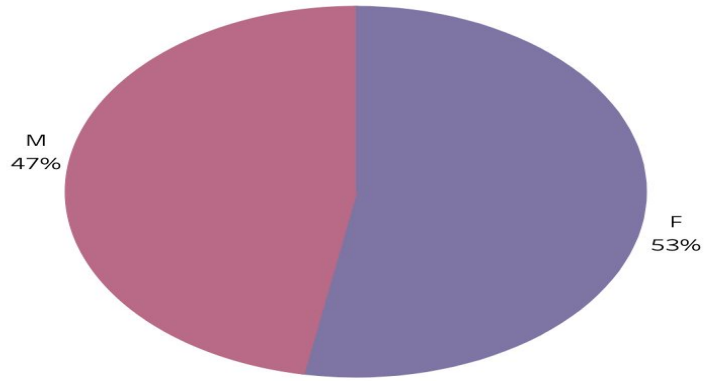
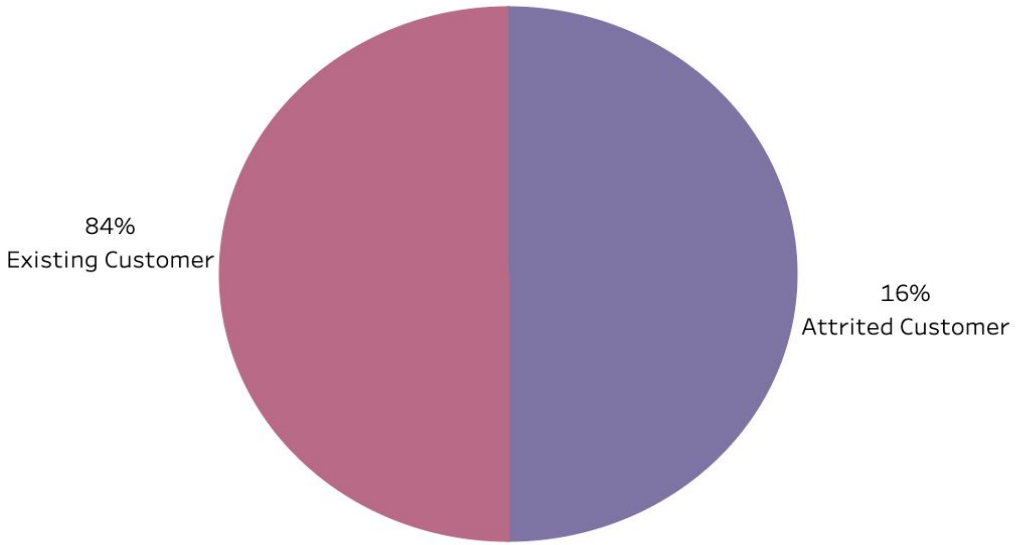
--===Credit Card Usage analysis using Excel,Tableau and R===--

