

EDA Phase - I

1. *Mohammed Zeeshan Ali*
2. *Shivangi Jagdishkumar Bhavsar*
3. *Yachen Cheng*
4. *Shreya Chauhan*



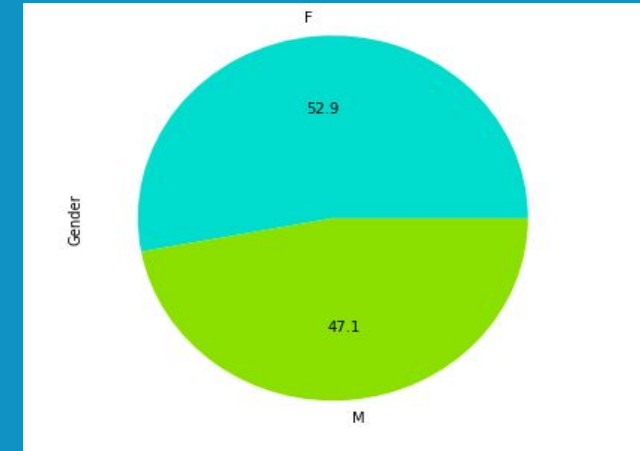
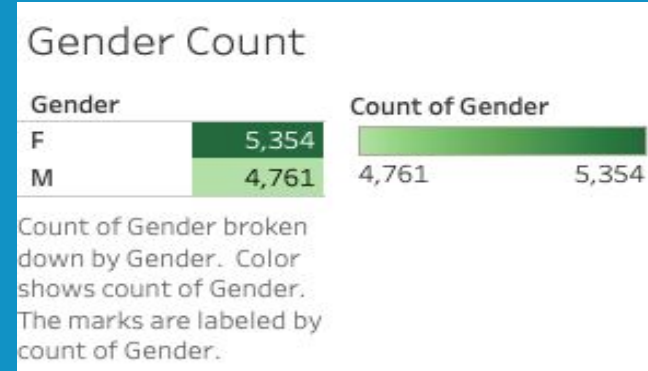
Exploratory Data Analysis Goal

Independent variables chosen for analysis by each member

- Gender & Income Category - Mohammed Zeeshan Ali
- Customer Age - Shivangi Jagdishkumar Bhavsar
- Education Status - Yachen Cheng
- Credit Limit & Marital Status - Shreya Chauhan

Individual Analysis – Gender Variable (Categorical) [Zeeshan]

- Fig 2 shows the comparative pie charts of gender variable
- Considering the Pie chart comparison female customers are more than male customers
- An absolute gender count table is shown in fig 1 which states the appropriate distribution of gender count



