Exploratory Data Analysis On Churn Dataset

Presented by
Srijan Rit and Pritam Taldhi

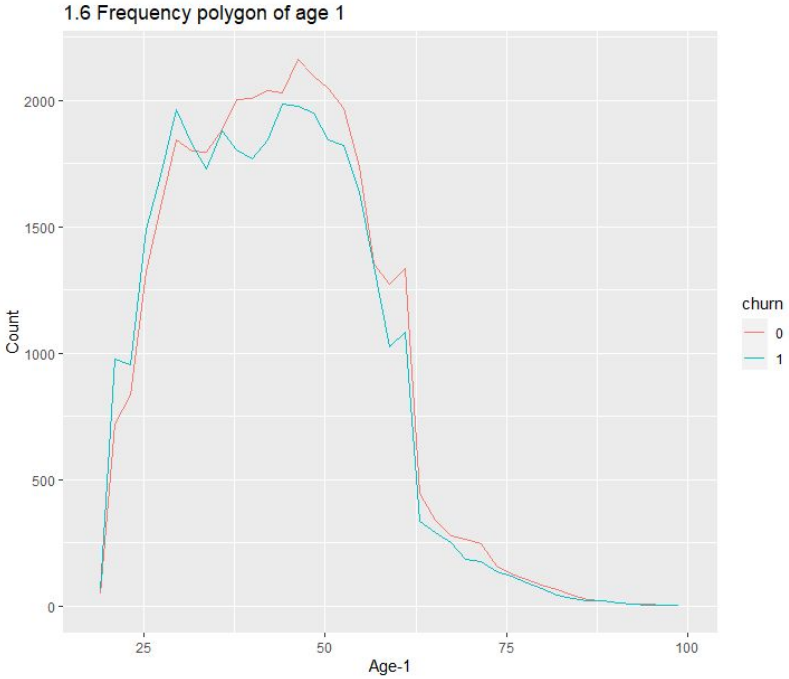
Data Description and Goal

The dataset contains information about 1,00,000 customers of a telecommunication company. There are 173 total columns with 46 variables that contain information about mean usage of the customers and 46 others with the corresponding mean range of the same parameter. The other 81 variable consists of general information about the customers. There are 128 numeric columns and 45 character columns.

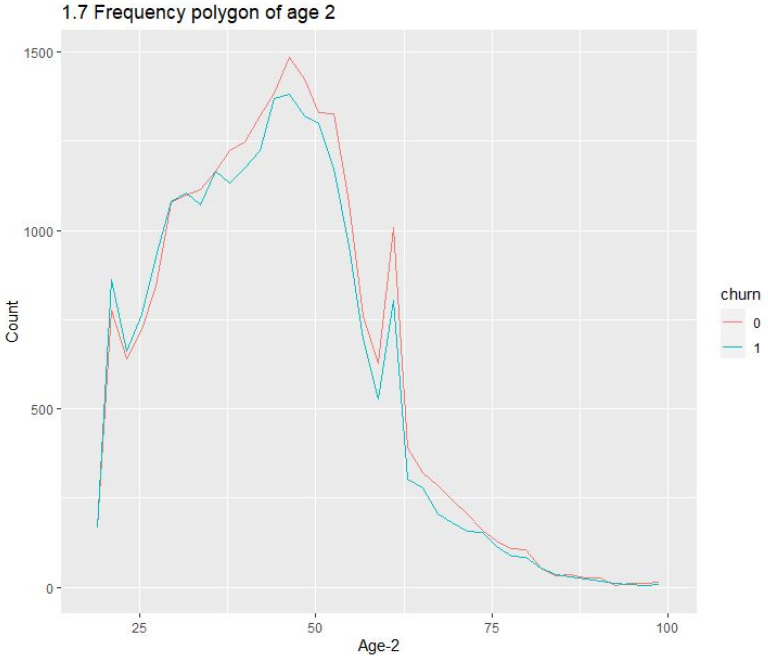
Of the 100,000 people 49,562 people have churned and 50,438 people were still a customer. The target variable is churn with two levels 0 and 1, 0 being churned customers and 1 being the customers still in service.

The company's aim is to retain as many profitable customers as possible. We try to find information from the dataset about which factors influence the likelihood of a customer to churn.

