Optimizing Customer Satisfaction for E-Commerce: A Logistics Approach

Team 20: Hongsen Qin, Michael Huang, Su Min (Olivia) Lee, Alycia Lee

1 Executive Summary

Due to the rapid growth and widespread adoption of the Internet since its inception, the booming e-commerce market has transformed the landscape of the retail industry. With its razor thin margins[1] and competitive landscape of hundreds of millions of sellers and countless products, we take the perspective of a seller and address a fundamental business question: how do we maximize revenues given constrained company resources?

We address this question by first looking at the source of the customer feedback via a natural language processing approach to determine the sentiment and content of customers. From our preliminary analysis, we note that specifically for Olist, but–more broadly–for general e-commerce marketplaces, delivery needs to be speedy and punctual, a realization that applies to both when competing with other sellers and physical stores. To that end, we study how logistics can improve deliveries and how they impact customer's downstream satisfaction. We study different treatment effects such as fulfilling orders earlier than promised and coordinating deliveries by leveraging fundamental causal inference techniques such as instrumental variables and propensity score matching to handle endogeneity in observational data.

Indeed, we show through our treatment groups that early arrivals and delivery coordination are highly relevant in influencing customer satisfaction. To corroborate these conclusions, we take an information theoretic approach and discover the underlying causal graphs of all of our interactions variables with the customer to map out how we might effect policies that will lead to improvements in customer satisfaction. The causal graph and the treatment effects all point to the need to improve shipping logistics.

To quantify the marginal benefit of any policy that we enact, we use random forests to select for important features in order to fit an interpretable predictive ratings model. We use a modified version of a generalized additive model by incorporating interacting terms. Furthermore, we also fit an interpretable tree classifier to model the effects of policy recommendations.

Given a clear understanding of the qualitative and quantitative links in a general e-commerce business model, we seek to produce actionables for our company, Olist. We know from studying other online sellers such as Amazon that warehouse logistics is pivotal to reducing shipping friction, as they can distribute packages faster and aggregate different items in the same order. Hence, we model the reduction in shipping time, the improvement in shipping coordination, and increase in customer satisfaction per warehouse installation.

To model financial impact of our optimized customer satisfaction, we extrapolate ordinary least squares model from literature to relate relative change in customer satisfaction to relative change in total equity. We further convert this relative change to an absolute change in USD by using the venture capital method to approximate total market cap and revenue of Olist in 2018. We follow up the economic analysis with a geographic projection of the market in the year 2025, and show that our policies have significant impact.

Finally, we combine the models and reformulate them as a mix integer programming problem to arrive at our optimal warehouse placement locations given a constrained development budget, providing the board with data driven, high impact policies.

2 Exploratory Analysis and Data Pre-processing

2.1 Data Pre-processing

Though we are not faced with significant missing numeric data, since all e-commerce numbers are recorded to arbitrary precision, we are challenged with the Portuguese review dataset since none of us knew the language. While pre-trained models exist for sentiment analysis (VADER[2]), they are typically trained on English. We implement a caching API to interact with Google translate web, and batch-translate our data into English.

We take advantage of an existing NLP model that could also process emojis, but for the other NLP model, we find that non alphanumeric text often adds noise to our sentiment aggregation, and thus we also devise pipelines to remove them.

Later in this work, we source GDP, HDI, and population data to corroborate the Olist dataset in a geographically weighted regression. We aggregate the data from a human readable format to CSVs.

2.2 Exploratory Geoplots

We begin looking at the Olist dataset due to its completeness of factors that describes the entire e-commerce transaction. We are interested in characterizing our revenue sources and how we can optimize them.

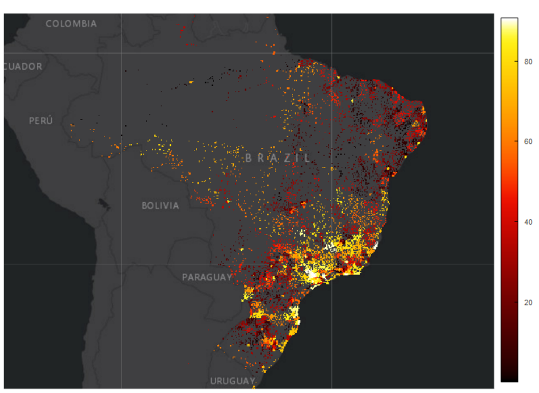

Figure 1: Geoplot of Revenue in Thousands (in Brazil Real)

we visualize sellers' cumulative revenue 1 earned by grouping together sum of order prices by zip code prefix, and observe that the southern and southeast areas of Brazil had the highest levels of revenue. This makes sense as the largest cities in Brazil by population, such as São Paulo, Rio de Janeiro, and Belo Horizonte are located in those regions. Since our revenue comes from such a geographically diverse range, we wanted to see whether there exists common indicators that people were satisfied with our service.

2.3 NLP of Customer Reviews

Following the business intuition that when customers are satisfied or dissatisfied, they will express their beliefs in feedback reviews, we use natural language processing techniques to analyze the Olist[3] dataset, which contains over forty thousand reviews from after sale surveys.

We process the reviews with state-of-the-art Valence Aware Dictionary and sEntiment Reasoner (VADER[2]) to arrive at sentiment scores for each review. VADER is a rule-based model that works well on analyzing text from social media (i.e. short clauses, emojis, capitalizations, etc) and generalizes well to multiple domains. We obtain a compound sentiment score that classifies sentiment as either good, bad, or neutral, and we use this score to categorize reviews into positive and negative reviews.

Filtering on sentiment, we analyze reviews using collocation, a well-known method in literature for text categorization[4] and summarization[5]. Taking the N-Gram frequency analysis approach outlined in literature[5], we clean the translated English texts and preform an N-gram frequency analysis on positive / negative reviews.

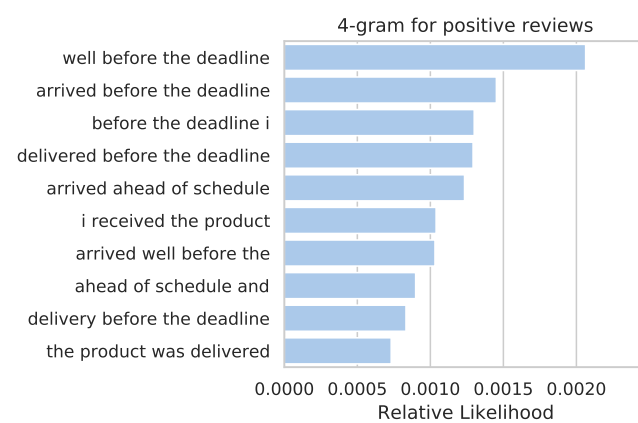

Figure 2: 4-Gram frequency table for all positive reviews. Note the overwhelming mention of delivery and other logistics related praise

We reason that since our customers seem to care so much about delivery, delivery-related coordination issues must also affect their reviews.

