

SMDMPROJECT

CODED

DSBA

Anand B R | Batch: PGPDSBA.O.JAN24

Contents

Contents	2
Introduction	3
Problem 1	4
Descriptive statistics to Summarize Data	5
Treat the Missing Values:	6
Analyze the Numerical and Categorical columns.	7
Analyze the Relationship between Variables	14
Analyze Relationship between Categorical Variables	20
Visualize Multiple field relationships	25
Key Questions to Answer:	30
Actionable Insights and Business Recommendations	37
Problem 2	38
Objective	38
Top 5 important Variables:	39

INTRODUCTION

Objective

Objective is to conduct Descriptive and Exploratory Data Analysis of the Data Set.

AIM

- Understanding Data Set
- Detect Anomalies and Outliers and Treat
- Perform Univariate, Bivariate and Multivariate analysis.
- Visualize with appropriate plots.
- Develop insights to convey the information, which helps Business in Decision Making

Data Set

- Sample Size 1581
- Fields 14

Fields

- **Numerical:** Age, No of Dependents, Salary, Partner salary, Total salary, Price
- **Categorical:** Gender, Profession, Marital Status, Education, Personal Loan, House Loan, Partner Working

Car Make Type

- Sedan
- SUV
- Hatchback

PROBLEM 1

Context

Explore Data and Reflect the insights.

Introduction

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. Borad decides to rope in analytics professional to improve the existing campaign.

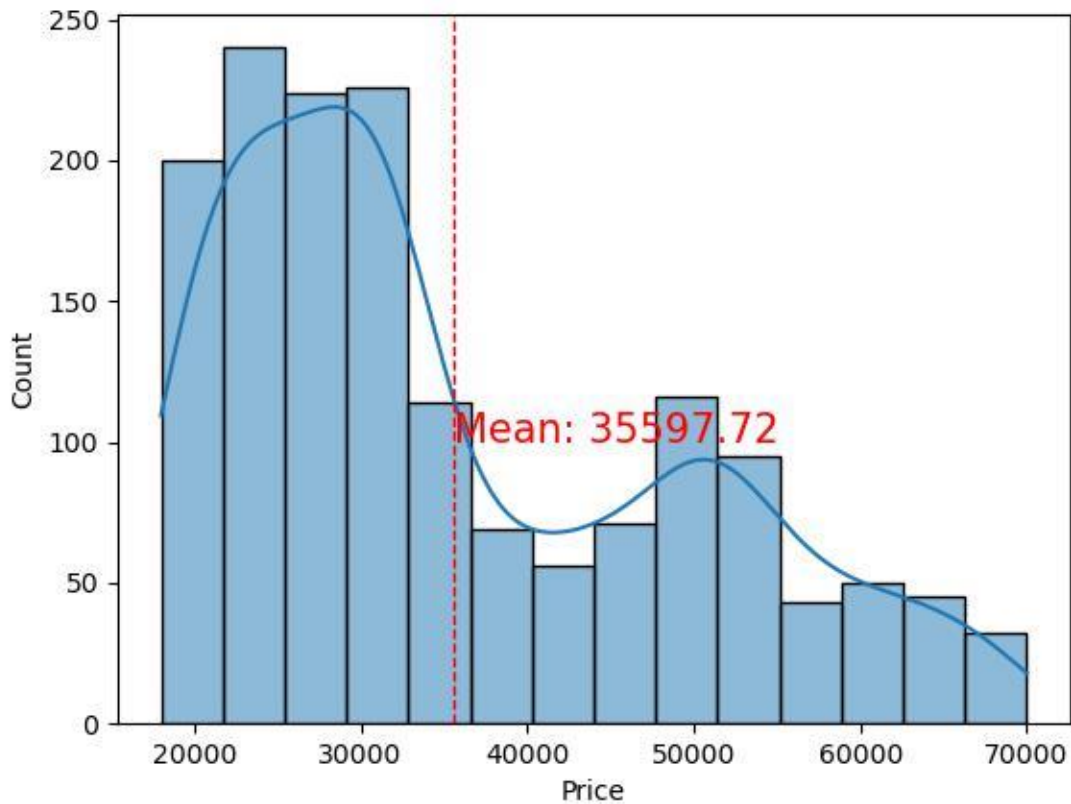
Objective

As Data Scientist expected to Analyze data to get fair idea about the demand of Customers which help Austo in enhancing Customer experience.

The objective of the analysis is to conduct descriptive and exploratory data analysis of a dataset related to car purchases. The analysis aims to understand the dataset, detect anomalies and outliers, perform univariate, bivariate, and multivariate analysis, visualize the data with appropriate plots, and develop insights to help improve the business of Austo Motor Company, a leading car manufacturer.

DESCRIPTIVE STATISTICS TO SUMMARIZE DATA

- Missing Values observed in Categorical Fields Gender and Partner Salary
- Age varies from 22 years to 54 years.
- Number of Dependents up to 4 for a person buying CAR
- Approx. 80% are Male.
- Approx. 91% are Married.
- 44% bought Sedan, 37% bought Hatchback and 19% bought SUV.
- 56% Salaried Professional and 43% Business.
- 62% with Postgraduate Education and 37% with Graduate Degree
- 50% with Personal Loan
- 66% with House Loan
- Mean Salary approx. 60392
- Mean Total Salary approx. 79625
- Mean price of the car approx. 35597



TREAT THE MISSING VALUES:

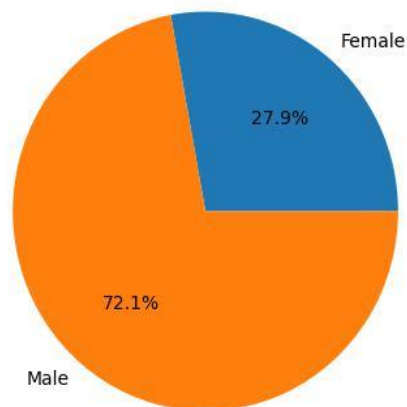
Partner Salary:

1. Fill the NULL Cells with Zero, where Partner Working Field is 'NO'
2. Fill the Cells with Difference of Total Salary with Salary to Obtain the Partner Salary

Gender:

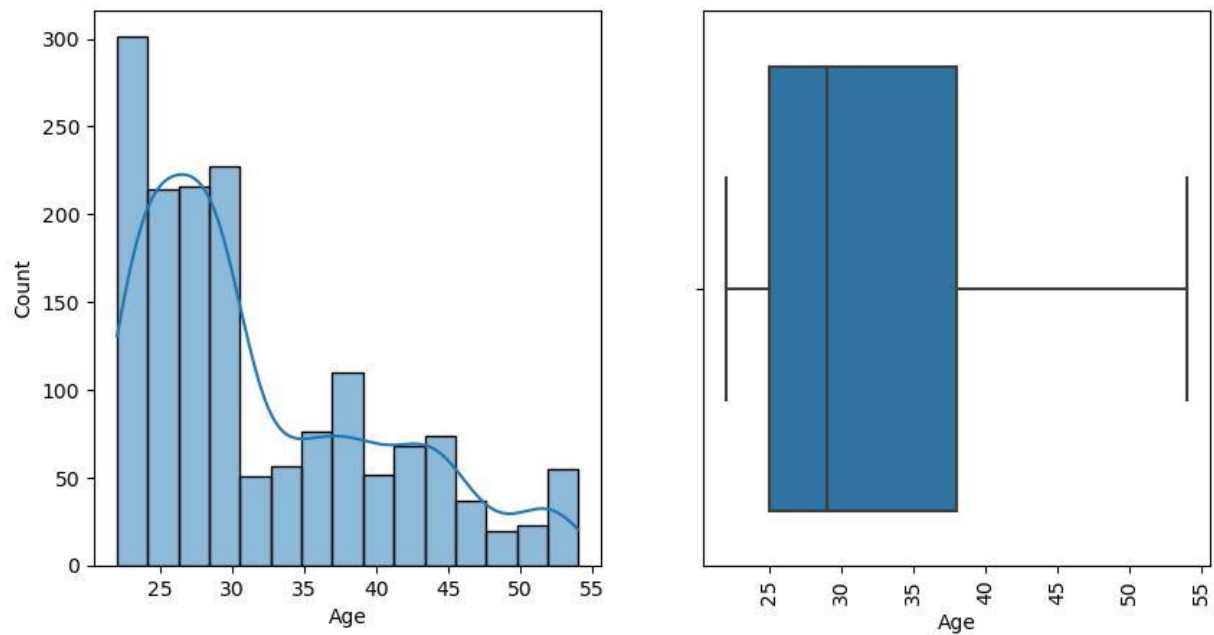
- 53 rows missing Gender Values, which contribute to 3.35 %.
- Either we can drop 53 rows or can Impute Central Tendency Mode to the Cells
- In the Context – considering the value addition from other fields of those 53 ROWS, Gender Mode is filled to Null cells.

After Treating the Null values, the below Pie Chart represent of Total PRICE on car, share of Price spent by Male and Female. Even though number of Female is less than 20%, but they contribute to 28% of Total SUM Price. This Hints Female BUY More expensive Cars. (Further will be Analyzed with Box Plots).



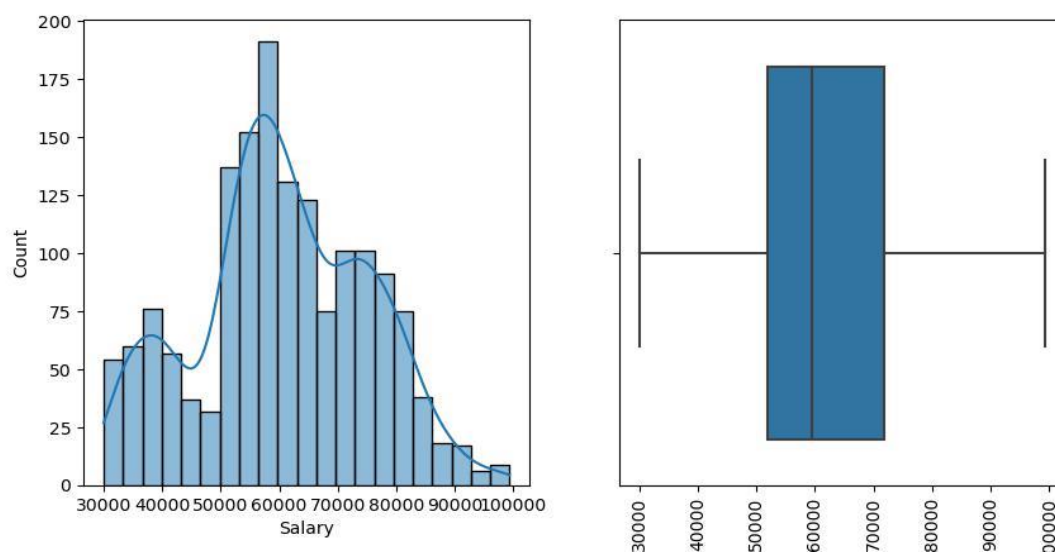
ANALYZE THE NUMERICAL AND CATEGORICAL COLUMNS.

(a) Age:



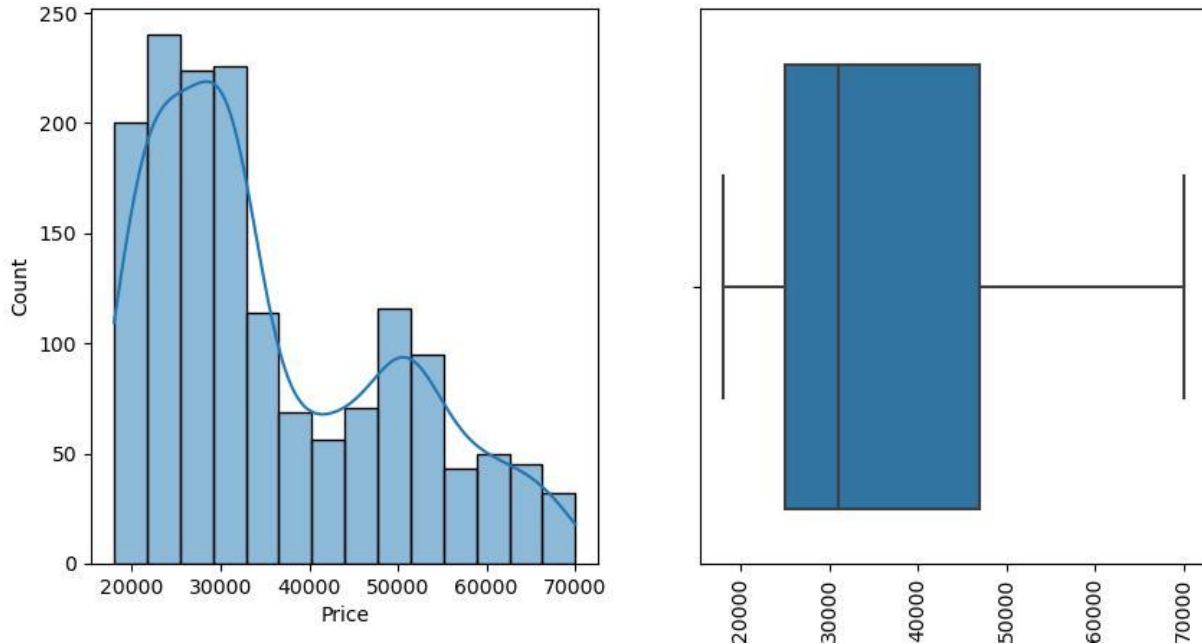
- More Buyers are less than 31 Years.
- No Outliers Detected

