Build an Optimal Pricing Model for TURO

BAX 452 – 01
Machine Learning
Final Project
March 17, 2022
Group 5

Authors:
Yuntong Xiao
Jiaqi Zhou
Gijeong Go

Table of Contents

Executive Summary

The largest car-sharing firm based in the United States, Turo, enables private car owners to rent out their cars via an online platform in over 56 countries. After analyzing the competitive advantages and disadvantages of Turo's website, we suggested adding a new category feature "Browse by category" to its website that filters out car listings in addition to "Browse by make" and "Browse by destination". Through cluster analysis, we created the following three vehicle categories: Spacious, Luxury and Economy. Moreover, we established a pricing system that helps car hosts to set the optimal prices that maximize their revenue. In this report, we discussed the past literature on dynamic pricing strategies for peer-to-peer businesses, described the insightful patterns and anomalies discovered in Turo's dataset, and presented our approach to solving the pricing problem in hope of establishing pricing competitiveness for Turo.

Business Background and Problem Formulation

Dubbed the "Airbnb of vehicle rentals," Turo began its operations in San Francisco and Boston in 2009 and has since spread to "more than 5,000 cities across the United States, Canada, Germany, and the United Kingdom" (Hamstra, 2019). It competes against well-known vehicle rental companies, including National, Enterprise, Dollar, Avis, and Hertz, and provides comparable peer-to-peer (P2P) car-sharing services like Getaround. Unlike other vehicle rental companies, Turo has distinguished itself and established its position in the market by changing the game on the supply side.

After comparing across top car rental websites, we can easily observe that Turo's webpage looks quite different, as it misses the vehicle category. Unlike other companies that provide broad categories such as economy, mini, compact, and standard, as shown below in Figure 1, Turo allows users to browse by the specific car brand. However, for users who are unsure of which brand they prefer, a more general filter might be helpful. This poses a typical clustering problem.