Bala Mira G

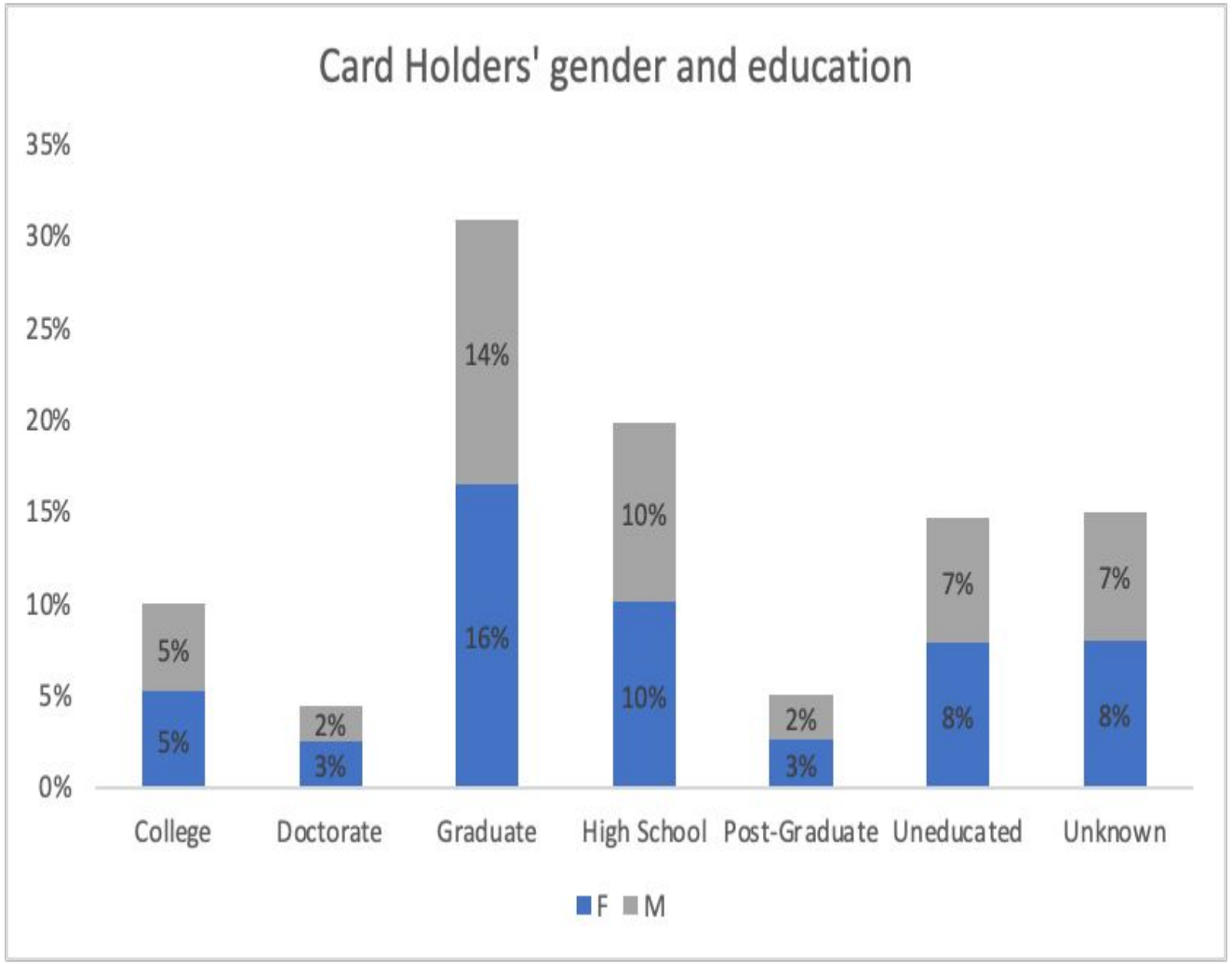
--===Basic EDA for customer data===--

Customer count:10,127



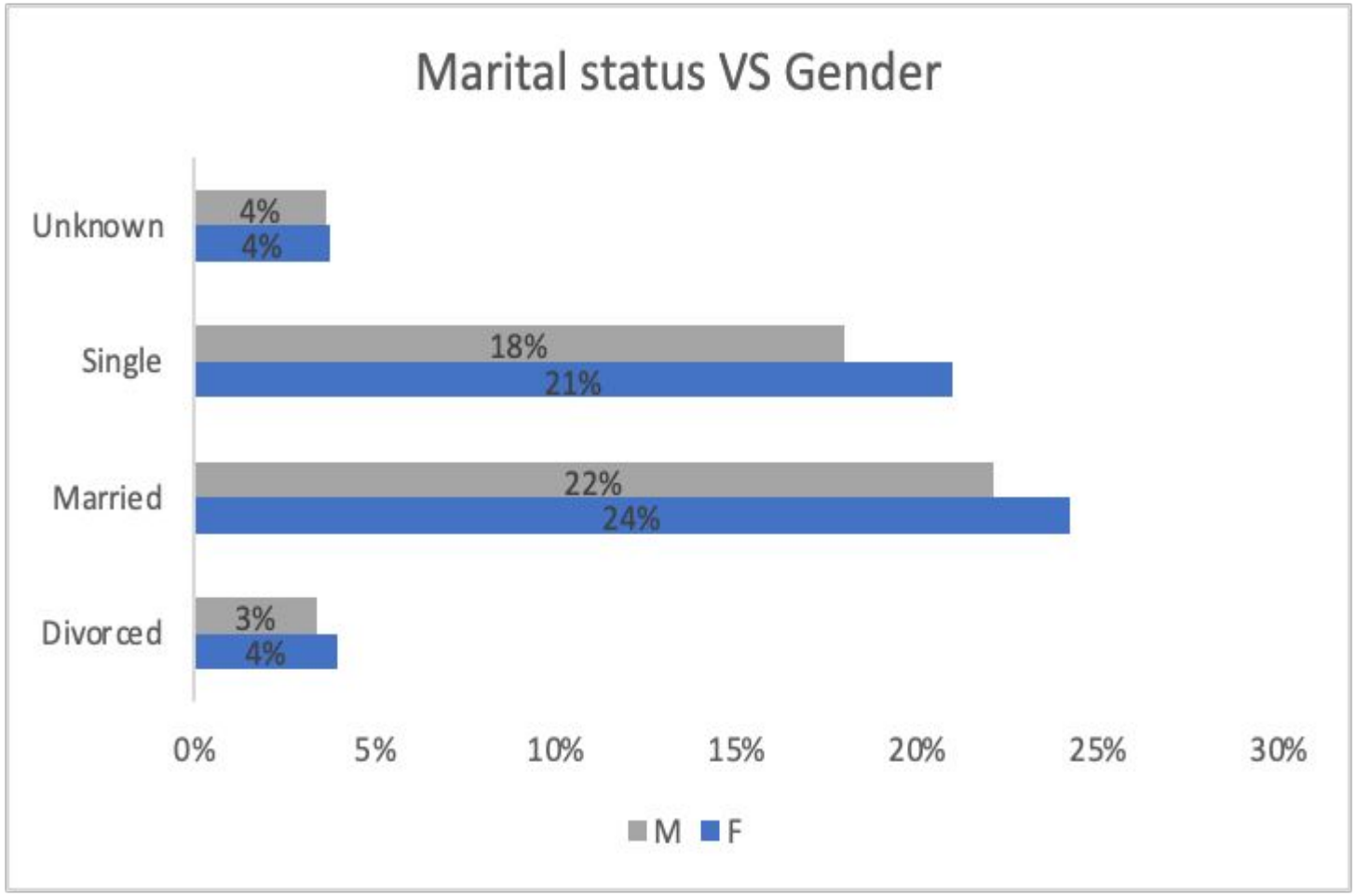


--===Basic EDA for customer data===--



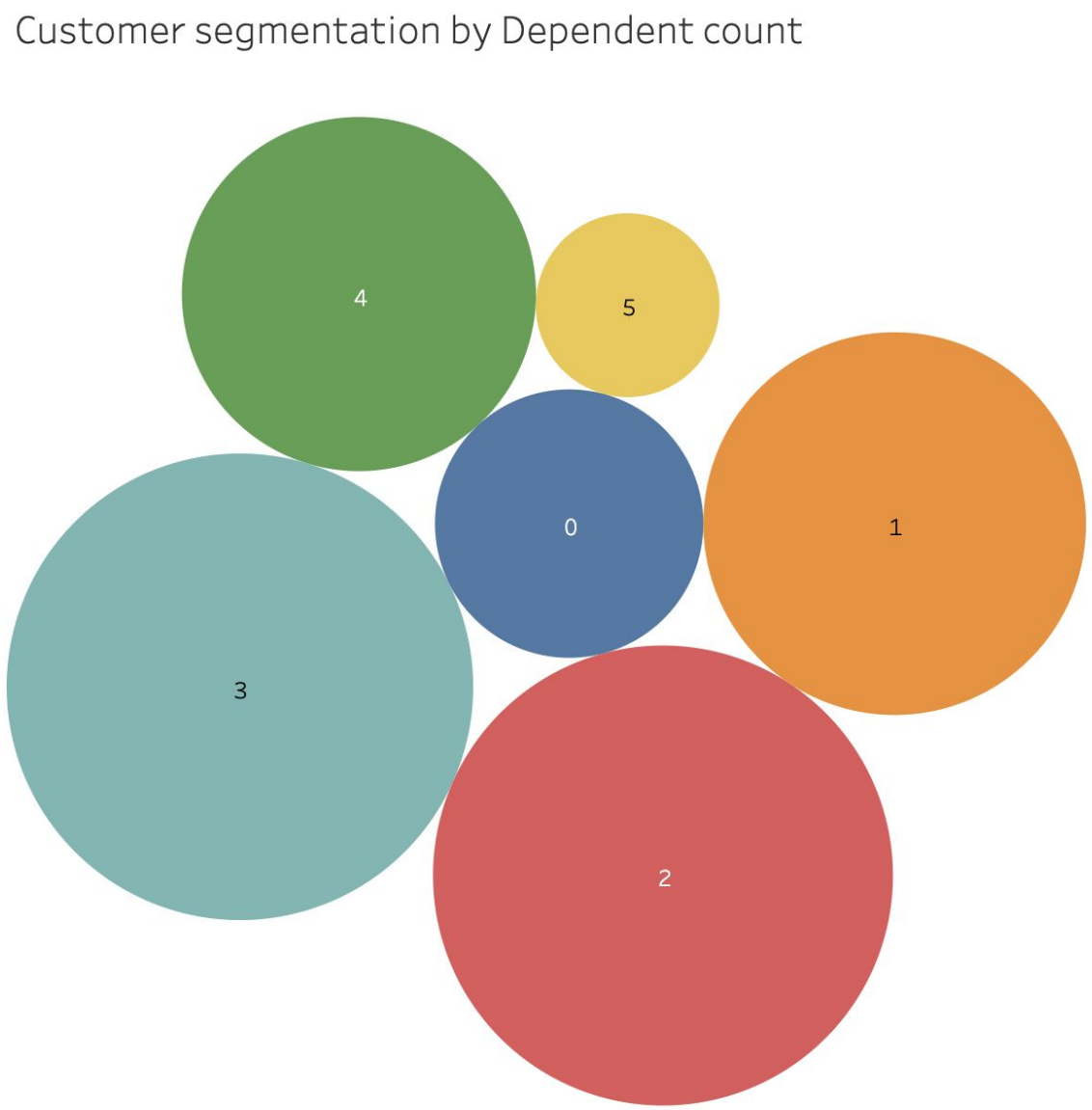
- As we can infer from the graph, we can notice that most of our customers (around 30%) are graduates.
- Understanding this distribution is essential for card companies as they design their product offerings which are personalized for them.
- In contrast, the Doctorate holders have the least amount of transactions done.
- Companies can also offer incentives whenever the transaction activity for a particular customer segment is very low.

