PROPOSAL

Rotterdam Accommodation Price Predictor (RAPP)

Prepared by Pham Nguyen An Phuong

Date: 21/04/2021

Contact: 426122@student.fontys.nl

Abstract

RAPP is a project inspired by Vietnamese students and the process of finding comfortable accommodations in the Netherlands, specifically Rotterdam. The aim is to provide this stakeholder group with a tool that helps them in viewing and balancing their standard of "comfortability" and personal budgets. To do this, rental price prediction will be carried out using techniques such as Random Forest or Regression. Potential features include surface area, number of rooms, and distant to a selected hub for commutation and leisure. Further provisioning could add trendiness of an area based on restaurants, ratings, and pricing of service. Main external stakeholders for the project includes Vietnamese students, general community, real estate agencies, data provider, and law experts in Rotterdam. The analysis of impacts has shown that there is no major issue that requires the re-consideration of RAPP. Hence, the project will be carried out according to the AI methodology, with the application of DOT framework for research, and will be finished within a timeframe of 3 weeks.

Table of Contents

1. PROJECT STATEMENT

This is the proposal document that provides contextual insights, impact analysis, and data overview for the project Rotterdam Accommodation Price Predictor (RAPP). For better comprehension of the initially collected data, it is recommendable to refer to the Exploratory Data Analysis (EDA) section in the attached 'EDA and Provisioning' notebook.

1.1. Context

Looking at the data provided by 'Worldometers' – a global statistics tracking website, the Netherlands is currently the seventh country for highest population density in Europe (*European Countries by Population (2021)*, n.d.). Furthermore, the housing shortage problem in the country is getting more serious and putting major cities such as Amsterdam, Rotterdam, Utrecht, and The Hague under great pressure (Lalor, 2021). These factors have made it difficult to find 'comfortable' housing for Vietnamese expats in the Netherland, especially apposite to the student community. In this context, the term 'comfortable' refers to reasonable price, and distant to city center or hub for commutation and leisure.

For further research into the problem, the project developer has contacted the association of Vietnamese Student in the Netherlands (VSNL). As a result, the following estimations could be presented.
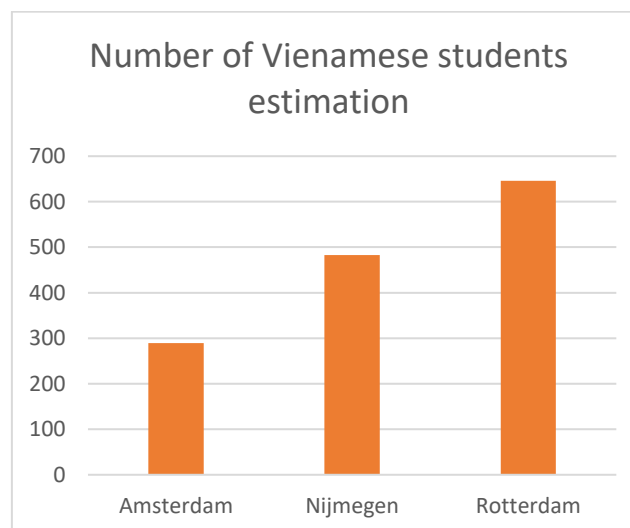
The graph above depicts the estimation of number of Vietnamese students in Amsterdam, Nijmegen, and Rotterdam – the cities with high concentration of Vietnamese in the association's record. The contact points have also shared that complaints regarding "accommodation hunting" and housing is regularly received.

With these factors in mind, RAPP's developer will address the issue by developing a product to predict rental price based on infrastructure facts (e.g. number of rooms, area) and comfort-related attributes (e.g. distant to city center). Via the solution, users will be able to find a balance between their comfort standard and personal allowance. Since Rotterdam has the highest number of Vietnamese students and is one of the major cities mentioned above, application will begin there.

1.2. Goals and deliverables

RAPP is a machine learning project aimed at predicting rental price of accommodations in Rotterdam using infrastructure and comfort-related attributes. Initial features include surface area, number of rooms, relative distance to a commutation/leisure hub (to be selected) in the center. They will be used to predict the target variable – price. This is the rental price in euros of accommodations per month. Further provisioning may add more features, and modelling techniques could encompass Random Forest or Regression. Further research will be carried out into this matter. The deliverables include:

- Proposal
- Scraper
- EDA and Provisioning notebook
- Modelling notebook
- Data requirement and collection ledgers
- Process report and demonstration material

* The deployment has yet to be finalized. Initial ideas include an interactive dashboard, or a prototype via Heroku. However, this may be subjected to changes and is not within the scope of this current project.