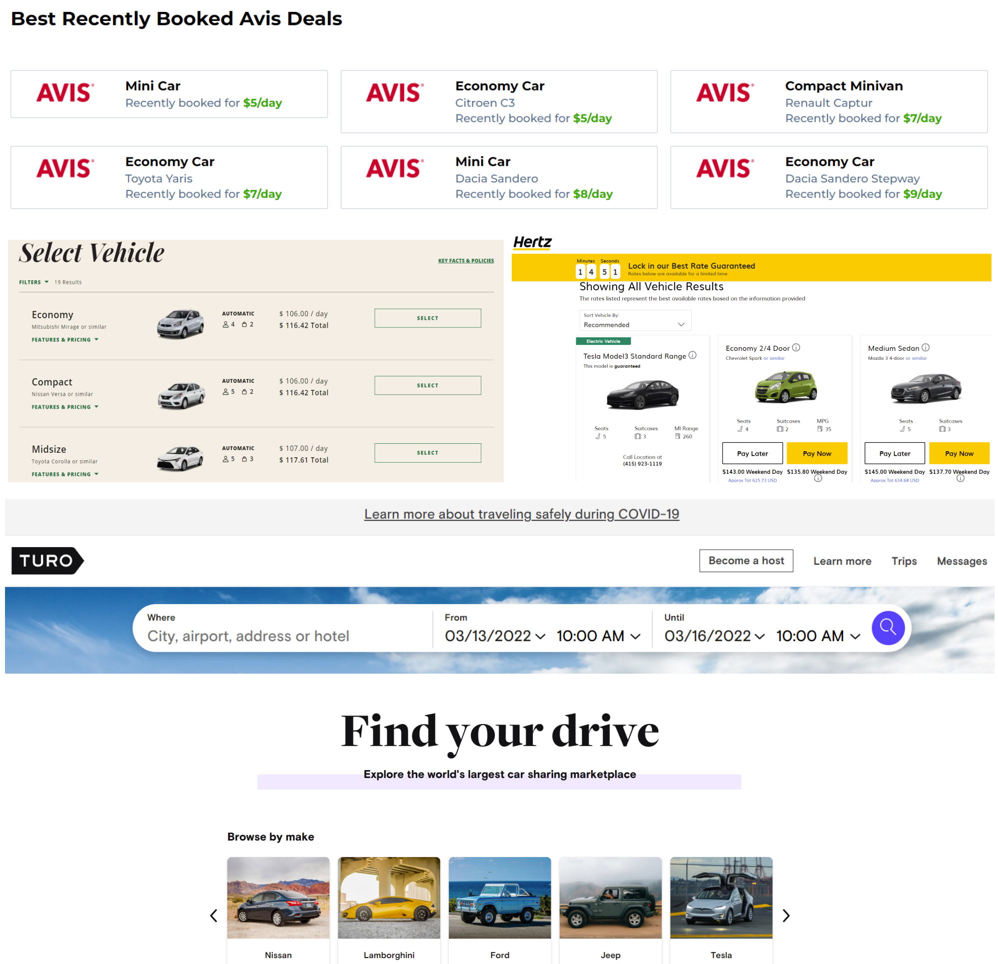

Figure 1: Websites from Turo's competitors (Avis, National, Hertz) and Turo

The second business problem we want to solve is how to price the car listings. As a platform provider, Turo allows hosts to choose rental prices for their listings. A problem hosts oftentimes encounter is setting the right listing price. To tackle this problem, we looked for ways to build an optimal pricing model that provides price suggestions. The optimization objective of the model is to maximize revenue, which requires the estimation of demand for a range of prices.

Traditional Approaches

Since Airbnb, a P2P home-sharing platform, operates under a similar business model to that of Turo, meaning they might face similar pricing challenges, we researched how Airbnb developed its pricing model. Peng et al. (2018) introduced three components in the Airbnb pricing system: 1) predict the booking probability for each night using a binary classification model, 2) predict the optimal price for each night using a regression model with a customized loss function, and 3) provide the final price suggestions by applying additional personalization logic, such as listing goals on top of the second model. Due to data accessibility issues, we referred to the first and second approaches when building our model.

Peng et al. (2018) used a k–d tree to divide listings into k-dimensional clusters/nodes before training the classification model per node. Clustering listings helps generalize listing characteristics while retaining variations among listings to some extent. Similarly, we would build a k-means clustering model to generate different car categories and create a demand function for each category. Previous studies have assumed that the market value of one product is linearly associated with the value of each product feature (Abbasi-Yadkori et al., 2011; Cohen et al., 2016). Thus, based on this assumption, we used linear regression to predict demand with price and other product features such as the age of the car. By estimating how demand changes with respect to price, we built a demand curve that was generally assumed to slope down (Krugman et al., 2017). Finally, we calculated the optimal price for each listing by maximizing Price p $\times$ Demand D (p).

Exploratory Data Analysis

Our dataset contains 21 variables and 36,279 records. The data definitions are shown in Appendix Table 2. Our variables of interest are averageDailyPrice and rentalTripsTaken. They represent price and demand respectively. We first explored the distribution of those variables (Figure 6 and Figure 7 in Appendix) to identify the outliers. While half of the listings were at prices below $64, there was an extreme outlier of $1,999, which was excluded from our analysis.

Figure 2 below shows a downslope curve pattern between price and the number of trips. This indicates that the higher the price, the lower the number of trips, which justifies using rentalTripsTaken to simulate demand. Looking at Figure 3, we see that 2012 was characterized by high trip demand. The sharp spike looks unnatural so we excluded 2013 data from the analysis. The daily demand continuously increases until mid-2017 and then starts going down. In addition, each year's seasonal change in demand is completely different, as shown in Figure 8 of the Appendix. Therefore, we only used 2017 data to build our demand curve instead of 6 years of data (2013-2018).

