

Starting Dabbawala Service at the city of Kolkata

1. INTRODUCTION

1.1 Background:

Dabbawala service is famous Tiffin delivery service of Mumbai, which caters to thousands of people in very low cost. Dabba means Tiffin in local language. It provides lunch boxes to people in their offices at a specified time. The meals are fresh and home cooked mostly by House wives, giving them immense opportunity to earn, working at home. This ensures quality food made with all possible care. Apart from the cooks/house wives, the next most important part of this service are the Dabbawalas or the delivery personals. The Dabbawalas are the one who collect the Dabbas/Tiffin from the household and deliver it to the clients directly to their Office spaces at Lunch time. They almost always commute through train as it is the cheapest mode of communication in India. Where one Dabbawala collect Tiffin prepared by house wives in one location and deliver it to another fellow Dabbawala at the nearest station at a specific time, who then commute through the train and at each specific location (as indicated by color coded stickers on the Dabbas), the Dabbawalas in the train would pass the Tiffin to other dabbawala in the station, who then delivers it to the destination by walking or by bicycle. This system of Delivery is unique and very much cost efficient, making the service very cheap and yet very high quality.



IMAGE 1: Image Credits: marathimedia.com

This service had been a topic of many research papers of top most Management Schools and even Harvard Business School added the case study *The Dabbawala System: On-Time Delivery, Every Time* to their compendium for its high level of service with a low-cost and simple operating system. This is a service which give employment to semi skilled workers and produce employment opportunity to women who cannot leave their house to work outside due to myriads of family responsibility. Therefore the social significance of Dabbawala service is huge.

1.2 Business Problem:

Despite the popularity of the service it is not replicated in other cities like Kolkata. There is huge potential of growth of the service in the city of Kolkata as Kolkata also have very good local train connectivity, a huge demand for food service and many semi skilled persons looking for employment. The Dabba service is known for very little investment required to kick start.

Now if someone wants to start a Dabba service in the city of Kolkata with minimum investment, where should he/she start with? Which are the location where the advertisement for the service be targeted? From which area should the cooks/house wives be recruited?

1.3 Audience:

- Individuals looking for a social entrepreneurship opportunity with minimum investment and good social benefits.
- House holders looking for opportunities to earn money cooking same food that they already cook for their own family.
- Semi skilled people looking for job in cities can earn decent money through delivery service.
- Government and NGOs can also invest in such a service as it provides immense opportunity for women to be included in work force of the country and thus giving a boost to the country's GDP.

2. DATA

2.1 Data Source:

- Wikipedia Page on list of Kolkata suburban railway stations
- FourSquare Location data.

2.2 How Data will be used:

- List of Local Train Stations in Kolkata is required as the Dabbawala service heavily rely on commutation through Train. Kolkata Suburban railway Station list is extracted from [WikipediaPage](https://en.wikipedia.org/wiki/List_of_Kolkata_Suburban_Railway_stations) (https://en.wikipedia.org/wiki/List_of_Kolkata_Suburban_Railway_stations) through web scrapping using Pandas bs4 flavor. We fetch the table in the Wikipedia page that consist of all train stations name along with their line.
- The coordinates (Latitude & Longitude) of each Station to be fetched using **Geopy API** of Python.
- Using **FourSquare API** all venues around these Stations would be fetched. The 'Explore' feature of the FourSquare API is used here. The request is made for each Station location.

The result is a JSON file for each request, which consist of top 100 venues within 1.5Km of each requested locations along with their Name, Category, Latitude and Longitude.

For example if the request is sent to FourSquare API for Ballygunj Station location, all the venues like Restaurant, Park etc within a specified radius(say 1.5KM) of the Ballygunj station would be returned. After all Venues around all the stations is retrieved we will filter out stations around which the Dabbawala service would probably flourish. We would need to find areas which are *office places with high demand for food services*. Also we would need to find *residential areas from where cooks would be recruited*. The Venue Category would be used to find such areas. Closer these locations are to each other and closer they are to a stations, better is the business opportunity as the commutation expense as well as time of commutation would be minimum.