(b) Salary:



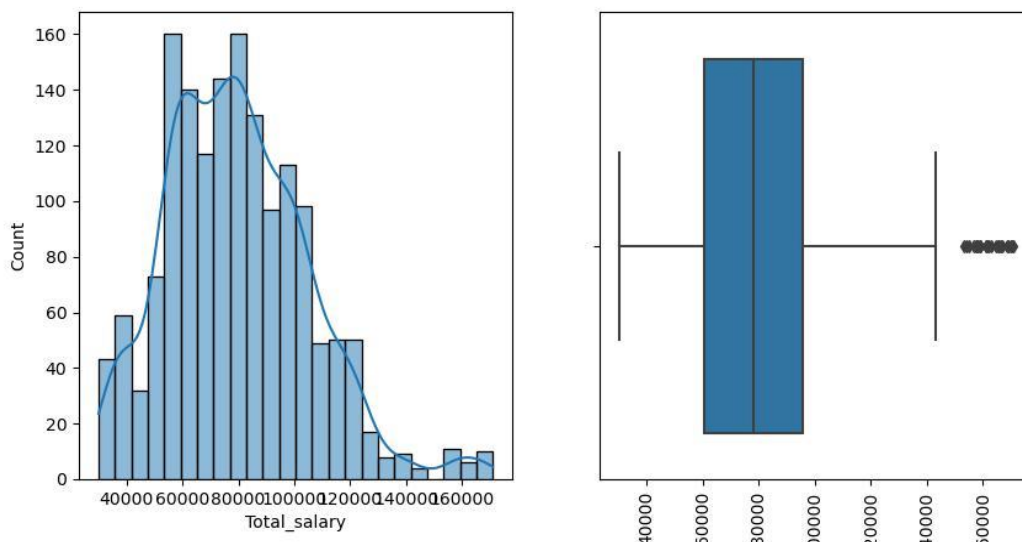
- Salary is Normally approximately Normally Distributed
- No Outliers detected.

(c) Price



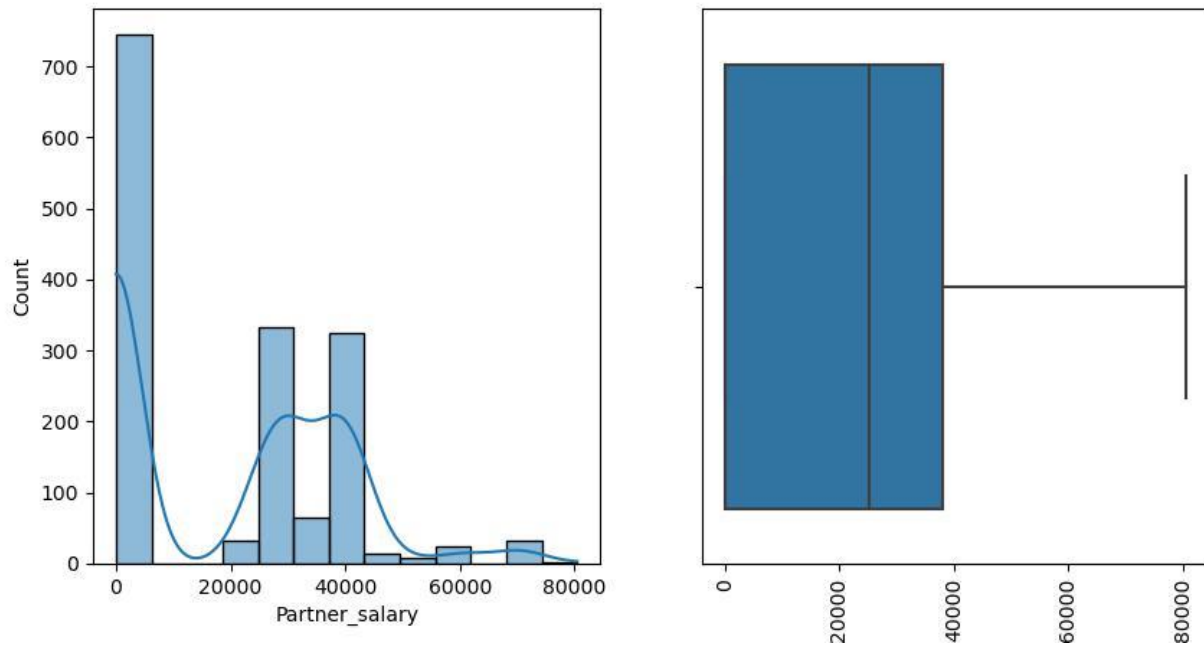
- Large number of Cars Priced around 20000 to 40000.
- No Outlier Treatment required.

(d) Total Salary



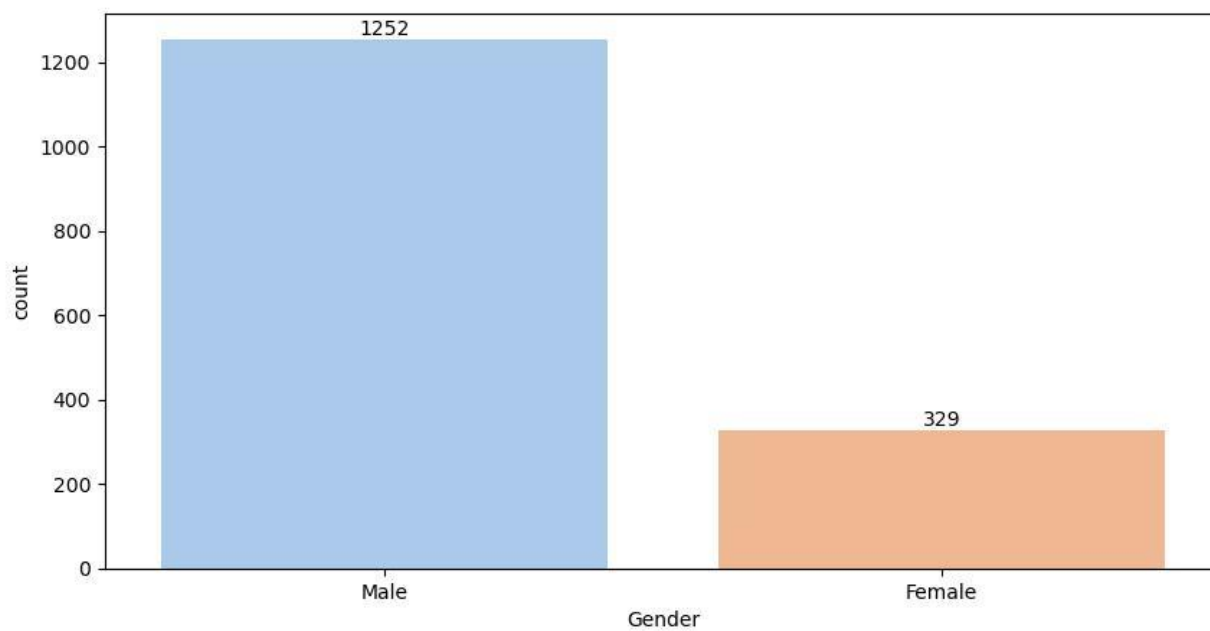
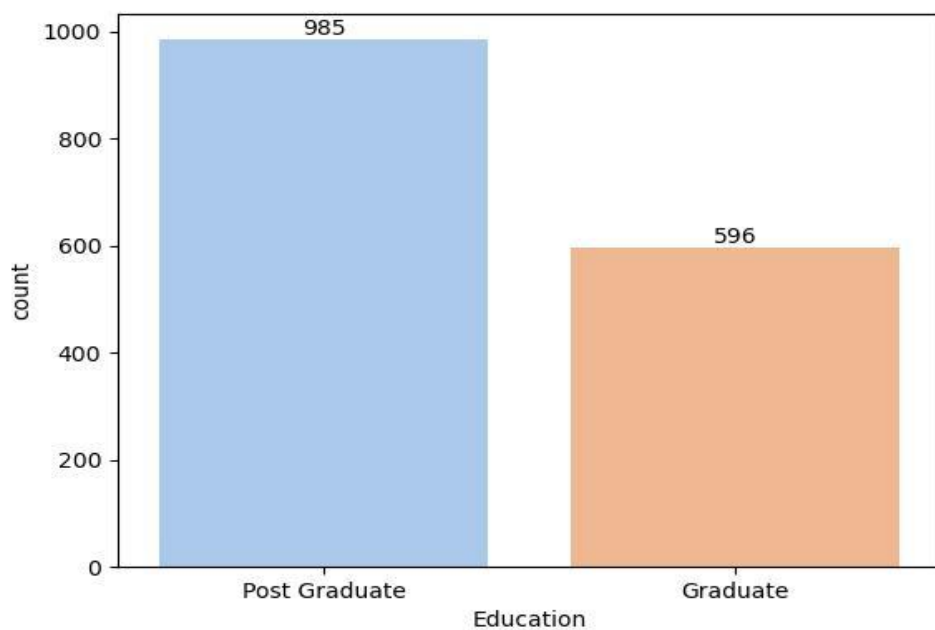
- 27 number of Total Salary greater than Whisker in the above Box Plot
- Even though Outliers are detected, since this is a UNSUPERVISED analysis, Outlier Treatment is not necessary.

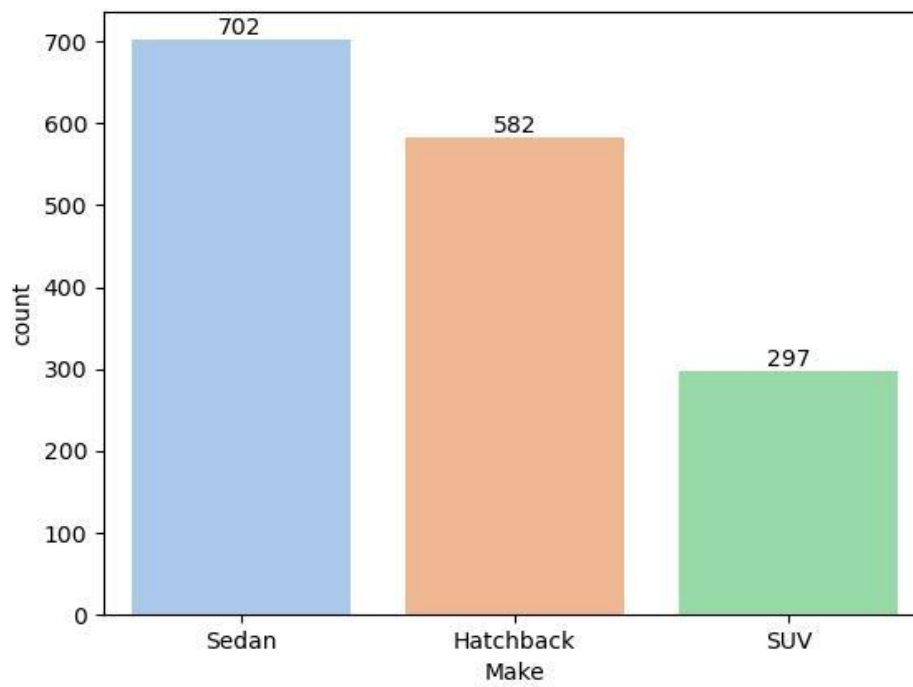
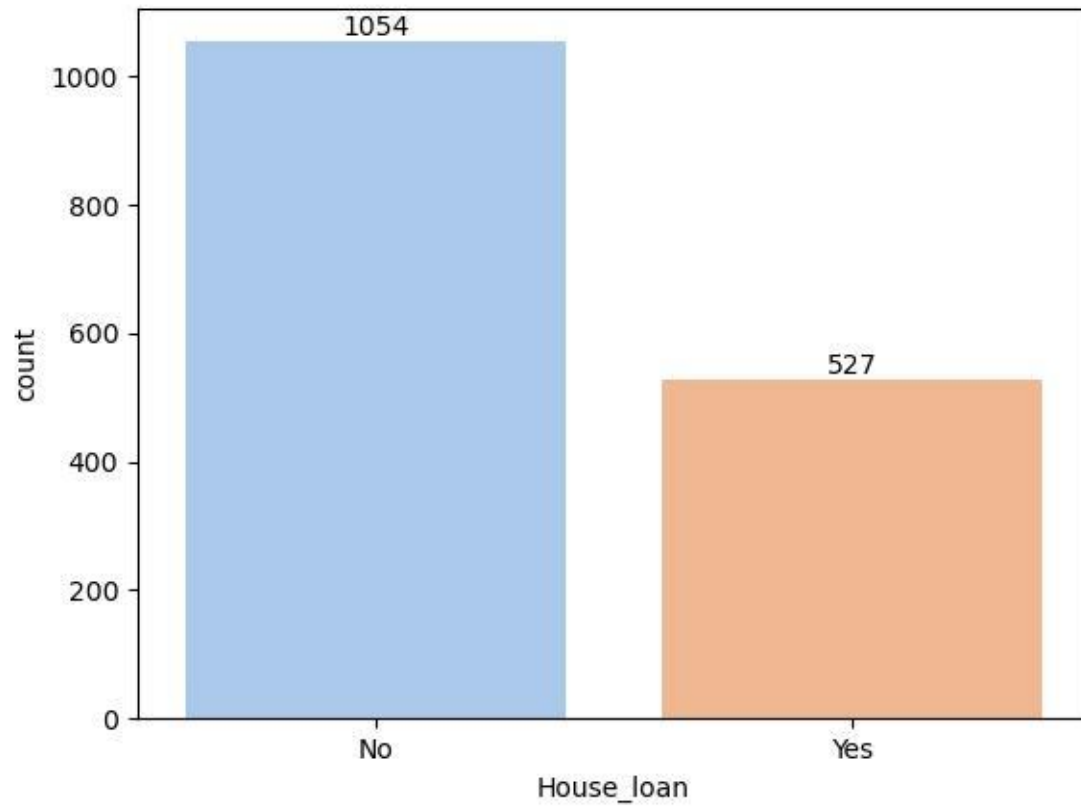
(e) Partner Salary