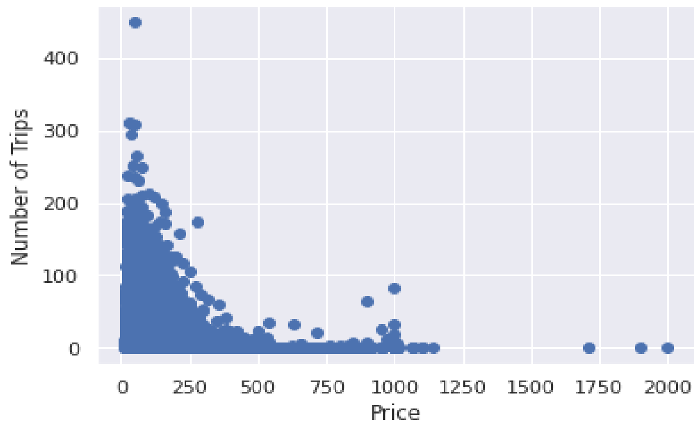

Figure 2: Scatter Plot of Price and the Number of Trips

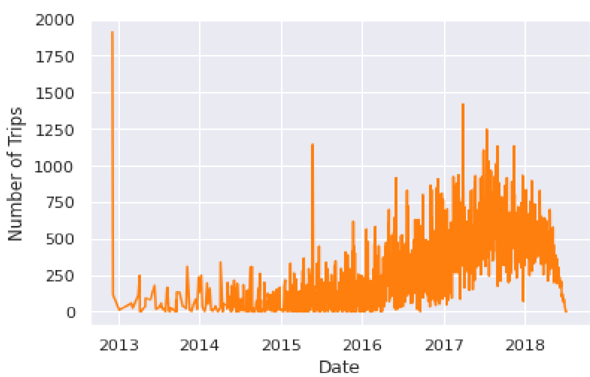

Figure 3: Trend of Daily Demand

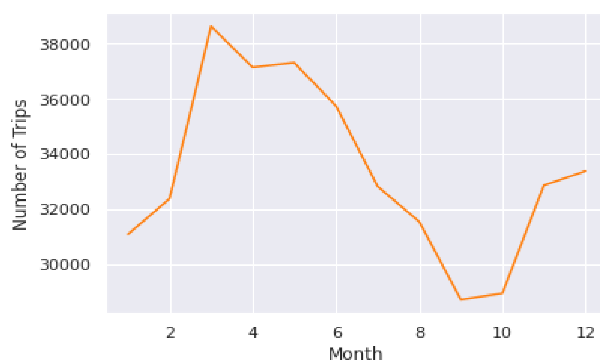

Figure 4: Overall Seasonal Trend of Demand
Note: This plot combines 2013–2018 results

Model Development

There are three steps involved in building our pricing model. First, we clustered cars into different categories using their ratings, price, brand, and other features. Second, we used log-log linear regression to find the price elasticity of each car category, based on which we estimated the demand curve. The demand function is denoted as D(p), where p is the average daily price. Lastly, we computed the optimal price (p*) that maximizes Revenue = D(p) × p.

- **Data description and cleaning**

We used the Turo data from Kaggle (Turo Rental Car Pricing Info, API code: kaggle datasets download -d theriley106/turo-rental-car-pricing-info). We extracted the data from the JSON file and transformed it into a data frame with informative variables that helped us formulate the following modeling. Since the car rental market is different across states in the U.S., we collected population data for each state from https://www.census.gov/. We used population as a confounding variable to control for the variance in the number of rental cars and economic levels among states.

- **Modeling and result interpretation**

  - **K-means Clustering**

To determine the optimal number of categories generated by k-means clustering, we computed the Within-Cluster-Sum of Squared Errors for different values of k to obtain the SSE-k scree plot. The KneeLocator function shows that k = 3 was the optimal classification parameter. The clustering result is shown in the table below:

Category		Review Count	Avg. Daily Price	Type
0	Spacious	8.170704	114.156455	3.869763
1	Luxury	5.599037	116.334889	1.062011
2	Economy	45.052825	52.689022	1.298117

Table 1: Means of Variables in Each Cluster

The algorithm clusters data points majorly by review count, average daily price, and car type. Based on the difference in cluster means, we labeled the clusters as Spacious, Luxury and Economy. Category 0 is labeled as Spacious due to its Type being larger than the other categories. We encoded car type with 1 for car, 2 for minivan, 3 for truck and 4 for SUV. The larger the value, the bigger the car. Category 1 is labeled as

Luxury because of its high average daily price. Category 2 is labeled as Economy since it has the lowest average daily price and highest review count.

