A Work Project, presented as part of the requirements for the Award of a Master's degree in Business Analytics from the Nova School of Business and Economics.

# THRIVING IN THE DIGITAL DINING ERA: A DATA-DRIVEN STRATEGY FOR OPTIMIZING UBER EATS MENUS

GONÇALO ALVES DE MELO

Work project carried out under the supervision of:

Samuel Pedro

Pedro Verruma

Nuno Barroso

Rodrigo Belo

20/12/2023

**Abstract:** How to stay relevant in the online food delivery services industry? It is within this dilemma that this thesis emerges. More precisely, how can a restaurant manager enhance a menu to better serve consumer needs? The core objective is integrating a data-driven and consumer-centric approach into a calculation method developed by Pleez, a food-tech startup. This method evaluates a client's menu against two competitors, analyzing visual and organizational attributes. Based on the comparison, the goal is to empower clients to improve their menus. Consequently, the thesis also evaluates the effectiveness of this initiative in prompting a "call-to-action" among Pleez clients.

**Keywords:** Data-Driven Business Decisions, Customer-Centricity, Regression Analysis, Competition Effects

**Acknowledgments:** To my advisors at Pleez Samuel Pedro, Nuno Barroso, Pedro Verruma, and all company members for the continued support provided throughout this thesis, for presenting me with this opportunity, and for guiding me throughout the process. Additionally, to my advisor at Nova School of Business and Economics, Rodrigo Belo, for the support provided.

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

## Table of Contents

## 1. Introduction

Online food delivery services (OFDS) refer to food ordering and delivery systems that connect restaurants with customers through digital channels (Ray et al., 2019). The development of easy-to-use applications, changes in consumer behavior, and the widespread use of smartphones coupled with tech-enabled driver networks were crucial factors in the growth of the Food Delivery Applications (FDA) Industry. The way people eat has evolved! This trend towards OFDS has become even more pronounced during the pandemic, which has given the FDA's industry an enormous boost (Ahuja et al., 2021). Subsequently, providing online delivery services through the FDA's has become mandatory. It is unlikely that people will go back entirely to traditional ways of interacting with the food industry. In addition to the repercussions of the pandemic, the age group that contributes most, about 70%, to online deliveries is 21 to 36, and it is undeniable the pivotal role of youth in shaping and defining current and future societal norms (Hospitality, 2022). As a consequence, restaurants that neglect the importance of technology will be unable to meet the digital natives' demand and be left behind (data and Yomoc, 2023b).

The FDA's industry already accounts for 30% of the revenue for numerous restaurant brands, a substantial increase from the period before the pandemic, which constituted less than 10% (Saraswat, 2023). However, restaurant businesses face challenges staying relevant in the FDA segment once it becomes highly competitive. The rapid growth of the number of players is shown by Uber Eats, the most relevant delivery platform in Portugal, where there were already 12 000 restaurants in 2022 (Queixa, 2022). Therefore, to survive and thrive in the digital dining era, restaurants must recognize the profound impact of collecting and aggregating raw data from the daily operations of the FDA's, supported by real-time data analysis. This allows the restaurants to make data-driven informed decisions, such as comparing competitor prices,

allowing for an adjustment of the sales strategies, and measuring the impact of menu changes, allowing to identify consumer-attractive options (data and Yomoc, 2023a).

This digital transformation presents immense opportunities, but it also brings challenges as for many restaurants, FDA's are heavy, demanding significant manual effort to extract critical data, translating to a large amount of work (Saraswat, 2023). It is within this dilemma that Pleez emerges, a Portuguese food tech startup that empowers restaurants to use the full potential of delivery platforms through data-driven decision-making. Pleez Sales Boost Platform integrates real-time data from delivery applications, offering restaurant managers a 360º view of market trends, sales performance, competitor activity, and price comparisons. It also provides machine learning recommendations for improving delivery offerings and AI-powered features like menu image creation.

## 2. Problem Identification and Motivation

Pleez's primary objective is to improve restaurants' presence on FDA's (Uber Eats, Glovo, and Bolt) with data-driven techniques. Pleez accomplishes this by delivering processed real-time data to restaurant managers via the Pleez Sales Boost Platform. In addition to this accessible processed data, each restaurant's account manager plays an essential role by providing insights for enhancing the restaurant's visibility across all the FDA's. These insights are drawn from private restaurant data and publicly available data from the FDA's. In addition to the recommendations being accessible on the Pleez Sales Boost Platform, as of April 2023, Pleez clients receive all recommendations through WhatsApp and email. The recommendation approval rate stands at 70%. However, the remaining 30% does not necessarily indicate rejection. Instead, it simply implies that the client has allowed the recommendations to expire once, it is active for one week. To assist clients in recalling these recommendations, Pleez is currently in a testing phase, started in September 2023, involving reminders, email

notifications, and automated calls to which the client only needs to respond with a vocal "yes." It is worth noting that, within the first eight months of 2023, 12 700 recommendations have already been extended to clients in Portugal and Spain.

Beyond helping clients recall recommendations, an additional objective of Pleez is to make these suggestions more engaging for clients and improve their “Call to Action.” As a result, these improved recommendations aim to help restaurants stay relevant within the FDA’s industry, consequently leading to an increase in their sales.

This motivated the creation of the Pleez Rating project, which will be the main subject of this thesis. The Pleez Rating aims to create a classification system to evaluate restaurant menus based on their visual and organizational attributes. This evaluation will assign a rating on a scale from 0 to 10. The purpose is to rate a restaurant's menu and make a comparative analysis with the menu rating of two competitors. The introduction of the Pleez Rating holds the potential to provide a more qualitative means of comparison between Pleez customer's menus and those of their competitors. This, in turn, hopefully empowers them to take proactive measures to enhance the quality of their menus. If this proves successful, it may become a feature on the Account Managers' Platform, enabling them to embed this comparison on insights and perform it for all their clients. Therefore, the first question that this thesis aims to answer arises:

**Question 1:** How does the adoption of the rating comparison system impact the likelihood of restaurant managers taking action to make changes to their menus?

When the project was conceptualized, a set of criteria was idealized, drawing on the experience of the Pleez product team. The classification method idealized for the Pleez Rating involved six different parameters, each assigned specific weight and score criteria in the classification process. However, the selection of parameters, respective weights, and score criteria were

determined through expertise judgment. Given these circumstances, after a meeting with the product team, concerns emerged regarding aligning the Pleez Rating calculation method with consumer needs, prompting the necessity to address and affirm this alignment. Consequently, the second question that this thesis aims to answer arises:

**Question 2:** What visual and organizational aspects of a menu influence consumers' perceptions?

## 3. Objectives of a Solution

To answer the questions addressed in the previous section, Problem Identification and Motivation, the Pleez Rating project will be divided into two distinct phases, each with its distinct set of objectives. The two stages of the project exclusively incorporated restaurants available on the Uber Eats FDA, serving as the project's initial focus point.

The first phase aims to answer the question: Does the Pleez Rating drive restaurant managers to make changes to their menus? Ideally, the Pleez Rating impact assessment should be conducted using the optimized consumer-centered calculation method. However, due to the limited time for the research, this phase will encompass the execution of the Pleez Rating classification methodology based on the in-house expertise, wherein 10 restaurants will be selected. Subsequently, two competitor restaurants will be identified for each of these selected establishments using a competitor algorithm. On that account, in the pursuit of assessing the competitive effects on restaurant managers, ratings will be assigned to these 30 restaurants (10 * ( [1 Pleez Client] + [ 2 Competitors] ) = 30) , and an email communication will be dispatched.

**Objective 1:** Assess whether the email dispatched to the 10 clients with the Pleez Rating effectively elicits a "call to action" among the restaurant managers.

The second phase aims to answer the question: What visual and organizational aspects influence consumers perceptions? This phase will embrace a data-driven approach, optimizing the Pleez Rating calculation method to pursue customer-centricity. This will be accomplished by building a dataset encompassing various parameters on the visual and organizational aspects of 150 restaurant menus. Each of these restaurants will be subject to a rating process by a sample of users via Google Forms. Afterward, a regression analysis will be conducted through the development of a predictive model. A predictive model is simply a tool used to estimate the value of an unknown variable of interest, known as the target variable (2013, Foster Provost, Tom Fawcett). Moreover, models incorporating a single target variable and multiple independent variables are named multilinear regressions (Uyanık & Güler, 2013). The second phase of the Pleez Rating Project will use a multilinear regression, with a menu's visual and organizational rating as the target variable. Regression analysis is a statistical method employed to estimate the relationship between variables characterized by a cause-and-effect relation (Uyanık & Güler, 2013). Consequently, it will be feasible to evaluate the cause-and-effect relation of each parameter with the rating given by the consumers and compare it with the expertise-driven calculation method. This comparison will allow for an in-depth analysis of whether the classification undertaken through expertise aligns with consumer preferences, thus contributing to a comprehensive re-evaluation of the Pleez Rating calculation method.

**Objective 2:** Conduct a regression analysis through the development of a multiple linear regression model to identify the visual and organizational elements within a menu that impact consumer perceptions. Additionally, reassess the calculation methodology based on the gathered data to ensure alignment with consumer preferences.

This Thesis delves into a comprehensive examination of the two distinct phases and their corresponding findings.

## 4. Methodology

### 4.1. Data Collection and Understanding

To fulfill the objectives abovementioned in Objectives of a Solution, information covering all restaurants for both phases of the Pleez Rating project had to be collected. The data for all restaurants was collected between October 10 and October 26, 2023.

#### 4.1.1. Data Collection: Assessing the Effectiveness of the Pleez Rating

For the first phase of the Pleez Rating, as mentioned, when the project was conceptualized, a classification method was idealized. The classification method for the Pleez Rating involved six different parameters, each assigned a score criteria and a specific weight in the classification process. Upon initiating the first phase, a secondary assessment of the calculation method was conducted with the product team. The meeting aimed to ensure that the expertise calculation method while leveraging internal knowledge, was well-designed for the accurate evaluation of a menu. While keeping the parameters, adjustments were made to the filters employed in data collection to acquire the intended information for scoring attribution.

Apart from parameters commonly associated with user-friendliness, such as the Number of Menu Items, Number of Menu Categories, Classification, and Number of Reviews, other parameters were maintained. For instance, the Number of Menu Items with Pictures since when a consumer is exposed to a visual stimulus, Ghrelin, a hormone secreted by the organism, is released, causing a physiological reaction. Ghrelin is essential to transmit an individual's appetite signals to the brain. Ghrelin is released into the bloodstream more quickly when food imagery is presented, which increases appetite (Mishra, 2020). Consequently, the inclusion of this variable was imperative. It is worth noting that certain restaurants have achieved a remarkable increase in their average customer expenditure, up to 30%, by simply incorporating

food images in combination with textual descriptions within their menus (Mishra, 2020). Therefore, the Number of Menu Items with Descriptions was also maintained. Information was manually collected from the Uber Eats FDA to accomplish this analysis. The entire set of variables used is provided in Appendix 1, Table 2 to provide a complete understanding.