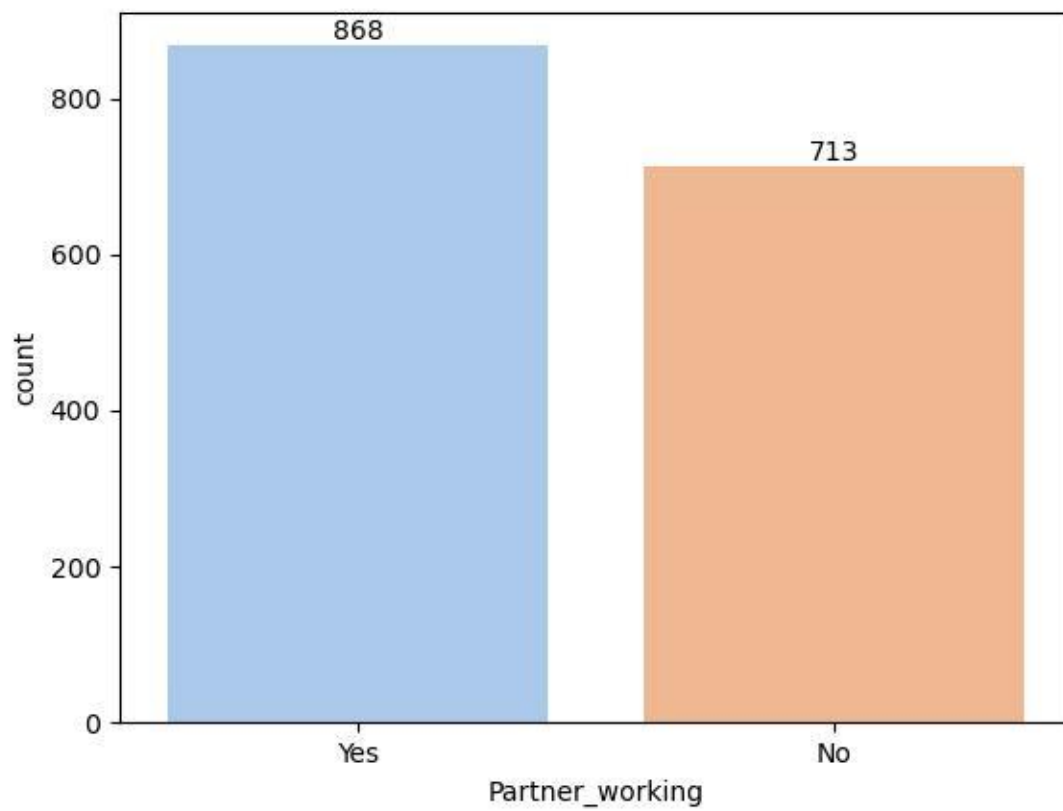
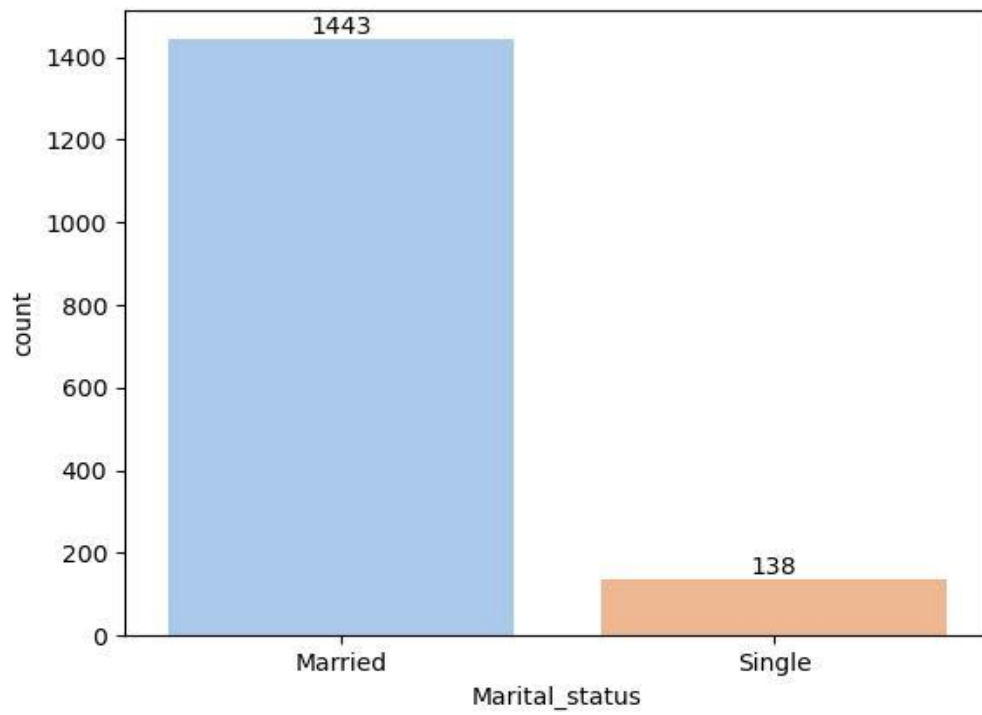


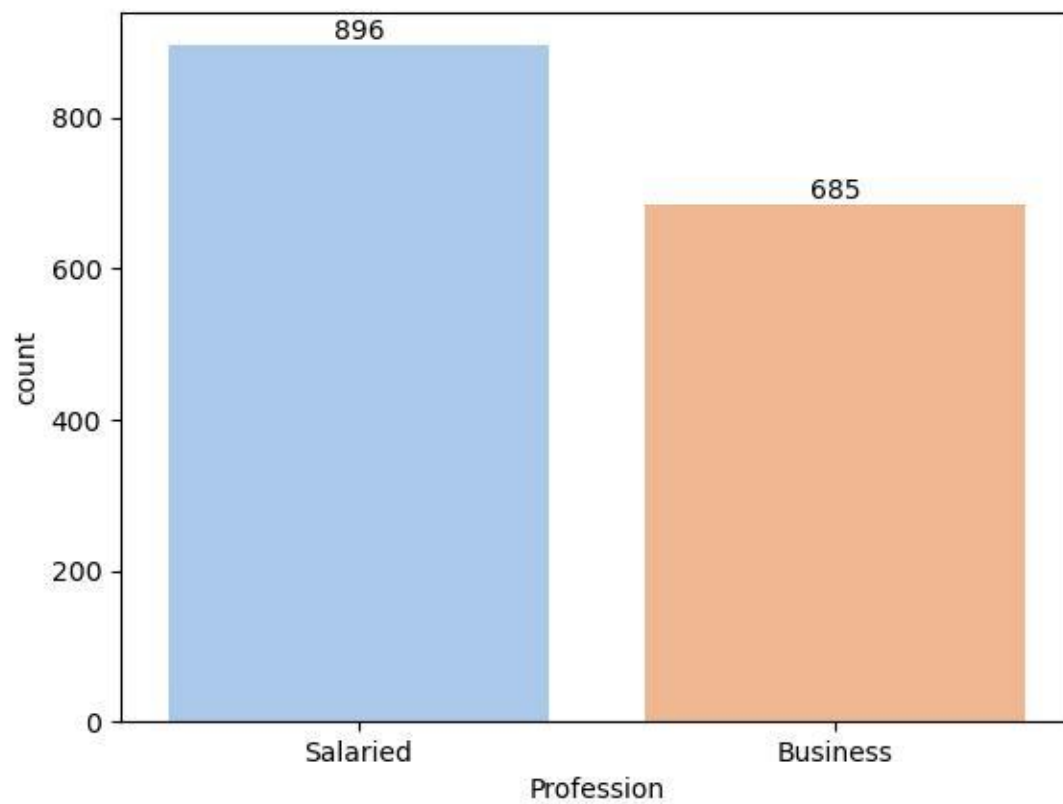
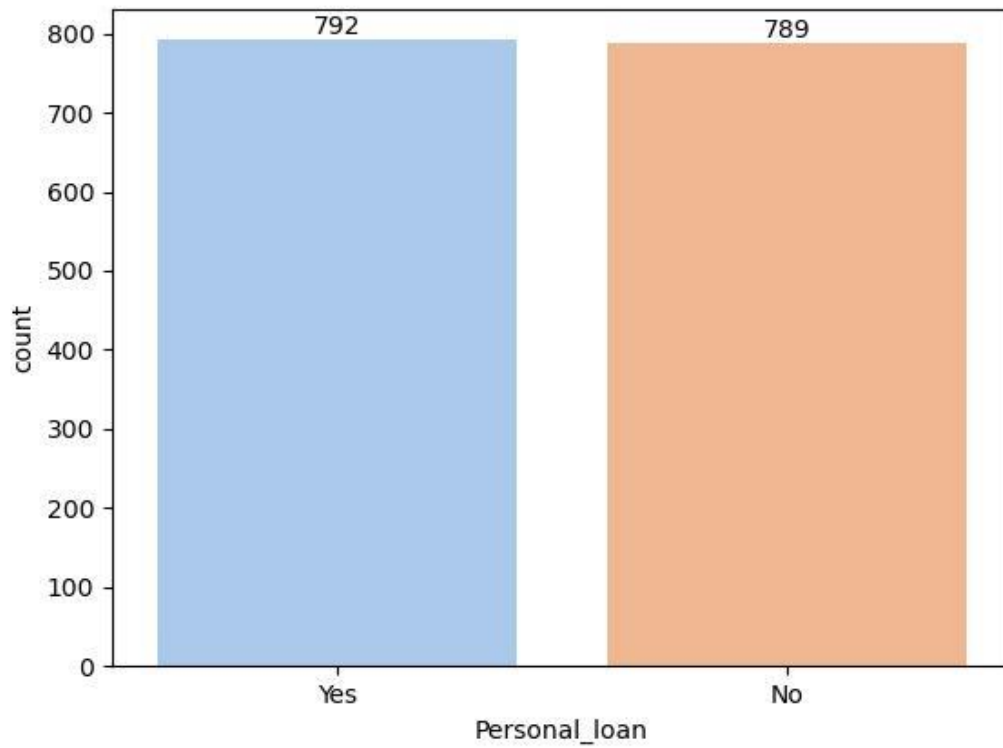
- Most number of Partner Contribution to Total Salary < 10000
- No Outlier Treatment Required

(f) Below is the Visualization of Categorical Columns and respective Value Counts:



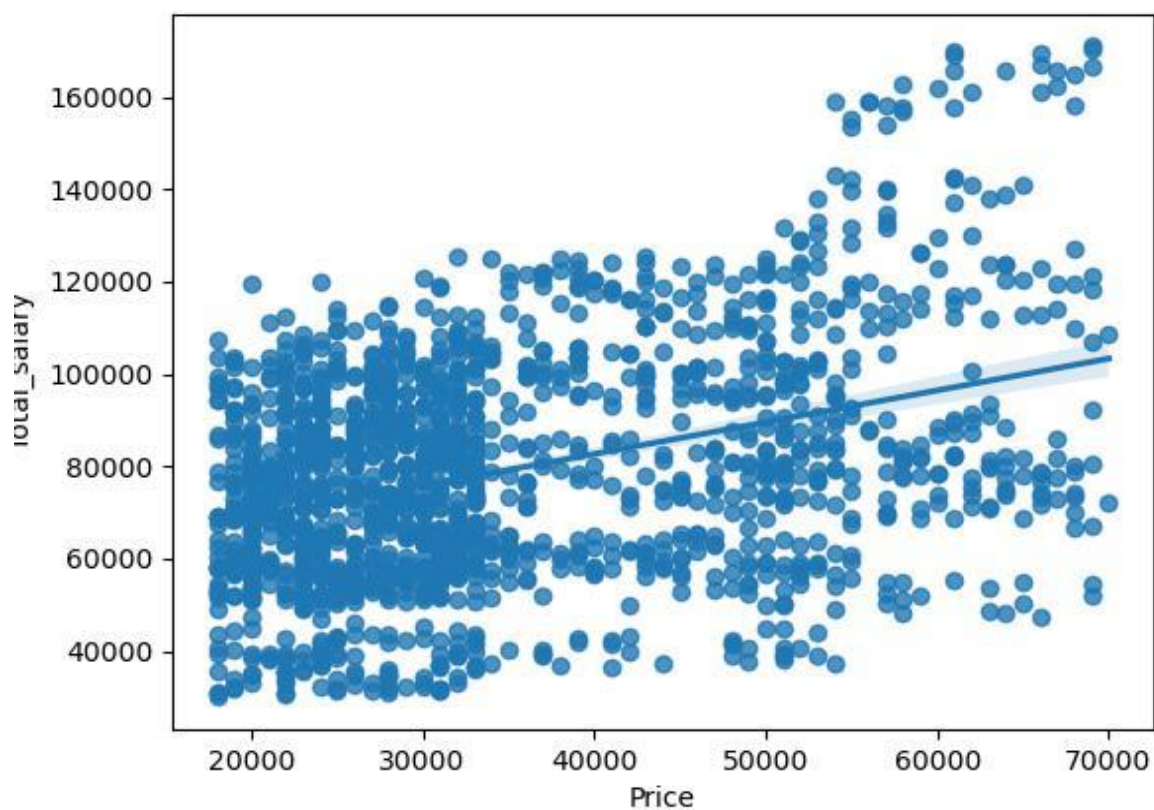






ANALYZE THE RELATIONSHIP BETWEEN VARIABLES

(a) Is there a Linear Relationship between Price of the Car and Total Salary

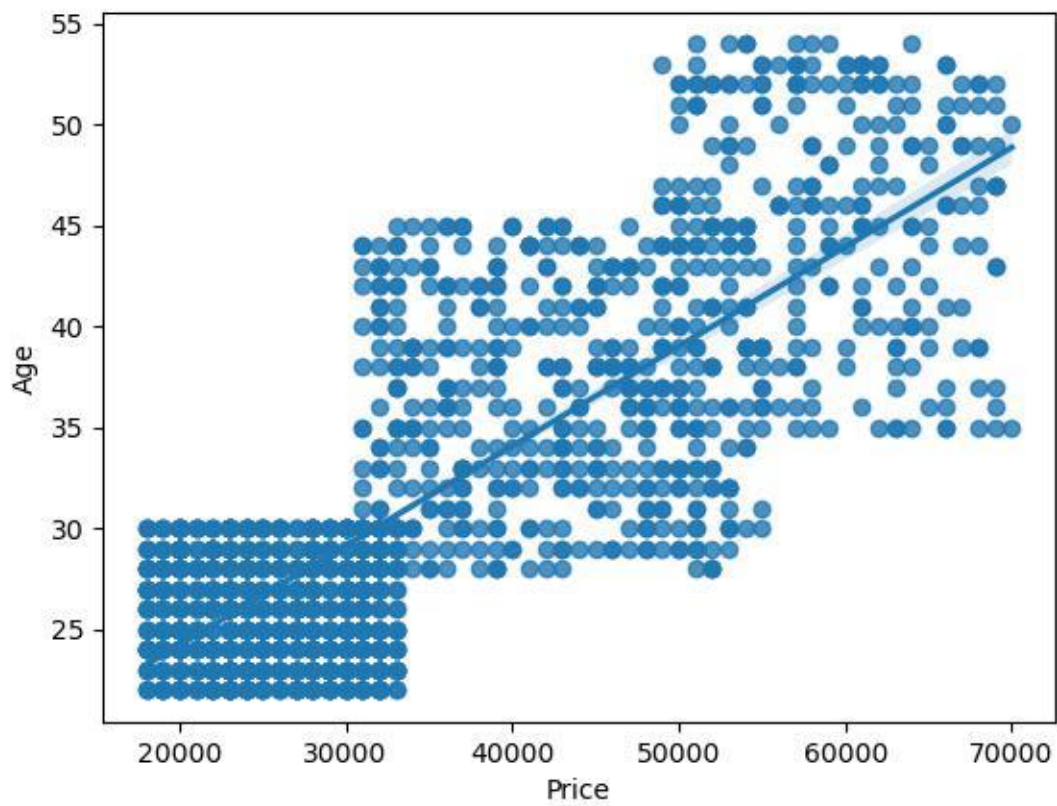
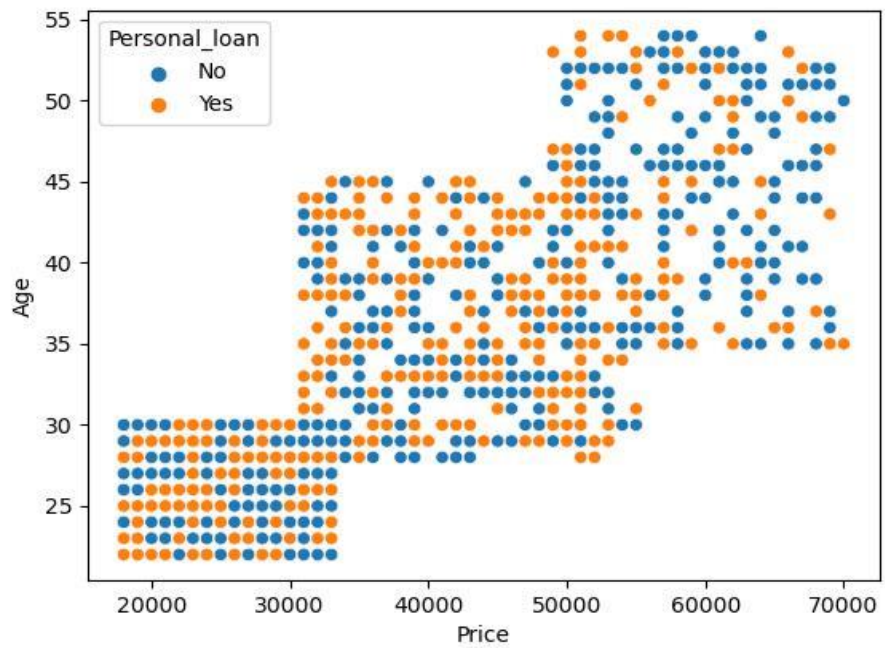


Correlation Matrix:

	Price	Total Salary
Price	1.0	0.367823
Total Salary	0.3678231	1.0

With 0.36 Correlation, not a Strong Linear Relation

(b) Age v/s Price



Correlation Matrix:

	Price	Age
Price	1.0	0.797
Age	0.797	1.0

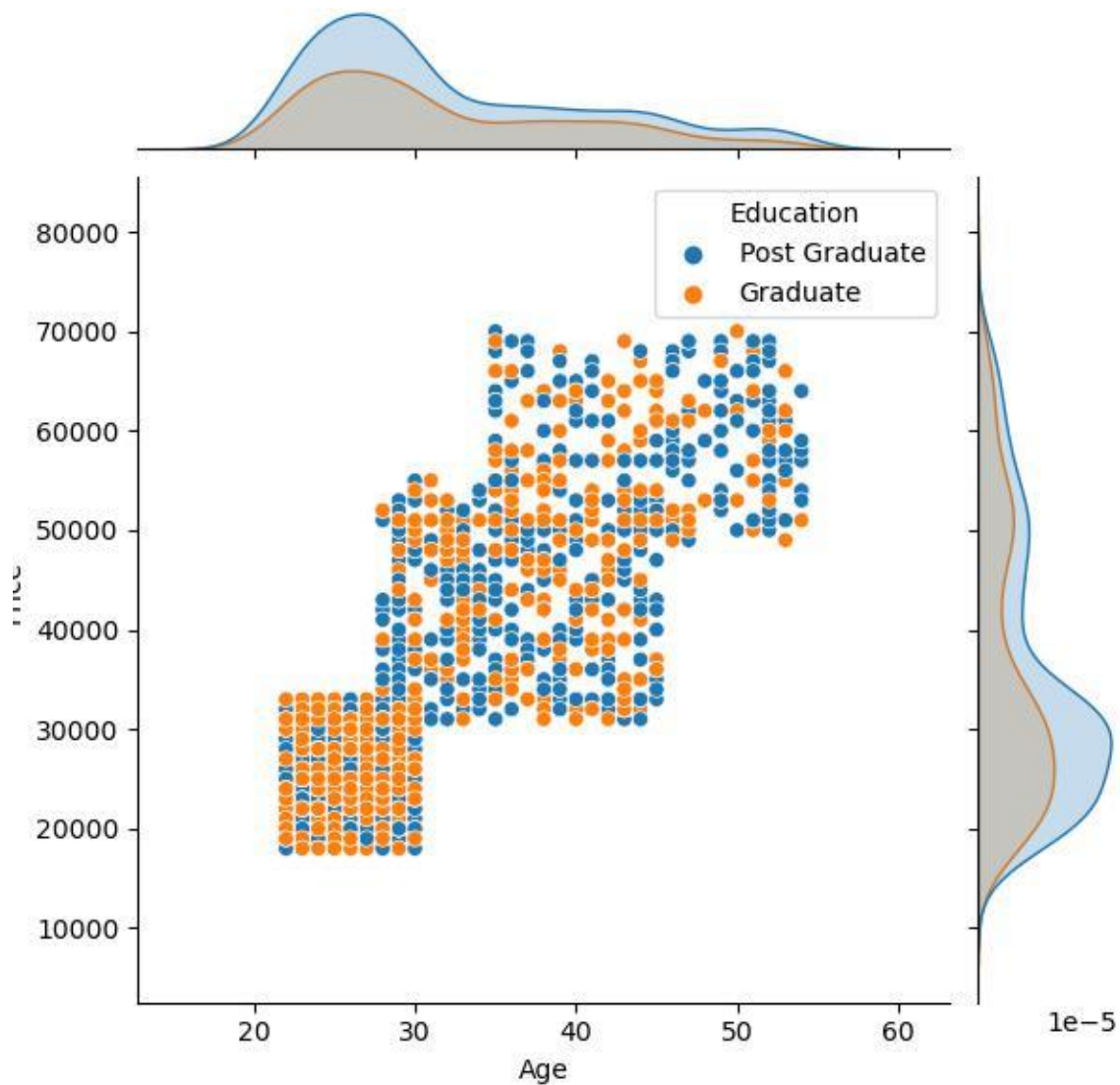
- **Age and Price are Strong Positively Correlated.**

inference:

- the Older Person buying expensive Cars and Younger People buy affordable Cars.

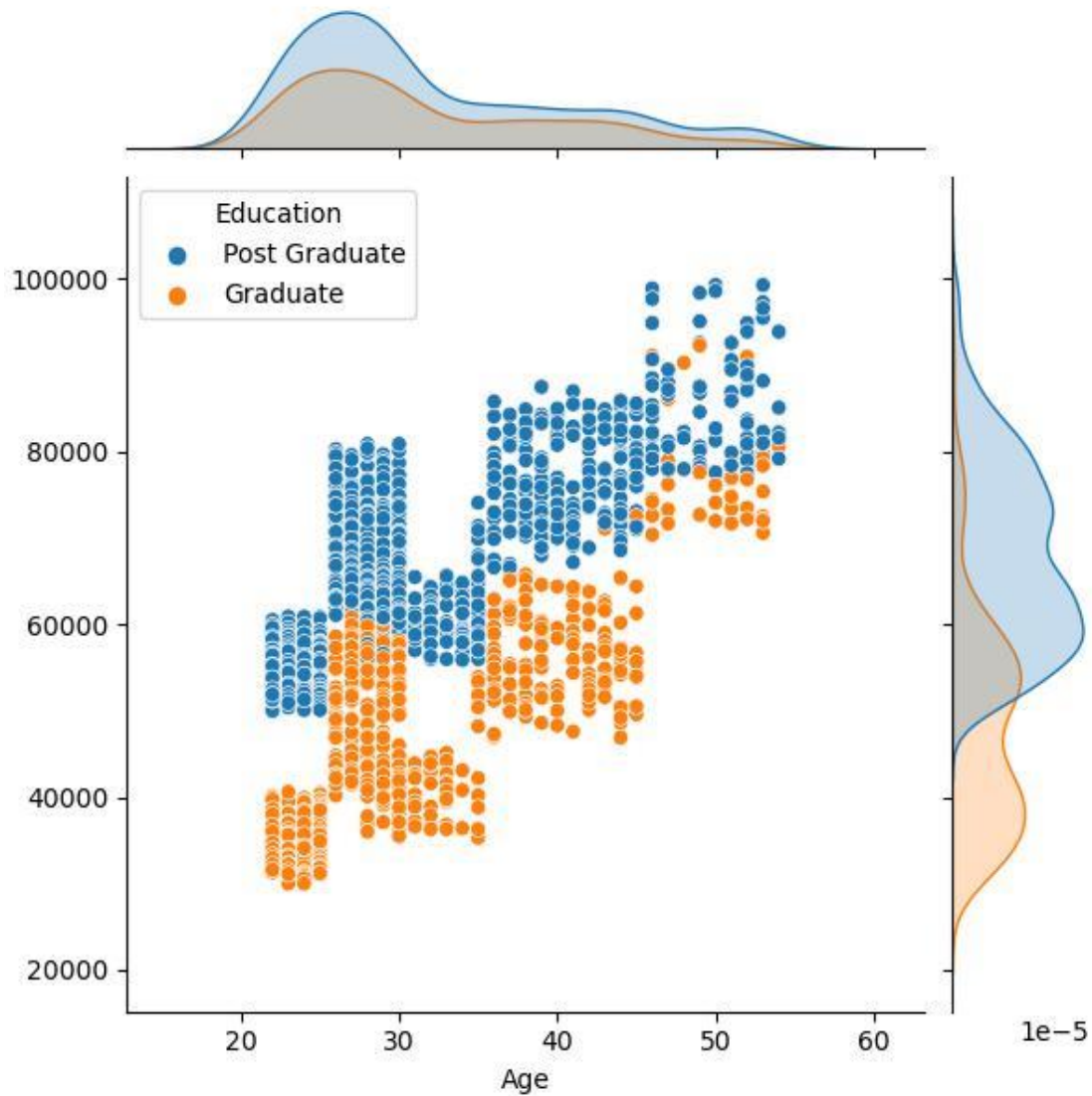
(c) Compare Age and Price along with Education.

