Safe area in Vancouver Canada
for the restaurant business
by Zholamanov Zhastalap

2020

Table of content

1. INTRODUCTION

1.1. Background

Vancouver is a major city in western Canada, located in the Lower Mainland region of British Columbia. As the most populous city in the province, the 2016 census recorded 631,486 people in the city, up from 603,502 in 2011. As in any big city, big crime. Vancouver, all in all, is a very safe place to visit or live but keeping a smart lookout while venturing around the city is key. Downtown is very busy for most hours of the day and therefore is fairly safe, particularly the West End. Personal safety is extremely high in Vancouver. Despite other statistics, it is necessary to identify the problem by analyzing crime data in Vancouver City and finding the safest area and the area in the area that best meets the requirements of our business task.

1.2. Business case

The goal of this project is to open a restaurant business of Russian cuisine in safe areas of Canada in the city of Vancouver, which would interest all interested parties of this business, including the investor, but the main priority is safety. The first task is to select the safest area, having analyzed crime data for the restaurant business and included in the list of neighborhoods, where a restaurant that represents Russian cuisine is not one of the most popular places, and at the same time are as close to the city as possible. To do this, we will use our data processing tools to analyze data and focus on the safest area and explore its surroundings.

2. DATA ACQUISITION AND CLEANING

2.1. Source

We will use the geographical coordinates of Vancouver to build areas in a safe area of the city, in the immediate vicinity of the city, and finally, group our areas and present our results. To extract / generate the necessary information, we need the following data sources:

• Using a real world dataset containing crime details in Vancouver from Kaggle. (https://www.kaggle.com/agilesifaka/vancouver-crime-report/version/2)

• Gathering additional information on the list of classified districts in Vancouver from Wikipedia. (https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Vancouver)

• Creation of a new cluster of data on neighborhoods, together with their areas, crime data and the corresponding coordinates of the neighborhood. This latitude and longitude data will be selected using a geocoder.

List of columns in dataset:

- Neighborhood: Name of the neighborhood in the Borough.
- Borough: Name of the Borough.
- Latitude: Latitude of the Borough.
- Longitude: Longitude of the Borough.

• Folium to visualize Maps, KMeans from SciKit will also be used.

• Creating a new data cluster by region, the data source is the Four Square API.

2.2. Data Cleaning

Data from the Kaggle data source was a heavy file that Git could not accommodate. Therefore, it was decided to focus on recent crimes of 2018, which significantly reduced the number of lines in csv file from 600,000 to 37,000.

I found out that not all the categories of data in a dataset are useful for this case. Therefore, information such as the month and hour in which the crime occurred were excluded.

```
In [52]: van_crime_df.head()
Out[52]:
```

	TYPE	YEAR	MONTH	DAY	HOUR	NEIGHBOURHOOD
0	Break and Enter Commercial	2018	3	2	6	West End
1	Break and Enter Commercial	2018	6	16	18	West End
2	Break and Enter Commercial	2018	12	12	0	West End
3	Break and Enter Commercial	2018	4	9	6	Central Business District
4	Break and Enter Commercial	2018	10	2	18	Central Business District

1.1.   Data set after cleaning

```
van_neighbour_borough.head()

Total Neighbourhood Count 24 Borough Count 4
Out[79]:
```

	Neighbourhood	Borough
0	West End	Central
1	Central Business District	Central
2	Hastings-Sunrise	East Side
3	Grandview-Woodland	East Side
4	Mount Pleasant	East Side

1.2.   Data set has been generated from Wikipedia

```
van_borough_crime.head()
Out[81]:
```

	Type	Year	Month	Day	Hour	Neighbourhood	Borough
0	Break and Enter Commercial	2018	3	2	6	West End	Central
1	Break and Enter Commercial	2018	6	16	18	West End	Central
2	Break and Enter Commercial	2018	12	12	0	West End	Central
3	Break and Enter Commercial	2018	3	2	3	West End	Central
4	Break and Enter Commercial	2018	3	17	11	West End	Central

1.3.   Data set after merging crime and Neighborhood datasets

```
van_ws_eighbourh_geo
```

Out[197]:

	Neighbourhood	Borough	Latitude	Longitude
0	Shaughnessy	West Side	49.251863	-123.138023
1	Fairview	West Side	49.264113	-123.126835
2	Oakridge	West Side	49.230829	-123.131134
3	Marpole	West Side	49.209223	-123.136150
4	Kitsilano	West Side	49.269410	-123.155267
5	Kerrisdale	West Side	49.234673	-123.155389
6	West Point Grey	West Side	49.264484	-123.185433
7	Arbutus Ridge	West Side	49.240968	-123.167001
8	South Cambie	West Side	49.246685	-123.120915
9	Dunbar-Southlands	West Side	49.253460	-123.185044

In addition to data analysis, it was also necessary to obtain data on latitude and longitude in order to display the surroundings on a map for a better visual image, for this we create a data cluster similar to the above.

3. METHODOLOGY

3.1. Data analysis

Pandas describe() is used to view some basic statistical details like percentile, mean, std etc. of a data frame or a series of numeric values. When this method is applied to a series of string, it returns a different output, which is shown in the examples below.
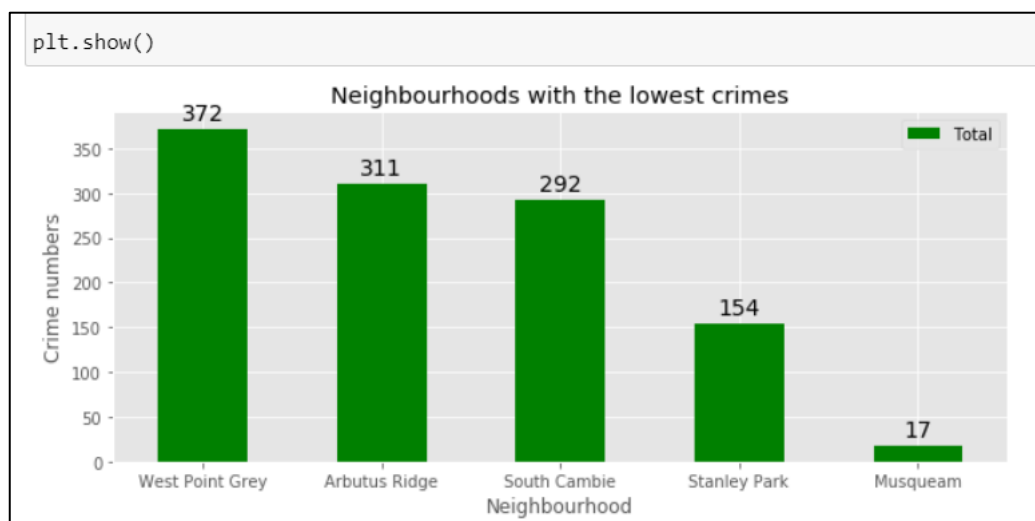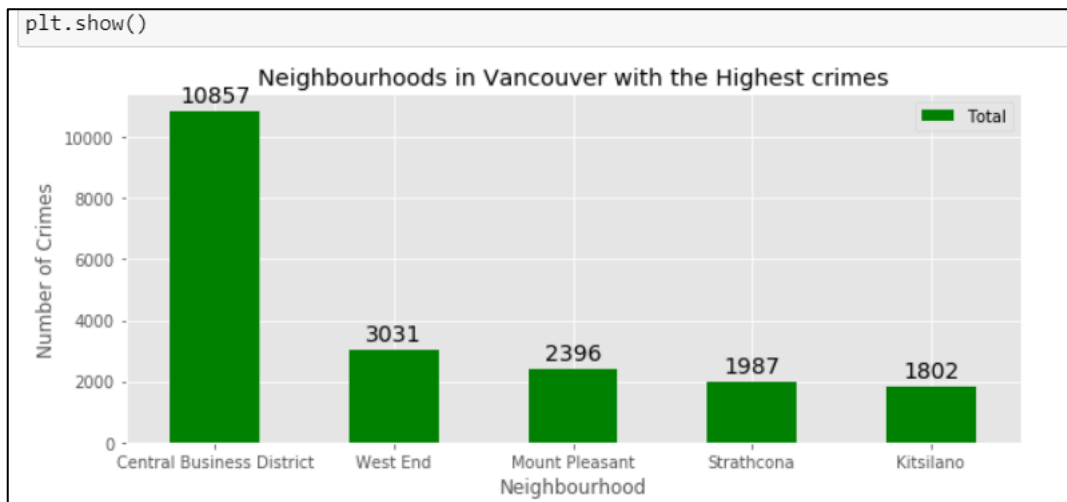
In [71]:
```
van_crime_d.describe()
```
Out[71]:

	YearBreak and Enter Commercial	YearBreak and Enter Residential/Other	YearMischief	YearOther Theft	YearTheft from Vehicle	YearTheft of Bicycle	YearTheft of Vehicle	YearVehicle Collision or Pedestrian Struck (with Fatality)	YearVehicle Collision or Pedestrian Struck (with Injury)	Total
count	4.000000	4.000000	4.00000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000
mean	506.250000	599.250000	1430.25000	1236.750000	3736.500000	539.750000	286.500000	3.250000	368.500000	8707.000000
std	354.409721	488.189427	997.26572	1060.087221	2723.536977	353.955153	226.117226	3.304038	227.060198	5801.870618
min	49.000000	156.000000	187.00000	88.000000	483.000000	36.000000	71.000000	1.000000	111.000000	1182.000000
25%	314.500000	187.500000	843.25000	544.000000	2249.250000	450.000000	186.500000	1.000000	263.250000	5698.500000
50%	594.500000	599.000000	1627.00000	1185.000000	3796.000000	633.000000	235.000000	2.000000	351.500000	9802.000000
75%	786.250000	1010.750000	2214.00000	1877.750000	5283.250000	722.750000	335.000000	4.250000	456.750000	12810.500000
max	787.000000	1043.000000	2280.00000	2489.000000	6871.000000	857.000000	605.000000	8.000000	660.000000	14042.000000

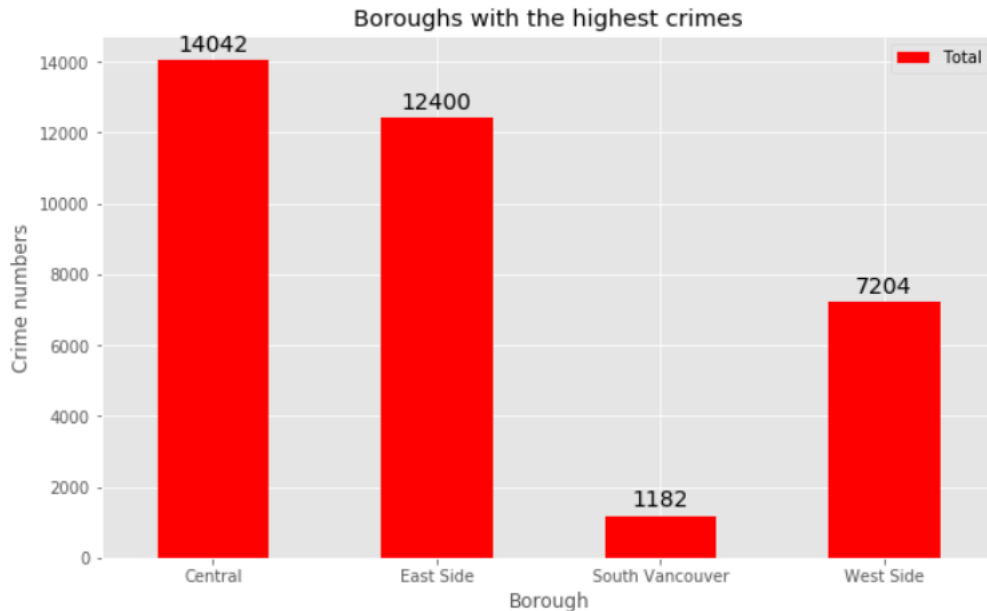## 3.2. Neighborhoods with the highest crime rates

Comparing the crime rate among all neighborhoods, we see that **Central** business occupies the bulk of criminal records, which explains why in the **Central** area of Vancouver there are the largest number of crimes that we will study in recent years, the only area on the **West Side**, which is one of the lowest in the top five.

```
plt.show()
```

