**SMDM Project**

**Business Report**

Yedupati Venkata Yamini

Table of Contents

**Problem1: Perform a thorough analysis of the data and coming up with insights to improve the marketing campaign.**

1.1 What is the important technical information about the dataset that a database administrator would be interested in? (Hint: Information about the size of the dataset and the nature of the variables)

1.2 Take a critical look at the data and do a preliminary analysis of the variables. Do a quality check of the data so that the variables are consistent. Are there any discrepancies present in the data? If yes, perform preliminary treatment of data.

1.3 Explore all the features of the data separately by using appropriate visualizations and draw insights that can be utilized by the business.

1.4 Understanding the relationships among the variables in the dataset is crucial for every analytical project. Perform analysis on the data fields to gain deeper insights. Comment on your understanding of the data.

1.5 Employees working on the existing marketing campaign have made the following remarks. Based on the data and your analysis state whether you agree or disagree with their observations. Justify your answer Based on the data available.

1.5.1 Steve Roger says “Men prefer SUV by a large margin, compared to the women”

1.5.2 Ned Stark believes that a salaried person is more likely to buy a Sedan.

1.5.3 Sheldon Cooper does not believe any of them; he claims that a salaried male is an easier target for a SUV sale over a Sedan Sale.

1.6 From the given data, comment on the amount spent on purchasing automobiles across the following categories. Comment on how a business can utilize the results from this exercise. Give justification along with presenting metrics/charts used for arriving at the conclusions.

Give justification along with presenting metrics/charts used for arriving at the conclusions.

1.6.1 Gender

1.6.2 Personal\_loan

1.7 From the current data set comment if having a working partner leads to the purchase of a higher-priced car.

1.8 The main objective of this analysis is to devise an improved marketing strategy to send targeted information to different groups of potential buyers present in the data. For the current analysis use the Gender and Marital\_status - fields to arrive at groups with similar purchase history.

**Problem 2 \*\*\*Framing An Analytics Problem\*\*\* Analyse the dataset and list down the top 5 important variables, along with the business justifications.**

Problem1: **Austo Motor Company is a leading car manufacturer specializing in SUV, Sedan, and Hatchback models. In its recent board meeting, concerns were raised by the members on the efficiency of the marketing campaign currently being used. The board decides to rope in analytics professional to improve the existing campaign.**

1.1. **What is the important technical information about the dataset that a database administrator would be interested in? (Hint: Information about the size of the dataset and the nature of the variables)**

Ans: The database administrator would be interested in knowing the scale and depth of the data which is the size of the dataset, features that the dataset contains and types of data that is present in the features

Overall summary of the data:

|  |  |  |
| --- | --- | --- |
| **Column Name** | **Comments on Data Quality** | **Type of column** |
| Age |  | integer |
| Gender | Has 53 missing/null values | categorical |
| Profession |  | categorical |
| Marital Status |  | categorical |
| Education |  | categorical |
| Number of Dependents |  | integer |
| Personal Loan |  | categorical |
| House Loan |  | categorical |
| Partner Working |  | categorical |
| Salary |  | integer |
| Partner Salary | Has 106 missing/null values | Float (decimal) |
| Total Salary |  | integer |
| Price |  | integer |
| Make |  | categorical |

Scale of the data: there are 14 columns.

Depth of the data: The initial dataset has 1581 rows.

As seen in the above table there are

6 numeric variables –

Discrete numeric variables - Age, Number of Dependents

Continuous numeric variables - Salary, Partner Salary, Total Salary, Price

8 categorical variables – Gender, Profession, Marital Status, Education, Personal Loan, House Loan, Partner working, Make

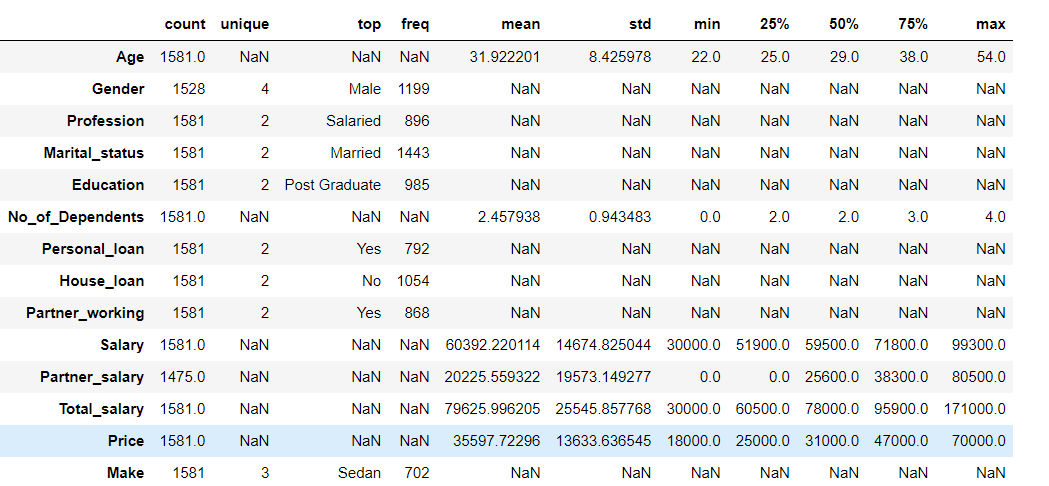
One dependent variable – Total salary depending on Salary and Partner Salary.

There is no datetime variable in the current context/dataset.

1.2. **Take a critical look at the data and do a preliminary analysis of the variables. Do a quality check of the data so that the variables are consistent. Are there any discrepancies present in the data? If yes, perform preliminary treatment of data.**

**Hint: First do detailed EDA and then do all steps of data cleaning/pre-processing – outlier treatment, missing value treatment etc..**

Ans: The statistical summary of the data present in the dataset is shown below.



* Check the relevance of data against objective or goal to be achieved
  + Scope of data – given data is well scoped
  + Time relevance of data – given data does not consider time relevance
  + Quantum of data - if the data is biased w.r.t any parameter or obtained as it is recorded – data
  + Features of data - data obtained is helpful in exploration

Points that can be noted from the above statistical summary:

* All the variables except Gender and Partner salary seem to have no discrepancies with respect to the categorical levels, minimum values, maximum values, mean and standard deviation values.
* Gender has 53 missing values.
* Partner\_salary has 106 missing values. It is of type float64 but in the given dataset there is no decimal value for Partner\_salary. The integers are represented as float numbers for this feature in the dataset.

Detailed EDA:

Categorical variables:

|  |  |
| --- | --- |
| **Feature/Column** | **Categorical levels and their counts** |
| Profession | Salaried 896 Business 685 |
| Marital Status | Married 1443 Single 138 |
| Education | Post Graduate 985 Graduate 596 |
| Personal Loan | Yes 792 No 789 |
| House Loan | No 1054 Yes 527 |
| Partner working | Yes 868 No 713 |
| Make | Sedan 702 Hatchback 582 SUV 297 |
| Gender | Male 1199 Female 327 Femal 1 Femle 1  Missing 53 |

EDA and data cleaning done on categorical variables:

1. ‘Gender’ column has bad values which are spelling errors for categorical level ‘Female’. Hence corrected the bad values to be ‘Female’
2. ‘Gender’ column has 53 missing values. They can either be replaced with another categorical level called ‘Unknown’ or they can be replaced with the mode of the data. In this case it is Male. In this business report, the first approach is followed, I.e., introduced another categorical level called ‘Unknown’ to describe the 53 missing values.
3. Statistics of Gender column after treating missing values:
   1. Male 1199  
      Female 329  
      Unknown 53

EDA and data cleaning/pre-processing on numeric variables:

|  |  |
| --- | --- |
| **Feature and statistical summary** | **Boxplot to analyze distribution and outliers** |
| Age  count 1581.000000 mean 31.922201 std 8.425978 min 22.000000 25% 25.000000 50% 29.000000 75% 38.000000 max 54.000000 Name: Age, dtype: float64  Missing values: 0 |  |
| No\_of\_Dependents  count 1581.000000 mean 2.457938 std 0.943483 min 0.000000 25% 2.000000 50% 2.000000 75% 3.000000 max 4.000000 Name: No\_of\_Dependents, dtype: float64  Missing values: 0 |  |
| Price  count 1581.000000 mean 35597.722960 std 13633.636545 min 18000.000000 25% 25000.000000 50% 31000.000000 75% 47000.000000 max 70000.000000 Name: Price, dtype: float64  Missing values: 0 |  |
| Salary  count 1581.000000 mean 60392.220114 std 14674.825044 min 30000.000000 25% 51900.000000 50% 59500.000000 75% 71800.000000 max 99300.000000 Name: Salary, dtype: float64  Missing values: 0 |  |
| Total\_salary  count 1581.000000 mean 79625.996205 std 25545.857768 min 30000.000000 25% 60500.000000 50% 78000.000000 75% 95900.000000 max 171000.000000 Name: Total\_salary, dtype: float64  Missing values: 0 |  |
| Partner\_salary  count 1475.000000 mean 20225.559322 std 19573.149277 min 0.000000 25% 0.000000 50% 25600.000000 75% 38300.000000 max 80500.000000 Name: Partner\_salary, dtype: float64  Missing values: 106 |  |