1.3. Stakeholders

Below are the main external stakeholders to be considered:

Stakeholder	Description
Vietnamese students in Rotterdam	These stakeholders are the main target group and inspirator for the development of RAPP. The developer hopes that the solution can help them in estimating rental price based on desired inputs (e.g. desired living area). They have moderate influence on the development process, but high interest.
General community in Rotterdam (Vietnamese and other professionals, and students of other nationality)	Although they are not the main target group, the solution is inclusive for all. Similar functions can be performed, with interest and influence being alike to the previous group.
Rental agencies and landlord in Rotterdam	By applying the solution, potential tenants will have already realized the suitable price range for what this stakeholder group offer. Hence the mitigation of bias complaints and waiting time for potential tenants to make a decision. Furthermore, this group may use the solution to check their own offered prices. They have low interest and low influence.
Data providers	Data is needed to carry out this project. Hence, providers (passive or not) are very important. The current provider for the pre-

	anchored data is Pararius. This group have low interest and high influence.
Real estate regulation experts	Beside complying with data regulations, no real estate law should be broken. Hence, consultation with experts in the field may be needed at some point of the project. They have low interest and possibly high influence.

2. IMPACT ANALYSIS

2.1. Societal

When a person is new or inexperienced with the real estate market in Rotterdam, it can be difficult to find a reasonably priced house that is 'comfortable'. This is due to population density, property shortage, and partly not knowing what to expect for a certain price range. RAPP aims to help users form an idea of what price they should expect with regard to desired housing conditions, or an actual property. By doing this, the user can balance their budget with the most suitable level of comfort.

The solution will use machine learning models to predict rental price based on features such as number of rooms, area, accommodation type, distant to city center, etc. So far, no negative effect can be evaluated for this category. Once completed, the product will contribute to making accommodation finding easier, especially for Vietnamese students in Rotterdam – the target group.

2.2. Hateful and criminal actors

Since no communication or transaction is performed when using the product, there should be no opportunity to perform thieving or scamming for valuable goods. As this is only a price predictor and not an e-commerce platform, there has yet to be any serious threats identified in this category.

2.3. Privacy

There should be no privacy issue as the indicator does not save personal information of an user. Even if the inputs are temporarily stored in RAM to proceed with model's code, user's identity will not be saved under any method or form, nor is it required when interacting with the product.

2.4. Human values

RAPP aims to imbue users with necessary insights to make their own decision and find the best balance in their own view. There are no particular issues to be mentioned within this category.

2.5. Stakeholders

- Developer: as the internal stakeholder, the developer is responsible for RAPP and any caused effects. The product will also be an opportunity for the internal stakeholder to develop in terms of professional skills.
- Vietnamese students in Rotterdam: they are the main targeted users according to the scope of this project. Possible impact and their position of the interest – influence scale has been mentioned in section 1.3.

- Other external stakeholder groups: description of the relationship between RAPP and other external stakeholder groups have also been mentioned in 1.3.

2.6. Data

The collection and application of data within this project is straightforward, independent of personal issues of any party, and simple. Hence, no pitfall has been identified as of yet. To keep a high standard of relevance, it is recommendable to re-scrape and train the predictor often to accommodate changes in the market. This will be the responsibility of the moderating/hosting party if the product passes the MVP stage.

2.7. Inclusivity

According to the scope, the data in use is specific to Rotterdam and does not include social housing as it is based on the renting process via real estate/landlord representative agencies in the city. Besides this, the main target user is Vietnamese students in Rotterdam. However, the potential users can come from a wider scope as mentioned in the 1.3 section.

2.8. Transparency

All working steps and reasoning will be recorded and presented in the clearest manner possible for readers and future users' comprehension. Any complaints or questions can be filed to the contact provided in the front page of this document.

2.9. Environmental sustainability

Since the data is not of great size, sustainable choice of storage solution is the local system. However, this is not a significant issue. No further subject is applicable in this category as of yet.

