

# Possibility of business success based on correlation between existing businesses

Essam Qabel

June 6, 2020

## 1. Introduction

### 1.1 Business Problem

Starting an business has been always risky, because after investing money, time and effort into opening the business, due to many factors, it may not succeed, that is why it is essential to study the market first, and the first thing to decide on is the type of industry you will enter. And the success of a certain industry is usually related to the existence of another, so we will use current opening businesses in Alexandria, Egypt to investigate what business work with what and in which neighborhood, and this process can be implemented with any city.

**1.2 Target Audience:** This project can be implemented in a business consultant firm, that give clients - based on their preferences, budgets, or desires - an idea on what might work and what might not.

## 2. Data acquisition and cleaning

### 2.1 Data Sources

We will need to gather the locations of the neighborhoods in Alexandria from Wikipedia and use those locations with foursquare API to gather venue information in each neighborhood.

- **BeautifulSoup:** is used to scrape the locations of the neighborhoods in Alexandria.
- **Foursquare API:** is used to get the venues data for each neighborhood.
- **Nominatim:** Nominatim is a search engine for openstreetmap.org data available for use in Python within the geopy.geocoders library; in the

context of this solution, Nominatim is used to perform geocoding of popular places' address

## 2.2 Data Acquisition and Cleaning

### 2.2.1 Neighborhoods Locations

The first thing to think about is to gather all the names, and locations of all neighborhoods in Alexandria, and we can do this using the information found on Wikipedia.

The URL [https://en.wikipedia.org/wiki/Category:Neighbourhoods\\_of\\_Alexandria](https://en.wikipedia.org/wiki/Category:Neighbourhoods_of_Alexandria) has the names and further links for each neighborhood in Alexandria, so using BeautifulSoup to get the names, and the links for the neighborhoods, and then further extracting the locations of each neighborhood.

I found some locations are missing; this is because either the further investigated link does not exist, or the neighborhood Wikipedia-page does not have the location for that neighborhood.

So I dropped the neighborhood that does not have locations, so that the information is consistent when used with foursquare.com, and after finishing the data cleaning process, I ended up with the following data frame:

```
alex_neighborhoods_df = pd.DataFrame(data)
alex_neighborhoods_df.head()
```

|   | Neighborhood      | Latitude  | Longitude |
|---|-------------------|-----------|-----------|
| 0 | Amreya            | 31.104538 | 29.766226 |
| 1 | Asafra            | 31.20502  | 29.906402 |
| 2 | Azarita           | 31.206121 | 29.914813 |
| 3 | Bahary            | 31.194999 | 29.897915 |
| 4 | Bakos, Alexandria | 31.235555 | 29.957213 |

Fig 1. Neighborhoods locations

### 2.2.2 Neighborhoods Venues

After locating the neighborhoods, and having their latitudes and longitudes, as showed by the following image:

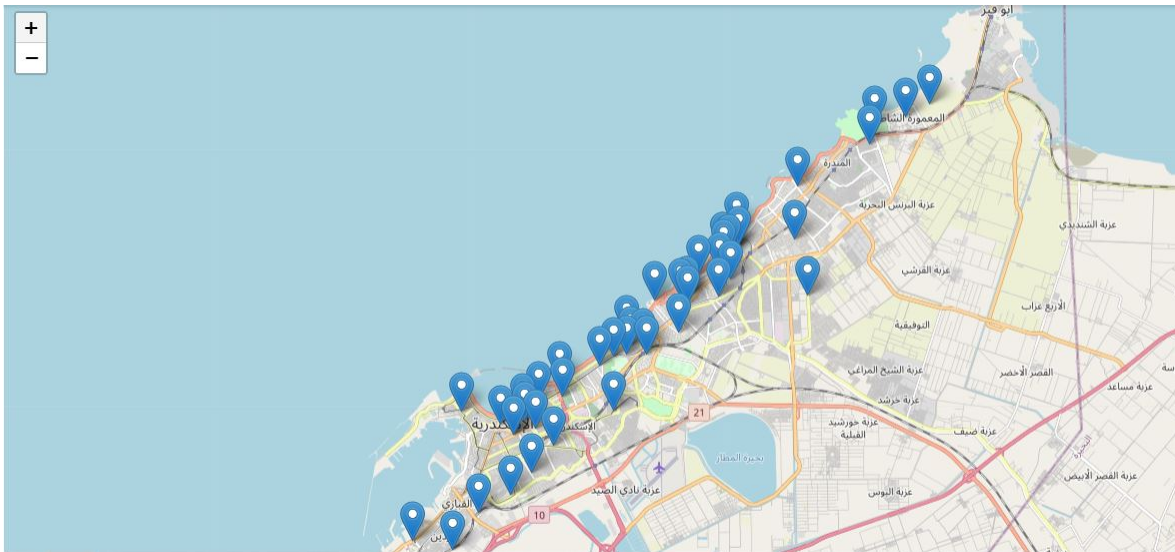


Fig 2. Neighborhoods in Alexandria, Egypt

We can now use Foursquare API to get the venues and their categories in each neighborhood, by submitting the Foursquare API credentials, and the locations of each neighborhood, then combining the data retrieved into a single data frame, as follows:

```
alexandria_venues.head()
```

(805, 7)

|   | Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue           | Venue Latitude | Venue Longitude | Venue Category |
|---|--------------|-----------------------|------------------------|-----------------|----------------|-----------------|----------------|
| 0 | Amreya       | 31.104538             | 29.766226              | Agami (المجمي)  | 31.106045      | 29.763894       | Neighborhood   |
| 1 | Amreya       | 31.104538             | 29.766226              | Hadota          | 31.106702      | 29.762998       | Café           |
| 2 | Amreya       | 31.104538             | 29.766226              | Copacabana Cafe | 31.104593      | 29.768400       | Café           |
| 3 | Amreya       | 31.104538             | 29.766226              | Zizo gym        | 31.101154      | 29.768342       | Gym            |
| 4 | Asafra       | 31.20502              | 29.906402              | Walker's        | 31.204390      | 29.906571       | Bakery         |

Fig 3. Venues of neighborhoods of Alexandria

And then formatting the data in a format that have its columns representing the venues categories, and the index represents the neighborhoods, and the interceptions represent the number of business categories in that neighborhood, as follows:

```
# Get the venue categories descriptions in the neighborhoods
ven_cats_dest_df = temp_df[['Neighborhood', 'Venue Category']].groupby('Neighborhood').apply(print_group).set_index('Neighborhood').fillna(0)
ven_cats_dest_df
```

|                   | Café | Bakery | Bar | Coffee Shop | Dessert Shop | Fast Food Restaurant | Juice Bar | Pizza Place | Egyptian Restaurant | Fried Chicken Joint | Hotel | Sandwich Place | Ice Cream Shop | Shopping Mall | Tram Station | Italian Restaurant | Sup |
|-------------------|------|--------|-----|-------------|--------------|----------------------|-----------|-------------|---------------------|---------------------|-------|----------------|----------------|---------------|--------------|--------------------|-----|
| Neighborhood      |      |        |     |             |              |                      |           |             |                     |                     |       |                |                |               |              |                    |     |
| Amreya            | 2.0  | 0.0    | 0.0 | 0.0         | 0.0          | 0.0                  | 0.0       | 0.0         | 0.0                 | 0.0                 | 0.0   | 0.0            | 0.0            | 0.0           | 0.0          | 0.0                | 0.0 |
| Asafra            | 0.0  | 1.0    | 2.0 | 2.0         | 1.0          | 1.0                  | 1.0       | 1.0         | 0.0                 | 0.0                 | 0.0   | 0.0            | 0.0            | 0.0           | 0.0          | 0.0                | 0.0 |
| Azarita           | 0.0  | 0.0    | 1.0 | 0.0         | 0.0          | 1.0                  | 1.0       | 0.0         | 0.0                 | 0.0                 | 0.0   | 0.0            | 0.0            | 0.0           | 0.0          | 0.0                | 0.0 |
| Bahary            | 4.0  | 0.0    | 1.0 | 1.0         | 2.0          | 0.0                  | 0.0       | 0.0         | 1.0                 | 1.0                 | 1.0   | 1.0            | 0.0            | 0.0           | 0.0          | 0.0                | 0.0 |
| Bakos, Alexandria | 1.0  | 0.0    | 0.0 | 1.0         | 1.0          | 1.0                  | 0.0       | 0.0         | 1.0                 | 0.0                 | 1.0   | 0.0            | 0.0            | 0.0           | 0.0          | 0.0                | 0.0 |
| Bolkly            | 1.0  | 0.0    | 0.0 | 1.0         | 0.0          | 1.0                  | 0.0       | 0.0         | 1.0                 | 0.0                 | 0.0   | 0.0            | 0.0            | 0.0           | 0.0          | 0.0                | 0.0 |