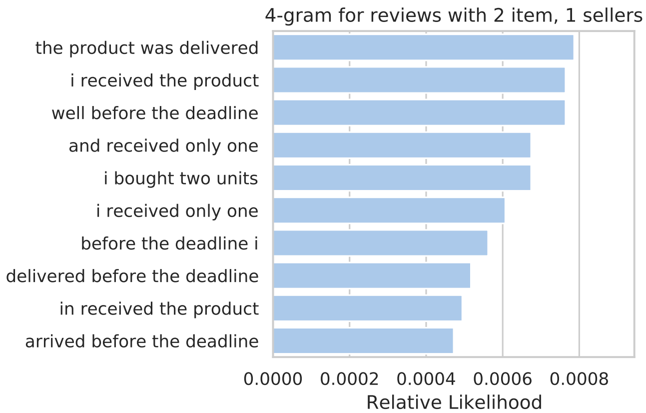

Figure 3: 4-Gram frequency table for all positive reviews. Note the mentioning of shipping issues

Figure 4: 4-Gram frequency table for all negative reviews. Note the observations about buying 2 but only getting 1

To that end, we observe that orders with 2 items tend to get split up, and that frustrates our customers. Furthermore, the issue complicates with more than 1 seller to further complicate coordination. This suggests that delivery logistics plays a significant role in customer satisfaction.

3 Causal Discovery

3.1 Causal Treatment Effects

From insights provided by our NLP analysis, specifically from the review text from the Olist data, we see that shipping logistics has a significant impact on customer satisfaction. To make these observations more rigorous, we estimate the causal effects of different operational/logistic "treatments" on customer satisfaction. Specifically, we investigate two different logistics-based treatments which were inspired from our NLP analysis: 1. arrival of orders before promise date and 2. coordination of order arrival. The former considers how the number of days an order arrives ahead of schedule improves customer satisfaction. The latter considers whether coordinating the arrival of items of the same order to arrive at the same time impacts customer satisfaction.

3.1.1 Early Order Arrival Effect

To investigate the effect of orders arriving early, we estimate the effect on review score for each day an item arrives early using linear regression. To handle endogeneity issues, we utilize an instrumental variable (IV) approach based on the day-of-week of a customers order. For our day-of-week IV to be valid, we see it must satisfy two criteria: (i) It causes variation in the treatment variable (review score); (ii) It does not have a direct effect on the outcome variable, only indirectly through the treatment variable (days early). We see this approach and choice of IV is common in operations literature[6,7]. In fact in [7], the authors work in a similar setting, that of e-commerce and logistics, and apply it to vary delivery time when the response variable is a review score. The intuition for why the day-of-week satisfies IV assumptions comes from the fact that customers do not consider and potentially do not remember the day of the week they ordered their package. However, if it was ordered on a later day of the week, delays may be caused by orders not getting shipped out before the weekend.

To implement the IV approach, we use two-stage least squares (2SLS) regression analysis. In the first stage, we fit a linear model for days early (DE),

$$DE_i = \hat{\beta} X_i + \hat{\gamma} C_i + \hat{\xi} S_i + \hat{\lambda} Z_i + \nu_i \qquad (1)$$

where i represents the i^{th} order. Here C_i are customer side controls like location and payment type. S_i are seller side controls like product type, seller location, and the number of different sellers. X_i are other general controls like time, distance between seller and customer, and other logistics related explanatory variables. Finally, $Z_i = 1$ if the order is made on Thursday, Friday, or Saturday and 0 otherwise.

In the second stage, a linear model is estimated

$$RS_i = \theta DE_i + \hat{\beta} X_i + \hat{\gamma} C_i + \hat{\xi} S_i + \varepsilon_i \qquad (2)$$

where RS_i is the review score for order i. The results can be seen in Figure 5. We see that the coefficient is positive so that the earlier the package arrives from the promise date, the higher the review score will be given by the customer.

This phenomenon is similar to the concept of under-deliver than over-promise which has been studied by marking and operations research literature[8]. We see that in [8], they study a game theoretic model that suggests how firms may choose the maximal delivery time to set customer expectations.

	(1) OLS	(2) IV OLS
Second Stage **(Review Score)**		
Days Early	0.02657***	0.04066***
	0.0006231	0.01225
	p < 2e-16	p = 0.000907
Day of Week IV	-0.00887	
	0.007654	
	p = 0.246680	
First Stage **(Days Early)**		
Day of Week IV		-0.6295***
		3.91E-02
		p < 2e-16
Num. of obs.	98843	98843
R-squared	0.1832	0.1822
Adj. R-squared	0.1681	0.1671

Figure 5: We compare the regressions with and without the IV. We see that in both cases days early is significant. Additionally, we verify that the IV causes variation in the treatment while it has not effect on the outcome in the original OLS.

3.1.2 Coordination of Order Arrival Effect

In our NLP analysis, customers commonly complain about orders not arriving or arriving late. But in some cases, we see that even when packages arrive on time customers still provide poor reviews. Clearly, there may be a variety of reasons that contribute to poor reviews scores. However, we found a significant contributor to these low scores comes from items in the same order not arriving simultaneously. Intuitively, having one package arrive early may remind customers they have to wait longer in the cases where items must be used together. One could also imagine that it builds worry that the other items in the order are lost. As Olist plans to expand (potentially from 7,000 to 100,000 sellers)[1], coordination of orders will become an increasingly bigger challenge.

We study a simple logistics solution to this problem, which is to ship items together. While the Olist data cannot tell us which orders have items shipped together, we can proxy the coordination based on whether an order comes from a single origin or from multiple origins. Looking at Figure 6, we see that a significant number of orders are multi-item and that they tend to have lower scores as the number of items increases. Simultaneously, we see in row 2 and row 4, when the number of sellers for a single order increases, the rating also drops dramatically. This provides evidence that order coordination is a potential issue.

# of Items	# of Sellers	Count	Avg. Review Score	Std. Error
1	1	88386	4.146358	0.00436
2	1	6509	3.746505	0.019433
3	1	1112	3.58723	0.048479
2	2	951	2.864353	0.054011
1	1	474	4.088608	0.053042
4	1	431	3.468677	0.080108
5	1	186	3.510753	0.120118
6	1	185	3.243243	0.12781
3	2	168	2.922619	0.132155
4	2	51	2.607843	0.241041

Figure 6: We display a subset of the scores and counts for orders sorted by the number of sellers and items. One can see that when the number of items and sellers goes up the review score decreases. Standard errors allow us to see the differences in means are significant for the top rows.

To strengthen our causal analysis, we consider utilizing propensity score matching[9] (PSM) with the number of sellers as the treatment to estimate the treatment effect on review score. The advantage of PSM is that it provides a way to reduce the bias in our estimation that comes from the likelihood of receiving the treatment. For example, in our case it could be that items sent from two locations will have slower delivery time and is more likely to be late. By constructing a control and treatment group with similar properties, we can avoid these issues.

As a first stage for propensity score matching, we construct a control group consisting of orders with 1 seller and 2 items and construct a treatment group consisting of orders with 2 sellers and 2 items. In this case, the treatment would be increasing the number of sellers from 1 to 2.

To perform propensity score matching, we first fit a logistic regression to predict the likelihood an order receives the treatment of coming from two sellers. We then utilize bipartite matching with the propensity scores as weights to construct equally sized control and treatment groups that should have similar covariates. Since we showed that being late or early affects the review score, we include covariates such as days early and predicted delivery time

[1]https://reut.rs/33VlH9Y

to factor them into the matching. The benefits of this matching can be seen in Figure 7 which shows that afterwards the first moments of key covariates become more similar. We see in the table that the treatment of increasing the number of sellers in an order from 1 to 2 decreases review scores by 0.99. This suggests that coordinating logistics can be a significant factor in customer satisfaction.

