Ömer Faruk Durmuş, 7406414

Programming for Data Analytics - Programming Assignment

Seminar Paper

Supervisor: Prof. Dr. Markus Weinmann
Second Examiner: Sercan Demir

Submitted for the Master's examination in the study program Information Systems at the Faculty of Management, Economics and Social Sciences of the University of Cologne

October 31, 2024

Contents

List of Figures

List of Tables

1 Introduction

This report examines customer satisfaction levels within the Ryanair dataset, utilizing various analytical approaches to identify key factors influencing satisfaction across different customer demographics and service experiences. The goal is to understand the primary determinants of satisfaction, highlighting differences in customer expectations and perceptions based on travel class, age, and travel purpose. In the competitive airline industry, customer satisfaction is crucial as it influences both customer retention and brand loyalty. For a low-cost airline like Ryanair, where services are often more limited, insights from this analysis can guide targeted improvements in areas with the greatest impact on satisfaction.

Through data preprocessing, distribution analysis, correlation assessment, and multivariate regression, this report provides a comprehensive view of the Ryanair customer experience. The analytical process begins by addressing the dataset's structure and key variables of interest, followed by exploratory data analysis to assess distributions and relationships between variables. Correlations and statistical tests help identify significant trends and predictors, with a regression model further quantifying each factor's influence on customer satisfaction. This report is structured as follows: an overview of the dataset and key variables, a detailed analysis of distributions and relationships, a discussion of findings, and finally, a conclusion summarizing insights and offering directions for future analyses.

2 Dataset

2.1 General description of the data

The Ryanair dataset includes a range of customer characteristics, travel details and satisfaction-related metrics which together offer insights into passengers' experiences and overall satisfaction with Ryanair's services.

2.2 Variables of interest

The main variable of interest in this analysis is *satisfaction*, which serves as the dependent variable. This variable categorizes passenger satisfaction as "satisfied", "neutral" or "dissatisfied", summarizing overall satisfaction with the flight experience. Independent variables relevant to satisfaction analysis include:

- Service and comfort ratings: these cover specific aspects of inflight and ground experiences, such as *Inflight.wifi.service*, *Gate.location*, *Food.and.drink*, *Ease.of.Online.booking*, *Online.boarding*, *Seat.comfort*, *Inflight.entertainment*,

On.board.service, *Baggage.handling*, *Checkin.service*, *Inflight.service*, *Cleanliness* and *Leg.room.service*. These variables reflect satisfaction across various aspects of service, each on a scale from 1 to 5.

- *Departure.Delay.in.Minutes*, *Arrival.Delay.in.Minutes* and *Flight.Distance* provide quantitative measures of the flight experience. Flight distance is particularly important, as longer flights can impact expectations for service and comfort levels.
- Variables like *Gender*, *Age*, *Type.of.Travel* (business or personal), and *Class* (business, eco, eco plus) can influence satisfaction based on customer expectations and individual travel purposes.

2.3 Data analysis strategy

To effectively analyse customer satisfaction levels, this study uses a step-by-step approach with data preprocessing, correlation analysis, t-tests, and regression modelling:

- Data preprocessing: first, the data is cleaned and prepared to ensure accuracy and completeness.
- Correlation Analysis: this step checks the relationships between satisfaction and other factors like service quality and demographics. Using visualizations, it highlights the strength and direction of linear relationships, identifying variables with potential predictive value for satisfaction levels.
- T-tests: we use them to compare satisfaction levels between groups, such as loyal vs. disloyal customers and business vs. personal travellers. This will help to identify statistically significant differences in satisfaction across these groups.
- Multivariate regression: a regression model with satisfaction as the dependent variable will be used to measure the impact of each factor on satisfaction. This approach will help identify key predictors of satisfaction, distinguishing between primary and secondary influences on the customer experience.

