Marketing Data Manual

Group Assignment - Open Source Programming 2022

By Lucero Fabrizio, Ramírez Luisa, Quiñones Doménica

déc. 22, 2022

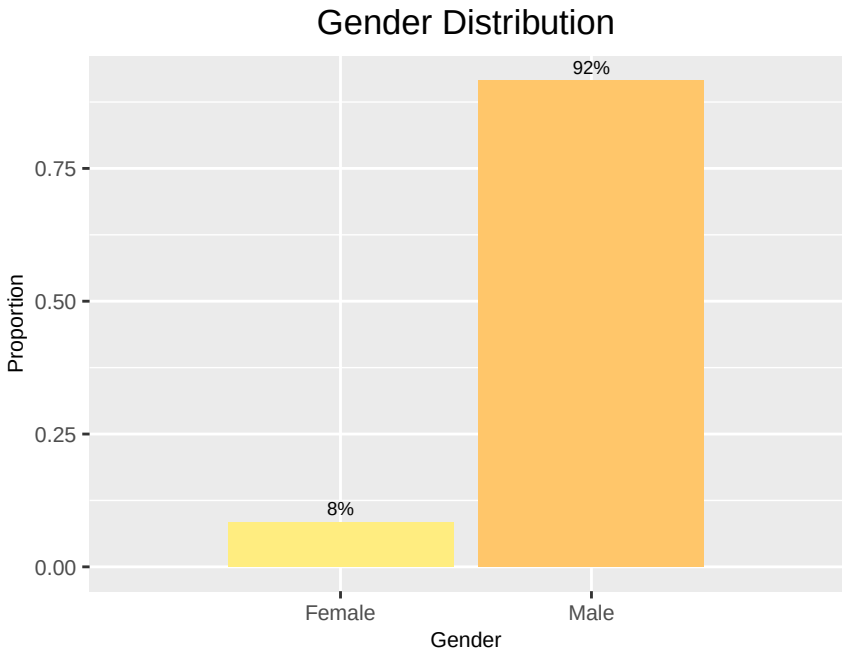
The manual provides the summary statistics and the detailed description of the gamblers' metrics.

These datasets are retrieved from an internet gambling research project including the Division on Addictions (DOA) and bwin Interactive Entertainment, AG (bwin), an Internet betting service provider headquartered in Vienna, Austria. The information provides evidence from the first eight months of the Internet sports betting behavior study that took place from February 1 to September 30, 2005.

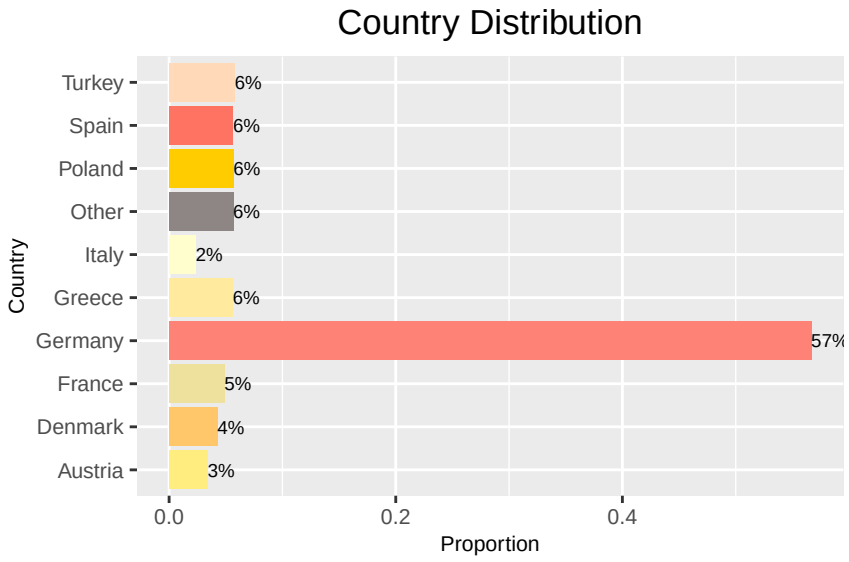
Summary statistics

The dataset used for these report consists of 24 variables, 12 demographics and 12 marketing insights, and 42620 rows or unique customers.