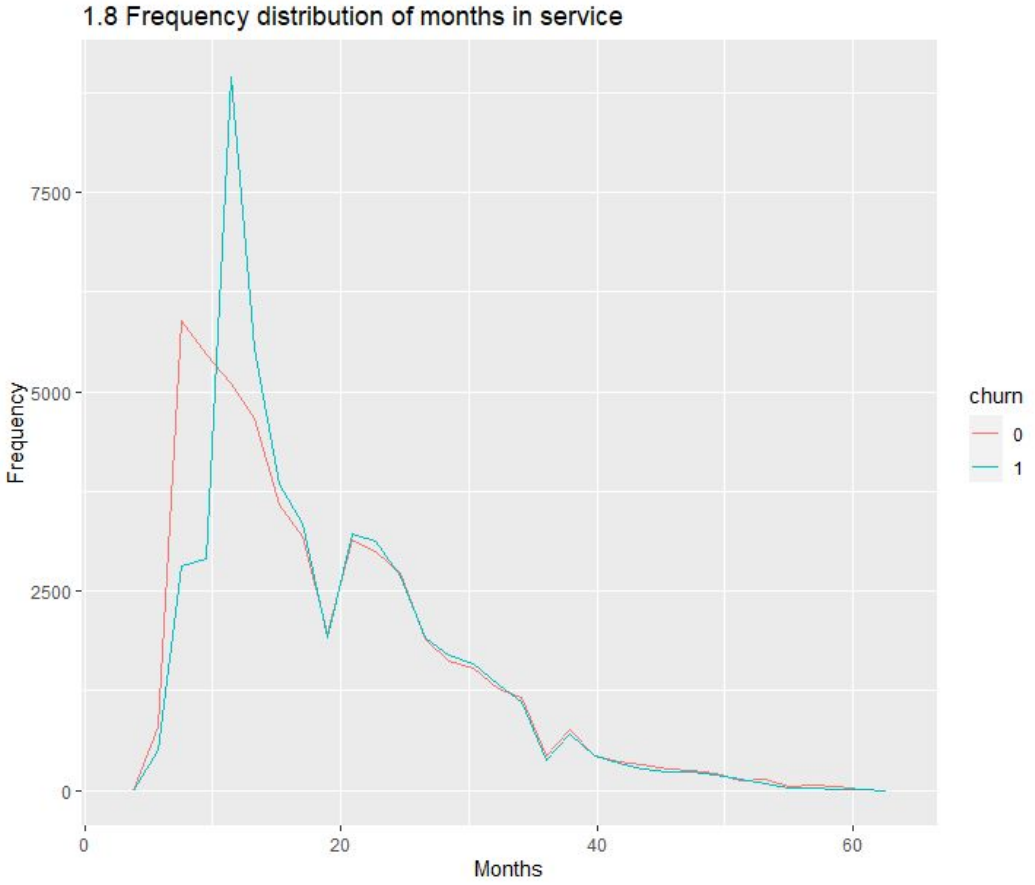
Softwares used: R and Excel



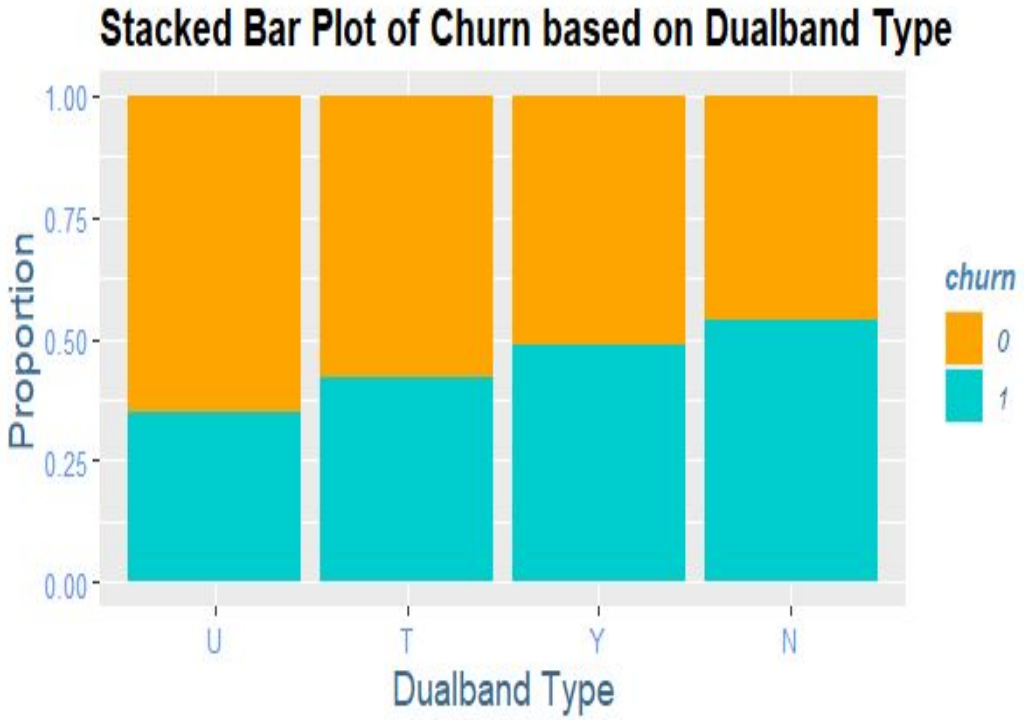
Upto age 30, more customers are in churn level 1 , however after age 30 more customers are in churn level 0.