3. METHODOLOGY

3.1 Exploratory data analysis:

As mentioned, the service in order to be cheap is usually conducted through train stations, so we need the data on all the train station within the city on Kolkata. Web scrapping the Wikipedia page on the list of Kolkata suburban Train Station provided us with a table containing details 411 such stations. We only kept two columns, one of Station Name and another of Line name. We dropped rest all columns. We used geopy to get the coordinates (latitude and longitude) of these stations. No coordinates could be fetched for 62 stations, hence these rows were deleted as without latitude and longitudes these data won't be of much use.

In order to see if the station coordinates are within the city of Kolkata we plotted the Stations on the Map.

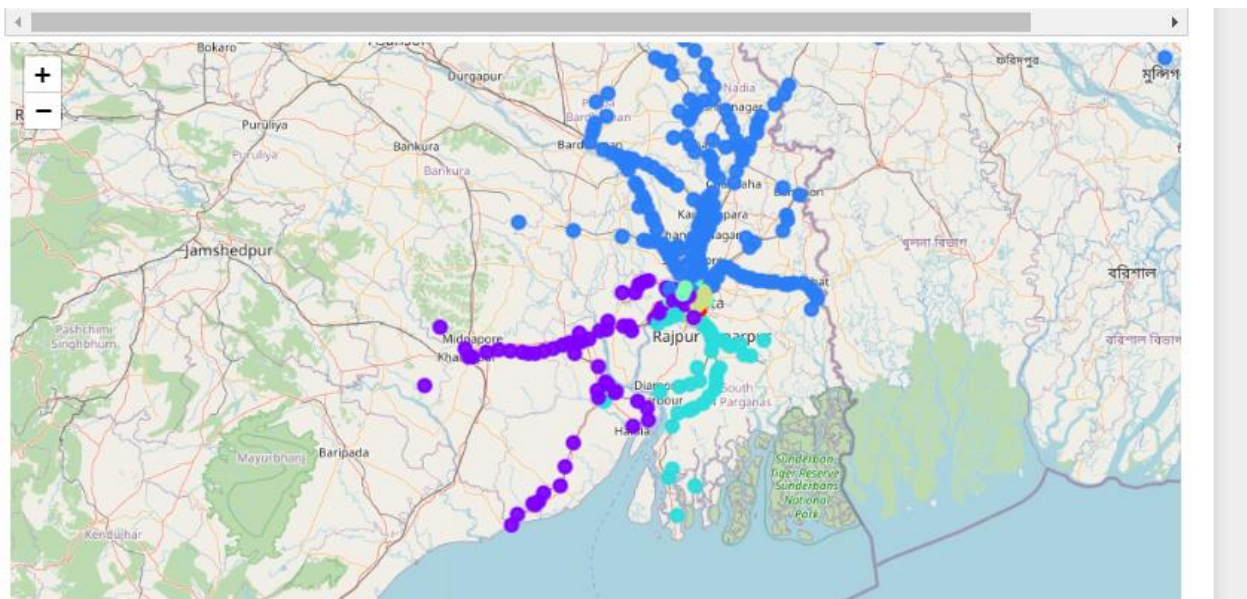


Image 2: The dots represents Train Stations and the color of the dot represent the Train Line.

As the plot reveals the Wiki page consist of all stations in the state of West Bengal rather than only the list of train stations inside Kolkata. Therefore we needed to find the stations which are within the vicinity of Kolkata. So we selected four Train Stations at the extreme boundaries in four directions of the Kolkata city from the map and then we selected all the stations that falls within that boundary in the following manner:

Boundary	Station at the extreme Boundary	Logic
South	Sonarpur Junction	Latitude should be \geq Latitude of Sonarpur Junction
North	New Barrackpore	Latitude should be \leq Latitude of New Barrackpore
West	Santragachi Junction	Longitude should be \geq Longitude of Santragachi Junction
East	Barasat Junction	Longitude should be \leq Longitude of Barasat Junction

The selection of the above stations is done in such a manner so as to include the whole stretch of the city of Kolkata as well as its outskirts. With this we are left with 65 stations, within the city of Kolkata and its outskirts.