The data was gathered for 10 Restaurants, and two competitors were selected for each restaurant. The selection of these competitors was made using an algorithm developed by Pleez. The restaurants chosen for this research were the top 5 restaurants in terms of monthly sales from 2023 until September and the lowest-performing 5 restaurants with sales exceeding 200€ during the period (among restaurants partnered with Pleez). The sales data was retrieved from the Pleez database.

Concerning the data collection for scoring attribution, specific filters were applied to extract relevant data and ensure the accuracy of the score calculation. Within the Number of Menu Categories parameter, a deliberate decision was taken to abstain from incorporating non-essential categories, notably not including “Featured Items” and “Promotional” Categories, within the calculation. This selective omission aimed to enhance the precision of the analysis and channel attention towards the core menu categories.

Moreover, it is essential to highlight that the count of items excluded those mentioned in the above categories. This exclusion was deployed to avoid duplication, as those items are included in other categories. Add-ons or contents nested within the items and those classified as drinks and cutlery were also excluded from the final count of the Number of Menu Items. The rationale behind prioritizing the principal menu items was to bring a more comprehensive and meaningful representation of the restaurant's offerings.

In the filtration process applied to the Number of Menu Items with Pictures and the Number of Menu Items with Description, specific items, such as drinks and cutlery, were also intentionally

excluded. The reasoning behind this exclusion was to focus attention on the primary menu items, which exert a more pronounced influence on customer perception.

In relation to the Number of Reviews, the filtration process did not consider the presence of the "+" sign. In instances where a restaurant displays, for example, "200+" reviews, the value for analysis purposes was treated as 200.

### 4.1.2. Data Collection: Improvement of the Calculation Method

For the second phase of the Pleez Rating, data was collected from the menus of 150 restaurants. A meeting with the development team was conducted to determine which parameters to include in the second phase. The goal was to incorporate as many parameters as possible, ensuring they were feasible for automated scrapping in the future. Scraping refers to identifying a source housing information and using software to retrieve the desired data from the source (Ot, 2021). The selection is also aligned with Pleez's current objectives, considering the time investment required to develop these scrapers. Consequently, a wide range of parameters were chosen to be included. Following the first phase, the Number of Menu Items, Number of Menu Categories, Number of Menu Items with Pictures, Number of Menu Items with Descriptions, Classification, and Number of Reviews were chosen. Additionally, the selected menus were analyzed for the count of the Number of Menu Drinks, the Number of Menu Desserts, the percentage of Menu Items without descriptions that were drinks or desserts, and the percentage of Menu Items without picture that were drinks. Moreover, the presence or absence of the category labeled "Featured Items" was also considered in the analysis. In order to provide a complete understanding, the entire set of variables used is provided in Appendix 1, Table 3. The model will assign the weights to each parameter.

In contrast to the first phase, the second phase lacks predefined score attribution as the ratings will be acquired through a survey. Consequently, the data collection process in this phase did

not involve applying any filters. The only exception was that any add-ons or contents inside the items were not included in the count for any of the parameters, following the rationale applied in the first phase of the project. Furthermore, the Number of Reviews also underwent a filtering process similar to the project's initial phase. This phase was intentionally unfiltered to allow the model to grasp the significance of raw variables.

In the context of the target variable, visual and organizational menu rating, a data collection process was essential to train the multiple linear regression model. In this way, the collection and understating of data were conducted through 30 Google Forms, each with 5 restaurants, for the purpose of gathering menu ratings from 150 different restaurants, encompassing a varied number of restaurant types. Respondents were instructed not to consider their personal food preferences or pricing considerations but to concentrate solely on the visual and organizational aspects. It is noteworthy that the data-gathering process involved participation from 50 respondents. In the process of collecting the five distinct Pleez Ratings for each restaurant, two metrics, the median, and the mode, were assessed. However, owing to the subjective nature of the evaluations, it was observed that for 11 restaurants, there were no similar ratings, making the application of the mode impractical. Consequently, the metric employed for data gathering was the median.

The Pleez Rating achieved through the expertise-based method will be denoted as the Calculated Pleez Rating. Conversely, the term User-Feedback Pleez Rating will be used for the ratings derived from this data collection. Henceforth, referencing restaurants with a User-Feedback Pleez Rating of 10 implies establishments with a median rating of 10 for the menu's visual and organizational aspects gathered through the conducted surveys.

#### 4.1.3. Data Collection: Assessing the Predictive Model Results

Additionally, for the second phase of the Pleez Rating, in the pursuit of achieving a Pleez Rating that accurately reflects consumer preferences and validates the possible modifications to the Calculated Pleez Rating motivated by the model results, among the thirty Google Forms conducted, ten surveys were selected to incorporate an additional survey. This supplementary survey utilized a scale from 0 to 10 to directly assess the respondents' opinions on the importance of specific parameters. These parameters included the Number of Menu Items, the Number of Menu Items with Pictures, Visually Appealing Images, the Number of Menu Items with Descriptions, Description Content, and Classification. In order to provide a complete understanding, the entire set of variables used is provided in Appendix 1, Table 4. The inclusion of specific parameters aimed to accurately determine the model's ability to predict individuals' genuine preferences in a menu. This was achieved by asking for evaluations regarding the importance of parameters included as independent variables in the model. Other questions, such as those related to the importance of visually appealing images and the importance of description content, aimed to evaluate individual's perceptions of more visual and organizational aspects of a menu, providing insights into possible enhancements to the Pleez Rating in the future, which were not intended to be used during the execution of this research. Opinions from 50 participants were gathered using the mode as the central tendency measure.

### 4.2. Data Compilation

Following the conclusion of the Data Collection, three datasets were constructed.

#### 4.2.1. Data Compilation: Assessing the Effectiveness of the Pleez Rating

The first dataset pertains to the first phase of the project. As mentioned, it encompasses 6 parameters, each with a specific weight. Each parameter also includes a specific score derived

from the information collected from 30 restaurants. Moreover, it includes the calculation of a final rating: Calculated Pleez Rating for each restaurant, achieved by summing the products of the scores and weights assigned to each parameter. The scores are attributed on a scale from 0 to 10. Table 2 of Appendix 1 outlines the parameter's weight and the criteria for calculating each parameter's score.

#### 4.2.2. Data Compilation: Improvement of the Calculation Method

The second dataset corresponds to the compilation designed for the multiple linear regression model. As mentioned, it incorporates data regarding 12 parameters gathered from 150 restaurants. However, certain restaurants feature multiple menus on Uber Eats, providing breakfast, lunch, and dinner. During the scraping process, only the first menu encountered was considered. Unfortunately, whether respondents base their ratings solely on the first menu or consider all menus remains unclear. Consequently, 9 restaurants were excluded from the dataset for the model, resulting in a total of 141 restaurants used. Therefore, the dataset includes information for the 12 parameters across 141 restaurants, along with the corresponding median ratings obtained from the Google Forms survey: User-Feedback Pleez Rating. Table 3 of Appendix 1 outlines the 12 parameters.

#### 4.2.3. Data Compilation: Assessing the Predictive Model Results

The third dataset consolidates the 10 additional surveys assessing the perceived importance of specific parameters. As mentioned, this dataset encompasses six parameters and the mode of the 50 respondents for each parameter. Table 4 of Appendix 1 outlines the 6 parameters.

### 4.3. Descriptive Statistics and Exploratory Data Analysis

The Exploratory Data Analysis will exclusively focus on the second dataset due to limited efficacy in conducting exploratory data analysis for the first and third datasets. It is imperative to comprehensively investigate the second dataset, examining variables and discerning potential interdependencies before proceeding to the model development stage and recommending potential enhancements to the Pleez Rating.

The distribution of the number of restaurants per User-Feedback Pleez Rating is presented in Appendix 2, Figure 3. The distribution reveals that User-Feedback Pleez Ratings 6, 7, and 8 have the highest counts of restaurants, with a notable peak at a User-Feedback Pleez Rating of 8, where 35 restaurants are represented.

Further exploration involves an investigation into the Number of Menu Categories. Restaurants with the highest User-Feedback Pleez Ratings exhibit a positive correlation with more categories, as demonstrated in Figure 4, Appendix 2. Compared to the Calculated Pleez Rating, a restaurant requires 6 categories to secure a 10 in the final rating. Despite the limited representation in the dataset, establishments with a User-Feedback Pleez Rating of 10 display an average number of categories of 16. This finding is based on only two instances with this particular rating. Those with an average number of categories of 6 had User-Feedback Pleez Ratings of 5 and 6, further emphasizing the contrast with the Calculated Pleez Rating.

Moreover, Figure 5 in Appendix 2 visually shows a positive proportionality between the User-Feedback Pleez Rating and the Number of Menu Items. This implies that establishments achieving the highest User-Feedback Pleez Rating demonstrate a higher product offering. Establishments with a User-Feedback Pleez Rating of 10 manifest the highest average number of drinks, precisely 33.50. Contrarywise, the average number of drinks for User-Feedback Pleez Ratings 1 through 9 falls within the range of 9 to 12. It is noteworthy to emphasize that a restaurant must present a menu featuring 25 to 30 items, excluding drinks, to secure a

Calculated Pleez Rating. However, an examination of restaurants with a User-Feedback Pleez Rating of 10 reveals an average number of items at 138.5. Although the average number of drinks totals 33.50, excluding drinks results in a count of 105 items. This underscores a distinctive criterion concerning the number of items, wherein a restaurant with 105 items in the Calculated Pleez Rating would receive a score of 0 in the parameter Number of Items. Despite the limited occurrence of a User-Feedback Rating of 10 in the dataset, even when considering User-Feedback Ratings of 9 and 8, the average number of items, excluding drinks, is 40 and 36, respectively.

Establishments with the highest count of desserts correspond to those with User-Feedback Pleez Ratings of 2, 8, 9, and 10. Except for a User-Feedback Rating of 2, the data illustrates that a higher number of desserts is correlated with higher User-Ratings, as shown in Figure 6, Appendix 2. Through a comprehensive analysis of Figure 7, Appendix 2, a positive correlation also exists between the User-Feedback Pleez Rating and the Number of Menu Items with Pictures, with the exception of a slight dip at a User-Feedback Pleez Rating of 5 and 10. Moreover, the relationship between the User-Feedback Pleez Rating and Menu Items with Description also exhibits a clear positive proportionality, as shown in Figure 8, Appendix 2. Specifically, among the restaurants classified with a User-Feedback Pleez Rating of 9, the Percentage of Menu Items With a Description was notably high at 96%.

