SMDM PROECT REPORT

PGPDSBA

11/08/2023

FASNA.PP

Table contents:

Table of Figures:

TABLES:

Problem 1

Context Analysts are required to explore data and reflect on the insights. Clear writing skill is an integral part of a good report. Note that the explanations must be such that readers with minimum knowledge of analytics is able to grasp the insight.

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 an analytics professional to improve the existing campaign.

Problem 1-1 - Data Overview

1-1-a) Import the libraries

In the problem we want work with data set so importing pandas library, Numpy library want import for performing mathematical operations on arrays. Visualizing the problem we want to import matplotlib.pyplot ,import seaborn library for making statistical graphics.

1-1-b) Load the data

We want read the file ,using python code (its csv file, read the csv file using python code)

1-1-c) Check the structure of the data

- Check the shape of data set(number of rows and number of columns of the data set). no.of rows: 1581
- no.of rows: 1581 ,no.of columns: 14
- then just check the our data set ,just print the data set with first 5 rows in the table using head function.

1-1-d)Check the types of the data

- check the data types of data using info() function
- data set have 8 object data types,5 int type,1 float type.
- missing values on Gender and Partner_salary columns.

1-1-e) Check for and treat (if needed) missing values

check missing values

- check null values in the data frame using isnull() function then take sum because number of zeros in each column check.
- in the Gender column have 53 missing values, Gender is object data type.
- partner_salary column have 106 missing values,partner_salary is float data type.

* treat missing values

- Gender is object data type ,this 53 null values so we can treat null values by mode() of the gender column.
- treat null values of gender column by mode of column, after treat check the null values in gender column
- There is no null values in gender column
- Then check the null values in entire data set.
- partner_salary is float data type ,have 106 null values ,treat null values by median of partner_salary value.
- plot boxplot partner salary for checking median
- replace null values by median of partner_salary, after check the null values in the entire data frame using isnull() function
- after there is no null values in the data set

1-1-f) Check the statistical summary

create data frame with integer and float data types of df data set
create data frame with object data types of df data set
describe the object data frame

	Gender	Profession	Marital_status	Education	Personal_loan	House_loan	Partner_working	Make
count	1581	1581	1581	1581	1581	1581	1581	1581
unique	4	2	2	2	2	2	2	3
top	Male	Salaried	Married	Post Graduate	Yes	No	Yes	Sedan

freq	1252	896	1443	985	792	1054	868	702

Table 1:Described table of categorical variables

Like describe the numerical data frame
We can see standard deviation of Price is 13633.636545 whereas the mean is 35597.722960. However, the max price is 70000.000000 and min is 18000.000000.

1-1-g)Check for and treat (if needed) data irregularities

- check the duplicated values using duplicated() function

- there is no duplicated values in the data set
- check data irregularities, so print first 5 rows.

****spelling mistakes on Gender column**

Gender column have spelling mistake in "Female" letter
check unique values in gender column

```
array(['Male', 'Femal', 'Female', 'Femle'])
```

- the Gender column have "Femal","Femle" are wrong data ,we want to replace by "Female"

"Femal","Femle" replace by "Female" using replace function

****change the integer data type of salary, total_salary, price to float data type**

change the integer data type of salary, total_salary, price to float data type using astype function.

1-1-g)Observations and Insights

describe data with all data type

Insights

- Dataset has 14 columns and 1581 rows
- In the dataset have 4 float variables, 2 integers variables, 8 object variables
- The dataset provided is for 'SUV', 'Sedan', 'Hatchback' cars
- partner salary depends on the partner working whether or not
- Total salary depends the salary and partner salary
- No duplicate records.

- No null values.
- most make car is sedan ,702 sedan total make.
- most customer are Male ,Salaried ,Married, Post Graduated, have 4 dependents , their have personal loan ,have no house loan, their partner are working.
- We can see standard deviation of Price is 13633.636545 whereas the mean is 35597.722960. However, the max price is 70000.000000 and min is 18000.000000.

Problem 1.2 - Univariate Analysis¶

Explore all the variables (categorical and numerical) in the data - Check for and treat (if needed) outliers - Observations and Insights

1.2.a)-Explore all the variables (categorical and numerical) in the data

Personal_loan,House_loan,Partner Histogram for every categorical variable s,Plot histogram of Gender,Profession, Marital_status, Education, partner working,make.