Next step would be to find out all the venues around these stations so as to determine the areas where the demand of the Dabbawala service would be highest.

For this we would be using the 'explore' functionality of the foursquare API where we will send the latitude and longitude of the stations along with the radius of our search in the url and in return we get a JSON file for each station containing the list of top 100 venues within 1.5Km of each Station along with their Latitude, Longitude and Venue Category. We compile all the entries in one Dataframe (named Kolkata_Venues). Total of 882 venues are fetched, with no one area having 100 entries, which is the maximum limit set by us on the number of venues to be fetched per location. This signifies that the result is all encompassing i.e., the result contains all possible venues within 1.5Km radius from each station that are present in Foursquare dataset.

Assumptions:

Now in order to pin point on areas where the service need to be targeted, we will make some **assumption**:

1. We will assume that Dabbawala service demand would be substantial at areas with high concentration of Office places, where people would need homemade food during their lunch time. These areas are places where due to high demand for food service, concentration of restaurants is also very high. Generally office areas are highly concentrated in commercial places which would have high concentration of Banks, ATMs and other Financial services. Therefore our target location would be areas with high concentration of Offices, Restaurants, Banks, ATM, and Other Financial services.
2. Similarly we will assume that areas from where house wives could be recruited for cooking are residential areas. And residential area are places where concentration of

residential apartments, parks, departmental stores(grocery stores in India), sweet shops (specially in case of city of Kolkata) are high.

First we will see for locations to target for the service delivery. Therefore based on the above mentioned Assumption 1 we will filter out all venue category which have 'Office', 'Restaurant', 'Bank', 'ATM' and 'Finance' in it.

[Since Foursquare data set is still growing in city like Kolkata and as of now very few data on 'Office', 'Financial service' and 'ATM' is available in most areas, we will limit our search only to 'Restaurants' and 'Banks'.]

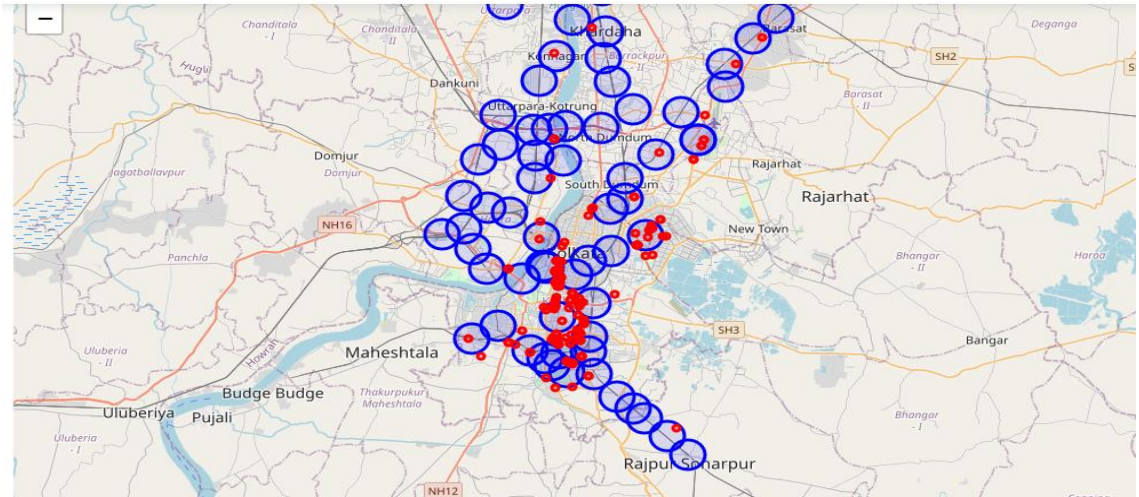


Image 3: In the map above the Blue circles are the Stations and their vicinity while the red dots are Restaurants and Banks.

