

# REPORT/SUMMARY OF THE CAPSTONE PROJECT

## Seattle vs. Vancouver (with Gdansk in the background) Battle of Neighbourhoods

---

### Background

**Jan Kowalski, PhD** is a professor of data science at the Gdańsk University of Technology. He specialises in the processing of natural languages, although other areas of data processing and data analysis are also of interest to him. He published numerous articles and books in his field of interest.

<https://pg.edu.pl>

and

a professor of applied linguistics University of Gdańsk.

<https://ug.edu.pl>

He specialises in the processing of natural languages, although other areas of data processing and data analysis are also of interest to him. He published numerous articles and books in his field of interest.

He is well known in international circles thanks to the participation in numerous seminars and other events where he delivers presentations of his extraordinary achievements.

He is 40 years old, single, devoted to his work.

He likes theatre and baroque music and contemporary art. From time to time he likes to watch a good film.

He also is an accomplished sailor.

Professor Kowalski regularly receives invitations from various foreign universities to come and to teach and run research there.

Recently he has received two invitations to work:

from

**University of Seattle, Seattle, WA**

and

from

**University of British Columbia, Vancouver, BC**

Professor Kowalski is supposed to run classes for students of the universities in the following majors:

- data science
- data science in computational Linguistics.

He will also be able to continue his work in the area of processing natural languages.

For the detailed course of work please see the notebook:

**SvV(3).ipynb**

also available in my github under the url:

[https://github.com/btillack/Capstone\\_project/blob/main/SvV\(3\).ipynb](https://github.com/btillack/Capstone_project/blob/main/SvV(3).ipynb)

The notebook contains all the maps, pictures and all the information related to this project.

---

## I. INTRODUCTION

The project covers the gathering of the information constituting the basis for the decision of choosing between Seattle, WA and Vancouver, BC as a settlement place.

The decision will be made on the basis of socio-cultural basis.

Within the project we are choosing the city having a wider offering the following fields of interest:

Field of interest	Type of venue
contemporary art	Art Gallery
baroque music	Concert Hall / Music Venue
drama	Theatre
film	Cinema
sailing	Yachtclub or Harbor / Marina

It is based on the fictitious character Jan Kowalski, PhD, professor of Polish universities in Gdańsk who was offered positions at competing universities, one in Seattle, and one in Vancouver.

The list presented above was later extended by additional venue types:

Field of interest	Type of venue
contemporary art	Arts & Crafts Store
	Museum
music and theatre	Performing Arts Venue

The analysis may also be useful to other people considering moving to the West coast of America.

---

## II. GEO-LOCATION DATA SOURCES AND ACQUISITION

### 1. General location

In the first step, we gather the location information about both cities.

The cities are shown on the maps to analyse the geographical aspects.

As a result we can conclude, the location of both cities make it easy for Prof. Kowalski to pursue his sailing activities.

Seattle is located on the shores of Puget Sound while Vancouver lies on Georgia Strait, areas known among sailors as very interesting and demanding.

### 2. Administrative division and clustering

In order to analyse the socio-cultural offerings of both cities it is necessary to find the location of the most interesting venues.

As a starting point the data base of the postal codes was taken into consideration.

There is a publicly available full register of all postal codes in the US and Canada.

Filtering the codes of Vancouver and Seattle gave back still enormous database without the connection of the code to the name of the neighbourhood.

Therefore a different approach has been assumed, gather the information about neighbourhoods from the sites of Seattle and Vancouver municipalities.

Such gathered data was augmented by the information from Google Maps.

Such gathered data was prepared in the form of a csv-file and uploaded to the notebook.