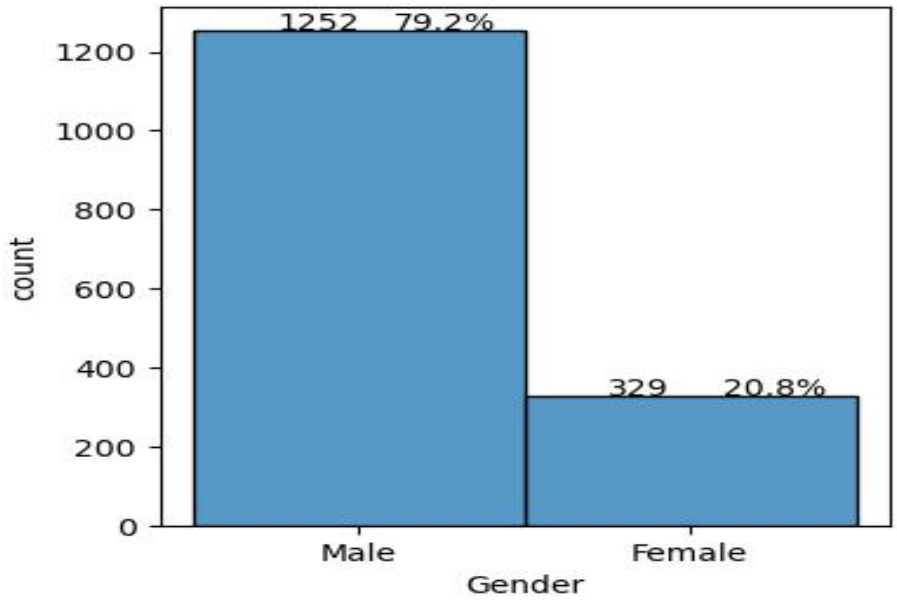


figure 1: Barplot of Gender column

like all categorical variables plot.
*most customers are Male, 20.8% customer are Female.
*56.7% of customers are salaried,43.3% customers are in Business
*91.3% of customers are married

*62.3% of customers are post Graduated
*most similar customers have or not personal loans.
* 33.3% customers have House loan.
* 45.1% customer have partner not working
* most make car is Sedan , 36.8% Hatchback is make ,18.8% SUV make

Univariate analysis for numerical variables

Plot histogram or boxplot of numerical variables ,for example Age column

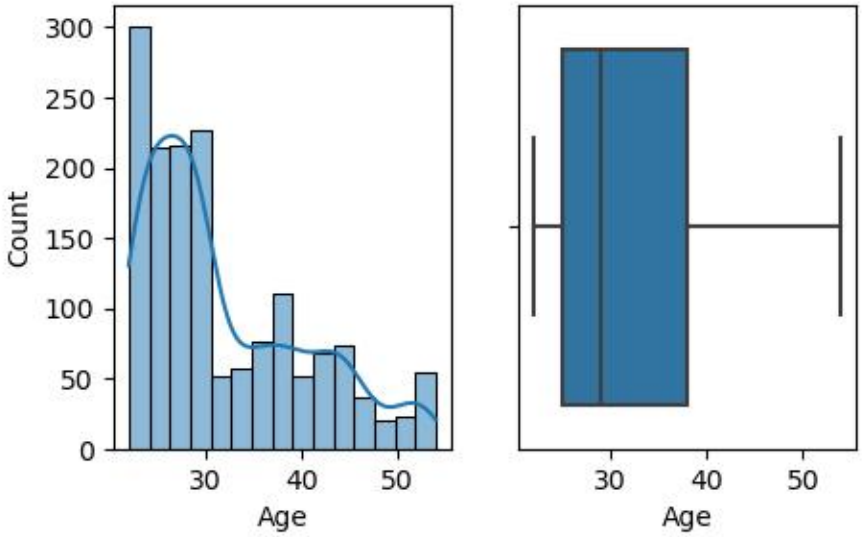


figure 2: Histogram&Boxplot of Age column

- most customers age in between 20 to 30
- check the skewness ,kurtosis of Age column
- skewness of age is 0.8930870865867485
- kurtosis of age is -0.24405428996418044
- age variable is right skewed so we want make it symmetrical because we want get normal distribution, using root of 10 Transformation. After get skewness 0.5911857058564791 , kurtosis -0.7600344747237493
- skewness is decreased .like plot the boxplot or histogram all numerical variables then check skewness ,most symmetrical skewness is in between -1/2 and 1/2.its not symmetrical then treat by root 10 method.

- most customers age in between 20 to 30
- median line on first quartile ,maximum of no of dependents of customers is 4,have a outlier in lower value.

- minimum of partner salary is 0.00 because 45.1% of customer have partner are not working.
- most customers have salary is 60000
- Total salary have many outliers in upper values most customers have total salary in between 60000 to 90000
- maximum of price for service is 70000,

1.2.b)- Check for and treat (if needed) outliers

Plot box plot of all numerical variables for checking outliers

No.of dependents variable have one outlier in lower value, total salary variable have many outliers in higher values, we want to treat outliers for accurate values,remove outliers using IQR method for example

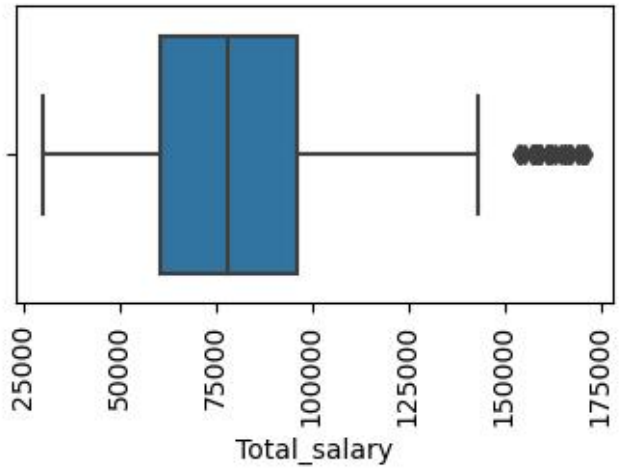


figure 3. Boxplot for Total_salary

remove outliers using IQR method(Check the lower range and upper range for the variable, total salary values less than lower then it will to lower else past value, total salary values greater than high range it will to high range else past value).like this do the No.of dependents column.