Fig 4. Formatted data frame to represent number of venues in neighborhoods

### 3. Exploratory Data Analysis

The main idea is to investigate the relation between existing businesses in the neighborhoods and relating the correlation to the number of businesses operating in its category, to come up with a new business that has a strong potential of success.

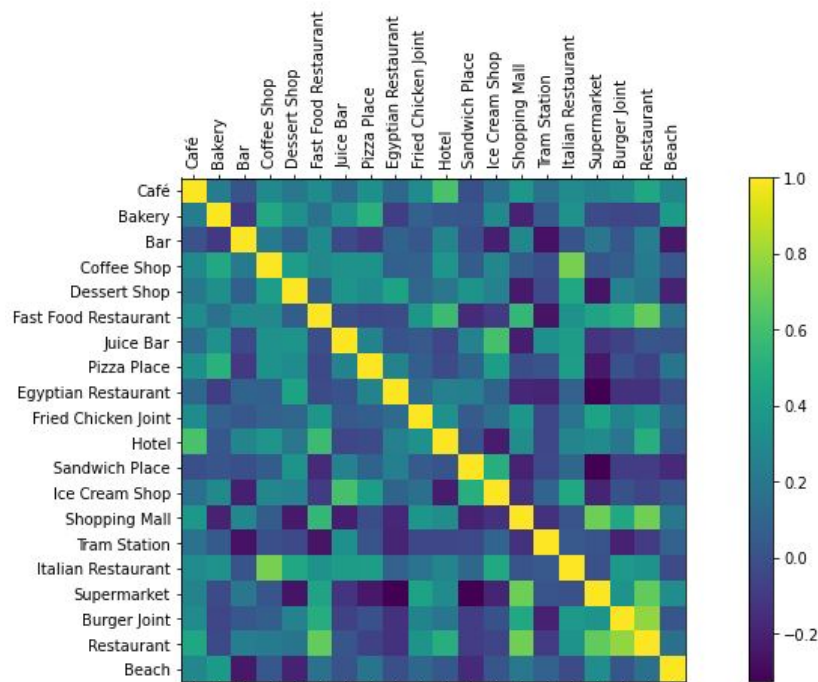


Fig 5. Correlations between venues in neighborhoods

The above figure, allowed us to find out multiple insights, like:

- Pizza Place and a Bakery have a some what high correlation factor (0.52)
- Hotel and a Café have a high correlation factor (0.62)

- Italian restaurant and a Coffee shop have a very high correlation factor (0.73)
- And at the bottom-right we can see multiple high correlations, that is because all of these venue categories [Restaurant, Burger Joint, Fast Food Restaurant], usually exists in shopping malls, or around them, due to high people turn out.

**Note:** We can eliminate the shopping mall related categories, due to high cost, and the client may be limited to a lower cost project.

