FeyNN Labs Final Assignment

Market Analysis of Tea Production and Consumption in South India

Contributors: Bitan Biswas, Siddharth Rajvanshi, Nikky Sam Mathew, Kinshuk Gaurav Singh
Project Link: https://github.com/bitanb1999/FeyNN

Background

Tea belongs to the family Camelliaceae and all the cultivated tea plants belong to two distinct species, viz., Camellia sinensis (L). O. Kuntze, the short leaved 'China' plants and Camellia assamica (Masters) Wight, the broad leaved 'Assam' plants. The 'Cambod' variety, a subspecies of the latter, is named C. Assamica lasiocalyx (Planchon exwatt) Wight. The 'China', 'Assam' and 'Cambod' and a large number of their hybrids are seen in many tea fields. It is believed that many wild species of teas have also contributed to the present-day hybrid population of cultivated tea plants.

The widespread occurrence of the leaf rust (Hemilia vastatrix Berk & Br.) Of coffee and the consequent decline of the coffee industry was a major factor responsible for the extensive planting of tea in south India. The tea growing tracks of south India, extending along the Western Ghats, vary in their elevation from 300 to 2,300 m above MSL and experience an annual rainfall ranging from 90-750 cm. These plantations, with their adjoining forest ecosystem contribute greatly to the maintenance of terrestrial ecology by providing extensive land cover and minimizing soil erosion.

In this paper, our motto is to obtain information on the present market of Tea in South India using Business Analytics and Market Segmentation techniques. A detailed analysis of the various aspects of South Indian tea and its customer base is carried out through machine learning techniques such as Clustering. The case study focuses on the clustering pipeline using limited amount of data obtained from several trusted platforms, including government open source.

Data

The data used in the report are obtained from the following sources:

1. Kaggle (https://www.kaggle.com/srinivas1/agricuture-crops-production-in-india)

2. Tea Board of India (http://www.teaboard.gov.in/WEEKLYPRICES/2021)

3. Mendeley Data (https://data.mendeley.com/datasets/rcgvn92yxx/1)

4. Open Government Data (https://data.gov.in/keywords/tea)

Information on exports, weekly prices, customer base of supplementary products such as coffee, consumption information and production information along with weather conditions is obtained from the platforms. Further, alternative data of tea market is also obtained for Kenya, due to easy availability and its similarity in weather conditions certain hot, sultry areas of South India. The data is partly used for visualization purpose and partly for clustering.

[47]: df

[47]:

	Date	Centre_Name	Commodity_Name	Price	year_x	zone	tea_loose	year_y
0	2009-07-01	CHANDIGARH	Tea Loose	NaN	2009	1	19500.0	2009
1	2009-07-01	DELHI	Tea Loose	12500.0	2009	1	12500.0	2009
2	2009-07-01	SHIMLA	Tea Loose	NaN	2009	1	9500.0	2009
3	2009-07-01	SRINAGAR	Tea Loose	NaN	2009	1	73200.0	2009
4	2009-07-01	JAMMU	Tea Loose	NaN	2009	1	NaN	2009
...	...	...	...	...	...	...	...	...
4377	2012-11-07	SHILLONG	Tea Loose	NaN	2012	3	NaN	2012
4378	2012-11-07	AGARTALA	Tea Loose	13250.0	2012	3	13250.0	2012
4379	2012-11-07	ERNAKULAM	Tea Loose	12300.0	2012	4	12300.0	2012
4380	2012-11-07	CHENNAI	Tea Loose	NaN	2012	4	NaN	2012
4381	2012-11-07	HYDERABAD	Tea Loose	17900.0	2012	4	17900.0	2012

4382 rows × 8 columns

Fig 1. Glimpse of Data used for Clustering

Python libraries such as NumPy, Pandas, Scikit-learn and SciPy are used for the workflow and the results obtained are ensured to be reproducible.

EDA

We start the Exploratory Data Analysis with some Principal Component Analysis of different parameters in the dataset obtained from the combination of all the data we have. PCA is a statistical process that converts the observations of correlated features into a set of linearly uncorrelated features with the help of orthogonal transformation. These new transformed features are called the Principal Components. The process helps in reducing dimensions of the data to make the process of classification/regression or any form of machine learning, cost-effective.

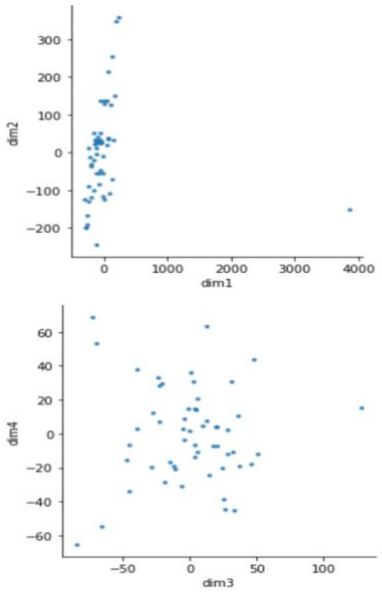

Fig 2. Unfruitful PCA

The data is sparse and has very few parameters(approximately 9 or 10) and hence, PCA doesn't prove to be much useful in dimension reduction.

