

# Market Analysis for Pressed Juice Chain in London

Nikhil Shah

June 25, 2019

## Introduction

### Background

Cold-pressed juice is made by extracting juice from fruits and vegetables using hydraulic press which gives 100% juice, as opposed to other methods. It can be stored in a refrigerator for longer period. And also, there are lot of other health benefits of pressed juices which could be very essential for today's world where people are suffering from so many health diseases. A healthy alternative to so many unhealthy diets has become a need of an hour. A chain of pressed juice outlets seems to have a lot of scope for business and we need to do market analysis to identify best place to start from.

### Problem

This project aims to identify the key markets to launch the cold-pressed juice chain in London for which it would be essential to take into consideration certain factors like demographics, such as Gross Income, Health concerns etc., competitors and the overall activeness of the consumers in the market.

### Interest

This analysis would certainly be of interest to young entrepreneurs who are excited to venture into this potential market of Juices.

## Data

### Data Sources

The demographic stats about each borough of London was readily available from London Govt. website [here](#). For the data around list of Boroughs, It was scrapped from Wikipedia with not so accurate co-ordinates hence the co-ordinates were fetched later using Google API. For competitor analysis, data about Juice Bars, Coffee Shops, and Cafés was taken using Foursquare API.

## Data Discussion

All the data was merged into one dataset after adequate cleaning. Important variables like Gross Pay, life expectancy, anxiety score, carbon emission, Obesity Index etc. were substituted from the entire Borough profile data, as they were most relevant to our analysis. Moreover, Foursquare API was leveraged to get venues of competitors, limited to venues in the category of Coffee Shops, Café, and Juice Bar, around the boroughs. The details of Venue like count of ratings, likes and tips was also collected, assuming that it shows the popularity of store as well as activeness of consumers in the market.

## Data Processing

London Borough Profile Dataset had 84 variables, not all variables were important to us. So, I identified key and most relevant variables that would be essential for the analysis.

Variables that were included are:

### Health Concerned variables

People\_aged\_17+\_with\_diabetes\_(%), Childhood\_Obesity\_Prevalance\_(%),  
Anxiety\_score\_2011-14\_(out\_of\_10), Happiness\_score\_(out\_of\_10),  
Female\_life\_expectancy,, Male\_life\_expectancy,, Total\_carbon\_emissions

With these variables we'll be able to measure the need of particular market as the assumption is that people who are prone to such unhealthy environment would be the right audience to target for Juices. Consumer would want to bring juices in their routine lifestyle in order to keep their health in control.

### Scope defining variables

Two-year\_business\_survival\_rates\_, Jobs\_Density, Gross\_Annual\_Pay,  
Employment\_rate\_(%)\_, Average\_Age, Population\_density\_(per\_hectare)

These variables would define scope/potential of the business. It signifies the purchasing capacity, assuming that market with higher gross pay, better employment, and large population, would certainly have better sales.

### Competitors analysis variables

Price, likes, tip\_count, rating

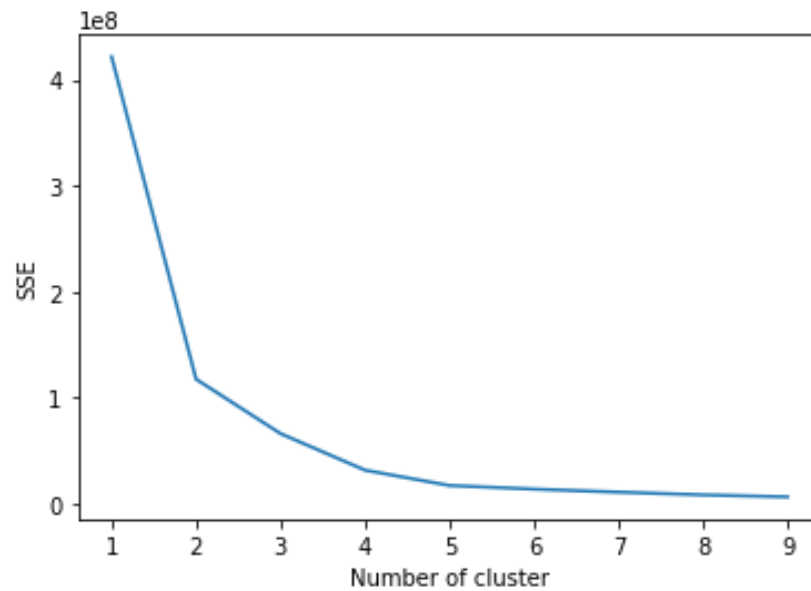
These variables are scraped using Foursquare API. Assuming that Coffee Shops, Cafés, and Juice Bars are the direct competition to our potential business. Lot of Café offers juices and also Coffee is the alternative of juice, hence it would be safe to assume them as competitors. Moreover, the variables such as ratings, likes, and tips shows the performance and popularity of the venues and also, in some sense it gives idea about how active the consumers are in that market. These details about each venue were fetched over the course of 3 days because of the limited premium calls per day which is 500

## Methodology

In order to derive best markets for launch of Juice Bar chain, I segmented the market based on the variables that we grouped together for our analysis. As we know, Market Segmentation can be done by clustering, I used K-Means clustering, very popular unsupervised machine learning method, which divided 32 boroughs of London in K similar clusters. Those clusters gave clusters with similar attributes and considering our idea about an Ideal market for Juice Bar, I decided the best cluster or the market.