As clear from the map, the concentration of these categories mentioned in assumption 1 are much higher near some stations as compared to others.

Following are the top 10 stations around which the demand for the service would probably be highest as these are the places where demand for food service is high as the concentration of restaurants is high and also these are the office area as the categories as mentioned in the assumption one are high here.

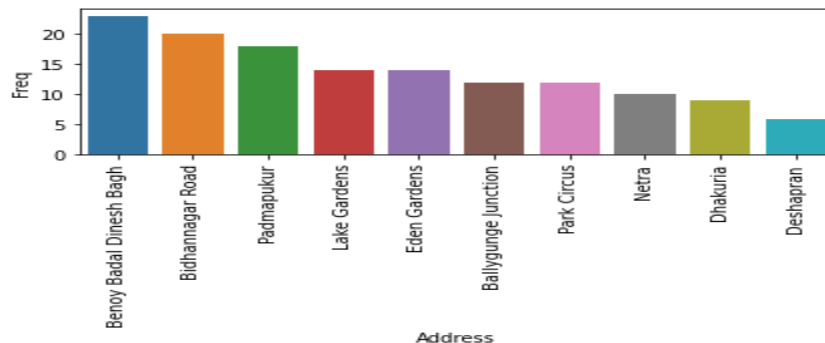


Image 4: The graph denotes top ten stations to target for Dabbawala service at X axis and sum total of Restaurants and Banks around each stations in Y axis.

Similarly we find areas where the categories as mentioned in Assumption 2 are high. These are the areas which are residential areas and cooks could be recruited from these areas. [Since Foursquare data set is still growing in area like city of kolkata and as of now there is a dearth of data on residential apartments, market, departmental stores and grocery stores in most areas, we will limit our search only to 'Parks' and 'sweet shops']. Following are the top ten stations around which venue categories indicating them to be a residential area are highest in frequency and from where cooks can be recruited.

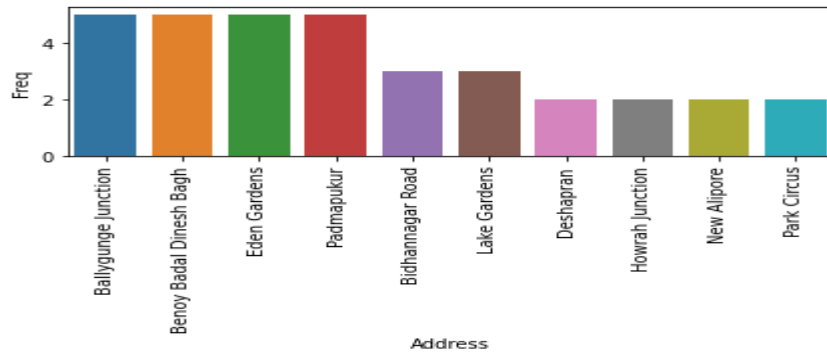


Image 5: The graph denotes top ten stations to target for Recruitment of Cooks at X axis and sum total of Parks and Sweet Shops around each stations in Y axis.

Therefore in accordance with the above two graphs (Image 4 and Image 5) the Station target for our business are as follow:

Station Name
Bidhannagar Road
Ballygunge Junction
Benoy Badal Dinesh Bagh
Eden Gardens
Lake Gardens
Park Circus
Barasat Junction
Padmapukur
Deshapran
Dhakuria
Netra

Table 1: Stations of importance

We see that the location of service target as well as that of residential area is either overlapping or very close to each other, which is very optimal for the business as it would further squeeze the commutation cost.