3 Data analysis

3.1 Pre-processing

The Ryanair dataset was loaded using `read.csv()`, with ; as the separator to match the file's delimiter. Initial inspection through `str()` provided insight

into variable classes. Missing values were identified using `colSums(is.na())` and removed with `drop.na()`. The unnecessary first column labeled 'X', which held only index values, was removed using `select(-1)`. Categorical variables such as *Gender*, *Customer.Type*, *Type.of.Travel*, *Class*, and *satisfaction* were converted to factors using `mutate_at()` and `as.factor()`. Additionally, a binary variable, *satisfaction_numeric*, was created by encoding the *satisfaction* variable as 1 for "satisfied" and 0 for "neutral or dissatisfied," preparing it for statistical modeling. Outliers in *Flight.Distance*, *Departure.Delay.in.Minutes*, and *Arrival.Delay.in.Minutes* were managed using the Interquartile Range (IQR) method, with bounds calculated as $Q1 - 1.5 * IQR$ and $Q3 + 1.5 * IQR$, removing extreme values for more reliable analysis. The cleaned data was saved as a new CSV with write.csv() for easy reloading, resulting in a structured, high-quality dataset ready for accurate distribution, correlation, and modeling analysis.

3.2 Distributions

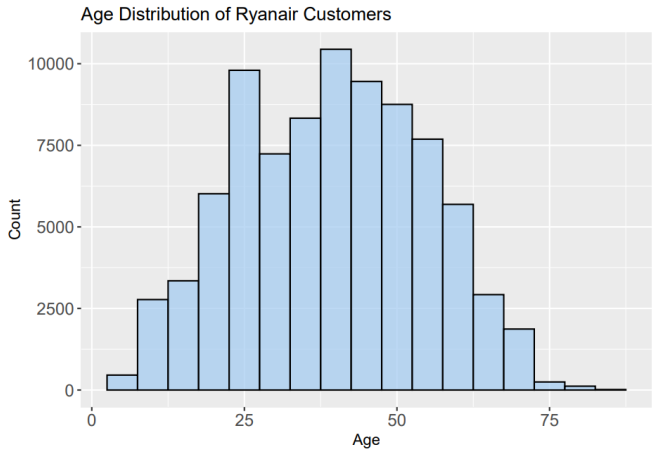

Figure 1: Age distribution of Ryanair customers

Figure 1 shows the age distribution of Ryanair customers. While Ryanair attracts customers of all ages, the distribution is denser between 25 and 50, accounting for over half of its customers. Ryanair should prioritize the preferences of this group without overlooking others, and investigate why representation from other age groups is lower and how it might be increased.

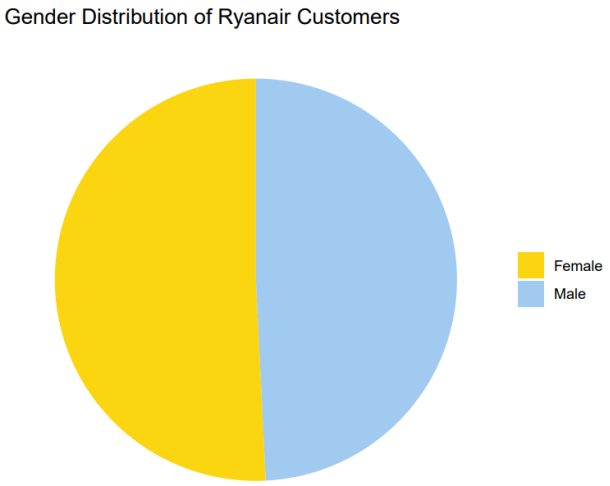

Figure 2: Gender distribution of Ryanair customers

As seen in Figure 2, Ryanair's customer data shows a near 50/50 split between males and females, indicating the company's appeal to both genders. Therefore, gender can be set aside in favor of focusing on other areas for policy improvement.

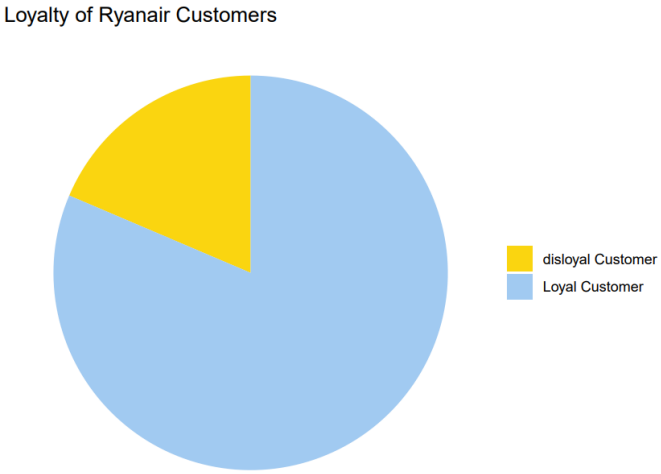

Figure 3: Loyalty of Ryanair customers

Figure 3 shows that the customer data primarily consists of loyal customers, introducing bias into the research. Service evaluation polls are sent electronically after each flight, but frequent Ryanair users are more likely to respond, while "disloyal" customers tend to ignore them. Thus, the data reflects the perceived service quality among repeat customers rather than an objective measure. However, this could also be advantageous, as regular customers may offer more informed assessments.