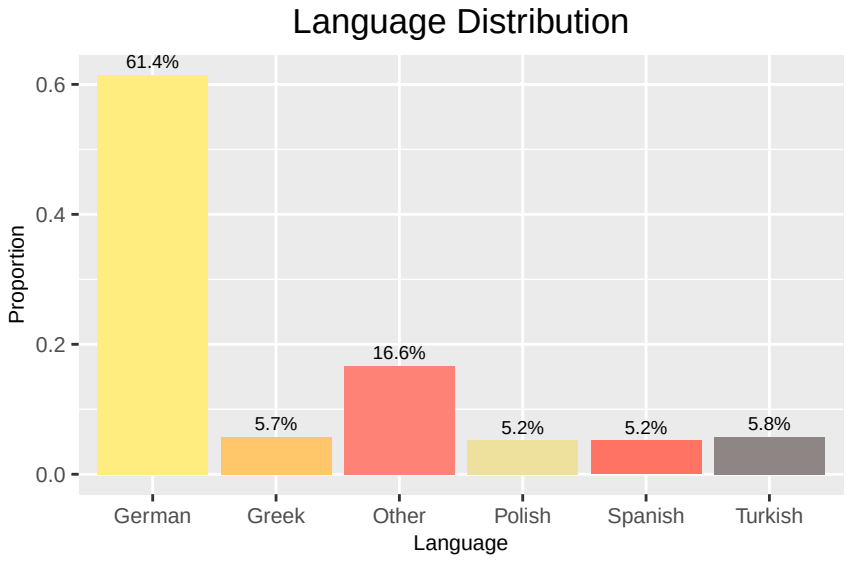
The customers are distributed by gender as follows:



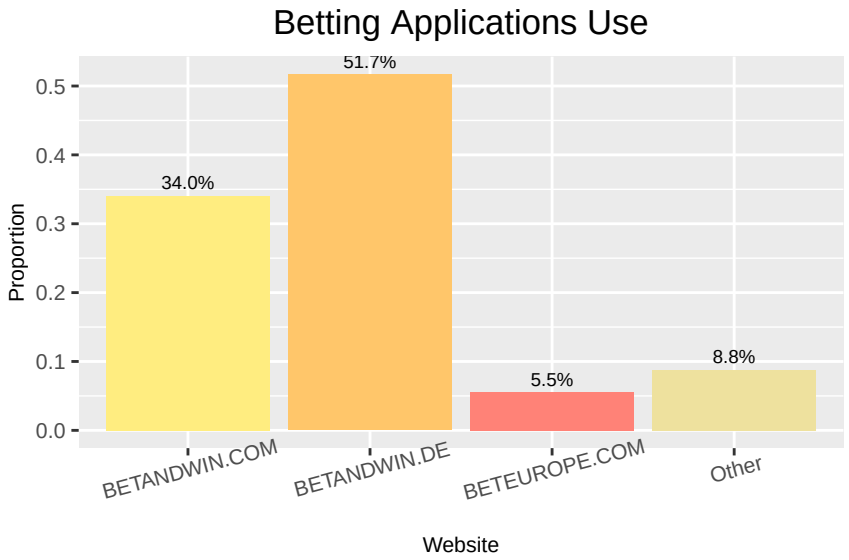
The data presents information of the residence of the customers, which includes 81 countries. The following chart shows the top 5 frequent countries, all other countries are classified as 'Other'.



The customers speak 17 languages, including: German, Greek, English, Turkish, Danish, Portuguese, Polish, Norwegian, Spanish, French, Italian, Catalan, Swedish, Hungarian, Czech, Russian, Dutch. The top 5 spoken languages are presented below, the rest are classified as "Other":



The three websites with most frequent usage represent more than 90% of the applications:

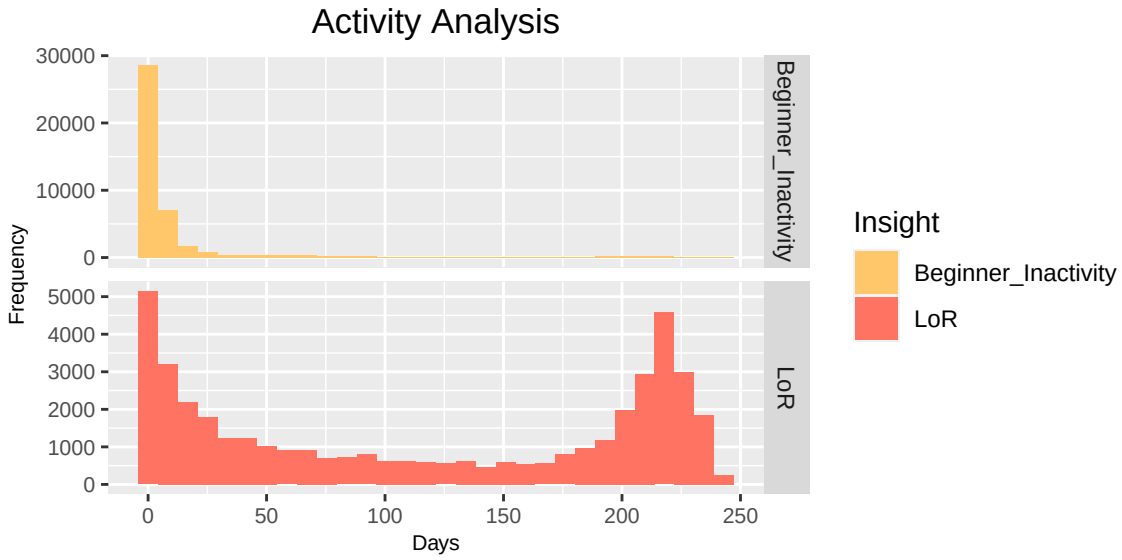


Description of the gamblers' metrics

Customer Participation

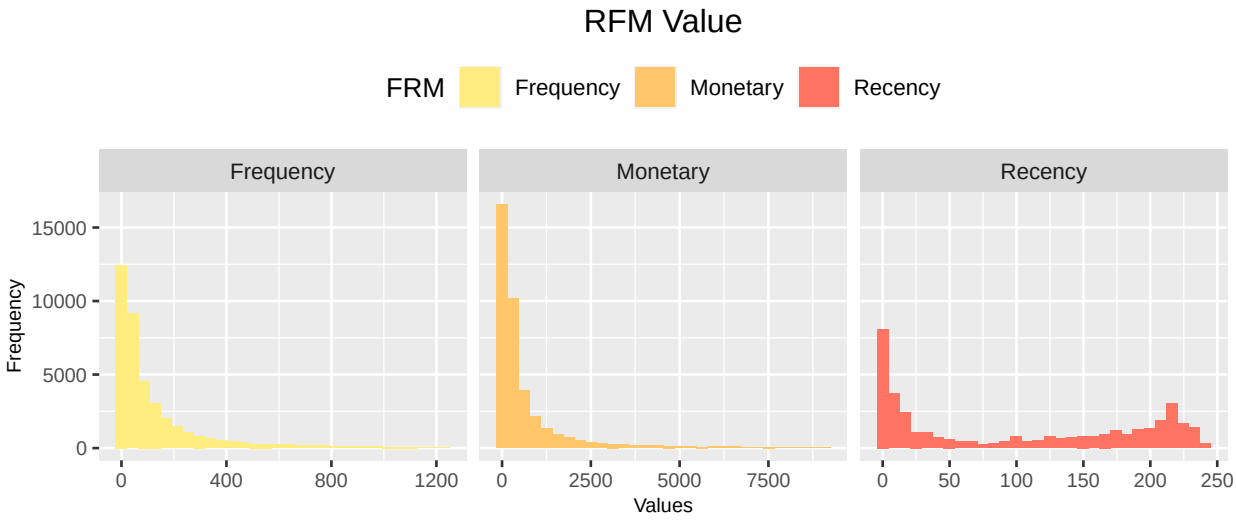
LoR Metrics The Length of Relationship measures the time a client has been engaged in the period of study. It is obtained finding the amount of days between the last and the first day of activity. The plot shows that the majority of the customers have a LOR between 0 and 25 days and 200 and 240 days.

Beginners Inactivity Is the amount of days between registration and first activity. It measures the time needed by each customer to start playing on the websites.