Further exploration reveals that, apart from two anomalous cases, where a restaurant with a classification of 4.40 and 108 reviews was classified with a User-Feedback Pleez Rating of 0 and another with a classification of 4.90 despite having a User-Feedback Pleez Rating of 1, there is a consistent positive correlation between the User-Feedback Pleez Rating, Classification, and the Number of Reviews, as shown in Figure 9, Appendix 2. Generally, higher Classifications and a higher Number of Reviews align with a higher overall User-Feedback Pleez Rating.

## 5. Design and Development

After completing all preliminary processes for the model, a final step was taken to ensure the data's suitability for fitting a multiple linear regression model. Recognizing that outliers in specific attributes could impact the model, the Robust Scaler from the sci-kit-learn library was applied to the features. The Robust Scaler removes the median and scales the data according to the quantile range (Sklearn.preprocessing.RobustScaler - Scikit-Learn Documentation, n.d.). This scaling technique was employed to enhance the model's resilience to outliers, a crucial consideration for the accuracy and stability of the multiple linear regression analysis.

Using a linear regression model was grounded in its simplicity and interpretability. It allowed for a direct assessment of the relationship between each feature, the selected parameters, and the target variable, the visual and organizational menu rating. Additionally, linear regression is well-suited for addressing collinearity. The coefficients derived from the model provided insight into how the target variable changes with a one-unit shift in each parameter, all while maintaining the other parameters at a constant level.

Predictions were rounded to the nearest integer, reflecting the nature of the target variable. To ensure a robust evaluation, 20% of the dataset was reserved for the test set. Consequently, the test set encompassed 29 restaurants, while the training set incorporated 112 establishments. The model's results and respective analyses are presented in the forthcoming chapter.

## 6. Interpretation of Results

### 6.1. Performance Metrics

The standard Mean Squared Error (MSE) was applied to provide a metric indicating the average squared disparity between predicted and User-Feedback Pleez Rating. Additionally, it was applied the Bounded MSE, which considers the natural range of the Pleez Rating (0 to 10),

aimed to provide a more tailored evaluation. However, adjusting the Bounded MSE did not lead to a deviation, implying that the linear regression model was well-suited for predicting Pleez Ratings within the specified limits. The Standard and Bounded MSE training error was 2.10, and the test error was 1.76. It is crucial to highlight that interpreting these results means that the reported error values are, in fact, the squared root of the values presented.

A residual analysis was conducted to enhance comprehension of the significance surrounding MSE. The residuals are calculated by subtracting the predicted value from the observed value. Figure 1 shows that the values adhere closely to a normal distribution, with the majority concentrated around zero, exhibiting a symmetrical and even dispersion. As deviations from zero increase, there is a visible reduction in the count of wrong predictions at higher residual values. Notably, both the training and testing datasets exhibit a similar distribution, implying a nonexistence of overfitting in the model. Turning attention to Figure 2, the observation that human ratings tend to avoid extremes validates that the model was trained on ratings centered in the middle of the scale. As a result, the model's predictions demonstrate a narrowing trend, aligning with the anticipated outcome. It is worthwhile to consider continuing to train the model on datasets with more consistent and varied user ratings to optimize model training.

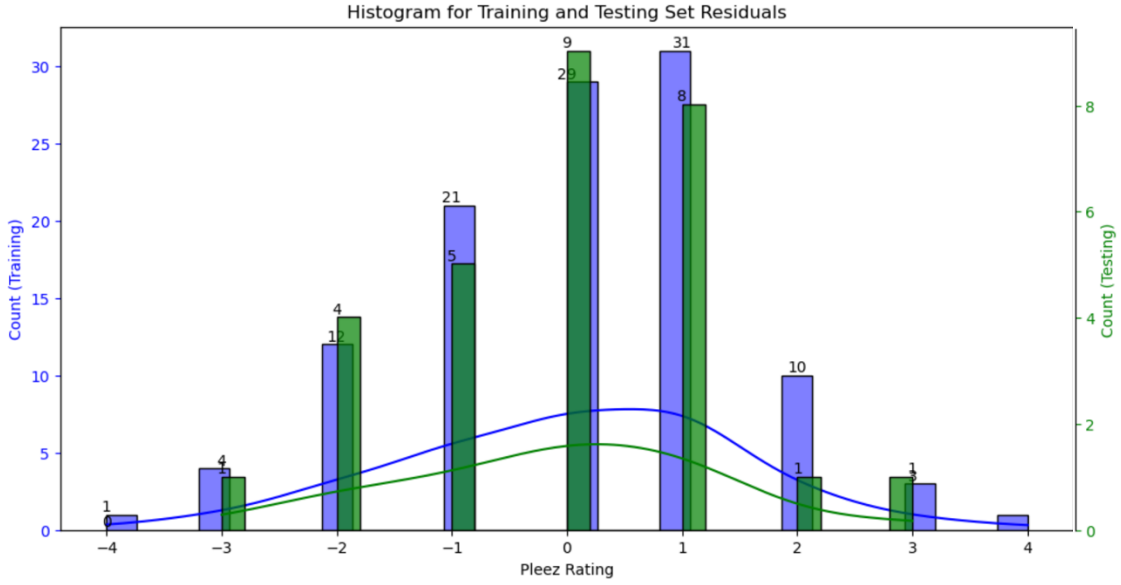



**Figure 1**: Distribution of Residuals in Training and Testing Sets

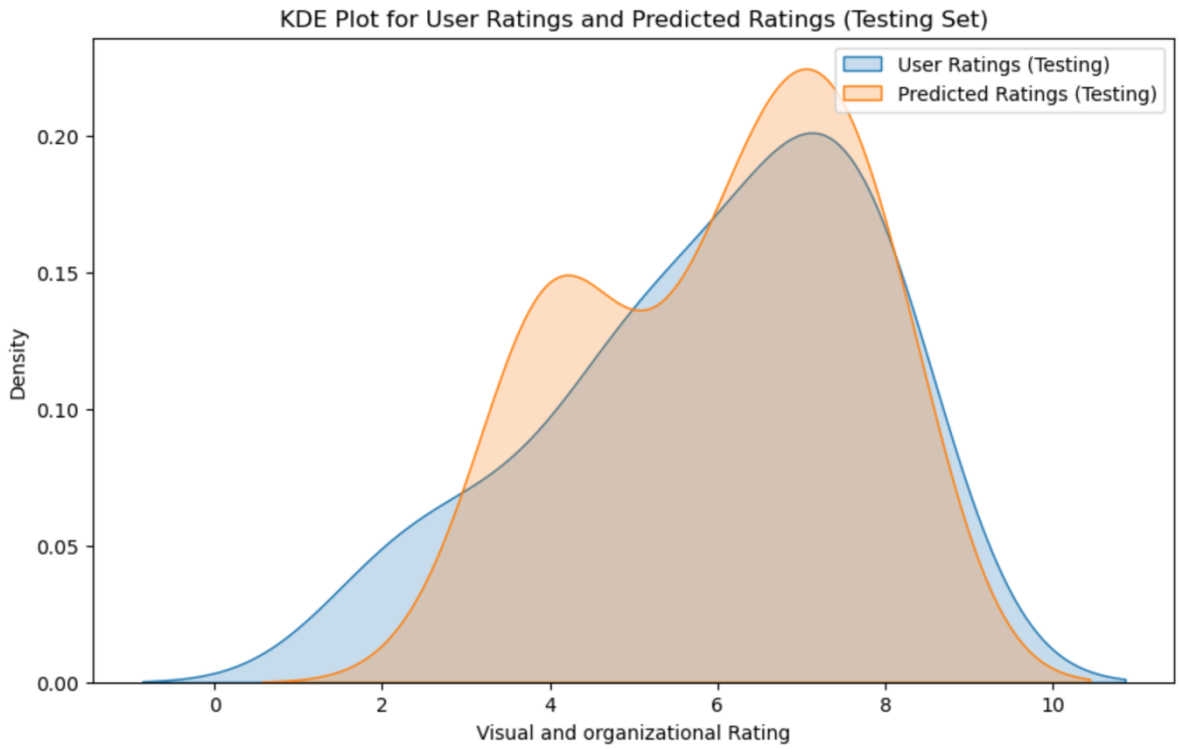



**Figure 2**: KDE for User Ratings and Predicted Scores in the Testing Set

### 6.2. Feature Importance

Examining the importance of features within the model provides insights into the parameters that significantly influence the predicted outcome. The coefficients associated with each feature clarify the strength and direction of their impact on the predicted rating. The coefficients are presented in Appendix 3, Table 5.

The dominant feature identified in this analysis is the Percentage of Menu Items with Pictures, characterized by a substantial coefficient of +1.83. This underscores a positive correlation between a higher percentage of items featuring pictures and an elevated predicted rating. Following is the Featured Items, which has a notable coefficient of +1.04, indicating a positive impact on the predicted rating when the menu has a Featured Items category. Moreover, the Number of Menu Items and Classification also exhibit positive relationships with coefficients of +0.79 and +0.54, respectively. These outcomes suggest that a more significant Number of

Menu Items and a higher Classification positively contribute to the User-Feedback Pleez Rating. On the contrary, the Number of Menu Categories feature negatively impacts the predicted rating, as reflected by its coefficient of -0.14. This implies that an increased number of categories adversely affects the User-Feedback Pleez Rating. Similarly, the Number of Drinks also demonstrates a negative relationship with a coefficient of -0.42. This observation suggests that a high number of drinks have a negative impact on the overall menu. The Percentage of Drinks Without Pictures demonstrates a coefficient of + 0.51 and the Percentage of Menu Drinks Without Description demonstrates a coefficient of +0.11. The positive coefficients underscore that a high number of drinks has a low impact on the User-Feedback Rating. This is evident when contrasted with the significant influence of the Percentage of Menu Items with Pictures, identified as the feature with the highest coefficient. The positive coefficient of +0.13 for the Percentage of Menu Items with Descriptions suggests that an increased number of products with descriptions is associated with higher User-Feedback Pleez Ratings. Other features, such as Nº Desserts, Nº of Reviews, and the Percentage of Desserts Without Description, all exhibit very low coefficients, indicating that the presence or absence of these elements has an insignificant impact on the overall rating.