1.2.c)-Observations and Insights

Insights

- Most individuals in the dataset are either 22 or 23 years old.
- The majority of individuals in the dataset are males.

- The dataset includes individuals with both salaried and business employment types.
- The majority of individuals in the dataset are married. Some individuals are also single.
- Most individuals in the dataset are post-graduates, with a smaller number having graduate degrees.
- Maximum of no.of Dependents of customer is 4.
- Most similar customer have ,or not personal loan. maximum price is 70000
- Salary and partner salary increases the total salary,most customer have salary in between 50000 to 80000--mean of total salary of customer is 78000 .
- Most 44.4% make car is Sedan,36.8% of make car is Hatchback,18.8 % SUV

Problem 1-3 - Bivariate Analysis

1-3-a)- Explore the relationship between all numerical variables

we can observe every relationship between all numerical variables
scatterplot(pair plot)

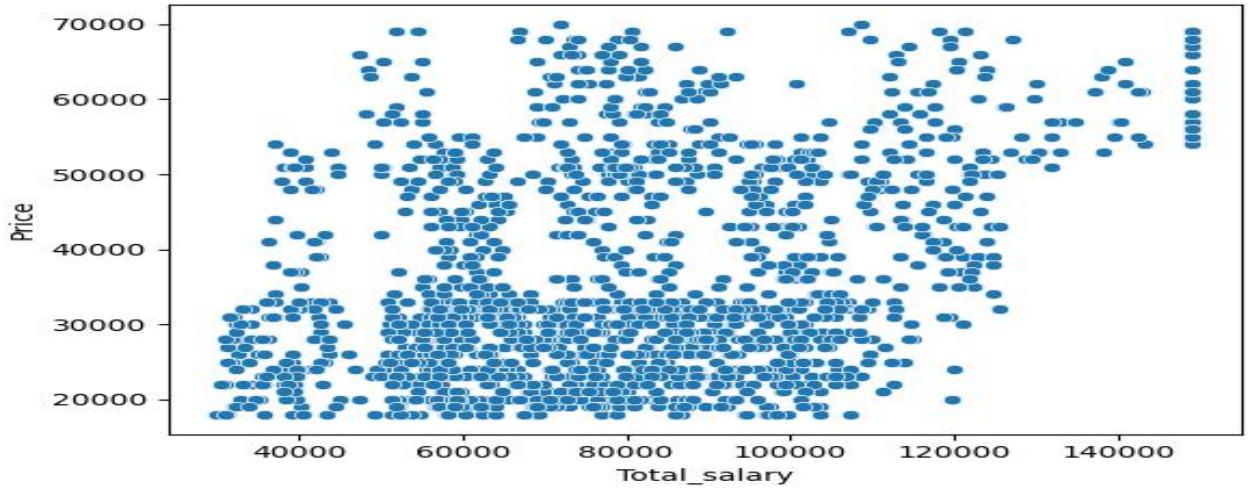


figure 4. Scatter plot for Total_salary and Price

From above plots we see that as the Total salary increases, there are some points where price also increasing proportionally.These are customers who have high total salary they are buying high price car.
Like plot every numerical variables scatter plot each other.

- all are variables are positively related except age and no.of dependents are negatively related to price and no.of dependents are negatively related

1-3-b)Explore the correlation between all numerical variables

check correlation of data set using corr() function
plot heatmap correlation b/w variables

figure 5. Heatmap for correlation between variables

Insights

- We observe price is highly correlated with Age each other. Especially the price have a correlation is 0.8
- There is a negative correlation between the Age and No.of.dependents (around -0.21) this means high age customer have low no.of dependents.
- price and no od dependents are negatively correlated

1-3-c) Explore the relationship between categorical vs numerical variables

Plot the box plot with numerical variables and categorical variables .for example Plot the boxplot of Age and Gender,like all plot.