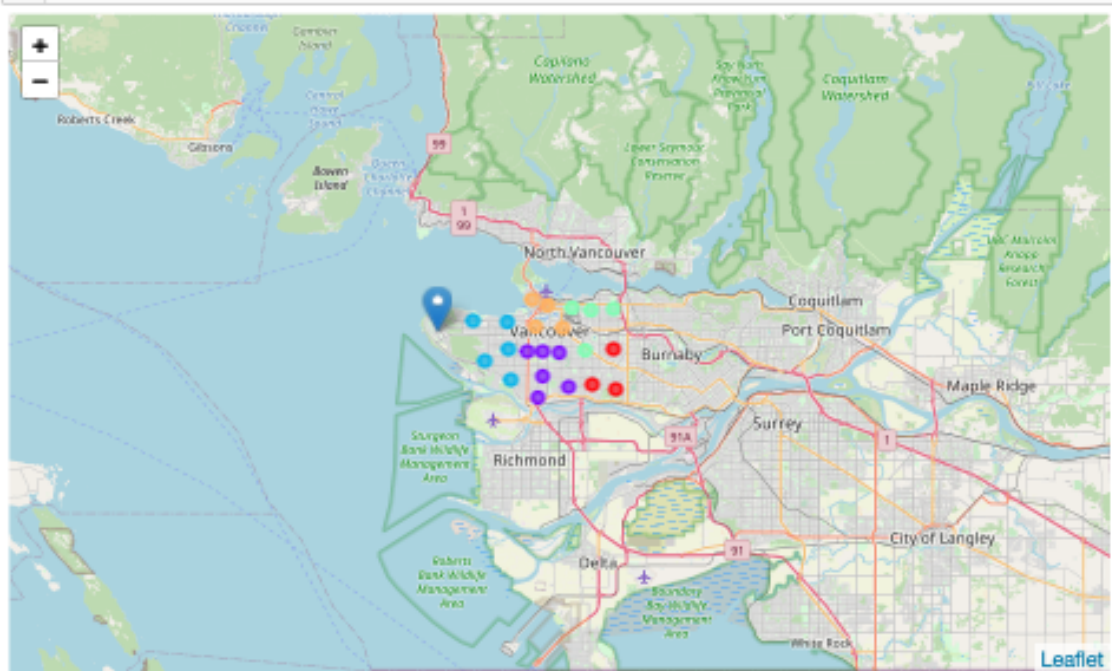
As a result the dataframes containing the information on neighbourhoods in Seattle and Vancouver were created.

### Clustering

The neighbourhoods were clustered and presented in relation to the Universities.

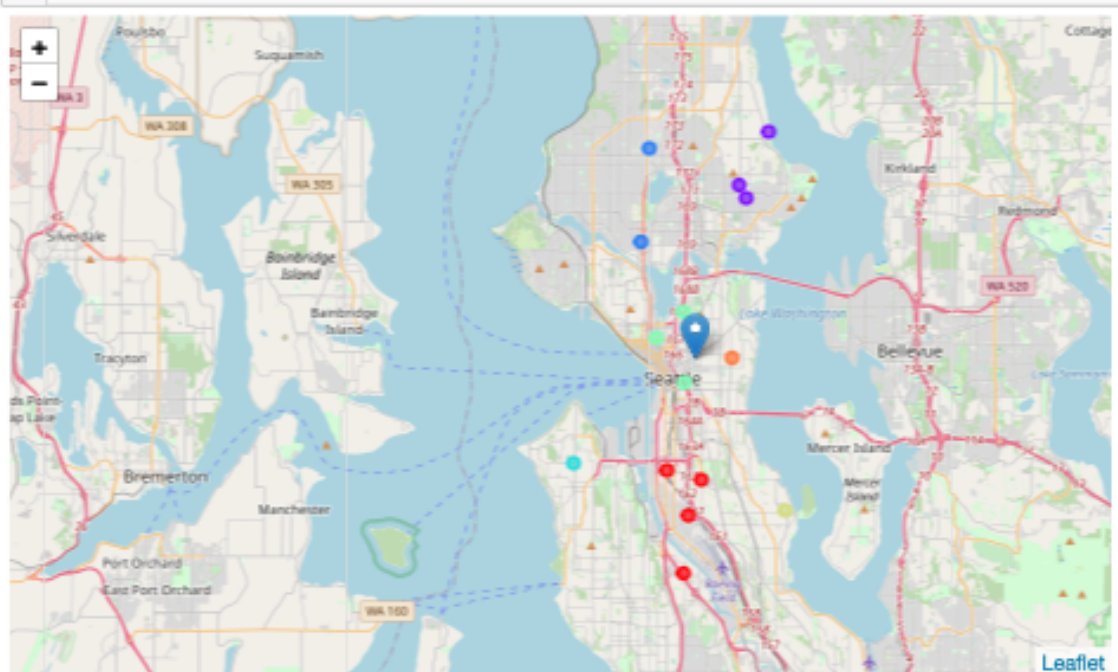
## Vancouver Neighbourhoods Clustering

Out[18]:



## Seattle Neighbourhoods Clustering

Out[161]:



This way the centroids of the clusters lying in the nearest vicinity to the future workplace will be the central point for data sourcing from Foursquare.