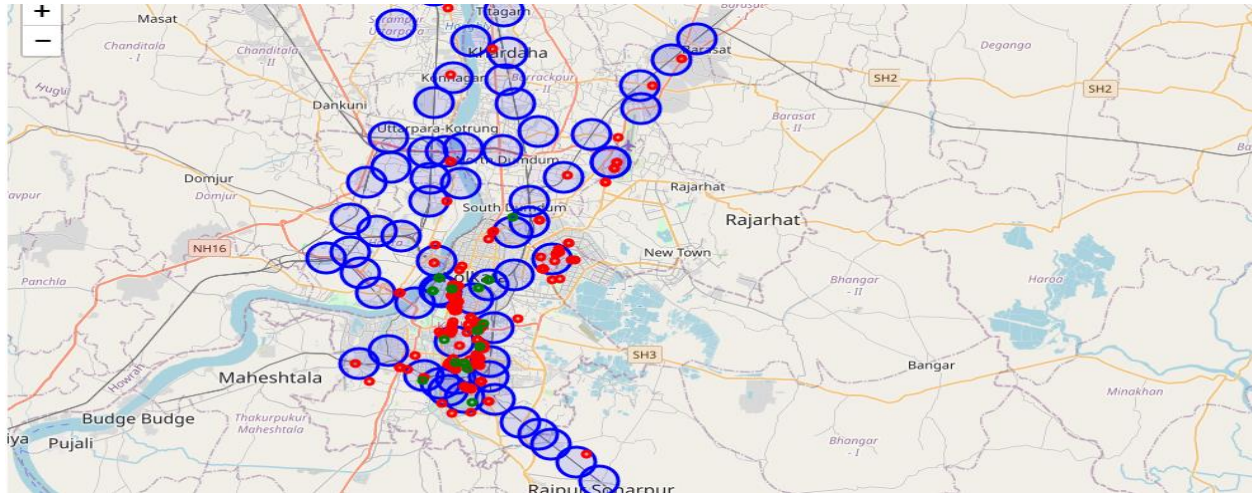


Image 6: The blue circle represents the stations and its vicinity; the red dots are venues which implies high demand for the Dabbawala service and the green dots represent categories which indicates that the areas are residential areas from where cooks could recruited.

3.2 Machine Learning:

Now in order to have a better understanding of different areas, understand the scope of future growth and better planning of the business we would group all the stations in three clusters based on the venues that are found within 1.5km of its radius. For this we will use KMean clustering algorithm.

For this we will use one hot encoding mechanism on the venue category field, thereby converting each category as a column, then merging the set with the station name. Once that's done we will use the KMean clustering algorithm to divide the stations in three clusters.

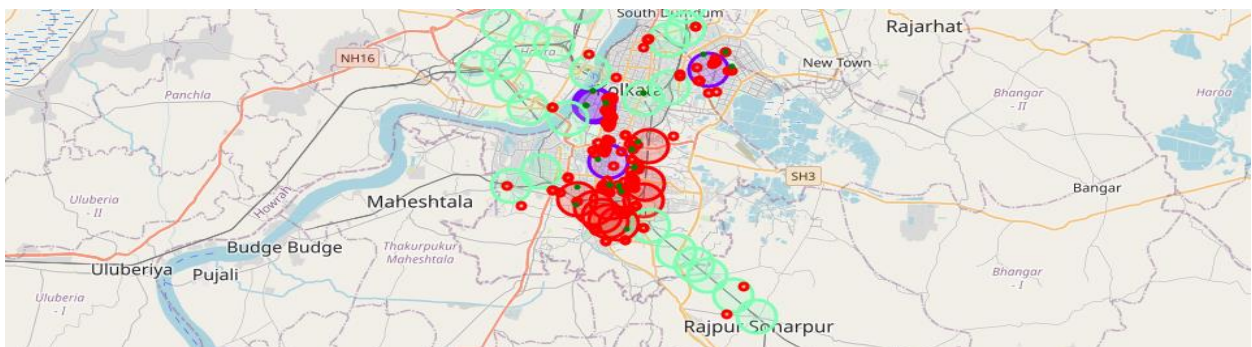


Image 7: The large circles denotes the stations and their color denotes different clusters; the red dots are venues which implies high demand for the service and the green dots represent residential areas from where cooks could recruited. [Cluster 0: Red circles, Cluster 1: Blue Circles, Cluster 2: Cyan Circles]