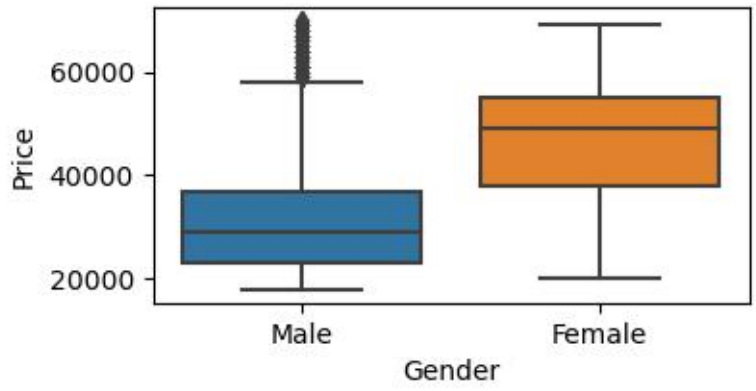


figure 6. Boxplot for Gender and Price

`Insights`

- We see that median value of the price for male is less compared to that of female. male have many outliers in high price.
- There is no difference in partner salary in between male and female. median value of total salary of female is higher than male.
- We see that median value of the price for business customer is less compared to that of salaried customers. business have outliers in high price.
- There is no difference in total salary in between business and salaried customers.
- We see that median value of the price for sedan ,Hatchback cars is less compared to that of SUV.
- There is no difference in median Partner salary amount between these 3 cars
- we can see that in make list of company the Highest price for SUV car,then Sedan, then Hatchback.

Problem 1-4-1) - Key Questions

1-4-1) Do men tend to prefer SUVs more compared to women?

create data frame with make car is equal to SUV ,then plot count plot with Make and gender

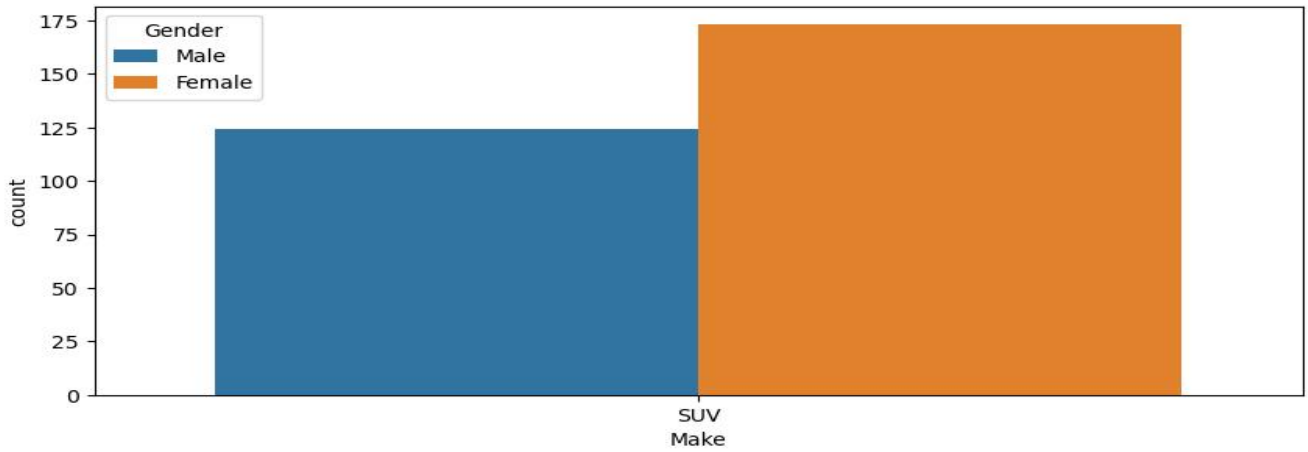


figure 7. Countplot for Make with Gender

we can see don't men to prefer SUVs more than women. Female customers are more prefer SUV.

1-4-2). What is the likelihood of a salaried person buying a Sedan?

- Likelihood = (Number of salaried individuals who bought Sedans) / (Total number of salaried individuals)
- *create data frame with profession is equal to salaried customer*
- create table with cross the profession ,make column

Make, Profession	Hatchback	SUV	Sedan	All
Salaried	0.325893	0.232143	0.441964	1.0
All	0.325893	0.232143	0.441964	1.0

Table 2: Cross Table profession ,make column of salaried customers

So, the likelihood of a salaried person buying a Sedan is approximately 0.4419, or about 44.2%.higher than other two cars.

1-4-3). 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?

Filter the Data Frame for salaried males, Count the number of SUVs and Sedans among salaried males.
Sedan 279
Hatchback 276
SUV 85
we can see that salaried male is an easier target for a Sedan sale over a SUV sale,so Sheldon Cooper's claim is wrong.

1-4-4). How does the amount spent on purchasing automobiles vary by gender?

plot the bar plot of Gender,Price with Make.