- ## Log-log Regression Model

For the above three categories of rental cars, we ran a log–log regression model separately to estimate the elasticity of demand and price. We chose our independent variables in three aspects:

1) Owner Features: rating, reviewCount, and responseRate.
2) Temporal Features: listingmonth (seasonality), age_before_listing (the age of car before listed), and population (state's population density level to represent economic level).
3) Supply and Demand Dynamics: Number of available listings in the neighborhood (number of available rental cars in the 10-mile neighborhood).

The coefficient of the average daily price in the log–log regression model indicates the elasticity of demand and price. For the three categories in Table 2, we observed a difference in the sensitivity of change in demand relative to price. For Category 0 (Spacious), when the price of rental cars increased by 1%, the demand for this category decreased by 0.1735%. For Category 1 (Luxury), when the price of rental cars increases by 1%, the demand for this category decreases by 0.0934%. For Category 2, when the price of rental cars increases by 1% (Ecnomy), the demand for this category decreases by 0.0108%. The demand for Category 0 (Spacious) was found to be the most sensitive to price adjustments.

Table 2: Demand–Price Elasticity for Three Categories

Category	**Elasticity**
Spacious	-0.1735
Luxury	-0.0934

Economy	-0.0108

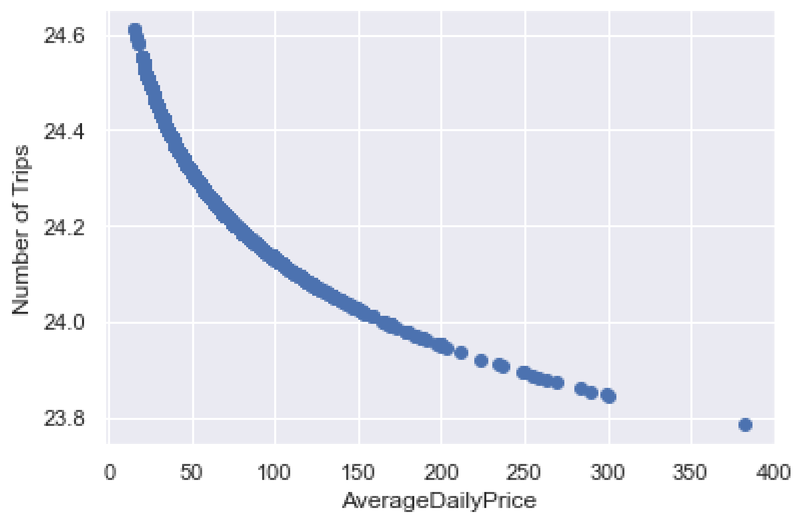

Figure 5-1: Demand–price curve for a car listing of Economy

The demand function is denoted as D(p, c) = $e^{(\beta_0 + \beta_1(ln(Price)) + \beta_2 rating + \beta_3 reviewCount + ... + \beta_8 num\_neigh)}$, where p represents the average daily price and c represents other control variables. For a car listing with rating = 5, reviewcount = 10, responserate = 100, age_before_listing = 1, population = 1424393, listingmonth = 5, num_neigh = 10, we plotted the demand curve as above. As we can see, the number of trips decreases very slightly, from 24.6 to 23.8, as the price increases from 0 to 399. This indicates that the demand is not quite sensitive to the price for this listing.

We then computed the optimal price (p*) that maximizes Revenue = D(p, c) × log(p). We took the log of price so that the magnitude of price does not override that of demand as the change in demand is negligible (23.8-24.6) compared with price in the graph above.

Here is a graph of the relationship between revenue and price for a listing of Spacious with the same control variables as the above listing. The optimal daily price is $318.