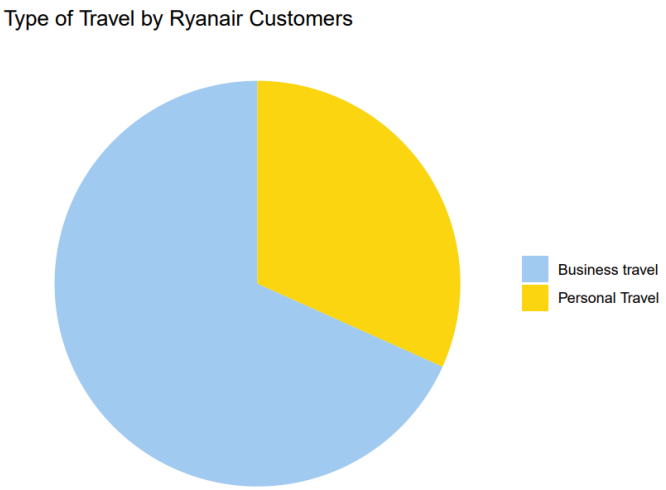

Figure 4: Type of travel by Ryanair customers

Business travel is the most common type of travel among customers, accounting for nearly two-thirds of all trips, as shown in the figure 4. Notably, business travel costs are covered by the employer, not the traveler, which may make these customers more lenient about service quality. To address this factor, we should assess satisfaction levels separately for business and personal travel to obtain more accurate results.

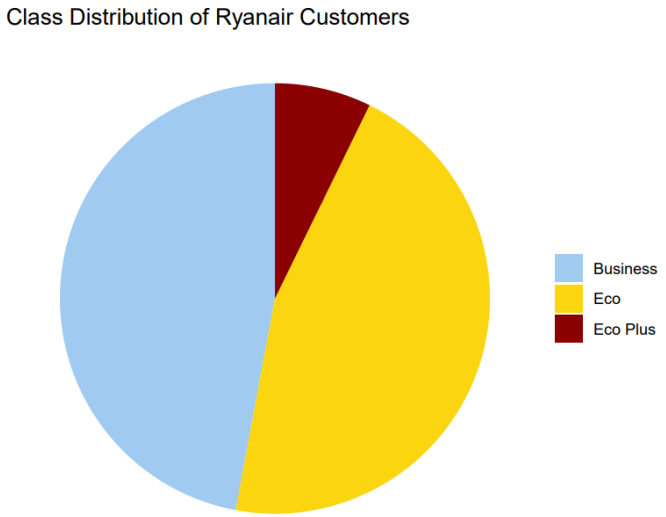

Figure 5: Class distribution by Ryanair customers

Figure 5 shows the distribution of flight classes preferred by Ryanair customers. Business class is the most popular, aligning with the trend in travel type preferences, followed closely by Eco class. However, Eco Plus is chosen by only a small minority. This suggests two points: first, policy evaluations should focus on Business and Eco classes, as they are nearly equally popular. Second, Eco Plus lacks distinctiveness and appeal, indicating a need to reconsider its design.

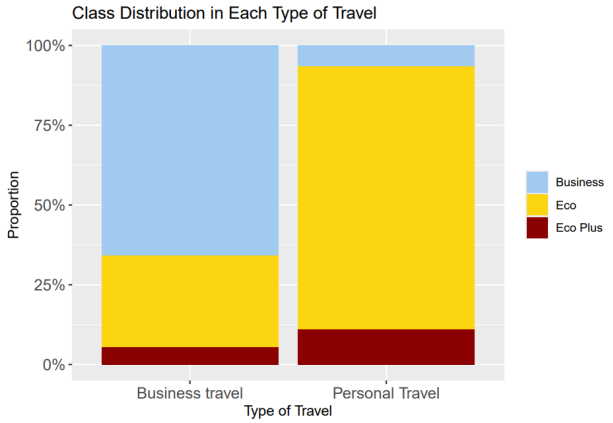

Figure 6: Class distribution in each Type of travel

The preferred flight classes are shown by travel type (Figure 6), with Eco Plus remaining the least popular. Interestingly, a significant number of business travelers also use Eco class, suggesting that some companies choose this option to reduce costs. This may also indicate limited availability or convenience of Business class at certain destinations.