	Pre-Match		Post-Match	
	1 Sellers	2 Sellers	1 Sellers	2 Sellers
Price of Order	218.94	191.16	185.66	190.96
Freight Cost	38.95	40.10	40.31	40.12
Avg. Distance	595.11	563.80	557.93	564.69
Prob. Late	0.074	0.015	0.019	0.015
Pred. Del. Time	24.17	25.40	25.31	25.38
Days Early	11.68	16.16	16.56	16.16
Avg. Review Score	3.90	2.88	3.87	2.88
t-score	17.878		13.05	
Degrees of freedom	1016.4		1764.3	
p-value	$< 2.2e{-16}$		$< 2.2e{-16}$	

Figure 7: Using propensity score matching, we see that important covariates related to the review score are more similar between the treatment and control group. We primarily focus on cost and delivery time which was shown to have a causal effect earlier.

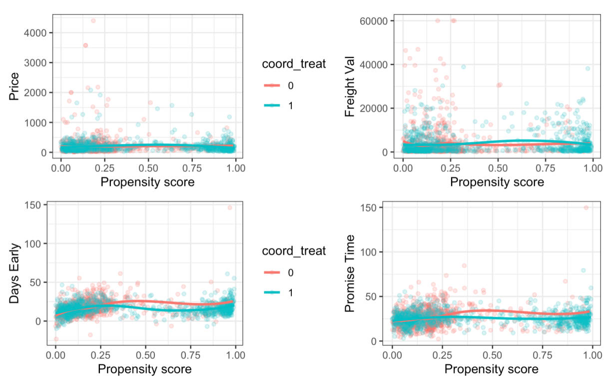

Figure 8: We graphically show the distribution of covariates is the same between the control and treatment group

3.2 AGES Algorithm

To further our understanding of the relevant causal variables, we fit AGES (Aggregated Greedy Equivalence Search) algorithm[10] on our data. AGES aggregates results from the GES algorithm with a greater range of skeletal undirected graphs to more accurately detect weak causal relations for smaller samples of data. Our main motivation for choosing AGES over other causal discovery algorithms include AGES's ability to detect weak edges and our data's compatibility with AGES's assumptions. Other algorithms such as GES[11], PC[12], GIES[12], and IDA[12] do not allow for hidden and selection variables, and thus it was important for us to choose a method that did not make these assumptions. In addition to our variables most closely resembling the Gaussian distribution, our data also satisfies δ_i-strong faithfulness (defined in Theorem 3.2[10]). In short, δ_i-strong faithfulness is satisfied when all variables are pairwise independent or significantly partially correlated conditional on all other variables not included in the pair. Prior to proceeding with the result of AGES, we note that the partial correlation calculations and background research on conditional independence of our variables satisfy the δ_i-strong faithfulness condition. From this causal diagram we observe both the causal factors that affect predicted and actual delivery times, and that predicted and actual delivery times are causal factors to review score.

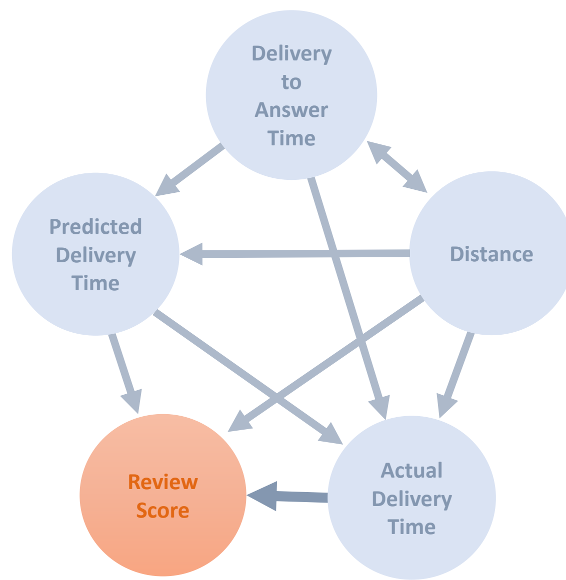

Figure 9: Causal Analysis of review score. The primary cause of review score (highest variable covariance) is the actual delivery time, followed by predicted delivery time.

4 Customer Satisfaction Model

4.1 Interpretable Machine Learning Model

In literature, generalized additive models have been used to offer intrepretability that black box models lack. Since GAMs are typically modeled as a sum of univariate models, it is easy to note the coefficients of each term and interpret the impact of each factor. Standard GAMs have the form

$$g(E[y]) = \sum f_i(x_i)$$

where g is a link function. GAMs produce clear partial dependence plots of each variable, but often at the cost of accuracy compared to black box models. In order to maintain black box accuracy while maintaining interpretability, we model bivariate interactions using the GA^2M[13] model.

$$g(E[y]) = \sum^{n} f_i(x_i) + \sum^{m} f_{i,j}(x_i, x_j)$$

The formulation of the GA^2M maintains the univariate regressor f_i, while adding in pairwise interactions $f_{i,j}$. While the model fits all n variables, we select the top m interacting features using a greedy forward stagewise selection strategy called FAST[13].

4.1.1 Insights from Interpretability

With an interpretable model, we gain insight into how each feature affects the final scoring criteria. The model learned that the higher the predicted vs actually delivery time difference, meaning that the package arrived earlier than expected, the higher the customer satisfaction. Furthermore, we learn that the faster the actual delivery time, the more satisfied the customer is. We also note some more marginal and unstable effects from payment value. Despite not knowing at first why customers had lower satisfaction from higher payment, we later discover from our other models that higher payment is positively correlated with more products per order, which–we described earlier–may not arrive together and therefore, frustrate customers. We use these plots to further tune the model by removing unimportant features such as order to shipping limit time. Finally, we observed that the linear terms are more significant in our interaction terms in our final factor ranking, which makes sense because separate timed measurements of the overall delivery time do not interact non-linearly.

Figure 10: Every additive component of the interpretable model. The shaded regions of the noninteracting components is the range containing the middle 95% of the models' predictions. Most importantly, the model learns that an increase in delivery time is associated with a decrease in score.

4.1.2 Model Construction

We first perform feature selection from all available features using an extra tree classifier[14]. We choose extra randomized tree due to its resistance to overfitting to the data and faster training time compared to ordinary random forests for our large dataset. After observing the most frequently occurring top level factor in our forest, we deduce 6 total significant variables to train the GA^2M on. Using grid search, we find optimal parameters for number of regressors and interaction, and validate using 10-fold cross validation. We obtain a training mean squared error (MSE) of 1.23 and a test MSE of 1.24, showing that we successfully controlled for overfitting of our model.

4.2 Other Models

4.2.1 LASSO Model

We also consider a simple LASSO model to predict customer satisfaction. We use the same linear model for review score (see Eqn. 2), but add a regularizer for the coefficients when fitting the data. Letting X be the matrix for all the covariates and the treatment variable, we fit the coefficients by solving

$$\min_{\beta} \|RS - X\beta\|_2^2 + \lambda \|\beta\|_2^2 \tag{3}$$

We pick $\lambda = 0.000924$ by tuning λ in the model using five-fold cross validation. Using a hold out test set we evaluate the out of sample MSE to be 1.39. While not shown, the non-zero coefficients corerspond to GA^{2}M aside from interaction terms which we did not include.