## 4. Investigating Options

### 4.1 Option one: Pizza place and a Bakery

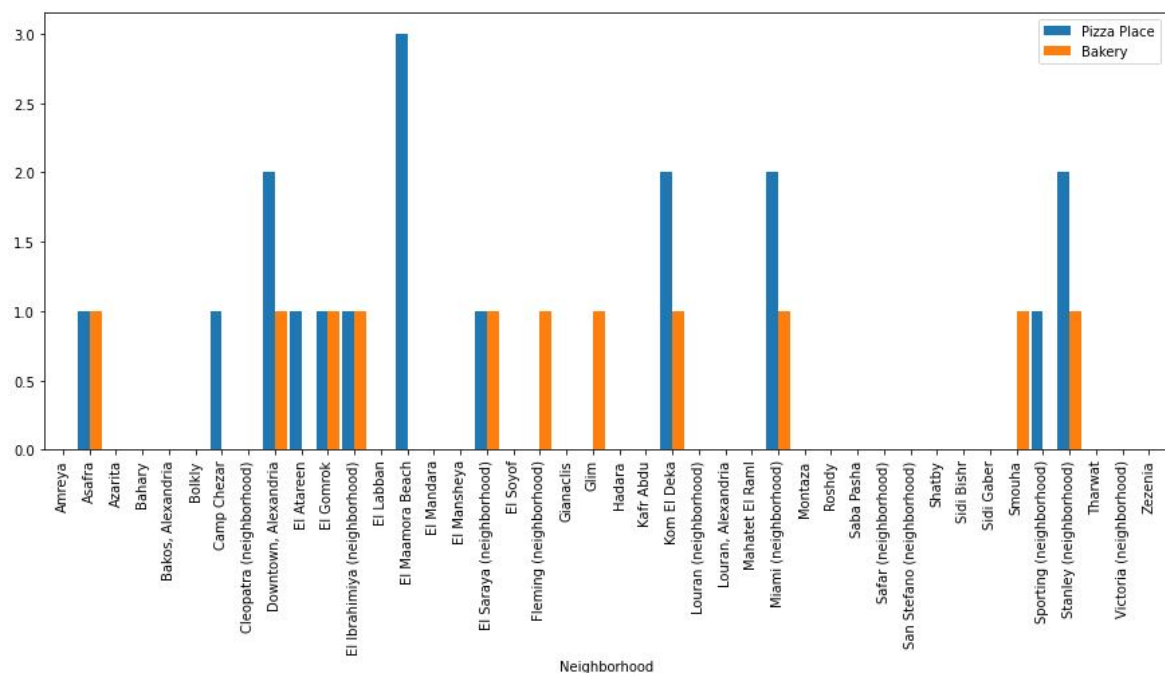


Fig 6. Pizza places and Bakery in Alexandria

As we can see there is some correlation that pizza place rely on the products of a bakery to serve all sorts of pizza types, and for each two pizza places, there usually one bakery shop. And **El Maamora Beach** has three pizza places, with no bakery.

## Conclusion:

The probability of a successful bakery project in El Maamora Beach neighborhood is **High**.