Recency, Frequency and Monetary Value

The RFM Value analyzes the consumers based on its habits. Recency determines how long ago the customer made the last purchase or bet (measured in days), Frequency evaluates how often they bet (measured in amount of bets in a determined period) and Monetary, the amount of money spent on bets. The chart has information of the quantile .95 and shows the general patterns of the bet consumers:



Customer Lifetime Value

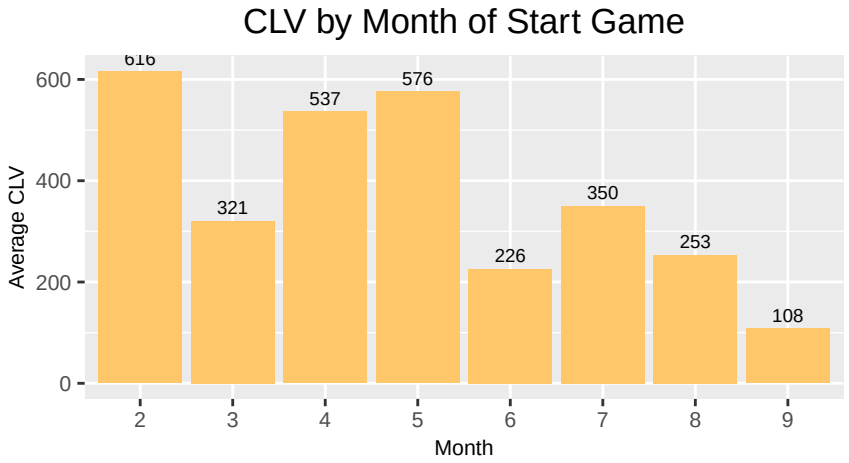
The CLV is the total revenue each customer brings to the company over the period of relationship. It is calculated as follows:

$$CLV = \frac{GrossMargin}{RetentionRate}$$

Where,

Gross Margin = Total Bets Revenue – Total Expenses

Retention Rate = Number of users that stayed betting after Sept 2005 / Total number of users

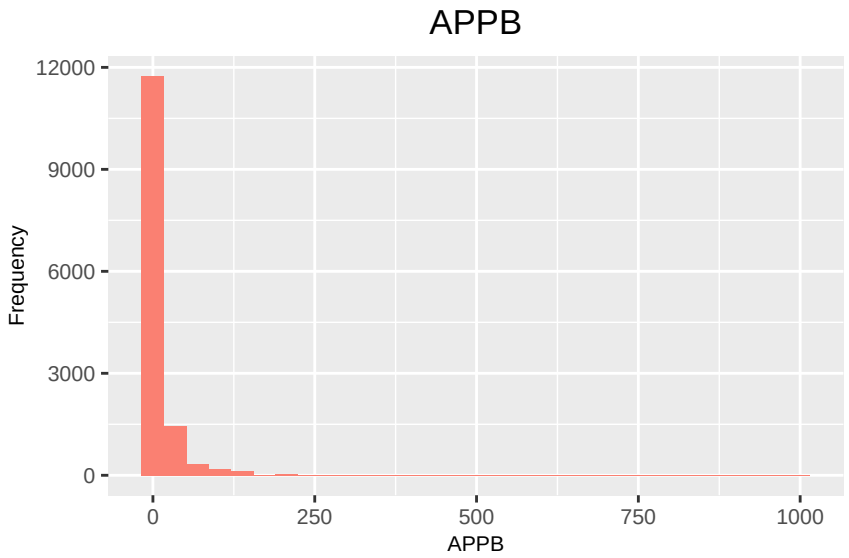


Average Profit Per Bet

APPB determines the revenue the company is getting for each bet the customer makes.