4.2.2 Optimal Tree Classification

Another model that we consider is Optimal Tree Classification from iAi package. Optimal classification trees attempt to improve upon other decision tree algorithms by creating the entire decision tree at once to achieve global optimum[15]. Among various models that the package provides, classification is the most logical choice as we are predicting categorical data (rating from 1 to 5). The out-of-sample accuracy of the model is 62% and MSE is 1.89. This model is less accurate compared to the interpretable machine learning model above, perhaps because the classification tree structure cannot capture the bidirectional causal factors that affect rating. Nonetheless, the model agrees with the feature selection of the models above, emphasizing the impact of earliness and coordination of delivery in customer satisfaction. Full diagram is attached in appendix.

4.3 Conclusions about Potential Policies

Through our exploratory analysis, we discover that late delivery and, in particular, lack of coordination on multi-item orders negatively impact customer satisfaction. We further confirm that lateness in delivery had negative impact on review rates from our causal discovery and the customer satisfaction models above. From these discoveries, we consider policies that can improve the probability of the order arriving on time and actionables that Olist, or any other e-commerce company, can take to improve customer satisfaction through shortening their delivery time. Furthermore, we consider logistical changes to promote coordination of the arrival of multi-item orders.

5 Policy Suggestion

From our analysis, we demonstrate that there are key operational and logistical levers that Olist can take advantage of to improve customer satisfaction. With their recent third-round financing round completed with a \$46.5 million investment from SoftBank[16], we propose potential logistics policies and projects that Olist could pursue to expand their business and market share.

5.1 Facility Location Problem

To improve Olist's logistics capabilities, we identify potential warehouse locations that could improve Olist's logistic capabilities. we Warehouses provide logistics companies like Olist a location to better coordinate deliveries and improve delivery speed. Given that Olist has limited resources, we formulate a general data-driven mixed-integer program in order to identify cost-effective warehouse c Warehouses provide logistics companies like Olist a location to better coordinate deliveries and improve delivery speed. Given that Olist has limited resources, we formulate a general data-driven mixed-integer program in order to identify cost-effective warehouse construction proposals. The formulation is as follows:

$$\begin{aligned}
\max \quad & \sum_{i,j} g_{ij} y_{ij} \\
\text{s.t.} \quad & \sum_{j \in W} y_{ij} \leq 1, \quad \forall i \in S \\
& \sum_{j \in C(i)} d_j y_{ij} \leq K x_i, \quad \forall j \in W \\
& \sum_{j \in W} c_j x_j \leq B \\
& \sum_{j \in W} x_j \leq 20 \\
& x_i \in \mathbb{Z}^+, 0 \leq y_{ij} \in \leq 1, \quad \forall i, j
\end{aligned}$$

- S,W: Set of locations of sellers and warehouses, respectively
- y_{ij}: Fraction of items from sellers in location i stored in warehouse j
- x_j: Binary decision variable to build warehouse in location j
- c_j: Cost of warehouse j
- d_j: Demand for location j sellers
- g_{ij}: Improvement in a metric, such as customer satisfaction or delivery speed
- K: Base size unit of warehouse
- B: Olist Budget

The first constraint ensures that we do not move more than the actual number of a seller's items. The second constraint is the capacity constraint for warehouses. We assume our choice of warehouses increase by size K. The third constraint is the budget constraint. The last constraint states that we cannot build warehouses 20x more than the size of K.

For our integer program, we consider two possible metrics to optimize over: Review Score and Days Early. We obtain estimates of the gains of moving items from location i to warehouse j by estimating the counterfactual performance of shipping from location j. Formally, we define the gains as

$$g_{ij} \equiv \frac{\sum_k d_{ik} m_{jk}}{\sum_k d_{ik}} - \frac{\sum_k d_{ik} m_{ik}}{\sum_k d_{ik}}$$

where d_{ik} is the demand going from location i to location k and m_{ik} is the respective performance metric (like average review score) for that location pair. This represents the counterfactual performance as both terms in the difference use the demand from location i. Since the Olist order data may not have data on the metrics for each location pair, we utilize the R package softImpute to apply matrix completion to estimate the missing values.

The remaining parameters such as K, c_j, and B are roughly estimated using back of envelope calculations, but, in a sense, can be adjusted to produce solutions with a certain number of warehouses that store items in a certain the number of locations.

5.2 Geographic Recommendations

Using the integer program and current Olist data without projections into the future, we can infer the order of locations which Olist should build warehouses. This can be done by increasing the budget in the budgeting constraint. Through this method, Olist should build the first five warehouses by state in the following order:

1. Mato Grosso do Sul (MS)
2. Ceará (CE)
3. Goiás (GO)
4. Rio Grande do Norte (RN)
5. Rio Grande do Sul (RS)

Interestingly, these recommendations line up with the order Amazon has chosen to enter the Brazilian market. In their case, they opened near Sao Paulo in 2018[2] and expect to open another warehouse in Pernambuco in 2020[3]. Our recommended states of Mato Grosso do Sul and Ceará border Amazon's chosen states and follow the same pattern of targeting the southeast region first and then the northeast region. A potential factor that may have led us to different conclusions than Amazon could be the fact that we do not factor in the benefits of port cities. However, since Olist primarily serves domestic customers, we see our recommendations may in fact be more suitable.

[2]https://reut.rs/2BOmOOE
[3]https://reut.rs/2shPPii

Figure 11: Recommended warehouses in Brazil. States shaded in green represent the northern region of Brazil; in blue represent northeastern region; in yellow represent central western region; in red represent southeast region; and in orange represent southern region. Black star-shaped markers denote warehouses.

Figure 12: Small blue markers denote sellers that will send order to black markers that denote warehouses. The paths from seller to warehouse are shown by orange lines for CE warehouse; purple lines for GO warehouse; red lines for MS warehouse; green lines for RN warehouse; and blue lines for RS warehouse.

5.3 Constraints on Budget and Capacity

To estimate still rough but more specific constraints on budget, we use statistics on average warehouse sizes[17,18] and average commercial construction cost in Rio de Janiero[19] and partially linearly scale to account for the differences in land cost using GNP per capita[20]. From the approximated cost per square feet, we calculate the cost of each warehouse for each small warehouse, 5000 square feet, to further optimize for warehouses of different sizes. For estimation of capacity, we calculate the average volume of the products from the average width, length, and height of the products[3] and approximate spacial efficiency of 70%. Note that 70% is a very rough estimate accounting for varying level of spacial efficiency dependent on additional investments in technology and labor.

5.4 Financial Impact

From literature-based background research, we discover two ways the customer satisfaction, represented by review score in our analysis, affects revenue for e-commerce[21,22]. The model extrapolated from literature utilizes ordinary least square to estimate the relative impact of changes in customer satisfaction in a company's market value using American Customer Satisfaction Index (ACSI)[23]. Also represented in the table below, a 1% change in customer satisfaction increases the company's market value by 4.6%.