Using the Foursquare profile we will import the information about the venues in both cities and analyse them regarding prof. Kowalski's personal preferences.

---

### III. DETAILED VENUE DATA AND ANALYSIS OF DATASETS

#### Data from Foursquare

To import the data from Foursquare the centres of particular centroids were chosen with the radius of 1000 m.

From the datasets received for both, Seattle and Vancouver only the venues of the particular type were selected.

Taking into account the interests of our fictitious character we analysed the number of venues of particular type in the set vicinity.

In order to measure the value of the results in light of the personal preferences of Prof. Kowalski we employed the table with weights of particular venue types. This will let to build the final mark best suited to Mr. Kowalski's liking.

Below, the datasets for Vancouver and Seattle present the grouping by the venue category. This layout allows for analysing the preferences of our character:

#### Vancouver Venues<sup>1</sup>

	Venue Category	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude
0	American Restaurant	1	1	1	1	1	1
1	Art Gallery	2	2	2	2	2	2
2	Arts & Crafts Store	5	5	5	5	5	5
3	Asian Restaurant	4	4	4	4	4	4
4	Athletics & Sports	2	2	2	2	2	2
...	...	...	...	...	...	...	...
149	Video Store	1	1	1	1	1	1
150	Vietnamese Restaurant	10	10	10	10	10	10
151	Wine Shop	2	2	2	2	2	2
152	Women's Store	2	2	2	2	2	2
153	Yoga Studio	2	2	2	2	2	2

154 rows x 7 columns

---

<sup>1</sup> for detailed datasets see the notebook SvV(3).ipynb in this repository on Github [https://github.com/btillack/Capstone\\_project/blob/main/SvV\(3\).ipynb](https://github.com/btillack/Capstone_project/blob/main/SvV(3).ipynb) or the notebook-report [https://github.com/btillack/Capstone\\_project/blob/main/SvV-Summary\\_Report.ipynb](https://github.com/btillack/Capstone_project/blob/main/SvV-Summary_Report.ipynb)

## Seattle Venues

	Venue Category	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude
0	American Restaurant	5	5	5	5	5	5
1	Antique Shop	1	1	1	1	1	1
2	Arcade	1	1	1	1	1	1
3	Art Gallery	2	2	2	2	2	2
4	Arts & Crafts Store	3	3	3	3	3	3
...	...	...	...	...	...	...	...
139	Vietnamese Restaurant	6	6	6	6	6	6
140	Warehouse Store	2	2	2	2	2	2
141	Wine Bar	2	2	2	2	2	2
142	Winery	2	2	2	2	2	2
143	Yoga Studio	5	5	5	5	5	5

144 rows x 7 columns

Extracting the right data from the above datasets allows to present the detailed information for each city, using the weighing factors as in the table below:

### Weighing factors

<b>Art Gallery</b>	20%
<b>Art Gallery</b>	20%
<b>Concert Hall</b>	15%
<b>Music Venue</b>	15%
<b>Theatre</b>	10%
<b>Cinema</b>	5%
<b>Harbour/Marina</b>	35%
<b>Total</b>	100%

You can find the tables summarising the result of the survey including the weights assumed for the particular areas of interest of prof. Kowalski as below:

### Results for Vancouver:

Venue Category	Weight	Number	Points
Art Gallery	20%	2	0,4
Concert Hall	15%	3	0,45
Music Venue	15%	1	0,15
Theatre	10%	3	0,3
Cinema	5%	0	0
Harbour/Marina	35%	1	0,35
Total	100%	10	1,65

### Results for Seattle:

Venue Category	Weight	Number	Points
Art Gallery	20%	2	0,4
Concert Hall	15%	0	0
Music Venue	15%	0	0
Theatre	10%	1	0,1
Cinema	5%	0	0
Harbour/Marina	35%	0	0
Total	100%	3	0,5

The above presented results obviously show that the **Vancouver's socio-cultural offer with 1.65 scoring is much richer than Seattle's with scoring of 0.5.**

As Seattle performed very poorly in comparison to the Vancouver in the first attempt, Prof. Kowalski reviewed the list of venue categories and extended the search with changed weights to accommodate the extended list of venues.

