BUSINESS REPORT ON SMDM PROJECT

(Austo Automobile + GoDigt Credit Card)

-RAHUL SHARMA

FIGURES: -

Problem1: Austo automobile-

A. DATA OVERVIEW: -

A1. Check the structure of the data (no. of rows and columns)

Answer:- (1581,14)

A2. Check the types of the data.

Answer:-

```
Data Types:
Age                  int64
Gender              object
Profession          object
Marital_status      object
Education           object
No_of_Dependents     int64
Personal_loan       object
House_loan          object
Partner_working     object
Salary               int64
Partner_salary     float64
Total_salary         int64
Price                int64
Make                object
dtype: object
```

Fig.(1): Data types

A3. Check the missing values and treat them if needed.

Answer:- There are only two columns which has missing values
i.e. Gender-53, Parter salary- 106

```
Missing Values:
Age                   0
Gender               53
Profession            0
Marital_status        0
Education             0
No_of_Dependents      0
Personal_loan         0
House_loan            0
Partner_working       0
Salary                0
Partner_salary      106
Total_salary          0
Price                 0
Make                  0
dtype: int64
```

Fig.(2): Missing Values

So, now we have to treat the missing values.

```
Age                0
Gender             0
Profession         0
Marital_status     0
Education          0
No_of_Dependents   0
Personal_loan      0
House_loan         0
Partner_working    0
Salary             0
Partner_salary     0
Total_salary       0
Price              0
Make               0
dtype: int64
```

Fig.(3): Treated Missing Values

A4. Statistical Summary

Answer:- Here we easily analyze min.,max.,mean,etc of different categories.

```
Statistical Summary:
                Age  No_of_Dependents        Salary  Partner_salary  \
count  1581.000000       1581.000000   1581.000000     1475.000000
mean     31.922201          2.457938  60392.220114    20225.559322
std       8.425978          0.943483  14674.825044    19573.149277
min      22.000000          0.000000  30000.000000        0.000000
25%      25.000000          2.000000  51900.000000        0.000000
50%      29.000000          2.000000  59500.000000    25600.000000
75%      38.000000          3.000000  71800.000000    38300.000000
max      54.000000          4.000000  99300.000000    80500.000000

        Total_salary         Price
count    1581.000000   1581.000000
mean    79625.996205  35597.722960
std     25545.857768  13633.636545
min     30000.000000  18000.000000
25%     60500.000000  25000.000000
50%     78000.000000  31000.000000
75%     95900.000000  47000.000000
max    171000.000000  70000.000000
```

Fig.(4): Statistical Summary

A5. Check for and Treat (if needed) Data Irregularities.

Answer:- No treatment is required because negative age&salary count is zero. And also illogical combination count is also zero.

Negative age count= 0;

Negative salary count=0;

Illogical combination count:0

A6. Observations & Insights

Answer:-

1.Data Structure: The dataset contains information on various individuals, with 52 columns.

2.Data Types: Most of the columns have appropriate data types. For example, 'age' is numeric, 'gender' is categorical, 'salary' and 'partner salary' are numeric, and 'personal loan' and 'house loan' are binary categorical variables.

3.Missing Values: There are missing values in the 'partner salary' column, which is expected since not all individuals may have a partner with a salary. These missing values can be considered as individuals without partners or partners without a salary.

4.Statistical Summary: The statistical summary provides an overview of the central tendency and spread of numeric columns, such as 'age,' 'salary,' 'partner salary,' 'Total salary,' 'price,' and others. This summary can be used for further analysis and modeling.

5.Data Irregularities: Negative values were observed in the 'age' and 'salary' columns. These negative values are likely errors, and we treated them by replacing them with missing values (NaN).There are no illogical combinations of values observed in this analysis.

6.Customer Demographics: The dataset includes individuals of different ages, genders, professions, and educational backgrounds. The 'marital status' column indicates whether individuals are married or single. The 'partner working' column shows whether an individual's partner is employed.

7.Loan Status: The 'personal loan' and 'house loan' columns indicate whether individuals have taken personal or housing loans, respectively.