In India, multiple crops are grown due to the favourable weather conditions and fertility of soil. Tea and Coffee in general are grown in the cooler parts of the country, hills, to be specific due to the easy terrace plantation

availability and slanting land which provides minimum water clogging facilities to the tea plantations. South India has ample number of such hilly areas for the tea production to flourish. These are:

	Hills
1	Wayanad(Kerala)
2	The Nilgiris (Tamil Nadu)
3	The Anamallais (Coimbatore District, Tamil Nadu)
4	Nelliampathy (Palghat, Kerala)
5	High Range (Iddukki District, Kerala)
6	Vandiperiyar and Peermade (Iddukki District, Kerala)
7	High Wavys (Madurai District, Tamil Nadu)
8	Trivandrum (Kerala)
9	Singampatty (Tirunelveli, Tamil Nadu)
10	Coorg (Karnataka)
11	Hassan (Karnataka)
12	Chikmagalur (Karnataka)

The overall of production of Tea in India is on the lower side compared to the other crop productions and this is because Tea plantations require high maintenance and aren't cost-effective in certain places due to unfavourable weather.

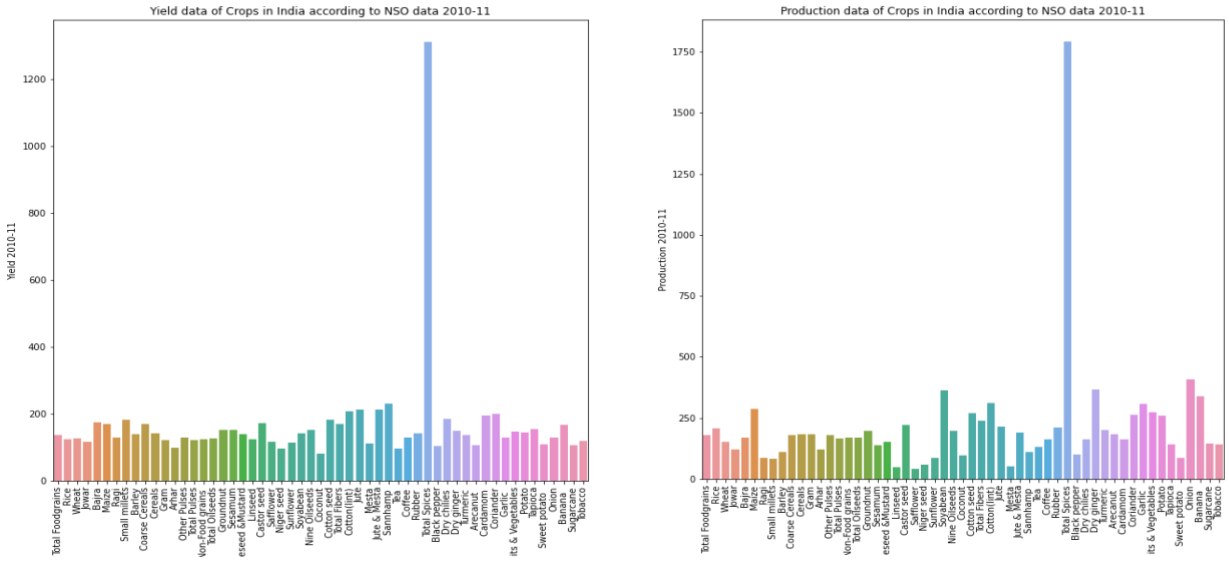

Fig 3. Yield and Production of all Crops in India with Tea being on the lower sides

The lower yield and production of tea in India as whole, indicates a lesser priority to the same in South India as well. This can be caused by various reasons such as the MSP or Minimum Selling Price allotment of tea by the government, weather conditions and production cost, human and financial capital availability and much more.

The Market share of Tea is 1% compared to 14-15% of Spices and condiments in India. This leaves scope for the growth in tea market in India. Further, we know that the richest condiments and spices markets are in Southern India. This intuitively means that due to higher profits in Spice productions, the southern farmers prefer spice and condiment production over Tea.

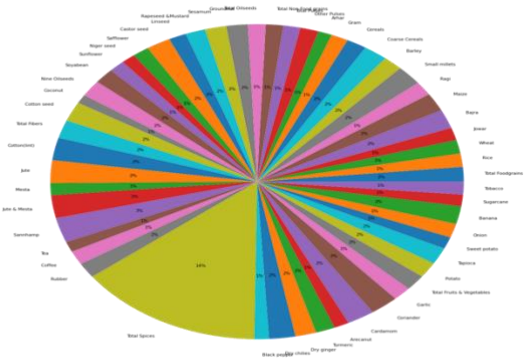

Fig 4: Market share of Tea is 1% compared to 14% of Spices

Further, alternatives to tea like coffee are also widely grown in South India. People who prefer coffee, particularly Arabica, may have a liking for tea due to similar caffeine content. South Indian coffee, popularly known as the South Indian filter coffee, has 10-30% chicory added to it along with Arabica coffee beans, thus making it comparable to green tea or milk tea caffeine content. Hence, a detailed look into Indian coffee market, also gives us an estimate of the South Indian tea market, given almost 90% of Indian coffee is produced in South India.