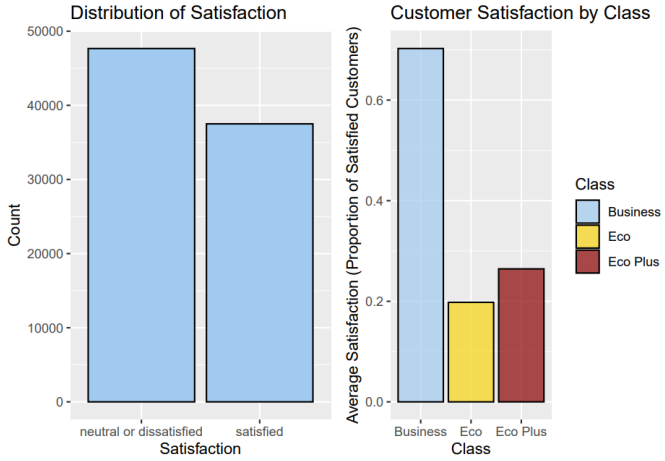

Figure 7: Distribution of satisfaction and average satisfaction by class

The results (Figure 7) show the number of dissatisfied/neutral versus satisfied customers, with dissatisfied/neutral responses significantly outnumbering satisfied ones—a concerning trend for Ryanair, especially given the loyalty of most respondents. This suggests a need for service improvements to boost satisfaction. Adjacent data reveals average satisfaction by class: as expected, satisfaction rises with class quality. Business class scores highest, at about 0.8 out of 1, while Eco and Eco Plus score around 0.2, indicating notably low satisfaction levels in these classes.

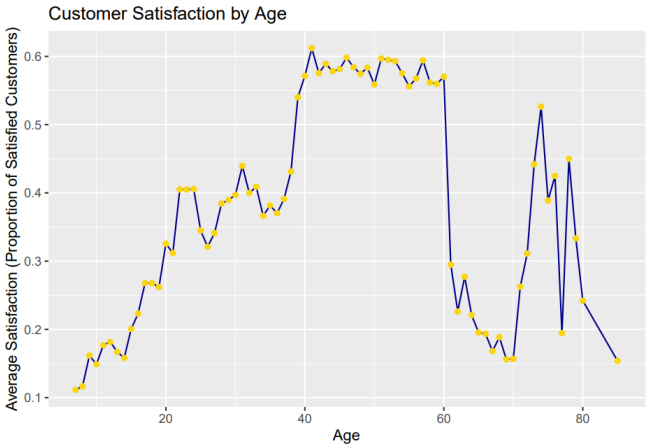

Figure 8: Customer Satisfaction by Age

As noted on Figure 8, the age range of Ryanair customers is broad, though concentrated between 25 and 50. The graph shows that average satisfaction is highest among customers aged 40-60, at over 0.5 out of 1. However, satisfaction drops significantly among those aged 20-40, with especially low ratings around age 25. Since the 25-40 age group represents a substantial portion of the customer base, improvements targeting their satisfaction are warranted. We will explore this further in the following sections.

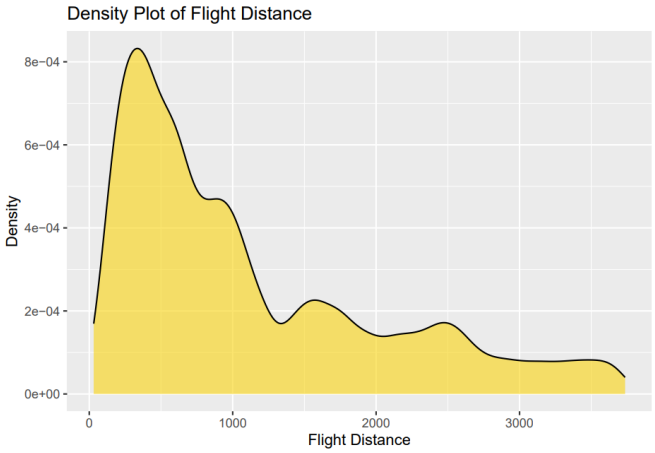

Figure 9: Density plot of Flight distance

Figure 9 shows flight density by distance, with most Ryanair flights falling between 1,000–1,500 km, averaging 2-3 hours. This reflects Ryanair's focus on short-distance travel, which aligns with its primarily European operations.