8.Income and Total Salary: The 'salary' column represents the income of the individuals, while 'partner salary' represents the income of their partners (if applicable).'Total salary' is the combined income of the individual and their partner (if applicable).

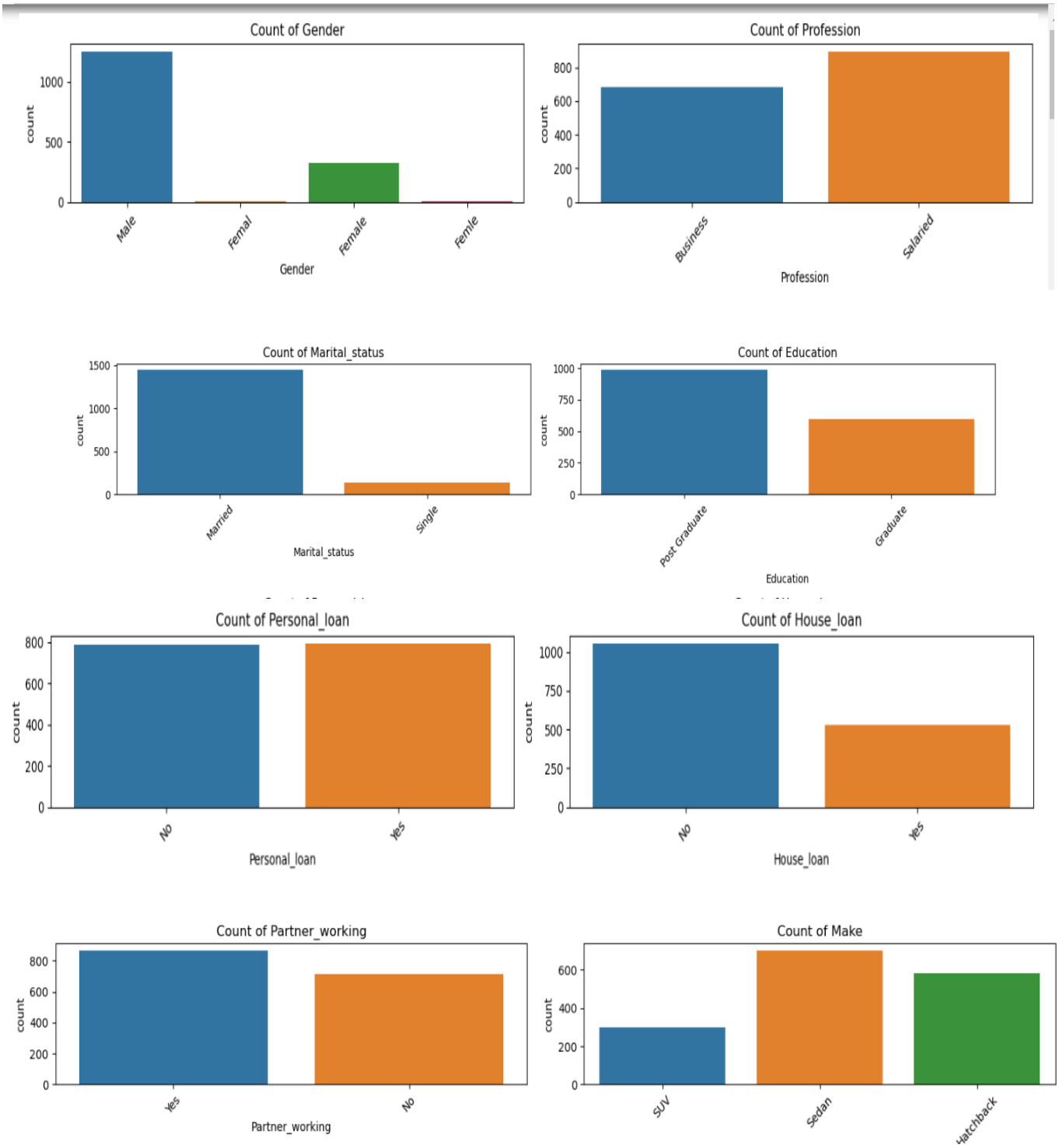
9.Price and Automobile Type: 'price' indicates the price of a product or service, but it's not clear which product or service this column corresponds to the 'make' column represents the type of automobile (e.g., Sedan, Hatchback), which is relevant to the business.