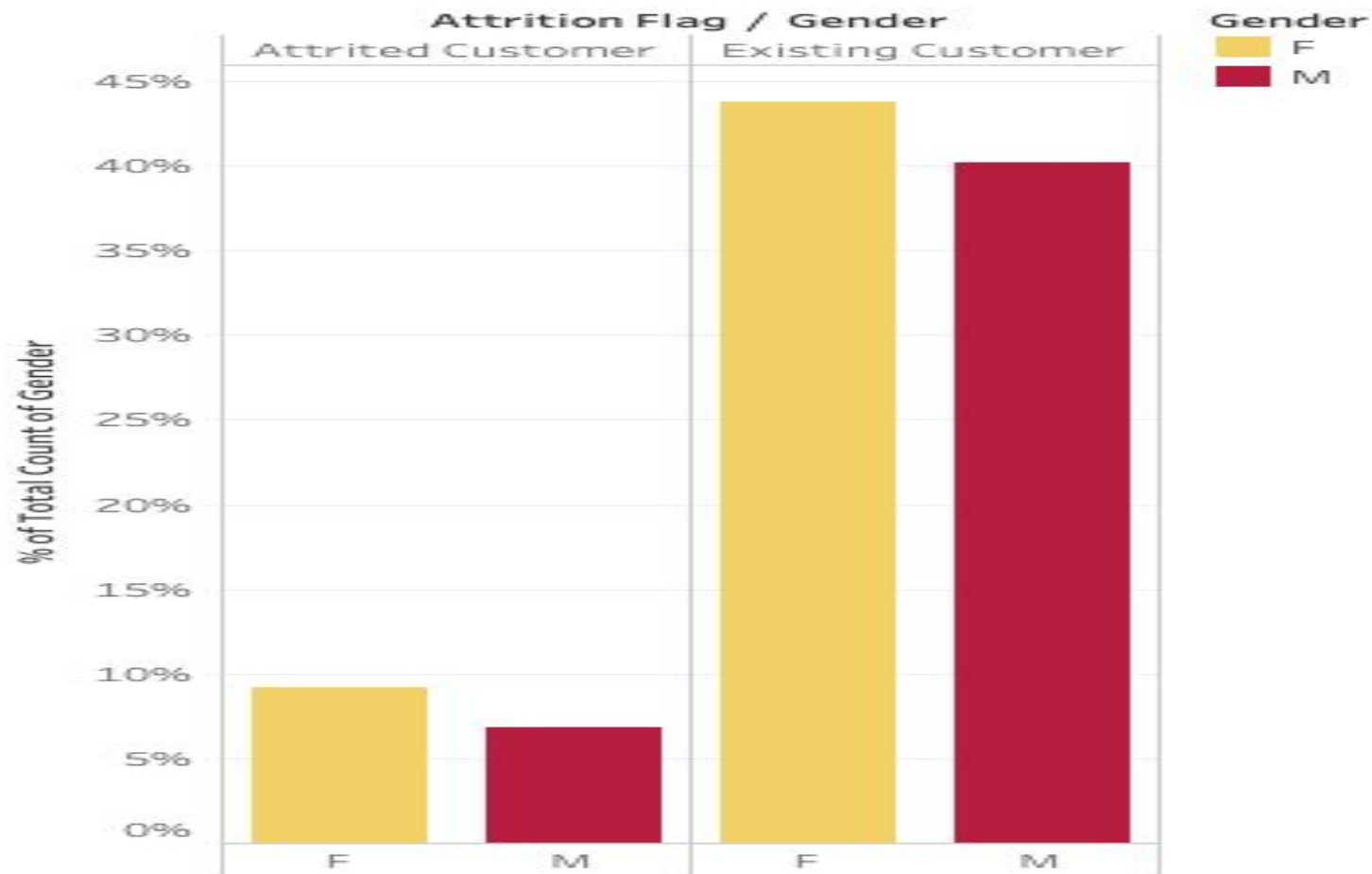
Individual Analysis – Gender Variable (Categorical) [Zeeshan]

➤ Relationship With Attrition Flag

- Attrited Female Customers – 927/1,621
- Attrited Male Customers – 694/1,621
- Existing Female Customers – 4,427/8,494
- Existing Male Customers – 4,067/8,494

Attrition Flag	Gender		% of Total Count of Gen..	
	F	M		
Attrited Customer	9.16%	6.86%	6.86%	43.77%
Existing Customer	43.77%	40.21%		

% of Total Count of Gender broken down by Gender vs. Attrition Flag. Color shows % of Total Count of Gender. The marks are labeled by % of Total Count of Gender.