- is Postgraduate or Graduate influence pattern on buying Price of Cars.

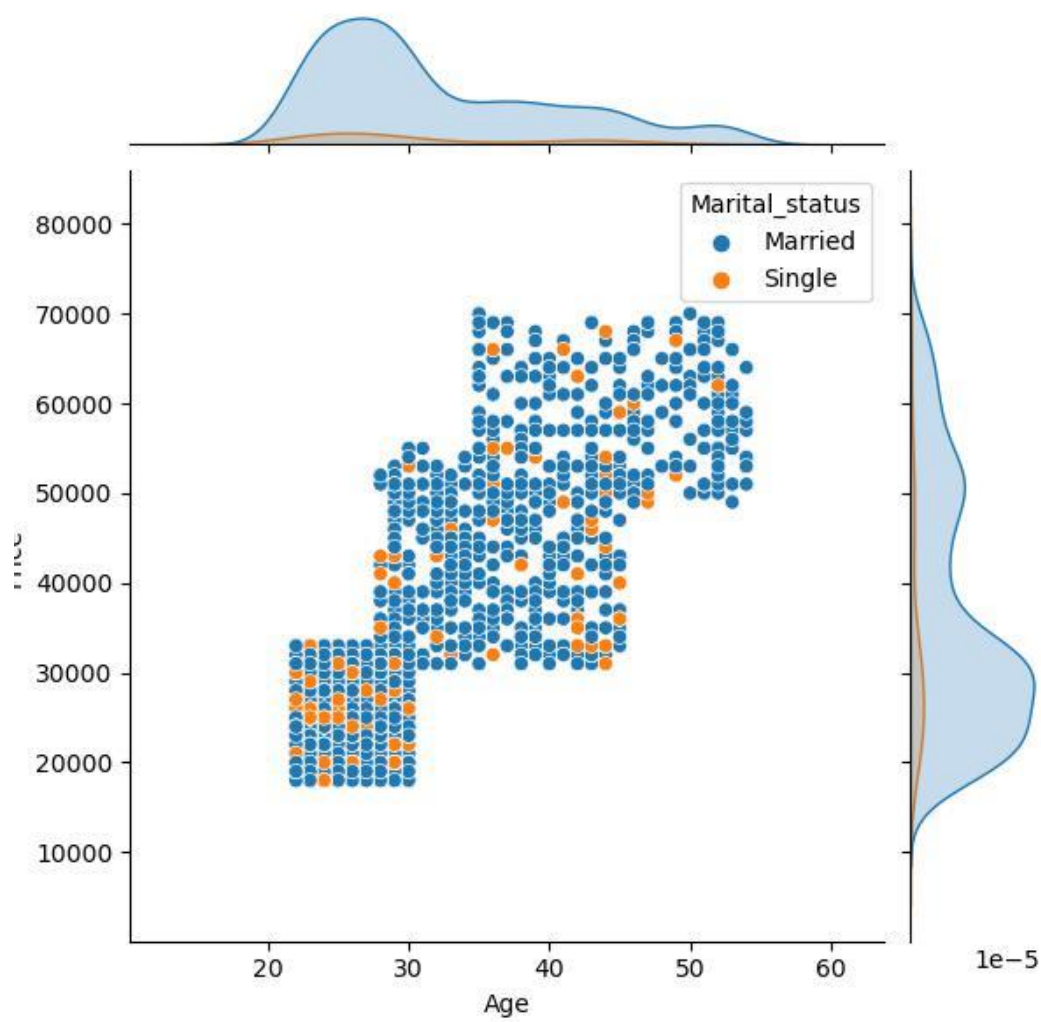


- Most Graduated Young People Buy affordable cars.
- Age greater than 30 years, no Strong Pattern on Purchasing with Education.

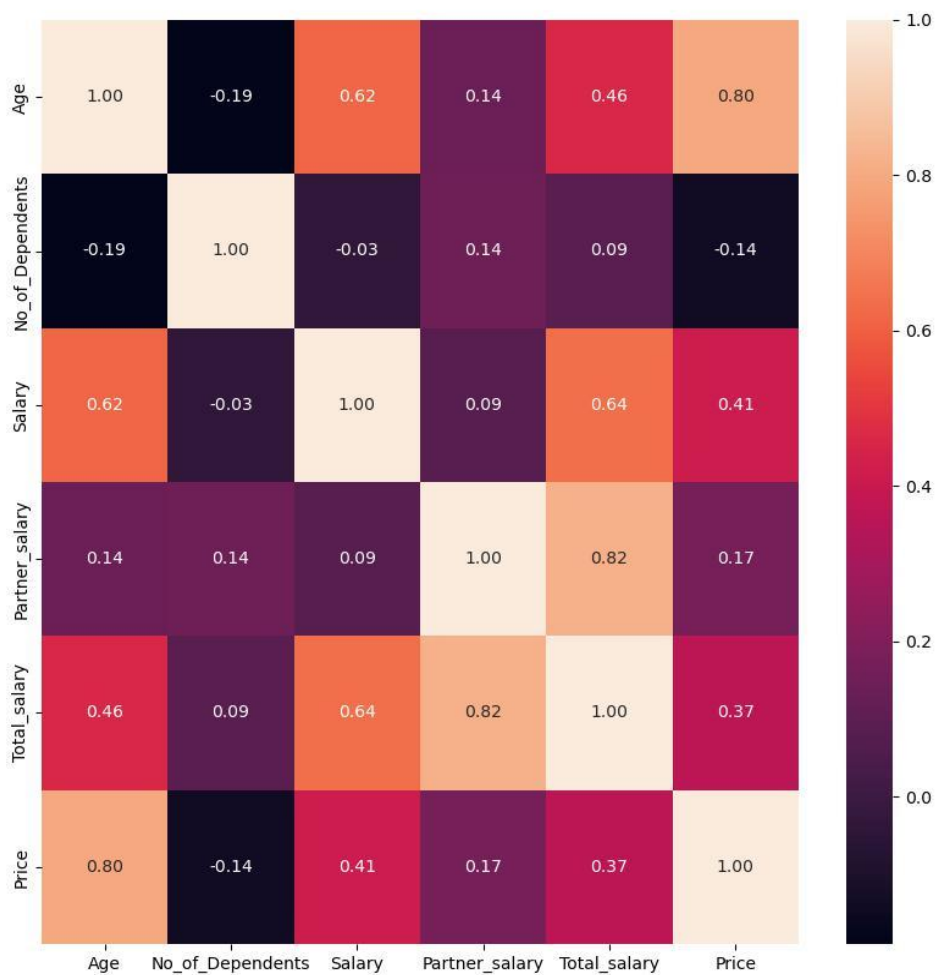
(d) Is Salary and Age Positively Correlated



- People with Postgraduate earn More than Graduate at similar Age.



(e) Visualizing Correlation Heatmap between All Numeric Fields.



Inference:

- With Age the Price of the Car purchased increases
- Salary also increases with Age.

From the Above analysis:

- People with High Total Salary not necessarily buying a Higher Price Cars
- People over 30 years are buying Higher Price Cars.
- Number of Dependents, Personal Loan, House Loan, Marital Status are not defining the Pattern on Car price of car purchased.

ANALYZE RELATIONSHIP BETWEEN CATEGORICAL VARIABLES

Tables Demonstrating Relationship

Profession/Gender	Business	Salaried
Female	6.64%	14.16%
Male	36.68%	42.50%

Marital Status/Gender	Married	Single
Female	19.41%	1.39%
Male	71.85%	7.33%

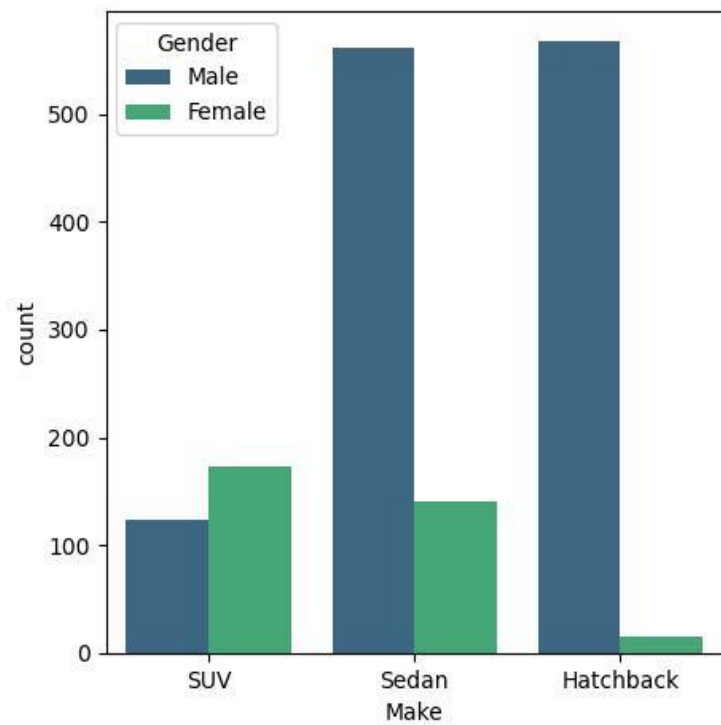
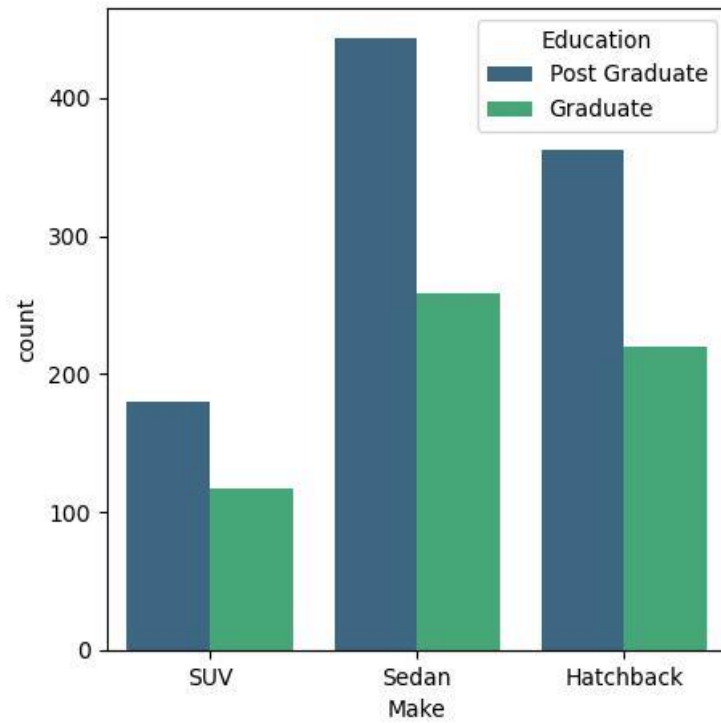
Education/Gender	Graduate	Postgraduate
Female	8.72%	12.08%
Male	28.96%	50.22%

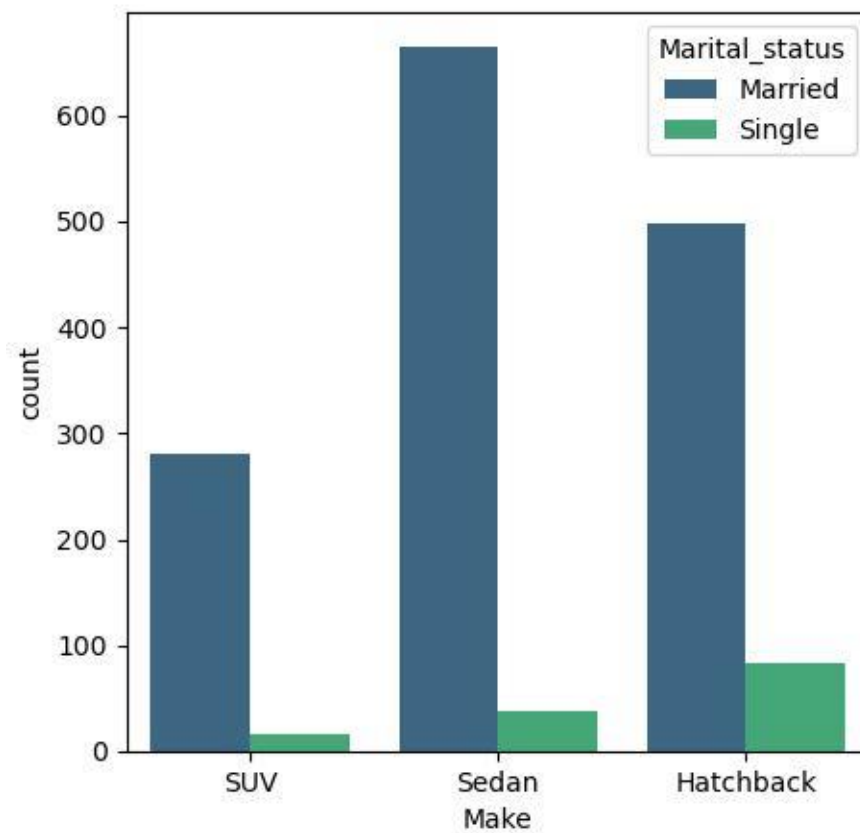
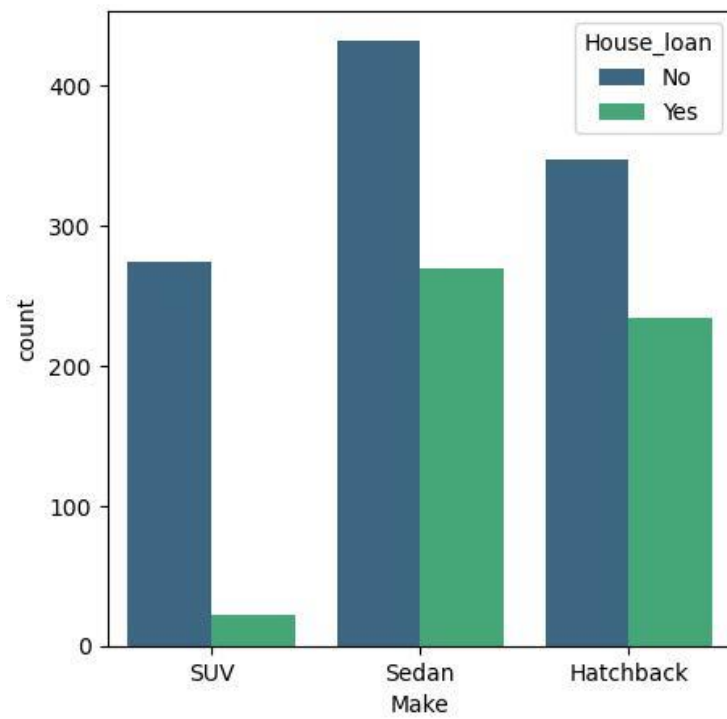
Education/Profession	Graduate	Postgraduate
Business	15.74%	27.57%
Salaried	21.94%	34.72%