1. The numeric variables Age, Price and Salary have no outliers or missing values.
2. No\_of\_Dependents is a numeric variable that has an outlier I.e., 0. Hence, treated the outliers by imputing the outlier with the min value I.e., Q1 - (1.5\*IQR) value which is 1.
3. Total\_salary has outliers but no missing values.
4. Partner\_salary has no outliers but 106 missing values.
   1. All the values which have salary and partner\_salary values equal to sum of total salary exactly
   2. Out of 106 missing value rows, 90 rows have salary same as total salary
   3. 16 rows which have Partner\_salary missing have difference in salary and total\_salary. Hence, replacing the missing values of Partner\_salary with the difference between partner salary and total salary
5. Check for duplicates – no duplicates present in the data, hence no removal of duplicates done

**C. Explore all the features of the data separately by using appropriate visualizations and draw insights that can be utilized by the business.**

**Ans:** Univariate analysis of all the variables:

Categorical variables:

|  |  |
| --- | --- |
| Feature and percentage distribution of categorical variables | Graphical representation |
| Gender  Male 0.758381 Female 0.208096 Unknown 0.033523 Name: Gender, dtype: float64 |  |
| Profession  Salaried 0.56673 Business 0.43327 Name: Profession, dtype: float64 |  |
| Marital\_status  Married 0.912713 Single 0.087287 Name: Marital\_status, dtype: float64 |  |
| Education  Post Graduate 0.623023 Graduate 0.376977 Name: Education, dtype: float64 |  |
| Personal\_loan  Yes 0.500949 No 0.499051 Name: Personal\_loan, dtype: float64 |  |
| House\_loan  No 0.666667 Yes 0.333333 Name: House\_loan, dtype: float64 |  |
| Partner\_working  Yes 0.54902 No 0.45098 Name: Partner\_working, dtype: float64 |  |
| Make  Sedan 0.444023 Hatchback 0.368121 SUV 0.187856 Name: Make, dtype: float64 |  |

Insights that can be utilized by business:

1. Males have purchased a greater number of cars than females. 75% of the people who purchased cars are Male.

2. Married people show more interest in purchasing cars than unmarried.

3. People having house loan did not purchase cars as much as people not having house loan probably due to financial commitments or restrictions.

4. Sedan is the most purchased car among the three Car Make models.

Numeric variables:

|  |  |
| --- | --- |
| Feature and percentage distribution of variables | Graphical representation |
| Number of dependents  3 0.352309 2 0.352309 1 0.157495 4 0.137887 Name: No\_of\_Dependents, dtype: float64 |  |
| Age |  |
| Salary |  |
| Partner\_salary |  |
| Total\_salary |  |
| Price |  |

Insights that can be utilized by business:

1. Though number of dependents is a numeric variable, it can be analyzed as if it is a categorical variable. 35% of the data has people who have 3 dependents which is the majority. People might be looking for cars that is best comfortable for daily use and provide right features that can satisfy his/her family needs.

2. Majority of the people who purchase cars are within 30 years of age. Around 75% of the people who purchase cars are within 38 years of age. Marketing campaigns can be designed in such a way that this age group is targeted to improve sales.

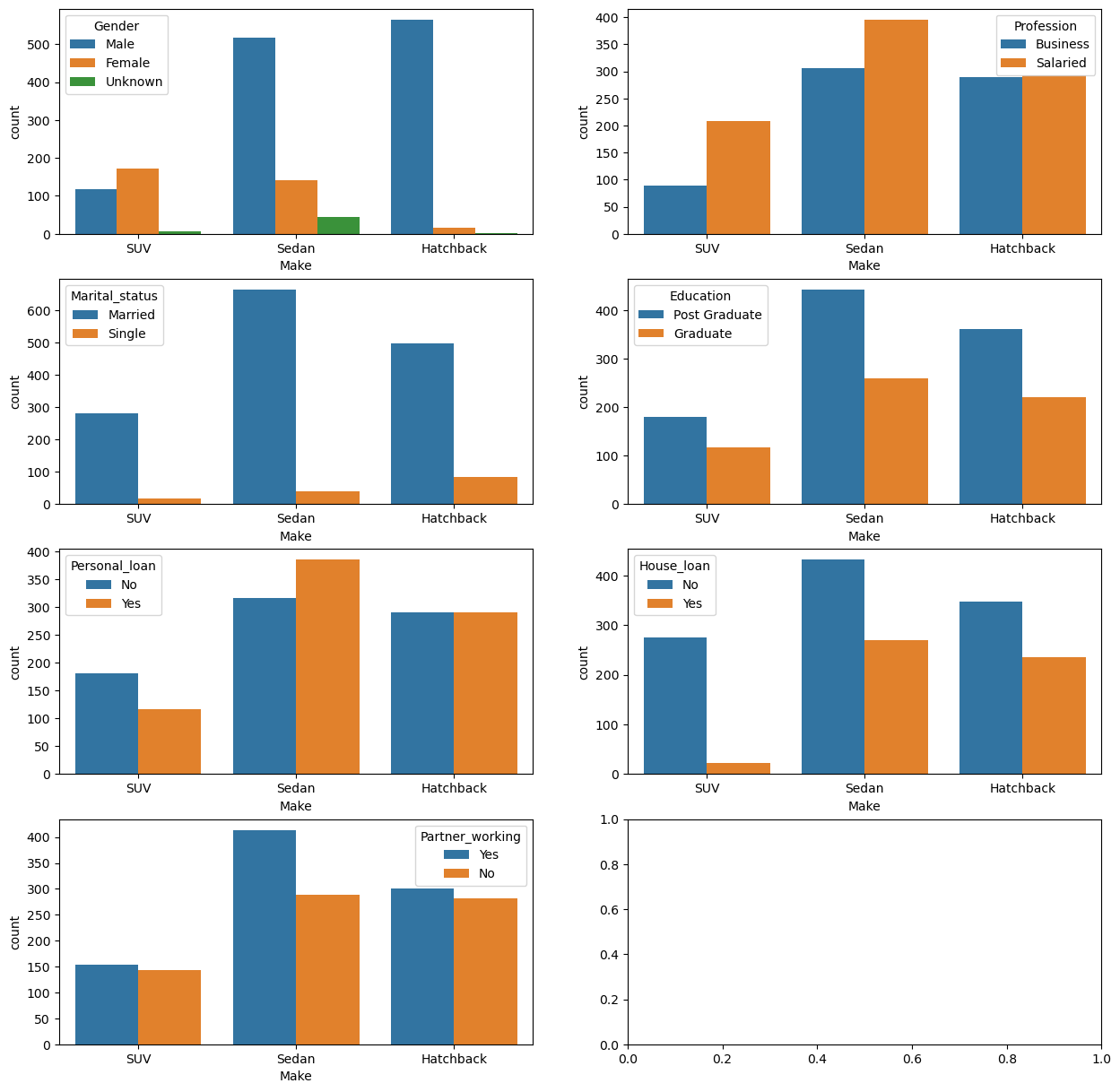
3. The average total salary of people who purchase cars is around $79,000. Majority of the people who purchase cars have non-working partner. This indicates that having working partner is added bonus but that does not refrain from the person to purchase cars.

4. The average price afforded by people in purchasing cars is $35,500. This can hint the business to run extensive marketing campaigns that advertise car make models with this price range.

**D. Understanding the relationships among the variables in the dataset is crucial for every analytical project. Perform analysis on the data fields to gain deeper insights. Comment on your understanding of the data.**

Ans: Performing bi-variate analysis on different variables to understand the relationships and dependencies of features with one another:

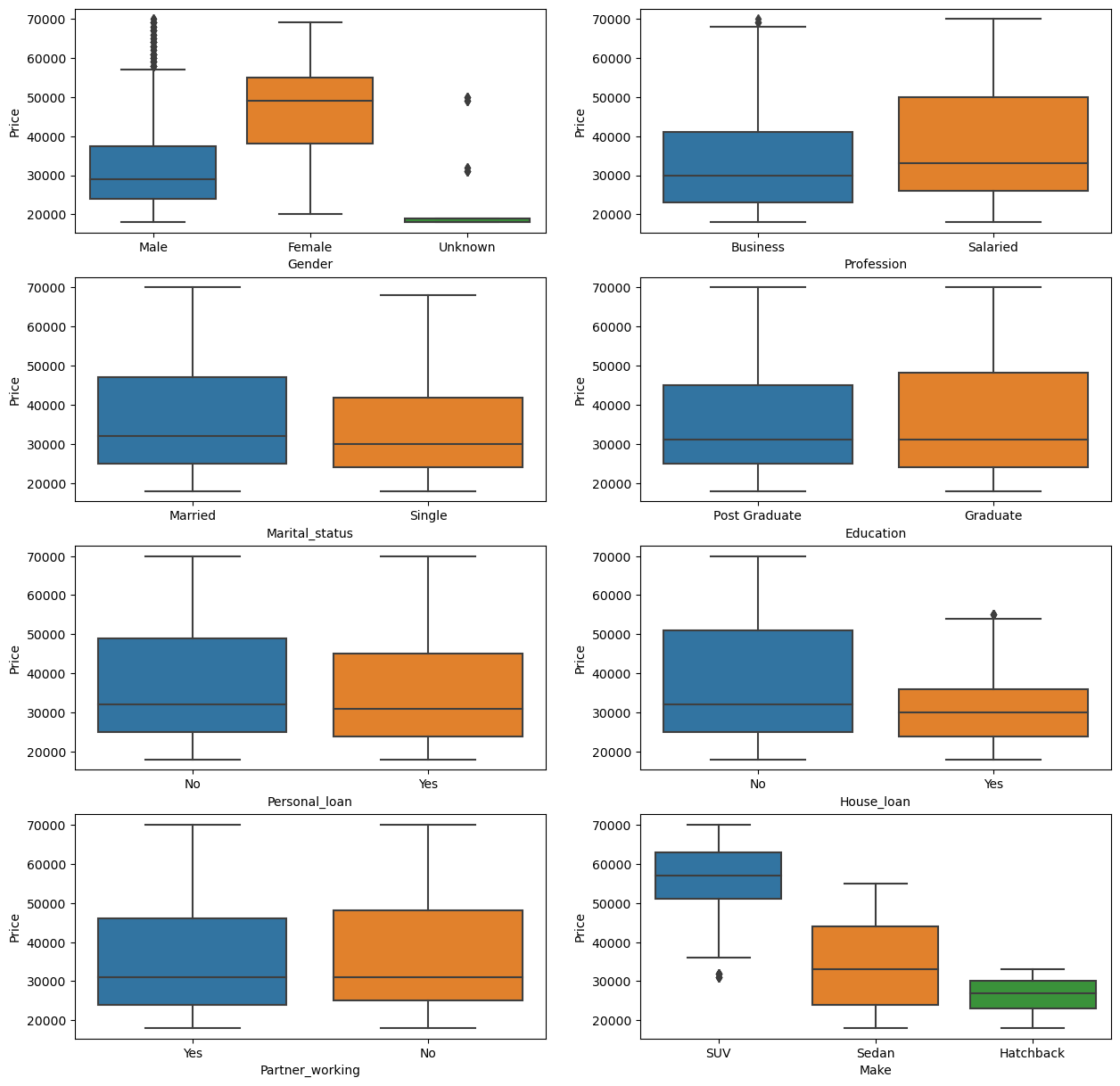
Understanding the relationship between different categorical variables:



Insights that can be utilized by business:

1. Majority of Male purchase Hatchback and Sedan. SUV is less purchased.
2. Most of the married people purchase Sedan and Hatchback models. Single people are least to target when it comes to running marketing campaigns.
3. Having a working partner tends people to purchase cars of any of the Make models.

Understanding the relationship between categorical and numeric variables:



Insights that can be utilized by business:

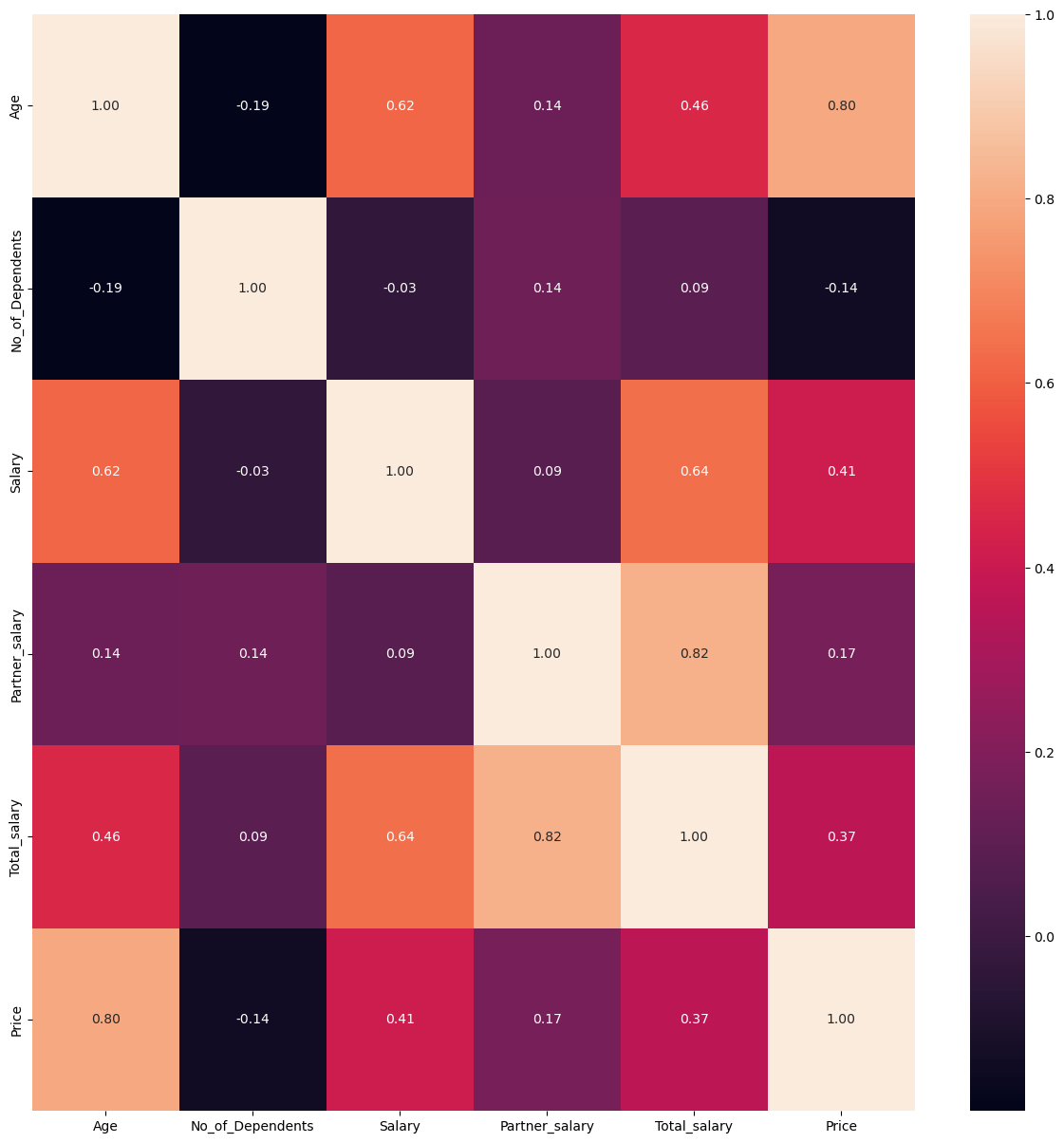
1. Male tend to purchase low budget cars and costliest cars. Median of price afforded by Female in purchasing cars is high compared to that of Male.

2. Salaried people tend to purchase high priced cars compared to single people.

3. People having no house loan tend to purchase high priced cars compared to people having house loan.

4. SUV cars cost highest compared to Sedan which is second and Hatchback which is least priced car.

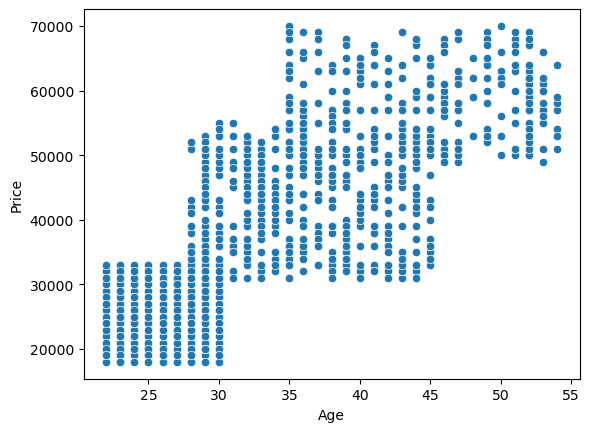
Comparing relationships among numeric variables:



Correlation between Age and price

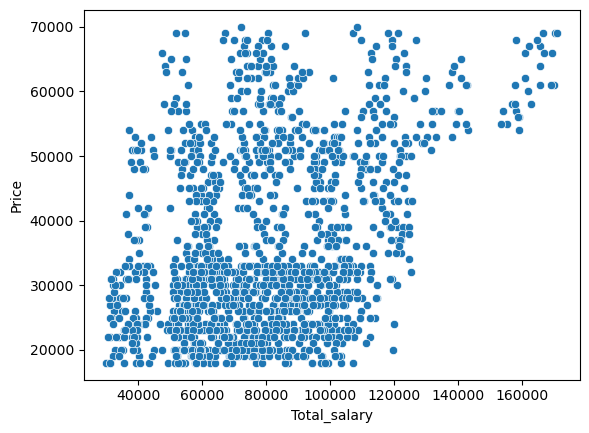
Age is positively correlated with price.

* As age progresses, people tend to purchase high priced cars.



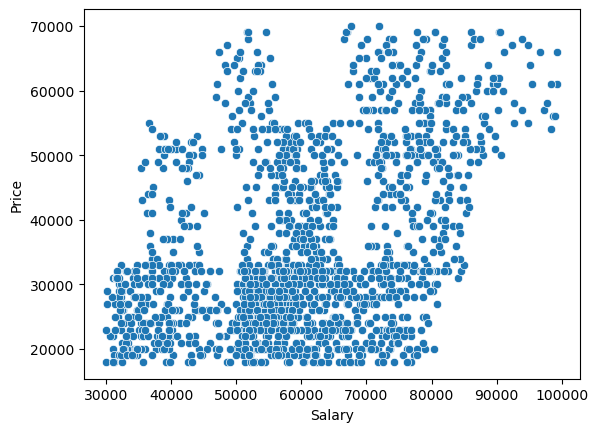
Correlation between total\_salary and price:

People with total\_salary high have capacity of purchasing cars with high cost.