## 7. Recommendations for Future Steps

### 7.1. Recommendation Regarding the Effectiveness of the Pleez Rating Project

To address question number one of this thesis: How does the adoption of such a rating system impact the likelihood of customers taking action to make changes to their menus? As mentioned previously, the initial steps involved dispatching emails to measure the efficacy of the "Call to Action". The email is available for reference in Appendix 4. Given the status of four clients, a decision was made to selectively email only six of the ten restaurants in the test phase to assess whether this approach effectively prompts a "call to action." Only four demonstrated

engagement during this preliminary stage by opening the email. However, none of them initiated communication with their respective account managers to request modifications to their menus. While this indicates a degree of disinterest, a more extensive sample is imperative during the testing phase to assess the effectiveness of the "call to action" precisely.

### 7.2. Recommendation Regarding the Calculation Method

To address question two of this thesis: What visual and organizational aspects of a menu influence consumers' perceptions? The final step in this process involves comparing the initial weights used in the calculation method with the feature importance identified by the model and the results of the supplementary surveys. This initial comparison will rely on the absolute values of the coefficients derived from the model, measuring the magnitude of a parameter's effect on the outcome. Additionally, the sign of the coefficient will indicate the direction of that effect, assisting in the evaluation of the criteria used in score calculation for each parameter coupled with the exploratory data analysis. To provide a comprehensive understanding, the results compilation of the two surveys employed in this thesis for assessing the significance of parameters and the initial and final weights applied in the classification method is provided below. The optimized calculation method is detailed in Appendix 5, table 6.

| | Feature Imp. | Initial Weights | Supp. Survey | Final Weights |
|---|---|---|---|---|
| Scale | - | 0% - 100% | 0 – 10 | 0% - 100% |
| Percentage of Menu Items With Picture | 1.83 | 21,67% | 10 | 20% |
| Featured Items | 1.04 | | | 10% |
| Number of Menu Items | 0.79 | 21,67% | 6 | 15% |
| Classification | 0.54 | 21,67% | 8 | 15% |

| | | | | |
|---|---|---|---|---|
| Percentage of Menu Drinks Without Picture | 0.51 | | | |
| Number of Menu Drinks | - 0.42 | | | |
| Number of Menu Categories | - 0.14 | 10% | | 10% |
| Percentage of Menu Items with Description | 0.13 | 15% | 5 | 10% |
| Percentage of Menu Drinks Without Description | 0.11 | | | |
| Percentage of Menu Deserts Without Description | 0.073 | | | |
| Number of Reviews | -0.030 | 10% | | 5% |
| Number of Menu Deserts | -0.025 | | | |
| Impact of Visually Appealing Images | | | 10 | 10% |
| Impact of Description Content | | | 7 | 5% |

**Table 1**: Overview of Survey Results and Initial and Final Weights in Classification Method

### 7.2.1. Comparison with Feature Importance Extracted by the Model

Through an in-depth examination of the calculation methodology employed in the first phase of the Pleez Rating, three categories have significant weight in the calculated Pleez Rating: Number of Menu Items, Percentage of Menu Items with Pictures, and Classification. In the second phase of the Pleez Rating, the development of the linear regression model and coefficients assigned to each feature further validate this weight, as the three elements are among the four features with the highest impact on the overall User-Feedback Pleez Rating. Notably, the model also assigns a high coefficient to the presence of the Featured Item category in the menu, indicating its potential as a crucial parameter in the classification process. However, a noteworthy deviation arises as the model emphasizes the relative insignificance of

the Number of Reviews from the User-Feedback Pleez Rating. This contrasts with its equal weight with the Number of Menu Categories in the Calculated Pleez Rating. Also, contrary to the Calculated Pleez Rating, the Number of Menu Categories carries a similar importance to the Percentage of Menu Items with Descriptions. This is noteworthy because, in the Calculated Pleez Rating, the Percentage of Menu Items with Descriptions carries a higher weight.

Consequently, to enhance the classification method, it is recommended that Pleez consider including a parameter for restaurants featuring the "Featured Items" category in their menus. Furthermore, a strategic adjustment involves diminishing the emphasis on the Number of Menu Items with Descriptions in the calculation method and standardizing with the Number of Menu Categories. Additionally, it would be recommended to reduce the weight assigned to the Number of Reviews.

### 7.2.2. Comparison with the Supplementary Survey

It is essential to acknowledge that what individuals express as their values or priorities in response to direct questions may not always align with their actual behaviors or choices when confronted with specific options. This phenomenon is called the "attitude-behavior gap" (Shaw et al., 2016). Therefore, while insights from the supplementary survey contribute valuable information, careful consideration of this gap is necessary.

Upon conducting an analysis of the responses from 10 selected surveys, it was observed that for the category Percentage of Menu Items with Picture, the model's predictions aligned with the perceived importance reported by participants. Interestingly, the Classification parameter surpassed the perceived significance of the Number of Menu Items. In contrast, the feature importance exhibited a higher positive coefficient for the Number of Menu Items than the Classification, indicating a disparity between participant perceptions and the model's emphasis on these parameters. Percentage Items with Description surfaced as one of the features with the

lowest significance according to the model's coefficients. Participants assigned it an importance rating of 5, being the least essential feature in the survey.

Consequently, to enhance the calculation method, after divergent prioritization of the Number of Menu Items and Classification by both the model and additional survey results, it is recommended to maintain equal weights in the calculation method. Furthermore, the need to reduce the weight assigned to the Number of Menu Items with Description is emphasized.

The two parameters, namely the Impact of Visually Appealing Images and the Impact of Description Content, which were not in the model but were asked in the additional surveys, emerged as essential features to be incorporated into the calculation. This revelation stemmed from participants assigning a significance rating of 10 to the impact of visually appealing images and 7 to the importance of description content. As such, it is crucial to integrate a supplementary model capable of predicting the appeal of both images and descriptions.

### 7.2.3. Refinement Proposals for the Weights in the Calculation Methodology

Based on the preceding analysis, the recommendation for the weights of the Pleez Rating calculation method is as follows - Number of Menu Items: 15%; Number of Menu Categories: 10%; Percentage of Items with Pictures: 20%; Percentage of Items with Description: 10%; Classification: 15%; Number of Reviews: 5%. For the newly proposed features, the recommended weight suggestions are as follows - Featured Items: 10%; Percentage of Visually Appealing Images: 10%; Percentage of Appealing Descriptions: 5%.

### 7.2.4. Refinement Proposals for the Score Calculation Criteria

Moreover, after the Descriptive Statistics and Exploratory Data Analysis and Feature Importance, not only the Criteria to be included in the calculation and respective weight should be adjusted but also the score calculation criteria.

The current criteria for the Pleez Rating involve thresholds for specific parameters to achieve a score of 10, such as the number of categories and items, which may not fully align with observed correlations. A suggested refinement would be to reconsider the scoring criteria for the number of categories, possibly rewarding establishments with a higher number of categories once the dataset indicates that establishments with a User-Feedback Pleez Rating of 10, 9, and 8 boast an average of 16.00, 7.29, and 7.71 categories. However, the model assigned a coefficient of -0.14, signifying that the rating decreases by 0.14 with each additional normalized category, assuming all other variables remain constant and considering the "ceteris paribus" principle. Therefore, further studies should be conducted to refine the scoring criteria concerning the number of categories.

The scoring methodology for evaluating the number of items could be modified to encompass a more extensive selection, given that establishments receiving high User-Feedback Pleez Rating present more items on average. Establishments achieving a User-Feedback Pleez Rating of 10, 9 and 8 showcase averages of 105, 40 and 36 or the number of items, excluding drinks. This is reassured by the positive coefficient assigned by the model of +0.79. Transitioning into the Calculated Pleez rating, a restaurant aiming for a 10 should ideally offer a menu featuring 25 to 30 items. Therefore, considering the limited instances with a User-Feedback Pleez Rating of 10, assigning a score of 10 is recommended when the menu comprises a number of items between 35 and 40.

Concerning the percentages of items with pictures and descriptions, the scoring calculation criteria should consistently target 100%, with adjustments to each criterion's weight. Similarly, the classification and Nº of reviews with a score of 10 in these criteria should be contingent on achieving a classification of 5 and a number of reviews of 500, with adjustments also made to weight and not the score calculation criteria. Concerning the newly incorporated features, the recommendation would be a score of 10 if the category Featured Items is included and a rating

of 0 if it is not in the Featured Items parameter. For the added criteria related to Visually Appealing Images and Descriptions, the goal should also consistently aim for 100% of images and descriptions deemed appealing by a forthcoming model.

## 8. Challenges and Limitations

The primary challenge in developing this thesis was data acquisition. Despite Uber Eats having 12 000 restaurants in 2022, the study focused on 141 restaurants and 50 participants. Although this model provides valuable insights and comparative analysis for the Pleez Rating calculation method in use, it is essential to acknowledge the need for a more extensive dataset for generalization. Additionally, during the menu evaluation surveys, respondents were explicitly instructed to disregard their personal food preferences or pricing considerations. Nevertheless, it remains challenging to definitively assert the absence of any bias in the rating of the menus. Finally, due to time constraints in the research, the optimal number of Pleez clients receiving the Pleez Rating still needs to be achieved, requiring additional efforts to fully understand the impact of the Pleez Rating in prompting clients to take action.

## 9. Conclusion

An extensive data collection process encompassing diverse restaurant data points delineated the methodology implemented to assess the project's impact and enhance the Pleez Rating calculation method. This method provided an initial insight into the project's impact and facilitated the identification of features that the calculation methodology should incorporate. Acknowledging the efficacy of regression analysis as a decision support tool, this methodology, fortified by a comprehensive analytical approach, allowed for overcoming primary challenges and limitations. Nevertheless, it is imperative to persist in testing the project's impact and acquire a more comprehensive dataset for broader generalization.

## 10. References

Ahuja, K., Chandra, V., Lord, V., and Peens, C. (2021). Ordering in: The rapid evolution of food delivery. [online] McKinsey & Company. Available at: https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/ordering-in-the-rapid-evolution-of-food-delivery [Accessed 18 Oct. 2023].

Chatzidakis, A., Carrington, M., Dhir, A., Goworek, H., & Shaw, D. (2016, July 28). Explaining the Attitude-Behaviour Gap: why consumers say one thing but do another | Ethical Trading Initiative. Available at: https://www.ethicaltrade.org/insights/blog/explaining-attitude-behaviour-gap-why-consumers-say-one-thing-do-another [Accessed 1 Dec. 2023].

da Queixa, P. (2022). 5 anos de Uber Eats em Portugal: o que pensam os consumidores da plataforma? [online] portaldaqueixa.com. Available at: https://portaldaqueixa.com/news/5-anos-de-uber-eats-em-portugal-o-que-pensam-os-consumidores-da-plataforma [Accessed 18 Oct. 2023].

data, pleez-You cook food W. cook and Yomoc (2023a). Delivery's Challenges and Opportunities for Small Restaurants: Strategies for Success. [online] trypleez.com. Available at: https://trypleez.com/en/blog/deliverys-challenges-and-opportunities-for-small-restaurants-strategies-for-success/ [Accessed 19 Oct. 2023].

data, pleez-You cook food W. cook and Yomoc (2023b). Sales through delivery platforms, profitable or loss? [online] trypleez.com. Available at: https://trypleez.com/en/blog/sales-through-delivery-platforms-profitable-or-loss/ [Accessed 18 Oct. 2023].