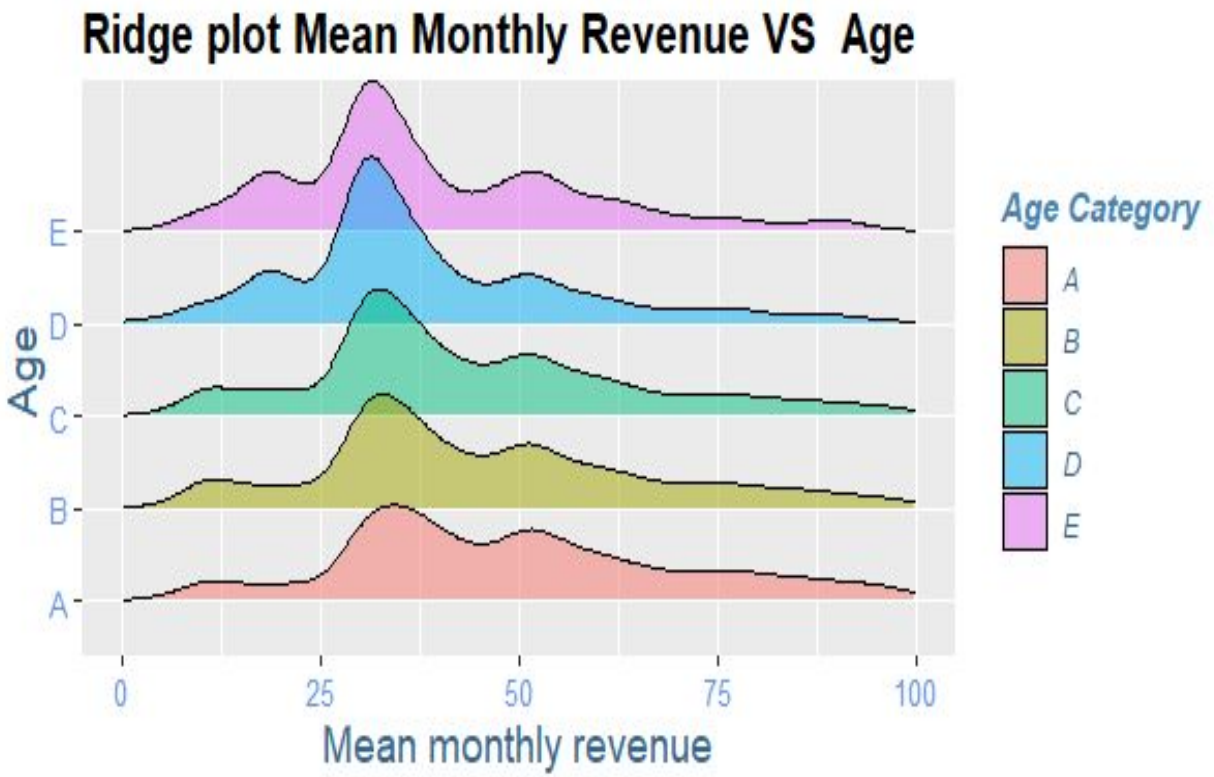
There is no significant difference between churn level with respect to age of second household member.



