

ADSC1910 Case 3

Neha Malhan(T00757554), Thiwanka Abeyweera(T00754871)

2024-09-29

Introduction

This report provides a deep analysis of a sample of approximately 500 customers, drawn from the provided customer dataset, to provide the telecommunications company with deeper insight into customer behavior, service quality, and pricing strategies. By understanding customer behavior across these critical factors, the analysis aims to support informed decision-making that enhances customer satisfaction and drives revenue growth.

The analysis is performed on the following attributes of the sample drawn from “Customers.csv”:

ID: A unique identification number for each customer.

Region: The geographical region within British Columbia where the customer resides.

Bundle: Whether the customer bundles their telecommunications services with other services (e.g., internet or television).

Fees: The monthly fee paid for telecommunications services.

Months: The length of time (in months) that the customer has held a contract with the company.

Methods

Two key methods used in this study are stratified random sampling and descriptive statistics to analyze the customer data and reveal key insights.

Stratified random sampling was applied by grouping customers based on *region* and *bundling behavior*, ensuring that different customer segments were proportionally represented in the selected sample while minimizing the sampling error. Stratified sampling is well-suited for this analysis as it provides better coverage of the population by ensuring all key subgroups have been represented in the sampling, improving accuracy and providing comprehensive insights into customer behavior across regions.

After sampling, **descriptive statistics** were used to examine the distribution of key variables such as *monthly fees* and *contract duration*. The mean was calculated to understand average customer spending and retention patterns. Additionally, variance was used to assess the diversity in customer fees and contract lengths. Descriptive analysis complements this by offering a clear summary of the customer base, helping the telecommunications company make informed decisions on pricing, retention, and bundling strategies.

Findings & Results

1) Regional differences in customer behavior

Table 1: Customer behaviour by region

Region	Customer Count	Total Fees	Average Monthly Fees	Variance in Fees	Bundling Rate
Cariboo	29	\$973	\$34	\$148	45%
Kootenay	48	\$1608	\$34	\$128	44%
Mainland-Southwest	220	\$7582	\$34	\$139	45%
Nechako	4	\$203	\$51	\$45	50%
North Coast	14	\$478	\$34	\$56	43%
Northeast	8	\$326	\$41	\$53	50%
Thompson-Okanagan	124	\$4183	\$34	\$137	44%
Vancouver Island-Coast	51	\$1763	\$35	\$140	45%

Table 1 reveals that most regions have similar average fees of \$34, with bundling rates ranging from 43% to 50%. The bundling rate refers to the percentage of customers who choose to purchase multiple services as a bundle rather than subscribing to individual services separately.

The **Mainland-Southwest** region has the largest customer base, which contributes significantly to overall revenue; **Thompson-Okanagan** is also a key contributor to total fees collected. Both **Kootenay** and **Vancouver Island-Coast** have high levels of customer engagement, with Vancouver Island-Coast having a slightly higher average fee than Kootenay but higher total revenue. **Cariboo** has high customer engagement but contributes less to the total revenue than these key contributors. The variance in fees among these regions is moderate, indicating that while most customers pay close to the average fee, there are some variations.

In contrast, **Nechako** has the fewest customers but charges higher average fees, reflecting a focus on a premium market. Its lower variance suggests that customers in this region pay more consistently. The **Northeast** region also has fewer customers and higher average fees, contributing to its premium market positioning. Finally, **North Coast** has a similar average fee but a larger customer count compared to Nechako and Northeast.

2) Effects of bundling on revenue and retention

Table 2: Summary of revenue & retention by bundling behavior

Bundle	Customer Count	Total Fees	Average Fees	Variance in Fees	Average Months	Variance in Months
No	275	\$9396	\$34	\$137	49	746
Yes	223	\$7721	\$35	\$134	47	777

Table 2 demonstrates that **non-bundled customers** have a greater customer count than **bundled customers**, resulting in a higher total fee for non-bundled customers. Non-bundled customers pay an average monthly price of around \$34, with a high variance of 137, showing significant fluctuation in individual fees. Bundled customers, on the other hand, pay an average price of \$35 with slightly lower variance of 134, indicating that while they experience variation, it is less apparent than that of non-bundled customers.

Regarding subscription duration, non-bundled customers typically remain subscribed for an average of 49 months, with a variance of 746 reflecting a wide range in individual subscription lengths. Bundled customers

have a lower average subscription duration of 47 months and a variance of 777, indicating a similar level of variability in how long they stay subscribed.