Correlation between individual salary and price:

* This relation also suggests that people with high individual salary purchase high priced cars.

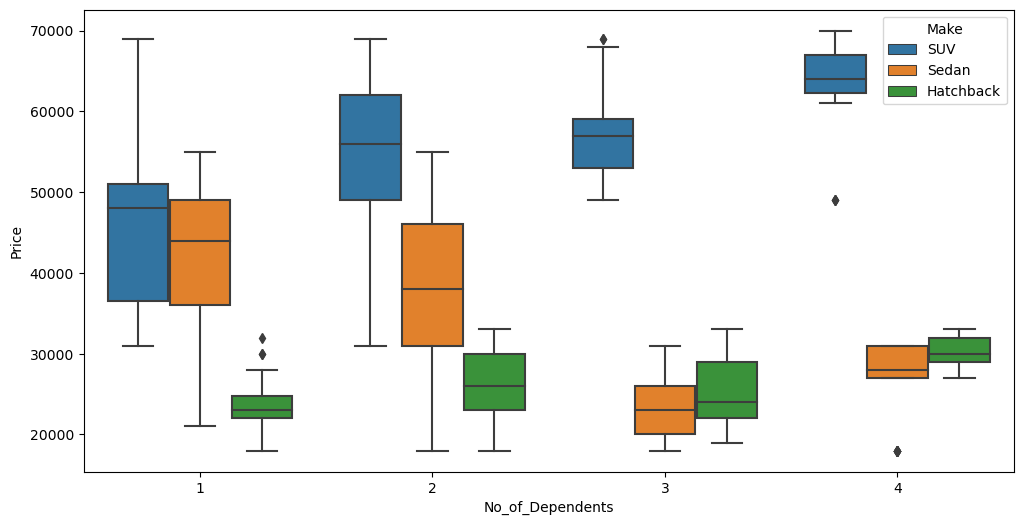


From the above graphs, the following relationships can be deduced:

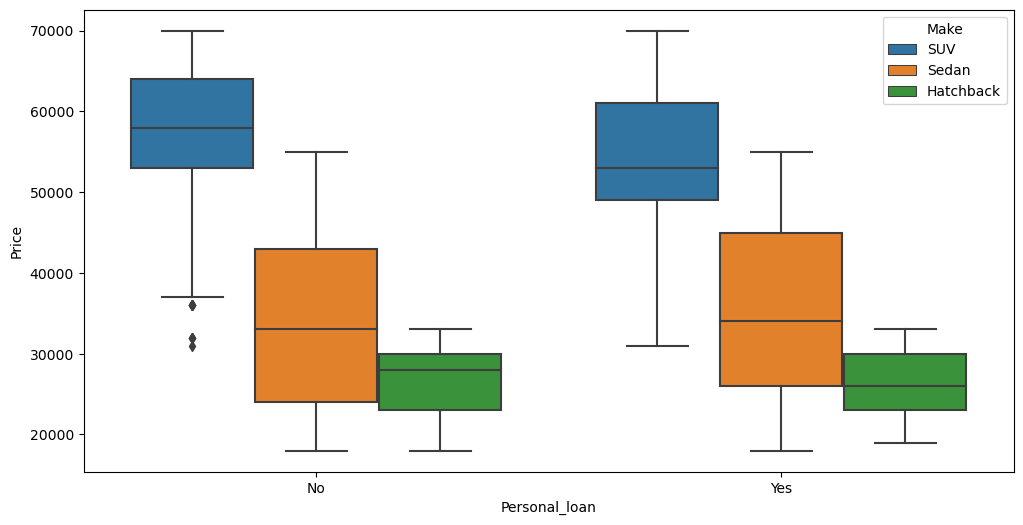
1. Age is highly overall correlated with Salary, Price, Gender, Make
2. Price is highly overall correlated with Age, Make
3. Education is highly overall correlated with Salary
4. Partner\_working is highly overall correlated with Partner\_salary, Total\_salary

Some more insights with multi variate analysis:

* People having 3 or 4 dependents purchase Sedan, Hatchback for low prices compared to people with a smaller number of dependents



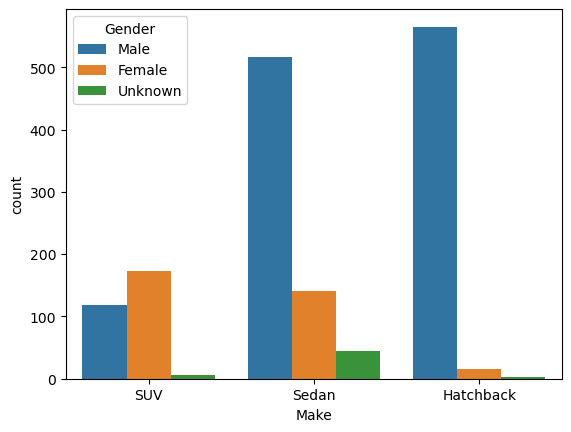
People having no personal loan afford high priced SUV and Hatchback cars as their median values are higher.



**1.5. Employees working on the existing marketing campaign have made the following remarks. Based on the data and your analysis state whether you agree or disagree with their observations. Justify your answer Based on the data available.**

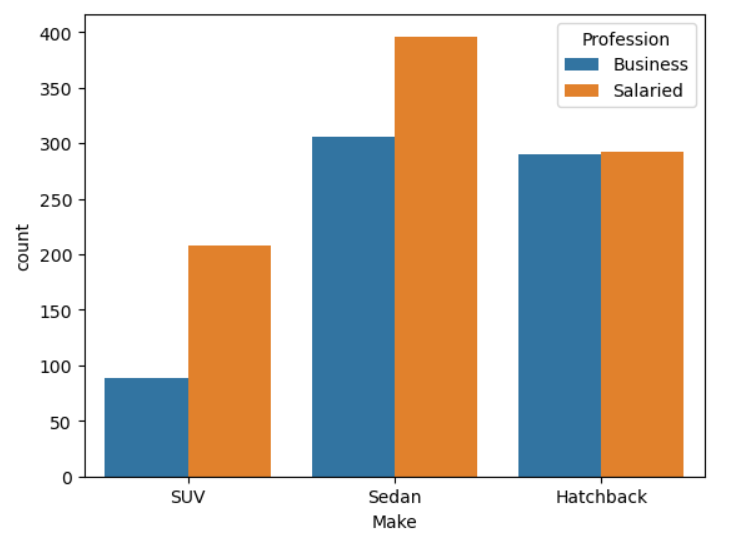
**1.5.1 Steve Roger says “Men prefer SUV by a large margin, compared to the women”**

According to the graph below, the count of women who purchased SUV is higher than that of men, So, statement made by Steve Roger is incorrect. Women prefer to by SUV by a slightly higher margin compared to men.



**1.5.2 Ned Stark believes that a salaried person is more likely to buy a Sedan.**

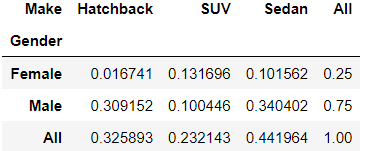
From the graph below, it is evident that the statement made by Ned Stark, ‘Salaried person is more likely to buy Sedan’ is justified compared to business professional.



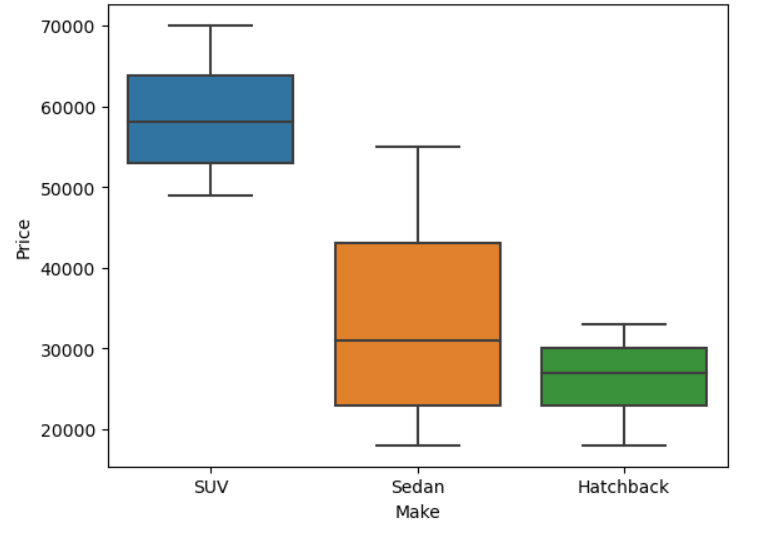
**1.5.3 Sheldon Cooper does not believe any of them; he claims that a salaried male is an easier target for a SUV sale over a Sedan Sale.**

The table below shows percentage of salaried people who purchased different Car Makes

Considering men, 10% of salaried men have purchased SUV whereas 34% of salaried men have purchased Sedan. This implies that salaried men are easier target to make them tend towards purchasing Sedan over SUV and hence they are easier target for Sedan Sale.