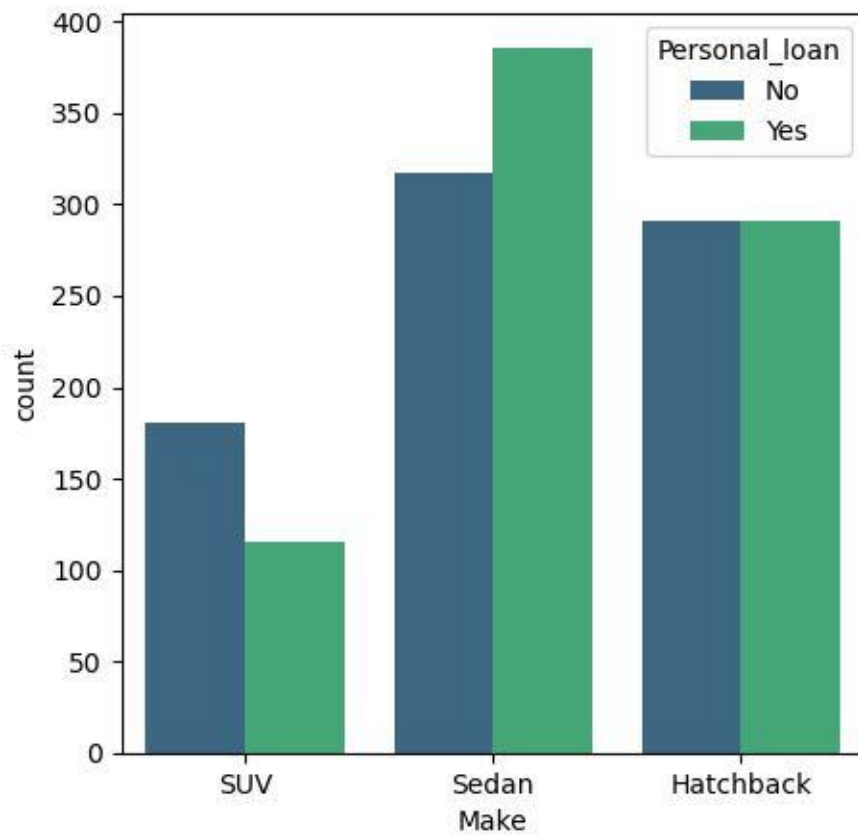
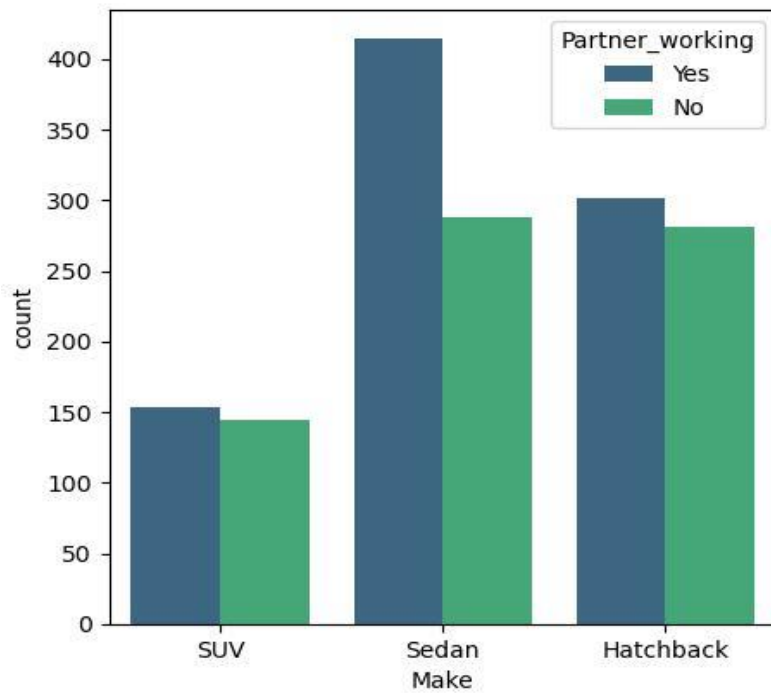
Profession/Make	Hatchback	SUV	Sedan
Business	18.34%	5.62%	19.35%
Salaried	18.46%	13.15%	25.04%

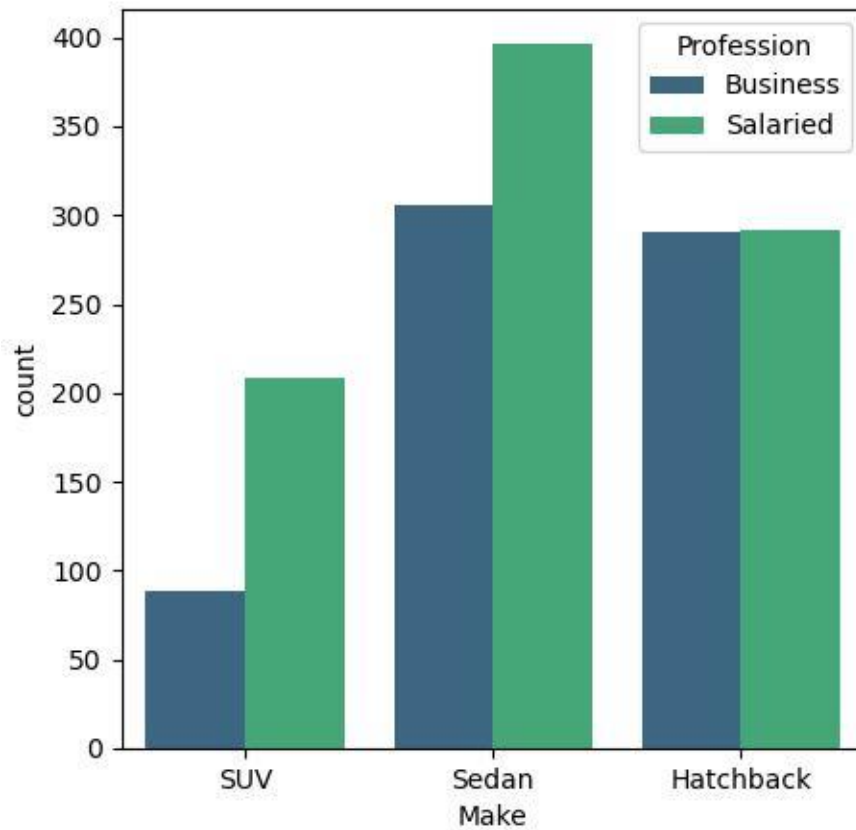
Gender/Make	Hatchback	SUV	Sedan
Female	0.94%	10.94%	8.91%
Male	35.86%	7.84%	35.48%

(a) Visualize the Categorical Field influencing the Make







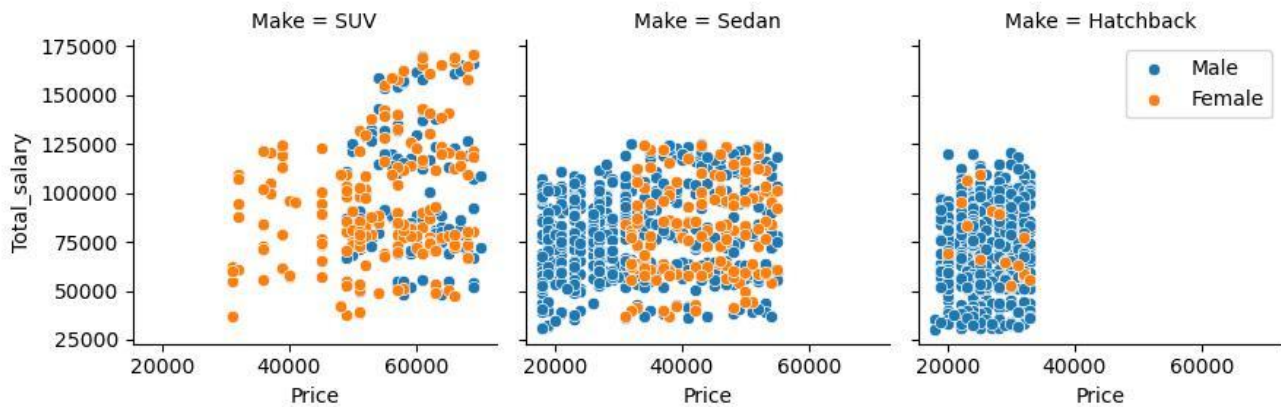


Inference:

- Male Purchase Sedan and Hatchback equally
- Female Purchase SUV more than they Purchase Sedan or Hatchback
- More number of Make Purchased is Sedan

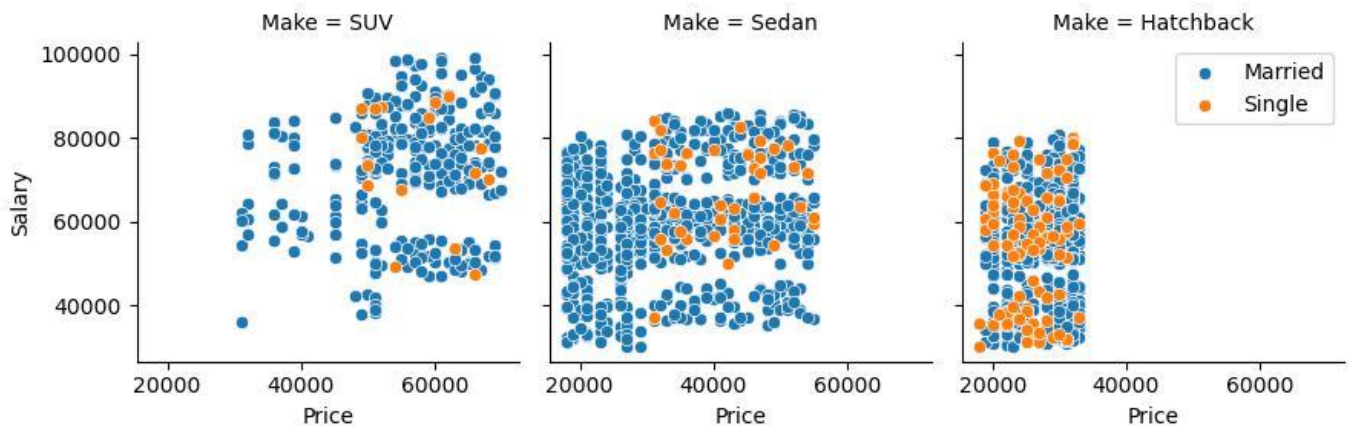
VISUALIZE MULTIPLE FIELD RELATIONSHIPS.

1 Price v/s Total Salary v/s Make v/s Gender.



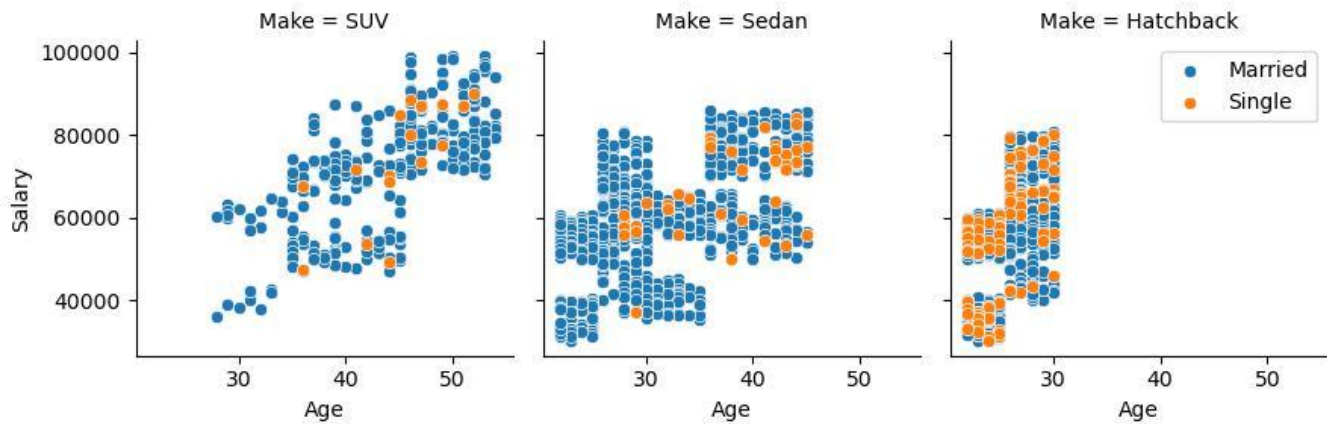
- SUV price is relatively higher than Sedan and Hatchback
- Females purchase More SUV and Sedan and Less Hatchback
- People with Total Salary greater than 125000 only purchase SUV.

