Optimizing Ink Tank Printer Retail Cost and Placement Locations in India

Senarto Salim

3 January 2021

Table of Contents

Introduction: Business Problem

Inkjet printers with ink tanks have been gaining popularity in emerging countries like India. This is due to the appeal of lifetime ink that comes with the printer, tackling the perennial issue of frequent purchases of costly ink supplies.

Epson is the market leader currently with HP catching up hard to gain market share. With limited go to market funds and retail costs, **how can HP effectively compete in different countries by being selective in the in-country choice of cities to promote their printers with targeted product placement?**

For this project data science will be used to answer the above question. This will be a valuable tool for HP to gain market share rapidly without eroding profits unnecessarily to achieve it.

India will be the chosen country as it is one of the main battlegrounds of the ink tank printer war.

Data

Google Trends with the following parameters is used to generate the choice of cities for analysis in India by the amount of search interest:

- Search term **'tank printer'**
- Time range will be from Jan 2016 to Jan 2021
- Due to limitations of **pytrends**, data will be imported for analysis using **Google Trends** generated csv file

GeoPy is used to generate latitude and longitude data for India and selected cities.

Foursquare API is used to generate top 10 most common places where printers can be displayed and purchased within a 30km radius from city centre.

Create dataframe from Google Trends csv file

Using the keywords "tank printer" the cities with search interest trend are listed and exported as a csv file.

pytrends unfortunately is unable to generate data by cities for India so it is not used in this analysis.

Below is a view of the top 5 rows of the data:

	City	trend_count
0	Kolkata	100
1	Ghaziabad	94
2	Noida	92
3	Mumbai	91
4	Gurgaon	84

Figure 1 Top 5 India Cities that Searched for "tank printer" through Google

Generate latitude and longitude data for the listed cities using GeoPy

Below is the table of cities with geolocation data generated by GeoPy:

	City	trend_count	gcode	lat	long
0	Kolkata	100	(Kolkata, Howrah, West Bengal, India, (22.5414...	22.541418	88.357691
1	Ghaziabad	94	(Ghaziabad, Uttar Pradesh, India, (28.711241, ...	28.711241	77.444537
2	Noida	92	(Noida, Dadri, Gautam Buddha Nagar, Uttar Prad...	28.535633	77.391073
3	Mumbai	91	(Mumbai, Mumbai Suburban, Maharashtra, India, ...	19.075990	72.877393
4	Gurgaon	84	(Gurgaon, Gurugram, Haryana, India, (28.428262...	28.428262	77.002700
5	New Delhi	62	(New Delhi, Delhi, India, (28.6138954, 77.2090...	28.613895	77.209006
6	Coimbatore	57	(Coimbatore, Coimbatore North, Coimbatore Dist...	11.001812	76.962842
7	Bengaluru	53	(Bengaluru, Bangalore North, Bangalore Urban, ...	12.979120	77.591300
8	Kochi	47	(Kochi, Ernakulam district, Kerala, 682005, In...	9.931370	76.267376
9	Chennai	46	(Chennai, Chennai District, Tamil Nadu, India,...	13.083694	80.270186
10	Guwahati	38	(Guwahati, Kamrup Metropolitan, 781015, India,...	26.180598	91.753943
11	Hyderabad	32	(Hyderabad, Telangana, India, (17.38878595, 78...	17.388786	78.461065
12	Lucknow	29	(Lucknow, Sadar, Lucknow, Uttar Pradesh, 22601...	26.838100	80.934600
13	Pune	25	(Pune City, Pune District, Maharashtra, India,...	18.521428	73.854454
14	Ahmedabad	24	(Ahmedabad, Ahmadabad City Taluka, Ahmedabad D...	23.021624	72.579707
15	Raipura	19	(Raipura, Raipura Tahsil, Panna, Madhya Prades...	23.900227	79.964884
16	Indore	16	(Indore, Madhya Pradesh, 452001, India, (22.72...	22.720362	75.868200

Figure 2 India Cities that Searched for "tank printer" through Google with Geopy Geolocation Data