## 4.2 Option two: Hotel and a Cafe

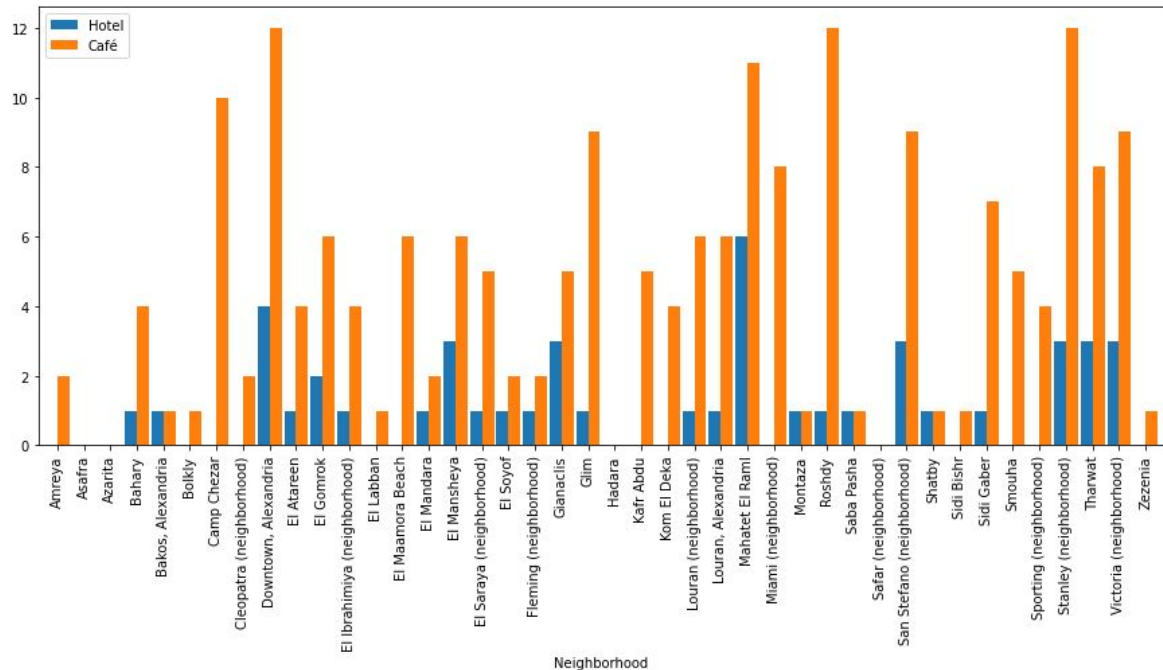


Fig 7. Pizza places and Bakery in Alexandria

While having some outliers like *Camp Chezar* neighborhood - that have no hotels but high number of cafes, that is because it is a campus area of Alexandria University so there are a lot of students going there -, except there is a general trend that with higher number of hotels, there is a higher number of cafes.

## Conclusion:

Due to the overwhelming number of cafes in Alexandria it is **NOT** recommended to open a cafe.