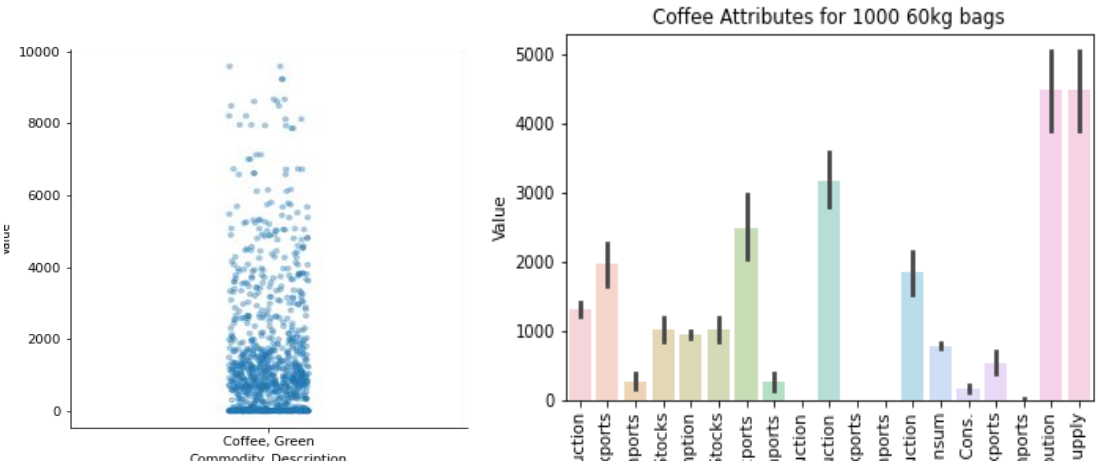

Fig 5. Type of Coffee Grown in South India and Auction, Export-import, Stocks, Supply and other attributes associated with coffee cultivation in India.

The value component essentially means the monetary capital (in lakhs) associated with 1000 60kg bags of Coffee grown per year in India (over a period of 5 years). Also, we notice that the export of coffee is much higher than the import of coffee and this implies the presence of a sustainable international customer base for Indian tea as well as coffee, grown in South India. Thus, to incentivize the farmers in South India, essential customer base knowledge should be provided to them to encourage them to produce more amount of tea than produced earlier or presently.

Indian export docks are located, mostly in the lower and eastern parts of India due to peninsular structure of the country. This ensures adequate exportation, importation, or transport facilities for tea producers, nationally as well as internationally.

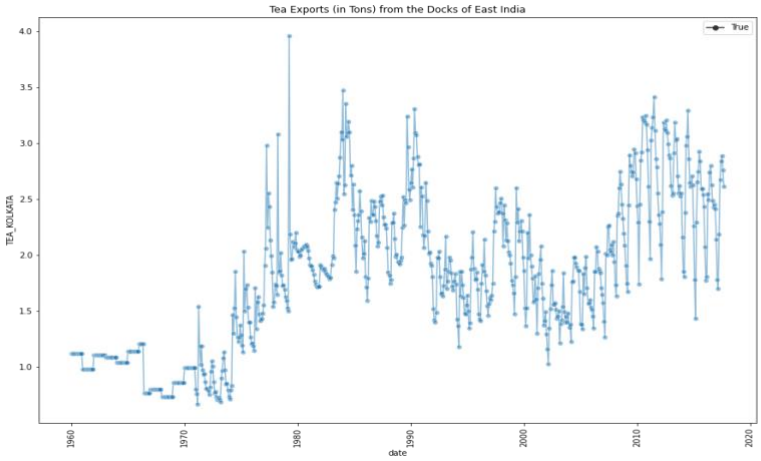

Fig 6. Tea exports from Eastern Dock of our country over past 60 years

The tea exports from eastern dock of our country are on an average in the range of 2-4 tons every year, implying that the average tea export from south India is also the same, if not more, given the abundant number of docks or harbors in Andhra Pradesh and Kerala.

An overview of the Tamil Nadu tea production is also obtained due to the availability of Tamil Nadu data on open government platforms. Tamil Nadu, as the graph suggests, has its highest production of tea in the Nilgiris and Tirunelveli. This might be due to the favorable weather conditions in these areas. However, the other parts of Tamil Nadu, too, on average allot 1000 acres of land to tea production, irrespective of presence of favorable conditions. This can act as a model to implement in the other South Indian states.

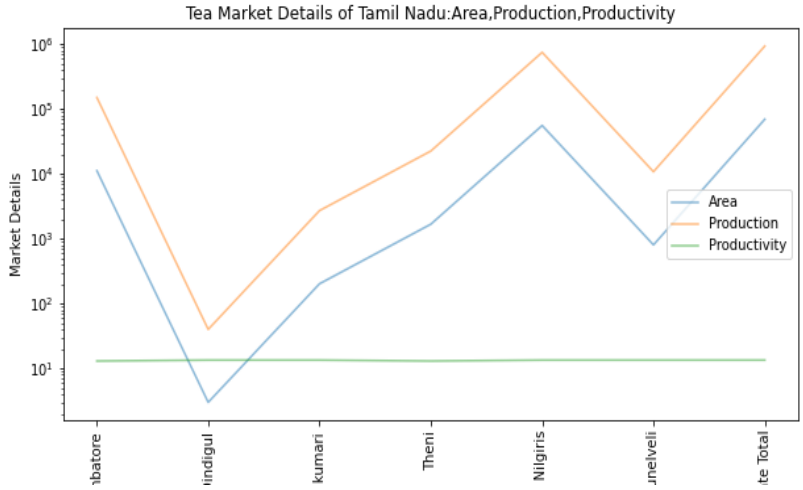

Fig 7: Area of Production of Tea in Tamil Nadu

Additionally, the weather, as stated earlier, in South India is perfect for tea cultivation. The temperature in South India is ideally around 20-25 degree Celsius and the weather is apt as it fulfills the three-criterion required for tea

growth: temperature between 20-29 degree Celsius, high humidity and rainfall conditions of 150-290 cm per year, on an average.