Clearly, Stations in Cluster 0 are in core of the city of Kolkata and Cluster 2 contains Stations mostly at outskirts of Kolkata.

4. RESULT

We can clearly see in the map above (Image 7) that the concentration of red dots and green dots (signifying the service target area and cook recruitment target areas) are high at Cluster 0 and Cluster 1. Also all of our Stations of Importance (Table1) falls in these two clusters. Thus the Dabba service can be targeted in these areas.

4.1 Discussion

4.1.1 Observation

In order to visually compare different clusters, we first find the number of each category of venues around each station (For example number of restaurants around Ballygunge Station) and then based on this frequency we find the top ten most frequent venue category for each station. This will make it easier to compare different clusters. Following are screen shots of entries from three clusters.

Cluster 0: Denoted by Red circles in above map (Image 7). If we see the image below, we can find that cluster zero has the areas with good demand for food as there are many Restaurants, cafes and food places. They also seem to have many clothing store, shopping malls and Movie Theaters. Also as could be seen in Image7 it has maximum numbers of red and green dots, making it very good locations for the business.

```
In [63]: kolkata_merged[kolkata_merged['Cluster Labels']==0]
```

	Area	Line	Latitude	Longitude	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
5	Ballygunge Junction	Circular Line	22.519162	88.372192	0	Café	Bengali Restaurant	Chinese Restaurant	Plaza	Vegetarian / Vegan Restaurant	Park	Boutique	Tea Room	Dhaba	Mughlai Restaurant
21	Dhakuria	South Lines	22.509408	88.371185	0	Café	Coffee Shop	Boutique	Plaza	Pizza Place	Chinese Restaurant	Clothing Store	Shopping Mall	Department Store	Sandwich Place
34	Lake Gardens	Circular Line	22.508055	88.353329	0	Café	Shopping Mall	Coffee Shop	Clothing Store	Chinese Restaurant	Movie Theater	Multiplex	Department Store	Indian Restaurant	Park
38	Netra	South Lines	22.496326	88.358423	0	Café	Clothing Store	Chinese Restaurant	Pizza Place	Dumpling Restaurant	Sandwich Place	Movie Theater	Mughlai Restaurant	Multiplex	Coffee Shop
39	New Alipore	Circular Line	22.509880	88.335954	0	Café	Asian Restaurant	Italian Restaurant	Plaza	Dhaba	Market	Pizza Place	Park	Ice Cream Shop	Frozen Yogurt Shop
Residential															

Image 8: Screenshot of Results in Cluster 0

Cluster 1: Denoted by Blue circles in the map above (Image 7). It also consists of areas with very high demand for food as it has mostly Restaurants, Hotels, Pubs and Nightclubs. As could be seen in Image7 these areas also have high concentration of red and green dots making it good for the business.


```

: kolkata_merged[kolkata_merged['Cluster Labels']==1]