--===Basic EDA for customer data===--



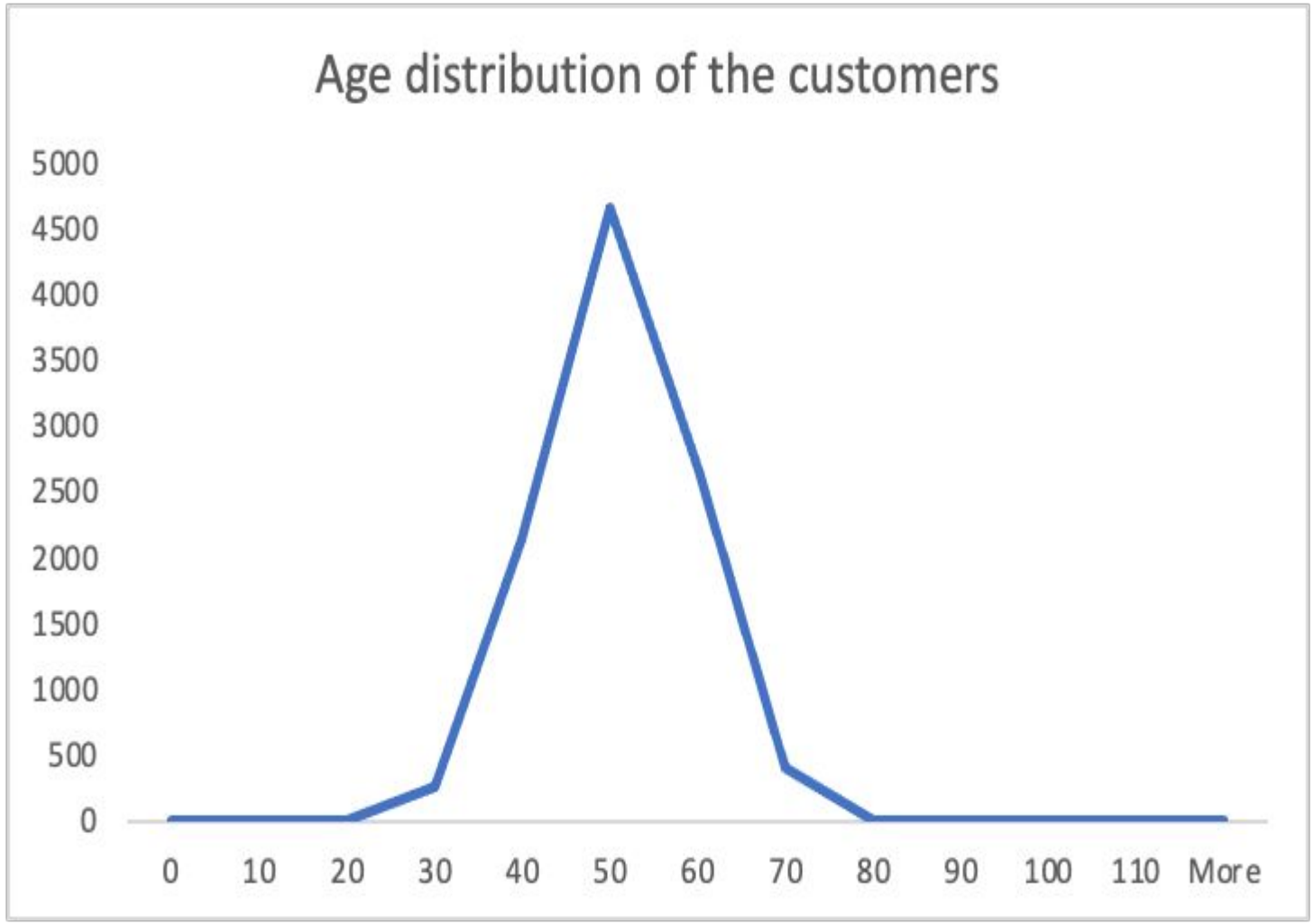
- The bar chart analysis provides valuable insights into the distribution of credit card holders by marital status.
- The largest segment of credit card holders is married individuals, accounting for a significant percentage of 46%.
- This indicates that married individuals are more likely to hold credit cards compared to other marital status categories.
- The next prominent category is singles, comprising 39% of credit card holders. This suggests that unmarried individuals, including those who are single or never married, form a substantial portion of the customer base.
- Understanding the distribution of credit card holders by marital status is crucial for tailoring marketing strategies, developing targeted financial products, and offering appropriate benefits that cater to the needs and preferences of specific marital status segments.

