Dynamic Pricing

Overview

Scope of this project is to design a pricing model that would determine a customer's propensity to pay and charge them accordingly while also catering to the supply/demand gap against each day of the week and hour of the day. This document covers the first phase of this project dealing with dynamic pricing on the customer level.

Strategy

The implementation comprises 2 descriptive and 1 predictive model. Results from both the descriptive models elaborating on a customer's transactional and demographic value are fed into the machine learning model to determine their propensity to pay and a surge multiple respective to each customer.

RFM Analysis *(Determines Transactional Behavior by scoring customers on Recency, Frequency and Monetary values of their trips)*
POI Modeling *(Infers Profession of Customers based on Frequency of Morning Trips to a specific Building POI)*
Machine Learning *(Neural Network based Prediction of Customer's Propensity to Pay and fare multiplier)*

With the aim of making this a supervised classification problem, customers are pre-labeled to train the model using the following definition.

Determining Propensity To Pay

A customer's willingness and ability to pay is gauged acoss 7 conditions and predictions are made on the 30 day active population against their lifetime behavior including Fulfillment, Cross Category Utilization, Promo Utilization and participation in historical marketing quest campaigns. 2.64% of our customers fulfilled the following conditions and this population imbalance is catered through oversampling in the machine learning phase.

Frequency Score = 3/4 *(Last 30 Days)*
Monitory Score = 4 *(Last 30 Days)*
Profession != 'Student' *(Last 15 days)*
No. of Low Bids < No. of HIgh Bids *(Last 30 Days)*
Avg KMs Per Trip > 5 KM *(Since Jan 22')*
Avg TPU > 8 *(Last 3 Months)*

Qualified Customers: 79,549

RFM Analysis

RFM Analysis is a quantitative ranking technique that scores the customers based on the recency, frequency and monetary value of their transactions. Data is distributed in 4 quartiles and the position of each customer with respect to the distribution determines their score which ranges from 1 to 4 with 4 being the highest. The scores are then concatenated and all possible permutations of these concatenated scores are grouped into the following categories:

Best Customer
Perfect RFM Score
RFM Score = 444

Big Value Provider
Good Recency and frequency, perfect monetary score.
RFM Score = 334/344/434

Loyal Customers
Perfect Recency and Frequency
RFM Score = 331/332/333/341/342/343/431/432/433/441/442/443

Need Attention
Average Recency and Frequency
RFM Score =231/232/233/234/241/242/243/244/311/312/313/313/314/321/322/323/324/411/412/413/414/421/422/423/424

About to Sleep
Poor Recency
RFM Score = 121/122/123/124/131/132/133/134/141/142/143/144/211/212/213/214/221/222/223/224

Lost
Poor RFM Score
RFM Score = 111/112/113/114

Based on Mobility trips from last 30 days (7th Oct 22 - 7th Nov 22'), following is the RFM categorization of our customers:

RFM Category	No. of Customers	% of Customers
Loyal Customers	19,832	6.00%
Big Earners	20,962	6.34%
Best Customers	44,234	13.37%
Lost	57,499	17.38%
About To Sleep	81,195	24.54%
Need Attention	107,080	32.37%
Grand Total	**330,802**	**100.00%**

POI Modeling

To infer the professions of our customers, closest POI mapping is done using data available through OpenStreetMaps. Out of more than 100k POIs tagged across Karachi in OSM, 1200 were buildings associated with the following categories:

School
College
University
Commercial Building
Office
Industrial
Stadium
Sports Center
Mall
Hospital
Government Office

Customers who took trips between 7:00 am to 11:00 am in the last 15 days (23rd Oct 22 - 07th Nov 22) are considered to distinguish morning users under the assumption that people travel to work or for education during these hours and the nearest POI is tagged with each trip's dropoff location. The coordinates with the highest frequency against each customer is considered as their primary dropoff location across trips and a profession is tagged to that corresponding POI. In the case of multiple POIs having the same frequency, multiple categories are tagged to the customers. Following is the distribution of the major segments identified:

Profession	No. of Customers	% of Customers
Working Profesional	31,089	52.39%
Student	13,665	23.03%
Health Professional	5,026	8.47%
Working Profesional, Student	2,666	4.49%
Working Profesional, Health Professional	1,284	2.16%
Athlete	866	1.46%
Student, Working Profesional	843	1.42%

As this exercise was initially conducted in march, Abbas made calls to a sample of these tagged customers and identified that in most cases 'Students' were being tagged accurately but 'Working Professionals' had cases of mistagging. Taking that into consideration, our outcome definition for dynamic pricing includes 'Not being a student' as a factor instead of considering 'being a working professional' to have more propensity to pay.

Machine Learning Model

Results from RFM and POI Modeling are used to determine our outcome definition which also includes average trips, avg KMs of customers and the frequency of high and low bids in the last 30 days. To determine the behavior of customers against this definition, following data points reflecting the customer's historical trends were considered.

Data Points

Days since Last TXN
Net Bookings Last 30 Days
Gross Bids
Net Bids
Lifetime Net Bookings
Lifetime Promo Trips
Acquisition Source
Age of Platform
Avg PC1 Commision per Trip
Frequent Morning Pickup Zone
Frequent Evening Drop Off Zone
Targeted In Marketing TPU Campaigns (Yes/No)
Targeted In Marketing ETA Campaigns (Yes/No)

Algorithm

MLP Neural Network

Imbalanced Class Problem

Since our definition involves 7 conditions, only 2.64% of our customers were initially labeled to have a propensity to pay which means that our algorithm would have been unable to identify the trends of the minority class. To cater to this issue, an oversampling technique called SMOTE was used. SMOTE creates synthetic (not duplicate) samples of the minority class. Hence making the minority class equal to the majority class. SMOTE does this by selecting similar records and altering that record one column at a time by a random amount within the difference to the neighboring records.

Results

Predictive Accuracy: *82%*

PropensityBands	No. of Customers	% of Customers
HighPropensity	2153	4.06%
LowPropensity	1479	2.79%
MediumPropensity	1478	2.79%
VeryHighPropensity	6462	12.19%
VeryLowPropensity	41415	78.15%
Grand Total	**52991**	**100.00%**

Abstract

WIth the aim of determining a customer's propensity to pay and apply dynamic pricing, we have implemented a combination of techniques to overcome our lack of demographic information pertaining to the customer. The RFM model gauges the transactional value of customers while frequency based POI tagging of customers allows us to infer their professions. Pairing this information with the historical trip level data of customers enables us to predict their propensity to pay and the probability of that outcome.

Based on this probability, a multiple is determined against each customer which would be added in the distance based estimated fare for each trip. This multiple would range from 0.01 to 1 meaning that the maximum surge a customer can incur would be 2X. However, we'll fix the upper limit to 1.5X and treat all customers with above 50% probability the same way.

Results:

Propensity	KHI	LHR	RWP	Total	**Multiplier**
Very High	5.49%	0.54%	0.54%	6.57%	**1.09**
High	5.06%	1.20%	1.28%	7.54%	**1.07**
Medium	3.54%	1.59%	1.85%	6.99%	**1.05**
Low	3.64%	1.83%	1.99%	7.47%	**1.03**
Very Low	32.12%	18.46%	20.85%	71.43%	**1**
% Total	49.86%	23.62%	26.53%	100.00%	

Propensity	KHI	LHR	RWP	Total	**Multiplier**
Very High	225,844	22,112	22,417	270,373	**1.09**
High	208,274	49,338	52,601	310,213	**1.07**
Medium	145,766	65,396	76,290	287,452	**1.05**
Low	149,710	75,335	82,067	307,112	**1.03**
Very Low	1,321,287	759,399	857,803	2,938,489	**1**
Total	2,050,881	971,580	1,091,178	4,113,639	
% Total	49.86%	23.62%	26.53%	100.00%	