Neighbourhoods in Vancouver with the Highest crimes


```
plt.show()
```

Neighbourhoods with the lowest crimes

3.3. Boroughs crime analysis

Comparing the crime report in four boroughs, South Vancouver has the lowest crime rate, probably due to its low neighborhood, followed by the West Side, which, despite having up to 10 districts, has fewer crimes than central Vancouver.

```
plt.show()
```


Since there are very few neighborhoods in South Vancouver, and opening a business would not be appropriate, we can choose the next area with the lowest crime rate, namely West Side. Despite the fact that the West side has crime, it is chosen because crime is relatively low and Commercial enter also low, which makes West Side an ideal place to open commercial establishments.

```
plt.show()
```
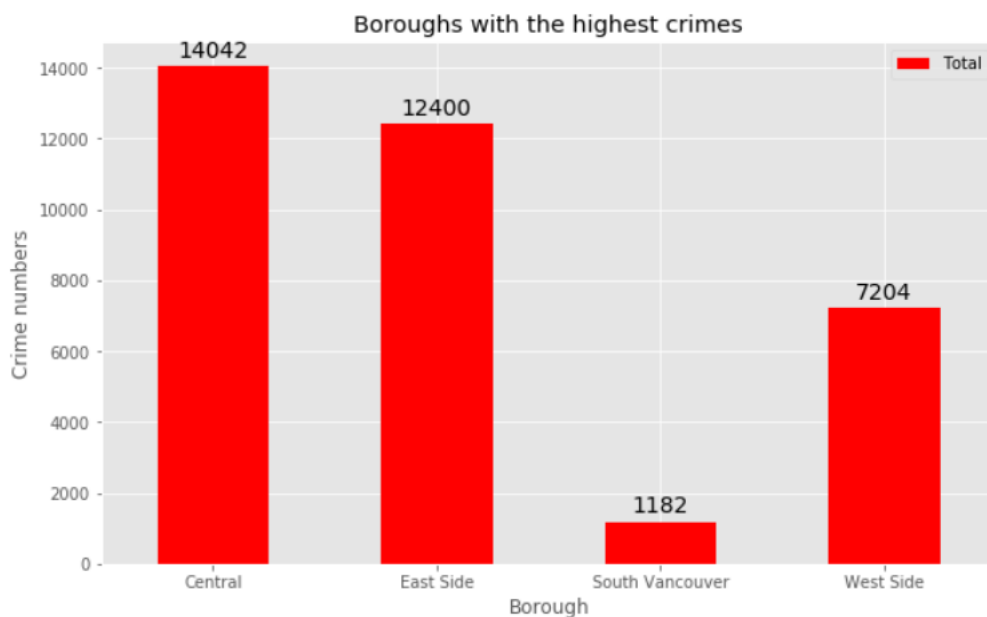
3.4.   West side borough neighborhoods
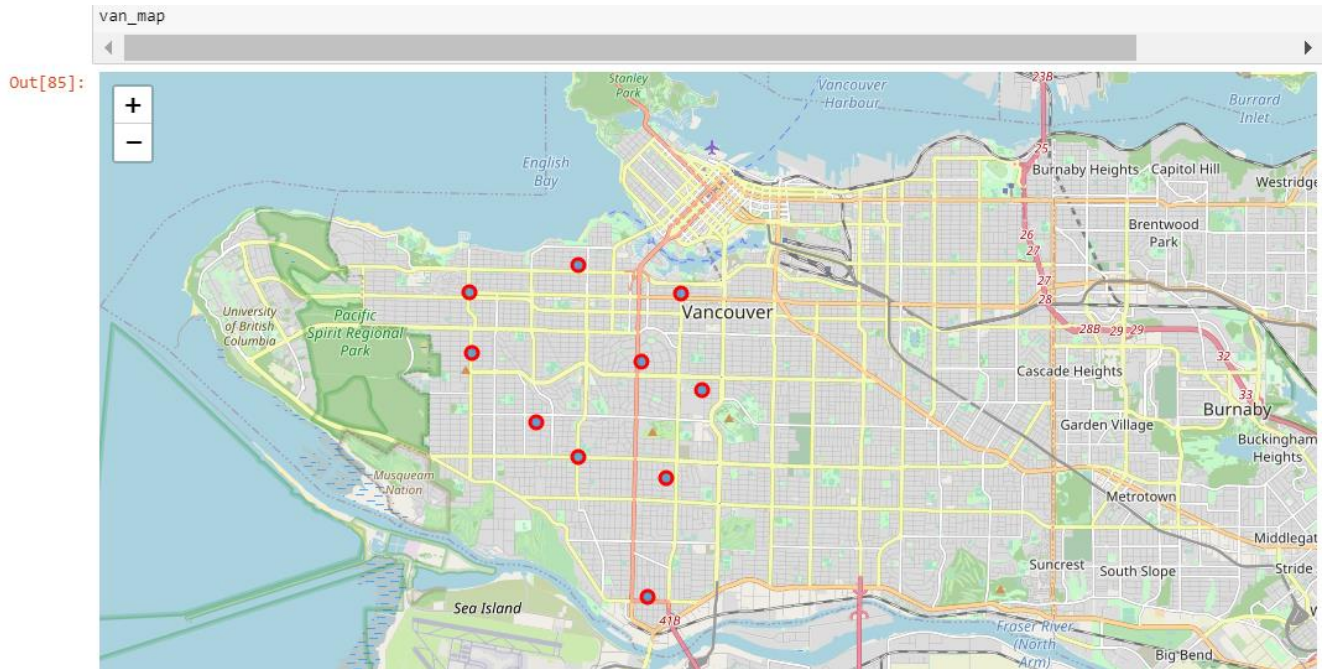
1.4.   There are 10 neighborhoods in the West Side borough