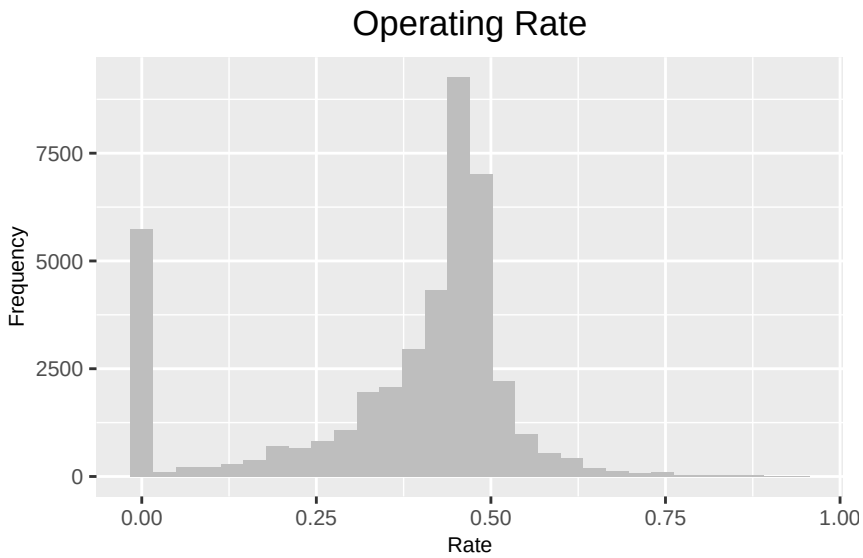
$$APPB = \frac{GrossMargin}{AmountofBets}$$

The following plot shows the distribution of the customers by APPB.



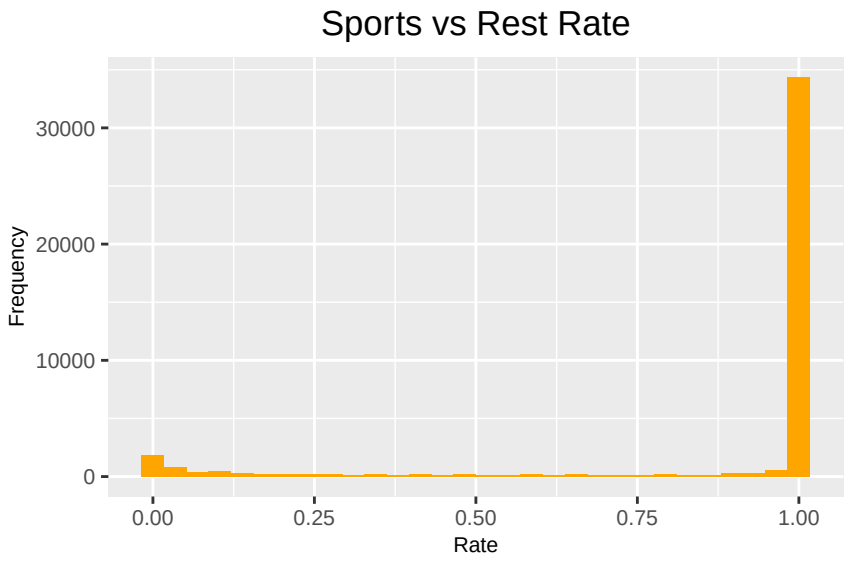
Operating Rate

The Operating Rate determines the ability of a business to employ its resources by comparing the expenses of a company to net sales.



Sports vs Rest Rate

Since Sports bets are more popular than all of the other games, the SvR Rate compare the users by this metric.

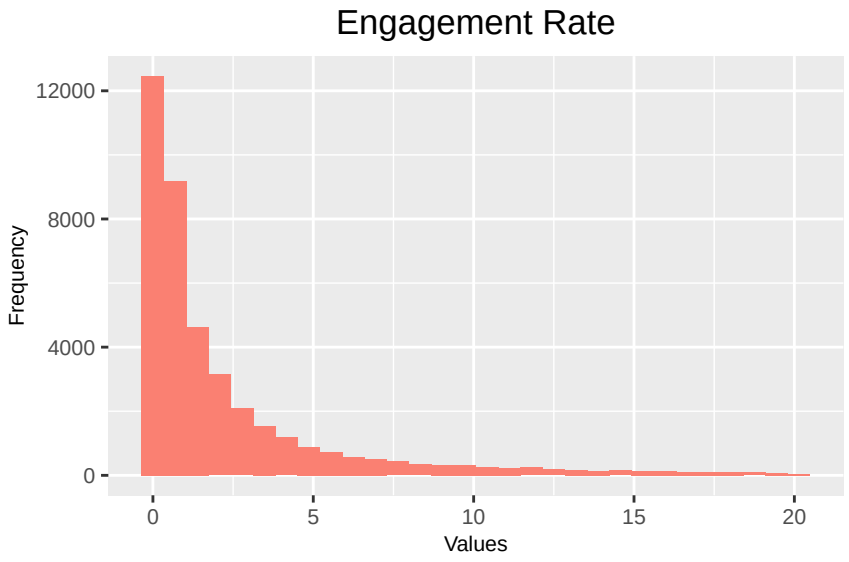


Engagement Rate

The Engagement Rate determine how engaged the users are are with the brand, which is useful for online businesses as the betting websites.