Mishra, S. (2020). Why are food images so important for your digital menu? [online] Medium. Available at: https://fabled.medium.com/why-are-food-images-so-important-for-your-digital-menu-3ddad9241332 [Accessed 10 Nov. 2023].

Ot, A. (2021, May 22). What is Data Scraping? | Best Web Scraping Strategies for Analytics. Datamation. Available at: htt
ps://www.datamation.com/big-data/data-scraping/ [Accessed 10 Nov. 2023].

Ray, A., Bala, P.K., Dhir, A., and Kaur, P. (2019). Why Do People Use Food Delivery Apps (FDA)? A Uses and Gratification Theory Perspective. Journal of Retailing and Consumer Services, [online] 51, pp.221–230. Available at: doi:https://doi.org/10.1016/j.jretconser.2019.05.025 [Accessed 18 Oct. 2023].

RobustScaler - scikit-learn 0.24.2 documentation. (n.d.). Scikit Learn.org. Available at: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.RobustScaler.html [Accessed 1 Dec. 2023].

Saraswat, P.S. (2023). The Crucial Role of Dispute Management and Data Consolidation for Restaurants. [online] QSR Magazine. Available at: https://www.qsrmagazine.com/growth/fast-casual/the-crucial-role-of-dispute-management-and-data-consolidation-for-restaurants/ [Accessed 19 Oct. 2023].

Uyanık, G. K., & Güler, N. (2013). A Study on Multiple Linear Regression Analysis. Procedia-Social and Behavioral Sciences, 106(1), 234–240. Available at: https://doi.org/10.1016/j.sbspro.2013.12.027 [Accessed 21 Oct. 2023].

# 11. Appendix

## 11.1. Appendix 1 – Data Collection Variables

| **1st Phase Pleez Rating** | **Score Calculation Criteria** |
|---|---|
| Number of Menu Items<br><br>**Weight:** 21,67% | If the menu falls within the range of **25 to 30 items**, inclusive, it is assigned a **score of 10**. Additionally, for every **2 items below** the minimum threshold of 25, the score **decreases by 1** until it reaches 0. Similarly, for every **2 items above** the maximum threshold of 30, the score **decreases by 1** until it reaches 0. |
| Number of Menu Categories<br><br>**Weight:** 10% | If a menu comprises exactly **6 categories**, it receives a **score of 10**. Moreover, for every category **below** the threshold of **6**, the score **decreases** by **-2** until it reaches 0. Similarly, for every category **above** the threshold of 6, the score **decreases** by **-2** until it reaches 0. |
| Number of Menu Items with Pictures<br><br>**Weight:** 21,67% | If a menu includes pictures for **100%** of its items, it is assigned a **score of 10**. Additionally, for every **10 percentage points below 100%** in terms of items with pictures, the score **decreases** by **-1** until it reaches 0. |
| Number of Menu Items with Descriptions<br><br>**Weight:** 15% | If a menu provides descriptions for **100%** of its items, it receives a **score of 10**. Furthermore, for every **10 percentage points below 100%** in terms of items with descriptions, the score **decreases** by **-1** until it reaches 0. |
| Classification<br><br>**Weight:** 21,67% | If the rating on UberEats is **5**, the **score is 10**. If the rating is **1**, the score is **0**. For ratings between 1 and 5, the score is calculated as the classification value multiplied by 2 (10/5), rounded to the nearest whole number. |
| Number of Reviews<br>**Weight:** 10% | If a restaurant has **500** or more reviews, it is assigned a score of **10**. Additionally, for **every 50** reviews below the threshold of 500, the score **decreases by -1** until it reaches 0. |

**Table 2:** Variables 1st Phase Pleez Rating

**2nd Phase Pleez Rating**

| |
|---|
| Featured Items |
| Number of Menu Items |
| Number of Menu Drinks |
| Number of Menu Deserts |
| Number of Menu Categories |
| Percentage of Menu Items With Picture |
| Percentage of Menu Drinks Without Picture |
| Percentage of Menu Items with Description |
| Percentage of Menu Drinks Without Description |
| Percentage of Menu Deserts Without Description |
| Classification |
| Number of Reviews |

**Table 3:** Variables 2nd Phase Pleez Rating

**Supplementary Survey**

| |
|---|
| Number of Menu Items |
| Percentage of Menu Items With Picture |
| Visual Appealing Images |
| Percentage of Menu Items with Description |
| Description Content |
| Classification |