---

## IV. VERIFICATION OF THE DATA QUALITY

After performing the above analysis Prof. Kowalski was very surprised with the modesty of the socio-cultural offer of both cities.

He compares the offer of Seattle and Vancouver to what he can find in the Tri-City area (Gdańsk-Sopot-Gdynia) where he leaves now (approx. 700 thousand inhabitants).

There are more theatres, an opera, the Philharmonic and numerous museums and galleries. There are many events even off-season, notwithstanding, the very sailor-friendly Gulf of Gdańsk and Southern Baltic Sea.

### General information

<https://www.gdansk.pl/en/about-gdansk>

### Cultural guide (unfortunately, only in Polish)

<https://kultura.trojmiasto.pl>

### Pictures

[https://galeria.trojmiasto.pl/#Kultura\\_i\\_rozrywka](https://galeria.trojmiasto.pl/#Kultura_i_rozrywka)

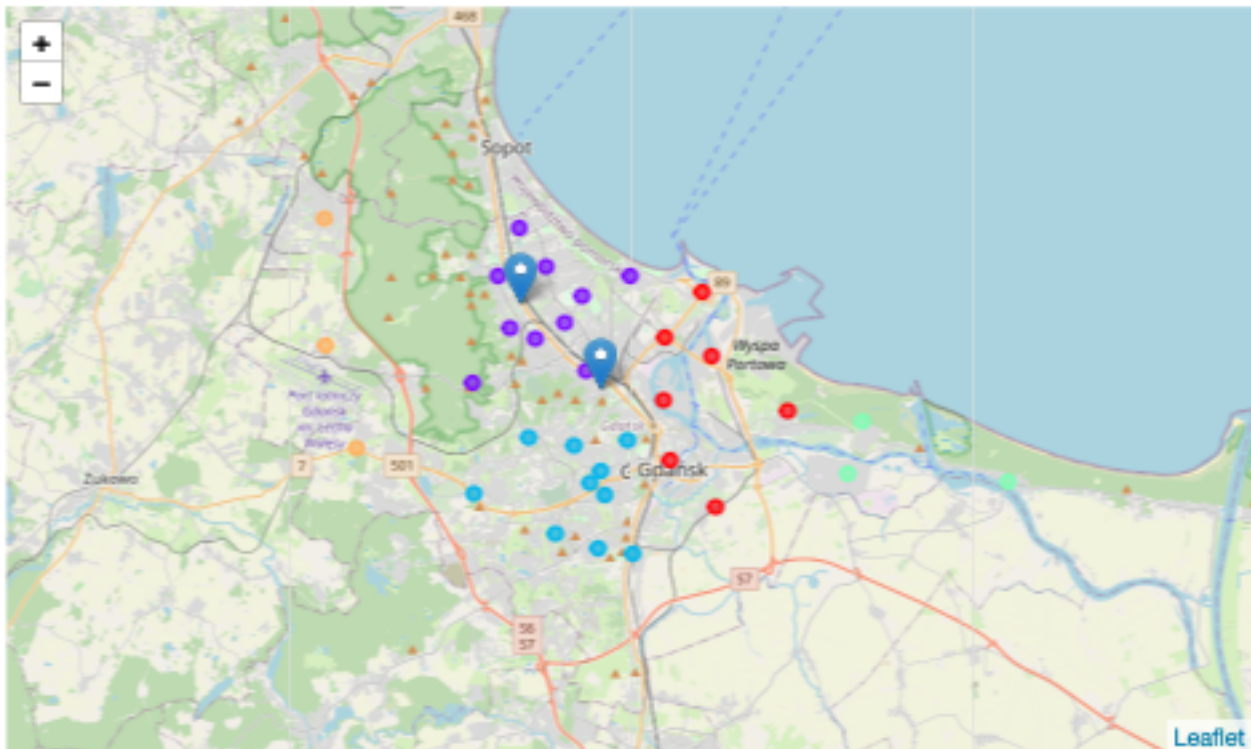
Therefore, to compare the quality of the data source, Prof. Kowalski decided to perform the analysis for his home town.

