



# Battle of neighborhood

CAPSTONE – WEEK 5

# Business goals

- ▶ Gdansk is one of the most popular tourist destination in Poland
- ▶ Due to large number of tourist, but also well paid workforce, Gdansk seems to be a good place for people interested in investing in the restaurant industry.
- ▶ The investor want to open a new dumpling restaurant.
  - ▶ **Where is the best place for establishing the restaurant?**

# Data sources

- ▶ List of postal codes/streets in Gdansk (webpage):
  - ▶ <https://trojmiasto.onet.pl/kod-pocztowy-gdansk-lista-kodow-pocztowych-artykul/7ftd6ws>
- ▶ Detailed locations:
  - ▶ <https://geocode.xyz/>
- ▶ Most popular venues (Foursquare API):
  - ▶ <https://developer.foursquare.com/docs/api/endpoints>

# Python packages

- ▶ Pandas - Library for Data Analysis
- ▶ NumPy – Library to handle data in a vectorized manner
- ▶ JSON – Library to handle JSON files
- ▶ Requests – Library to handle http requests
- ▶ Matplotlib – Python Plotting Module
- ▶ Sklearn – Python machine learning Library
- ▶ Folium – Map rendering Library
- ▶ BeautifulSoup – Library for parsing HTML sources

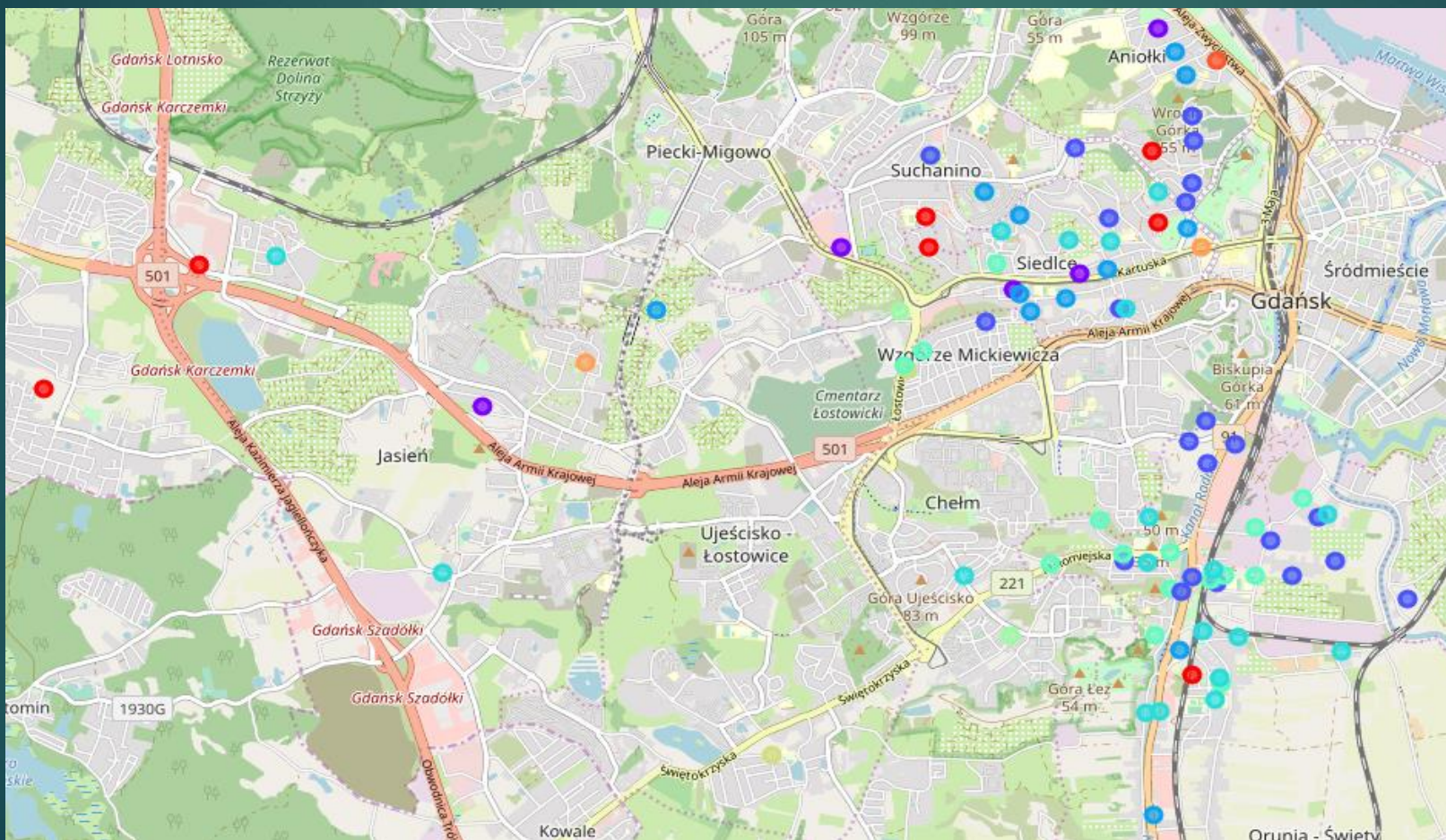
# Methodology

- ▶ For each location check the list of 10 most common venues using Foursquare API
- ▶ Create clusters using machine learning k-means algorithm basing on most common venues
- ▶ Select a cluster where dumpling restaurant is among the most popular venues
  - ▶ Select all addresses within the cluster where the dumpling restaurant is still not popular



# Results

- ▶ 10 clusters based on most common venues among residents was found





# Results

- ▶ One cluster was selected as best place for new dumpling restaurant
- ▶ The best address was found