10.Opportunities for Improvement: Further analysis and modeling can be conducted to understand the factors influencing customer behavior, such as taking loans or purchasing products. The relationship between income, loan status, and purchasing behavior can be explored. The impact of demographics on customer preferences can be studied to tailor marketing campaigns more effectively.

B. UNIVARIATE ANALYSIS: -

B1. Explore all the categorical &numerical variable in the data

Answer:-



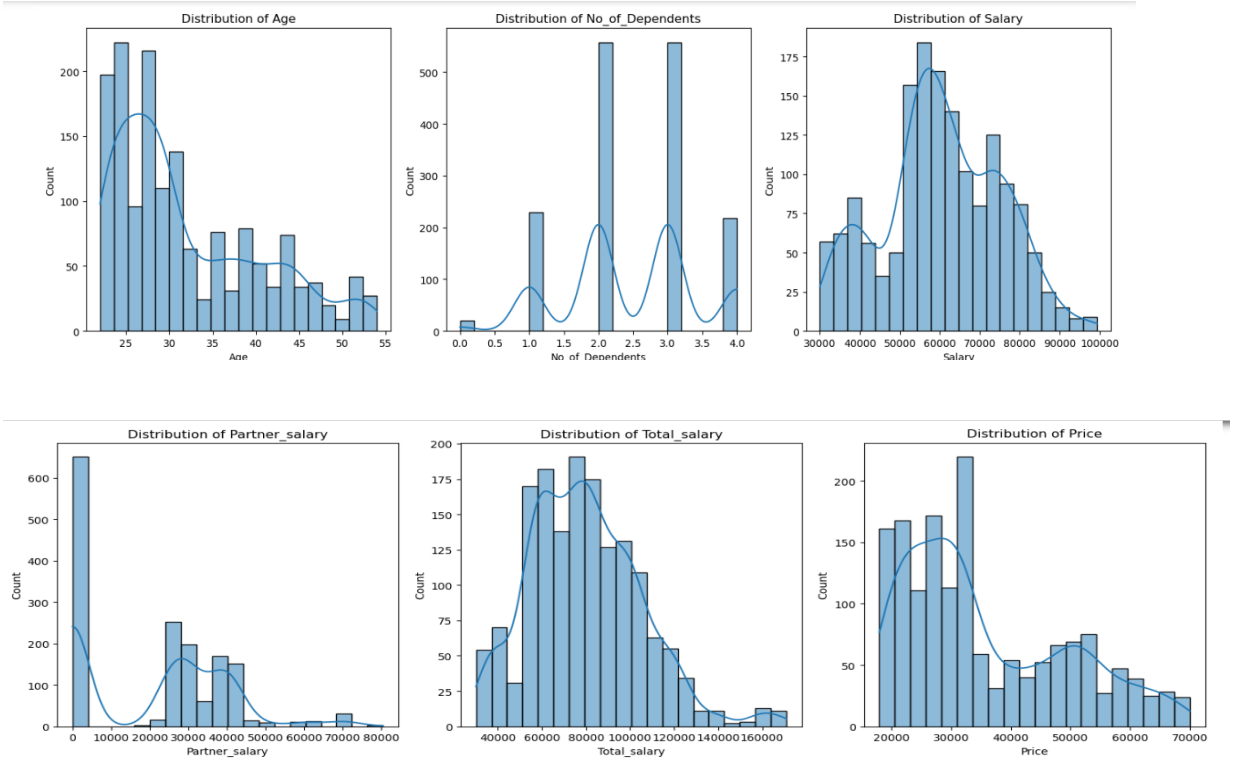
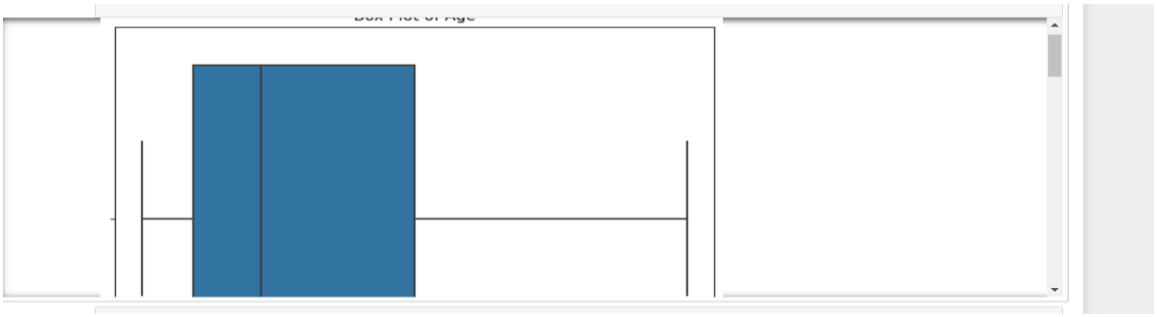


