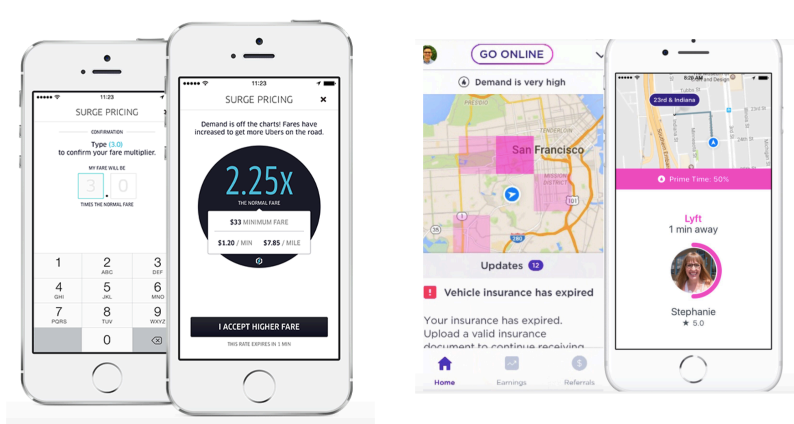
Dynamic Pricing using Xgboost and Dual Annealing

Abhinav Khare, Suirong Dong, Diala Gammoh, Amruth Sivelenka
University at Buffalo (SUNY)
abhinavk@buffalo.edu, Suiron.Dong@golfchannel.com ,Diala.Gammoh@golfchannel.com edu, AmruthSivalenka@golfchannel.com

Introduction

- Dynamic Pricing is a pricing strategy in which businesses set flexible prices of products and services based on market demands in order to maximize revenue.
- Dynamic Pricing was traditionally popular in airline and hotel businesses.
- However with advent of e-commerce a lot of digital businesses like retail & ridesharing apps do regular dynamic pricing.

Popular Brands that use Dynamic Pricing



Popular Ride Sharing Services use Dynamic Pricing

Price optimization problem

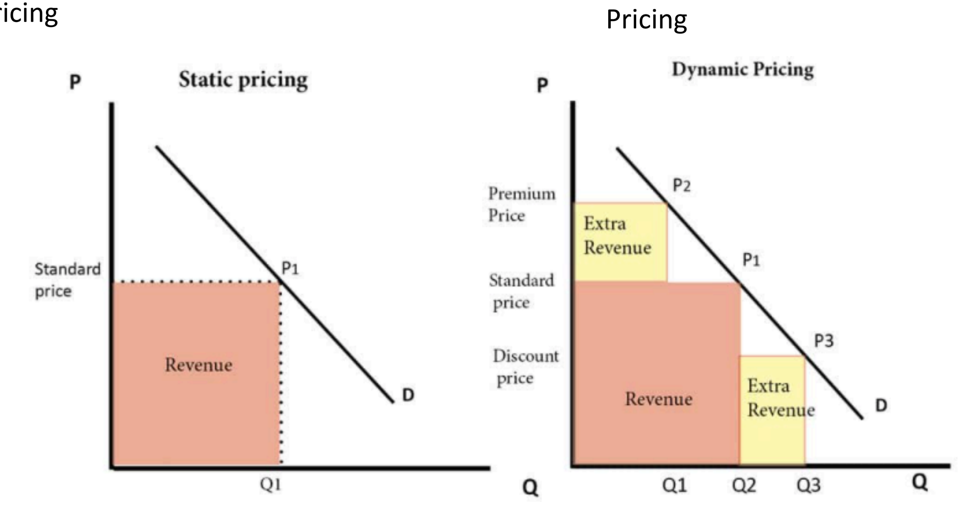


Illustration explaining how dynamic pricing leads to higher revenues

$$\max \sum_{t=1}^{T} r(t, d(t))$$

Maximize Revenue in sales horizon *(1…T)* where *r(t,d(t))* is revenue in period *t*

$$\text{s.t.} \sum_{t=1}^{T} d(t) \leq C$$

Demand *d(t)* in a period *t* should not exceed capacity *C*

$$d(t) \geq 0.$$

Demand induced in period *t* by the price *p(t)*

Marginal Revenue = $MR(t) = \delta r(t)/\delta d(t) = p(t) + \delta p(t)/\delta d(t) * d(t)/p(t) = p(t)\ (1 + 1/e)$

