# Shopping mall analysis in city of Warsaw

## Coursera capstone project

# Table of Contents

# 1. Introduction

## 1.1 Business Context

Statistics and conclusions from numerous analyses of Polish commercial property cannot be wrong. The first quarter of 2019 indicates continued strong interest in investing in Poland. What is more, our practice and market data show the group of active investors has grown since last year.

One such investor- our client is a commercial real-estate developer who is impressed by the growing potential of Warsaw as an attraction for retail trade. With the ambition to take advantage of this wave, our client is looking to open/acquire a shopping mall in Warsaw.

## 1.2 Business Problem

In order to understand the commercial real-estate market better, the client has approached us to answer the following key questions:

- What is the current landscape of the shopping centers/retail parks/high streets in the city of Warsaw?
    - What is the typical size?
    - Are they relatively modern construction or older?
    - What is their distribution in terms of their location (city-center, office district, residential district, suburbs etc.)
- How are malls that are situated in city center/commercial district different from those situated in residential districts in terms of size and store formats?
- Given, the client is looking to open/acquire a large shopping mall (60000 sq.m) near city center, what should be the typical stores in the mall based on current environment

# 2. Data acquisition and cleaning

## 2.1 Data sources

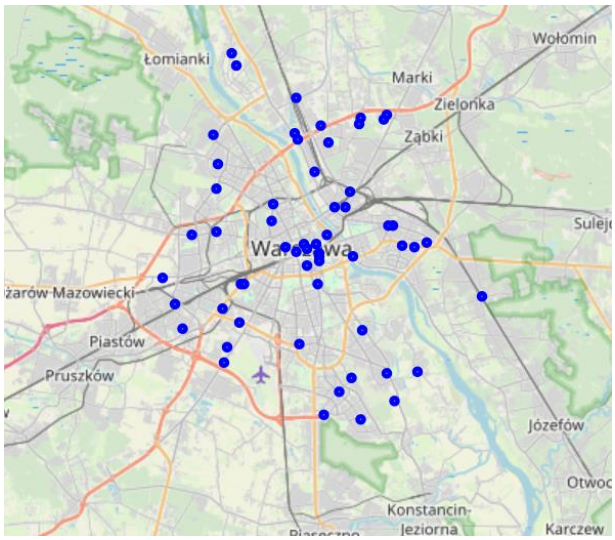
For the given analyses, we will use the dataset of shopping malls/high streets/retail parks available on https://prch.org.pl/en/sc-catalog. The dataset was creating by scraping the contents of the given url. The shopping mall dataset contains the following fields:

- Name of the mall
- Address
- Type of mall (traditional/Speciality/retail-park/high-street/mixed-use)
- GLA (Gross Leasing Area in sq.m)
- Status (Open, In Construction, Closed, Planned)
- Opening year

In addition we will be using the location data from Foursquare API in order to analyze the venues around the location of a particular shopping mall. With the radius of 250 m, we can assume that the venues provided by Foursquare- search query are related to the venues within and around the mall. This will help us in understanding the prevailing store formats of the given shopping mall.

| Type of mall | Number (#) |
|---|---|
| Traditional | 38 |
| Mixed-use | 7 |
| Retail park | 5 |
| Specialty mall | 4 |
| High street | 4 |
| Outlet | 2 |
| **Total** | **60** |

| Status | Number (#) |
|---|---|
| Open | 50 |
| In construction | 7 |
| Planned | 2 |
| Closed | 1 |
| **Total** | **60** |




From the tables above, we see that there are 60 shopping centers in the data, with 38 as traditional malls, 7 as mixed-use, 5 as Retail parks, 4 as Speciality malls and 4 as High streets. Looking at the status of the malls, 1 is closed. We will remove this entry from our data.