--===Basic EDA for customer data===--



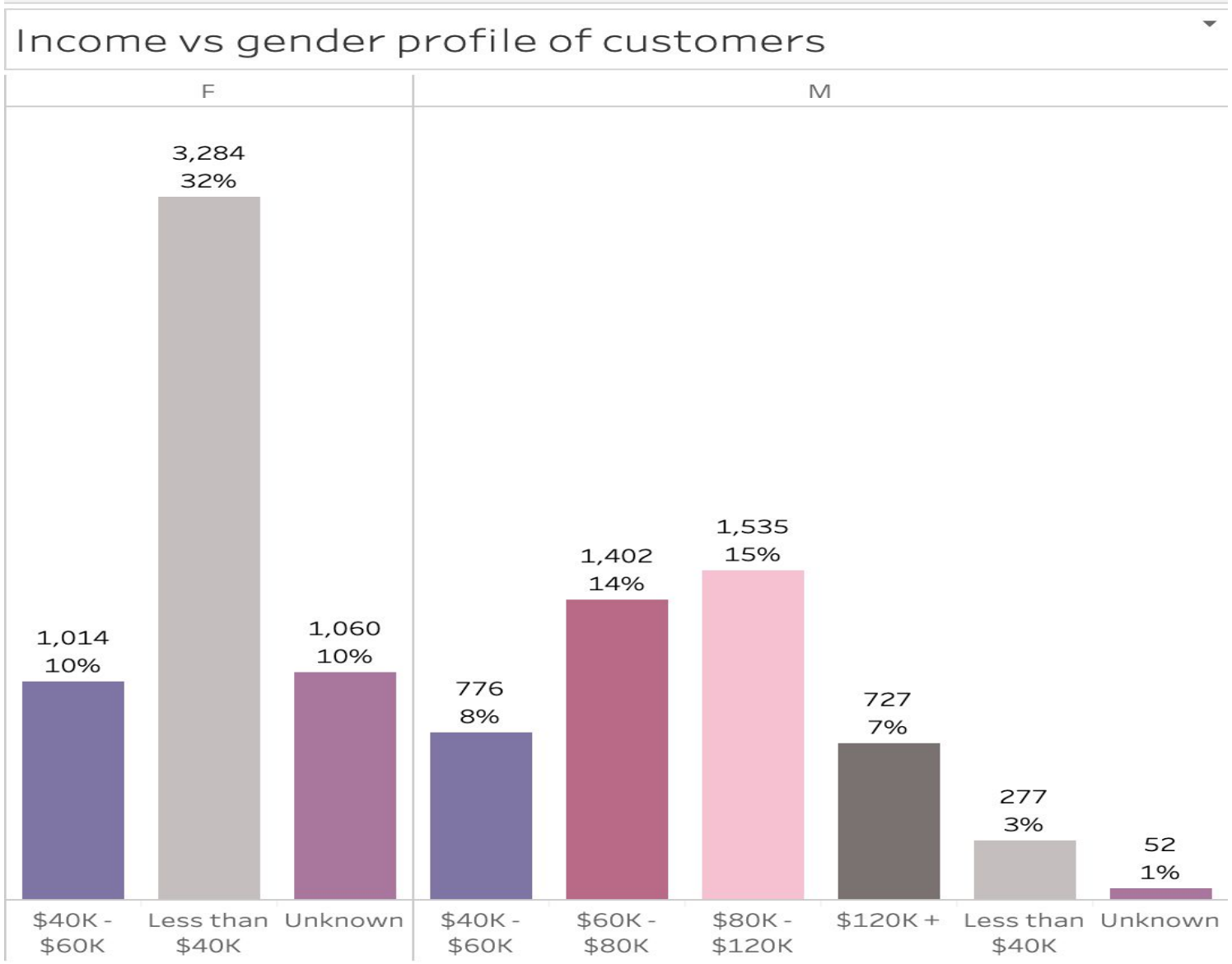
- The bubble chart analysis provides valuable insights into the dependent count of credit card holders.
- Based on the graph, we can clearly see that people with more dependents form most of our customers.
- Companies can also introduce more ‘Family focused plans’ .
- A joint credit card account can make it easier to manage bills each month. That can help account owners, like a married couple, simplify their finances.
- In case of family plans, if one of the cardholders has less positive credit history than the other, they can take advantage of the joint account holder's higher credit scores to gain access to better interest rates and higher credit limits on a credit card.

--===Basic EDA for customer data===--



- The line chart analysis provides valuable insights into the age distribution of credit card holders.
- Based on the graph, The majority of your customers fall within the age range around 50 years old.
- This peak suggests that the customer base is concentrated in the middle-age group.
- There is a significant drop in the number of customers as age decreases below 50 and increases above 50.
- Very few customers are below age 20 or above age 80.
- The graph indicates that the most of the credit card users are middle aged people.
- Companies can devise strategies to attract the attract younger customers (below 30) and older customers (above 60).
- Promotions or customized offerings for these age groups could be implemented to widen the customer base.

--===Basic EDA for customer data===--

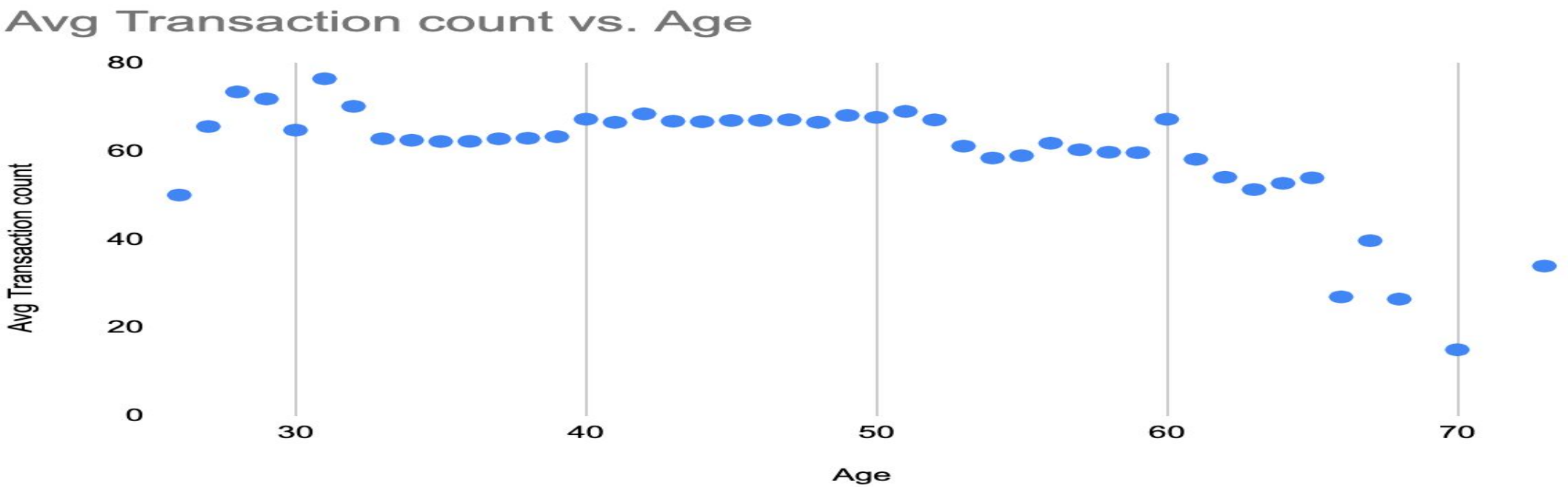


- The bar chart analysis provides valuable insights into the gender and income levels of credit card holders.
- Based on the graph, The majority of your customers are females (who earn less than 40k) who make up 32% of our customer base.
- This might be due to the fact that they are borrowing to meet their daily ends.
- On the contrary, the high income earners do not rely highly on credit cards because they might either be financially settled or would prefer other methods for financial management.