**Table 4:** Variables Supplementary Survey

## 11.2. Appendix 2 –Descriptive Statistics and Exploratory Data Analysis

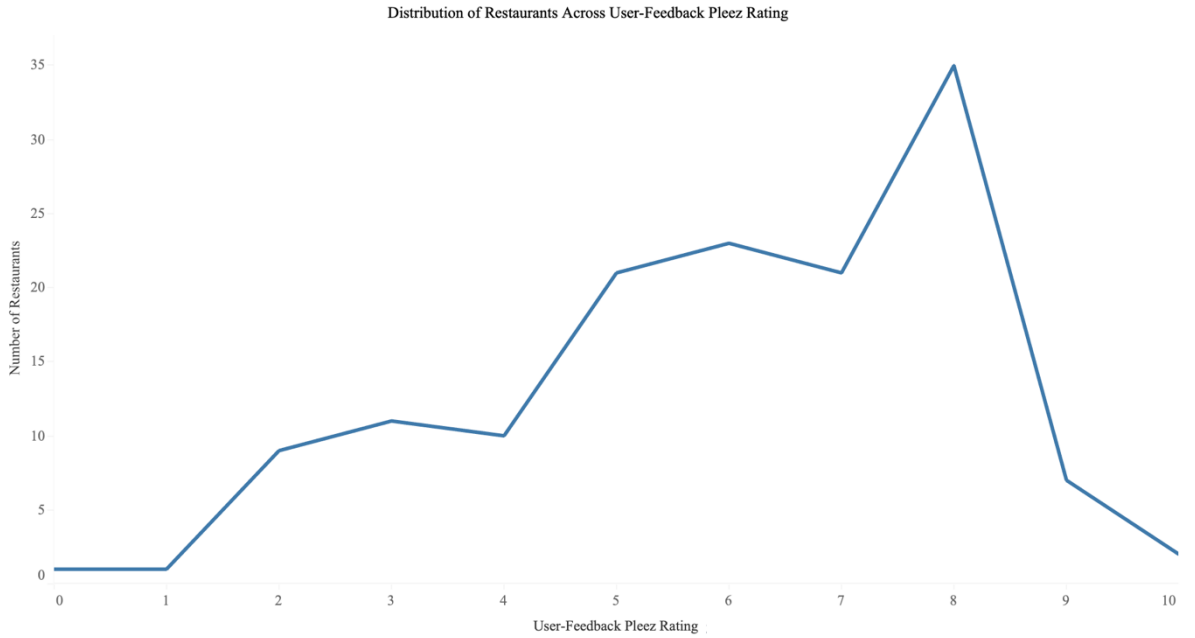



**Figure 3**: Distribution of Restaurants Across User-Feedback Pleez Rating

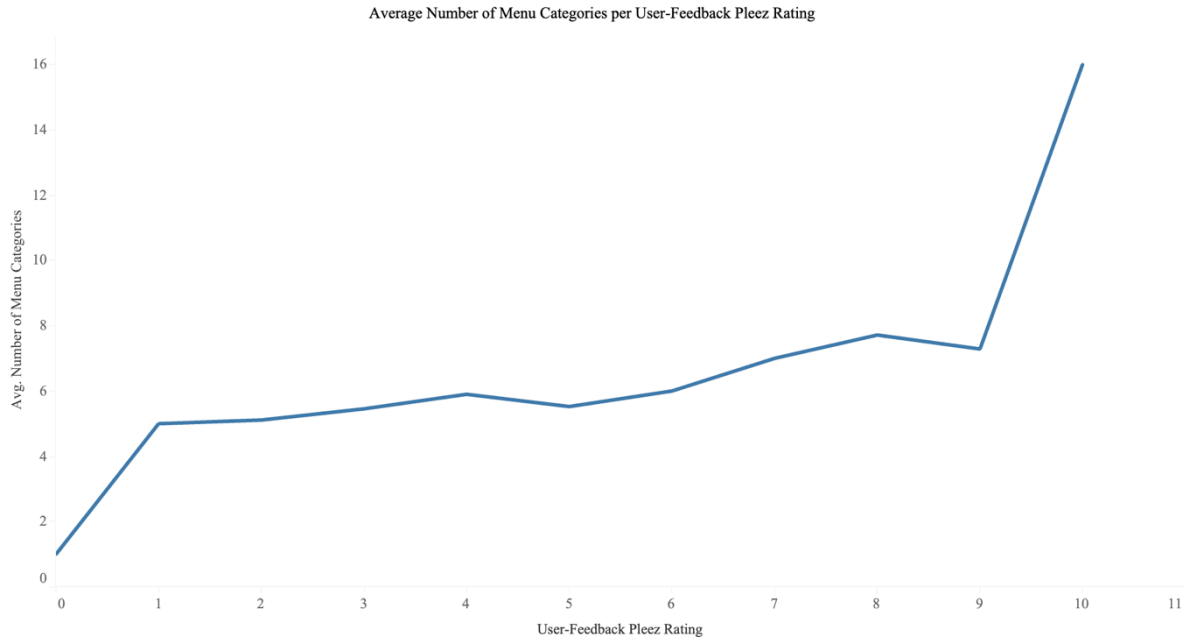



**Figure 4:** Average Number of Menu Categories per User-Feedback Pleez Rating

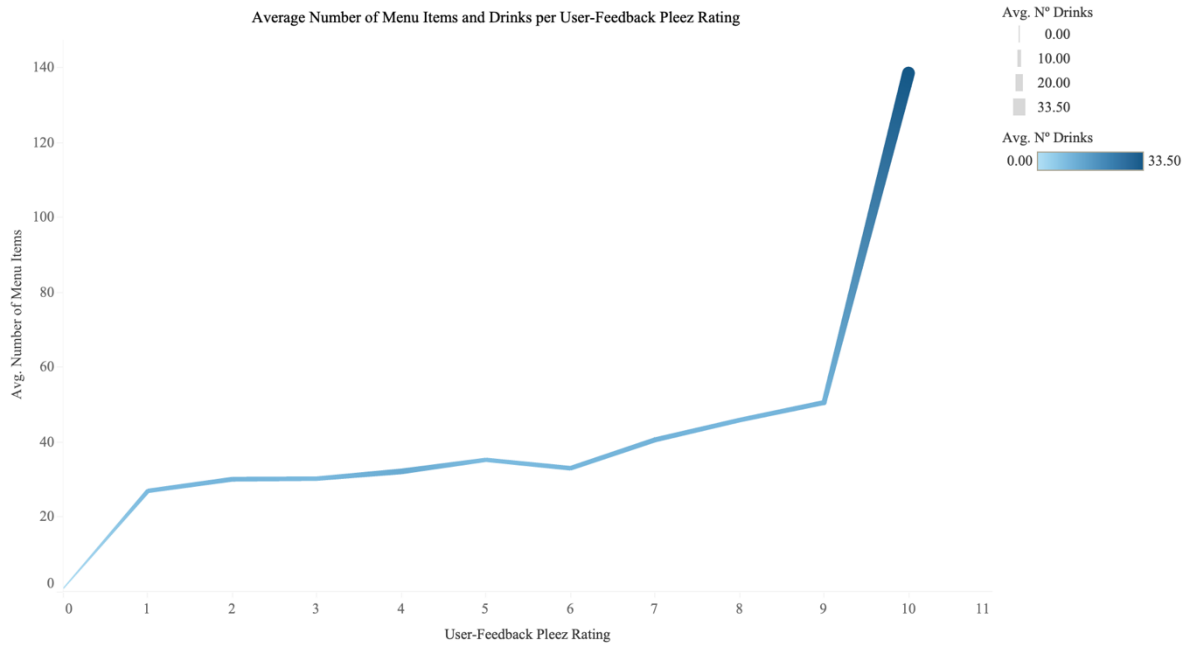



**Figure 5**: Average Number of Menu Items and Drinks per User-Feedback Pleez Rating

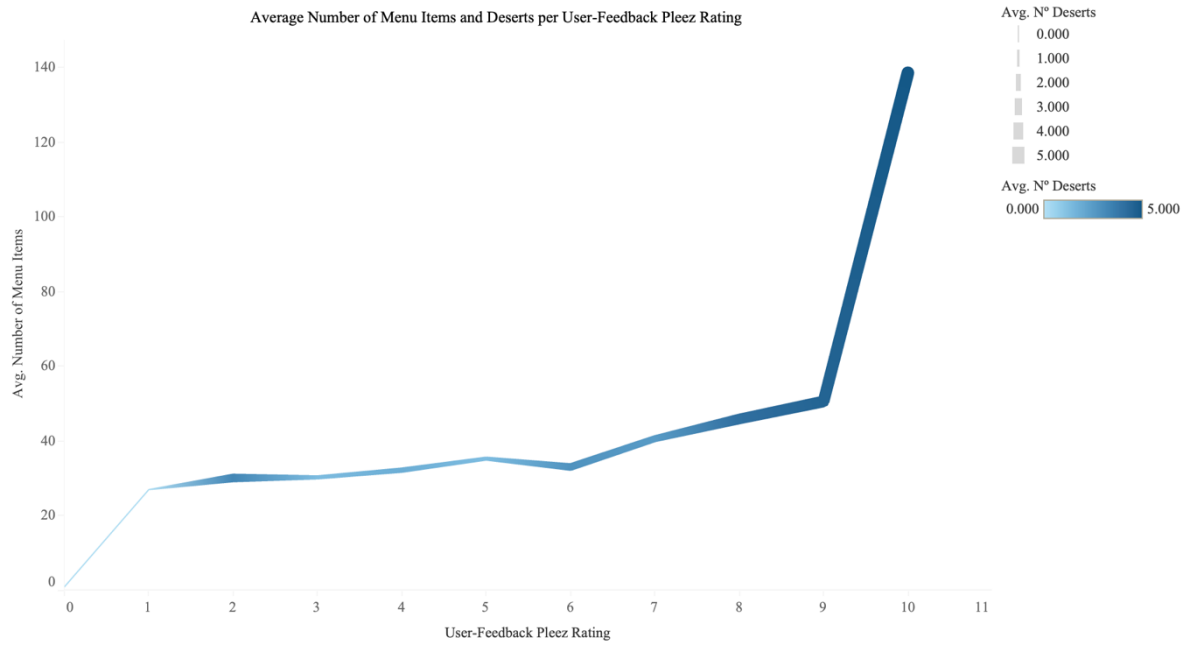



**Figure 6**: Average Number of Menu Items and Deserts per User-Feedback Pleez Rating

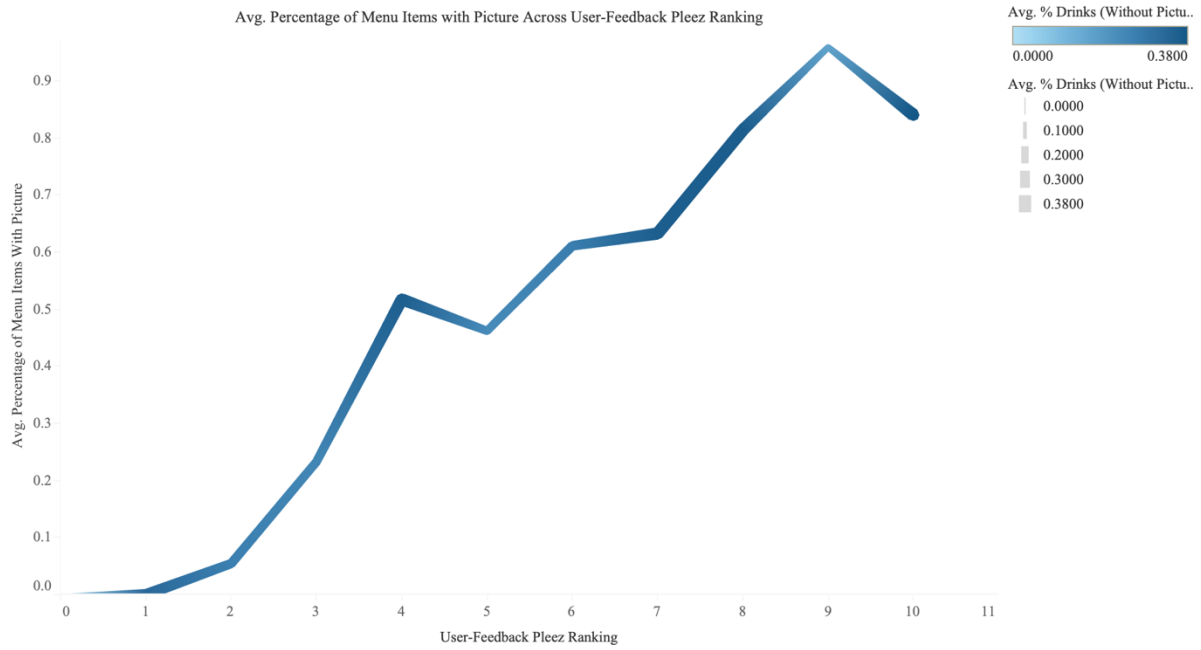



**Figure 7:** Avg. Percentage of Menu Items with Picture Across User-Feedback Pleez Rating

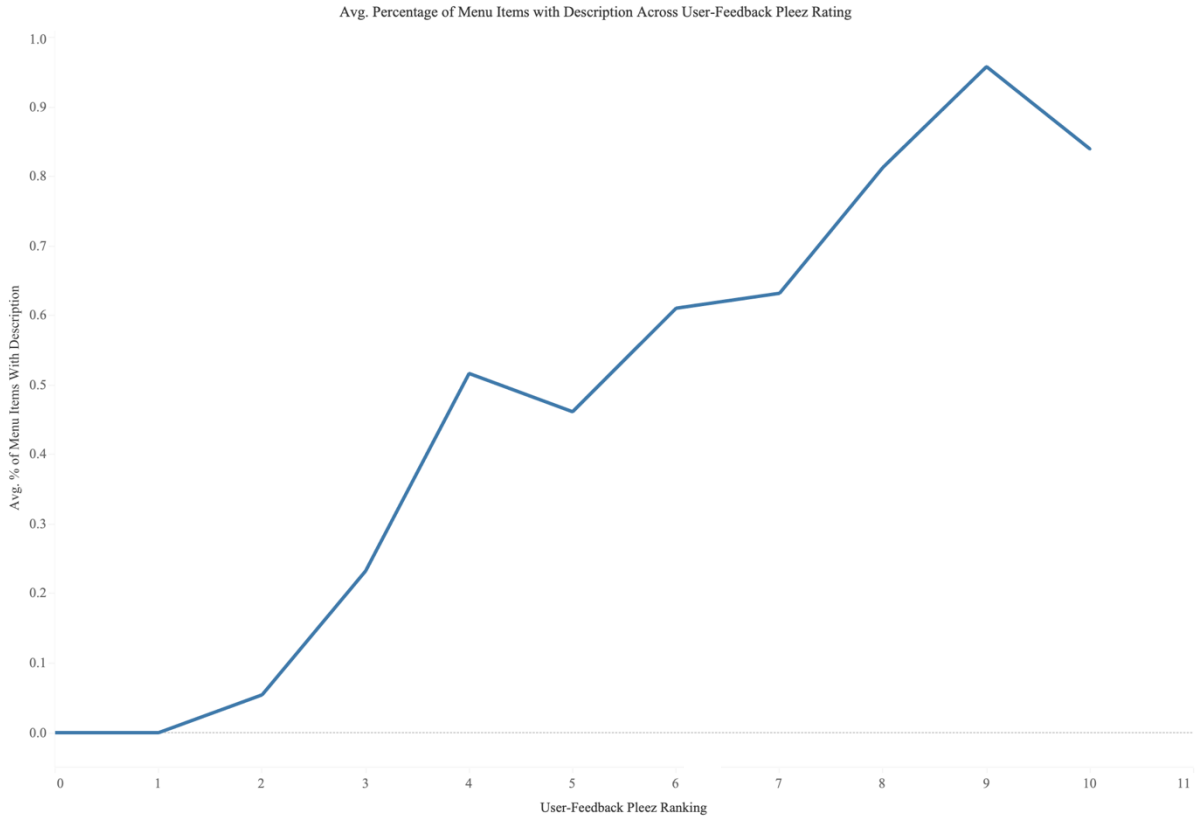



**Figure 8:** Avg. % of Menu Items with Descriptions Across User-Feedback Pleez Rating

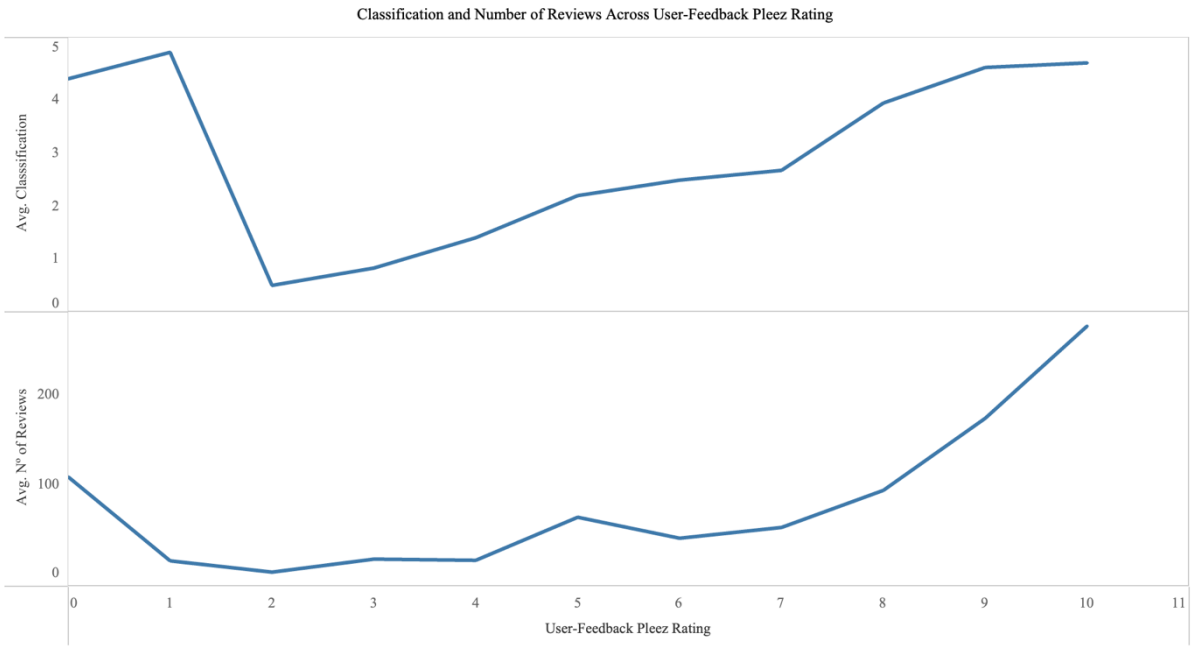



**Figure 9**: Classification and Number of Reviews Across User-Feedback Pleez Rating

### 11.3. Appendix 3 – Feature Importance

**Model Equation:** $y = 6.61 + [1.04 \times F1] + [0.79 \times F2] - [0.42 \times F3] - [0.03 \times F4] - [0.14 \times F5] + [1.83 \times F6] + [0.51 \times F7] + [0.13 \times F8] + [0.11 \times F9] + [0.07 \times F10] + [0.54 \times F11] - [0.025 \times F12]$.

| Feature | Importance |
|---|---|
| Percentage of Menu Items With Picture | **1.83** |
| Featured Items | **1.04** |
| Number of Menu Items | **0.79** |
| Classification | **0.54** |
| Percentage of Menu Drinks Without Picture | **0.51** |
| Number of Menu Drinks | **-0.42** |