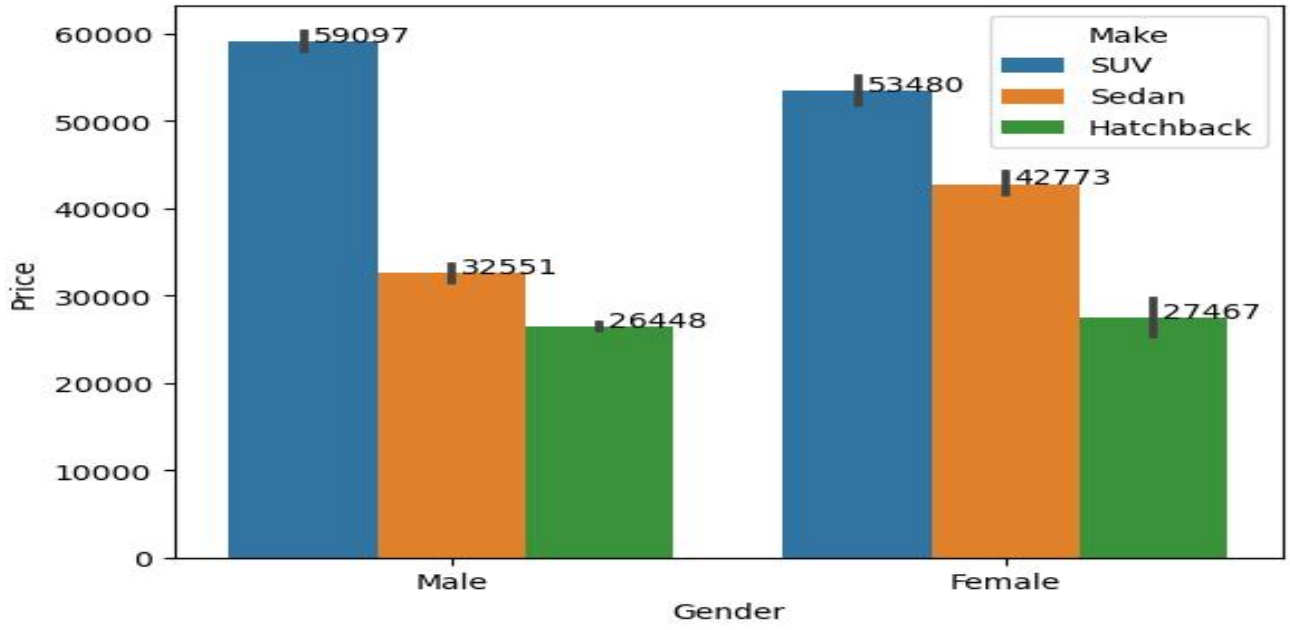


figure 8. Bar plot of Gender,Price with Make.

- we can see that the highest amount was spent on purchasing an SUV, then a Sedan, then a Hatchback
- *Males spend more amount spent on purchasing an SUV than Females.
- *Females have a higher amount on purchasing Sedans than Males.

- *Females spend a little higher amount spent on purchasing Hatchback than males.

1-4-5). How much money was spent on purchasing automobiles by individuals who took a personal loan?

- creating data frame with customers have personal loan
- check the sum of price in customers who have a personal loan
- 27290000.0
- people who took Personal loans spent a total of 27290000.0 amount spent on purchasing automobiles

1-4- 6). How does having a working partner influence the purchase of higher-priced cars?

create a data frame this have partners are working
plot bar plot of working partners and price with make

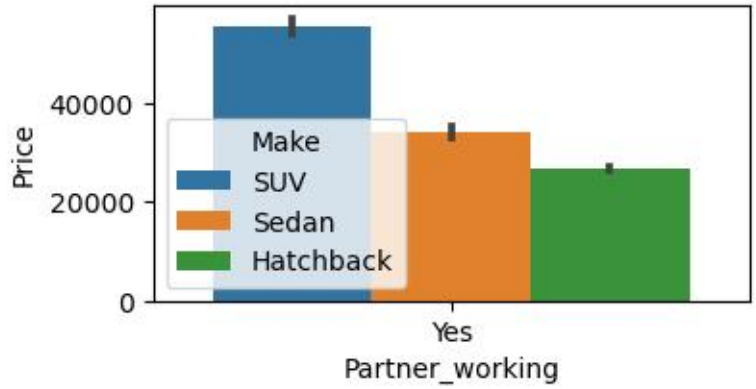


figure 9. Bar plot of working partners and price with make

more working partners influence purchasing highest cost for SUV. second Sedan then Hatchback. Individuals with working partners tend to purchase higher-priced cars.
OR
plot bar plot of partner working and price(working partners and non working partners influence purchasing high priced cars approximately both are same. little higher non working partners than working partners.

Actionable Insights:

Salaried Individuals Prefer Sedan:The data indicates that a significant number of salaried individuals prefer Sedan cars. This insight could be used for marketing strategies, such as offering promotions or discounts on Sedan models to attract salaried customers.

Marital Status Affects Car Choice:Marital status seems to influence the choice of car. Married individuals are more likely to purchase sedans, while single individuals often choose hatchbacks. Tailoring advertising and promotions to specific marital statuses can be effective.

*Post Graduates Prefer Sedans:*Post-graduate individuals tend to opt for sedans, which are often considered more upscale. Marketing campaigns could target this educated demographic with features and benefits that align with their preferences.

Working Partners Impact Car Prices:Individuals with working partners tend to purchase higher-priced cars. This suggests that dual-income households may have more purchasing power and are willing to invest in more expensive vehicles. Dealerships can cater to this segment by offering premium car options and financing solutions.