The Effect of ACSI on Market Value of Equity

		Ln(Market Value)	Ln(Market Value)
ACSI (lnACSI)	β_1	4.592 (.000)	
Total assets (lnBVA)	β_2	1.943 (.000)	2.039 (.000)
Total liabilities (lnBVL)	β_3	−1.020 (.003)	−1.157 (.003)
Constant	α	−20.056 (.000)	.235 (.704)
N		601	601
R^2		.70	.61
F		(.000)	(.000)

Notes: We estimated the models by including dummy variables (not shown) for years 1995 through 2002 (1994 was the reference year); *p* values are in parentheses.

Figure 13: Results of Ordinary Least Squares[23]

Note that the right column represents regression without customer satisfaction as a variable. The R-squared value decreases by 0.09 when customer satisfaction is not included, further supporting the impact of our model on Olist's financial growth.

5.5 Valuation and Revenue Estimation

Since Olist is a private company, its financial statements are not publicly available. We therefore estimate Olist's current equity market capitalization and its revenue in USD. We know that in 2019 SoftBank invested $46.6million in Olist[16]. Using the venture capital (VC) method in 14, a valuation approach used by VC and private equity investors, we utilized this information to back calculate an estimated valuation that SoftBank used to valuate Olist. The inputs include: the investment amount; an ownership stake, which we assume to be 25% and is computed as an average of SoftBank's ownership stakes in other companies in its Vision Funds I and II[24]; the time to exit, which is typically 5 years for VC funds and is the case for SoftBank Vision Fund I; the VC cost of capital, which is typically assumed to be 15%; the target return, which is typically assumed to be 30-40% and here we assume 40%; and the retention rate, which is roughly 60% for VC funds. These inputs produce an implied exit valuation of $1.67 billion for SoftBank in 5 years. Backtracking, the post-money valuation, which is the valuation after SoftBank's investment is $186.6 million. We use this valuation as a proxy for Olist's market capitalization.

VC Method: SoftBank Investment (Olist)	
Investment today	$46,650,000
Ownership stake	25%
Time to Exit	5
VC cost of capital	15%
Target return	40%
Retention	60%
Implied probability	37%
Implied post-money	$ 186,600,000.00
Implied exit valuation	$1,672,632,640.0

Figure 14: Venture Capital (VC) Method for SoftBank's 2019 investment in Olist

To estimate Olist's current revenue, we compiled financial information on the market caps and total revenue (TTM) of public comparable companies, such as HubSpot, Weimob, and Lightspeed POS, from Pitchbook[25]. We computed the average market cap/revenue ratios of these comparables to be 2.782. We then divided this value from the market cap computed previously to arrive at a current revenue of $67 million.

From this estimation of market cap and revenue, we calculate the scale factor that relates order volume and revenue according to the subset represented in the data sets provided and the actual corresponding values. From Olist's business model of taking 20% of sales and subscriptions we estimate the actual demand and revenue to be approximately 200 times the values represented in the subset of data.

5.6 Final Optimized Policy

In this section, we present a proposal for Olist for locations to lease or construct warehouses to improve their logistics. We first assume they spend $10 million of their $50 million investment towards the proposal. Under this budget constraint, we see that our integer program prescribes building a 100,000 square feet warehouse in Mato Grosso do Sul, a 25,000 square feet warehouse in Ceará, and 10,000 square feet warehouse in Rio Grande do Norte.

6 Impact and Evaluation

6.1 Short Term Impact

To evaluate the short term financial impact of our policy, we estimate the monetary value of the increase in total equity of Olist for varying levels of budget using 2018 data. For calculations regarding converting increase in rating to dollars, we utilize the financial model and numbers estimated in the optimization section above: 4.1%, 186 million dollars of market value, multiplier of 200 for scale demand from data to actual. The average rating of 4.09[3] was used to calculate the percent increase in rating. Below are the results with monetary data in million USD.

Budget	Increase in Rating	Increase in Total Equity
1	0.0216	4.517
5	0.0773	16.176
10	0.136	28.379
20	0.055	47.070
30	0.276	57.712
40	0.276	57.712
50	0.276	50.571

In 15, we visualize the average customer score by zip code prefix. The southern and southeast areas of Brazil have higher average customer scores, whereas the lower average customer scores of 1-2.5 are spread across southern, eastern, and northern areas. The northern areas have primarily low review scores of 1-2.5. In 16, we visualize the average customer review score with the customer review gains added in from the policy we propose. Compared to 15, 16 shows that there is a cluster of perfect customer review scores in the northeastern region of Brazil and a spread of perfect customer review scores across the northern and western regions.

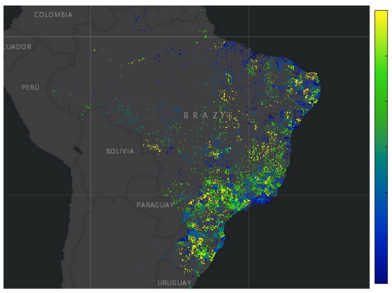

Figure 15: Geoplot of Average Customer Review Scores from Olist data

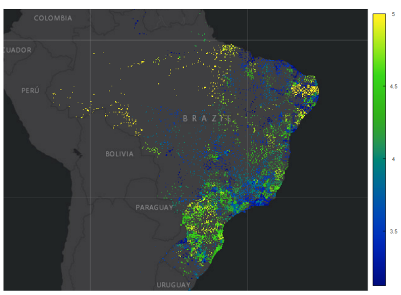

Figure 16: Geoplot of Average Customer Review Scores with Gains from Policy Proposed (Short Term)

6.2 Long Term Projection

We evaluate the long term benefit of putting warehouses using a geographically weighted regression to both evaluate demand for the future and the benefit derived from the demand in each state in Brazil. Since commerce is mainly an indicator of economic activity and buying power, we include variables such as GDP, HDI (human development index), and income index and regress on the number of orders from that region[26,27,28]. We intuitively know that regional demand should be correlated with regional economic indicators, hence we test the spatial auto correlation of these factors with Morgan's I[29] at 95% confidence. Based on the results of the significance test, we train the model on 2018 data and use cross validation to determine the kernel bandwidth, obtaining an R^2 value of 0.91 with a kernel bandwidth of 25.