### 4.3 Option three: Italian restaurant and a Coffee shop

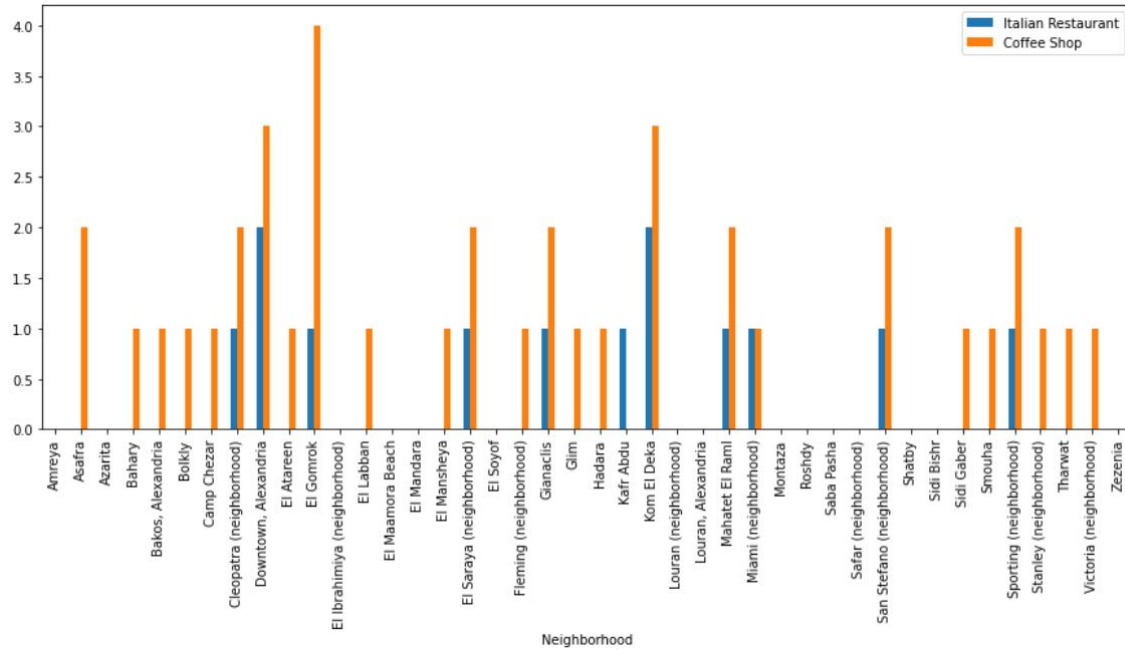


Fig 8. Italian restaurant and a Coffee shop in Alexandria

Going deeper and unveiling the insights of the numbers of Italian restaurants and the coffee shops, there is no clear relation between the two.

#### Conclusion:

There is **no clear indication** on the success or failure of opening an Italian restaurant or a coffee shop.

### 4.4 Option four: Juice bar and an Ice cream shop

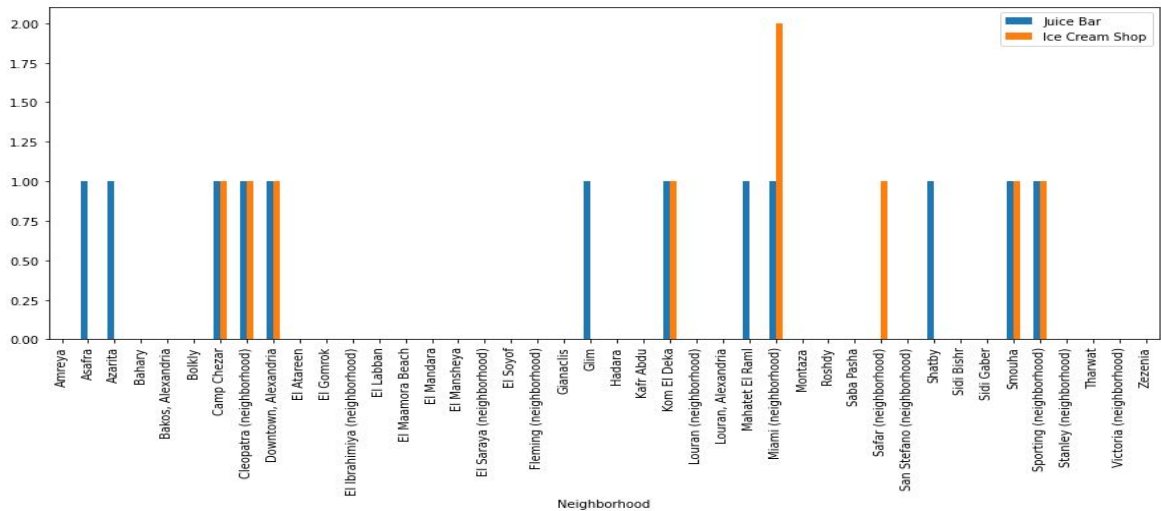


Fig 9. Juice bar and an Ice cream shop in Alexandria

In this option, opening an ice cream shop, or a juice bar appears to be new, or at least not popular yet, because their exist a number of both all-around Alexandria, so opening an ice cream shop, or a juice bar in one of the neighborhoods that don't have one might be an advantage.

### Conclusion:

The probability of opening a successful ice cream shop or a juice bat in a neighborhood that don't have either one of them is **Medium**.

## 5. Conclusion

In this project we walked-through the idea of using correlation between existing different categories of businesses to identify the possibility of a successful business.

And we figured out that there are four possible options to investigate:

- Pizza Place and a Bakery
- Hotel and a Café
- Italian restaurant and a Coffee shop
- Juice bar and an Ice cream shop



And after investigation, we found that **opening a bakery shop in El Maamora Beach neighborhood** is the best option, if we applied the idea of using the correlation between businesses.