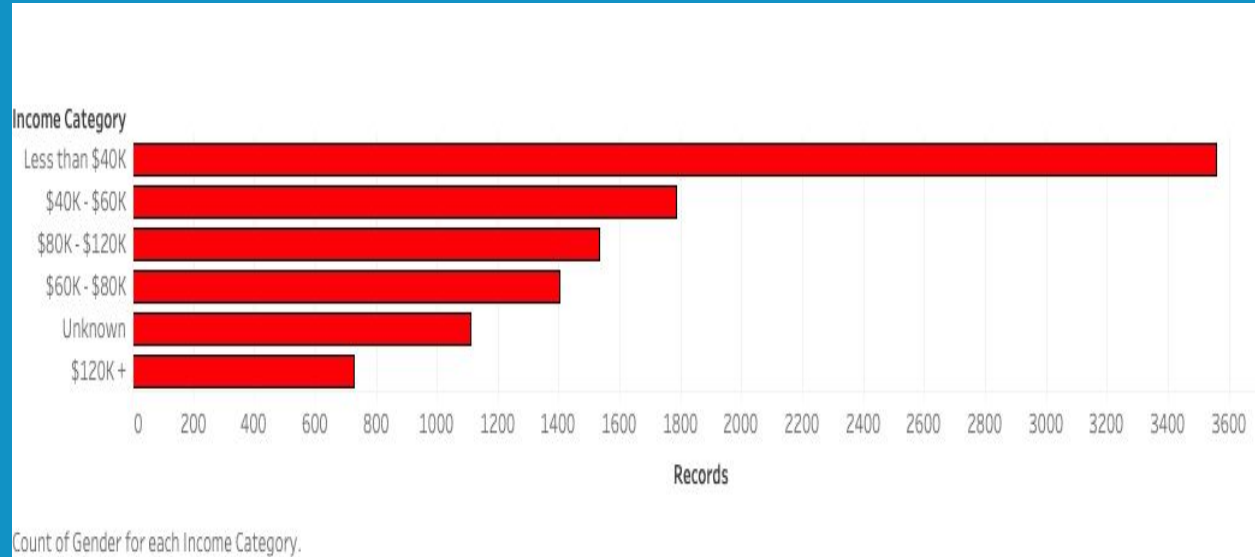


% of Total Count of Gender for each Gender broken down by Attrition Flag. Color shows details about Gender.

Summary

- ❑ There are more female customers than male customers in the list of existing customer records (53%)
- ❑ Based on the discussion of bar graph there are higher attrited female customers
- ❑ The attrited female customers are 9.16% while the attrited male customers are 6.86% which is shown in the contingency table.

Individual Analysis – Income Category Variable (Continuous) [Zeeshan]



- The bar graph discusses the distributed income category ranges from 'less than \$40k to \$120k +.'

Individual Analysis – Income Category Variable (Continuous) [Zeeshan]