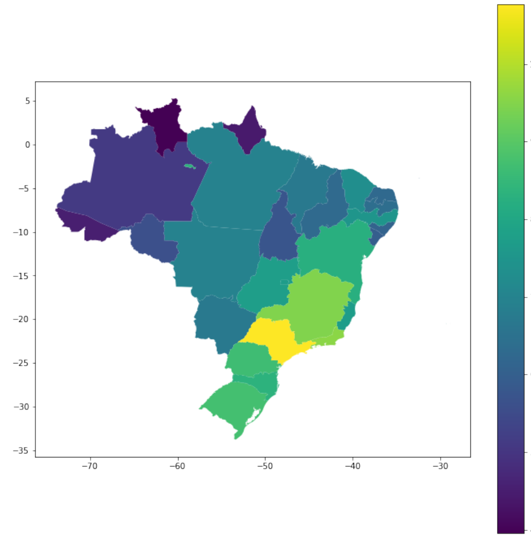

Figure 17: Demand in 2018 per state

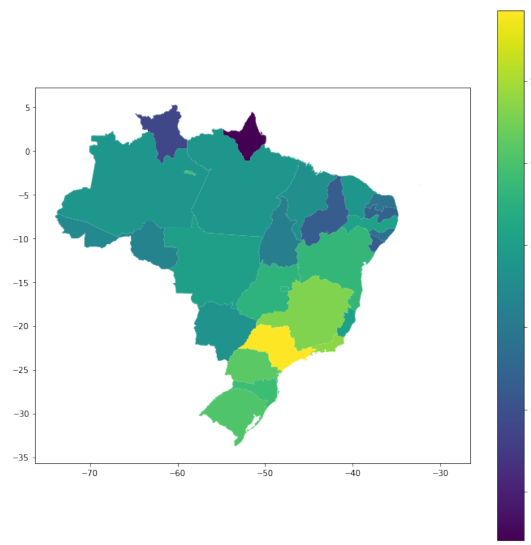

Figure 18: Demand in 2025 per state

We evaluate a sensitivity analysis on the demand for the year 2025 under an assumption of 0.7 percent population growth per year[30], and a GDP growth rate of 1.17% according to the world bank[31]. Using the projections in demand, we further evaluate the impact of our policy in long term. Below aggregates our percent growth in total equity for varying values of budget.

Budget	Increase in Rating	% Increase Total Equity
1	0.136	15.3
10	0.257	29.0
20	0.410	33.8
30	0.376	42.2
40	0.376	42.2
50	0.376	42.4

6.3 Sensitivity Testing

To further strengthen our model we stress-test our model by varying the demand scale factor. We consider the case of demand multiplier of 100 and demand multiplier of 300 to account for both overestimation and underestimation of Olist's actual sales volume. The monetary impact represented by increase in total equity are presented in the tables below, each representing stress-test on short term model and stress-test on long term model. Note the units for monetary data are in millions of USD.

Budget	Multiply by 100	Multiply by 300
1	6.741	3.714
5	29.796	11.626
10	54.022	19.831
20	68.107	32.285
30	68.107	43.689
40	68.107	50.220
50	68.107	50.571

Budget	Multiply by 100 (%)	Multiply by 300 (%)
1	16.6	13.8
10	35.5	26.8
20	46.1	30.0
30	46.1	32.5
40	46.1	37.9
50	46.1	38.5

Notice how the model projects non-negligible profit despite potential underestimation and overestimation in demand. This result further support our policy proposal.

7 Conclusion

Since the first invention of the internet more than 25 years ago, the e-commerce space has become a bustling competitive landscape with razor thin margins[1]. With a large variety of products and competitive algorithmic pricing, sellers online invest tremendous efforts to maintaining customer relations and buyer retention. In light of this, we start our analysis directly from the source of customer happiness by performing an translation and aggregation of review and review sentiments. Among the variety of feedback, we saw a recurring theme of people frustrated at late and separate packages, and due to this expectation of shipping hassle, customers are very please when packages arrive early. All this seems to confirm what we intuitively know to be true, that for e-commerce to replace brick and mortar, people need to get their stuff and get it right when they want it.

In our causal analysis, we leverage instrumental variables and propensity score matching to show improving logistics and supply chain for Olist's e-commerce business has a positive effect on customer satisfaction. Our causal inference insights reveal two important operational levers to target: 1. early order arrivals and 2. coordinating deliveries. The first confirms past work that studies how to manage customer expectations, while the second is a less studied phenomenon. From a managerial implications perspective, we show that there are significant gains towards investing in logistics as a business scales. By using Olist as a case study, we construct an interpretable predictive customer satisfaction model and use integer programming to model the potential impacts of expanding their warehouse supply chain in a targeted fashion. After fitting Olist data to our model, we generate recommendations that focus on improving customer satisfaction by constructing warehouses to improve Olist's fulfillment capabilities. Impressively, we show that our model makes recommendations that align with past and future plans of the e-commerce and logistics juggernaut Amazon but tailored to meet the specific needs of Olist. Finally, we show through geographically weighted regression that these proposals are not only sensible but also would significantly boost Olist's growth.

In conclusion, we identify essential insights into the importance of logistics, construct a data-driven pipeline to exploit their long term implications, and ultimately show the fundamental importance of logistics. Moreover, we provide a general framework for reasoning about operations problems that can be generalized, making it especially relevant in an increasingly quantitative economy.

Appendix A

Additional Exploratory Analysis

A.1 Further Geoplot Analysis

In addition to the geoplots included in Section 2, we visualize other features of the Olist dataset to get a better sense of how different characteristics on the customer or seller side vary by location.

In 19, we visualize the average price per order by zip code prefix. The southern and southeast areas of Brazil had lower average prices compared to the northeastern region. To explore why this may be, we first examine if customers in the northeastern area have to pay more for freight, which we confirm in 20

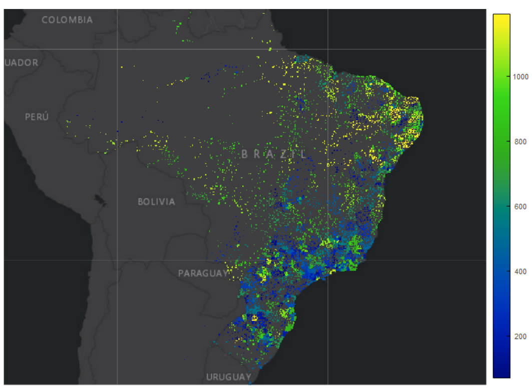

Figure 19: Geoplot of Average Price of Orders (in Brazil Real)

In 20, we compute and visualize the freight ratio of each zip code, where the freight ratio is the ratio of the freight value by the order price. This represents the proportion of order price that a customer has to pay in order to receive their delivered order. The southern and southeast areas of Brazil have lower freight ratios (0.1-0.2) compared to the ratios of the northeastern region (0.6-0.8). This is reasonable, given that logistics (i.e. shipping, handling, etc.) costs imply that highly populated regions may have lower freight ratios compared to sparse regions of low population. This–along with the geoplots of proportion of delayed orders and average review score–may suggest why the northern region of Brazil contains customers with lower average review scores.

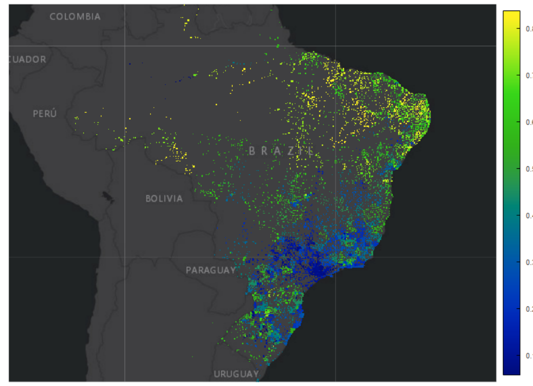

Figure 20: Geoplot of Freight Ratio

In 21, we visualize the locations of the sellers of the orders by displaying the total count of sellers by zip code prefix. Most sellers are aggregated in the southern and southerneast areas, which could influence delivery time–at least the time it takes for the seller to hand the order to the logistics provider.