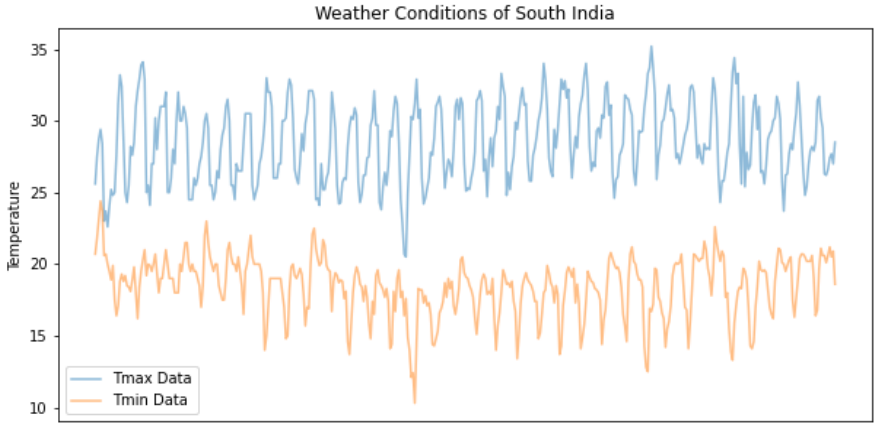

Fig 8. Temperature min and max in South India

India is one of the largest producers of tea in the world with 3/4th of its produce being consumed locally. Nilgiris in South India account for a large percentage of tea production in the country.

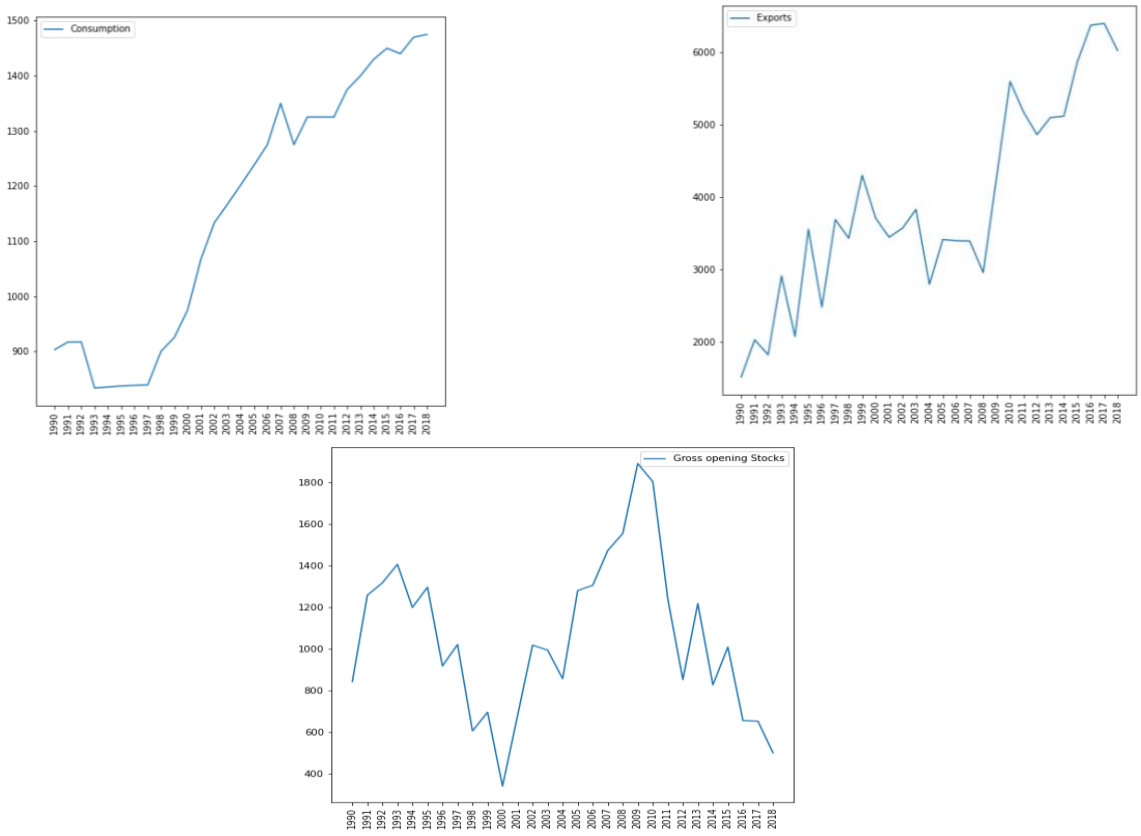

Fig 9. Consumption, Export and Gross Opening stocks of Tea and Coffee from 1990 to 2018 in India.

Due to the high contribution of South India in non-alcoholic beverage production of India, we can holistically assume that the consumption has increased, exports decreased recently, and gross opening stocks have severely decreased for the India tea and coffee market.

Due to the closeness of Sri Lanka to the southern regions of India, a further analysis of tea and coffee production is also done for the same.