| Number of Menu Categories | **- 0.14** |
| Percentage of Menu Items with Description | **0.13** |
| Percentage of Menu Drinks Without Description | **0.11** |
| Percentage of Menu Deserts Without Description | **0.073** |
| Number of Reviews | **-0.030** |
| Number of Menu Deserts | **-0.025** |

**Table 5:** Feature Importance

**Linear Model with Normalized Data - RobustScaller**

```
X = data.iloc[:, 1:13]
y = data.iloc[:, 13]

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20, random_state=42)

scaler = RobustScaler()

XTrainNormalized = scaler.fit_transform(X_train)

XTestNormalized = scaler.transform(X_test)

linearModel = LinearRegression()

linearModel.fit(XTrainNormalized, y_train)

yTrainPred = linearModel.predict(XTrainNormalized)

yTestPred = linearModel.predict(XTestNormalized)

yTrainPred = yTrainPred.round().astype(int)

yTestPred = yTestPred.round().astype(int)
```

**Figure 10:** Linear Model Development Code

**Standard MSE**

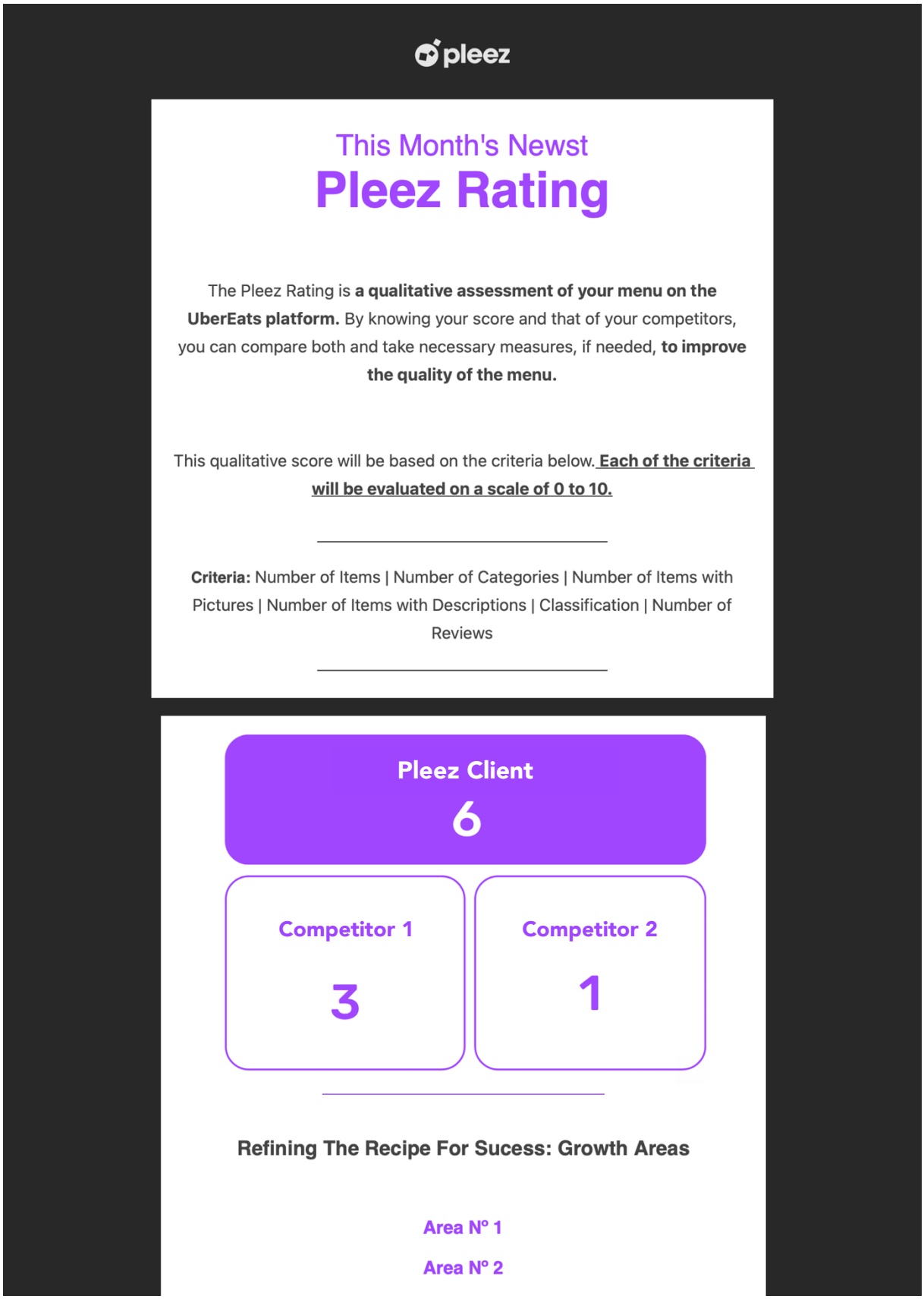
```
trainMSE = mean_squared_error(y_train, yTrainPred)

testMSE = mean_squared_error(y_test, yTestPred)

print("Training error (Mean Squared Error):", trainMSE)

print("Test error (Mean Squared Error):", testMSE)
```

**Bounded MSE**

```
def mseBounded(yActual, yPredicted):

    error = np.square(np.clip(yActual - yPredicted, -10, 10))

    return np.mean(error)

trainMSENounded = mseBounded(y_train, yTrainPred)

testMSEBounded = mseBounded(y_test, yTestPred)

print("Training error (Bounded Mean Squared Error):", trainMSENounded)

print("Test error (Bounded Mean Squared Error):", testMSEBounded)
```