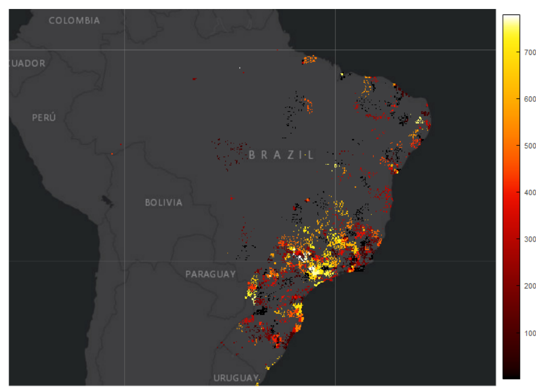

Figure 21: Geoplot of Sellers (Count)

In 22, we visualize locations of the sellers of delayed orders by zip code prefix. Most sellers whose orders have been delayed are aggregated in the southern and southerneast areas. This is intuitive given that most sellers are located in those areas.

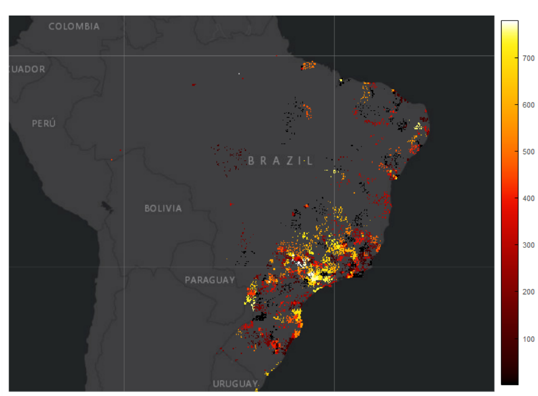

Figure 22: Geoplot of Sellers of Delayed Orders (Count)

A.2 Order Growth Over Time

By grouping the dataset on customer city and summing up the order count, we identify the top customer cities by total order count: Sao Pualo (17934), Rio de Janeiro (7887), Belo Horizonte (3171), Brasilia (2419), Curitiba (1758).

Then we visualize the growth of orders over time in these cities by plotting total order count per month over time. We use the order approved at dates to place the orders into buckets of months. For Sao Paulo, we do not include the last data point in September 3, 2018, since that is the latest timestamp of order data collected by Olist.

Across all plots, around Dec 2017 to Jan 2018, there is a spike in order count, which indicates that ecommerce orders reached a peak during the Christmas season. For the retail sector in Brazil, Christmas is the most profitable shopping season, while Mothers' Day, is a close second. Mothers' Day occurs in May, which could explain the peak in order count in May 2018.

Figure 23: Time Series of Sao Paulo Order Count per Month

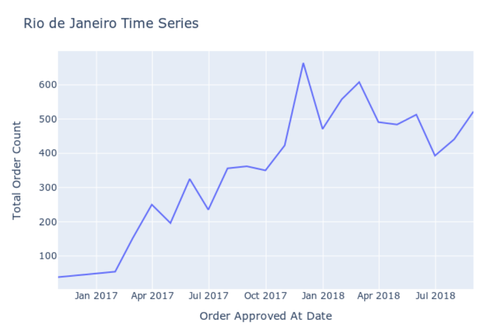

Figure 24: Time Series of Rio de Janeiro Order Count per Month

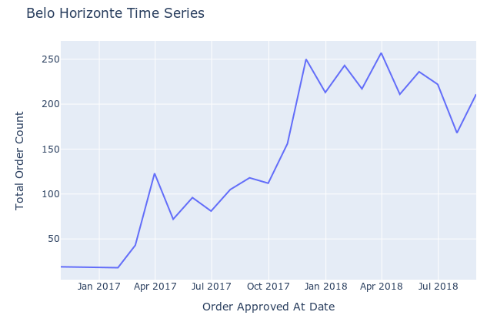

Figure 25: Time Series of Belo Horizonte Order Count per Month

Figure 26: Time Series of Brasilia Order Count per Month

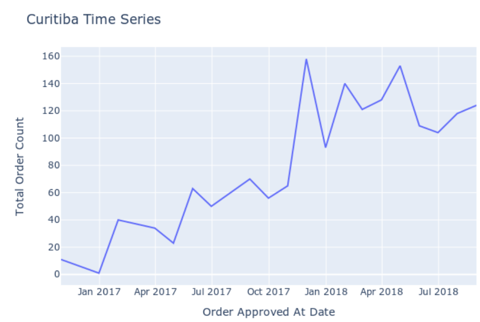

Figure 27: Time Series of Curitiba Order Count per Month

Appendix B Supplementary Literature Review

B.1 Financial Impact of Customer Service

Firstly, the empirical analysis of factors that impact revisiting customers for Alibaba support delivery risk's negative impact on customer satisfaction and retention aka the likelihood of customers revisiting the e-commerce platform. Secondly, field experiments on eBay conclude that customers are willing to pay on 8.1% more to buy from sellers with strong ratings . From these results we can model the impact of rating on e-commerce revenue through increase in price and quantity of the orders.