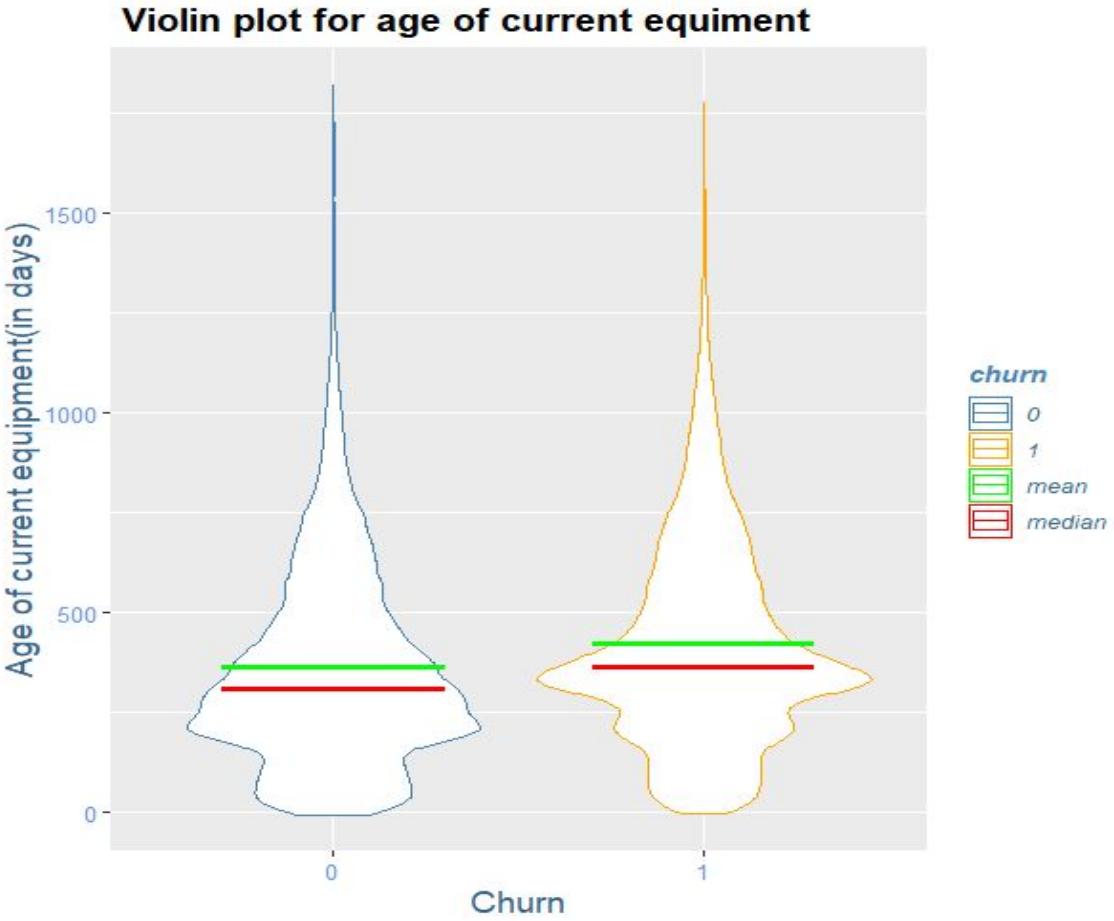
Most customers churn around 5 to 10 months of service.
Majority of current customers have been with the company for 11 months.
There are no customers who have been with the company for more than 5 years.



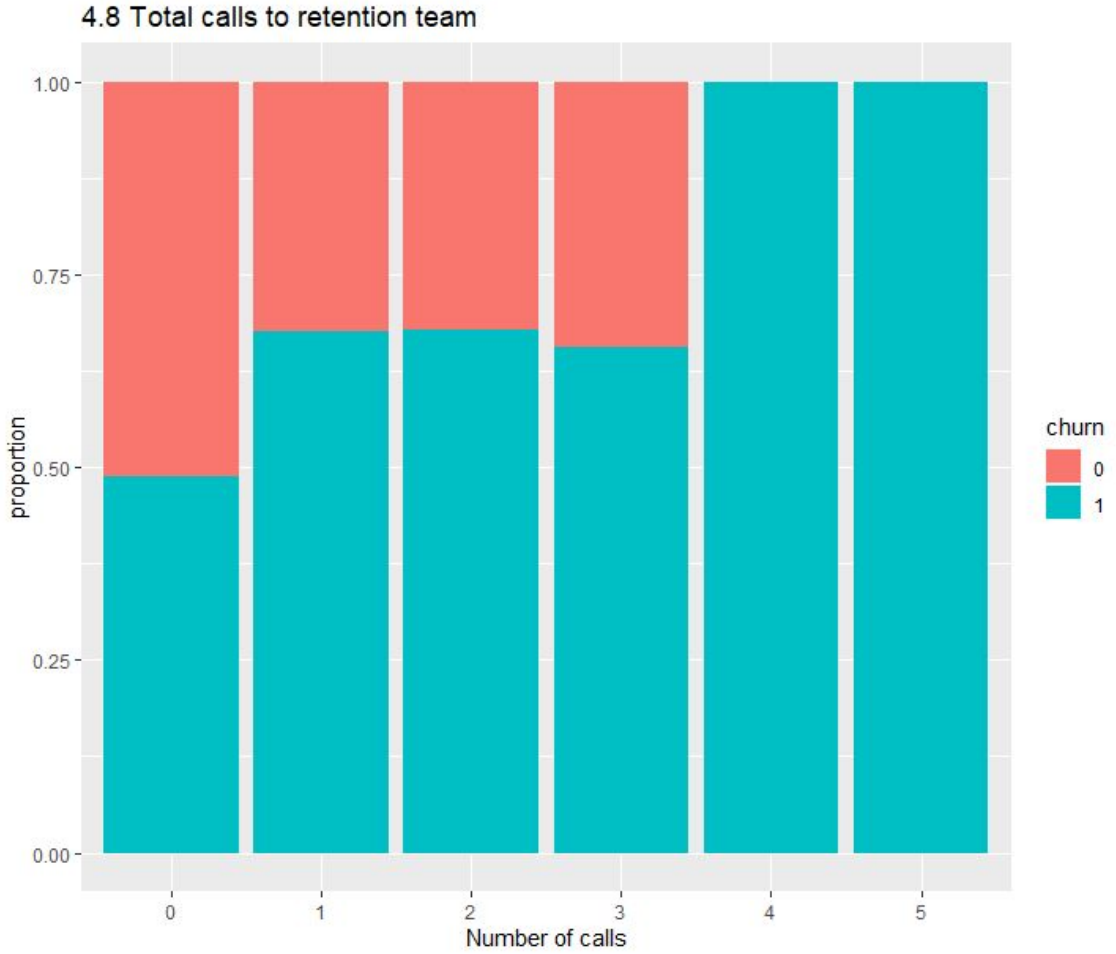
Dualband Type N were more likely to be in churn 1.
Dualband Type U were more likely to be in churn 0.