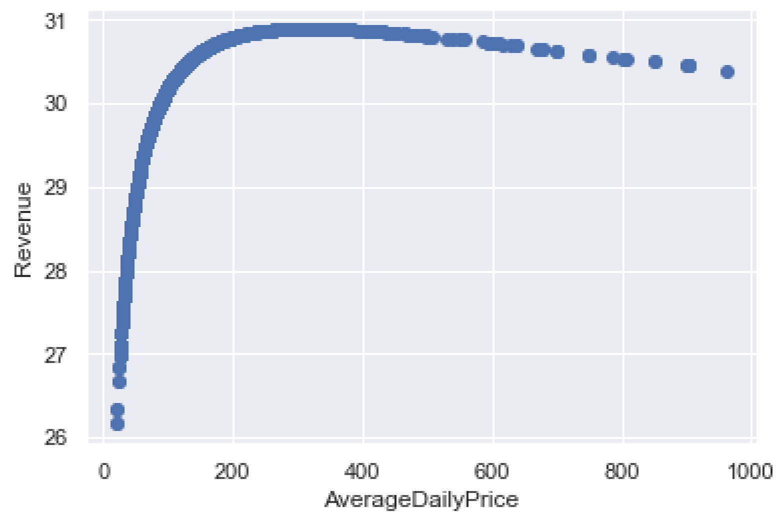

Figure 5-2: Revenue–price curve

Recommendation and Business Impact

This project generates two business impacts. First, Turo can deploy our clustering model and add Spacious, Economy, and Luxury filters to its search page. The category filter can help the renters find their ideal cars more quickly, especially when the users are unfamiliar with car models and brands, thus enhancing the overall customer experience.

Moreover, we created a demand function at a listing level to solve the price optimization problem. By inputting the features of a listing such as a host's response rate, the age of the car, and other control variables, the final output is the optimal price that maximizes the revenue. With this pricing model incorporated into its website, Turo is able to attract more users and improve its user experience.

Conclusion

In this report, we described the pricing challenges that Turo might be facing and introduced an optimal pricing model that consists of cluster analysis and demand curve estimation to assist the hosts with their pricing decisions. We applied K-means clustering to group the car listings into three categories:

Spacious, Economy and Luxury and built one log-log regression model per cluster to estimate the price elasticity for each car category. The final output of the pricing model is the optimal average daily price that maximizes the revenue given the estimated demand curve for a particular car listing. We will continue perfecting the pricing model by training the model on post-Covid data and exploring different demand functions that will potentially improve the accuracy of demand estimation.

References

Hamstra, Mark. 2019. Car Sharing Service Turo Relies on Data to Get Ahead in Crowded Rental Market. U.S. Chamber of Commerce. https://www.uschamber.com/co/good-company/launch-pad/turo-car-sharing-technology

Krugman, Paul; Wells, Robin; Graddy, Kathryn (2007). *Economics: European Edition*. Palgrave Macmillan. ISBN 978-0-7167-9956-6.

Maxime Cohen, Ilan Lobel, and Renato Paes Leme. 2016. Feature-based Dynamic Pricing. In Proceedings of the 2016 ACM Conference on Economics and Computation. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2737045

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. 2011. Improved Algorithms for Linear Stochastic Bandits. In Proceedings of the 24th International Conference on Neural Information Processing Systems (NIPS'11). Curran Associates Inc., USA, 2312–2320

Ye, P., Qian, J., Chen, J., Wu, C. H., Zhou, Y., De Mars, S., ... & Zhang, L. (2018, July). Customized regression model for airbnb dynamic pricing. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 932-940).DOI:https://doi.org/10.1145/3219819.3219830