The median price spent by salaried men for Sedan sale is less compared to SUV sale. Hence this can be used to target salaried men for Sedan sale.



**1.6. From the given data, comment on the amount spent on purchasing automobiles across the following categories. Comment on how a business can utilize the results from this exercise. Give justification along with presenting metrics/charts used for arriving at the conclusions.**

**1.6.1 Gender**

**1.6.2 Personal\_loan**

**Gender:**

* Male tend to purchase most cars within range of 60,000$ compared to females. More number of males tend to buy low budget cars. Females have tendency of purchasing high budget cars.

How it helps the business: Car makes with affordable budget can be advertised as affordable and for wider use.

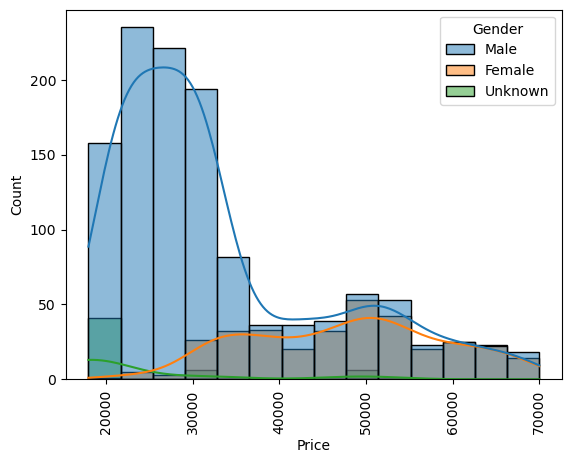
* Highest number of males have purchased cars worth around 25,000$

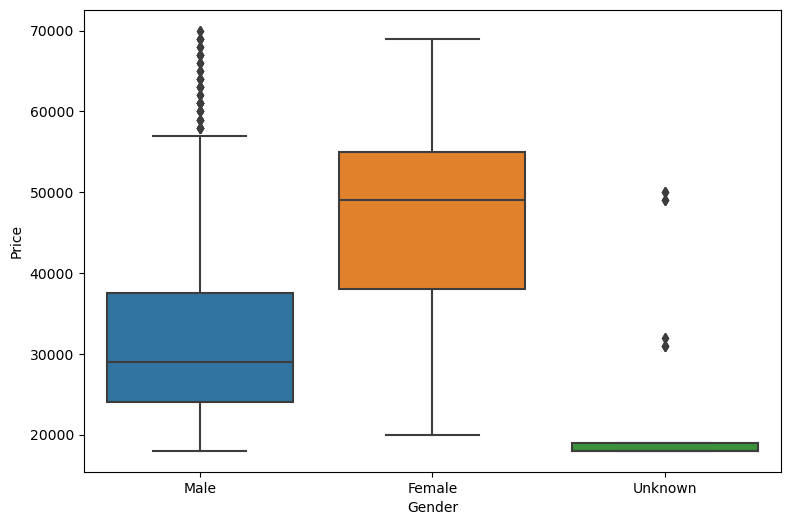
How it helps business: It is the budget cars that sale more and contribute to good amount of revenue for the business.

* Highest number of females have purchased cars worth around 50,000$

How it helps business: Females tend to spend higher than males. So, females can be target for marketing campaigns involving costly cars. Males tend to have low spending than females. So, males can be targeted for marketing campaigns involving budget cars.

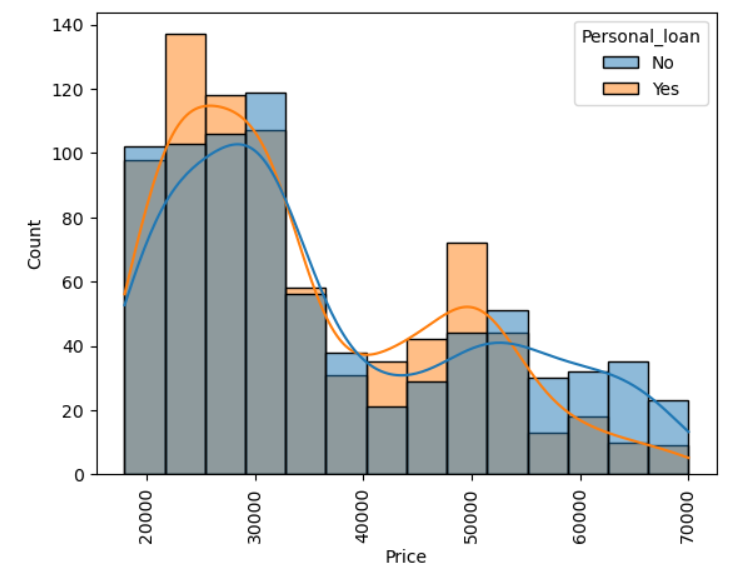
* Median is higher for females than males. Females tend to buy costly cars compared to Males
* Outliers are huge for males than females. There is considerable section of Males who purchase costliest cars. Probably males with highest salaries.

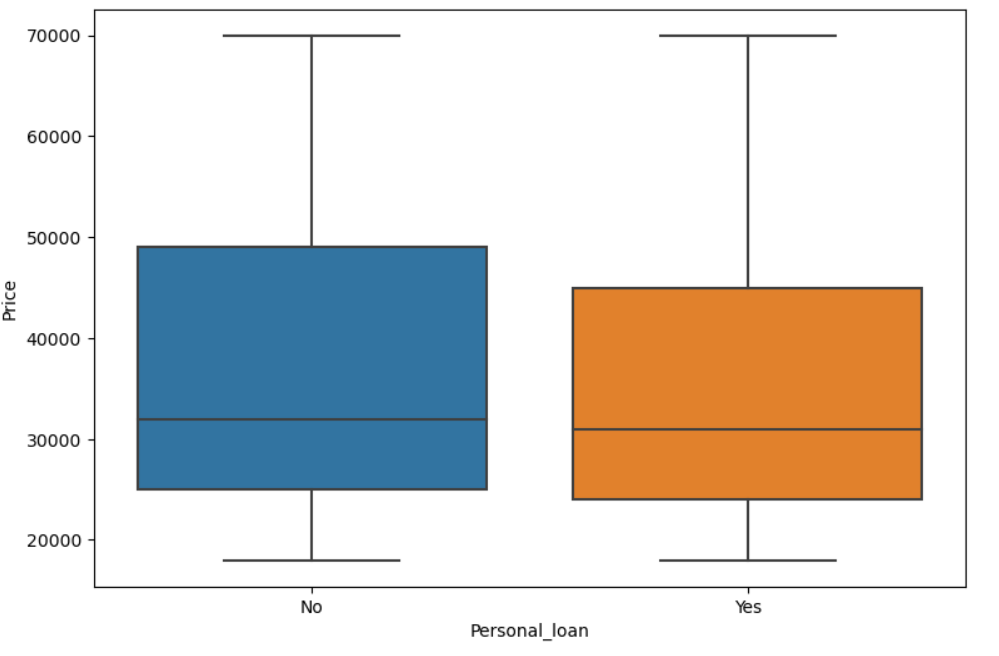




b. Personal Loan

* People having personal loans tend not to buy costly cars.
  + How this can help business: Attracting people with personal loans and exhibiting the affordable car makes to this section in the marketing campaign can increase the sale of affordable car make types.
* People having personal loans tend more to buy costly cars
* The relation between personal loan and number of people who spent more amount on purchasing cars is not consistent through every price range.
* Median price spent by people with and without personal loan lies in the same range. But the median price spent by people having no personal loan is a bit higher than the people having personal loan