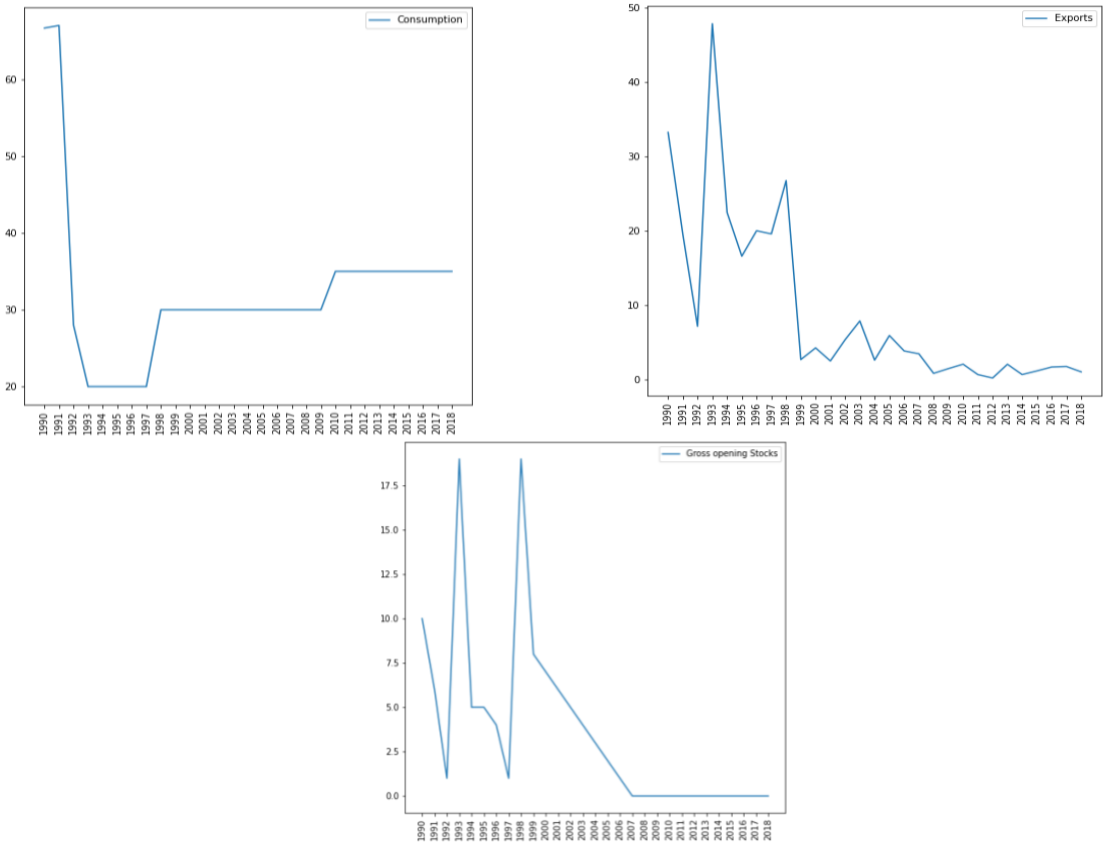

Fig 10. Consumption, Exports and Gross Opening stocks of Coffee and Tea in South India

Sri Lanka has a comparatively different scenario of non-alcoholic beverage market base, and this is based on the farmer behavior and country's economic dependencies. Finally, visualizing the prices paid to the Indian farmers for their tea/coffee productions shows that it has decreased drastically over time, which provides a justified explanation for the decrease in tea/coffee production over the past years.

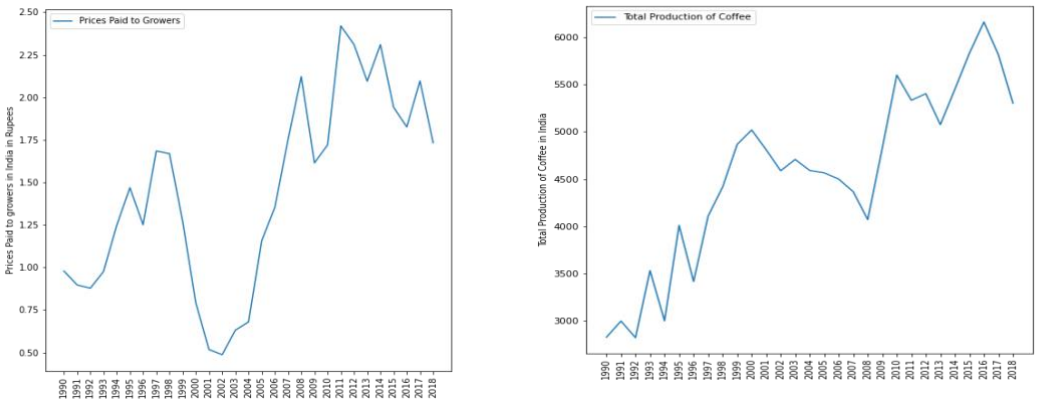

Fig 11. Reduction in Prices paid to Customers causing reduction in coffee/tea production

K-Means Clustering

K-Means Clustering is an unsupervised learning algorithm that is used to solve the clustering problems in machine learning or data science. It allows us to cluster the data into different groups and a convenient way to discover the categories of groups in the unlabeled dataset on its own without the need for any training. It is a centroid-based algorithm, where each cluster is associated with a centroid. The main aim of this algorithm is to minimize the sum of distances between the data point and their corresponding clusters. The algorithm takes the unlabeled dataset as input, divides the dataset into k-number of clusters, and repeats the process until it does not find the best clusters. The value of k should be predetermined in this algorithm.

We start by pre-processing the data and cleaning it. This essentially involves null-handling and label encoding the ordinal parameters of the data. The data is then passed into the Scikit-Learn K-Means Clustering model to obtain the elbow curve for the ideal number of clusters. Using the "elbow" or "knee of a curve" as a cutoff point is a common heuristic in mathematical optimization to choose a point where diminishing returns are no longer worth the additional cost. In clustering, this means one should choose a few clusters so that adding another cluster doesn't give much better modeling of the data. The intuition is that increasing the number of clusters will naturally improve the fit (explain more of the variation), since there are more parameters (more clusters) to use, but that at some point this is over-fitting, and the elbow reflects this.

```
clusters = range(3,31)
inertia = []
for n in clusters:
    kmeans = KMeans(n_clusters=n)
    kmeans.fit(X)
    inertia.append(kmeans.inertia_)

fig, ax = plt.subplots(figsize=(10,8))
ax.plot(clusters, inertia);
plt.show()
```