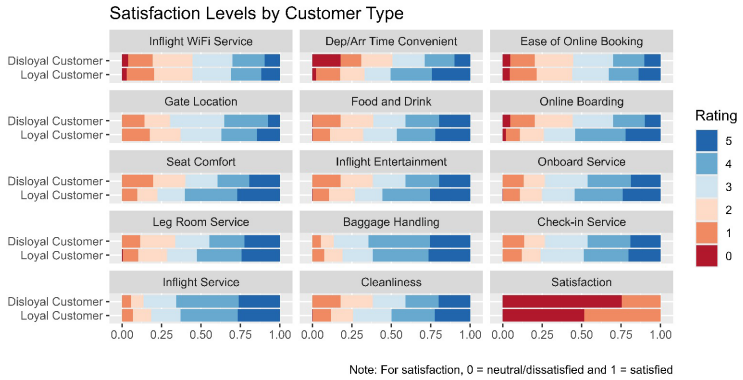

Figure 10: Satisfaction level by Customer type

Figure 10 displays satisfaction levels across various services. Loyal customers consistently rate aspects like inflight entertainment, Wi-Fi, and online boarding higher than disloyal customers, highlighting a strong link between service satisfaction and loyalty. Minimal differences in areas like gate location and departure/arrival time convenience suggest these factors are less influenced by loyalty. Addressing satisfaction gaps in inflight amenities could boost overall loyalty by better meeting the needs of disloyal customers.

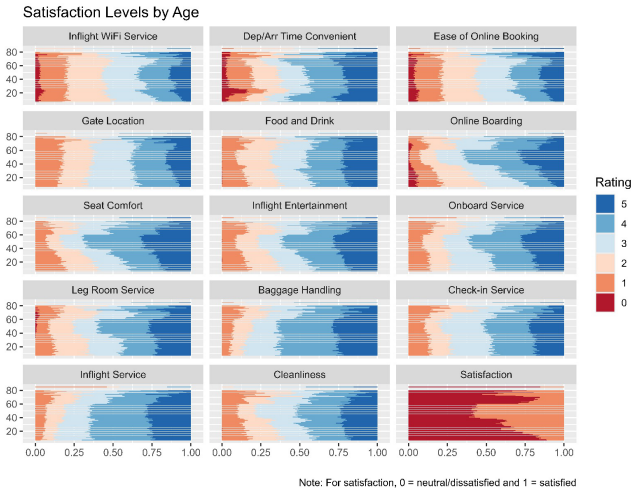

Figure 11: Satisfaction levels by Age

Figure 11 compares service ratings across age groups, showing that younger travelers generally rate services like legroom, food and drink, and seat comfort lower, while older travelers rate areas such as inflight entertainment, service, and cleanliness higher. Services like online boarding, booking ease, and baggage handling show little variation across ages, suggesting they're valued similarly by all. Overall, older travelers report higher satisfaction (closer to 1), while younger travelers are more critical, especially regarding comfort. Enhancing comfort-focused services could improve satisfaction among younger travelers.