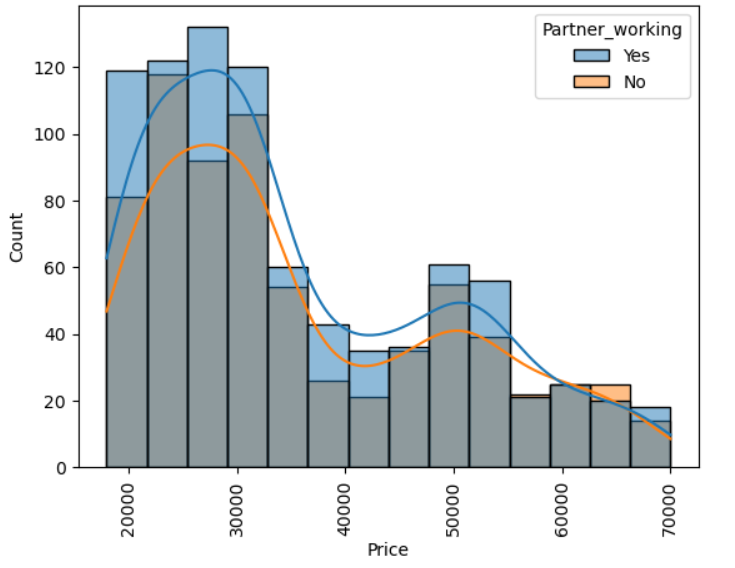


**1.7. From the current data set comment if having a working partner leads to the purchase of a higher-priced car.**

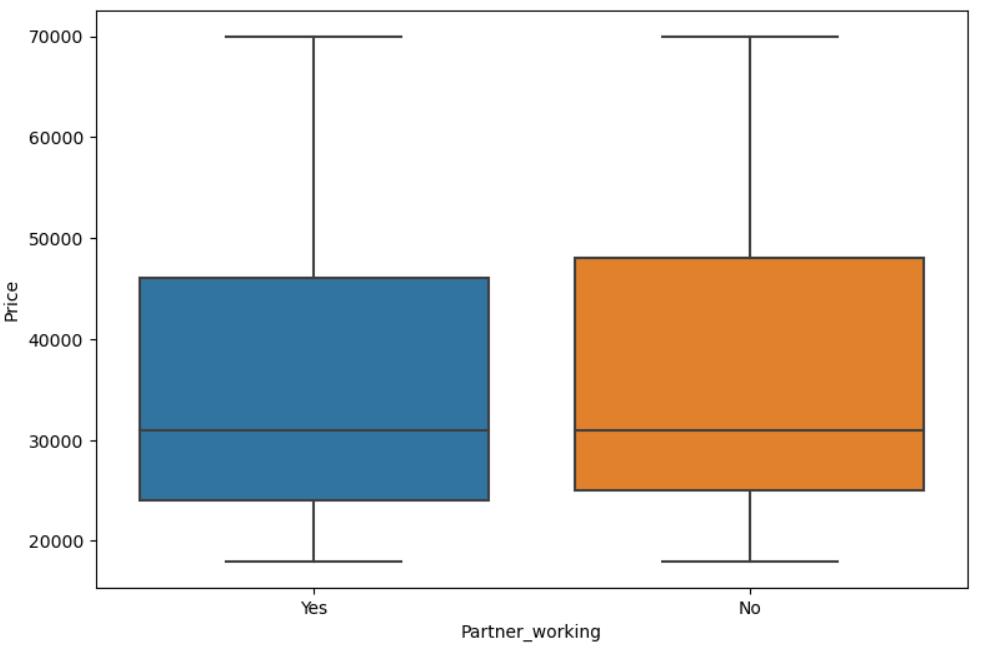
From the chat below, people with working partner tend to purchase high priced cars than people without working partner up to the price range of around 60,000$.

Beyond this range, people without working partner tend to purchase slightly higher priced cars than people with working partner up to around 67,000$.

Beyond that, people with working partner tend to purchase higher priced car compared to people without working partner



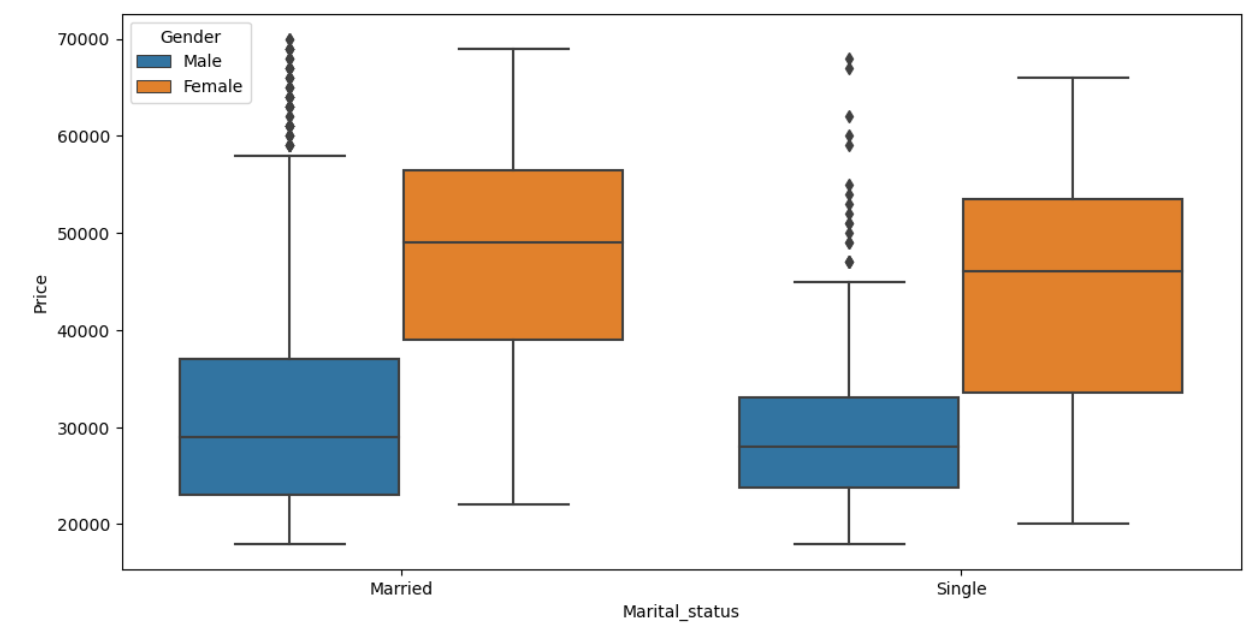
Interestingly, median price spent by people with and without working partner is almost the same



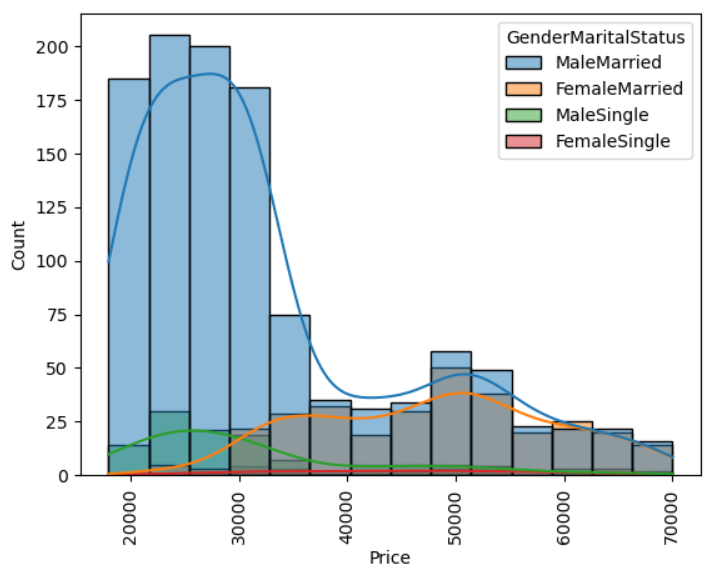
**1.8. The main objective of this analysis is to devise an improved marketing strategy to send targeted information to different groups of potential buyers present in the data. For the current analysis use the Gender and Marital\_status - fields to arrive at groups with similar purchase history.**

Use gender, marital status and share your insights and recommendations based on data analysis

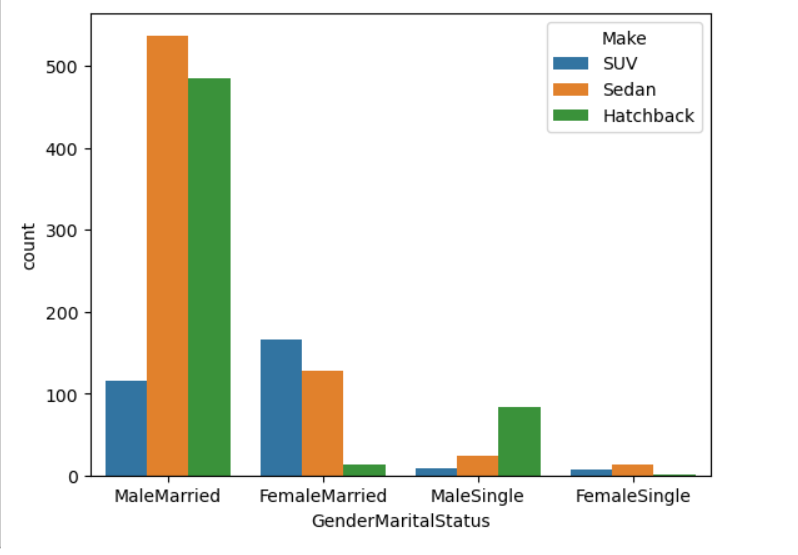
* Either single or married, majority of females tend to have high priced car purchase history than males
* But males either single or married purchased cars with highest price



* Married Male tend to purchase highest number of cars.
* Purchase history of single females is considerably the least.
* Female married and Male married tend to purchase cars more than unmarried people



* Married males purchase Sedan and Hatchback the most whereas married females purchase Sedan and SUV the most.



**Problem 2:**

**\*\*\*Framing An Analytics Problem\*\*\* Analyse the dataset and list down the top 5 important variables, along with the business justifications.**

**Ans:**

Important points to be noted:

1. Interest charged on the capital that the bank lends out to customers has historically been the most significant method of revenue generation.

2. The bank makes money only if customers spend more on credit cards

3. the Bank wants to revisit its credit card policy and make sure that the card given to the customer is the right credit card.

4. The bank will make a profit only through the customers that show higher intent towards a recommended credit card. (Higher intent means consumers would want to use the card and hence not be attrite.)