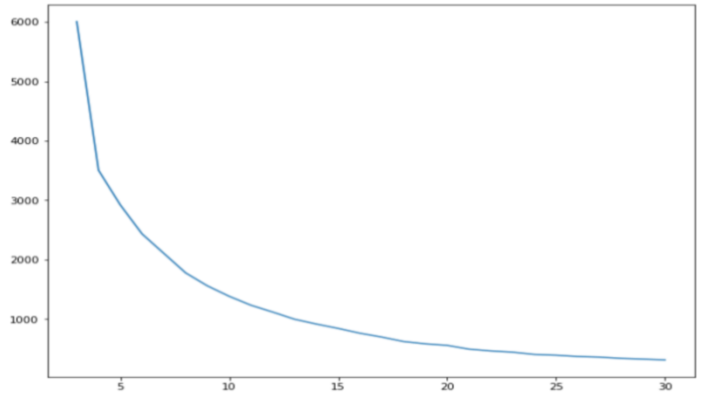

Fig 12. Elbow curve showing a cluster of 10 at most to be the ideal number of clusters

Based on the elbow curve, we assume the number of clusters to be optimally around 10. We further obtain a silhouette score for our data model. Silhouette analysis can be used to study the separation distance between the resulting clusters. The silhouette plot displays a measure of how close each point in one cluster is to points in the neighboring clusters and thus provides a way to assess parameters like number of clusters visually. This measure has a range of [-1, 1]. Silhouette coefficients (as these values are referred to as) near +1 indicate that the sample is far away from the neighboring clusters. A value of 0 indicates that the sample is on or very close to the decision boundary between two neighboring clusters and negative values indicate that those samples might have been assigned to the wrong cluster. In our case, the silhouette score seems to be rising drastically, beyond the value of 10 and is optimally low within the range of 7.5 and 10.

Fig 13. Silhouette score for different values of clusters up to 20

The cluster validity index (CVI) is an indicator by which to provide a way of validating the quality of clustering algorithms and determine the correct number of clusters in datasets. Here too, we obtain a cluster validity index for our clustering algorithm.

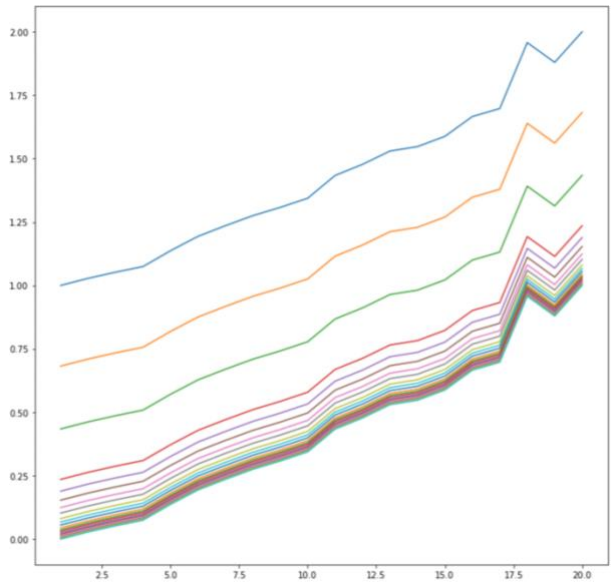

Fig 14. Validity index for various number of clusters from 0 to 20.

Thus, we finalize our clustering model to have 10 clusters where the clusters represent states of India with the similar kind of production and consumption parameters and then we specifically analyze the clusters with the south Indian states in them.

Inference and Supplementary material

Fig 15. Clusters obtained in K-Means

The Clusters are closely related to each other and we find that the clusters 1,2,4,5,6 and 9 have states 'CHENNAI', 'HYDERABAD','KARNAL','ERNAKULAM', 'PANAJI' and given the clusters aren't much distinguishable we can estimate the average price of tea in these areas to be approximately 12500 per kilo and these areas have similar market base for tea as in the Western Ghat and Eastern India port cities.

A wide range of variables and their impact is measured that affects consumer buying behaviour. Some of them are price, aroma, quality, packaging, celebrity endorsement, etc. Different statistical methods have been used by the researchers like PCA and factor analysis to achieve the desired results. Let us examine this paper and try to derive some useful insights for our problem statement.

The following tables are obtained from a questionnaire designed on Likert Scale based on 15 attributes. For sampling, out of 100 units 50 were from urban and 50 from rural areas. Convenience sampling was used for sampling the data.

Table 1 : Age of the respondents

Sr. No.	Age (in years)	Urban		Rural	
		Number	Per cent	Number	Per cent
1.	Up to 30 years	13	26.00	17	34.00
2.	31-40 years	19	38.00	14	28.00
3.	41-50 years	12	24.00	16	32.00
4.	Above 50 years	6	12.00	3	6.00
Total		50	100.00	50	100.00

Source:Internat. J. Com. & Bus. Manage., 6(2) Oct., 2013 : 199-205 HIND INSTITUTE OF COMMERCE AND BUSINESS MANAGEMENT

Age is a crucial factor that affects buying behaviour for any foodstuffs. From table 1 we come to know that about 38% of urban respondents were of the age group 31-40 years. While 34% of the age group were less than 30 years

and from rural areas. In table 2 we observe most respondents who completed graduation were from urban areas. Education influences consumer purchasing behavior. From table 3 we see that there are differences in the income level of urban and rural families. This could be the result of the difference in the type of occupation of the family members.