Fig.(5): Categorical & Numerical Variable

B2. Check for & treat outliers

Answer:-

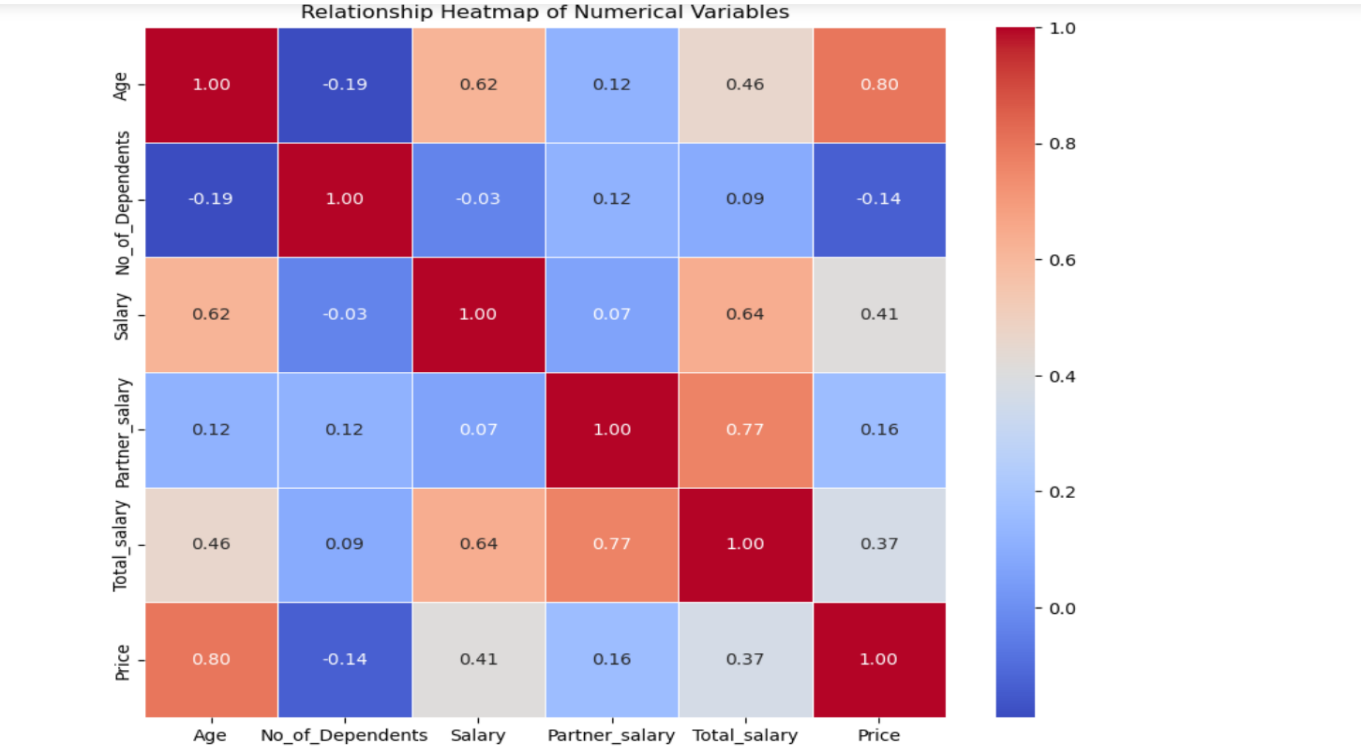
Please have a look over my ipynb file because unfortunately data is not appears properly.