This way he will be able to see how "filtering" the limits imposed by Foursquare really are.

## GDANSK

Prof.Kowalski repeated the whole procedure for his home-town, i.e.

1. he imported geo location data for the neighbourhoods (Wiki, gdansk.pl, Google Maps)
2. Clustered the neighbourhoods (as in the picture below)





Knowing his home-town he would rather change the clustering in regard to Śródmieście-Główne Miasto neighbourhood.

This is so called "Old Town", a main historic centre of Gdańsk rich in historic and cultural venues and it was clustered together with the industrial districts.

The location of this neighbourhood was arbitrarily chosen as a centre of search in Foursquare.

However, at this moment we are testing the influence of the limits imposed by Foursquare and not the results of clustering.

### Sourcing and analysing the data from Foursquare

As a result of the analysis it turned out that the socio-cultural offering of Gdańsk is not much better than the offering of Seattle.

Gdansk - socio-cultural offering

Venue Category	Weight	GDAŃSK	
		Number	Points
Art Gallery	15%	6	0,9
Concert Hall	10%	1	0,1
Music Venue	10%	0	0
Theatre	8%	1	0,08
Cinema	5%	0	0
Harbour/Marina	25%	0	0
Int. total	73%	8	1,08
Arts & Crafts Store	12%	0	0
Museum	5%	8	0,4
Performing Arts Venue	10%	0	0
<b>TOTAL</b>	<b>100%</b>	<b>16</b>	<b>1,48</b>

This was a surprise to Prof. Kowalski because he is abreast of the cultural events taking place in Gdańsk and knows most of the cultural venues of the types used in this analysis. In his opinion Gdańsk should have gathered much more points but he observed the following shortcomings of the data set:

- the list did not contain OPERA as a separate category,
- CINEMAs, non-existent and although the cinematic multiplexes are usually parts of the shopping malls, there are still independent one-hall (studio) cinemas in Gdańsk,
- his preferred venues - MARINAs - are generally located on the Eastern part of the city farther then 1 km from the city-centre although there is one representative yacht-harbour on the Motława River in the very city centre,
- the dataset cannot include Sopot and Gdynia, smaller members of the Tricity agglomeration, which can contribute at least three theatres (two dramatic and one musical) and numerous art galleries and other interesting venues, including a marina in the city centre of both cities.

---

## V. CONCLUSION

Bearing on the gathered data the conclusion is pretty obvious.

The ranking shows that Vancouver gathered 2.04 points while Seattle - 1.37.

Vancouver is, therefore, a better choice than Seattle assuming everything else being equal, especially academic opportunities.

**Therefore Prof. Kowalski should accept the proposal from the University of British Columbia in Vancouver, BC.**

The decision will be easier to make as the comparison with his home-town showed the advantage of Vancouver over Gdańsk, too.

Knowing Gdańsk and its socio-cultural life, in Prof. Kowalski's opinion the data from Foursquare does not show the full potential of the city of Gdańsk. This same conclusion must be applied also to both contenders, i.e. the results must be deemed acceptable. This, however, does not change the picture of Vancouver's advantage over Seattle.

---

## VI. Observations

This notebook contains venue data for the radius  $r=1000$  m.

But the test was done with the radius  $r=2500$  m in an attempt to widen the area in search for POIs. The results received were the same as in the case of radius  $r=1000$  m. Increasing the radius did not bring the expected results as the limit on number of venues is 100 and cannot be increased and is the main limiting factor for this analysis.

---

## VII. PRESENTATION

This document is a final report from the Capstone project: Seattle versus Vancouver, the battle of neighbourhoods.