# The Battle of Neighborhoods – Final Report

1. Introduction
   1. Background

The Nuremberg Metropolitan Region comprises 3.5 million people on 21,800 square kilometres. It consists of the cities of Nuremberg, Fürth, Erlangen, Bayreuth and Bamberg and is one of Germany’s strongest economic areas. Due to a decline in historically prevalent industry, such as consumer electronics the area has lacked behind in economic development compared to other more famous German regions, such as Munich or Stuttgart.

However, this is also means that real estate and wages are lower compared to its contemporaries. Thus, potential investors find a large pool of well-educated workers, consumers and relatively cheap real estate.



* 1. Business Understanding/Problem Description

The optimal location for an investor would maximize population density, while minimizing real estate prices and competition. These values vary significantly from district to district and from city to city.

Therefore, we want to create a map, which charts all areas according to its real estate values, population and venue density.

Afterwards, each district is clustered according to the density of venues and business opportunities.

1. Data Description

The following data sources were identified to tackle the business problem:

* The number of venues within the certain radius of each district (Foresquare API)
* The net income per citizen per district. Source:

<http://www.boeckler.de/pdf/wsi_vm_verfuegbare_einkommen.xlsx>

* The population and the population density of the district. Source:

<http://www.daten.statistik.nuernberg.de/geoinf/ia_bezirksatlas/atlas.html>

* The housing prices per district. Source:

<https://www.wohnungsboerse.net/mietspiegel-Nuernberg/2176>

* The coordinates of each district. Source: Open Street Map

<https://nominatim.openstreetmap.org/ui/search.html?q=nuremberg>

* 1. Dataset
  2. Neighbborhood

1. Methodology
   1. Data Collection
   2. Analytic Approach
2. Analysis
   1. Exploratory Data Analysis
   2. Clustering
3. Result
4. Discussion
5. Conclusion
6. References

[1] <https://en.wikipedia.org/wiki/Nuremberg_Metropolitan_Region>

**[2]** <http://www.daten.statistik.nuernberg.de/geoinf/ia_bezirksatlas/atlas.html>

**[3]** **http://www.boeckler.de/pdf/wsi\_vm\_verfuegbare\_einkommen.xlsx**

[4] [Forsquare API](https://developer.foursquare.com/" \t "_blank)

[5] <https://nominatim.openstreetmap.org/ui/search.html?q=nuremberg>

Links:

<https://www.wsi.de/de/einkommen-14582-17976.htm>

<https://www.capital.de/immobilien-kompass/nuernberg>

<http://www.daten.statistik.nuernberg.de/geoinf/ia_bezirksatlas/atlas.html>

<https://www.metropolregionnuernberg.de/daten-fakten>

<https://www.boeckler.de/pdf/wsi_vm_verfuegbare_einkommen.pdf>

<https://github.com/zhumazik/IBM-Capstone-Project/blob/master/Report.pdf>

<https://github.com/zhumazik/IBM-Capstone-Project/blob/master/Capstone_project.ipynb>

<https://github.com/rupeshvins/IBM-Applied-Data-Science-Capstone/blob/master/Data%20Science%20Project%20Week%205th%20-%20IBM.pdf>

<https://www.linkedin.com/pulse/housing-sales-prices-venues-data-analysis-ofistanbul-sercan-y%C4%B1ld%C4%B1z/>

For this week, you will required to submit the following:

1. A description of the problem and a discussion of the background. (**15 marks**)
2. A description of the data and how it will be used to solve the problem. (**15 marks)**

For the second week, the final deliverables of the project will be:

1. A link to your Notebook on your Github repository, showing your code. (**15 marks**)
2. A full report consisting of all of the following components (**15 marks**):

* Introduction where you discuss the business problem and who would be interested in this project.
* Data where you describe the data that will be used to solve the problem and the source of the data.
* Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.
* Results section where you discuss the results.
* Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.
* Conclusion section where you conclude the report.

3. Your choice of a presentation or blogpost. (**10 marks**)