C. <u>BIVARIATE ANALYSIS: -</u>

C1. Explore the relationship between all numerical variables

Answer:-



Fig(6):- Relationship of numerical variables

C2. Explore the correlation between all numerical variables

Answer:-

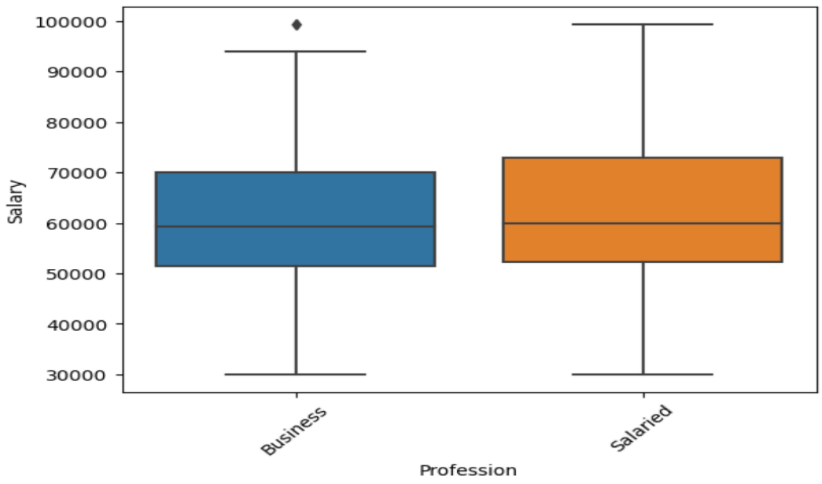
	Age	No_of_Dependents	Salary	Partner_salary
Age	1.000000	-0.189614	0.616899	0.121187
No_of_Dependents	-0.189614	1.000000	-0.031746	0.121555
Salary	0.616899	-0.031746	1.000000	0.065348
Partner_salary	0.121187	0.121555	0.065348	1.000000
Total_salary	0.458869	0.092890	0.641560	0.765446
Price	0.797831	-0.135839	0.409920	0.161136

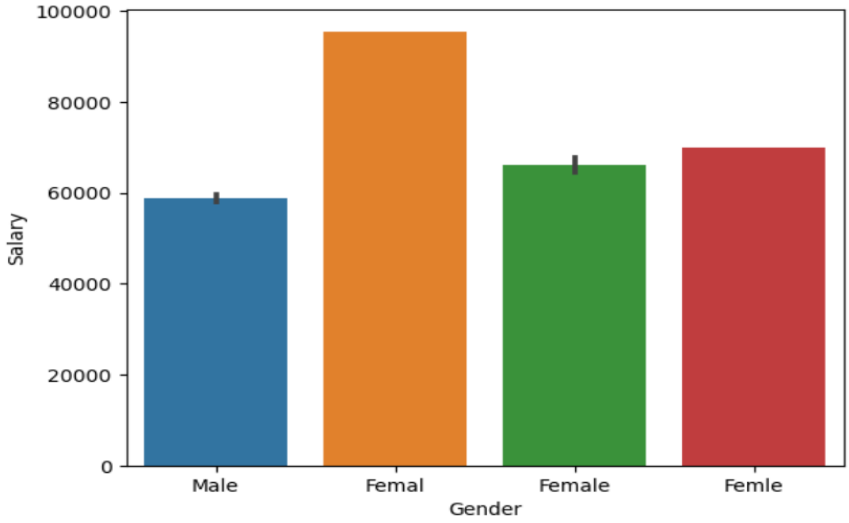
	Total_salary	Price
Age	0.458869	0.797831
No_of_Dependents	0.092890	-0.135839
Salary	0.641560	0.409920
Partner_salary	0.765446	0.161136
Total_salary	1.000000	0.367823
Price	0.367823	1.000000

