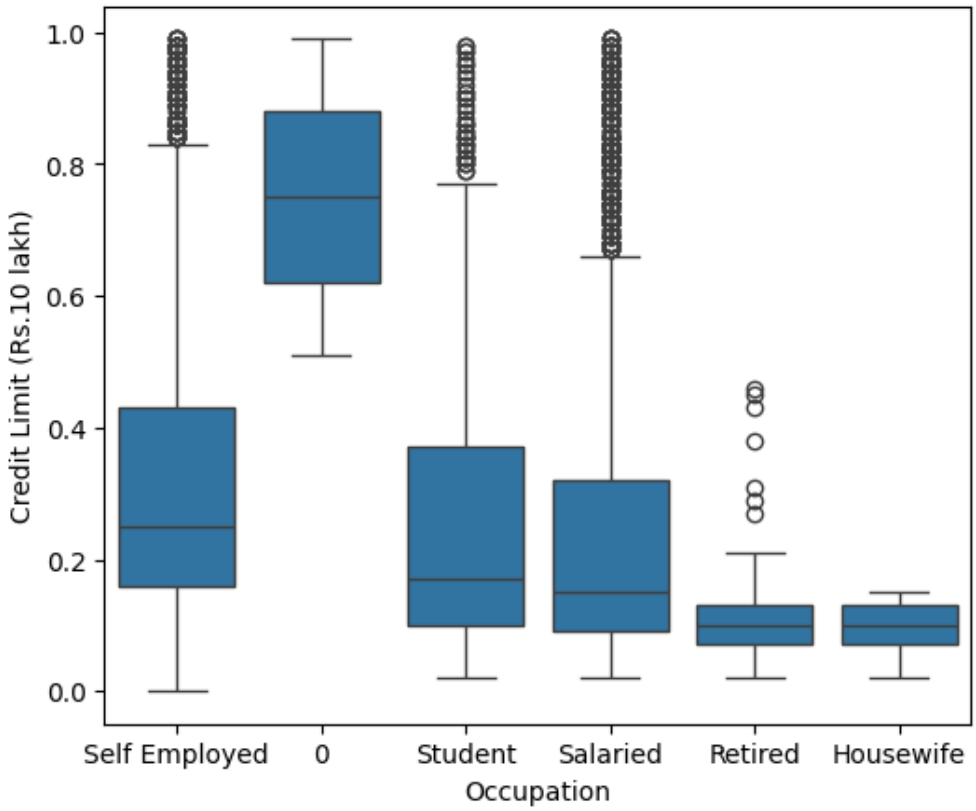


Fig. 61 - Credit Limit by Occupation (0 is an anomaly)

Credit Limit would be important when compared with other variables. In the distribution we see that 59% of Credit limits are below Rs. 2 lakh and only 14% are above Rs. 5 lakh. The histogram follows a similar shape to that of Annual Income of networth categories. When put against Last 3 Months Spend and dividing it by networth we see that category A and B have a wider range, which means that those with lower limit and higher expenditure are likely to default. In Credit limit by occupation, the doubtful part is that students have been given more limit than salaried people, but their income is also higher as shown by the income chart.