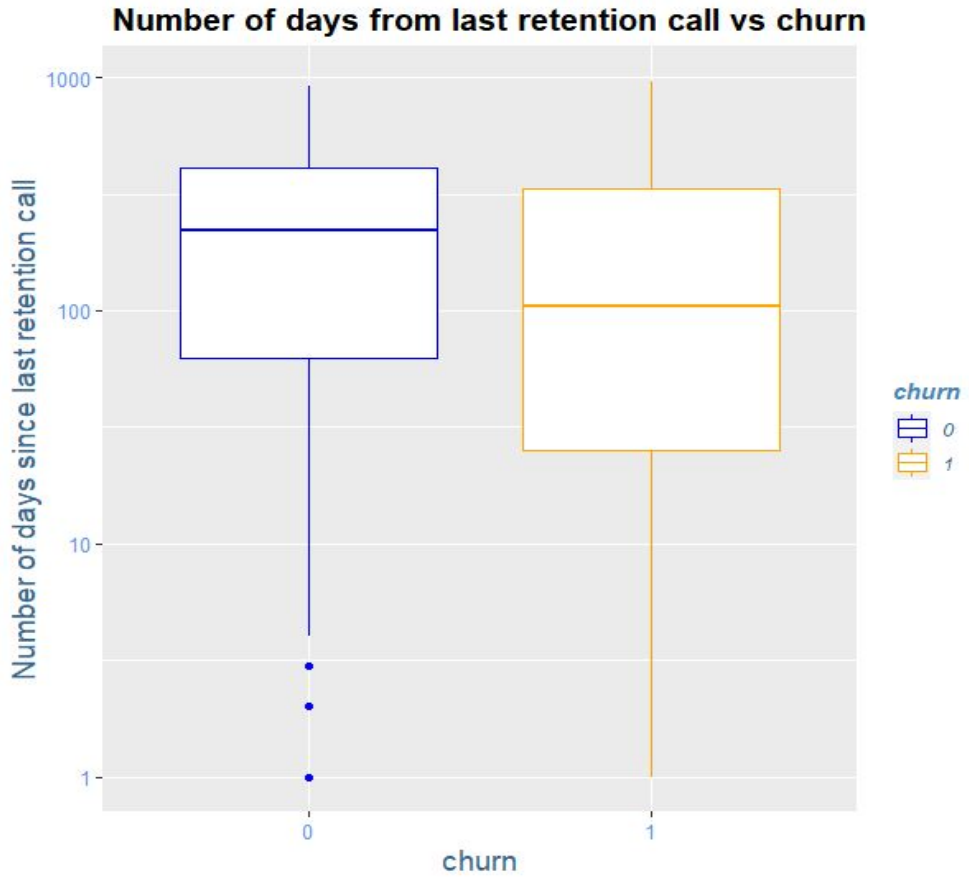
Median monthly revenue decreased as the age of customer increased.