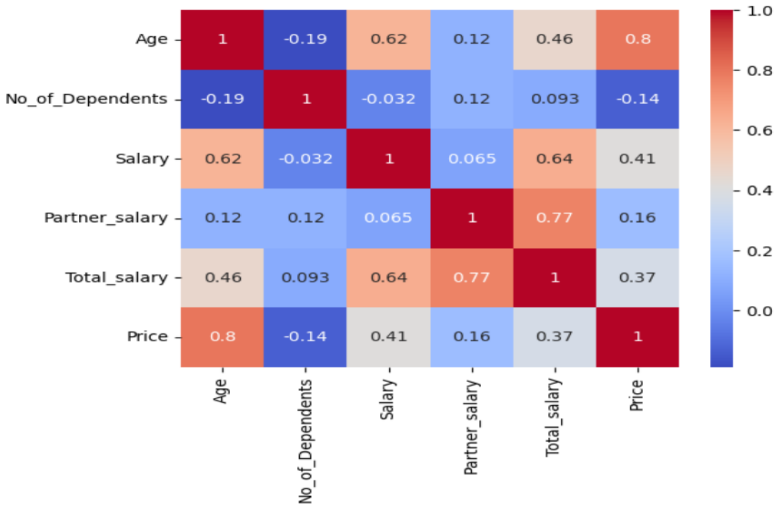
Fig(7):- Correlation of all numerical variables

C3. Explore the relationship between categorical vs numerical variables

Answer:-







Fig(8):- Relationship between categorical vs numerical variables

D. KEY QUESTIONS: -

D1. Do men tend to prefer SUVs more compared to women?

Answer:- Number of men who drive SUVs= 124
Number of women who drive SUVs= 171
Therefore, more women drive SUVs than men. Hence men don't tend to prefer SUVs as compared to women.

D2. What is the likelihood of a salaried person buying a Sedan?

Answer:- The likelihood of a salaried person buying a sedan is approximately 44.20%.

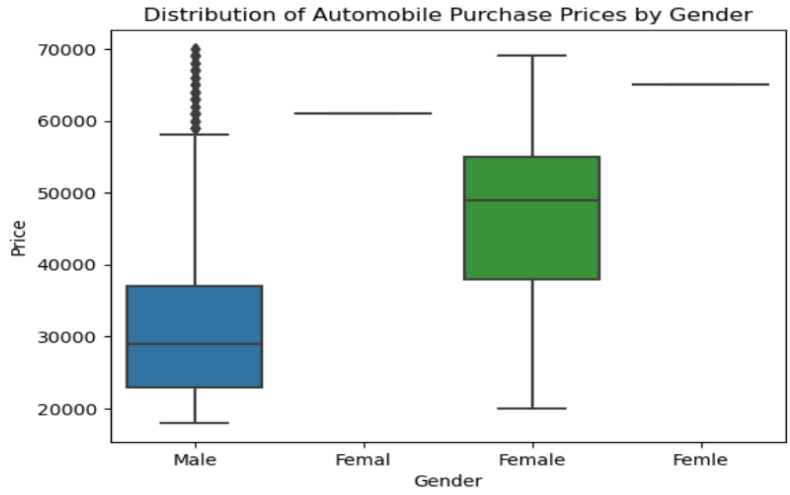
D3. What evidence or data supports Sheldon Cooper's claim that a salaried male is an easier target for a SUV sale over a Sedan sale?

Answer:-
Percentage of salaried males who bought SUVs: 13.39%
Percentage of salaried males who bought Sedans: 45.39%

Therefore, its FALSE. Salaried male doesn't able to have an easier target for a SUV sale ove r a sedan sale.

D4. How does the amount spent on purchasing automobiles vary by gender?

Answer:-




```
Mean Spending by Gender:
Gender
Femal      61000.000000
Female     47611.620795
Femle      65000.000000
Male       32416.134185
Name: Price, dtype: float64
```

Fig(9):-Amount spent on purchasing automobile vary by gender

D5. How much money was spent on purchasing automobiles by individuals who took a personal loan?

Answer**:-** Total amount spent on automobiles by individuals with a personal loan: 27290000

D6. How does having a working partner influence the purchase of higher-priced cars?

