# Coursera Capstone

## IBM Applied Data Science Capstone

*"Where should I open a grocery store with a restaurant inside in Brooklyn?"*

**May 2020**

# Business Case Study

☐ Shoppers today are craving for new experiences

☐ This has led many retailers to look for opportunities of opening restaurants inside their grocery stores

☐ Business question: Where would we recommend that the retailer should open a new grocery store/restaurant?

☐ Deciding which is the best location to open it, is the objective of the analysis

☐ Additional prerequisites: The retailer is interested in offering Italian/Mexican cuisines in the restaurant and must be located close to the center of Brooklyn

# Data

➢ <u>Data</u>

❑List of neighborhoods in Brooklyn, NY

❑Latitude and longitude coordinates of these neighborhoods

❑Venue data, more specifically data related to grocery stores and restaurants

➢ <u>Sources</u>

❑New York dataset page free on the web (https://geo.nyu.edu/catalog/nyu_2451_34572)

❑Geopy library to get the latitude and longitude of Brooklyn

❑ Foursquare API for venue data

# Methodology

➢ <u>Exploratory Analysis</u>

❑ Subtracting the area of Brooklyn of the initial New York dataset

❑ Loading and exploring data for newyork_data.json file

❑ Transforming the list of python nested dictionaries into a pandas dataframe which consists of the geographical coordinates of New York city neighborhoods

❑ Using Foursquare API to get venue category data

❑ Using geopy and Folium libraries to create a map of New York and its neighborhoods superimposed on top

❑ Subtracting the area of Brooklyn of the initial New York dataset and repeating the above process

# Methodology

➢ <u>K-means Clustering</u>

❑ Grouping the data by neighborhood and taking the mean of the frequency of occurrence of each venue category

❑ Keeping only the grocery store, Italian and Mexican restaurant venue categories in a dataframe

❑ Running K-means clustering on the data

❑ Algorithm K-means aims to allocate $n$ observations into $k$ clusters in which each observation belongs to the cluster with the nearest mean

❑ Selecting the optimum cluster based on the lowest frequency of grocery/Italian/Mexican restaurants

❑ Focusing on the selected cluster and re-running K-means to finally conclude to the nearest, to the center of Brooklyn, location

❑ Visualizing the clusters in a map using Folium and finally identifying the best places for opening the grocery store/restaurant

# Results

➢ Running K-means => four clusters based on the lowest on average frequency of occurrence of grocery stores/Italian/Mexican restaurants:

|  | Grocery Store | Mexican Restaurant | Italian Restaurant |
|---|---|---|---|
| **Cluster 0** | **0.68%** | **0.38%** | **0.28%** |
| **Cluster 1** | 1.36% | 7.43% | 0.81% |
| **Cluster 2** | 7.92% | 1.11% | 0.00% |
| **Cluster 3** | 1.72% | 1.03% | 5.46% |

❑ Cluster 0: Neighborhoods with the lowest frequency of grocery stores

❑ Cluster 1: Neighborhoods with the highest frequency of Mexican Restaurants

❑ Cluster 2: Neighborhoods with the highest frequency of grocery stores

❑ Cluster 3 : Neighborhoods with the highest frequency of Italian Restaurants

# Results

➢ Focusing on cluster 0 and by running K-means once again => four clusters based on the distance from Brooklyn

➢ Need to identify the optimum location among neighborhoods of cluster 1 which includes Brooklyn

| Neighborhood | Latitude | Longitude | Cluster Label |
|---|---|---|---|
| Ditmas Park | 40.643675 | -73.961013 | 1 |
| Manhattan Terrace | 40.614433 | -73.957438 | 1 |
| Flatlands | 40.630446 | -73.929113 | 1 |
| Marine Park | 40.609748 | -73.931344 | 1 |
| Prospect Lefferts Gardens | 40.65842 | -73.954899 | 1 |
| East Flatbush | 40.641718 | -73.936103 | 1 |
| Midwood | 40.625596 | -73.957595 | 1 |
| Crown Heights | 40.670829 | -73.943291 | 1 |
| Borough Park | 40.633131 | -73.990498 | 1 |
| Wingate | 40.660947 | -73.937187 | 1 |
| **Brooklyn** | **40.650104** | **-73.949582** | 1 |
| Ditmas Park | 40.643675 | -73.961013 | 1 |
| Manhattan Terrace | 40.614433 | -73.957438 | 1 |