## 2.2 Data transformation

In order to answer the first question, we will create 2 additional variables: Distance of the mall from the city center (Distance_center) and Years of operation (Years_operation). In order to calculate the distance from the city center, we take Warsaw's main post office as the point of center. Historically, in Europe, locations of main post offices of the cities- which were typically situated in the city center, were taken to calculate the distance between any two given cities.

| Name | City | Voivodeship | Postal_Code | Address | GEO_X | GEO_Y | Type | Year | GLA | Distance_center | Years_operation |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Arkadia | Warszawa | Mazowieckie | 00-175 | Al. Jana Pawla II 82 | 20.981652 | 52.256346 | Traditional Mall | 2004 | 117000 | 3.1 | 15 |
| ArtN | Warszawa | Mazowieckie | 00-841 | ul. zelazna 51/53 | 20.991977 | 52.232447 | mixed-use | 2020 | 24000 | 1.3 | 0 |
| Atrium ›menada | Warszawa | Mazowieckie | 04-175 | ul. Ostrobramska 75C | 21.106915 | 52.232908 | Traditional Mall | 1996 | 93000 | 6.6 | 23 |
| Atrium | Warszawa | Mazowieckie | 02-326 | Al. Jerozolimskie | 20.951726 | 52.212434 | Traditional | 1999 | 40700 | 4.8 | 20 |

# 3. Methodology

As part of methodology, we first perform clustering on the shopping malls, based on size, distance from center and years since operation. We use the KMeans method to perform such clustering.

We assume the number of clusters to be 4. However, before performing the clustering, we first standardize and normalize the data.

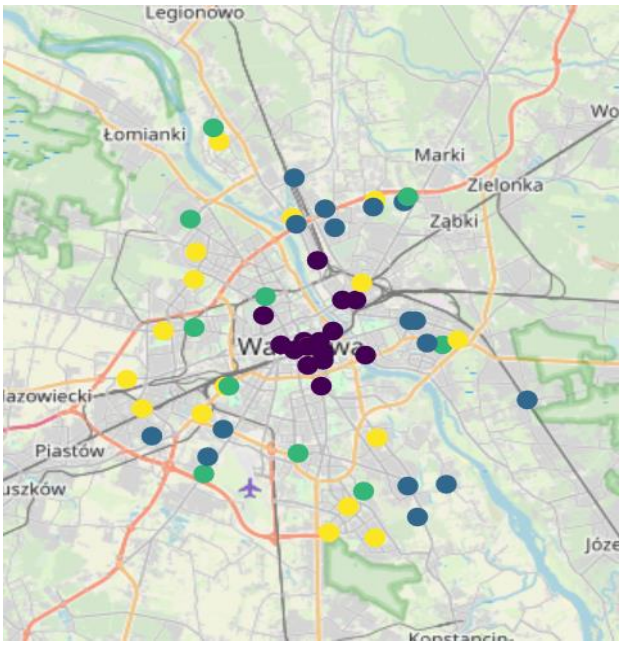
Once segments of shopping malls based on above mentioned parameters are obtained, we access Foursquare data to get nearby venues around each shopping mall. Assuming the radius of 250m, we can assume that the nearby venues provided by Foursquare are the stores within the particular shopping mall.

We compare and contrast the shopping malls characteristics and store formats to answer the questions posed by our client.

# 4. The current landscape of the shopping centers in Warsaw

In order to understand the characteristics of the shopping centers in Warsaw, we take 3 features- size (GLA), location (Distance_center) and years of operation (Years_operation). Let's segment the shopping centers based on these features using KMeans method.

| Cluster Labels | index | Status | Name | City | Voivodeship | Postal_Code | Address | GEO_X | GEO_Y | Type | Year | GLA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 0 | Open | Arkadia | Warszawa | Mazowieckie | 00-175 | Al. Jana Pawla II 82 | 20.981652 | 52.256346 | Traditional Mall | 2004 | 11700 |
| 1 | 1 | In construction | ArtN | Warszawa | Mazowieckie | 00-841 | ul. zelazna 51/53 | 20.991977 | 52.232447 | mixed-use | 2020 | 2400 |
| 3 | 2 | Open | Atrium Promenada | Warszawa | Mazowieckie | 04-175 | ul. Ostrobramska 75C | 21.106915 | 52.232908 | Traditional Mall | 1996 | 9300 |