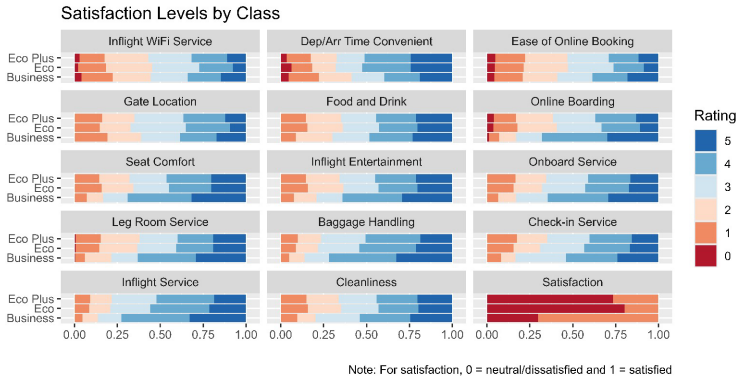

Figure 12: Satisfaction levels by Class

Figure 12 illustrates passenger satisfaction ratings by travel class (Business, Eco Plus, and Economy) across various airline service aspects, using a color gradient from red (low ratings) to blue (high ratings). Business class consistently receives higher ratings, particularly in *Seat.comfort*, *Leg.room.service* and *Inflight.service* as reflected by the darker blue bars. Eco Plus ratings typically fall between those of Economy and Business, often performing better than Economy in comfort-related areas. Economy class shows a predominance of lower ratings, especially in *Leg.room.service* and *Inflight.wifi.service* indicating dissatisfaction. Overall, Business class passengers report the highest satisfaction, followed by Eco Plus and Economy. Categories like *Ease.of.online.booking* and *Online.boarding* display a wider range of ratings across all classes, reflecting diverse passenger experiences. To enhance overall satisfaction, airlines should consider improving legroom and seat comfort in Economy class, as well as inflight WiFi and online services, which impact passengers across all classes. Focusing on these key areas could foster a more positive travel experience, particularly for Economy and Eco Plus passengers, ultimately building customer loyalty.

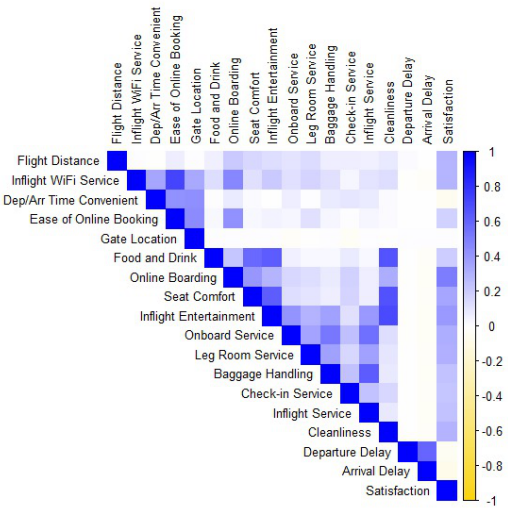

Figure 13: Correlation matrix

Figure 13 illustrates the relationships between individual variables. For instance, inflight Wi-Fi service is highly correlated with ease of online booking and online boarding, indicating a customer focus on internet connectivity. Some services complement each other; for example, food and drink service shows positive correlations with seat comfort, cleanliness, and inflight entertainment, suggesting that these services should be improved collectively.

While no single variable solely determines overall airline satisfaction, it is most closely correlated with online boarding. Therefore, the company should prioritize investment in online and internet services. Although most variables have a weak but measurable impact on general satisfaction, it is surprising that delays do not significantly affect satisfaction levels, and gate location appears to be of little importance.

3.3 Statistical modeling

1. **T-test modeling: Differences in satisfaction levels by Gender**

   For the first empirical analysis, we examined whether there is a statistically significant difference in satisfaction levels between male and female passengers. To this aim, we ran a T-test which is a prevalent type of statistical test that is used to compare the means of two groups.[1] Given that satisfaction is an important metric in customer service research, we argue that understanding any gender-based differences could provide insights into potential areas for Ryanair service improvement. For this purpose, we con-

[1]Kim TK. *T test as a parametric statistic.* Korean J Anesthesiol. 2015 Dec;68(6):540-6. doi: 10.4097/kjae.2015.68.6.540. Epub 2015 Nov 25. PMID: 26634076; PMCID: PMC4667138

verted satisfaction into a numeric variable (0 for neutral/dissatisfied, 1 for satisfied), this allowed us to calculate the mean satisfaction levels for each gender (male and female) and test whether any observed difference was due to chance or statistically meaningful.

The T-test analysis offers a statistically grounded comparison of mean satisfaction levels between male and female passengers while taking into account potential variance differences between these groups. As depicted in Figure 14, our findings indicate that female passengers report an average satisfaction score of 0.4337, while male passengers have a slightly higher mean satisfaction score of 0.4472. Although this difference may appear modest, the t-test statistic of -3.98, combined with a statistically significant p-value ($p < 0.01$), suggests that this variation is unlikely to be attributed to random chance. As a result our analysis highlights a gender-based distinction in satisfaction levels that Ryanair might consider when refining its strategies to enhance customer satisfaction across its diverse passenger base.