3.5.   Modelling

Based on the data set of the neighborhoods and areas, as well as the latitude and longitude of the western part of Vancouver, we can find all objects within a radius of 500 meters from each area by connecting to FourSquareAPI, where json was used that contains all the places in each neighborhood.

```
In [161]: print(van_ws_venue.shape)
          van_ws_venue.head()

          (227, 5)
```

Out[161]:

|	Neighbourhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Category
0	Shaughnessy	49.251863	-123.138023	Angus Park	Park
1	Shaughnessy	49.251863	-123.138023	Crepe & Cafe	French Restaurant
2	Fairview	49.264113	-123.126835	Gyu-Kaku Japanese BBQ	BBQ Joint
3	Fairview	49.264113	-123.126835	CRESCENT nail and spa	Nail Salon
4	Fairview	49.264113	-123.126835	Charleson Park	Park

3.6. Clustering

Clustering of K-means access to each created cluster was obtained to see which neighborhoods were assigned to each of the five clusters. Here's what the map looks like

	Borough	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
1	West Side	Coffee Shop	Park	Asian Restaurant	Malay Restaurant	Diner	Pharmacy	Nail Salon	Falafel Restaurant	Chinese Restaurant	Camera Store
3	West Side	Sushi Restaurant	Chinese Restaurant	Vietnamese Restaurant	Pizza Place	Dessert Shop	Dim Sum Restaurant	Coffee Shop	Falafel Restaurant	Plaza	Massage Studio
4	West Side	Bakery	American Restaurant	Thai Restaurant	Japanese Restaurant	Coffee Shop	Sushi Restaurant	Ice Cream Shop	Food Truck	French Restaurant	Grocery Store
5	West Side	Coffee Shop	Chinese Restaurant	Tea Room	Pharmacy	Sushi Restaurant	Sandwich Place	Thai Restaurant	Bakery	Gym Pool	Pizza Place
6	West Side	Coffee Shop	Japanese Restaurant	Sushi Restaurant	Café	Bakery	Bookstore	Pub	Asian Restaurant	Vegetarian / Vegan Restaurant	Women's Store
8	West Side	Coffee Shop	Bus Stop	Juice Bar	Vietnamese Restaurant	Light Rail Station	Liquor Store	Bank	Café	Malay Restaurant	Sushi Restaurant

