Section of the conference: Economic Sciences | Sekcja konferencji: Nauki ekonomiczne

How to cite: Sanin-Villa, D., Hernandez, C. M, & Martinez, A. F. (2024). Analyzing Travel Service Costs and Customer Behavior Through ANOVA and Chi-Square Tests. *World conference on future innovations and sustainable solutions*. Futurity Research Publishing. https://doi.org/10.5281/zenodo.14034082

Analyzing Travel Service Costs and Customer Behavior Through ANOVA and Chi-Square Tests

Daniel Sanin-Villa[1], Cristian Mateo Hernandez [2], Adrian Felipe Martinez[2]

[1]Ph.D., Professor, Universidad Pontificia Bolivariana (UPB), Medellín, Colombia, https://orcid.org/0000-0001-6853-340X

[2]M.Sc., Professor, Instituto Tecnológico Metropolitano (ITM), Medellín, Colombia

Accepted: October 17, 2024 | **Published**: November 1, 2024 | **Language:** English

Abstract: This work analyzes the impact of variables such as loyalty status, number of travelers, average rating, and expected travel duration on customer behavior using statistical techniques. Descriptive analysis, correlation tests, and ANOVA were employed to identify patterns, associations, and significant differences across customer groups (Gold, Regular, Silver). The findings reveal statistically significant differences in customer behavior based on loyalty status, suggesting that these variables influence service experience and customer retention. The results offer insights for tailoring customer management strategies to optimize loyalty and service.

Keywords: Customer behavior, loyalty status, ANOVA, correlation, service management.

Introduction

Analyzing customer behavior through quantitative measures is critical in understanding the factors influencing service experience and loyalty. Previous research has consistently highlighted the role of various customer attributes in shaping behavioral patterns, including loyalty status,

service ratings, and travel expectations. For instance, studies have found significant correlations between loyalty status and service satisfaction levels, where customers with higher loyalty statuses tend to report more consistent satisfaction due to tailored service offerings (Dam & Dam, 2021; Souza et al., 2023). These relationships suggest that customer segmentation based on loyalty status can significantly impact perceptions of service quality. Correlation analysis, a statistical method used to measure the strength and direction of associations between variables, has been widely employed to investigate these behavioral patterns. Research by (Kim & Kim, 2022; Pio Leonardo Cavaliere et al., 2021) demonstrated that service rating scores often correlate with customer expectations and retention rates, indicating that higher ratings are linked to greater customer loyalty. Additionally, the use of Analysis of Variance (ANOVA) in prior studies has revealed significant differences in customer behavior across segments, particularly in terms of service usage and travel frequency (Kevin J. Zeng, Irina Y. Yu, Morgan X. Yang, 2022; Tripathy & Sharaff, 2023). ANOVA helps to identify whether observed differences between groups are statistically significant, providing a foundation for further exploration of behavioral variations among customer categories such as Gold, Regular, and Silver. Moreover, the Chi-Squared test has been utilized to assess the independence between categorical variables, such as loyalty status and travel-related behaviors. For example in (Dis Chonsalasin, S Jomnonkwao, K. Chanpariyavatevong, W Laphrom, 2022; Zikienė et al., 2024), applied the Chi-Squared test to examine associations between customer loyalty programs and travel frequency, finding significant dependencies that suggest loyalty programs effectively influence customer travel habits. The statistical techniques employed have effectively clarified complex relationships in customer behavior, providing insights into how certain attributes influence service expectations and engagement. However, there is a noted gap in research on the combined use of correlation analysis, ANOVA, and Chi-Square tests to explore the interconnections among variables like loyalty status, traveler count, average ratings, and trip duration. The current study aims to address this gap by employing a comprehensive statistical approach to analyze these variables to uncover patterns and relationships that can inform strategies for service improvement and customer retention. The scientific interest in this study lies in its potential to provide a more nuanced understanding of customer behavior through a multi-method statistical analysis, which could guide the development of more effective management practices in service-oriented industries. The dataset used in this study is sourced from a publicly available collection titled "Dynamic Pricing Dataset" on Kaggle (https://www.kaggle.com/datasets/arashnic/dynamic-pricing-dataset), licensed under CC0: Public Domain. This dataset contains historical ride information from a ride-sharing company seeking to implement a dynamic pricing strategy to optimize fares based on real-time market conditions. The dataset includes features such as the number of riders, number of drivers, location category, customer loyalty status, average ratings, time of booking, vehicle type, expected ride duration, and historical cost of rides.

Research Aim and Research Questions

This research analyzes customer behavior and service performance within a travel service to identify patterns and relationships that can inform strategic improvements. By examining variables such as customer loyalty status, booking times, vehicle preferences, number of travelers, customer satisfaction ratings, expected ride durations, and historical ride costs, the study seeks to uncover factors influencing customer satisfaction and operational efficiency. Understanding these factors is essential to address the challenges of optimizing service offerings and enhancing customer loyalty. Recognizing the interplay between different service aspects and customer experiences can lead to more effective strategies that improve overall service quality and competitiveness in the travel industry. This research question appears: How does the expected ride duration correlate with the historical cost of rides in the travel service, and what implications does this relationship have for pricing strategies and service optimization?