2.10. Future impact

If successfully deployed, the solution will be a new tool to help potential users, especially the Vietnamese student community in Rotterdam to find suitable and comfortable housing more efficiently. As mentioned before, the relevance of this solution will be kept by periodical update of data. All code and steps will be presented so that the re-setup can be effortless for future hosting/moderating party.

3. DATA OVERVIEW

To carry out RAPP, the first step is to gather data of accommodations in Rotterdam from rental websites. There are many candidates; however, Pararius stands out among them. This website is dedicated to the rental domain and is rated as the third most popular real estate portal for Dutch housing (the top 2 are purchase oriented) with 1.6 million visits monthly (*Top Real Estate Websites in the Netherlands*, 2021). For this task, a web scraper built in python has been developed, it can be found in the Scraper notebook.

After examining and initial wrangling of the scraped data (1279 rows – 7 columns), the following visualizations can be derived for the target variable (i.e. price) and potential features – surface area (m^2) and number of rooms:
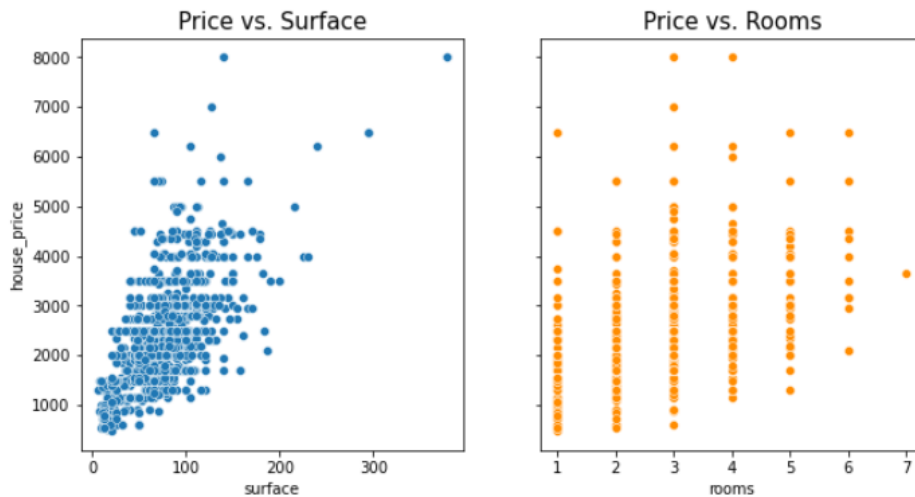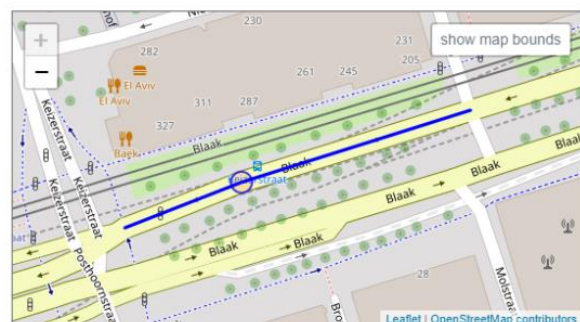
Looking at the first illustration, most of the accommodations in this dataset have area less than or equal to 200m^2. It is observable that the rental price generally increases when surface area of accommodation increases. However, this does not mean that a larger surface will always cost more per month than a smaller one. For example, there are still accommodations of 100m^2 that have higher price than those of 200m^2. The same could be observed for price increase with regards to number of rooms. Some one-room accommodations can cost more than others with 2 or 3 rooms. These insights suggest that there is more to rental price in Rotterdam than just area and number of rooms, and more features are needed.