2 Price v/s Salary v/s Make v/s Gender.



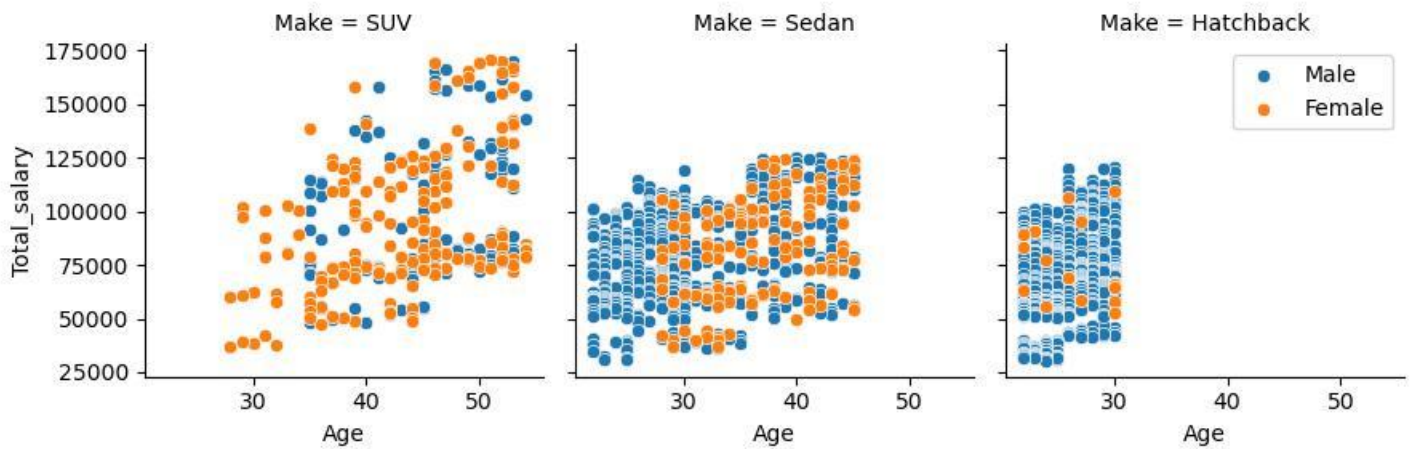
- Single People Mostly Purchase Hatchback
- Married People whose Salary above 90000 only end of purchasing SUV.

3 Age v/s Salary v/s Make v/s Marital Status.

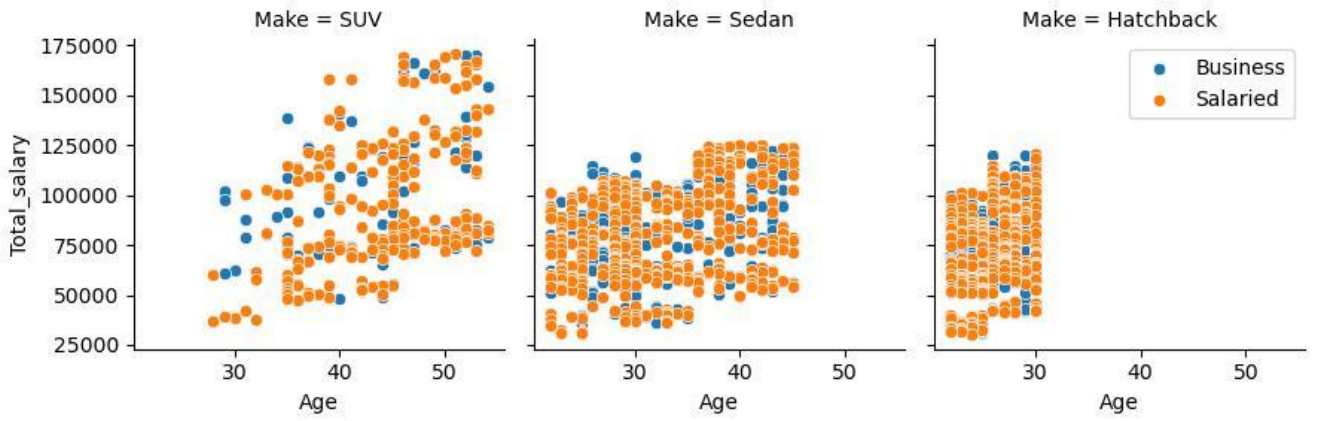


- Most People above age 30 are Married.
- Single People Mostly purchase Hatchback.

4 Age v/s Total Salary v/s Make v/s Gender.



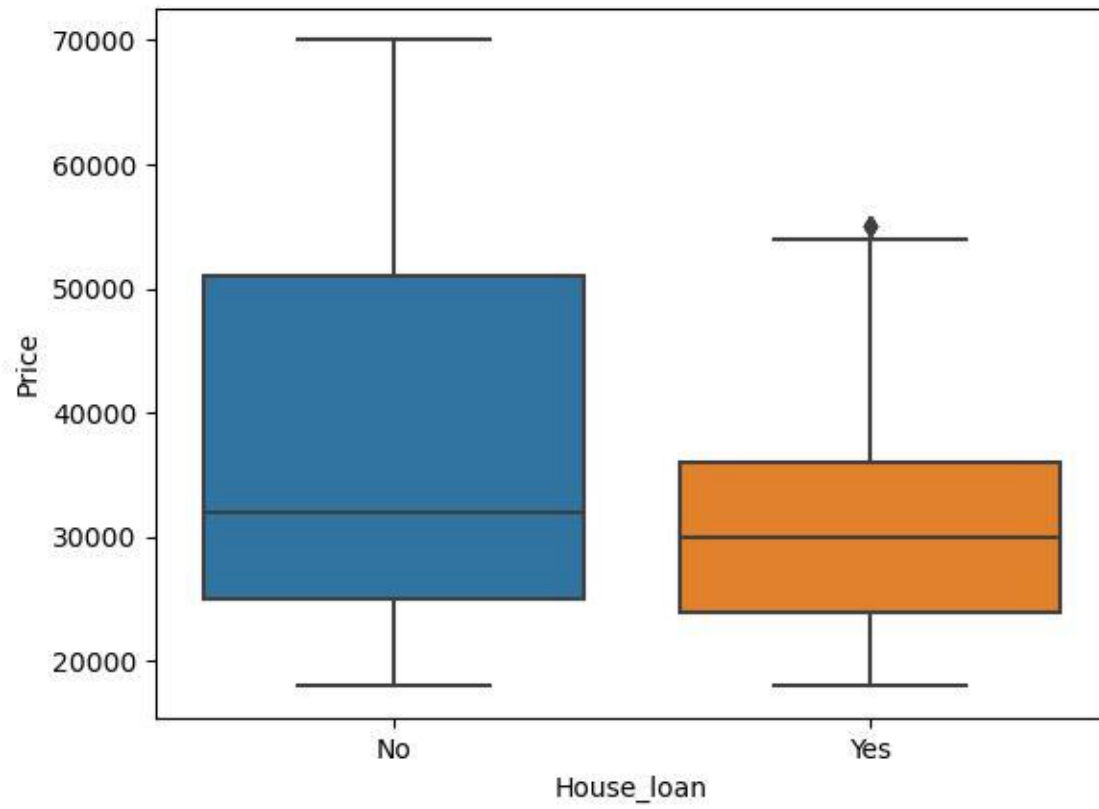
5 Age v/s Total Salary v/s Make v/s Profession.

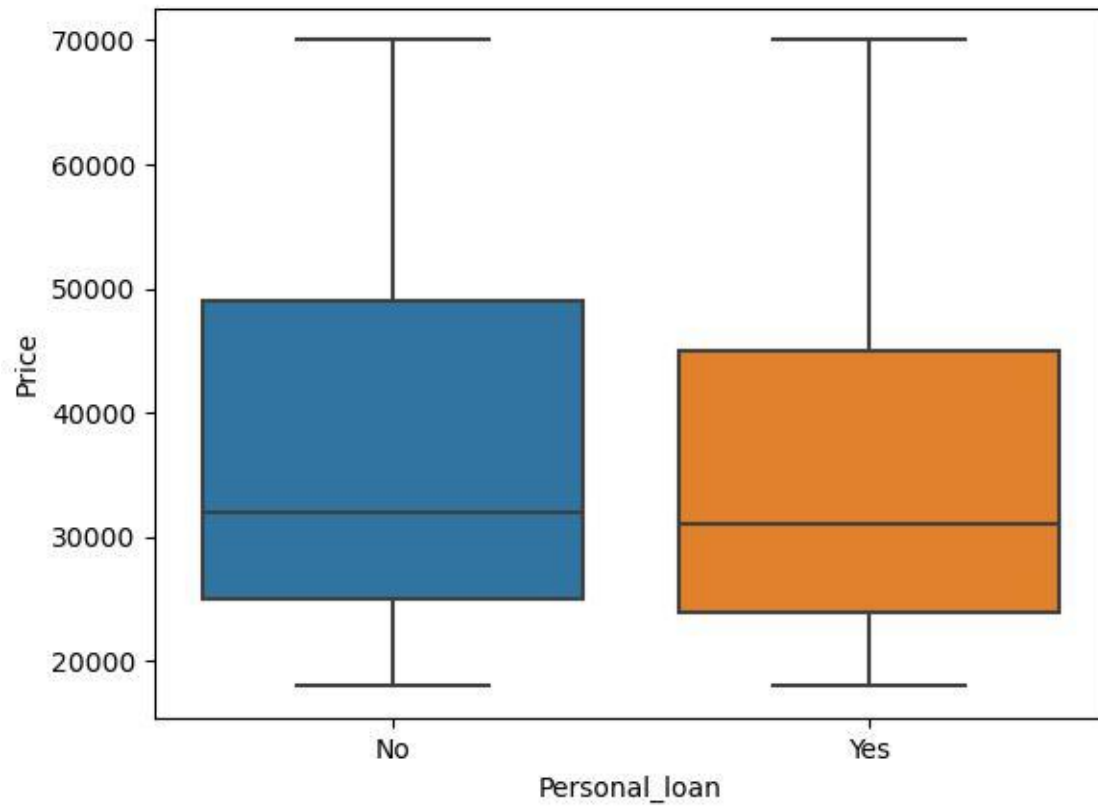


Inference:

- Single Salaried Male – Mostly purchase Hatchback.
- Females purchase SUV more than Sedan and Male Purchase Sedan More.
- Education fields do not define a pattern in purchase of a particular Make.
- Number of Dependents do not define a Strong correlation pattern of a Particular Make.

6 Median Spend of People with Personal and House Loan:





Inference:

- Median Purchase Price of People with and without personal Loan is almost similar.
- Median Purchase Price and Upper Purchase Price of People without House Loan is Higher than People with House Loan.

KEY QUESTIONS TO ANSWER:

1. Do Men Tend to Prefer SUVs more compared to Women?
 - No, justification below.

Make/Gender	Hatchback	SUV	Sedan
Female	15	173	141
Male	567	124	561

With Rank Dense and Descending Mode:

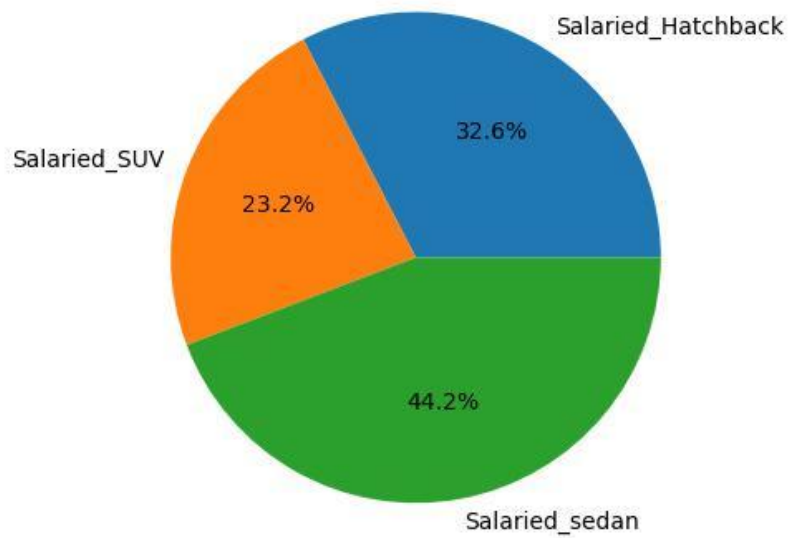
Make/Gender	Hatchback	SUV	Sedan
Female	2.0	1.0	2.0
Male	1.0	2.0	1.0

- Total number of SUVs bought by Female is greater than SUVs bought by Male
- Also, the relative number - rank within the Gender Category SUVs bought by Female is More.