# Results

➢ Visualization of clusters in a map – indicating Brooklyn and three of the closest locations:
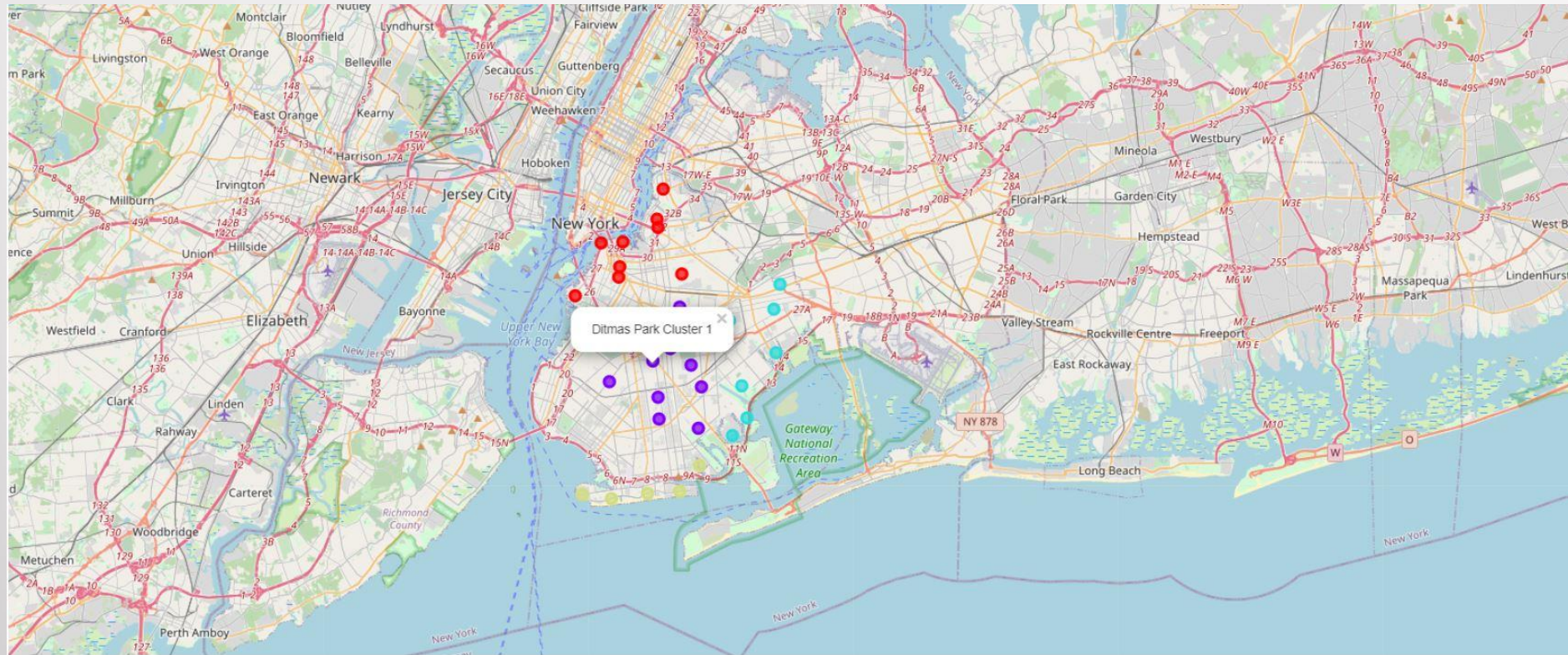
❑ Brooklyn
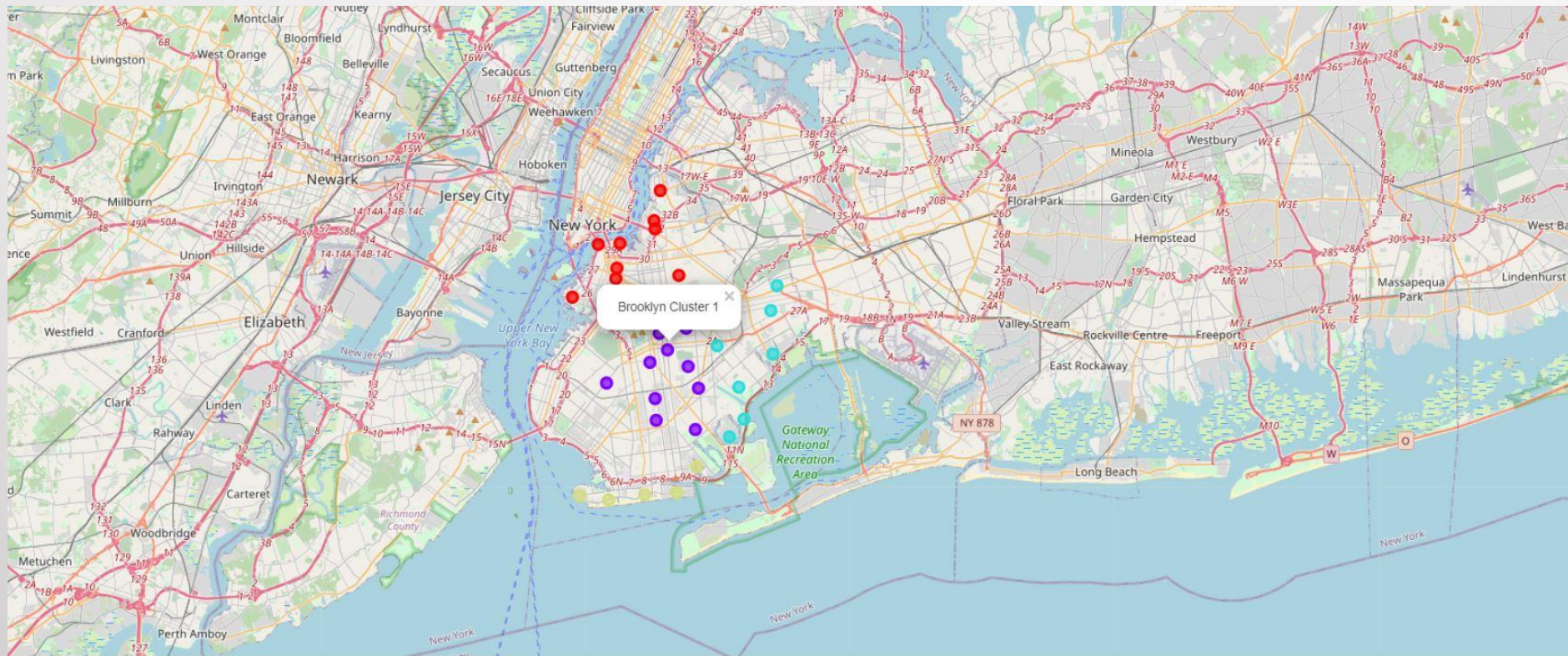
# Results