➤ Relationship With Attrition Flag

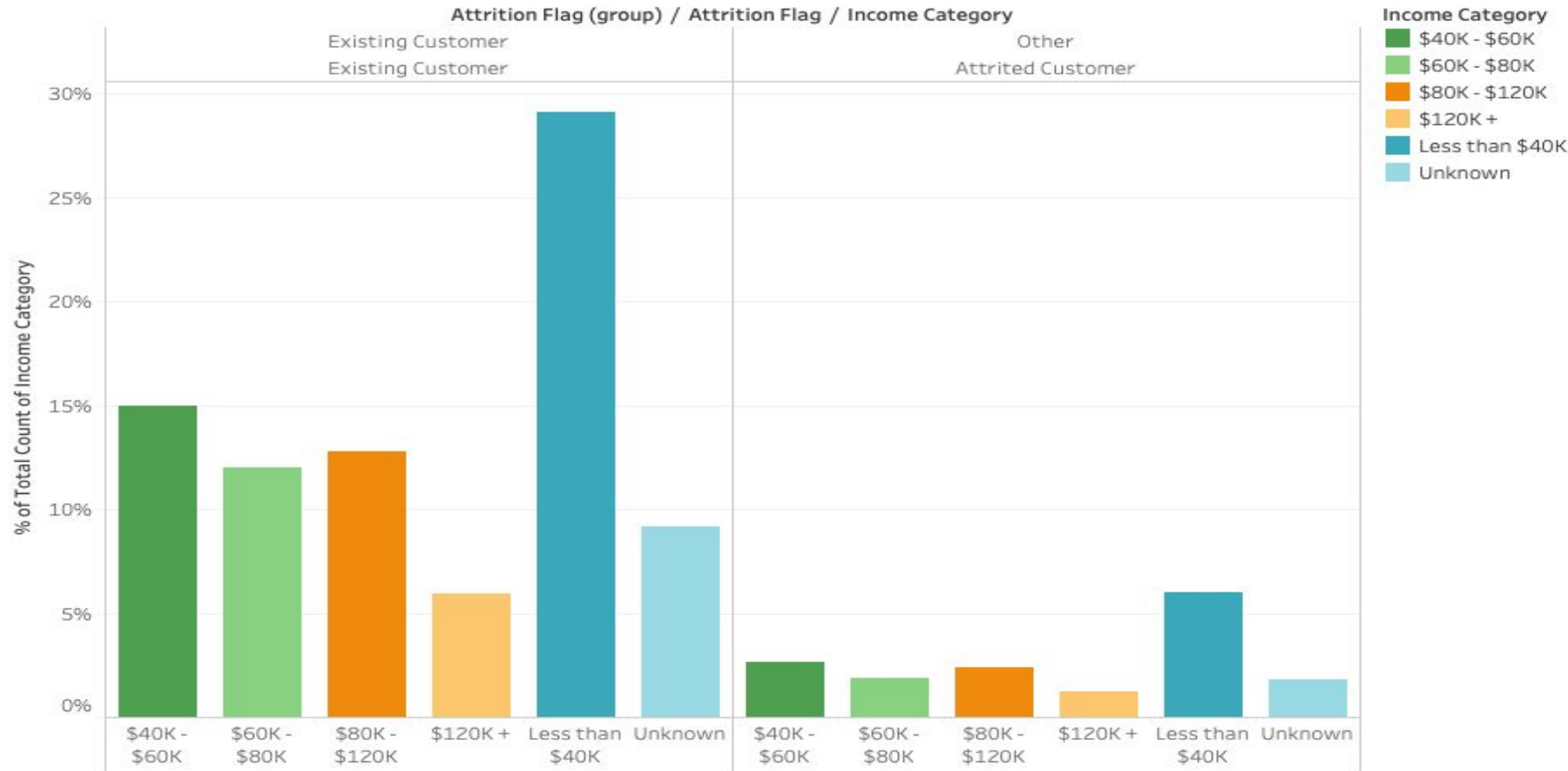
Customer Vs Income Category

		Income Category						Count of Income Categ..	
Attrition Flag (group)	Attrition Flag	\$40K - \$60K	\$60K - \$80K	\$80K - \$120K	\$120K +	Less than \$40K	Unknown	126	2,948
Existing Customer	Existing Customer	1,517	1,212	1,291	601	2,948	925		
Other	Attrited Customer	269	189	241	126	610	186		

Count of Income Category broken down by Income Category vs. Attrition Flag (group) and Attrition Flag. Color shows count of Income Category. The marks are labeled by count of Income Category. The view is filtered on Income Category, which keeps 6 of 6 members.

- Discussion of Attrition flag as target variable with Income category as a continuous variable
- To quantify the relationship a contingency table is used which shows customers with income of less than \$40k being existing customer and attrited customer.
- Out of 1,621, 610 customers having less than \$40k are attrited customers.

Customer Vs Income Category



% of Total Count of Income Category for each Income Category broken down by Attrition Flag (group) and Attrition Flag. Color shows details about Income Category. The view is filtered on Income Category, which keeps 6 of 6 members.

Summary

- ❑ Customers having an income of less than \$40k are given the credit card at a higher rate.
- ❑ Customers having income of less than \$40k have attrited the most
- ❑ Some of the records are unknown which is second highest among income category
- ❑ Considering the income the attrition of the customers can be improved