The first cluster is the largest cluster with 6 out of 10 quarters in the Westside area. After carefully examining these areas, we see that the most common places in these areas are restaurants, snack bars, parks, and food trucks. We understand that restaurants are not among the most common places, which makes this cluster of neighborhoods an ideal place to create a restaurant business.
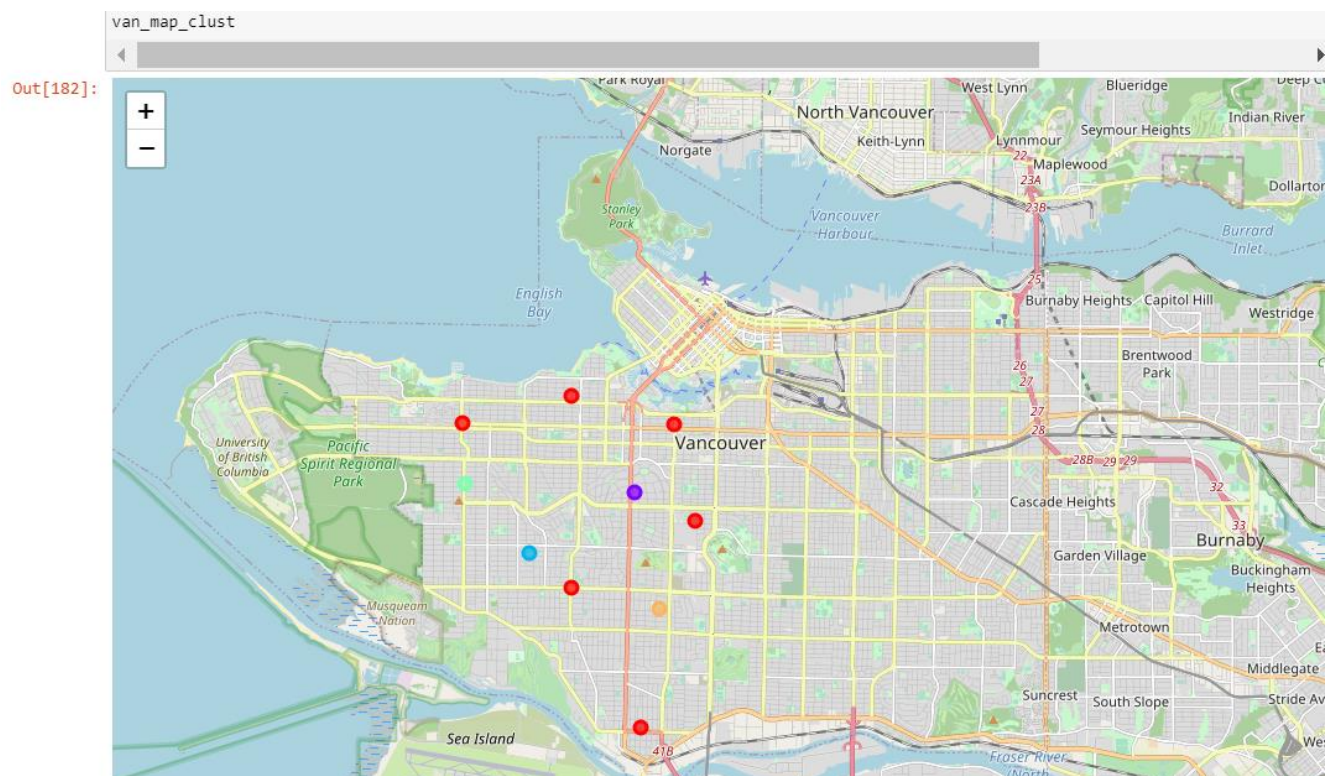
Each cluster is color coded for ease of presentation, we can see that most of the neighborhoods fall into the red cluster, which is the first cluster. Each of the remaining areas is part of the remaining four clusters and is represented by different colors.

4. BRIEF DISCUSSION

The task of the business task was to help interested parties determine one of the safest areas and whether there is an analogue of the business being launched in this area to create a restaurant business of Russian cuisine. This was achieved thanks to the first use of crime data in Vancouver to identify a safe area with a significant number of neighbors so that any business is viable. After choosing the area, it was necessary to choose the right area where the planned business was not among the places in close proximity to each other. We achieved this by grouping the areas in clusters to help stakeholders by providing them with relevant information about the locations and security of the area.

5. CONCLUSION

Crime data was studied to understand the different types of crimes in all areas of Vancouver, and then categorized them into different areas, which helped us group the areas into areas and first choose the safest area. After we confirmed that the number of areas to be considered is also decreasing, we additionally selected areas based on common locations to select the area that is best suited for the restaurant business. In the future, the scale of this project can be taken into account by the population of the region, which is an additional factor that will have a big impact for further positive decisions.