➢ Visualization of clusters in a map – indicating Brooklyn and three of the closest locations:

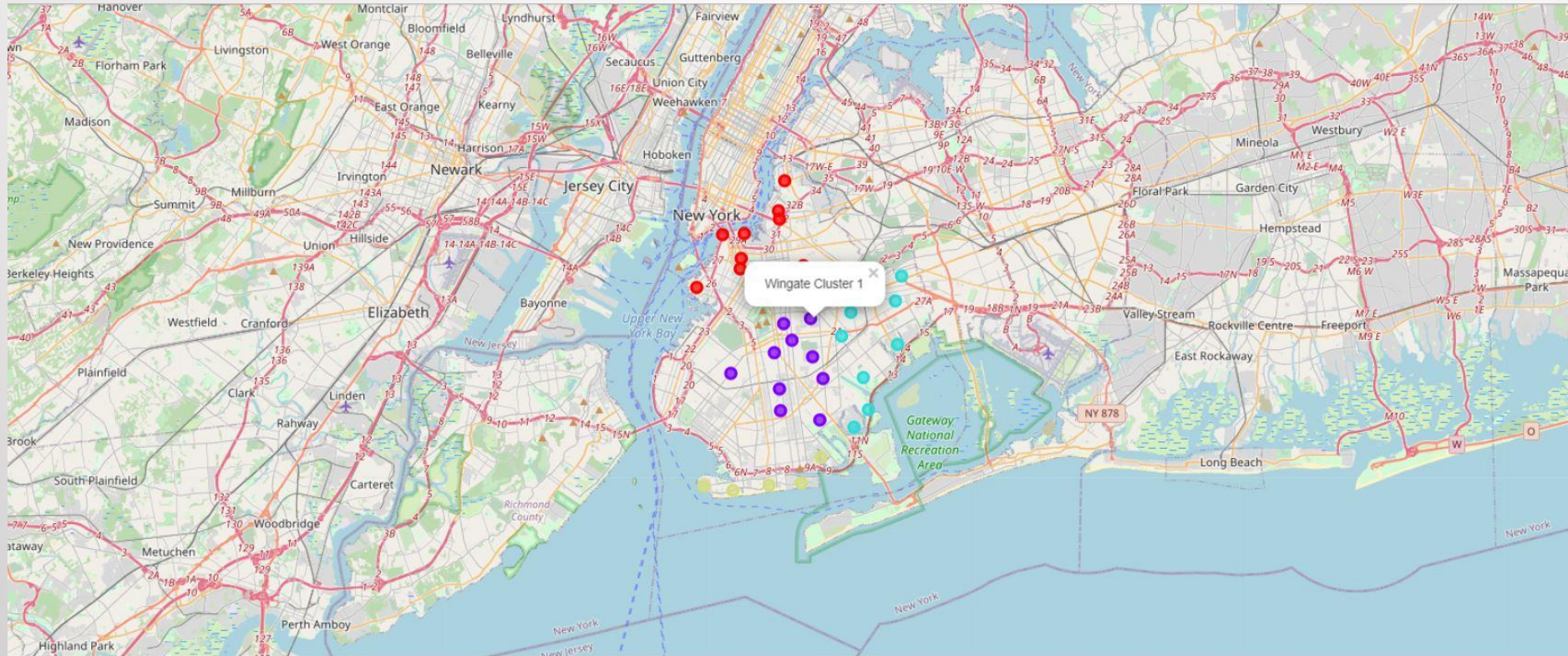❑ Prospect Lefflerts
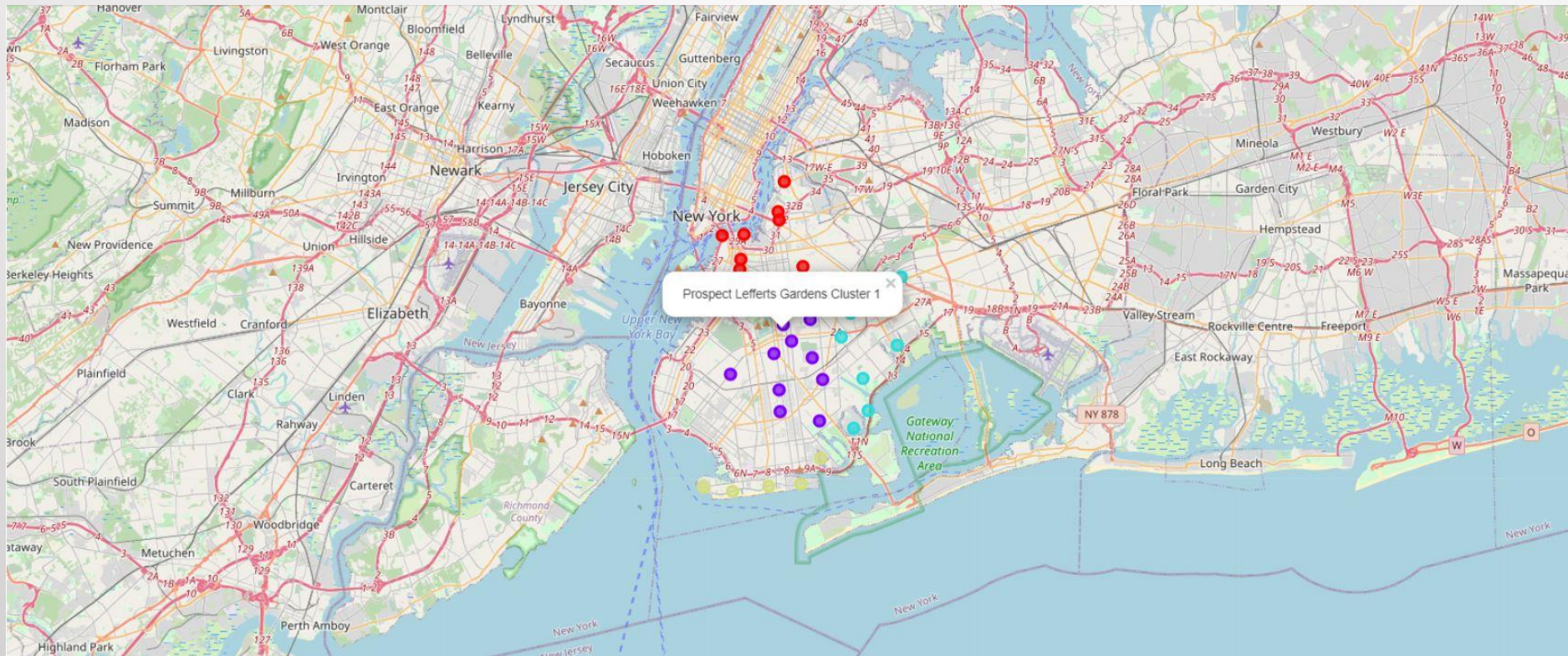
# Results