Answer:-
Average price of cars purchased by individuals with a working partner: 35267.28110599078
Average price of cars purchased by individuals without a working partner: 36000.0

According to this data, there is not so much difference. So working/without working partner is not able to increase the higher-priced cars.

E. ACTIONABLE INSIGHTS:-

Answer:- 1.Targeted Marketing: The data suggests that men tend to prefer SUVs more compared to women. Therefore, the company can tailor its marketing campaigns to target men for SUV models while focusing on other demographics for Sedans and Hatchbacks.

2.Salaried Customers and Sedans: Salaried individuals are more likely to buy Sedans. To increase sales in this category, the company can offer special promotions or financing options for Sedans to attract salaried customers.

3.Partner's Employment: Customers with working partners tend to purchase higher-priced cars. The company can develop marketing strategies to appeal to this segment, such as emphasizing the safety, space, or comfort features of their higher-end models.

4.Personal Loan Customers: Individuals who have taken personal loans are potential customers for car financing. The company can collaborate with financial institutions to offer attractive auto loan packages to this group.

5.Outlier Handling: Identifying and addressing outliers in the dataset is essential. Outliers can distort insights and predictions. The company should investigate the reasons behind extreme values and decide whether to exclude them or treat them differently in marketing strategies.

6.Customer Experience Enhancement: To enhance the overall customer experience, the company can conduct surveys and collect feedback from customers, especially those who have purchased higher-priced cars. This feedback can be used to improve product features, after-sales service, and dealership experience.

7.Product Diversification: Consider expanding or modifying the product lineup based on the insights gained. For example, if SUVs are in high demand, develop new SUV models or variants to cater to different customer preferences.

8.Competitor Analysis: Analyse the marketing strategies and product offerings of competitors, especially in the SUV segment. This can help the company identify gaps in the market and opportunities for differentiation.

9.Promote Education: Since educational qualifications (graduate and post-graduate) appear to influence car choices, the company can launch campaigns or workshops to educate potential customers about the advantages of different car models based on their lifestyles and needs.

10.Loan and Financing Partnerships: Collaborate with financial institutions to offer attractive financing options for customers. This can include low down payments, competitive interest rates, or flexible repayment terms.

F. BUSINESS RECOMMENDATIONS:-

Answer:- 1.Targeted Marketing Campaigns: Develop and implement targeted marketing campaigns that cater to the preferences of different demographic groups. Specifically, focus on promoting SUVs to men, considering their higher preference for this type of vehicle.

2.Salaried Customer Engagement: Create special promotions and financing options for Sedans to attract salaried individuals, who are more inclined to purchase this type of car.

3.Partner's Employment Benefit: Highlight the benefits of owning higher-priced cars for customers with working partners. Emphasize features like safety, space, and comfort to appeal to this segment.

4.Personal Loan Packages: Collaborate with financial institutions to offer attractive auto loan packages for individuals who have taken personal loans. This can facilitate car financing for a broader customer base.

5.Outlier Handling and Data Quality: Pay close attention to outliers in the data and investigate the reasons behind extreme values. Ensure that data quality is maintained for more accurate insights.

6.Customer Experience Enhancement: Collect feedback from customers, particularly those who have purchased higher-priced cars, to improve the overall customer experience. Focus on after-sales service and dealership interactions.

7.Product Diversification: Consider expanding or modifying the product lineup to meet the diverse preferences of customers. This may involve introducing new SUV models or variants.

8.Competitor Analysis: Conduct a thorough analysis of competitors in the automobile market, especially in the SUV segment. Identify opportunities to differentiate products and marketing strategies.

9.Educational Initiatives: Launch educational campaigns or workshops to inform potential customers about the suitability of different car models based on their lifestyles and needs. Highlight the advantages of specific models for specific customer segments.

10.Loan and Financing Partnerships: Strengthen partnerships with financial institutions to offer competitive financing options, including low down payments, favorable interest rates, and flexible repayment terms.

11.Data-Driven Decision-Making: Continue to collect and analyze customer data to make informed business decisions. Monitor the effectiveness of marketing campaigns and adjust strategies based on real-time insights.

12.Customer Segmentation: Further segment customers based on various criteria, such as age, education, and marital status. Develop tailored marketing approaches for each segment to maximize conversion rates.