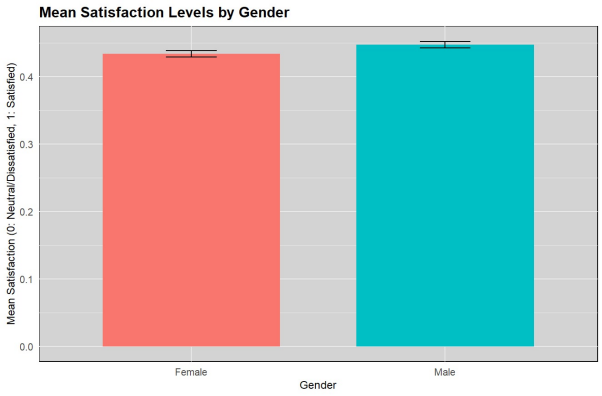

Figure 14: Mean satisfaction levels by Gender

2. **Regression modeling: Factors effecting overall customer satisfaction**

In this model, we applied multivariate regression analysis to measure the impact of various service variables on overall passenger satisfaction. Multivariate regression is a statistical tool used to evaluate the effect of multiple independent variables (predictors) on a single dependent variable. In this case, our dependent variable is passenger satisfaction, while the independent variables are all service-related factors. Consequently, this approach allows us to uncover which specific service elements influence satisfaction.

Process and findings

The pathway of the analysis involved fitting a logistic regression model where *Satisfaction* was coded as a binary outcome (satisfied vs. not satis-

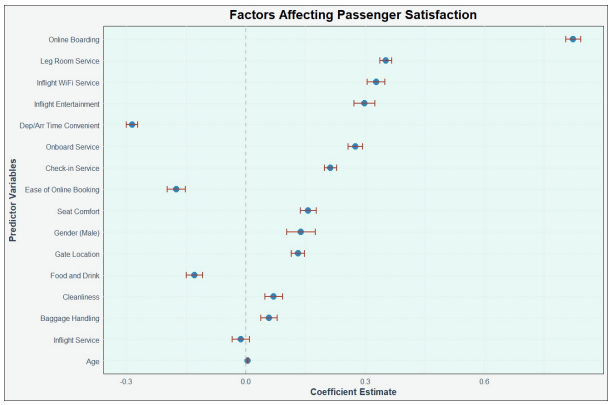

Figure 15: Factors affecting passenger satisfaction

fied). We included variables such as *Inflight.wifi.service*, *Online.boarding*, *Seat.comfort* and *Food.and.drink.service* among others. This method enabled us to identify both positively and negatively correlated factors with *Satisfaction*.

As shown in the coefficient plot (see figure 15), *Online boarding* emerged as the strongest positive predictor (coefficient = 0.823), suggesting that a smooth boarding experience is highly valued by passengers. Other impactful factors included *Inflight.wifi.service* (0.327) and *Leg.room.service* (0.352), indicating that onboard amenities play a significant role in shaping overall satisfaction. In contrast, *Ease.of.Online.booking* and *Food.and.drink.service* showed negative associations (-0.175 and -0.129, respectively), which suggest that issues with these services may detract from passenger satisfaction. Moreover, the model's statistical significance ($p < 0.01$ for most variables) confirms the robustness of these relationships (see also table 1. in appendix).

4 Discussion

4.1 Limitations for empirical analysis

While our multivariate regression analysis explores how each service factor relates to *Satisfaction*, it's important to remember that these results show associations, not causation. Including all service variables at once may introduce multicollinearity, where some variables overlap in their predictive power, and there may also be bias from unobserved factors impacting satisfaction. Despite these limitations, our analysis highlights key areas for improvement across the service experience, offering valuable insights for Ryanair's customer service department to evaluate and potentially enhance their CSAT (Customer Satisfaction) Score.

4.2 Discussion of Results

The findings of this report reveal a complex landscape of factors influencing customer satisfaction each of which intersecting across demographics and service categories. The data also reveals a diverse passenger base, predominantly aged 25 to 50, split almost evenly between male and female with a strong representation of business travelers who generally report higher satisfaction levels compared to personal travelers. This pattern is particularly noticeable among business travelers in Business or Economy class, where overall satisfaction trends higher. However, loyalty among frequent customers appears insufficient to offset dissatisfaction in specific service areas: for instance, satisfaction scores remain consistently low for Economy and Eco Plus classes, especially concerning food and drink service and seat comfort. Moreover, even among loyal customers, there is a clear demand for improved online services, specifically in booking and boarding processes.