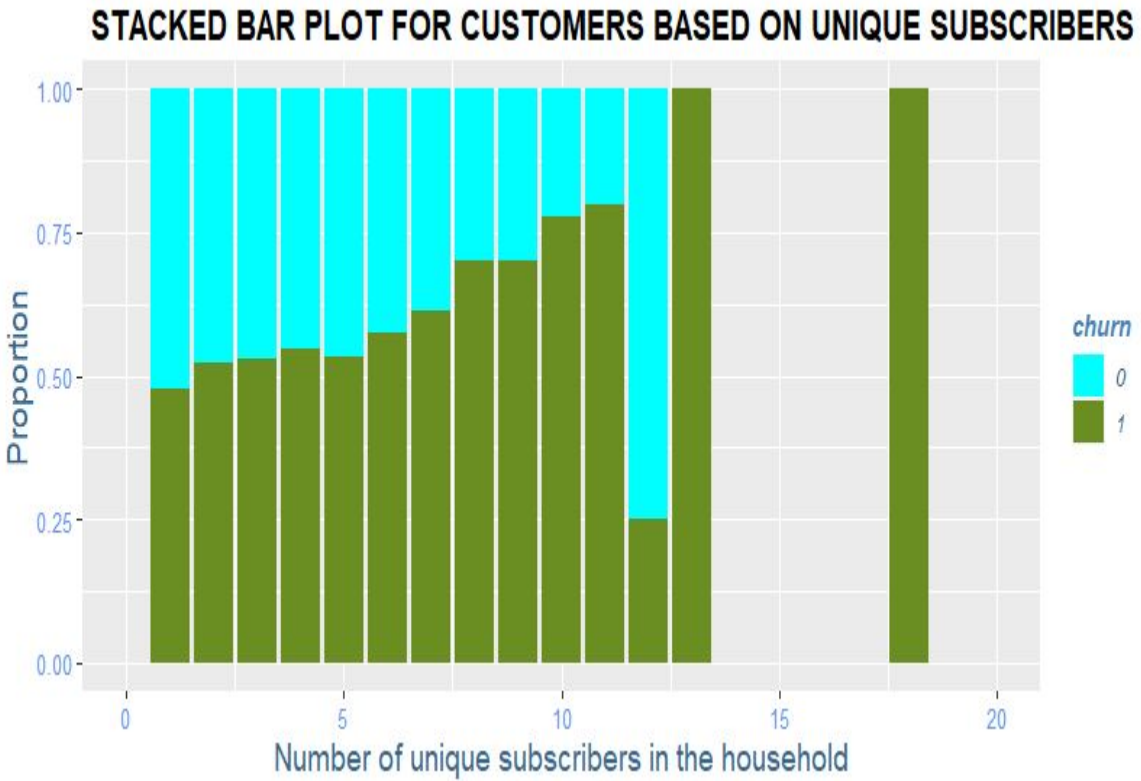
Customers in churn level 1, were more likely to have a higher equipment age.



