Evaluating Rental Prices in Utrecht City

1. Introduction

1.1. Background

Nowadays rental houses is a business continuously growing in most of Europe, with a higher rate in cities where expats' market is predominant, thanks to presence of mainstream Universities and High Tech hubs.
Utrecht, the fourth largest city of the Netherlands located in the very center of the mainland, surely belongs to this category.
For this reason, with the purpose of getting - all expats and/or any person going to move - acquainted with the current scenario of rentals in the area, a well detailed analysis is carried out.

1.2. Purpose

This project's purpose is to give a clear overview of rental prices in Utrecht area and evaluate mutual influence of multiple variables such as neighborhoods, house surface and typology.

2. Data

2.1. Data source

Most dataset are quite recent and have been recovered from Kaggler here. Also, Foursquare API is used to get information related to the neighborhoods of the houses listed in the csv.

2.2. Data Wrangling

Data will be reported as dataframe from Jupyter Notebooks and then manipulated in order to get expected results.
Firstly, a plot of Utrecht map with Folium of dataset *Utrecht_postcode_v1* is performed in order to have a clear perspective of Utrecht geospatial coordinates and its districts' distribution.
Then, with reference to Utrecht central districts (Binnenstad, Noordoost, Zuidwest, Zuid, Noordoost) **the** dependent variable Rental Price is evaluated. Way of working adopted is Machine Learning method typically used for continuous values (i.e. **Multiple Linear Regression**), rental price vs following independent variables xi:

- Influence of neighborhoods
    - the foursquare API is applied to the dataset *rental_central.csv*;
    - venues frequency is plotted with a bar chart
    - given the amount of data (more than 5000 entries) the 10% of the entire dataset having less impact is not considered. For the remaining, macro groups context based are generated
    - quantity of the groups related to each house will be compared with the price

- Size of apartment

- Type of apartment
    - Apartment indexed 0
    - House indexed 1
    - Studio indexed 2
    - Villa indexed 3

- Furnished
    - Furnished indexed 0
    - Upholstered indexed 1
    - Shell indexed 2

After the effect of each variable is evaluated, the xi dataset will be splitted in training and testing data, the xi variable will be inserted in a Pipeline and the fitting model completely developed/evaluated.

3. Exploratory Data Analysis

TBD

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

TBD

where you discuss the results.

5. Discussion

TBD

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

6. Conclusion

TBD

where you conclude the report.