The Battle of Neighbourhoods

Helsinki (Finland) vs. Tallinn (Estonia)

Prepared by: Norbert Juhász

For: Applied Data Science Capstone by IBM/Coursera

Version date (Week 1): May 25th, 2020

1. Introduction/Business Problem

In this final course assignment, I will compare neighborhoods of two cities:

- Helsinki, Finland (where I live) and
- Tallinn, Estonia (which is the closest other EU capital, just on the other side of the Gulf of Finland)

As a fact to raise the interest, there are even plans to connect the two cities by an underwater tunnel in the future. For more about this, please see https://en.wikipedia.org/wiki/Helsinki%E2%80%93Tallinn_Tunnel

Understanding the similarities of neighbourhoods between the two cities may help people who consider moving from one city to the other. For someone who, for instance, would be looking for a similar neighborhood to live in Tallinn as compared to where they live in Helsinki, the results of my analysis may be helpful.

In this project assignment, the neighborhoods of the two cities will be clustered, analyzed and described.

2. Data

Based on definition, my primary interest is profiling the neighourhoods of Helsinki and Tallinn with the purpose to cluster them based on similarities.

The following data sources are used:

- neighborhoods of Helsinki and Tallinn are defined based on postal codes from **Opendatasoft.com**'s API
  - dataset name: "geonames-postal-code%40public-us"
  - data of interest: postal code, area name, latitude, longitude, administrative structure
  - **Helsinki** is defined as Greater Helsinki area, including the administrative (but not geographically) separate cities of Espoo, Kauniainen and Vantaa
  - **Tallinn** city is also extended by the region of Viimsi vald which encompasses Tallinn towards the direction of Helsinki
- the central coordinates of Helsinki and Tallinn are retrieved from **geocoders Nominatim**
  - in order to center the folium map and make both cities are visible, the coordinates of the two cities are averaged
- the venues, their type and location in every neighborhood will be obtained from the **Foursquare API** (as required by the assignment specifications)

3. Methodology

TBA for Week 2

"... 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."

4. Results

TBA for Week 2

"... where you discuss the results."

5. Discussion

TBA for Week 2

"... where you discuss any observations you noted and any recommendations you can make based on the results."

6. Conclusion

TBA for Week 2

"...where you conclude the report."