Price elasticity of demand = $e = \delta d(t)/\delta p(t) * p(t)/d(t)$

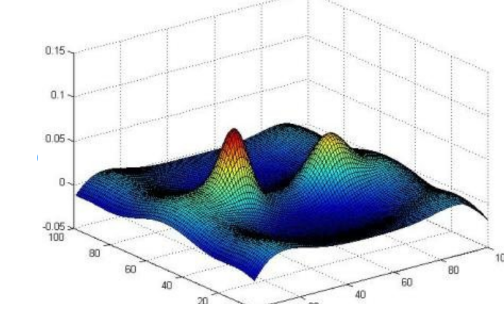
At Optimality Condition : Marginal Revenues are equal in all periods

Motivation & Objective

- Improve the methodology of pricing a digital product in the sports and recreation industry.
- Current method is a greedy allocation algorithm that allocates demands/ sets price such that the marginal revenue for each period is equal.
- The elasticity capturing the relationship between price, demand & marginal revenue is determined by analysis of historic data.
- Factors influencing elasticity like weather, season, market etc. are determined manually by experts analyzing data.
- We propose a machine learning method that captures the relationship between price, demand and other factors and eliminate explicit calculation of elasticity and factors affecting it.

A Two-Stage Methodology for Dynamic Pricing

Stage 1: Machine Learning Model



- **Input :** Price for given Location, Weather, Market Tier (factors affecting elasticity)
- **Output :** Demand/Revenue for Location, Weather, DaysOut, Tier etc.
- **Function :** Learn the mathematical function (complex demand curve) to map input to output from historical training data.

Stage 2 : Optimization Model

- **Input :** Range for price values, values of variable like weather, location, machine learning model
- **Output :** Optimal Price(maximizes revenue)
- **Function :** Searches the price demand curve stored in the machine learning model to find the optimal price.

Stage 1: Xgboost

Xgboost is a library that implements machine learning algorithms under the gradient boosting framework.

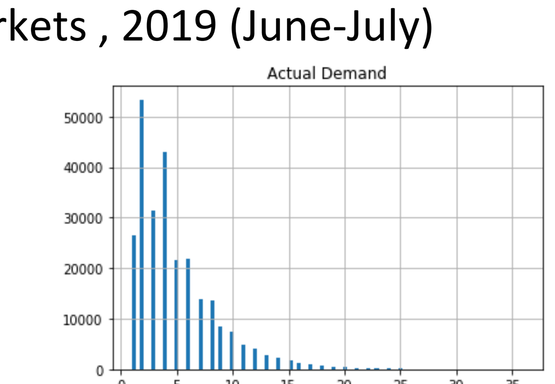
Features/ Independent Variables

- **Factors of time:** WeekNumber, Month, Day , DayNameBucket, IsPlaydayWeekend
- **Factors of Weather:** WeatherPlayableDays, WeatherPlayableRound, MaxTemperatureDegF , WeatherRainDay
- **Factors of Location:** Latitude, Longitude
- **Other Factors:** Tier, DaysOut
- **Price:** Trade Price of each DaysOutGroup

Training Data : All Play Locations , All markets , Year 2018 (all months) 2019 (Jan-May)
Test Data : All Play Locations , All markets , 2019 (June-July)

Predicted Variable : Demand

Grouping Level : PlayLocationKey, Year, WeekNumber, DaysOut, DayNameBucket,



Frequency Distribution of Actual Demand

Result on Test Data RMSE on test data = 1.65, MAPE = 30 %

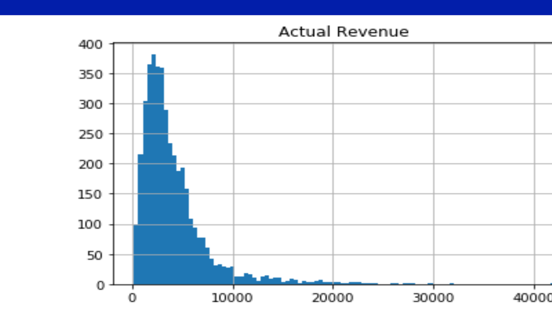
MarketName	LocationKey	Year	WeekNumber	DaysOut	DayNameBucket	Tier	Actual Demand	Predicted Demand
Oregon	2	2019	22	2	Tue_Thu	1	1	1.4441173
Phoenix	4	2019	23	4	Sat_Sun	1	3	3.3126905
Las Vegas	6	2019	23	0	Tue_Thu	1	13	13.655126
Dallas	30	2019	25	0	Sat_Sun	2	5	4.5798197
San Francisco	26	2019	23	0	Mon_Fri	1	3	2.7931218
Orlando	254	2019	28	1	Sat_Sun	1	8	8.320353