Discussion & Conclusion

The Mainland-Southwest region has the highest customer base and overall revenue. Following that, Thompson-Okanagan and Vancouver Island-Coast are significant contributors, with Vancouver Island-Coast marginally exceeding Kootenay in terms of average fees and overall revenue. Cariboo serves a huge number of customers, although its contribution to total revenue is very low. In contrast, Nechako distinguishes out due to its higher average costs while having the fewest customers, indicating its focus on a premium market. The Northeast region has a comparable premium positioning, with fewer customers and higher fees. Finally, although having a similar average fee, North Coast has a greater customer base than Nechako and Northeast.

According to findings of Table 2, bundled customers pay slightly higher fees on average but have shorter contract duration compared to non-bundled customers. Non-bundled customers contribute more to overall revenue due to their higher customer count. Both groups show similar variability in their fees and contract duration.

Recommendations

To enhance customer retention and optimize revenue, the company should offer targeted discounts and bundled packages in high-revenue regions like Mainland-Southwest and Thompson-Okanagan to encourage longer contracts and higher spending. Additionally, the company should adjust pricing strategies to reflect the regional differences in spending and bundling behaviors, ensuring competitive pricing in areas like Vancouver Island-Coast and Kootenay. Promoting bundled services across all regions will increase customer satisfaction and revenue. To enhance customer retention and revenue in regions with lower customer counts, such as Nechako, North Coast, Northeast, and Cariboo, the company should conduct market research to understand local needs and preferences.

References

- 1) Hellingman, S. (2024). ADSC 1000: Statistical sampling
- 2) Timbers, T., Campbell, T., & Lee, M. (2024-08-21). Data Science: A First Introduction.
- 3) sprintf funtion

Appendices

This case study was collaboratively developed by:

- 1) Thiwanka Abeyweera: Prepared the Introduction, methods, analyzed the regional differences in customer behaviour
- 2) Neha Malhan: Analyzed the effects of bundling on revenue and retention, prepared conclusion and recommendation

```

knitr::opts_chunk$set(echo = TRUE)
# Importing the dat

library(dplyr)

# Load the dataset from the working directory
customer_data <- read.csv("Customers.csv")

# Check the structure of the dataset
str(customer_data)

# Stratified sampling

# Set a random seed for reproducibility
set.seed(100)
data("customer_data")

# Perform stratified random sampling based on Region and Bundle
customer_sample <- customer_data %>%
  group_by(Region, Bundle) %>%
  sample_frac(size = 500/ nrow(customer_data))
# View the sampled data
customer_sample

library(knitr)

# Calculate regional differences in bundling and monthly fees
regional_fees <- customer_sample %>%
  group_by(Region) %>%
  summarize(
    customer_count = n(),
    sum_fees = sum(Fees),
    avg_fees = mean(Fees),
    variance_fees = var(Fees),
    bundle_rate = mean(Bundle == "Yes") * 100,
  )

regional_fees <- regional_fees %>%
  mutate(sum_fees = sprintf("%.f", sum_fees),
         avg_fees = sprintf("%.f", avg_fees),
         variance_fees = sprintf("%.f", variance_fees),
         bundle_rate = sprintf("%.f%%", bundle_rate)) # format bundle rate as percentages

# View regional differences
kable(regional_fees, caption = "Customer behaviour by region",
      col.names = c("Region", "Customer Count", "Total Fees", "Average Monthly Fees",
                    "Variance in Fees", "Bundling Rate"))

library(knitr)

```

```

# Calculate descriptive statistics on Monthly Fees and Contract Duration
summary_bybundle <- customer_sample %>%
  group_by(Bundle) %>%
  summarize(
    customer_count = n(),
    sum_fees = sum(Fees),
    avg_fees = mean(Fees),
    variance_fees = var(Fees),
    mean_months = mean(Months),
    variance_months = var(Months)
  )

summary_bybundle <- summary_bybundle %>%
  mutate(sum_fees = sprintf("$%.f", sum_fees),
         avg_fees = sprintf("$%.f", avg_fees),           # Format avg fees as currency
         variance_fees = sprintf("%.f", variance_fees),
         mean_months = sprintf("%.f", mean_months),
         variance_months = sprintf("%.f", variance_months))

# Display summary_fees as a table
kable(summary_bybundle,
      caption = "Summary of revenue & retention by bundling behavior",
      col.names = c("Bundle", "Customer Count", "Total Fees", "Average Fees",
                    "Variance in Fees", "Average Months", "Variance in Months"))

```