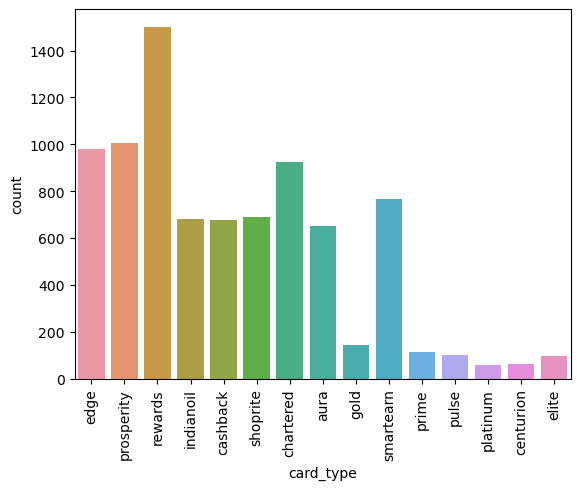
**Data Hygiene check:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Row** | **Dataype - comments** | **Null values and other remarks** | **Description** |
| userid | 8448 non-null | int64 | Unique customer bank id – discrete number |
| card\_no | 8448 non-null | object | Masked card number – alpha numeric |
| card\_bin\_no | 8448 non-null | int64 | Card IIN number |
| Issuer | 8448 non-null | object | Card issuer -  Visa 7279 Mastercard 728 Amex 441 |
| card\_type | 8448 non-null | object | Card type  rewards 1502 prosperity 1007 edge 980 chartered 923 smartearn 765 shoprite 688 indianoil 680 cashback 676 aura 652 gold 145 prime 112 pulse 101 elite 96 centurion 62 platinum 59 |
| card\_source\_date | 8448 non-null (from which date card has been active) | datetime64[ns] | Date from which card has been active |
| high\_networth | 8448 non-null | object | Customer category basis networth value  A 1740 D 1696 E 1693 B 1660 C 1659 |
| active\_30 | 8448 non-null | int64 | Account activity in last 30 days  0 5978 1 2470 |
| active\_60 | 8448 non-null | int64 | Account activity in last 60 days  0 4268 1 4180 |
| active\_90 | 8448 non-null | int64 | Account activity in last 90 days  1 5424 0 3024 |
| cc\_active30 | 8448 non-null | int64 | 0 6048 1 2400 |
| cc\_active60 | 8448 non-null | int64 | 0 4355 1 4093 |
| cc\_active90 | 8448 non-null | int64 | 1 5342 0 3106 |
| hotlist\_flag | 8448 non-null | object | N 8410 Y 38 |
| widget\_products | 8448 non-null | int64 | Number of products customer holds  Dc, cc, netbanking, mobile banking.. |
| engagement\_products | 8448 non-null | int64 | Number of enagement products -  FD, RD, personal loan, auto loan.. |
| annual\_income\_at\_source | 8448 non-null | int64 | Annual income record during credit card application and creation |
| other\_bank\_cc\_holding | 8448 non-null | object | Holding other bank credit card  Y 4728 N 3720 |
| bank\_vintage | 8448 non-null | int64 | Vintage with bank in months as on Tth month (vintage - bank account opening date) |
| T+1\_month\_activity | 8448 non-null | int64 | 0 7508 1 940 |
| T+2\_month\_activity | 8448 non-null | int64 | 0 8043 1 405 |
| T+3\_month\_activity | 8448 non-null | int64 | 0 7769 1 679 |
| T+6\_month\_activity | 8448 non-null | int64 | 0 8373 1 75 |
| T+12\_month\_activity | 8448 non-null | int64 | 0 8368 1 80 |
| Transactor\_revolver | 8410 non-null | object | Revolver– who carries bills to next month  Transactor– who pays bills on time  T 7115 R 1295 |
| avg\_spends\_l3m | 8448 non-null | int64 |  |
| Occupation\_at\_source | 8448 non-null | object | Occupation at the time of credit card application  Salaried 3918 Self Employed 2175 Retired 1089 Student 621 Housewife 384 0 261 |
| cc\_limit | 8448 non-null | int64 | Current credit card limit |
|  |  |  |  |

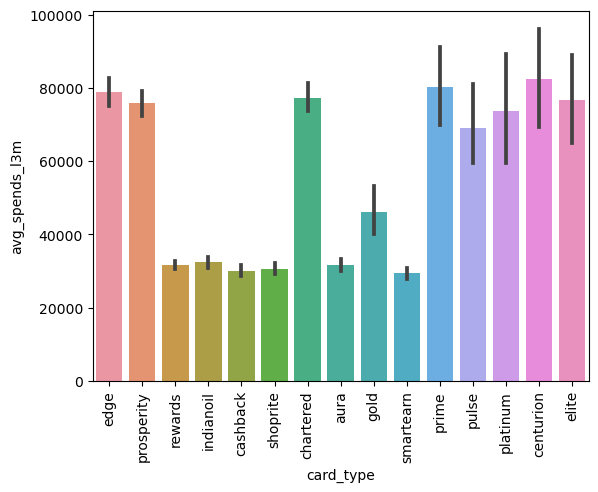
**Goal: Increase cc spendings of credit card holders**

**Questions and answers:**

1. What is the target variable in the data?
   1. Avg\_spends\_l3m
2. Find out all the important variables to consider at an initial glance
   1. Card type, high networth, cc\_active30, cc\_active60, cc\_active90, annual\_income\_at\_source, other\_bank\_cc\_holding, T+1,2,3,6,12 months activity, transactor\_resolver, cc\_limit
3. Do widget and engagement products, Occupation at source have a role to influence avg\_spends\_l3m
4. Which card\_type is seeing highest average\_l3m\_spend?
   1. avg spend of centurion and prime card types are higher but the count of people possessing those card types is very low.
   2. avg spending of rewards category is low but highest number of customers are possessing this card type.

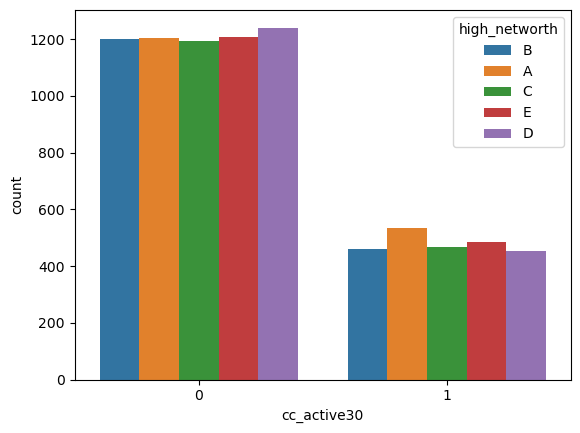


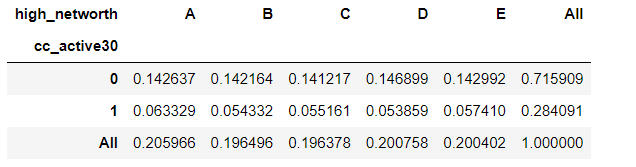
rewards 0.177794  
prosperity 0.119200  
edge 0.116004  
chartered 0.109257  
smartearn 0.090554  
shoprite 0.081439  
indianoil 0.080492  
cashback 0.080019  
aura 0.077178  
gold 0.017164  
prime 0.013258  
pulse 0.011955  
elite 0.011364  
centurion 0.007339  
platinum 0.006984  
Name: card\_type, dtype: float64



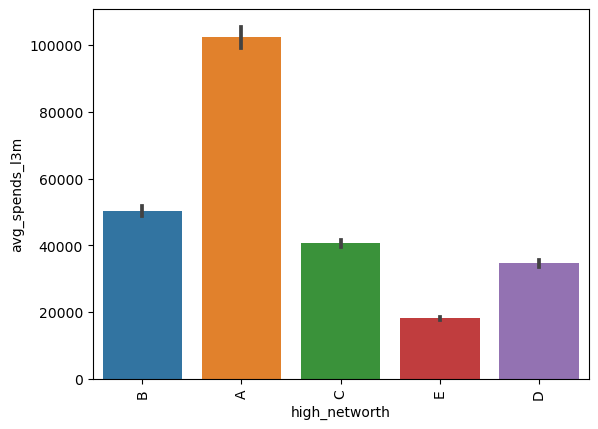
card\_type  
centurion 82440.564516  
prime 80403.357143  
edge 78981.873469  
chartered 77330.923077  
elite 76668.229167  
prosperity 75789.428004  
platinum 73599.711864  
pulse 69031.188119  
gold 46164.641379  
indianoil 32356.322059  
aura 31580.196319  
rewards 31516.928096  
shoprite 30667.816860  
cashback 30049.690828  
smartearn 29384.892810  
Name: avg\_spends\_l3m, dtype: float64

1. Which category of net worth people use credit cards the most in last 30, 60, 90 days
   1. 71% of people have no cc activity in last 30 days and there is no dominating high net worth category here.





* 1. high net worth cat A has highest average l3 month spends

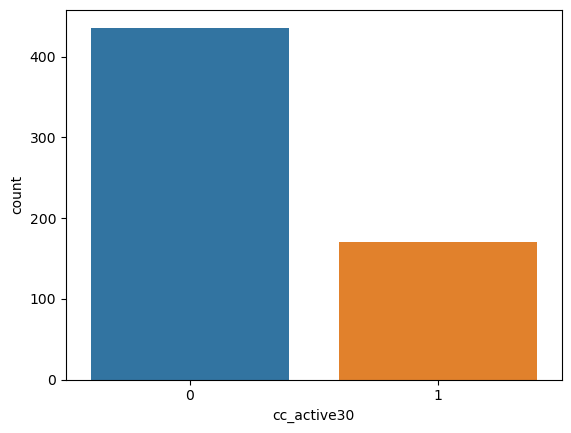


1. Are low annual\_income\_at\_source people attriding more?
   1. 72% of people with less than 5,00,000 annual incomes at source are not active with credit card since last 30 days.