Use Folium to visualize the chosen India cities with the search trend ranking shown in pop up

Below is a visualisation of the data overlaid over India map and each city shows the search count of the city expressed in percentage of total, generated by using Folium:

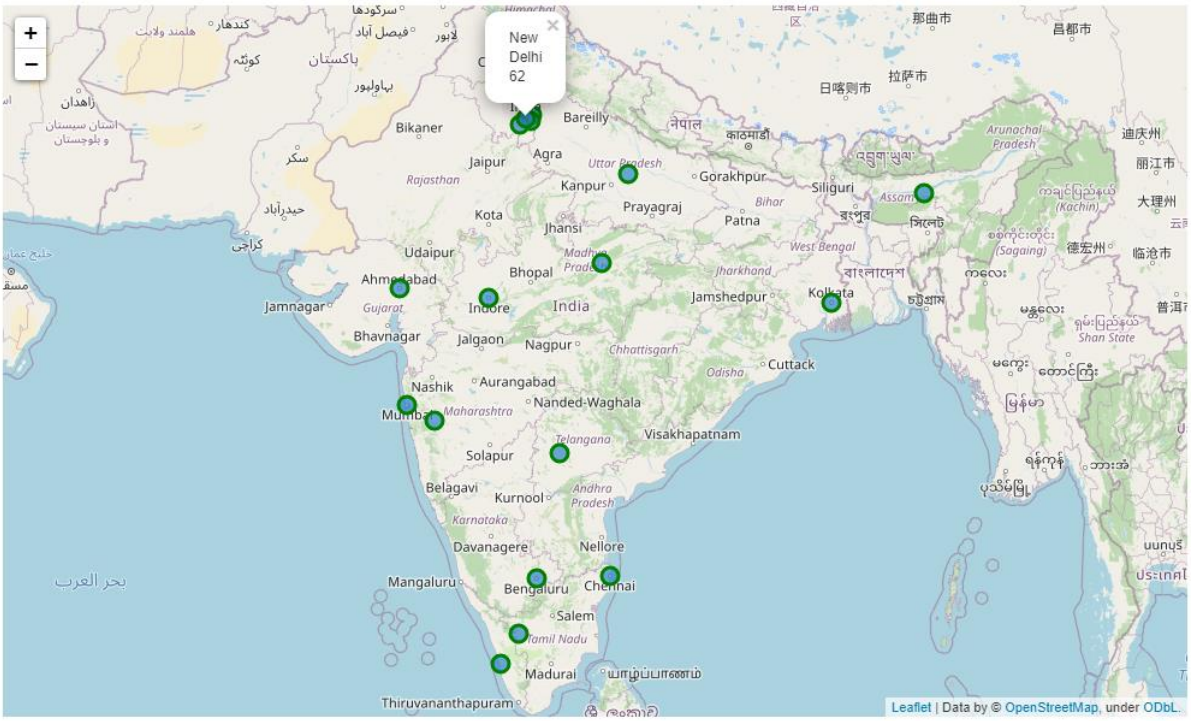

Figure 3 Folium Data Visualisation with Each City Popup Showing the Google Search Count in Percentage

Methodology

- Through **Foursquare API** generate for each city, venues where printers can be displayed and purchased within a 30km radius from city centre.
- Group the venues by cities.
- Conduct hot encoding so each integer value is represented as a binary vector that is all zero values except the index of the integer, which is marked with a 1.
- Group the hot encoded rows by city and by taking the mean of the frequency of occurrence of each venue category, the higher the mean of a venue category means high number of venues around the city centre.
- Create a dataframe of 10 most common venues around each city centres.
- Cluster the cities by their 10 most common venues:
    - Find a suitable number for cluster using Elbow Method for distortion and inertia:
        - distortion: calculated as the average of the squared distances from the cluster centres of the respective clusters. Typically, the Euclidean distance metric is used.
        - inertia: sum of squared distances of samples to their closest cluster centre.
    - Cluster the cities
- Visualize the city clusters.
- From the city clusters, determine suitable venues to market and display the printers.