--===Business Case Scenarios===--

	Age	*Transaction count*
Age	1	
Transaction count	-0.6658988	1

- The company was worried that the transaction count is decreasing as age increases and wanted to find ways to prove this hypothesis.
- Correlation and the scatter plots reveal that as age increases the transaction count decreases .
- But correlation does not always imply causation.
- We can use regression analysis or any other statistical tests to test this hypothesis.



--===Regression analysis between age and Transaction count===--(Moderate Negative)

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.66589884
R Square	0.44342127
Adjusted R Square	0.43047758
Standard Error	9.65856018
Observations	45

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	3195.8262	3195.8262	34.2577135	5.9799E-07
Residual	43	4011.37474	93.2877848		
Total	44	7207.20095			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	90.5812791	5.46120517	16.5863168	2.673E-20	79.5677092	101.594849	79.5677092	101.594849
X Variable 1	-0.6411789	0.10954689	-5.8530089	5.9799E-07	-0.8621013	-0.4202566	-0.8621013	-0.4202566

- **R-Square (R^2):** The R^2 value is 44.34%, which tells that approximately 44.34% of the variation in total transactions count can be explained by age. In other words, knowing a customer's age helps predict a significant portion of the variation in transaction count.
- This can be explained by the fact that younger customers spend more for socializing, shopping for personal needs and other activities that involve transaction count whereas the older people are already settled into certain a certain lifestyle thus lowering their transaction frequency.
- In summary, the regression model indicates that age plays a role in predicting total transactions count. As customers get older, their transaction count tends to decrease.

--===Regression analysis between age and Transaction amount====--(Moderate Negative)

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.65933733
R Square	0.43472571
Adjusted R Square	0.4215798
Standard Error	924.129879
Observations	45

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	28241682.9	28241682.9	33.0692655	8.421E-07
Residual	43	36722689.4	854016.033		
Total	44	64964372.4			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	6896.04942	522.527455	13.1974872	1.015E-16	5842.27238	7949.82646	5842.27238	7949.82646
X Variable 1	-60.274408	10.4814333	-5.7505883	8.421E-07	-81.412232	-39.136583	-81.412232	-39.136583

Analysis explanation:

R Square:

This is the proportion of the variation in the average transaction amount that can be explained by age. In our case, approximately 43.5% of the variability in the average transaction amount is explained by customers' age.

X Variable 1 (Age): The coefficient tells you how much, on average, the average transaction amount changes for each unit increase in age. Here, it's -60.27, indicating a negative relationship. In simpler terms, as age increases, the average transaction amount tends to decrease.

Summary of the analysis

Summary of the analysis:

Customers' Age and Transaction Count:

Finding: When you analyzed the relationship between age and transaction count, you found a negative relationship. In simpler terms, as customers get older, the number of transactions tends to decrease.

Example: A teenager makes frequent small purchases like snacks, gadgets, or entertainment. As people grow older and perhaps settle into more stable lifestyles, they might have fewer small transactions.

Customers' Age and Transaction Amount:

Finding: On the other hand, when looking at the relationship between age and transaction amount, we found a moderate negative relationship. This suggests that, on average, as customers get older, the amount of money spent per transaction also tends to decrease.

Example: A person in his/her 20s may spend a significant amount on one or two big purchases like a smartphone or travel. As people age and possibly prioritize saving or have different spending priorities, the average amount spent per transaction might decrease.

In summary, the credit card data is telling that, as customers age, they tend to make fewer transactions overall.

Additionally, while the number of transactions decreases, the average amount spent per transaction also tends to decrease, indicating potential shifts in spending patterns as people grow older.

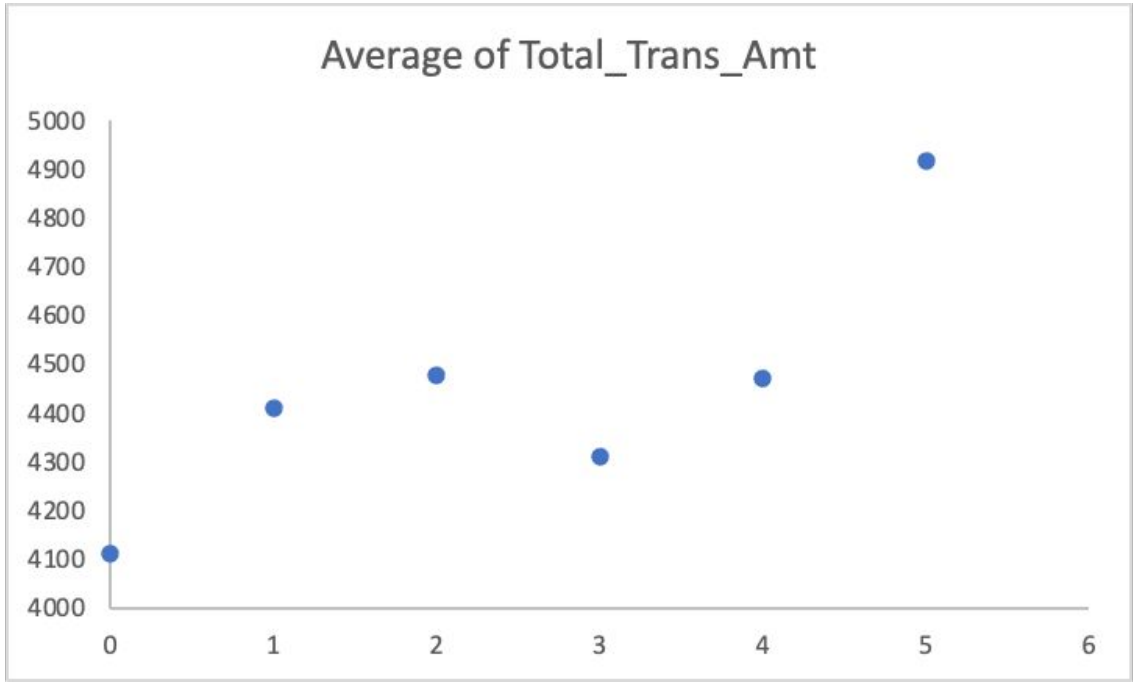
Recommendations:

Consider introducing Loyalty programs,enhanced customer service,feedback survey analysis from the older people to improve the products, services, and marketing strategies to better cater to their needs.

-----Regression analysis between dependent count and Avg Transaction amount (Positive)===-

	Dependent count	***Avg Transaction amount***
Dependent count	1	
Avg Transaction amount	0.81003536	1

- The company wanted to check if dependent count was in any way connected to the average transaction amount.
- Correlation and the scatter plots reveal that as dependent count increases the average transaction amount increases .
- But correlation does not always imply causation
- We once again use regression analysis to delve deep into the data to gain insights.



----Regression analysis between dependent count and Avg Transaction amount (Positive)===-

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.81003536
R Square	0.65615728
Adjusted R Square	0.5701966
Standard Error	174.990344
Observations	6

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	233741.746	233741.746	7.63322592	0.05070226
Residual	4	122486.482	30621.6204		
Total	5	356228.228			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	4160.04372	126.648713	32.8471063	5.1225E-06	3808.41052	4511.67692	3808.41052	4511.67692
X Variable 1	115.571066	41.830693	2.76282933	0.05070226	-0.5695572	231.711689	-0.5695572	231.711689

Analysis explanation:

R Square:

This is the proportion of the variation in the average transaction amount that can be explained by age. In our case, approximately 66% of the variability in the average transaction amount is explained by dependent count.

X Variable 1 Coefficient: 115.5710657

This coefficient represents the estimated change in the "Avg Transaction amount" for a one-unit change in the "Dependent count." In simpler terms, it indicates the units by which average transaction amount is expected to increase in the dependent count associated with a credit card account

Significance F (p-value): 0.050702261

The p-value tests the null hypothesis that all the coefficients are equal to zero.

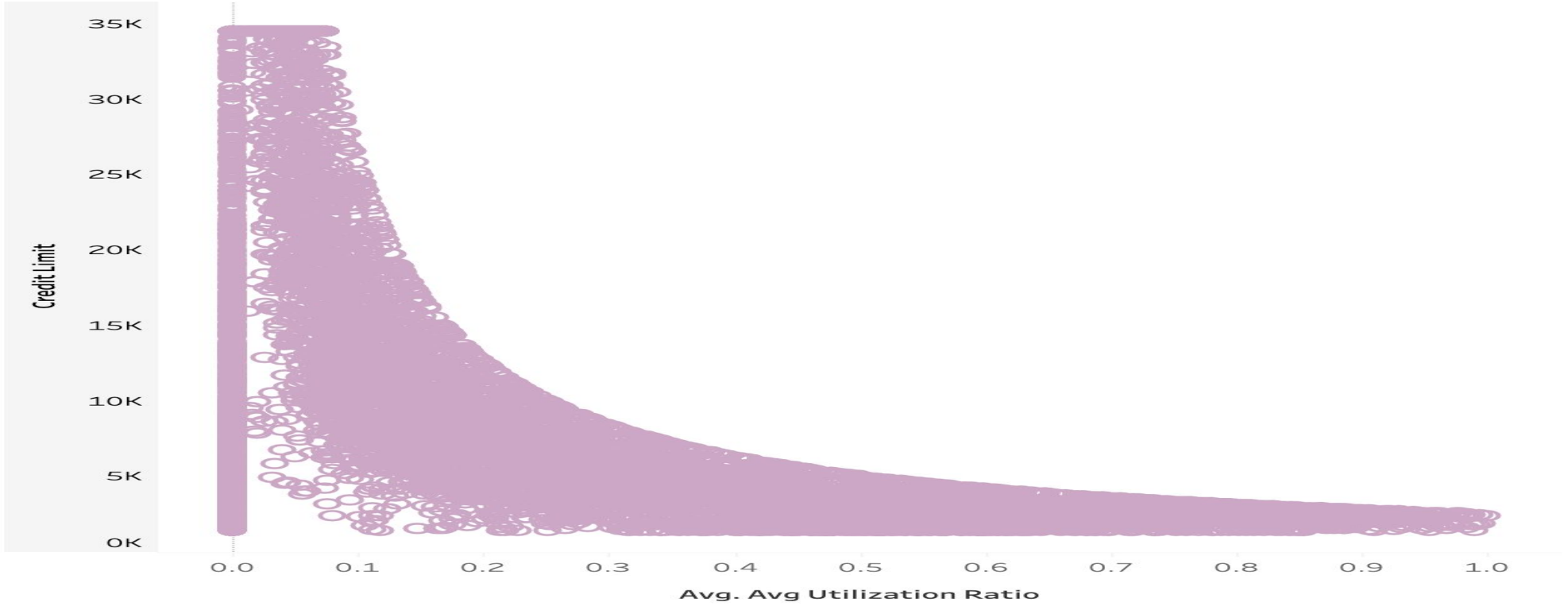
A p-value less than the significance level (commonly 0.05) tells us the overall regression model is statistically significant.