➢ Visualization of clusters in a map – indicating Brooklyn and three of the closest locations:

❑ Ditmas Park

# Results

➢ Visualization of clusters in a map – indicating Brooklyn and three of the closest locations:

❑ <u>Wingate</u>

# Discussion/Recommendations

➢ <u>K-means clustering (first time)</u>

❑ Cluster 0 is the one with the lowest on average frequency for grocery and Mexican restaurants

❑ Cluster 2 is suffering from competition due to high concentration of grocery stores (7.92%), whereas cluster 1 has the higher frequency in Mexican restaurants (7.43%)

❑ Cluster 3 is characterized by the highest frequency in Italian restaurants (5.46%)

➢ <u>K-means clustering (second time)</u>

❑ Prospect Lefferts Gardens, Ditmas Park and Wingate => the closest to Brooklyn location

❑ Wingate is the suggested area because of zero frequency of groceries/Italian and Mexican restaurants comparing to the other two

➢ <u>Recommendations for future studies</u>

❑ Specific case study's factors => frequency of grocery stores, Mexican and Italian restaurants and distance from center

❑ Potential additional factors => population of each neighborhood, average income of residents, unemployment rate per region, parking availability nearby, cost of renting in each location, cuisine preferences of residents

# Conclusion

➢ Business question: Where should I open a grocery store with a restaurant inside in Brooklyn?

➢ Answer: Most of the neighborhoods in cluster 1 excluding the ones that are in the seaside could be potential selected locations of the retailer with prevalent one Wingate because of zero frequency both in groceries and Mexican/Italian restaurants

➢ Retailers who look for relevant opportunities could benefit from the insights of this study considering that their goal is to expand their grocery stores to more than one area and maximize their revenue by eliminating competition and building their own loyal consumer base

# Thank you!