Graduate and Post Graduate Customers:Customers with higher education levels (graduate and post-graduate) are more likely to purchase sedans. Dealerships can leverage this information by showcasing the advanced features and luxury aspects of sedans to attract this educated customer base.

*Business Recommendations:

Segmented Marketing:

Tailor marketing campaigns based on customer demographics such as age, gender, marital status, and education level. Create targeted advertisements highlighting features that appeal to specific segments, like safety for families or luxury for well-educated individuals.

Offer Financing Options:

Recognize that customers with working partners may have higher purchasing power. Offer flexible financing options and packages for higher-priced cars to make them more accessible to this segment.

Promote Sedan to Salaried Individuals:

Develop promotions and incentives for salaried individuals to consider Sedan models. Emphasize affordability, fuel efficiency, and practicality in marketing campaigns.

Educate Customers:

For graduate and post-graduate customers, provide information about the advanced technology, safety features, and luxury aspects of sedans. Highlight the long-term benefits and value of these vehicles.

Market to Single Individuals:

Since single individuals tend to choose hatchbacks, create marketing messages that resonate with their lifestyle, emphasizing flexibility, style, and affordability.

Customer Loyalty Programs:

Implement customer loyalty programs to retain customers and encourage repeat purchases. Offer benefits such as discounts on maintenance services, accessories, or future car purchases.

Data-Driven Decision-Making:

Continuously analyze customer data to refine marketing strategies and product offerings. Stay updated with changing customer preferences and adjust business strategies accordingly.

automotive dealerships can better meet the needs and preferences of their target customer segments, ultimately driving sales and customer satisfaction

Problem 2

Context

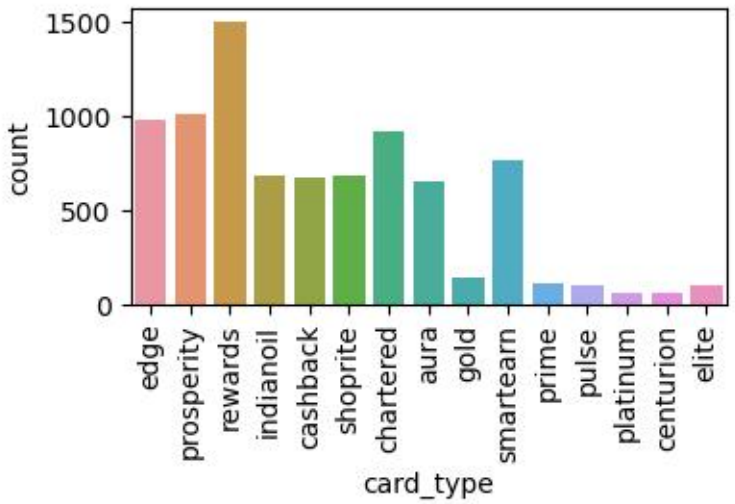
A bank generates revenue through interest, transaction fees, and financial advice, with interest charged on customer loans being a significant source of profits. GODIGT Bank, a mid-sized private bank, offers various banking products and cross-sells asset products to existing customers through different 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 relationships.

Problem 2 - Framing Analytics Problem

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

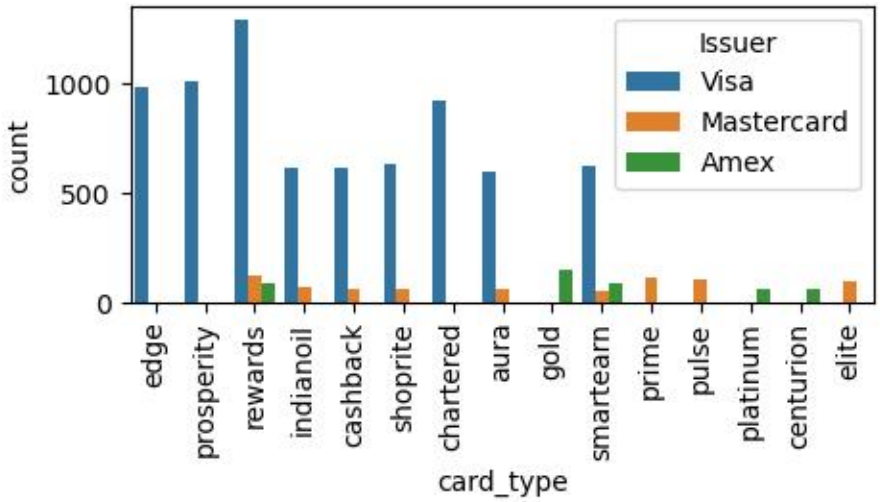
- Doing the data preprocessing
- Read the data set,
- check number of columns and number of rows- no.of columns: 8448 no.of rows: 28

- check the data types of columns
- missing values in the Transactor_revolver column, we want to treat the missing values. other data are good.
- treat missing values in the Transactor revolver column to mode data of the column
- we can seen that high values on credit limit,average spend in last 3 month, annual income we want treat them using log function.
- check duplicates-there is no duplicates.
- univariate analysis-
- plot count plot of card type



- rewards credit card type is most using. platinum credit card is less using.
- Like plot the count plot for all variables.we get insights.
- Insights
- rewards credit card type is most using. platinum credit card is less using.
- most number of issuers are visa card.
- more customers are A (high worth)type . other customers are approximately the same
- go to 90 to 30 we see that account activity is decreasing, inactivity is increasing. same as 90 to 60. *This means the last 30 days' account activity is much less than inactivity. *The credit card activity same as the account activity.
- most credit card have no problems.
- most customers have 7 Number of convenience products customer holds. then 4,5 number of convenience products.