$$EngagementRate = \frac{TotalRevenue}{MedianRevenue}$$

The Engagement Rate distribution of the customers of the dataset is the following:



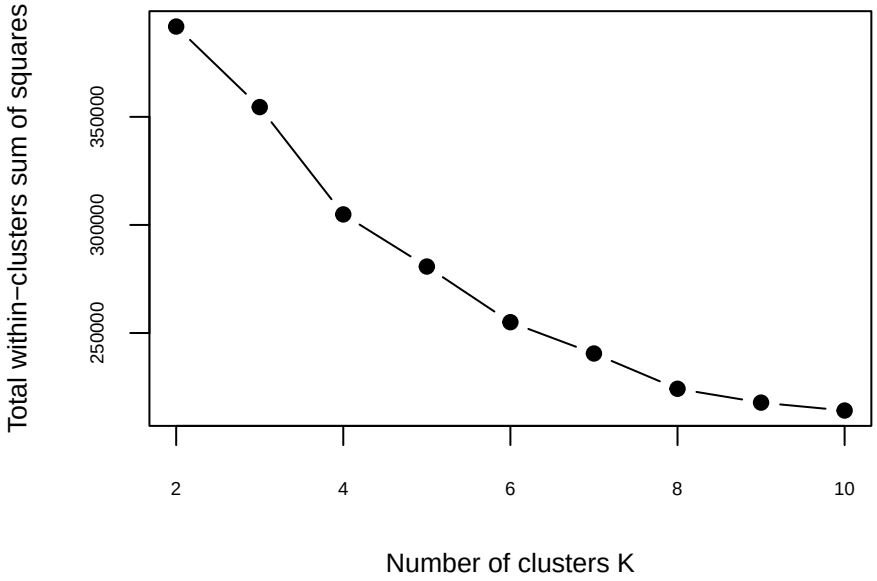
Clusters

As an additional insight, a cluster analysis was developed with the purpose of identifying the customers. All of the insights mentioned before were used: Beginner_Inactivity, Recency, Frequency, Monetary, Length of Relationship, Customer Length Value, Average Profit Per Bet, Operating Rate, Engagement Rate and SvR Rate.