In summary the regression model tell us that as dependent count increases, the average transaction amount would also increase.

-----Regression analysis between Avg Utilization ratio and Credit Limit

	Credit Limit	*Avg Utilization ratio*
Credit Limit	1	
Avg Utilization ratio	-0.5780663	1

- We plot the Correlation between Avg Utilization ratio and Credit limit in a scatter plot.
- The scatter plot presents an inverse relationship between the credit limit and the average utilization ratio.
- We can use R to perform further deeper analysis by usign Regression.



-----Regression analysis between Avg Utilization ratio and Credit Limit using R(Moderate negative)====-

```
Residuals:
     Min       1Q   Median       3Q      Max
-0.37060 -0.11178 -0.02088  0.10463  0.63161

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   3.950e-01  3.675e-03   107.5   <2e-16 ***
Credit_Limit -1.699e-05  3.045e-07   -55.8   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.185 on 6203 degrees of freedom
Multiple R-squared:  0.3342,    Adjusted R-squared:  0.3341
F-statistic:  3113 on 1 and 6203 DF,  p-value: < 2.2e-16
```

- **R-squared:** It denotes that approx 33.42% of the variability in the Credit Limit is explained by age factor.
- This model indicates that as the average utilization ratio decreases, the credit limit tends to increase, and vice versa.
- This model also suggests that customers who use a smaller proportion of their credit limit tend to have higher credit limits. This could be an indicator of good financial management, where maintaining a low utilization ratio leads to increased credit trustworthiness and, consequently, higher credit limits. The credit card company might also be rewarding such behavior with higher credit limits to incentivize low utilization ratios.
- These insights can be used by companies to to encourage responsible credit usage among cardholders.

Analysing the association between categorical and numerical variables

-----Linear regression between Avg Utilization ratio and Education Level using R====-

```
model <- lm(Avg_Utilization_Ratio ~ Education_Level, data = credit)
print(model)
```

```
all:
m(formula = Avg_Utilization_Ratio ~ Education_Level, data = credit)

oefficients:
                 (Intercept)      Education_LevelDoctorate       Education_LevelGraduate
                    0.265399                     -0.004780                     0.012320
 Education_LevelHigh School  Education_LevelPost-Graduate     Education_LevelUneducated
                    0.012876                     0.015173                      0.006191
     Education_LevelUnknown
                    0.011070
```

Intercept (0.265399): This represents the estimated average utilization ratio when the educational level is zero. In this context, it doesn't have a meaningful interpretation.

Education_LevelDoctorate (-0.004780): On average, individuals with a Doctorate education level have a slightly lower average utilization ratio compared to the reference group.

Rest other individuals with education levels other than Doctorate , have higher average utilization ratio

The above analysis may indicate that those with higher educational qualifications beyond a graduate degree may not see the need for credit cards or prefer other means of financial management. These insights could be valuable for credit card companies to strategize their marketing efforts based on education levels.

Analysing the association between categorical and numerical variables

-----Linear regression between Transaction amount and Education Level using R====-

```
> model <- lm(Total_Trans_Amt ~ Education_Level, data = credit)
> print(model)

Call:
lm(formula = Total_Trans_Amt ~ Education_Level, data = credit)

Coefficients:
                  (Intercept)      Education_LevelDoctorate       Education_LevelGraduate
                      4255.81                        -62.79                        162.68
  Education_LevelHigh School  Education_LevelPost-Graduate     Education_LevelUneducated
                       147.93                        259.61                        228.53
      Education_LevelUnknown
                       164.25
```

- I**ntercept (4255.81):** This represents the estimated total transaction amount when the educational level is zero. In this context, it doesn't have a meaningful interpretation.
- **Education_LevelDoctorate (-62.79):** On average, individuals with a Doctorate education level have a slightly lower total transaction amount compared to individuals with other educational qualifications.
- Rest other individuals with education levels other than Doctorate have higher transaction amount.
- Credit card companies can consider using this information to enhance customer engagement strategies, such as personalized recommendations or exclusive offers for particular education categories.

-===Important Takeaways===-

- Females have s slightly more usage of credit cards than their male counterparts.
- A significant portion of credit card holders are graduates.
- Married individuals form (almost 46%) of the total credit card users.
- People with more dependents or middle aged individuals (between 40 to 60) are more likely to use credit cards than the others.
- Lower income people might as well use credit cards more.
- Statistical analysis found that the association age and transaction amount/count is moderate negative suggesting that as age increases the transaction amount/count would decrease as older age people see the less need for using credit cards on a daily basis.
- Further analysis found that as Average Utilization Ratio decreases, the credit limit increases.The credit card companies can also consider rewarding such behavior with higher credit limits in order to encourage lower utilization ratios.
- Other than Doctorates, people from other educational levels such a Graduates,High schoolers tend to have high transaction amount and Average Utilization Ratio.