Appendix

Table 2: Summary of Data Description

Variables	Definition	Example
averageDailyPrice	Rental price	28
rating	Rating of the car	5
renterTripsTaken	Number of renters	1
reviewCount	Number of reviews	1

responseRate	The percentage of host responding to a request	100
newListing	Whether it is new listing or not	TRUE
freeDeliveryPromotion	Whether the host offers free delivery	TRUE
instantBookDisplayed	Whether the car could be booked without wait time	TRUE
deliveryLabel	Delivery type	FREE DELIVERY
scalar	The distance driven during the rental	19
city	The location of the listing	Raeford
longitude	Longitude	-79.161022
latitude	Latitude	35.023165
state	State of the listing	NC
model	Vehicle model	Civic
make	Car brand	Honda
id	Car id	440235
listingCreatedTime	The time the listing is posted	1406583813600
year	Year of the car	2012
type	Type of the vehicle	car
automaticTransmission	Whether the car has automatic transmission	TRUE
population	State population	3109350

Table 3: Descriptive statistics of renterTripsTaken

rentalTripsTaken	
count	36279.0
mean	11.0
std	22.5
min	0.0
25%	0.0
50%	2.0
75%	11.0
max	450..0

Table 4: Descriptive statistics of averageDailyPrice

averageDailyPrice	
count	36279.0
mean	98.6
std	112.6
min	10.0
25%	39.0
50%	64.0
75%	110.0
max	1999.0

Figure 6: Bar chart showing the distribution of renterTripsTaken

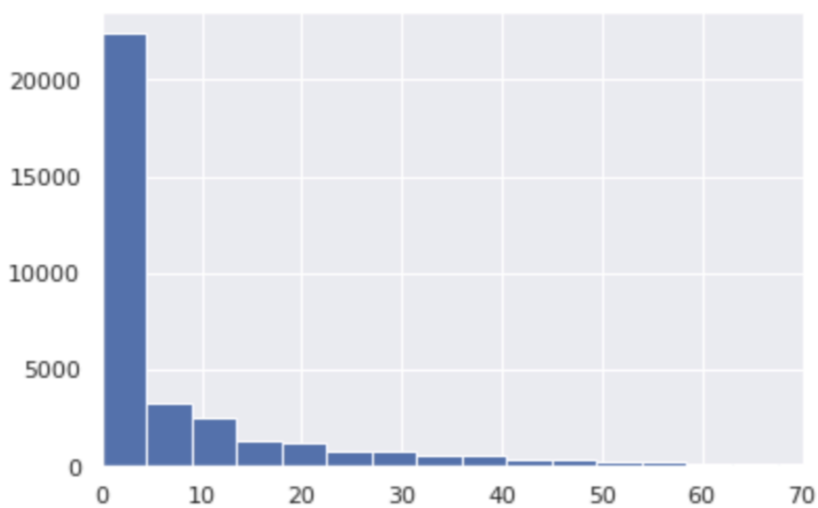

Figure 7: Bar chart showing the distribution of averageDailyPrice

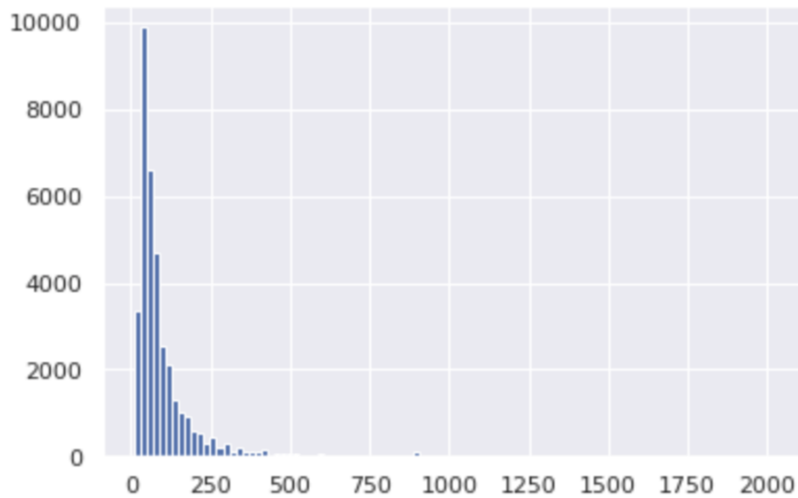

Figure 8: Seasonal Change in Demand by Year