Analysis

Create a function to repeat the same process to all the cities in India and generate venues where printers can be displayed and purchased within a 30km radius from city centre.

Generate common places where printers can be displayed and purchased in a 30km radius around the city centre using Foursquare

The following Foursquare categories are chosen to narrow down the data generated:

- Shopping Mall, 4bf58dd8d48988d1fd941735
- Shopping Plaza, 5744ccdfe4b0c0459246b4dc
- Outlet Mall, 5744ccdfe4b0c0459246b4df
- Outlet Store, 52f2ab2ebcbc57f1066b8b35
- Paper/Office Supplies Store, 4bf58dd8d48988d121951735
- Electronics Store, 4bf58dd8d48988d122951735
- Department Store, 4bf58dd8d48988d1f6941735
- Bookstore, 4bf58dd8d48988d114951735

Below is a view of top five rows of the common venues within 30km radius around city centre generated through Foursquare:

	City	City Latitude	City Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Kolkata	22.541418	88.357691	Quest Mall	22.539068	88.365525	Shopping Mall
1	Kolkata	22.541418	88.357691	South City Mall	22.501758	88.361726	Shopping Mall
2	Kolkata	22.541418	88.357691	Acropolis Mall	22.514823	88.393235	Shopping Mall
3	Kolkata	22.541418	88.357691	City Centre	22.587921	88.408098	Shopping Mall
4	Kolkata	22.541418	88.357691	Mani Square Mall	22.577823	88.400591	Shopping Mall

Figure 4 A Sample of Common Venue Data 30km Around City Centre Generated Through Foursquare

Conduct one hot encoding on the above data and group rows by city and by taking the mean of the frequency of occurrence for each venue category

Conduct hot encoding so each integer value is represented as a binary vector that is all zero values except the index of the integer, which is marked with a 1, this is shown in the table below:

	City	Accessories Store	American Restaurant	Bakery	Big Box Store	Bookstore	Bowling Alley	Building	Bus Line	Business Center	Butcher	Café	Clothing Store	Coffee Shop	College Cafeteria	Convenience Store	C
0	Kolkata	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
1	Kolkata	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	Kolkata	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	Kolkata	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	Kolkata	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

Figure 5 Top 5 Rows of the Data After Hot Encoding

By taking the mean of the frequency of the occurrence and grouping to each venue category in the above table, each venue category can be ranked as illustrated in below table for some cities:

```
----Ahmedabad----
                venue  freq
0     Shopping Mall  0.38
1  Electronics Store  0.19
2   Department Store  0.19
3          Bookstore  0.08
4     Clothing Store  0.02


----Bengaluru----
                venue  freq
0     Shopping Mall  0.34
1   Department Store  0.27
2  Electronics Store  0.17
3          Bookstore  0.08
4     Clothing Store  0.03


----Chennai----
                venue  freq
0   Department Store  0.39
1     Shopping Mall  0.27
2  Electronics Store  0.10
3          Bookstore  0.05
4     Clothing Store  0.03
```

Figure 6 A Sample of Top 5 Venue Mean Frequency for Each City

Display the top 10 venues for each city

A dataframe is then generated containing the top 10 venues for each city as shown in the table below:

	City	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Ahmedabad	Shopping Mall	Department Store	Electronics Store	Bookstore	Miscellaneous Shop	Building	Clothing Store	Cosmetics Shop	Paper / Office Supplies Store	Coffee Shop
1	Bengaluru	Shopping Mall	Department Store	Electronics Store	Bookstore	Clothing Store	Café	Hotel	Paper / Office Supplies Store	Butcher	Market
2	Chennai	Department Store	Shopping Mall	Electronics Store	Bookstore	Clothing Store	Flea Market	Multiplex	Café	Market	Men's Store
3	Coimbatore	Department Store	Shopping Mall	Electronics Store	Outlet Store	Shopping Plaza	Bookstore	Clothing Store	Food & Drink Shop	Flea Market	Factory
4	Ghaziabad	Shopping Mall	Department Store	Market	Electronics Store	Miscellaneous Shop	Plaza	Clothing Store	Convenience Store	Paper / Office Supplies Store	Coffee Shop

Figure 7 Dataframe Containing Top 10 Venues for Each City

Using k-means Clustering to Generate City Clusters by Common Venues

Cluster the cities by their 10 most common venues:

- Find a suitable number for cluster using Elbow Method for distortion and inertia:

- distortion: calculated as the average of the squared distances from the cluster centres of the respective clusters. Typically, the Euclidean distance metric is used.