Sample of the results comparing actual demand and predicted demand

Best Model using Cross Validation

Predicted Variable : Revenue

Grouping Level : PlayLocationKey, Year, SeasonID, DayNameBucket, DaysOutGroup

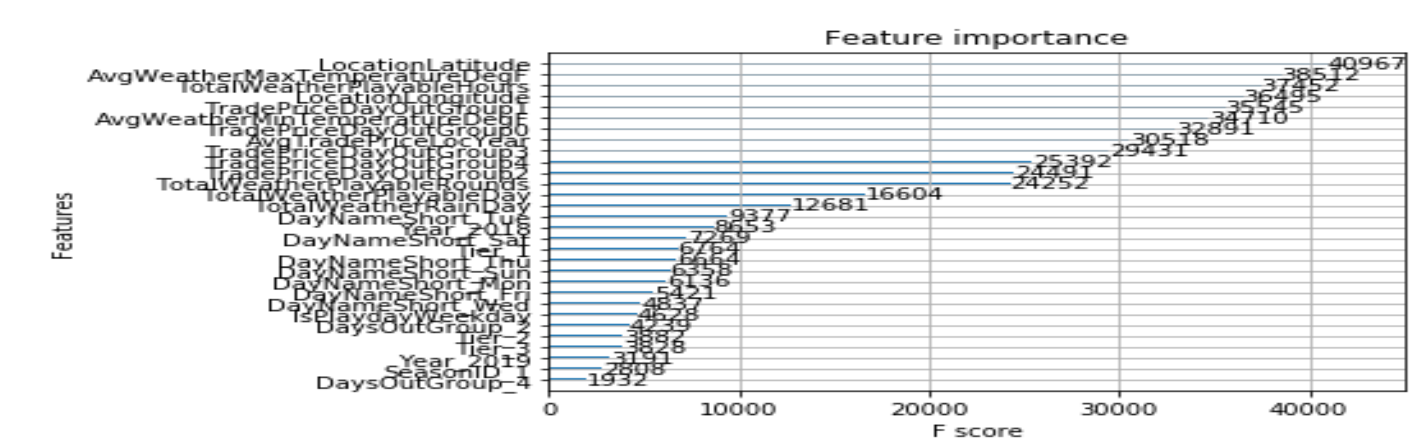


Frequency Distribution of Actual Revenue

- **Result on Test Data :** RMSE = 56.26 , MAPE = 12.73 %

Table comparing actual revenue and predicted revenue

Market Name	Location Key	Year	SeasonID	DayName	Tier	DaysOutGroup	Actual Revenue	Predicted Revenue
Orlando	2161	2019	1	Thu	1	0	120	120.024
Austin	2006	2019	2	Wed	2	2	224.4	763.4081
Chicago	11399	2019	2	Sun	2	0	502.4	502.477
Houston	6541	2019	2	Sat	1	0	358	358.0023



Stage 2 : Dual Annealing

Dual Annealing is a global optimization heuristic that combines the generalization of classical simulated annealing (CSA) with fast simulated annealing (FSA) coupled to a strategy to apply local search on accepted locations.

Results of optimization on all Locations in all markets , 2019 (June- July)

	Mean Absolute Percentage Change in Price	Root Mean Squared of Change in Price
TradePrice for Days Out Group 0	29.94	9.14
TradePrice for Days Out Group 1	30.3	10.42
TradePrice for Days Out Group 2	29.89	11.6
TradePrice for Days Out Group 3	32.49	13.05
TradePrice for Days Out Group 4	30.68	14.06

Increase In Revenue estimated	22%
With error adjusted	20%

Sample Results for Orlando

MarketName	LocationKey	Year	SeasonID	DayNameBucket	Tier	DaysOutGroup
Orlando	2161	2019	1	Sat_Sun	1	1

Optimal Trade Price Days Group 0	13.94971
Current Trade Price	10
LowerBound 0	5
UpperBound 0	15
Optimal Trade Price group 1	15.63965
Trade Price Days Out Group1	10.72222
LowerBound 1	5.361111
UpperBound 1	16.08333

Optimal Trade Price Days Group 2	19.23456
Current Trade Price DayOut Group2	16
LowerBound 2	8
UpperBound 2	24
Optimal Trade Price Days Group 3	21.6847
Current Trade Price DayOut Group3	16
LowerBound3	8
UpperBound3	24

Optimal Trade Price Days Group 4	23.84759
Trade Price Days Out Group4	18
LowerBound 4	9
UpperBound 4	27
Demand	31
Optimal Demand	36.61115
Revenue	326
Optimal Revenue	572.5855