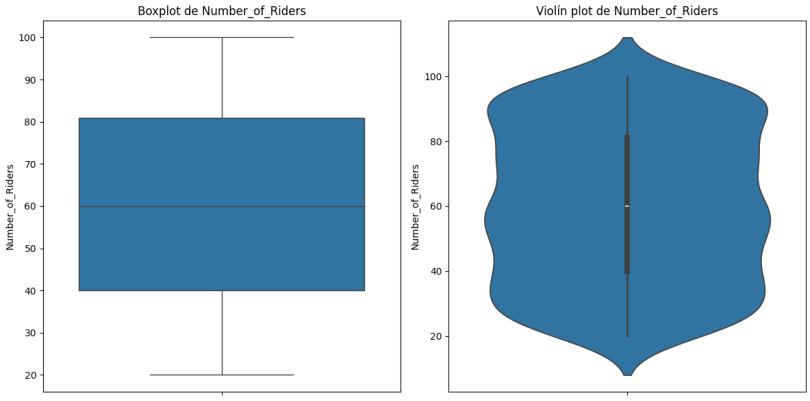
Research Results

The analysis of travel service data showed a balanced distribution of use across urban (346 observations), rural (332), and suburban areas (322), with urban regions leading, likely due to higher demand. Loyalty levels were dominated by "Silver" status users (367), followed by "Regular" (320) and "Gold" (313), indicating frequent ridership but also opportunities for enhancing loyalty programs to encourage more users toward higher tiers. Reservations peaked at night with 276 bookings, compared to 247 in the afternoon, 246 in the morning, and 231 in the early afternoon, suggesting a strong demand for travel outside typical work hours, possibly for personal activities. "Premium" vehicles had a slight edge in preference, with 522 bookings compared to 478 for "Economy," reflecting a preference for quality and comfort, likely shaped by demographics or travel needs. Numerical data showed an average of 60.37 travelers per trip with a standard deviation of 23.70, indicating varied group sizes. The average trip rating was 4.26 out of 5, with a standard deviation of 0.44 and a minimum of 3.5, suggesting consistent high satisfaction across customers. Trip durations averaged 99.59 minutes (SD=49.17), with times ranging from 10 to 180 minutes, accommodating both short and extended trips. The historical trip cost averaged 372.50 monetary units, with a standard deviation of 187.16 and prices spanning 25.99 to 836.12 units, reflecting pricing variability based on factors like duration and vehicle type. Distribution graphs, including boxplots and violin plots, showed a bimodal pattern in traveler numbers, suggesting distinct segments in group size preferences that may benefit from tailored services. This suggests that the service offers options for different budgets and that cost may be influenced by factors such as trip duration, vehicle type, or location. The distribution graphs presented in [Figure 1] using boxplots and violin plots of the number of travelers showed a relatively uniform distribution, with a median of around 60 and some extreme values close to 100. The violin plot indicated a smooth bimodal distribution, suggesting the existence of groups that prefer to travel in smaller or larger numbers. This implies considerable variability in group sizes, possibly reflecting different types of customers or travel purposes. The graphical analysis of

average ratings shows a consistent concentration around 4.3, with minimal variability and a slight skew towards higher values, suggesting generally positive user satisfaction.

Figure 1

Examplo of relatively even distribution of travelers

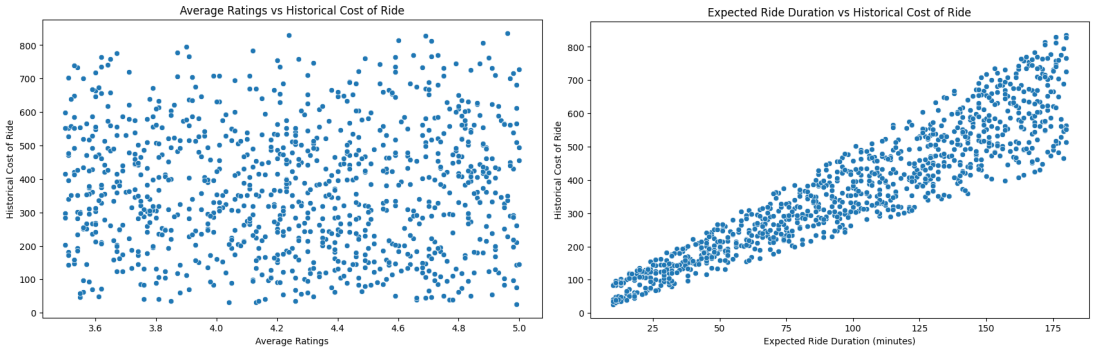


Source: authors own development

The violin plot of expected trip durations reveals a wide range, centered around 100 minutes but with a long tail, indicating that while most trips are moderate in length, some are significantly longer. Similarly, trip cost displays an asymmetric distribution with a concentration of around 350 monetary units and a tail extending toward higher costs, suggesting the presence of high-cost trips. Correlation and regression analyses presented in [Figure 2] showed no clear relationship between the number of travelers and trip cost, nor between average rating and trip cost, suggesting that these factors do not directly influence expenses. However, a significant positive relationship was found between expected trip duration and cost, indicating that longer trips tend to be more expensive, aligning with the intuitive expectation that duration impacts cost. This trend, despite some variability, underscores duration as a key cost determinant.