```

Area	Line	Latitude	Longitude	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
Benoy Badal Dinesh Bagh	Circular Line	22.563288	88.345868	1	Café	Hotel	Indian Restaurant	Mughlai Restaurant	Pub	Pizza Place	Fast Food Restaurant	Nightclub	Restaurant	Riv
Bidhannagar Road	Chord link Line	22.584431	88.405524	1	Café	Fast Food Restaurant	Indian Restaurant	Department Store	Park	Bookstore	Bakery	Bengali Restaurant	Shopping Mall	Dumpli Restaura
Eden Gardens	Circular Line	22.564561	88.344242	1	Hotel	Café	Pub	Mughlai Restaurant	River	Pizza Place	Park	Nightclub	Indian Restaurant	Indi Swe Sh
Padmapukur	Eastern Line	22.531543	88.352457	1	Café	Bakery	Dhaba	Hotel	Indian Sweet Shop	Ice Cream Shop	Bengali Restaurant	Chinese Restaurant	Indian Restaurant	Asi Restaura

Image 9: Screenshot of Results in Cluster 1

Cluster 2: Denoted by Cyan circles in the map above. They seem to be in the outskirts of Kolkata city and do not have much data to contribute.

```

In [65]: kolkata_merged[kolkata_merged['Cluster Labels']==2]

```

Area	Line	Latitude	Longitude	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
Agarpara	Eastern Line	22.682878	88.385365	2	ATM	Auto Workshop	Bakery	Pharmacy	Train Station	NaN	NaN	NaN	NaN	NaN
Baghajatin	South Lines	22.480517	88.388760	2	Department Store	Multiplex	Bus Station	Shopping Mall	NaN	NaN	NaN	NaN	NaN	NaN
Bally	Eastern Line	22.642591	88.318419	2	ATM	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Bally Ghat	Chord link Line	22.652442	88.347904	2	ATM	Train Station	Business Service	Bus Station	Boat or Ferry	NaN	NaN	NaN	NaN	NaN
Bally Halt	Chord link Line	22.652341	88.338636	2	ATM	Train Station	Business Service	Bus Station	Electronics Store	NaN	NaN	NaN	NaN	NaN

Image 10: Screenshot of Results in Cluster 2

The Cluster 0 and Cluster 1 are good targets, but since the Dabbawala would commute through Train, it is better if they do not have to change multiple lines (i.e, change multiple Trains while commuting). In the initial stage of the business it is best that the stations we target are at same train line. Checking out the 'Line' for our Stations of Importance (Table 1), we find that Most Stations (5 out of 11) in our sorted list, falls in 'Circular Line'. See the Table below:

Station Name	Line	Cluster
Bidhannagar Road	Chord link Line	1
Ballygunge Junction	Circular Line	0
Benoy Badal Dinesh Bagh	Circular Line	1
Eden Gardens	Circular Line	1
Lake Gardens	Circular Line	0
Park Circus	Circular Line	0
Barasat Junction	Eastern Line	2
Padmapukur	Eastern Line	1
Deshapran	Eastern Line	0
Dhakuria	South Lines	0
Netra	South Lines	0

5 RECOMMENDATION

It is recommended that initially business of Dabawala service be targeted at those stations in Cluster 0 and Cluster 1 which falls in 'Circular Line'. Therefore Aggressive advertisement be targeted in Stations of **Binoy Badal Dinesh Bagh, Lake Gardens, Park Circus, Eden Garden and Ballygynj Junction**. For future growth of the network, more stations from Circular Line in Cluster 0 and Cluster 1 could be targeted. Eventually Stations in Cluster 0 and Cluster 1 in Lines like Eastern Line and South Line could be targeted.

6. CONCLUSION

We find that the area where the service would have the highest demand are not very far from each other and are mostly concentrated around stations in Circular Line. Also Stations along this very Circular Line has many residential area which are potential ground to recruit cooks. This is the best case scenario for starting this business. Also Dabbawala service has great prospect in the city of Kolkata due to many reasons like good intracity train connectivity, huge demand for quality food and opportunity for many women to rightfully become a part of the work force cooking the same quality food that they already cook for their own family. This service with its homemade food, made with all care and low cost due to effective management via use of public transport like Train and also huge social benefit, could be a good competition for online delivery services. Although the Dabbawala service would mostly be a niche service catering mostly to men and women who would need a regular lunch service with guaranteed good quality.