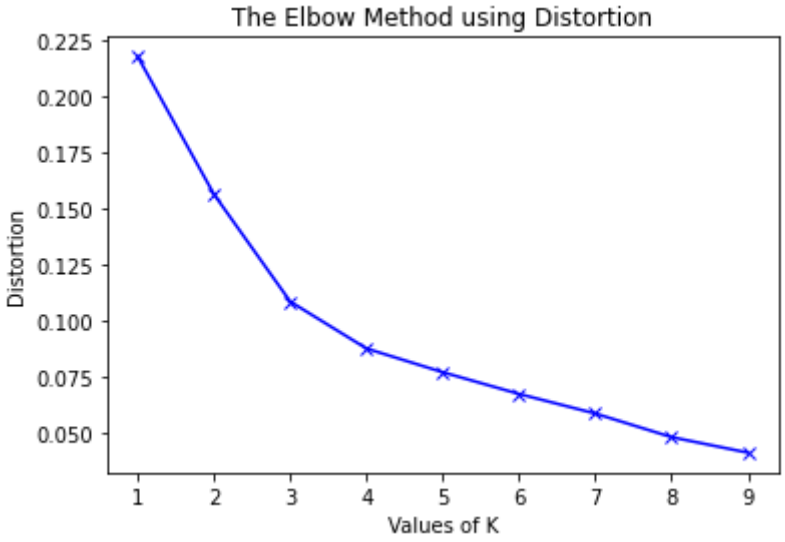

Figure 8 Chart of The Elbow Method Using Distortion

- inertia: sum of squared distances of samples to their closest cluster centre.

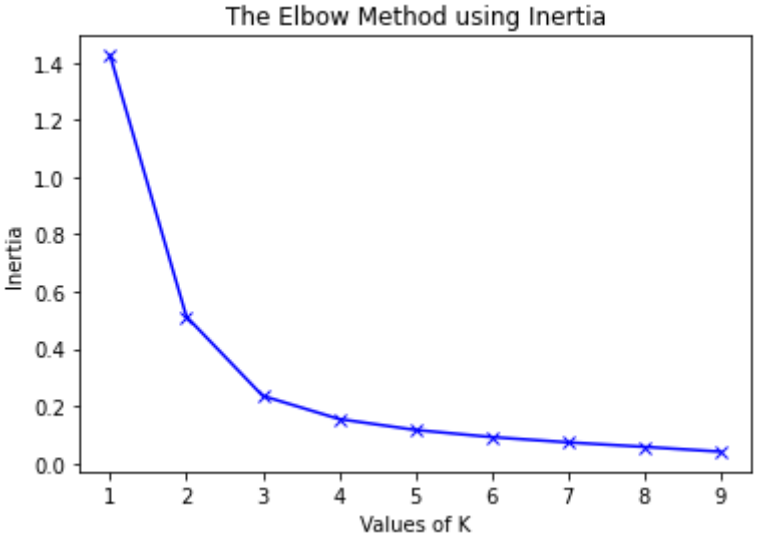

Figure 9 Chart of The Elbow Method Using Inertia

- From both Elbow Methods, it is determined the number of clusters should be 3.