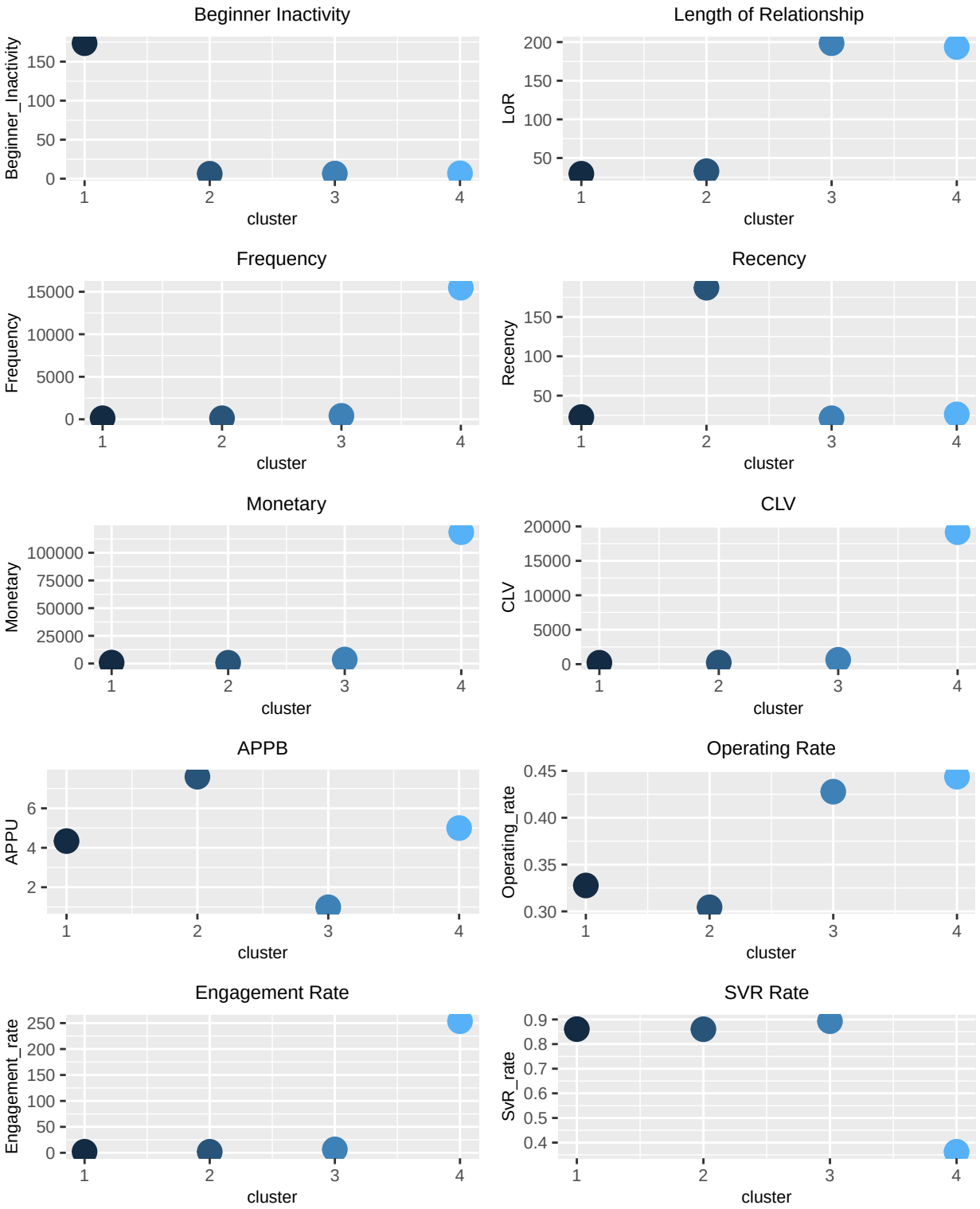
First the standardization process was applied with the purpose of putting all the variables on the same scale to be able to compare the scores between the different types of variables.

The technique used was K-means, which is each observation of the dataset is placed into one of K clusters.

The second step is to choose a value for K through the elbow method, which identifies where a bend appears in the plot.



Finally, the kmeans() function is used to perform k-means clustering on the dataset using K = 4:



Sources

Datasets Division on Addictions, Cambridge Health Alliance, a teaching affiliate of Harvard Medical School PI(s): Dr. Howard J. Shaffer, Sponsor(s): bwin, Interactive Entertainment, AG

LaBrie, R. A., LaPlante, D. A., Nelson, S. E., Schumann, A., & Shaffer, H. J. (2007). Assessing the playing field: A prospective longitudinal study of Internet sports gambling behavior. Journal of Gambling Studies, 23, 347-362.

ggplot configuration https://www.geeksforgeeks.org/change-color-of-bars-in-barchart-using-ggplot2-in-r/

Variable report https://community.rstudio.com/t/how-to-include-formatted-list-of-variable-definitions/8255/2

Customer Lifetime Value https://www.leadsquared.com/learn/sales/clv-customer-lifetime-value/

Engagement rate https://clevertap.com/blog/how-to-measure-customer-loyalty/

Cluster https://www.statology.org/elbow-method-in-r/#:~:text=One%20of%20the%20most%20common%20ways%20to%20

RFM Value https://www.investopedia.com/terms/r/rfm-recency-frequency-monetary-value.asp

APPU https://supersonic.com/learn/blog/the-3-most-important-kpis-for-testing-your-hyper-casual-prototype/#:~:text=APPU%2C%20which%20stands%20for%20average,that%20lead%20to%20greater%20profitability.

Formats https://www.montana.edu/rotella/documents/502/MarkdownEqnExamples.Rmd https://ggplot2.tidyverse.org/reference/geom_point.html

Operating rate https://www.investopedia.com/terms/o/operatingratio.asp