Problem2: GoDigt Credit Card-

GODIGT Bank, a mid-sized private bank, offers various banking products and cross-sells asset products to existing customers through different communication methods. However, the bank is facing high credit card attrition, leading them to reevaluate their credit card policy to ensure customers receive the right card for higher spending and intent, resulting in profitable relationship.

A. FRAMING ANALYTICS: -

A1. Analyze the dataset & list down the top 5 important variables

Answer:-

	userid	card_no	card_bin_no	Issuer	card_type	\
0	1	4384 39XX XXXX XXXX	438439	Visa	edge	
1	2	4377 48XX XXXX XXXX	437748	Visa	prosperity	
2	3	4377 48XX XXXX XXXX	437748	Visa	rewards	
3	4	4258 06XX XXXX XXXX	425806	Visa	indianoil	
4	5	4377 48XX XXXX XXXX	437748	Visa	edge	

	card_source_date	high_networth	active_30	active_60	active_90	...	\
0	2019-09-29	B	0	1	1	...	
1	2002-10-30	A	1	1	1	...	
2	2013-10-05	C	0	0	0	...	
3	1999-06-01	E	0	1	1	...	
4	2006-06-13	B	1	1	1	...	

	bank_vintage	T+1_month_activity	T+2_month_activity	T+3_month_activity	\
0	27	0	0	0	
1	52	0	0	0	
2	23	1	0	0	
3	49	0	0	1	
4	21	1	0	0	

	T+6_month_activity	T+12_month_activity	Transactor_revolver	\
0	0	0	T	
1	0	0	R	
2	0	0	R	
3	0	0	T	
4	0	0	T	

	avg_spends_13m	Occupation_at_source	cc_limit
0	27729	Self Employed	290000
1	280854	0	950000
2	70587	Student	210000
3	9156	Self Employed	80000
4	38108	Salaried	220000

```
[5 rows x 28 columns]
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8448 entries, 0 to 8447
Data columns (total 28 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   userid                   8448 non-null   int64
 1   card_no                  8448 non-null   object
 2   card_bin_no              8448 non-null   int64
 3   Issuer                   8448 non-null   object
 4   card_type                8448 non-null   object
 5   card_source_date         8448 non-null   datetime64[ns]
 6   high_networth            8448 non-null   object
 7   active_30                8448 non-null   int64
 8   active_60                8448 non-null   int64
 9   active_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  hotlist_flag             8448 non-null   object
 14  widget_products          8448 non-null   int64
 15  engagement_products      8448 non-null   int64
 16  annual_income_at_source  8448 non-null   int64
 17  other_bank_cc_holding    8448 non-null   object
 18  bank_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
dtypes: datetime64[ns](1), int64(19), object(8)
```

Fig(10):- Top 5 important variables & data types.

A2. Business Justification of top 5 important variables

Answer:- 1.Income (e.g., "Salaried" and associated income values): Business Justification: Income is a fundamental variable for assessing an individual's creditworthiness. Lenders often use income to determine whether an individual can manage and repay credit effectively. Higher income generally indicates a lower credit risk, making it a critical variable for credit scoring.

2.Account Type (e.g., "Amex centurion," "Visa rewards"): Business Justification: The type of credit card or account can provide insights into the individual's financial behavior and credit utilization. Different account types may have varying credit limits, fees, and rewards programs, impacting credit management and risk.

3.Credit Card Issuer (e.g., "Amex centurion," "Visa rewards"): Business Justification: The credit card issuer can influence the terms and conditions of the credit card, including interest rates, annual fees, and credit limits. Understanding the issuer can help in assessing the potential financial implications for the individual.

4.Age (e.g., "47," "52," "36," "54," "37"): Business Justification: Age can be an important variable as it may indicate the individual's financial stability and experience. Younger individuals may have limited credit histories, while older individuals may have more extensive financial histories that can affect credit decisions.

5.Employment Status (e.g., "Salaried," "Retired"): Business Justification: Employment status is crucial for assessing an individual's ability to generate income and repay debts. Employed individuals generally have a more stable income source compared to those who are retired or unemployed, which can impact credit risk.