- Cluster the cities into 3 clusters as visualised below:

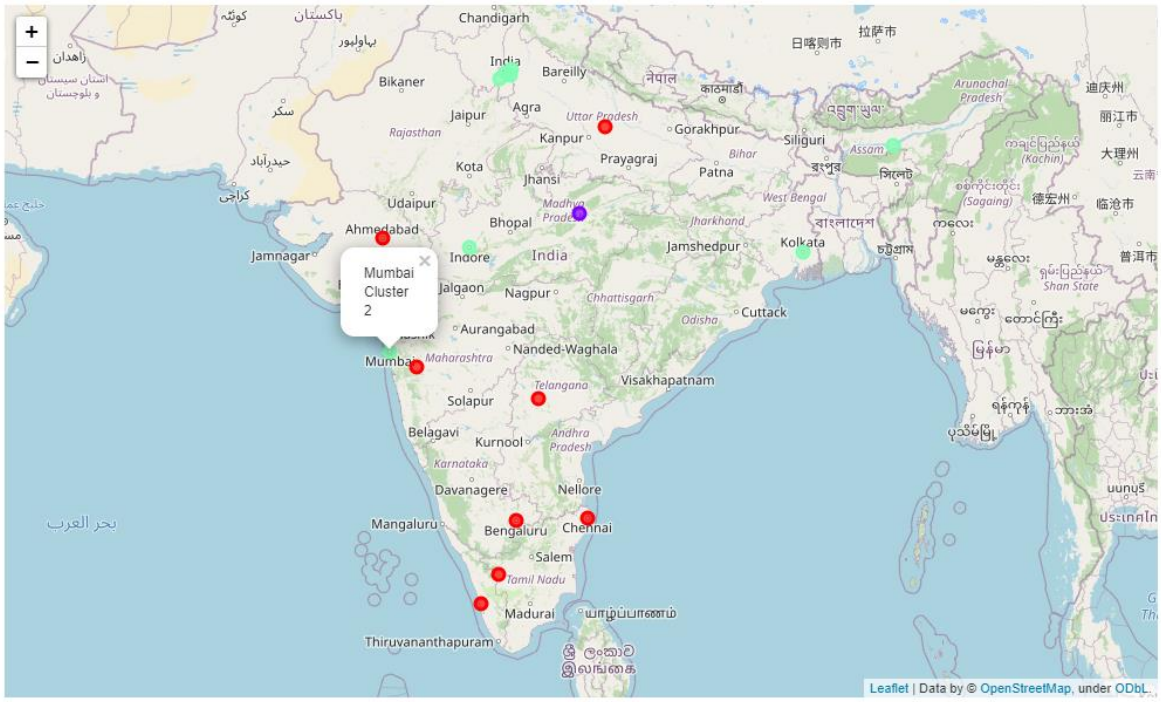

Figure 10 Folium Visualisation of the Cities Clustered by Common Venues, e.g. Mumbai is Cluster 2

The clusters are shown as per below:

- Cluster 0

	City	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
6	Coimbatore	0	Department Store	Shopping Mall	Electronics Store	Outlet Store	Shopping Plaza	Bookstore	Clothing Store	Food & Drink Shop	Flea Market	Factory
7	Bengaluru	0	Shopping Mall	Department Store	Electronics Store	Bookstore	Clothing Store	Café	Hotel	Paper / Office Supplies Store	Butcher	Market
8	Kochi	0	Electronics Store	Shopping Mall	Department Store	Bookstore	Road	Bus Line	Paper / Office Supplies Store	Convenience Store	Café	Clothing Store
9	Chennai	0	Department Store	Shopping Mall	Electronics Store	Bookstore	Clothing Store	Flea Market	Multiplex	Café	Market	Men's Store
11	Hyderabad	0	Shopping Mall	Department Store	Electronics Store	Bookstore	Clothing Store	Coffee Shop	Furniture / Home Store	Miscellaneous Shop	Mobile Phone Shop	Neighborhood
12	Lucknow	0	Shopping Mall	Electronics Store	Department Store	Neighborhood	Women's Store	Coffee Shop	Food Court	Food & Drink Shop	Flea Market	Factory
13	Pune	0	Shopping Mall	Department Store	Electronics Store	Bookstore	Paper / Office Supplies Store	Clothing Store	Other Nightlife	Factory	Coffee Shop	Outlet Mall
14	Ahmedabad	0	Shopping Mall	Department Store	Electronics Store	Bookstore	Miscellaneous Shop	Building	Clothing Store	Cosmetics Shop	Paper / Office Supplies Store	Coffee Shop