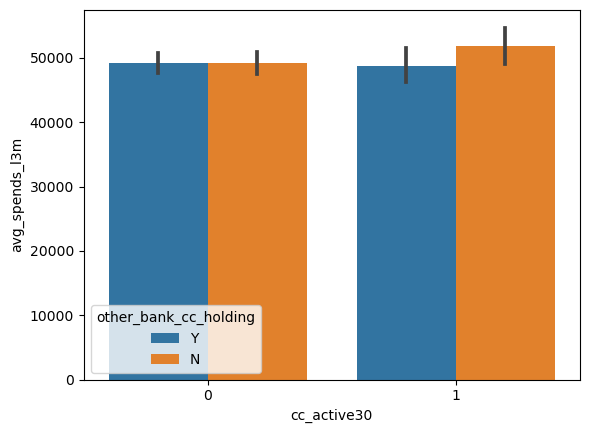
Distribution of cc\_active30 variable for people with annual income less than Rs 5,00,000/-

0 0.719472  
 1 0.280528  
 Name: cc\_active30, dtype: float64

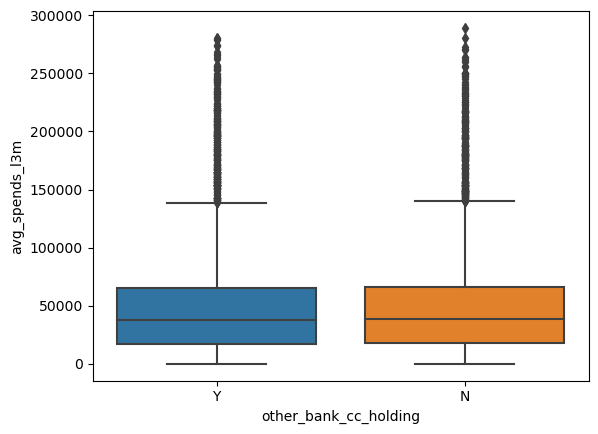
Graphical representation of the above metric:



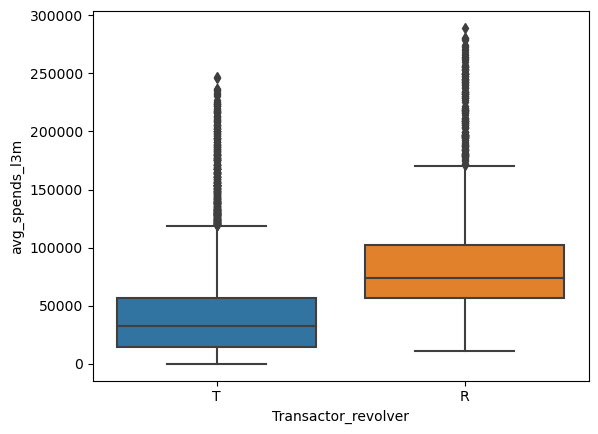
1. Avg spend l3m trends of customers who hold other bank credit cards
   1. 55% of customers hold other bank credit cards.
   2. Avg spends is more for people who are active since last 30 days that do not hold any other bank’s credit card.



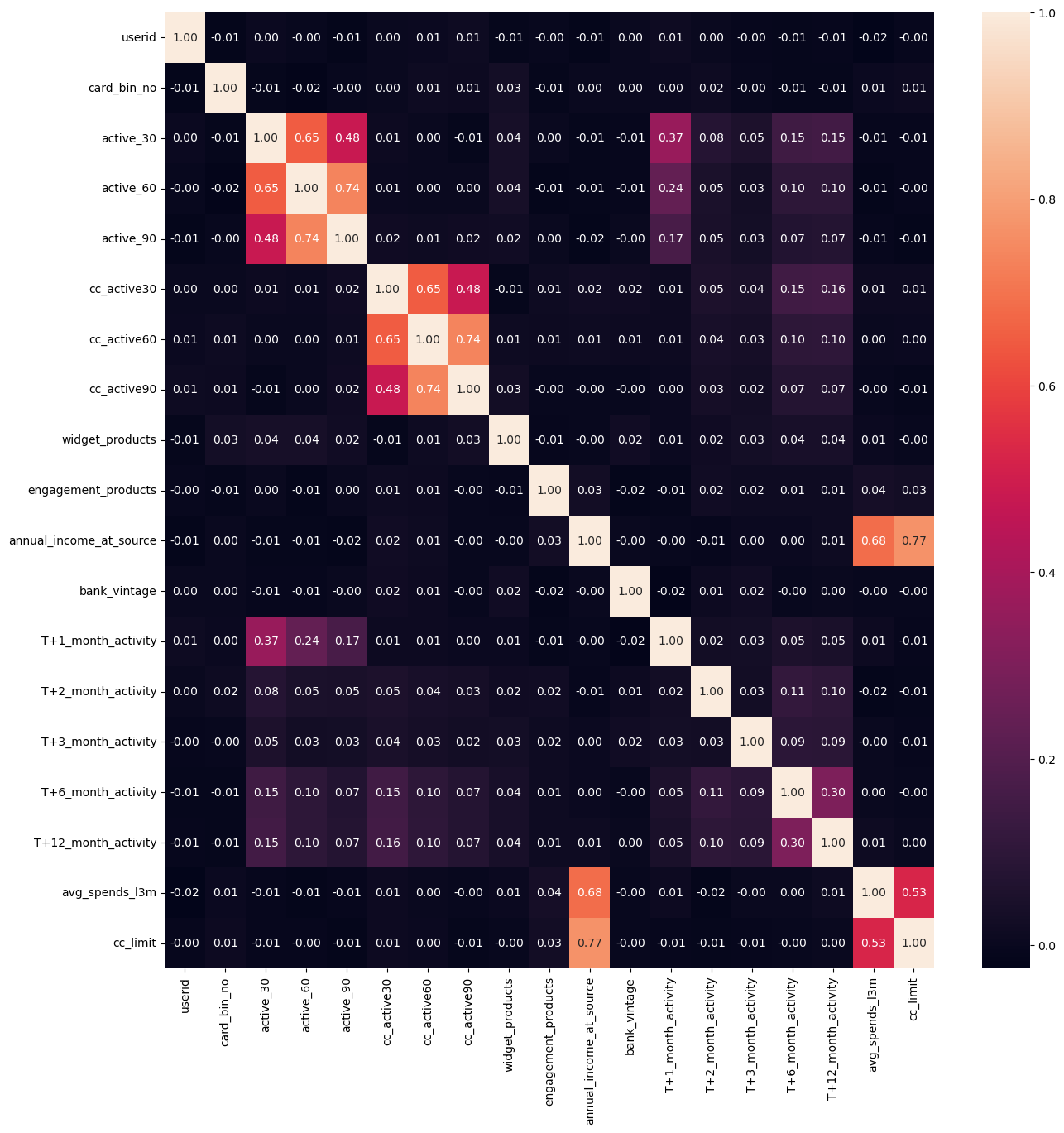
* 1. Median avg spends in last 3 months is the same for persons with/without holding other bank’s credit card.



1. How much does the customer has to spend in next 1,2,3,6,12 months vs avg spend last 3 months
   1. People who spent highest in prev 3 months have been inactive since T+1,2,3,6,12 months.
2. Are the people who are having high/low avg\_spend\_l3m transactors or revolvers?
   1. Revolvers spent more in last 3 months than transactors.



1. Cc\_limit vs avg\_spend\_l3months. Are the customers spending more than their limit?
   1. There are only a few customers who have roughly spent more than their cc\_limit.
   2. From correlation matrix, cc\_limit is slightly more correlated to averagespends. If high cc limit, more tendency to spend on credit card.
2. Findings from correlation analysis
   1. Cc\_limit and annual\_income\_at\_source have high correlation with avg\_spends\_l3m



1. Along with avg\_spends \_l3m, goal of the bank is to reduce attrition
   1. According to the statistics below, 89% customers stop using credit card from next month
   2. 95% 2 months after T+2nd month
   3. 91% 3 months after T+3rd month
   4. 99% and even above afterT+6th month T+12th month

|  |  |  |
| --- | --- | --- |
| **T+1\_month\_activity** | 0 7508 1 940 | 0 0.888731 - 89% 1 0.111269 - 11% |
| **T+2\_month\_activity** | 0 8043 1 405 | 0 0.95206 - 95% 1 0.04794 - 4% |
| **T+3\_month\_activity** | 0 7769 1 679 | 0 0.919626 - 91% 1 0.080374 - 8% |
| **T+6\_month\_activity** | 0 8373 1 75 | 0 0.991122 - 99.1% 1 0.008878 - 0.9% |
| **T+12\_month\_activity** | 0 8368 1 80 | 0 0.99053 - 99% 1 0.00947 - 1% |

* 1. Insights: 99% of customers stop spending on credit card and attride in or around 12 months from a given point of time.

From the above EDA top 5 imp variables that seem to be important to be considered by the bank to increase spendings of the credit card holders are:

Avg\_spends\_l3m

Cc\_active30

T + 6, T + 12 month activities

Cc\_limit

Annual\_income\_at\_source

High networth