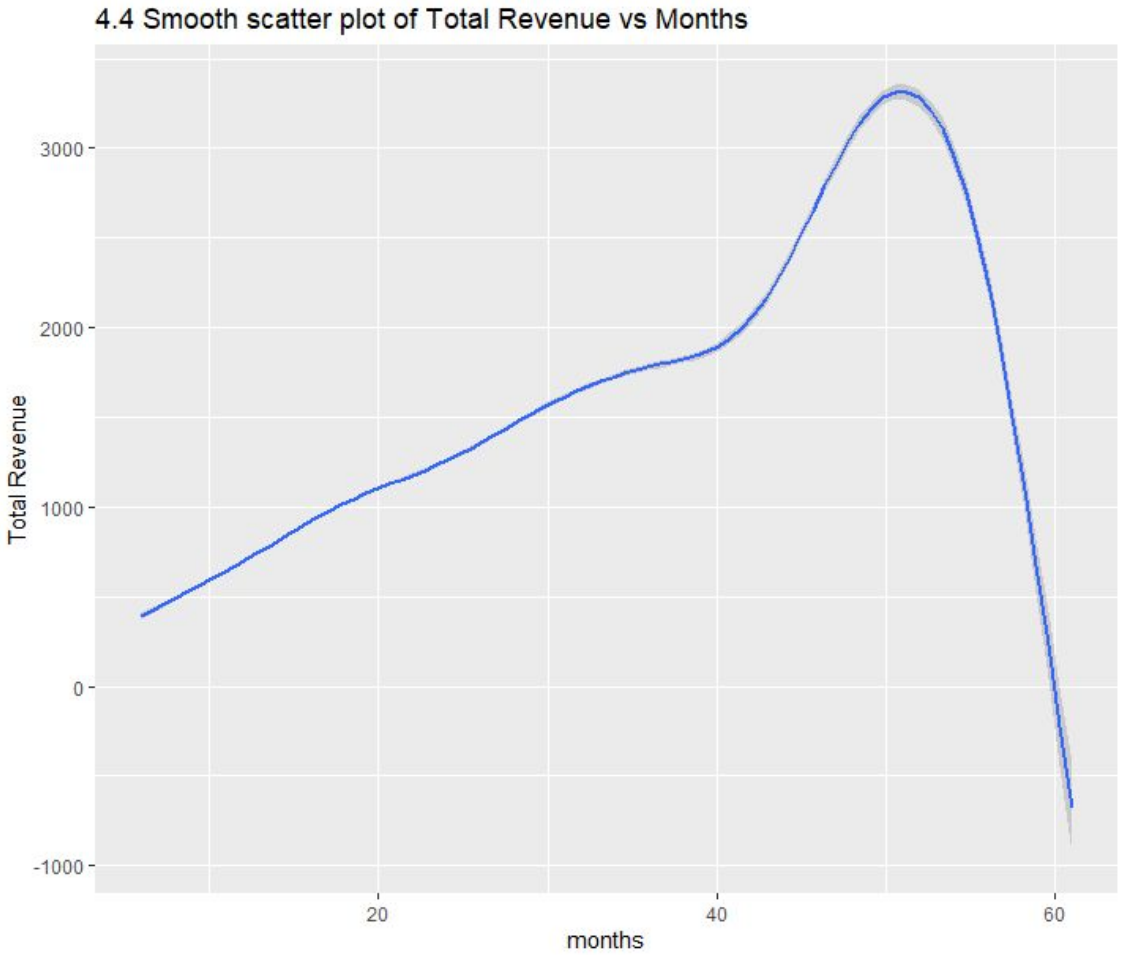
We have replaced missing values in Total Retention Calls with zero. Assuming zero calls have been made to them.
Clearly, no customer who received more than 3 retention calls had churned.
Whereas, customers who received no retention calls were much more likely to churn.



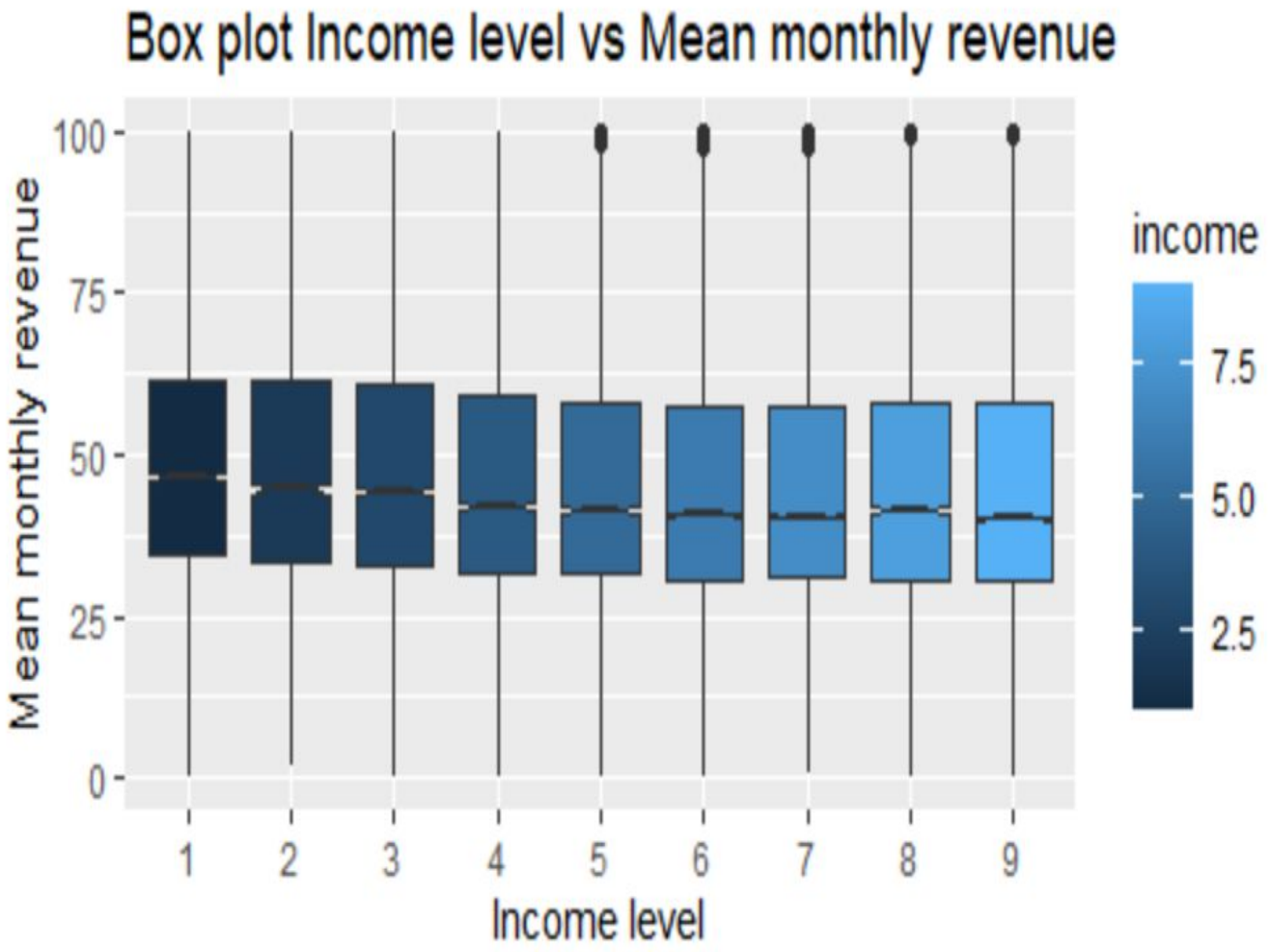
Customers who churned had much longer duration since they received a retention call. Customers in churn level 1 had lesser duration since their last retention call..