Appendix C Additional Figures

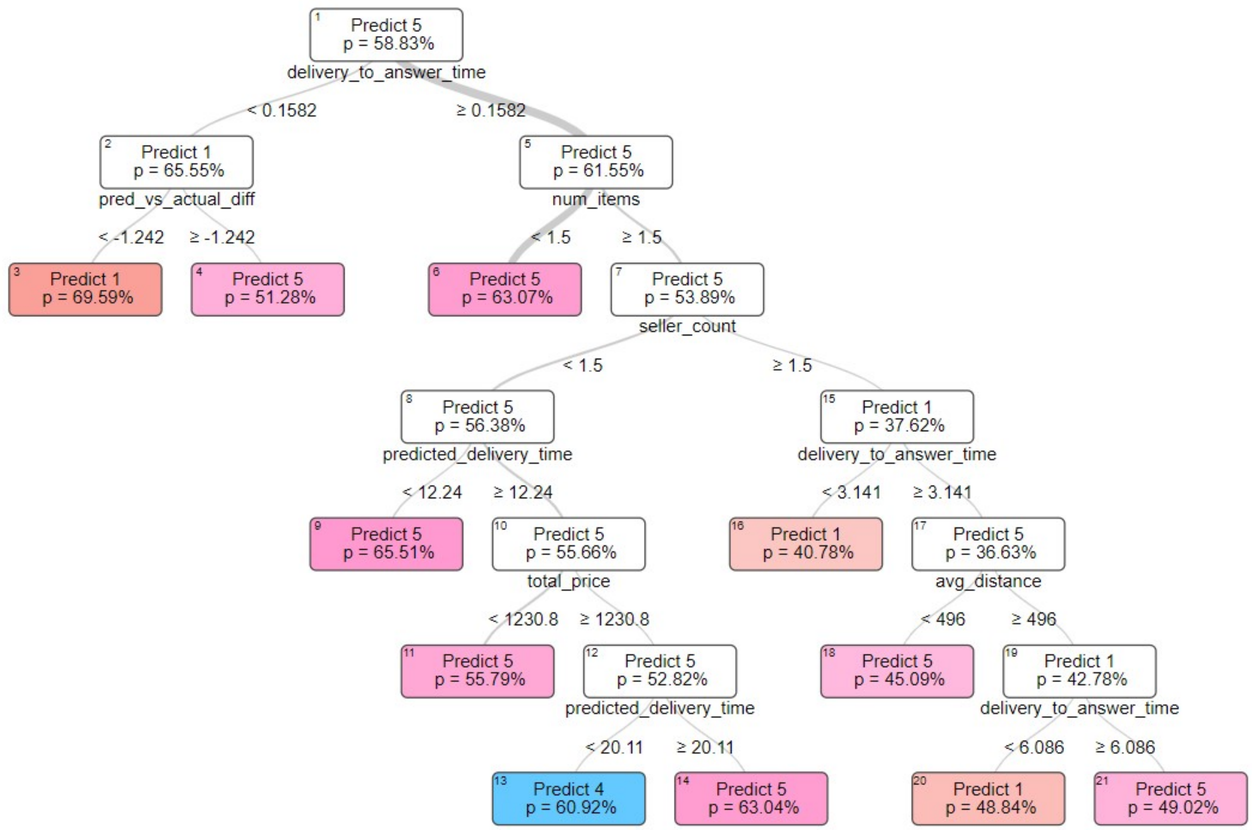

Figure 28: Results of optimal tree classification to predict review score

References

[1] *amazon profit*. 2019. URL: https://money.cnn.com/2017/12/19/investing/amazon-profit/index.html.

[2] Hutto et al. "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text". In: Jan. 2015.

[3] Olist and André Sionek. *Brazilian E-Commerce Public Dataset by Olist*. Nov. 2018. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1949412/.

[4] Michele Banko and Lucy Vanderwende. "Using N-Grams To Understand the Nature of Summaries". In: *Proceedings of HLT-NAACL 2004: Short Papers*. Boston, Massachusetts, USA: Association for Computational Linguistics, May 2004, pp. 1–4. URL: https://www.aclweb.org/anthology/N04-4001.

[5] William B. Cavnar and John M. Trenkle. "N-Gram-Based Text Categorization". In: *In Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval*. 1994, pp. 161–175.

[6] Ann P Bartel, Carri W Chan, and Song-Hee Kim. "Should hospitals keep their patients longer? the role of inpatient care in reducing postdischarge mortality". In: *Management Science* (2019).

[7] Robert Bray. "Operational transparency: Showing when work gets done". In: *Available at SSRN 3215560* (2019).

[8] Teck H Ho and Yu-Sheng Zheng. "Setting customer expectation in service delivery: An integrated marketing-operations perspective". In: *Management Science* 50.4 (2004), pp. 479–488.

[9] Paul R Rosenbaum and Donald B Rubin. "The central role of the propensity score in observational studies for causal effects". In: *Biometrika* 70.1 (1983), pp. 41–55.

[10] Marco F. Eigenmann, Preetam Nandy, and Marloes H. Maathuis. *Structure Learning of Linear Gaussian Structural Equation Models with Weak Edges*. 2017. arXiv: 1707.07560 [stat.ME].

[11] Max Chickering. "Optimal Structure Identification With Greedy Search". In: *Journal of Machine Learning Research* (Nov. 2002). URL: https://www.microsoft.com/en-us/research/publication/optimal-structure-identification-greedy-search/.

[12] Clark Glymour, Kun Zhang, and Peter Spirtes. "Review of Causal Discovery Methods Based on Graphical Models". In: *Frontiers in Genetics* 10 (2019), p. 524. ISSN: 1664-8021. DOI: 10.3389/fgene.2019.00524. URL: https://www.frontiersin.org/article/10.3389/fgene.2019.00524.

[13] Harsha Nori et al. "InterpretML: A Unified Framework for Machine Learning Interpretability". In: *arXiv preprint arXiv:1909.09223* (2019).

[14] Pierre Geurts, Damien Ernst, and Louis Wehenkel. "Extremely Randomized Trees". In: *Machine Learning* 63 (Apr. 2006), pp. 3–42. DOI: 10.1007/s10994-006-6226-1.

[15] Dimitris Bertsimas and Jack Dunn. "Optimal classification trees". In: *Mach Learn* 106 (2017), pp. 1039–1082. DOI: 10.1007/s10994-017-5633-9.

[16] Carolina Mandl. *Japan's SoftBank invests in Brazilian marketplace integrator Olist*. Oct. 2019. URL: https://www.reuters.com/article/softbank-latam-olist/japans-softbank-invests-in-brazilian-marketplace-integrator-olist-idUSL2N2771HD.

[17] Cisco-Eagle. *What is the Average Warehouse?* 2020. URL: https://www.cisco-eagle.com/industries-served/order-fulfillment/the-typical-warehouse.

[18] DC Velocity Staff. *Average size of U.S. warehouses has doubled since 2002, CBRE report finds*. 2020. URL: https://www.dcvelocity.com/articles/29330-average-size-of-u-s-warehouses-has-doubled-since-2002-cbre-report-finds.

[19] Compass International. *International Warehouses / Logistics Center Costs*. 2020. URL: https://compassinternational.net/international-warehouse-logistics-center-costs/.

[20] Brazil Consulting. *Brazilian States, Abbreviations, and Information*. 2015. URL: https://brazil-help.com/brazilian_states.htm.

[21] Shahzad Ahmad Khan, Yan Liang, and Sumaira Shahzad. "An Empirical Study of Perceived Factors Affecting Customer Satisfaction to Re-Purchase Intention in Online Stores in China". In: *Journal of Service Science and Management* 8 (2015), pp. 291–305. DOI: 10.4236/jssm.2015.83032. URL: https://www.scirp.org/journal/CTA.aspx?paperID=56198.

[22] Paul Resnick et al. "The value of reputation on eBay: A controlled experiment". In: *Experimental Economics* 9(2) (2006), pp. 79–101. DOI: 10.1007/s10683-006-4309-2Âă.

[23] Claes Fornell et al. "Satisfaction and Stock PRiceS: High Returns, Low Risk." In: *Journal of Marketing* 7(1) (2006), pp. 3–14. DOI: 10.1509/jmkg.70.1.003.qxd.

[24] *SoftBank Group*. 2019. URL: https://en.wikipedia.org/wiki/SoftBank_Group.

[25] *Pitchbook Olist Profile*. 2019. URL: https://pitchbook.com/profiles/company/102473-65?fbclid=IwAR1DVGRG31gRzDFNHvqrC473PeQ3lRen0FBaqyoN2Ax6esQuIS60PiPEozc.

[26] *States of Brazil*. 2019. URL: https://en.wikipedia.org/wiki/States_of_Brazil.

[27] *Brasil IBGE: GDP of Municipalities Estimates*. 2019. URL: https://www.ibge.gov.br/en/statistics/economic/national-accounts/19567-gross-domestic-product-of-municipalities.html?edicao=26417&t=destaques.

[28] *Brasil IBGE: Population Estimates*. 2019. URL: https://www.ibge.gov.br/en/statistics/social/population/18448-estimates-of-resident-population-for-municipalities-and-federation-units.html?=&t=downloads.

[29] "Notes on Continuous Stochastic Phenomena". In: *Biometrika* (1950).

[30] United Nations. "world population prospects". In: (2019). URL: https://population.un.org/wpp/Download/Standard/Population/.

[31] *World Bank Open Data*. 2018. URL: https://data.worldbank.org/.