Figure 2

Distribution correlation between variables



Source: authors own development

The Pearson correlation analysis found a strong positive correlation ($r = 0.9275$, $p < 0.05$) between expected trip duration and historical cost, confirming a significant linear relationship between these variables. In the contingency analysis, Chi-Square tests revealed no significant relationship between loyalty status and reservation time (p-value = 0.5894), indicating that reservation timing is independent of customer loyalty level. "Gold" and "Silver" customers showed a slight preference for nighttime reservations, whereas "Regular" customers exhibited a more uniform distribution. Similarly, no significant association was found between loyalty status and vehicle preference ("Economy" or "Premium"), with a p-value of 0.4744, suggesting that loyalty status does not influence vehicle type choice. Further, there was no marked preference for vehicle type based on reservation time (p-value = 0.3206), implying that customers choose between "Economy" and "Premium" vehicles independently of when they book. ANOVA results showed no significant differences among loyalty statuses ("Gold," "Regular," and "Silver") for the number of travelers, average rating, or expected trip duration, as all p-values exceeded the 0.05 threshold. This consistency across loyalty groups suggests that customer loyalty status does not impact these operational metrics.

Conclusion

The uniform distribution of riders across loyalty levels suggests predictable demand, allowing efficient fleet management and resource allocation without significant adjustments for different customer groups. The high, consistent customer satisfaction across all loyalty tiers reflects stable service quality, though introducing exclusive perks for top-tier members could enhance perceived value and foster retention. ANOVA results indicate no significant differences in rider count, satisfaction, or expected trip duration by loyalty status, supporting a uniform approach to daily operations. However, further analysis with additional variables may reveal deeper insights to guide strategy development.

References:

Dam, S. M., & Dam, T. C. (2021). Relationships between Service Quality, Brand Image, Customer Satisfaction, and Customer Loyalty. *Journal of Asian Finance, Economics and Business*, *8*(3), 585–593. https://doi.org/10.13106/jafeb.2021.vol8.no3.0585

Dis Chonsalasin, S Jomnonkwao, K. Chanpariyavatevong, W Laphrom, V. R. (2022). Modeling of airline passenger loyalty: A comparison of leisure and business travelers. *Research in Transportation Business & Management*, *43*(100735). https://doi.org/https://doi.org/10.1016/j.rtbm.2021.100735.

Kevin J. Zeng, Irina Y. Yu, Morgan X. Yang, H. C. (2022). Communication strategies for multi-tier loyalty programs: The role of progress framing,. *Tourism Management*, *91*(1). https://doi.org/https://doi.org/10.1016/j.tourman.2021.104460.

Kim, Y. J., & Kim, H. S. (2022). The Impact of Hotel Customer Experience on Customer Satisfaction

through Online Reviews. *Sustainability (Switzerland)*, *14*(2), 1–13. https://doi.org/10.3390/su14020848

Pio Leonardo Cavaliere, L., Khan, R., Jainani, K., Sundram, S., Bagale, G., Kalyan Chakravarthi, M., Regin, R., & Suman Rajest, S. (2021). The Impact Of Customer Relationship Management On Customer Satisfaction And Retention: The Mediation Of Service Quality. *Turkish Journal of Physiotherapy and Rehabilitation*, *32*(3), 1–16. www.turkjphysiotherrehabil.org22107

Souza, A. A. de, Damacena, C., Araújo, C. F., Carvalho, B. P., & Seibt, N. G. (2023). The influence of customer participation and convenience on customer satisfaction: a multiple mediation. *Revista de Administração Da UFSM*, *16*(2), 5. https://doi.org/10.5902/1983465971688

Tripathy, G., & Sharaff, A. (2023). AEGA: enhanced feature selection based on ANOVA and extended genetic algorithm for online customer review analysis. *Journal of Supercomputing*, *79*(12), 13180–13209. https://doi.org/10.1007/s11227-023-05179-2

Zikienė, K., Kyguolienė, A., & Kisieliauskas, J. (2024). The Influence of Creative and Innovative Loyalty Programs' Features on Customers' Attitudinal, Conative, and Behavioural Loyalty. *Creativity Studies*, *17*(1), 254–273. https://doi.org/10.3846/cs.2024.20130