The correlation and regression analyses reveal critical factors affecting satisfaction, with online boarding emerging as the most significant positive predictor, suggesting that Ryanair customers highly value seamless boarding processes. Additionally, in-flight amenities such as WiFi and legroom services also contribute positively, indicating that these features are critical to overall satisfaction. Conversely, dissatisfaction with ease of online booking and food and drink services negatively correlates with satisfaction, highlighting these as areas where customer experience falls short. Interestingly, delay-related factors exhibit minimal correlation with satisfaction, suggesting that punctuality holds less sway in shaping passenger perceptions compared to in-flight and service-based amenities.

4.3 Recommendations for Ryanair

1. **Enhance Digital and Online Services:** Our analysis underscores the importance of digital service quality, with online boarding scoring as the strongest satisfaction predictor and online booking showing negative associations with satisfaction. Therefore, we suggest that Ryanair would benefit from streamlining digital experiences, particularly in booking, check-in, and boarding processes, to appeal to the needs of younger, tech-savvy passengers.

2. **Upgrade Comfort in Economy and Eco Plus:** Consistently low satisfaction scores in Economy and Eco Plus classes, especially concerning seat comfort and legroom, signal the need for targeted improvements. For this reason, we argue suggest that an incremental upgrades in cabin layout or seat options may significantly enhance satisfaction in these classes without

major cost increases.

3. **Revise Food and Drink Offerings Across All Classes:** We observed that a low satisfaction with food and drink services spans across customer segments. A revision of in-flight menu options focusing on variety, quality, and affordability could improve satisfaction levels, benefiting both frequent and infrequent travelers who find these offerings lacking.

4. **Prioritize Service Enhancements Over Flight Punctuality:** Our findings indicate that punctuality has limited impact on satisfaction compared to digital services and in-flight amenities. Ryanair might more effectively boost satisfaction by focusing on connectivity, entertainment options, and other service-based features that more directly shape customer experience.

5 Conclusion

Another team can expand on the findings of our report by conducting deeper analyses on specific service aspects that showed significant correlations with passenger satisfaction by controlling other variables (e.g. confounder). For instance, further exploration into online boarding and inflight WiFi service could involve segmenting the data based on customer demographics, flight routes, or seasonal patterns to uncover more detailed insights. Additionally, a more detailed examination of negative predictors, such as ease of online booking and departure/arrival time convenience, could involve surveys or focus groups to understand the underlying reasons behind customer dissatisfaction, leading to more targeted solutions.

Furthermore, the data analysis revealed several key factors that significantly influence customer satisfaction for Ryanair passengers. The strongest positive predictors identified were online boarding and inflight WiFi service, highlighting these areas as core strengths of Ryanair's service offering. Enhancing these features further could lead to higher satisfaction scores and increased customer loyalty. On the other hand, aspects such as time convenience and ease of online booking emerged as significant areas needing improvement, as they had a negative impact on satisfaction. Additionally, demographic analysis showed that older passengers and male travellers tend to report higher satisfaction levels, suggesting the need for more personalized services to meet diverse customer expectations. Therefore, this report underscores the importance of operational efficiency and service quality in shaping the customer experience. By addressing the areas identified for improvement we believe that Ryanair can better meet passenger expectations and sustain its competitive advantage in the airline industry.

A Appendix

Table 1: Regression Results for Factors Affecting Satisfaction

	Dependent variable:
	Satisfaction
Age	0.005*
	(0.001)
Gender (Male)	0.139*
	(0.018)
Inflight WiFi Service	0.327*
	(0.011)
Departure/Arrival Time Convenient	−0.286*
	(0.007)
Ease of Online Booking	−0.175*
	(0.012)
Gate Location	0.131*
	(0.008)
Food and Drink	−0.129*
	(0.010)
Online Boarding	0.823*
	(0.010)
Seat Comfort	0.157*
	(0.010)
Inflight Entertainment	0.298*
	(0.013)
Onboard Service	0.275*
	(0.009)
Leg Room Service	0.352*
	(0.008)
Baggage Handling	0.058*
	(0.010)
Check-in Service	0.213*
	(0.008)
Inflight Service	−0.012
	(0.011)
Cleanliness	0.070*
	(0.012)
Constant	−7.644*
	(0.067)
Observations	85,161
Log Likelihood	−38,007.360
Akaike Inf. Crit.	76,048.710
Note:	p<0.1; p<0.05; p<0.01

B References

Kim, T. K. (2015, December). T test as a parametric statistic. *Korean Journal of Anesthesiology*, *68*(6), 540–546. (Epub 2015 Nov 25) doi: 10.4097/kjae.2015.68.6.540