- more customer have 4 Number of investment/loan products the customer holds . more customer have Number of investment/loan products the customer holds .
- most customers holding other bank credit card
- *very low customers uses credit card in T+1,T+2,T+3,T+6,T+12 month(future) . future activity is decreasing T+1 to T+2,month activity increasing T+2 to T+3, then decreasing T+3 t0 T+6,then decreasing T+6 to T+12. *we can see that month activity is much less than inactivity
- more customer customers who pays off their balances in full every month (T) less customers carry balances over from one month to the next(R)
- more customer are salaried ,low customer have no occupation
- we can see most customers are having low credit limits.so credit will lead to card attrition..
- *Bivariate analysis*



- most number of card rewards mostly produced by visa companies. most credit cards are produced by visa companies. all 3 companies are producing rewards, and smart earn credit cards. The lowest number of cards is platinum is produced by Amex company.
- Like plot count plot ,boxplot ,scatter plot analysis of variables each other and get the insights.
- *most number of card rewards mostly produced by visa companies. most credit cards are produced by visa companies. all 3 companies are producing rewards, and smart earn credit cards. The lowest number of cards is platinum is produced by Amex company

- *rewards card is most customers use, the credit limit of rewards card is very low. very less customers using platinum cards have higher credit limits than rewards. high credit limits for the edge, prosperity, and chartered cards are approximately same.
- *rewards card type is more active and inactive in the last 90 days. activity and inactive are directly proportional centurion card is less active and inactive
- *A,C,D using levels are approximately same. most numbers of reward cards are mostly using C,D,E (,middle, low worth) type customers. A type customers most using edge credit card, B type customers most using prosperity credit cards. less using platinum card only using A,B type customers.
- *rewards card using retried customers have high card problems.
- *the cc limit is very low for both the transactor and revolver. the median line of cc limit of the transactor and revolver are same.
- *high worth type customers have high credit limit. low worth customer type E have low credit limit
- *more active A Type customers in last 90 days
- *all number of convenience products hold customers are active and inactive. more customers are active who have 7 products. inactive all are approximately the same
- *slightly higher is 3,4 number of engagement products .customer who have loans ,other investments their have less credit limit it will leads to card attrition
- *average spends in last 3 months is directly proportional to credit limit
- *credit limit of activity in the last 90 days of customer has or not other bank credit card are same. the median line of inactive in last 90 days of customers having other bank credit cards is a little higher.
- *the credit limit is positively correlated with annual income. this means credit limit dependent the annual income
- *most number of customers are salaried the credit limit of salaried persons are low. the credit limit zero job (jobless)customers are very high compare to other customers
- *median line of annual income at source of jobless customer are high
- *salaried customers are more active and inactive in last 90 days. less active customers are house wife. less inactive is jobless customers

find correlation

annual income and credit limit are highly correlated.
annual income and active in last 90 days are very negatively correlated.

Insights

*__Occupation and Credit Limits__: High numbers of customers are salaried, and most have low credit limits. To prevent card attrition, businesses can explore options to increase credit limits for these customers responsibly.

*__Card Types and Rewards__: Visa is the dominant provider of credit cards and rewards cards. Consider diversifying card offerings and rewards programs to cater to a broader range of customer preferences.

*__Engagement Products__: Customers with loans and other investments tend to have lower credit limits. Provide additional incentives to these customers to prevent them from attrition,.

*__Usage Patterns Over Time__: Understanding that future credit card activity is decreasing over time (T+1 to T+2, T+3 to T+6, and T+6 to T+12) implies that customer engagement may drop as time progresses. Businesses should devise retention strategies for long-term customers.

*__Income and Credit Limit__: Recognize that credit limits are positively correlated with annual income. Businesses should regularly review and adjust credit limits based on income changes to maintain customer satisfaction.

*__Account Activity and Inactivity Trends__: most customers are active in last 90 days who have rewards visa type card type high worth , salaried customers. more inactive customers are rewards type . least active and least inactive on platinum type cards. Business should focus on strategies to re-engage customers and prevent attrition.

*__Balance Payoff__: Customers who pay off their balances in full every month are more common than those who carry balances. Consider providing incentives to the latter group to increase their engagement.

*__Credit Card Usage__: Most credit card users do not seem to have problems, which is a positive sign. However, it's essential to identify the factors contributing to customer satisfaction and replicate them across all customer segments. card problems on very less customers.