**FOUR SEGMENTS OF SHOPPING CENTERS**

**Cluster 0: Older, medium sized malls, located in residential districts**

**Cluster 1: Smaller sized units, located near city-center**

**Cluster 2:New/upcoming centers, mainly medium sized, located in residential districts**

**Cluster 3:Middle aged, large sized malls, located in business districts of Warsaw**

### 4.1 Results

First cluster seems to be group of shopping centers that are relatively older and are situated in the residential districts of the city. Almost all the malls within this cluster are medium sized, with GLA ranging between 20k sq.m to 40k sq.m

Second cluster seems to be group of shopping centers that are situated in the city center. Most of the shopping malls are relatively newer and smaller sized.

Third cluster consists of medium sized malls situated in residential districts of the city, and are newer or upcoming constructions.

The last cluster comprises of shopping malls that large sized and are situated in the business districts of Warsaw. Moreover, these malls seem to be middle-aged.

## 5. Difference in store formats of residentially and commercially located shopping centers

As we analyzed in the previous section, cluster 0 and 2 belong to shopping centers in the residential districts, while cluster 1 and 3 belong to shopping centers located in city center or business districts of the city.

In order to understand the difference in the store formats of these location, we can leverage the Foursquare data to get nearby venues of shopping centers in these areas. Assuming a radius of 250 m, we can assume that most of the venues provided by Foursquare would be located within the shopping malls.

### 5.1 Exploring store formats of all shopping centers of Warsaw

In order to understand the difference in store formats of residential and city center/commercial malls, we apply one hot encoding method to determine the frequency of each venue category in each shopping mall. We then create a variable to identify a particular mall as residential or commercial. Lastly, we determine the average frequency of particular venue category for residential and commercial shopping malls to identify the difference in the store formats

| Res_Comm_label | 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 |
|---|---|---|---|---|---|---|---|---|---|---|
| commerical | Coffee Shop | Café | Cocktail Bar | Clothing Store | Fast Food Restaurant | Restaurant | Dessert Shop | Hotel | Shopping Mall | Plaza |
| residential | Coffee Shop | Dessert Shop | Supermarket | Clothing Store | Shopping Mall | Fast Food Restaurant | Electronics Store | Sporting Goods Shop | Café | Pizza Place |

### 5.2 Results

What we see from the above table is that the category of stores in shopping malls located in city center/commercial districts of Warsaw are more 'entertainment based' with focus on cafes, cocktail bars, restaurants. This could be influenced by the factor that these districts also attract a large volume of tourists.

On the other hand, the shopping malls located in residential districts dominate in 'utility based' stores such as Supermarket, Electronics store and sporting goods shop. These stores are family oriented and exist to cater to household needs of the city.

## 6. Store formats for large mall near city center

### 6.1 Identifying the relevant cluster for the client

Based on the identified characteristics of shopping malls in the city, the typical store format of a large shopping mall near city center should resemble to Cluster 3.

The prevailing store formats of cluster 3 shopping malls is as per the table below:

| Cluster | 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 | Coffee Shop | Electronics Store | Clothing Store | Pizza Place | Dessert Shop | Supermarket | Shopping Mall | Sandwich Place | Fast Food Restaurant | Restaurant |
| 1 | Coffee Shop | Café | Cocktail Bar | Restaurant | Hotel | Plaza | Dessert Shop | Bar | Polish Restaurant | Hostel |
| 2 | Bus Station | Shopping Mall | Dessert Shop | Supermarket | Fast Food Restaurant | Coffee Shop | Sporting Goods Shop | Café | Furniture / Home Store | Clothing Store |
| 3 | Clothing Store | Fast Food Restaurant | Coffee Shop | Shopping Mall | Pizza Place | Electronics Store | Bookstore | Cosmetics Shop | Supermarket | Café |

### 6.2 Results

As we see from the table, the cluster 3 shopping malls have a wide variety of store category including fast food restaurants, clothing store, Electronic stores and supermarket. These malls,

owing to their size and location, are designed to cater to most of the needs of the visitors, that is both, entertainment based and utility based.