Figure 11 Dataframe of Cluster 0

- Cluster 1

	City	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
15	Raipura	1	Electronics Store	Women's Store	Coffee Shop	Furniture / Home Store	Food Court	Food & Drink Shop	Flea Market	Factory	Department Store	Cosmetics Shop

Figure 12 Dataframe of Cluster 0

- Cluster 2

	City	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Kolkata	2	Shopping Mall	Department Store	Electronics Store	Bookstore	Market	Residential Building (Apartment / Condo)	Paper / Office Supplies Store	Women's Store	Coffee Shop	Food & Drink Shop
1	Ghaziabad	2	Shopping Mall	Department Store	Market	Electronics Store	Miscellaneous Shop	Plaza	Clothing Store	Convenience Store	Paper / Office Supplies Store	Coffee Shop
2	Noida	2	Shopping Mall	Electronics Store	Department Store	Market	Bookstore	Restaurant	Clothing Store	Paper / Office Supplies Store	Nightlife Spot	Toy / Game Store
3	Mumbai	2	Shopping Mall	Electronics Store	Department Store	Paper / Office Supplies Store	Bookstore	Clothing Store	Grocery Store	Sporting Goods Shop	Miscellaneous Shop	Coffee Shop
4	Gurgaon	2	Shopping Mall	Department Store	Electronics Store	Market	Bookstore	Clothing Store	Café	Restaurant	Plaza	Music Store
5	New Delhi	2	Shopping Mall	Department Store	Electronics Store	Bookstore	Market	Clothing Store	Restaurant	Business Center	Paper / Office Supplies Store	Nightlife Spot
10	Guwahati	2	Shopping Mall	Clothing Store	Electronics Store	Department Store	Women's Store	Food Court	Food & Drink Shop	Flea Market	Factory	Cosmetics Shop
16	Indore	2	Shopping Mall	Electronics Store	Department Store	Grocery Store	Food Court	Restaurant	Office	Coffee Shop	Food & Drink Shop	Flea Market

Figure 13 Dataframe of Cluster 2

Results and Discussion

K-means clustering despite being unsupervised machine learning algorithms works well in this case study. Grouping similar data points together and discovering underlying patterns. Elbow Method provides an effective way to optimize the number of clusters.

Cluster 1 has only the city of Raipura (should be Raipur). The first most common venue is electronics store and it seems to be the only venue suitable for printer placement. For this city, HP will need to have a local channel partner that has large connections and frequent dealings with local electronics stores. This will ease printer demonstration and training of store owners.

Cluster 2 has a very consistent first common venue which is shopping mall. For these cities, HP might want to setup a storefront in the malls as it will command presence.HP will need trained regular employees to man the stores with knowledge of printer specifications, performance and operations.

Cluster 0 has a mixed first, second and third common venues, which consists of shopping malls, department stores and electronic stores. Due to the mix, there is opportunity to explore HP booths with promoters trained with knowledge of printer specifications, performance and operations. The booths can be targeted for department stores as there is regular traffic of shoppers and might capture browsers. The promoters should also distribute brochures to spread the word.

Conclusion

From the above results the below table shows the recommended printer placement and retail channels for the cities that have shown high Google queries on printers with ink tanks.

Cluster	Cities	Recommended Retail
Cluster 0	Coimbatre, Bengaluru, Kochi, Chennai, Hyderabad, Lucknow, Pune, Ahmedabad	HP booths at department stores
Cluster 1	Raipura (Raipur)	Electronics stores through HP channel partner
Cluster 2	Kolkata, Ghaziabad, Noida, Mumbai, Gurgaon, New Delhi, Guwahati, Indore	HP stores at malls

Figure 14 City Clusters and their Recommended Retail Locations

References

- Google Trends
- Foursquare API
- GeoPy
- Folium