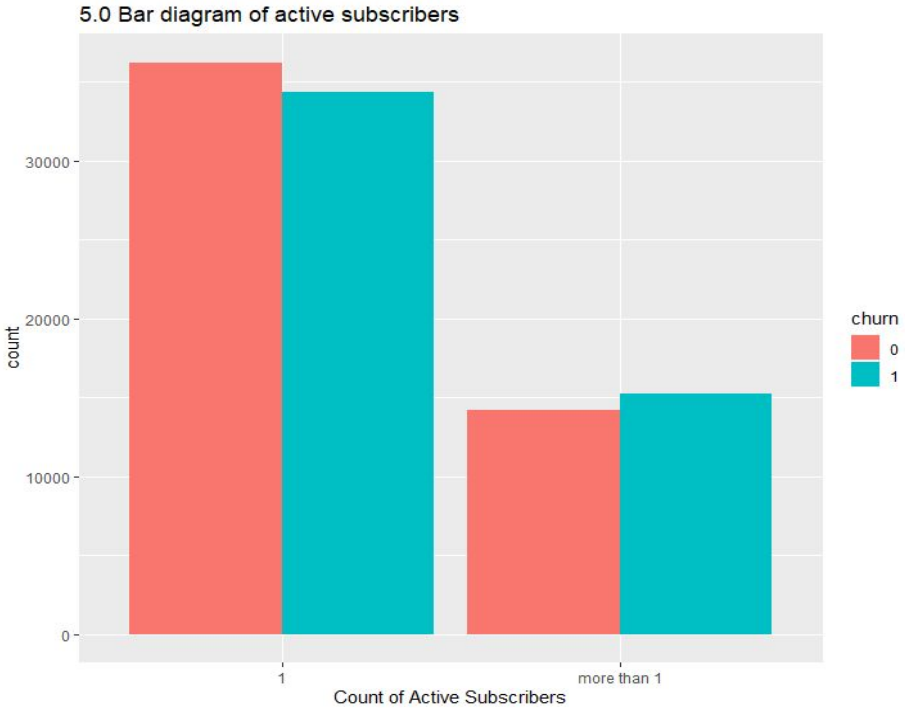
As the number of unique subscribers in a house increased, they were more likely to be in churn level 1.



Total revenue generated was highest from customers who were using the service for 50 months.
Revenue generated increased from 0 to 50 months of service then it steeply declined upto 60 months of service.



Mean monthly revenue decreased as the income level increased.



Active subscribers were categorised into 2 groups.
If there was only 1 active subscribers, then they were more likely to churn.
If there were more than 1 active subscribers then they were more likely to be in churn level 1

Important Insights

Top factors influencing likelihood of churn at the company:

1) Age of the current equipment: Customers with higher equipment age are less likely to churn.
2) Retention calls: Customers with no retention calls were more likely to churn.
3) As the number of subscribers in a household increased, customer was less likely to churn.

However, Cost and Billings and Network Service and Quality has no significant impact on churn rate.