**Figure 11:** Standard and Bounded MSE Code

## 11.4. Appendix 4 – Assessment of the Impact of the Pleez Rating



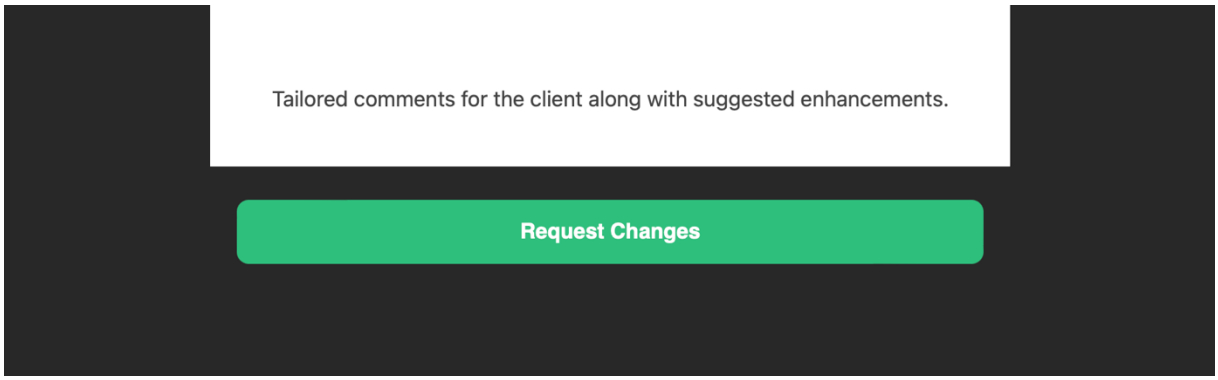



**Figure 12:** Pleez Rating Email Template

## 11.5. Appendix 5 – Consumer-Centered Calculation Method

| **Consumer-Centered Calculation Method** | **Score Calculation Criteria** |
|---|---|
| Number of Menu Items<br><br>**Weight:** 15% | If the menu falls within the range of 3**5 to 40 items**, inclusive, it is assigned a **score of 10**. Additionally, for every **2 items below** the minimum threshold of 35, the score **decreases by 1** until it reaches 0. Similarly, for every **2 items above** the maximum threshold of 40, the score **decreases by 1** until it reaches 0. |
| Number of Menu Categories<br><br>**Weight:** 10% | If a menu comprises exactly **6 categories**, it receives a **score of 10**. Moreover, for every category **below** the threshold of **6**, the score **decreases** by **-2** until it reaches 0. Similarly, for every category **above** the threshold of 6, the score **decreases** by **-2** until it reaches 0. |
| Number of Menu Items with Pictures<br><br>**Weight:** 20% | If a menu includes pictures for **100%** of its items, it is assigned a **score of 10**. Additionally, for every **10 percentage points below 100%** in terms of items with pictures, the score **decreases** by **-1** until it reaches 0. |
| Number of Menu Items with Descriptions<br><br>**Weight:** 10% | If a menu provides descriptions for **100%** of its items, it receives a **score of 10**. Furthermore, for every **10 percentage points below 100%** in terms of items with descriptions, the score **decreases** by **-1** until it reaches 0. |
| Classification<br><br>**Weight:** 15% | If the rating on Uber Eats is **5**, the **score is 10**. If the rating is **1**, the score is **0**. For ratings between 1 and 5, the score is calculated as the classification value multiplied by 2 (10/5), rounded to the nearest whole number. |
| Number of Reviews<br><br>**Weight:** 5% | If a restaurant has **500** or more reviews, it is assigned a score of **10**. Additionally, for **every 50** reviews below the threshold of 500, the score **decreases by -1** until it reaches 0. |

| | |
|---|---|
| Featured Items<br><br>**Weight:** 10% | If a menu has the category "Featured Items", it receives a **score of 10**. If it doesn't' have receives a score of 0. |
| Percentage of Visually Appealing Images<br><br>**Weight:** 10% | If a menu provides visually appealing images for **100%** of its items, it receives a **score of 10**. Furthermore, for every **10 percentage points below 100%** in terms of visually appealing images, the score **decreases** by **-1** until it reaches 0. |
| Percentage of Appealing Descriptions<br><br>**Weight:** 5% | If a menu provides appealing descriptions for **100%** of its items, it receives a **score of 10**. Furthermore, for every **10 percentage points below 100%** in terms of appealing descriptions, the score **decreases** by **-1** until it reaches 0. |

**Table 6:** Consumer-Centered Calculation Method

## 11.6. Appendix 6 – Conducted Survey to Gather Menu Ratings

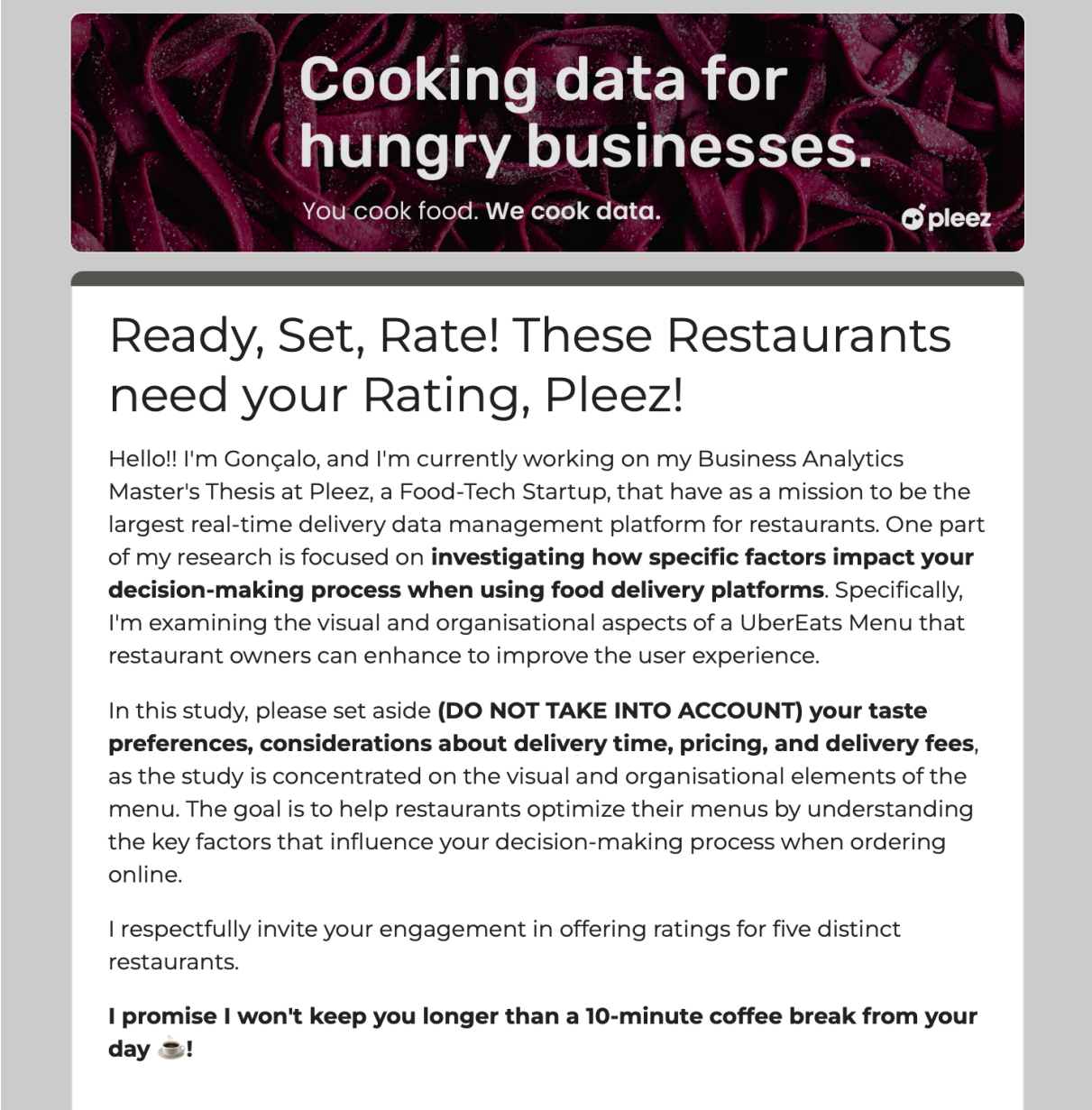



**Figure 13:** Pleez Rating Survey Template - Cover Page

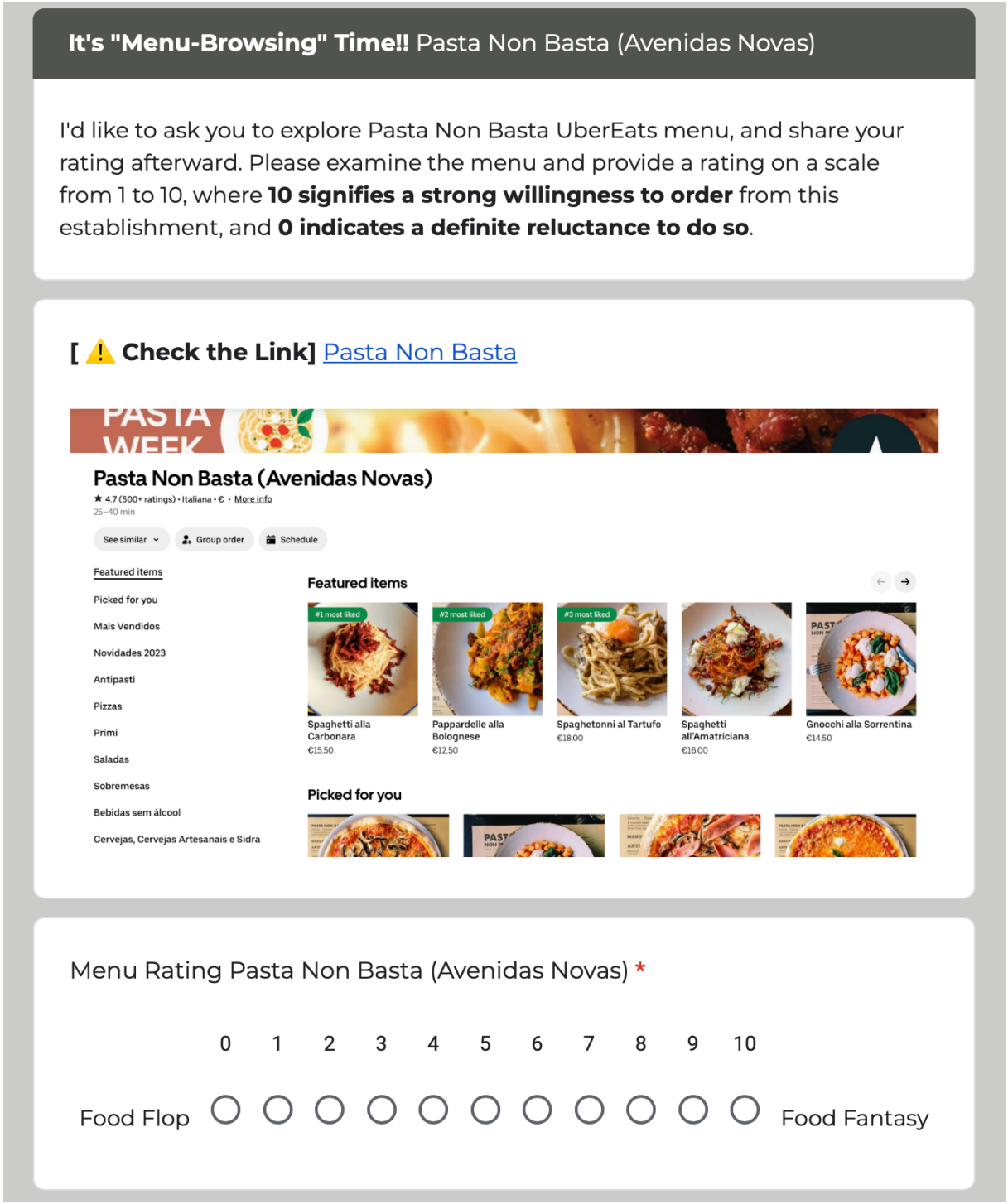



**Figure 14:** Pleez Rating Survey Template - Restaurant Rating Page