

MARKET ANALYSIS RESULTS

Problem Statement:

Perform Market Analysis to outline data integrity and consistency issues, Derive meaningful insights using visualizations and outline recommendations and action plan to improve the market growth by nearly 20% in 2 months

Did I come across any data integrity and inconsistency issues?

- While diving deep into the dataset, I came across many such instances that could have created problems in terms of data consistency and integrity.

My initial steps working on this assessment includes:

- Removing the duplicates
- Finding the Faulty data records and performing swapping operations to make the dataset more meaningful
- Working with Null value records: I performed meaningful operations to impute the feature null values

To handle the null values, I performed some statistical analysis and logical reasoning before working with **K-means cluster analysis** technique. These operations helped me to impute values for how much quantity of item can be ordered for a given pickup place and how long it takes to place an order. I decided to group the clusters based on the Pickup_place latitude and longitude values as it helped me to find better estimates for the pick-up locations that have null values.

The remaining feature columns: 'When Jumpman arrived Pickup' and 'when Jumpman left pickup' having null values were replaced keeping the logical order in mind i.e.

Datetime for:

When the delivery started < When the jumpman arrived at pickup < When the jumpman left the pickup < When the jumpman arrived at the drop-off

For every single pandas operation that I have implemented for the above logic in the jupyter notebook(.ipynb), I ensured data consistency so that our final data that I work with would be free of all the Data related issues and in logical terms 'makes sense'.

Where the data issues would have arised if not handled?

There were nearly 500 records where the Started delivery time (i.e when the order was placed) is after when the jumpman arrived at the pickup. This in real scenario doesn't make sense and would have impacted my analysis quite badly.

There were duplicates present as well. It was essential to remove these records to ensure consistency in the dataset

I had also come across one discrepancy in the Item Name. 'Shack Burger' AND 'Shackburger' FROM pickup_place 'Shake Shack'. I realized both these entrees have 29 records. So I wanted to cross confirm if these are the same records or different. Though they were different when I checked but still it is necessary for us to determine if there can be any inconsistency involved in such a scenario.

Few Recommendations that I can think of at the moment:

What plan do we need to increase the market growth in the coming months?

While performing the analysis, I realized that most of the dataset includes information about the '**Manhattan**' region, Few about the '**Brooklyn**' region and handful of '**Queens**' Region records. There are many measures that we can take to increase the market growth of the Jumpman23 services:

1. Based on the number of times of the **items ordered** and **item category preferred**, we can ensure to target and include more pick-up places around Brooklyn and Queens regions (the sub-regions that are close to Manhattan) to ensure more options for the preferred item type and item categories
2. As per the Map analysis, I could see that maximum pick-up places are concentrated towards the tail of Manhattan. Though the options availability is more in that region, we can also ensure to include those pickup places that can be also accessible to the residents living in the **Bronx** region as well to increase a market cap by little (sub-regions surrounded by Manhattan)
3. The **customer spread out** for few of the pick-up locations is more as compared to the others (compare pick-up and customer spread out marked red clusters in both the maps). It can help us determine if Jumpman23 can tie up with those pickup places to reach out to the much wider network across the nearby Brooklyn region Queens region. Also, there are few clusters which show that the customers count and preferences are only concentrated in a particular region (For e.g. clusters marked in Black and Green). We can see what is stopping those pick-up places to increase their customer delivery network and if there is any possibility to widen those networks or not
4. If the CEO is willing to invest resources to widen the outreach to attract new customers and to use it's jumpman delivery services, it would be essential to assign jumpman based on their proximity (if at all that is a challenge considering NYC is a crowded place). That means investing a little bit to hire a few Jumpmans that can quickly deliver to the widened network.
5. For the **already existing customers**, The Jumpman services can provide few attractive offers to retain the customers and ensure they order more through their delivery platform. For e.g. Few of my visualizations using bar plots helped me understand that there are customers that have the same food item preferences multiple times. Then there are few customers that choose to order from the same restaurants multiple times (different food items). What we can do is to provide them with some discount offers and complimentary offers after ordering a few more times. Complimentary can be some different food item that they haven't experimented with before so that to widen customers choices in terms of food selection. This sometimes can help customers order more items through the Jumpman23 Platform.
6. The other aspect that we could focus on is '**Time**'. I have visualized that Sundays are the busiest days when the maximum number of orders has been placed , especially in the afternoon. We can analyze if the number of items placed on Fridays (especially evenings) and Saturdays has anything to do with the Pick-up place

timings. If yes then we can choose to include those pick-up places that can have delivery options on Fridays and Saturdays as well.

7. We can also focus on investing resources to ensure more suitable vehicle types. Analysis shows that **Bicycles** are the most preferred vehicle type (which makes sense based on the location we are targeting). But it would be good to see if there are any challenges involved while delivering certain preferred items using bicycles. And if we can come up with any alternative mode of transport to deliver those items so as to increase the number of food items preferences that can be delivered through Jumpman services
8. There are few very popular pick-up places through which the items have been ordered many times by different customers. What we do here is to check why this place is so popular (whether it is because of the quality of food or the type of food etc.) and after identifying those metrics can target near similar restaurants that are almost equally good in terms of the quality and other services but cannot attract more customers due to less advertisement in the market. Jumpman23 platform can tie-up with those restaurants willing to utilize it's services and advertise those places as the '**new**' or '**recommended places**' on its online portal. This ensures a better boost in terms of services with many other options that can be chosen by the existing customers on the Jumpman23 platform