Public Capstone Project Summary

Statistical Analysis on Data Collection on the Status of Existing Buildings in the City of Innsbruck

CEU Business Analytics MSc
Péter Szilvási
2024 June

Project Outline

The project was part of a larger initiative to develop a strategy to carbon neutralize Innsbruck's heat supply network by 2040. As the first step of the analysis, researchers must understand the outlook of the city building's characteristics focusing on various parameters such as construction periods, building types, renovation need, and others that might influence the buildings' energy efficiency levels. My project utilized data, which was collected via an expert inventory assessment by the contracted companies, with the goal of providing insights into the current state of the building stock and to identify potential areas for improvement.

Work Done

Data Collection and Validation: I was given both data sources, one from the city of Innsbruck containing data about the real number of buildings aggregated on city, district, and city district level along with real construction periods, and information about residential and non-residential buildings. The second dataset was the one collected by researchers including parameters mentioned in the project outline part. The initial goal was to do data processing on the collected dataset, as for each entry the number of representing entrances was missing and that data came from a separate dataset. I was asked to first join the representing entrances with their corresponding entries which proved more problematic than initially thought, as the coordinates extracted from the google maps links were not unique. So, to begin the project, we had to do a separate analysis on why that column is not unique and find another way to merge the collected data. After solving this problem, I validated the collected data by comparing the construction periods and district level data to the ones collected.

Data Analysis: After, the clean collected sample was analyzed to identify trends and patterns in the building stock. This mainly included descriptive statistics on the data collected, such as the distribution of representing buildings by districts, share of construction periods and building types, along with examining the relationship between renovation need and other parameters.

For this I utilized Python's ggplot with custom colors and a custom theme to make the plots coherent.

Map Creation: Using Python's Folium library, I also created an interactive map of buildings in Innsbruck. The map had multiple layers with the user's ability to change the background, or the features presented. These features included construction period, renovation needs, and thermal insulation statuses. With the help of this map, researchers can identify which buildings, or areas need further attention, whether it comes to gathering additional data or picking buildings for renovation. Upon clicking on buildings, pop-ups also help providing additional information such as district, city district or the exact building identifiers. This map is now exported to HTML format and is best used in a web browser.

Regression Analysis: Logistic regression was performed to identify significant predictors of renovation need, including construction period, thermal insulation, window type, and other elements. The regression proved that we only have evidence to say that the presence of thermal insulation greatly reduces the need for urgent renovation.

Benefits to the client

My study provides valuable insights into the building stock in Innsbruck, which can be used by the client to:

Identifying Data Processing Issues: Prior to the analysis, I was given some steps to follow which turned out not to be the right way, and with the cooperation of my sponsor we had to find another way of proceeding to get the right data without any biases.

Validation of Speculated Ideas: Going into the project researchers already had some speculations and expectations about the results, like how older buildings will have a higher percentage of renovation need and how the Innenstadt (old town part of Innsbruck) will be the one needing further attention. It turned out that there might have been a previous wave of renovations as the oldest period was not the most problematic. Innenstadt also turned out to be relatively needing low attention and others previously not suspected being the ones with renovation needs.

Prioritize renovation efforts: Focus on buildings that require urgent renovation, with older buildings lacking thermal insulation and worse window types.

Target Specific Districts: Identified renovation efforts towards districts with higher needs. I also paid attention here to where we have a good proportion of coverage compared to the actual district numbers.

Key Outcomes

Construction Periods: According to the data collected the majority of the buildings in Innsbruck were constructed before 1945 with a significant proportion of buildings from the period between 1961-1980.

Renovation Need: Up to 5% of buildings need urgent renovation in the next 10 years, while 16.4% do not need renovation in the next 20 years. This leaves the rest of the city, around 78%, needing renovation in the next 20 years.

Thermal Insulation: High need of renovation buildings do not have thermal insulation in 75.5% of cases, while in the best category in terms of renovation need, 86.8% of buildings have thermal insulation.

Districts: Pradl and Wilten are the main districts with a higher need for renovation, while Innenstadt and Saggen have more older buildings.

Learning Experience & Lessons Learned

I really enjoyed working on my capstone as this was my first real-life data analytics project. As for technical aspects, I think I was already proficient in using Python, but this further proved my adequacy. Active coding while sharing my screen to my project sponsor really boosted my confidence. I think I also managed to create eye-pleasing plots with attention to detail and give insights into my findings. I have yet to present my findings to the whole company, but I am looking forward to it. Other than these, my project really highlighted the importance of:

Data Quality: Ensuring that the data is accurate and consistent is crucial for reliable analysis. I had a lot of categories which were messy (‘Innenstadt oder Mühlau’).

Data Cleaning: Cleaning and preprocessing data to remove these inconsistencies and handling not unique IDs was also something we did not have in school and proved to be really useful lessons.

Skewed Samples: The collected data had a higher proportion of buildings from the period before 1945 compared to the actual distribution of construction periods in the city. Accounting for these biases is also something I had to learn to cope with when creating my analysis.