Table 2 : Educational status of the respondents

Sr. No.	Education	Urban		Rural	
		Number	Per cent	Number	Per cent
1.	Illiterate	0	0.00	8	16.00
2.	Primary	7	14.00	17	34.00
3.	Secondary	14	28.00	11	22.00
4.	Higher Secondary	16	32.00	9	18.00
5.	Collegiate	13	26.00	5	10.00
Total		50	100.00	50	100.00

Table 3 : Annual family income of the respondents

Sr. No.	Annual Income (Rs.)	Urban		Rural	
		Number	Per cent	Number	Per cent
1.	Upto 50000	6	12.00	17	34.00
2.	50000 to1 lakh	17	34.00	21	42.00
3.	Above 1 lakh	27	54.00	12	24.00
Total		50	100.00	50	100.00

Table 4 : Average monthly expenditure on tea

Sr. No.	Monthly expenditure on tea	Urban		Rural	
		Number	Per cent	Number	Per cent
1.	<200	13	26.00	35	70.00
2.	201-400	24	48.00	12	24.00
3.	>400	13	26.00	3	6.00
	Total	50	100.00	50	100.00

Source: Internat . J. Com. & Bus. Manage., 6(2) Oct., 2013 : 199-205 HIND INSTITUTE OF COMMERCE AND BUSINESS MANAGEMENT

We can observe table 4 to conclude that 70% of respondents from rural areas spend less than 200 rupees per month. While in urban areas 48% of respondents spend 201-400 rupees per month.

Table 5 : Frequency of tea consumption

Sr. No.	Frequency (Per day)	Tea		Coffee	
		Urban	Rural	Urban	Rural
1.	Occasionally	7 (14.00)	4 (8.00)	27 (54.00)	31 (62.00)
2.	One time	11 (22.00)	7 (14.00)	17 (34.00)	15 (30.00)
3.	Two times	13 (26.00)	10 (22.00)	4 (8.00)	3 (6.00)
4.	Three times	9 (18.00)	14 (28.00)	2 (16.00)	1 (2.00)
5.	Four times	7 (14.00)	9 (18.00)	0 (0.00)	0 (0.00)
6.	Five times	3 (6.00)	4 (8.00)	0 (0.00)	0 (0.00)
7.	More than five times	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
Total		50	100.00	50	100.00

Table 6 : Frequency of buying tea

Sr. No.	Frequency	Urban		Rural	
		Number	Per cent	Number	Per cent
1.	Weekly once	5	10.0	17	34.00
2.	Fortnightly	14	28.00	13	26.00
3.	Once in a month	18	36.00	11	22.00
4.	Once in two months	9	18.00	8	16.00
5.	When required	4	8.00	1	2.00
Total		50	100.00	50	100.00

Source:Internat. J. Com. & Bus. Manage., 6(2) Oct., 2013: 199-205 HIND INSTITUTE OF COMMERCE AND BUSINESS MANAGEMENT

Tea and coffee are substitutes from generations. In table 5 we see that both rural and urban respondents consume tea twice and thrice per day. Coffee has a lesser frequency of consumption twice a day in both rural and urban areas (8% in urban and 6% in rural). From table 6 we can observe that urban respondents bought tea once in a month (36%) followed by fortnightly (28%). But when we analyze the rural population, we can see that 34% of the respondents bought tea weekly once and 26% fortnightly. This behavior may correspond to differences in wage type income of the respondents in rural and urban areas. Respondents in rural areas tend to buy tea more frequently, this could also be because of a bigger family size in such areas.

Table 8 : Component matrix

Items	Components						
	1	2	3	4	5	6	7
Quality	-0.215	-0.580	-0.111	0.193	-0.134	0.156	0.252
Aroma	-0.290	0.031	0.207	-0.287	0.671	0.126	0.335
Flavour	-0.323	-0.649	-0.098	0.000	-0.371	0.337	0.270
Taste	-0.056	-0.592	0.007	0.524	0.265	0.032	-0.282
Colour of end product	-0.040	0.147	-0.524	0.018	-0.122	-0.657	0.000
Brand image	-0.048	0.627	-0.012	0.262	-0.205	0.406	-0.094
Value added tea	-0.374	-0.120	0.706	-0.028	0.260	-0.007	-0.118
Retailers influence	-0.219	0.468	0.445	0.291	0.021	-0.249	-0.047
Influenced by others	0.167	0.108	0.382	0.318	-0.264	-0.060	-0.048
Reasonable price	0.923	-0.154	0.222	-0.132	-0.011	0.068	0.036
Timely availability	-0.010	0.674	-0.111	-0.168	-0.050	0.420	0.026
Attractive packing	-0.010	0.113	0.392	-0.079	-0.262	-0.390	0.627
Effective advertisement	0.282	0.197	-0.326	0.312	0.647	-0.096	0.090
Celebrity endorsement	0.923	-0.154	0.222	-0.132	-0.011	0.068	0.036
Gift/Promotion/Strategies	0.158	0.123	-0.221	0.349	0.169	0.177	0.547

Extraction method: Principal component analysis.
a. 7 components extracted

Source:Internat. J. Com. & Bus. Manage., 6(2) Oct. 2013: 199-205 HIND INSTITUTE OF COMMERCE AND BUSINESS MANAGEMENT

It can be concluded from table 8 that gift/promotion/ strategies, retailer's influence, and influenced by others are the three attributes, which were removed in the case of people who consumed tea from urban areas. Variables such as quality, brand image, value added tea, aroma, timely availability, attractive packing, flavor, taste, color of product, effective advertising, reasonable price and celebrity endorsement were retained in urban consumer.