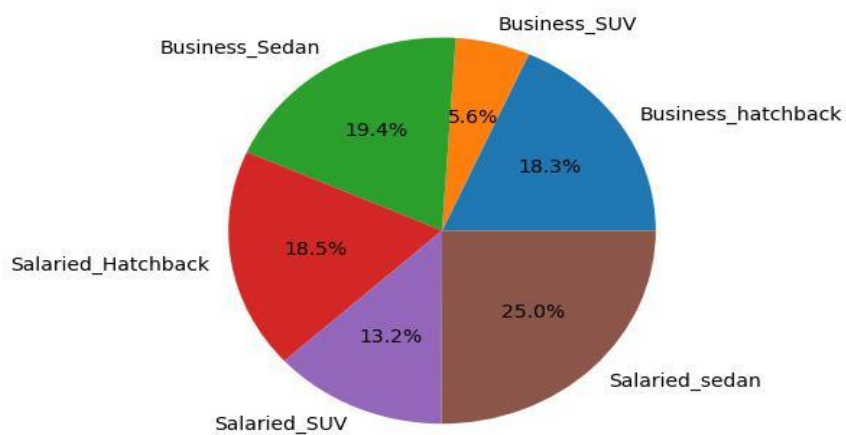
2. What is the Likelihood of a Salaried Person Buying a Sedan?

- 2 Methods used:

- Person likely to buy Sedan within Salaried Group is 44%



- Salaried Person likely to Buy Sedan in the Entire is 25%



3. 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?

Table with SUM of Price v/s Category

	Make	Hatchback	SUV	Sedan	All
Profession	Gender				
Business	Female	NaN	2944000.0	2121000.0	5065000
	Male	7629000.0	2074000.0	8079000.0	17782000
Salaried	Female	412000.0	6308000.0	3910000.0	10630000
	Male	7367000.0	5254000.0	10182000.0	22803000
All		15408000.0	16580000.0	24292000.0	56280000

Table Below with Count of Make v/s Category.

	Make	Hatchback	SUV	Sedan
Gender	Profession			
Female	Business	NaN	55.0	50.0
	Salaried	15.0	118.0	91.0
Male	Business	290.0	34.0	256.0
	Salaried	277.0	90.0	305.0

Inference:

- Both Table dictate Salaried Men Tend to Purchase Sedan v/s SUV

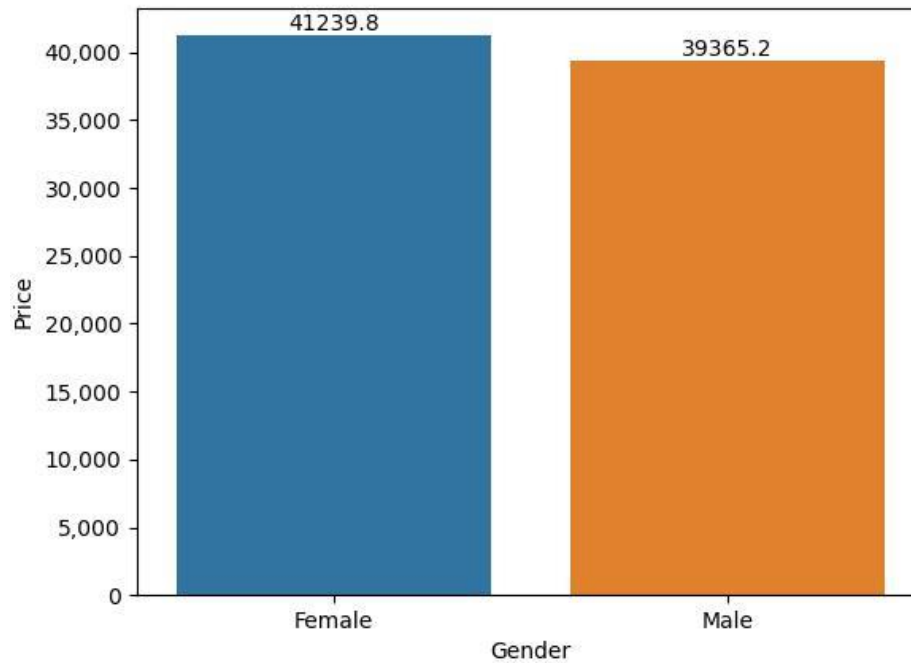
4. How does the Amount Spent on Purchasing Automobiles Vary by gender?

Table with Sum of Amount Spent vary by Gender.

Make	Hatchback	SUV	Sedan	All
Gender				
Female	412000	9252000	6031000	15695000
Male	14996000	7328000	18261000	40585000
All	15408000	16580000	24292000	56280000

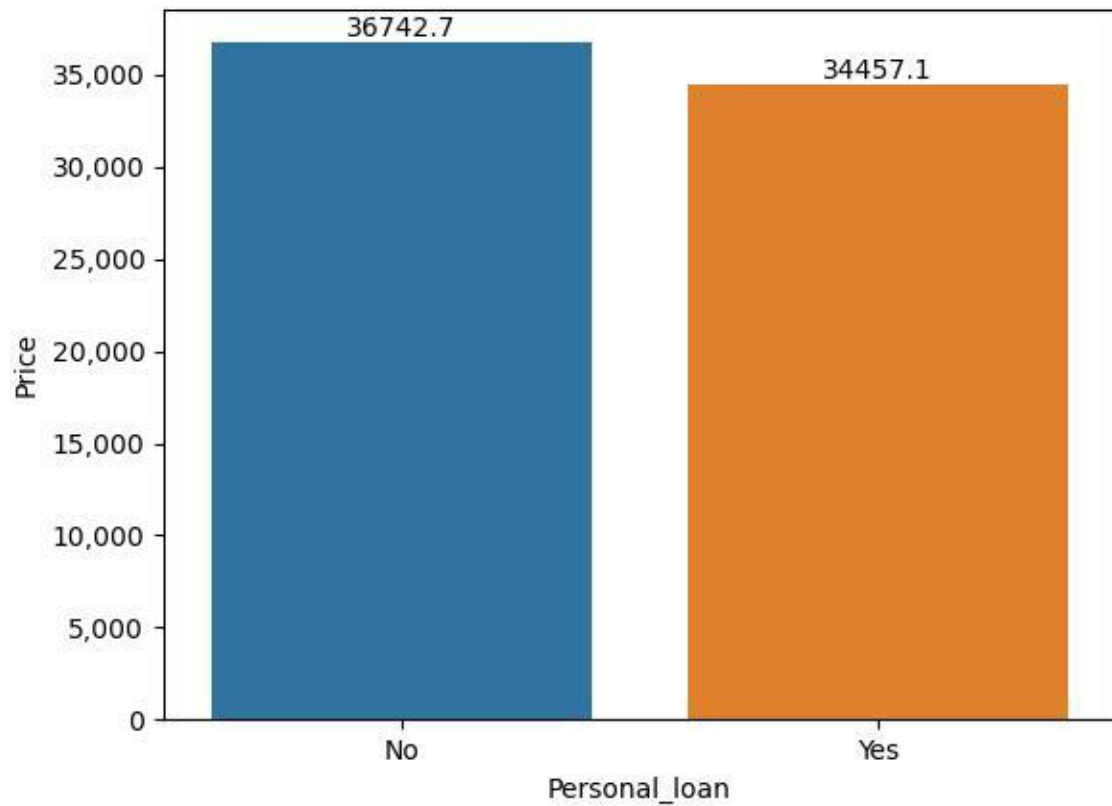
Table with Average Amount Spent vary by Gender.

	Gender	Make	Price(mean)
0	Female	Hatchback	27466.666667
1	Female	SUV	53479.768786
2	Female	Sedan	42773.049645
3	Male	Hatchback	26447.971781
4	Male	SUV	59096.774194
5	Male	Sedan	32550.802139



- On Average Female Avg amount spent on Car Purchase is relatively higher than Male Spending. This delta could be due to multiple reason, one of the reasons – Single Male mostly purchase less expensive Hatchback which left skews the overall spend.

5. How much Money was spent on purchasing Automobiles by Individuals who took a Personal Loan?



- People with Personal Loan on an Avg Spend relatively less than People without Personal Loan. But the difference is less and do not add significance to define a Pattern or Conclude an insight.

6. How does having a working partner influence the Purchase of higher-priced car?

- Car Prices are divided to 3 buckets -> LOW, MEDIUM, HIGH

Table with Mean Partner Salary

Partner working	No	Yes
price bucket		
LOW	0.0	30532.103321
MEDIUM	0.0	36945.544554
HIGH	0.0	51590.322581

Table with Count of Partner Working when Person bought a Car.

Partner working	No	Yes
price bucket		
LOW	440	542
MEDIUM	159	202
HIGH	114	124

- People who Bought High Price Car do not Critically Depend on Partner Working. Since the difference between Partner working and non-working is Less. For Medium Priced Car – a working Partner adds significance value.

ACTIONABLE INSIGHTS AND BUSINESS RECOMMENDATIONS

Customer Segmentation based on Price Bucket:

- Price Bucket segment customers to groups based on Purchasing power. This approach can be used to develop Tailored strategies to meet the preferences of each segment.

Targeted Marketing campaigns:

- Females prefer purchasing SUV, Single Male prefer Hatchbacks, People below 30 years prefer Hatchbacks, People 30+ years prefer Sedan or SUV, People Total Salary greater than 125000 only prefer SUV.
- Based on the above probabilities deploy targeted Marketing campaigns and Value propositions.

Customer Feedback:

- Customer satisfaction score is not available in this data set. This is one the Key field to understand the customer trend and index retaining or migration to a different make.

Financial Services:

- Customers with House Loan mean spend is relatively Lower than Customers without House Loan. Offer Financial Aid to such customers to motivate them to purchase higher Price Cars

Innovations:

- Customers with Business Profession prefer Sedan or Hatchback v/s SUV. Product R&D team to develop features in SUV which benefit Business Professionals (e.g. Internet in the Car, features to use Laptop effectively in the Car, Long driving, and camping functionalities).

PROBLEM 2

OBJECTIVE

- Analyze the dataset and list down the top 5 important variables, along with the business justifications:

Fields in the Dataset:

#	Column	Non-Null Count	Dtype
0	userid	8448 non-null	int64
1	card_no	8448 non-null	objects
2	cards_bin_no	8448 non-null	int64
3	Issuer	8448 non-null	object
4	cards_type	8448 non-null	object
5	cards_source_date	8448 non-null	datetime64[ns]
6	high_networth	8448 non-null	object
7	actives_30	8448 non-null	int64
8	actives_60	8448 non-null	int64
9	actives_90	8448 non-null	int64
10	cc_active30	8448 non-null	int64
11	cc_active60	8448 non-null	int64
12	cc_active90	8448 non-null	int64
13	hotlists_flag	8448 non-null	object
14	widgets_products	8448 non-null	int64
15	engagements_products	8448 non-null	int64
16	annuals_income_at_source	8448 non-null	int64
17	others_bank_cc_holding	8448 non-null	objects
18	banks_vintage	8448 non-null	int64
19	T+1_month_activity	8448 non-null	int64
20	T+2_month_activity	8448 non-null	int64
21	T+3_month_activity	8448 non-null	int64
22	T+6_month_activity	8448 non-null	int64
23	T+12_month_activity	8448 non-null	int64
24	Transactor_revolver	8410 non-null	object
25	avg_spends_13m	8448 non-null	int64
26	Occupation_at_source	8448 non-null	object
27	cc_limit	8448 non-null	int64

TOP 5 IMPORTANT VARIABLES:

1. Active_30/60/90 (Activity status for the user in last 30,60,90 Days) & CC_active_30/60/90 (Credit card activity status for the user in last 30,60,90)
 - Both fields jointly to be compared to understand the user's engagement on spending during a time frame.
 - Higher activity levels indicate user satisfaction, while declining activity levels may indicate potential churn.
 - Since both are in Numerical in nature, can be used directly to ML Train the data set.
2. Transactor Revolver (Transactor or revolver):
 - Identifying the users to assess credit risk. Revolvers spending and paying the minimum due amount are profitable to banks as they end up paying interest, at the same time they are at risk and might become defaulters.
 - Understanding the spend amount of Revolvers, vary the credit limit proportionately – so that risk of default is reduced and profits due to interest can be obtained.
3. T+1_month_activity, T+2_month_activity, T+3_month_activity, T+6_month_activity, T+12_month_activity:
 - This field can be used to understand the seasonal purchase behaviors of the users.
 - If respective users are active during festival periods or only during E-commerce sale period etc.
4. Occupation at Source & Annual income at Source & high Net worth:
 - These 2 fields together reflect the annual income of the user, which can be used to perform Risk assessment, provide personalized offers and financial services, different methods to engage customers and conduct income distribution.
5. Other Bank CC Holding:
 - Insights if user holds multiple credit cards or requires diverse financial needs:
 - E.g. if a USER has 2 cards, one card offers cash back on e-commerce and one offers better cashback for fuel. This is diverse financial needs. If one bank can provide a personalized offer to the USER reflecting cashback for both needs and analyze the USER retention.
 - Enhance Customer Engagement and Cross selling Opportunities.