*__Number of Convenience Products and Investment/Loan Products__: Customers with 7 convenience products appear to be the most active. Businesses should consider cross-selling or upselling additional products to customers to increase their engagement and overall satisfaction.

*__Customer Segmentation__: Understand which customer segments prefer specific card types (e.g., A-type customers with Edge cards). Tailor marketing and retention efforts accordingly

*__Salaried Customers__: As most customers are salaried and have lower credit limits, businesses should monitor this segment closely and introduce measures to enhance their card usage and satisfaction.

*__Customer Activity by Occupation__: Tailor engagement strategies to the specific needs and preferences of customers based on their occupations.

*__Annual Income and Credit Limit__: Businesses can segment customers based on their income levels and offer credit limits that align with their financial capacity.

list down the top 5 important variables, along with the business justifications

1-card type: most customers are using visa type rewards cards.
Business Justification:
The type of credit card (e.g., Visa, Mastercard) is crucial for understanding customer preferences and behavior. It can help in tailoring marketing strategies, offering relevant rewards, and assessing the popularity of different card types among customers.
2-Occupation at Source: most customers are Salaried
Business Justification:
Knowing the occupation of cardholders provides insights into the income levels and employment status of customers. This information is valuable for targeting

specific customer segments, setting credit limits, and designing financial products that suit their needs.

3-High Net Worth: more type customers have high worth

Business Justification:

Identifying high-net-worth individuals (HNWIs) among cardholders is essential for offering personalized financial services, wealth management products, and investment opportunities. It can also influence credit decisions and risk assessment.

4-Active 90 Days: Salaried ,rewards visa card type ,high worth customers are more active in last 90 days

Business Justification:

The variable indicating whether a cardholder has been active in the last 90 days is critical for assessing customer engagement and retention. It helps in identifying inactive customers who may need re-engagement strategies or account management interventions.

5-Credit Limit: credit limit of edge visa type ,high worth ,salaried customers have high credit limit , most customers are Transactor.

Business Justification:

The credit limit assigned to cardholders directly impacts their spending capacity and financial behavior. Monitoring credit limits is crucial for risk management, ensuring customers' ability to repay, and preventing overextension of credit.

Justifications:

Segmentation and Targeted Marketing:

Justification: Segmentation allows the financial institution to understand the diverse needs and behaviors of cardholders. Targeted marketing ensures that promotional efforts are focused on individuals most likely to respond positively. This can lead to increased customer engagement and higher returns on marketing investments.

Product Customization:

Justification: Offering customized credit card products aligns with the principle of meeting customer needs effectively. It provides cardholders with options that match their preferences and lifestyle, increasing the likelihood of card usage and customer satisfaction.

Customer Engagement:

Justification: Engaging infrequent card users is essential for maximizing revenue and retaining customers. Encouraging them to use their cards more frequently

through personalized offers can lead to increased transaction volumes and revenue for the institution.

Retention Strategies:

Justification: Retaining high-value customers is more cost-effective than acquiring new ones. Implementing retention strategies ensures that the institution doesn't lose its most valuable clients, thereby maintaining a stable revenue stream.

Risk Assessment:

Justification: Assessing risk associated with different card types and usage patterns is critical for prudent risk management. It helps the institution allocate resources effectively, monitor potentially risky accounts, and make informed decisions regarding credit limits and lending practices.

Market Expansion:

Justification: Expanding the product portfolio to cater to specific segments can open up new revenue streams. Targeting underserved markets, such as students or self-employed individuals, can help the institution tap into previously untapped customer segments.

Cross-Selling:

Justification: Cross-selling other financial products to existing cardholders can boost the institution's profitability and deepen customer relationships. It's often easier to sell additional services to customers who are already engaged with the institution.

Customer Support:

Justification: Providing dedicated customer support ensures that customers have a positive experience with the institution. It can lead to higher customer satisfaction, increased loyalty, and positive word-of-mouth referrals.

Compliance and Security:

Justification: Maintaining compliance with data protection regulations is not only a legal requirement but also crucial for building trust with customers. Data breaches or security lapses can lead to reputational damage and legal consequences.

Feedback Mechanism:

Justification: Gathering feedback directly from cardholders helps the institution understand customer preferences and pain points. This information is invaluable for making data-driven decisions and continuous improvement. Competitive Analysis:

Justification: Monitoring competitors allows the institution to stay competitive and responsive to market changes. Understanding what others are offering can help identify gaps and opportunities in the market. Education and Awareness:
Justification: Educating cardholders about their credit cards can lead to responsible card usage and increased customer satisfaction. Informed customers are less likely to make mistakes or misunderstand the features of their cards.
Partnerships and Alliances:
Justification: Strategic partnerships can provide additional value to cardholders, making the institution's cards more attractive. Collaborations with businesses that align with customer interests can drive card usage and loyalty. Credit Limit Adjustments:
Justification: Periodically reviewing and adjusting credit limits based on customer behavior and financial status helps mitigate credit risk. It prevents overextension of credit and reduces the likelihood of defaults.