Dynamic Pricing - Supply/Demand Clustering

Business Challenge

Uniform Fares across the platform for all customers has become a challenge which brings the logic of meeting supply and demand into question.To solve the problem, the basic principle of **Marginal Propensity to pay** is to be determined using a combination of **descriptive** and **Predictive** approach.This Document focuses on the Descriptive part of the Model.To understand the **Spread of demand and supply** across different **Zones, Hours** and **days** an **unsupervised descriptive** approach is used to create **clusters** of data in order to **apply** different multipliers.

Methodology

We have used a hierarchical approach for clustering using Machine learning,This Algorithm was used because it gives the user the most control over the clusters and provides a clearer picture of the clusters.One Drawback is that it can not be used on Big Data as the computational Power is very big.This will help us in setting a different multiplier based on the demand and supply in a specify zone, hour and day.Agglomerative hierarchical clustering (AHC) is a clustering method that merges two objects or nearby clusters into one cluster, recursively, so that one big cluster that contains all the objects in the data is formed

Aggolomerative approach for Hierarchical clustering using Ward's Method and Average linkage is used in the Model.

Multiplier Matrices

To determine the Multiplier we have grouped the data into different clusters based on 5 different Matrices.These Matrices define the supply and demand.Data is collected over a span of 30 Days.The data is grouped by Day,Hour and Zone.

Gross Trips : Average Gross Trips

Gross Customer : Average Gross Unique Customers

Fulfillment : NET/Gross

Real Fulfillment: Net trips / Unique Gross Customers

Active Drivers : Average Drivers who received Ringer

Expired Trips: Average Expired Trips

Canceled Trips: Average Canceled Trips

Machine Learning Model

Agglomerative hierarchical clustering (AHC) is a clustering method that merges two objects or nearby clusters into one cluster, recursively, so that one big cluster that contains all the objects in the data is formed.There are two parameters that are used to measure the proximity between two objects or clusters, namely the distance measure and the linkage method. Distance measure is used to measure the proximity between two objects or clusters which contain only two objects if it is merged. The linkage method is used to measure the proximity between two clusters which when merged the number of objects is more than two. Output of AHC is a dendrogram which is a hierarchy diagram that summarizes the process of clustering. T.

The model uses euclidean distance between data points to create clusters.We have used two methods to create clusters,**Average linkage** method takes the average distance between all points and clusters points with minimum distance.This Method is strong against outliers in the data as outlier effect is canceled by averaging out the distance.**Wards Method** takes all points as clusters and then takes the centroid of the two nearest clusters and uses the centroid as the new point of computation for the next cluster formation.

Input: A set X of objects $\{x1, \ldots, xn\}$, distance function $dist(c1, c2)$

Procedure: for i = 1 to n

$c1 = \{xi\}$

end for

$C = \{c1, \ldots, cn\}$

$I = n + 1$

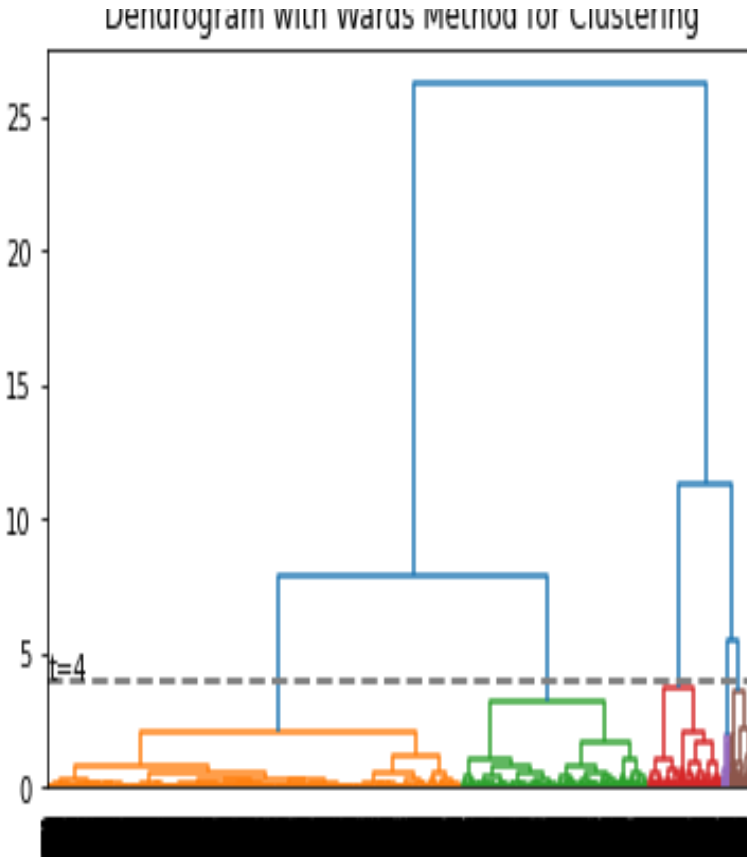
while $C.\ size > 1$ do

$(cmin1, cmin2)$ =minimum $dist(ci, cj)$for all ci , cj in C

Remove $cmin1$ and $cmin2$ from C Add $\{cmin1, cmin2\}$ to

C $I = I + 1$ end while

Output: Dendrogram



Results

5 Clusters were created with Below Mentioned Characteristics.

Clusters	Gross Trips	Gross Customer	Expired Trips	Canceled Trips	Active Drivers	Net
1	406	268	38	184	433	184
2	604	417	41	252	700	311
3	31	25	1	11	66	19
4	234	190	4	70	447	159
5	423	331	10	141	696	272

Abstract

Dynamic Pricing is a mode which brings together Demand and Supply based upon individual Customer level and Geo location. The problem is how to compute the equilibrium price with the large number of participants in short a need for a self learning model. For m driver n rider at the same time period,in a specific location there is a need for varying prices based upon the characteristics of customer and Location. This Document proposed Agglomerative Hierarchical Clustering (AHC) to be applied for determining clusters of Zones with the same characteristics. By applying AHC, the zone which is not feasible based on their day and time is proposed to be a Multiplier in price.

To decide pairs that will have a higher Multiplier Real FF, Gross Trips and are taken into account. As a result, we obtain a number of optimum clusters.The multiplier will be applied against each trip.The multiplier will range from 5% to 25%.

Clusters	Gross Trips	Net	FF	Real FF	Surge
4	234	159	68%	83.67%	5%
3	31	19	61%	76.63%	10%
5	423	272	64%	82.31%	15%
2	604	311	51%	74.62%	20%
1	406	184	45%	68.36%	25%

Reference

Application of agglomerative hierarchical clustering

https://iopscience.iop.org/article/10.1088/1742-6596/1821/1/012016/pdf

Fact Table for

Pipeline

1.Extraction for Input

Data Extracted from Bigquery For EDA and Model Training

2.Push to Git

Directory created on git and VS. Model saved in Artifactory and synced with Tellos Live surge pricing implementation.