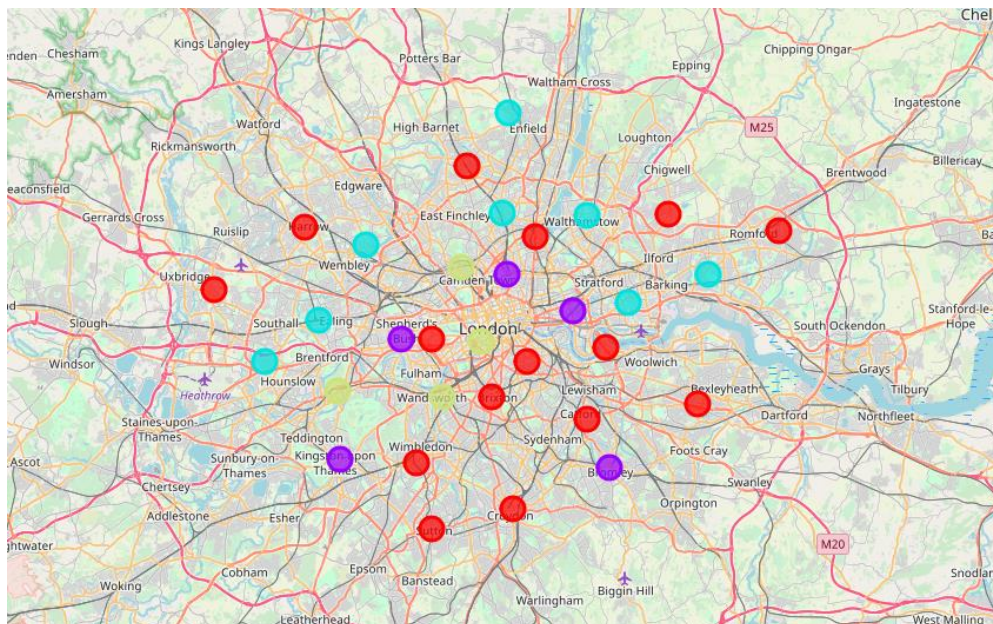
Looking at K-means clustering algorithm, I needed to decide the optimal K, which is number of clusters. The optimal k is the one which has minimum sum of distance of the points from the center its cluster. There are several ways to find the optimal K, one would be the trial and error, wherein you try putting certain range of K and check each sum of distance to decide on value of K. One which I used is elbow method, which does pretty much same thing except it automates the process using loop and gives you graph plotted with Ks on X-axis and SSE on y-axis, it gives you quick idea on best K values.

Here's the graph from the elbow method,



Looking from the graph 4 clusters seems the best

Based on the elbow method I chose 4 as the value of K to put in algorithm. After running the K means clustering, I got the following results



## Results

Variables	Cluster 0	Cluster 1	Cluster 2	Cluster 3
price	1.078493	1.086406	1.047780	1.077778
likes	26.728167	36.290222	25.461422	36.563953
tip_count	9.815906	12.633945	8.627013	13.215009
rating	6.203524	6.642259	6.722566	7.332531
Café	0.472533	0.565814	0.515335	0.580949
Coffee Shop	0.430799	0.379635	0.415988	0.307730
Juice Bar	0.096668	0.054551	0.068677	0.111321
People_aged_17+_with_diabetes_(%)	6.153333	5.220000	6.937500	4.050000
Childhood_Obesity_Prevalance_(%)	22.626667	20.820000	25.450000	19.525000
Anxiety_score_2011-14_(out_of_10)	3.300000	3.340000	3.162500	3.450000
Happiness_score_(out_of_10)	7.260000	7.260000	7.225000	7.225000
Female_life_expectancy,	84.133333	83.920000	83.887500	85.675000
Male_life_expectancy,	80.393333	80.040000	79.625000	81.500000
Total_carbon_emissions	1027.800000	1049.600000	1062.125000	1404.500000
Two-year_business_survival_rates_Jobs_Density	74.206667	74.200000	73.625000	74.250000
Gross_Annual_Pay	33324.000000	37344.600000	30187.125000	41269.250000
Employment_rate_(%)_Average_Age	73.720000	74.040000	70.725000	73.300000
Population_density_(per_hectare)	36.733333	35.840000	34.887500	36.975000
	69.073333	98.240000	68.500000	88.025000

The above table contains the mean values in each cluster that we got from K-Means clustering. Looking at clusters, it can be observed that Cluster 3 looks very vibrant in terms of responses the venues get from the consumers, considering likes, tip counts, ratings. Also, looking at the demographic factors of Cluster 3, it has highest life expectancy across genders. Looking at the high gross pay in cluster 3 it can be believed that people are rich in this clusters and generally rich people would a car which could be the reason for high carbon emission, although can't blame the rich people without the data about no. of car owners in this clusters. The job Density is also high, which could also be one more reason to say that cluster 3 has high purchasing power. Almost everything is being high in Cluster 3, however, the most important factors which could

stand out as USP's for our business which are people with diabetes and childhood obesity, are the lowest. Those people could be one of our target audience, however low diabetes and obesity could also mean that people are health conscious who would be easy to cater with health concerned products. This hypothesis could be confirmed by looking at the data about no. of gym users.

Hence, considering all the factors from all the clusters, the Cluster no. 3 clearly stands out, with everything being high, although there's high proportion of Juice Bars as well in Cluster 3 which signifies high competition but with other factors in place, competition could be well defeated with right strategy.

The list of Boroughs in cluster 3 are

**Camden, Richmond upon Thames, Wandsworth, Westminster**

From the map we can see that those are kind of in the center of London as well.

## Conclusions

In this study, I did pretty basic market analysis of the Boroughs of London for the launch of Cold-pressed Juice Bar chain. I identified the key markets to start with, by clustering using K-means cluster algorithm. By clustering, I grouped together the similar boroughs, through which I compared various factors that could affect our potential business. In the comparison, I observed Cluster 3 clearly being better for the Launch of our business as compared to other clusters. The second market, although, to look at should be cluster 1 after the success in Cluster 3 market.