Segmentation & Profiling Project

Merrimack College

Data Exploration

Amol Gote

(Data science student)

July 3rd, 2020

**Executive Summary**

This reports provides the customer segmentation for customers on telecommunications company which will help the organization in making effective economical customer retention strategies. Based on the data analytics exercise customers have been categorized on four categories Platinum, Gold, Silver and Bronze. These four categories have been primarily driven by the revenue generated by the customers over the tenure of their association with Telecommunications Company. Key value factor is revenue generated over the tenure and average monthly revenue generated. Apart from the value generation factors other attributes related to demographics line age, gender, education, employment length, marital status and household size have been factored in. Financial features like household income, debt to income ratio, total debt has been factored in as well. Distribution categories is as shown below

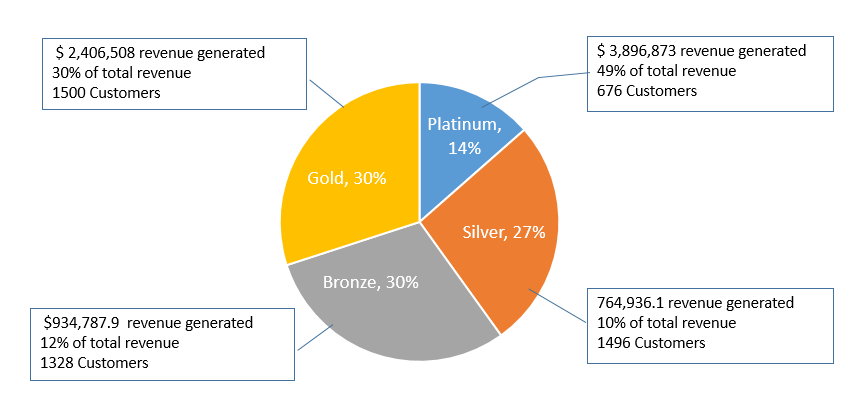


Figure 1: Customer segmentation distribution

Platinum customers who count for only 14% of the total customer base generate almost 50% of the total revenue, followed by gold with 30% of total revenue, silver and bronze generate very minimal revenue with over all percentage hovering around 10-12%. Average monthly revenue generated by platinum customers is $103, by gold customer is $27.50, for silver and bronze it is almost same which is $24.

As far as demographics is concerned, for platinum customer is dominated by customers aged in between 40 to 60, with average age of 53, gold customers are dominated by the customers with age greater than 60, with average age of 63, while silver and bronze is dominated by customers aged between 18 to 40 years of age, with average age of 35 and 37 respectively. Customers below age of 18 years are either in silver of bronze category. Gender distribution is equal across all customer segments, there is no bias. For platinum marital status is not of significance as there are nearly equal married and unmarried customers, with married ones on higher side a bit. Silver customer segment comprises of married customer and with very less unmarried customers. Bronze customer hardly contains any married customers. Median household size for platinum and gold customers is 2, while that for silver is 4, for bronze household size is hovers around 1 which is self-evident from the fact that majority of bronze customer is dominated by unmarried customers.

For household income average income for platinum is $90,000, for gold customers it is $65,000 for bronze and silver the average household income is around $40,000. Debt to income ratio for platinum, gold and silver customers is around 10% while that for bronze customers it shoots up to 37%. Nearly all the customers in the platinum segment are multiline customers, while the 65% of the gold customer are multiline, silver and bronze customers are dominated by non-multiline customers. Platinum and gold customer have been associated with an average of 56 months, followed by silver for 28 months and bronze by 20 months.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| # of Customers | 676  14% | 1500  30% | 1496  12% | 1328  10% |
| $ Revenue | $3,896,873  49% | $2,406,508  30% | $934,788  12% | $764,936  10% |
| $ Revenue (Average Monthly) | $103 | $28 | $25 | $24 |
| House hold Income (Avg) | $90,000 | $65,000 | $41,000 | $40,000 |
| Household size (Avg) | 2 | 2 | 4 | 1 |
| Marital status |  |  |  |  |
| Education (Avg) | 17 Years | 13 Year | 14 Years | 14 years |
| Age (Avg) | 54 years | 65 years | 35 years | 37 years |
| Association with company (Avg) | 56 months | 57 months | 28 months | 20 months |
| Debt to income ratio | 10% | 10.5% | 9.5% | 37% |
| Gender | Male: 355  Female: 321 | Male: 752  Female: 748 | Male: 577  Female: 751 | Male: 805  Female: 691 |
| Multiline |  |  |  |  |
| Loan Default |  |  |  |  |
| Car Ownership |  |  |  |  |

# Bibliography

Redman, T. (2018, April 02). *If Your Data Is Bad, Your Machine Learning Tools Are Useless.*

Retrieved from Harvard business review: <https://hbr.org/2018/04/if-your-data-is-bad-your-machine-learning-tools-are-useless>

*Definition: What is predictive modeling?* Retrieved from MicroStrategy website:

<https://www.microstrategy.com/us/resources/introductory-guides/predictive-modeling-the-only-guide-you-need>

*Data Matching.* Retrieved from Microsoft docs: [https://docs.microsoft.com/en-us/sql/data-](https://docs.microsoft.com/en-us/sql/data-quality-services/data-matching?view=sql-server-ver15)

[quality-services/data-matching?view=sql-server-ver15](https://docs.microsoft.com/en-us/sql/data-quality-services/data-matching?view=sql-server-ver15)

Sarfin, R (2019, Aug 7). *Data Quality Dimensions: How Do You Measure Up? (+*

*Downloadable Scorecard).* Retrieved from syncsort: [https://blog.syncsort.com/](https://blog.syncsort.com/2019/08/data-quality/data-quality-dimensions-measure/)

[2019/08/data-quality/data-quality-dimensions-measure/](https://blog.syncsort.com/2019/08/data-quality/data-quality-dimensions-measure/)

Swalin, A. (2018, Jan 30). How to Handle Missing Data. Retrieved from towards data science:

<https://towardsdatascience.com/how-to-handle-missing-data-8646b18db0d4>