Table 9 : Total variance explained

Components	Initial eigenvalues			Extraction sums of squared loadings		
	Total	% of variance	Cumulative (%)	Total	% of variance	Cumulative (%)
1.	2.037	13.577	13.577	2.037	13.577	13.577
2.	1.858	12.387	25.964	1.858	12.387	25.964
3.	1.688	11.256	37.220	1.688	11.256	37.220
4.	1.328	8.851	46.071	1.328	8.851	46.071
5.	1.298	8.655	54.726	1.298	8.655	54.726
6.	1.109	7.391	62.117	1.109	7.391	62.117
7.	0.964	6.425	68.542			
8.	0.806	5.373	73.915			
9.	0.794	5.296	79.211			
10.	0.692	4.611	83.822			
11.	0.635	4.236	88.058			
12.	0.575	3.831	91.889			
13.	0.504	3.359	95.248			
14.	0.379	2.529	97.777			
15.	0.333	2.223	100.000			

Source:Internat. J. Com. & Bus. Manage., 6(2) Oct., 2013: 199-205 HIND INSTITUTE OF COMMERCE AND BUSINESS MANAGEMENT

From table 9 we can interpret those 15 variables were diminished to 6 factors which contributed for 62.117% of the total cumulative variance.

Table 10 : Component matrix

Attributes	Components					
	1	2	3	4	5	6
Quality	-0.269	0.061	-0.363	0.597	0.133	0.463
Aroma	0.627	0.367	-0.249	-0.007	0.166	-0.097
Flavour	0.381	0.659	0.296	0.137	0.072	-0.088
Taste	0.134	0.594	0.350	0.056	0.233	0.034
Colour of end product	0.190	-0.198	-0.570	0.390	-0.377	0.080
Brand image	-0.067	0.499	-0.054	-0.300	-0.173	0.002
Value added tea	-0.383	0.205	0.250	0.482	0.246	-0.206
Retailers influence	0.094	0.087	0.181	-0.420	0.329	0.691
Influenced by others	0.572	-0.244	-0.030	-0.172	-0.340	0.241
Reasonable price	-0.328	0.287	0.629	0.002	-0.389	0.002
Timely availability	0.283	-0.008	0.550	0.130	0.304	-0.031
Attractive packing	0.551	0.059	-0.200	-0.227	0.221	-0.329
Effective advertisement	-0.368	0.207	-0.470	-0.205	0.299	-0.260
Celebrity endorsement	0.582	-0.146	0.269	0.477	-0.198	-0.026
Gift/Promotion/Strategies	0.155	0.621	-0.101	0.210	0.132	0.338

Extraction Method: Principal Component Analysis.
a. 6 components extracted.

Source:Internat. J. Com. & Bus. Manage., 6(2) Oct. 2013: 199-205 HIND INSTITUTE OF COMMERCE AND BUSINESS MANAGEMENT

It can be concluded from table 8 that value added tea, gift/ promotion/strategies, effective advertisement, and brand image are the four attributes, which were removed in the case of people who consumed tea from rural areas. The attributes such as celebrity endorsement, flavor, color of end-product, quality, aroma, taste, retailers influence, influenced by others, attractive packing, reasonable price, and timely availability were retained in rural consumers.

Insights and Recommendation

The quality, aroma, flavor, and taste rather than the influence by retailer's influence, influenced by others and gift/ promotion/strategies influenced urban people more while buying tea. People in urban areas are certainly more aware of the tastes and quality factors than people in rural areas. The loose tea dust rather than the branded ones was chosen by people in rural areas. One of the main reasons for low factor loading in the case of effective advertisement, value added tea, gift/promotion/strategies and brand image could be this. When opening a tea market in South India, manufacturers of tea brand must consider giving more importance to the quality, flavor, aroma of the beverage. More factor loadings in both the urban and rural areas was obtained for these four variables.

People must be educated and made aware about the health benefits of green tea. Choosing, the right price for the tea powder should be given utter importance in both rural and urban areas. It should be competitive but affordable. Eventually, developing tea bars to encourage out-of-home consumption can be a great idea.

Acknowledgment

We would like to thank the administration of FeyNN Labs for providing us with the opportunity of carrying out such a detailed analysis and providing us with the exposure to industry business analytics.

We would also like to thank our parents, friends and associates for their constant support, aid, and advice.

The project has required no form of funding and is absolute independent research by the Machine learning Interns of FeyNN [dot] AI.

References

- K.C. PRAKASH AND P. PARAMASIVAM, “Factors affecting buying behaviour of tea in Nilgiris district of Tamil Nadu” Volume 6 | Issue 2 | October, 2013 | 199-205 http://www.researchjournal.co.in/upload/assignments/6_199-205.pdf
- Upasi Tea Research Foundation http://www.upasitearesearch.org/history-of-tea-in-south-india-2/
- Global Coffee Production/Consumption Kaggle dataset https://www.kaggle.com/michaellight/global-coffee-production-consumption
- Kaggle customer Segmentation and Business Analytics https://www.kaggle.com/harihar365/customer-segmentation-and-business-analytics
- Modelling the impact of agrometeorological variables on regional tea yield variability in South Indian tea-growing regions: 1981-2015 Esack Edwin Raj et al. https://scholar.google.co.in/citations?view_op=view_citation&hl=en&user=C4TUPaMAAAAJ&citation_for_view=C4TUPaMAAAAJ:bz8QjSJIRt4C
- Esack, Edwin Raj (2019), “Meteorological and Yield Data of Tea at Monthly, seasonal and Annual scales by UPASI, South India.”, Mendeley Data, V1, doi: 10.17632/rcgvn92yxx.1
- Tea Board India http://www.teaboard.gov.in/WEEKLYPRICES/2021
- Coffee Vs Tea Sales https://www.kaggle.com/najibmh/coffee-vs-tea-sales/data
- Elbow Clustering https://en.wikipedia.org/wiki/Elbow_method_(clustering)
- K-Means Algo Javapoint https://www.javatpoint.com/k-means-clustering-algorithm-in-machine-learning
- Statista Total production of tea from 2014-2020 in India https://www.statista.com/statistics/652078/tea-production-in-south-india/
- Tea Producing Regions in India https://ratetea.com/region/india/2/