Because the inspiration for RAPP is to help Vietnamese students in Rotterdam find 'comfortable' housing (taking proximity to commutation/leisure hub into account - as defined in the proposal), distance to Blaak will be a good start for anchoring. This is because Blaak is a notable street in Rotterdam center and the selected point (on Blaak) is next to an area that has many commutation and leisure facilities. The point is as below:
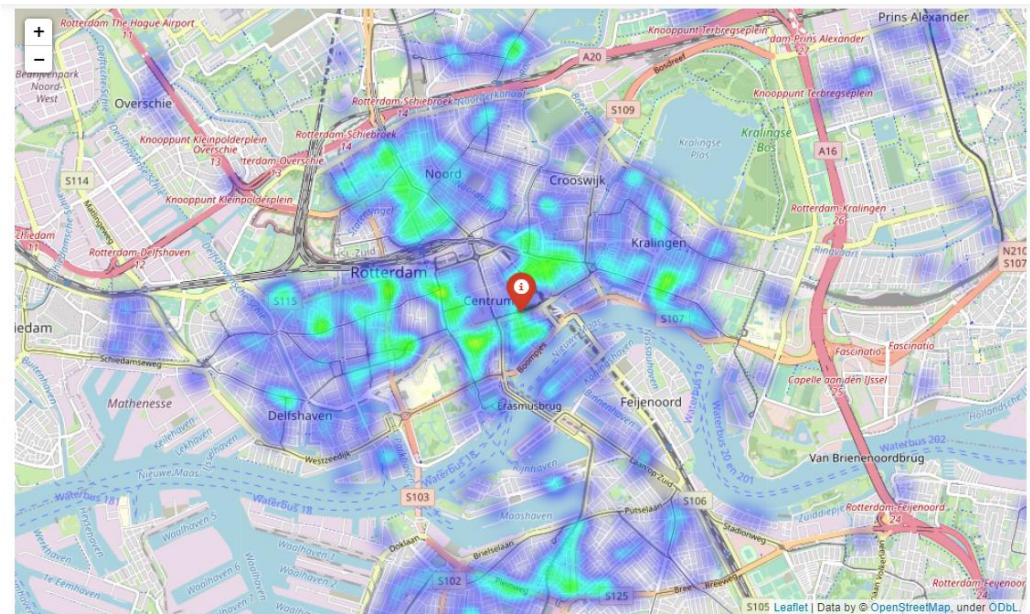

After finding the coordinates for accommodations in the dataset and get the approximate flat distance to the hub's point selected above. The following map with housing distribution could be checked.

The heatmap above depicts the distribution of accommodations in the dataset around Rotterdam. It is noteworthy that the more accommodations appear in one area, the brighter the color tone. The map shows that the majority of accommodations is within the proximity of the city center, with the 'city square' point (red marker) being very close to 3 of the concentrated areas. There are also other concentrated areas outside of the center. Those that are right next to the center have more significant concentration than those further away, with the exception of "Pleinweg" (across Nieuwe Maas - to the South). A scatterplot will now be used to check for relationship between distance (derived above) and price.
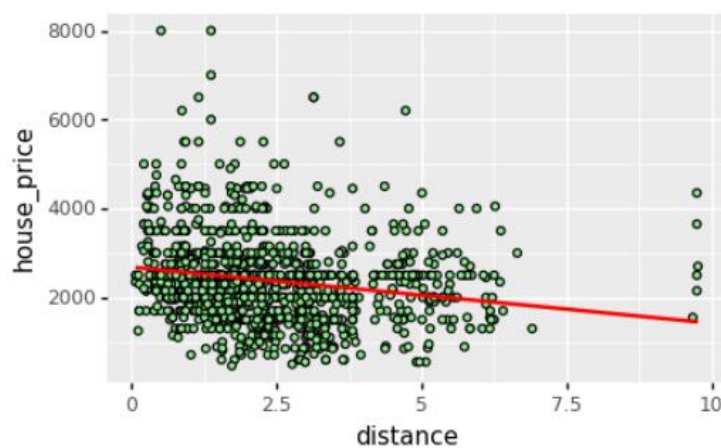

The relationship here is inverse to that of price and area or number of rooms. This is as expected. Overall, the EDA has looked at the initial scraped dataset, anchored distance to a center point on Blaak for each accommodation, and visualized relationships between the target variable (price) and other features. The distribution of the accommodations around Rotterdam has also been looked at. To conclude, there are relationships that could be investigated and used for modelling; before that, there is also the potential for more data anchoring (e.g. trendiness of surrounding area based on restaurants and ratings). More detail can be found in the EDA and Provisioning notebook.

4. METHODOLOGIES

4.1. General approach

RAPP will be carried out using the AI working methodology, which means that there will be four main phases: proposal, provisioning, prediction, delivery.
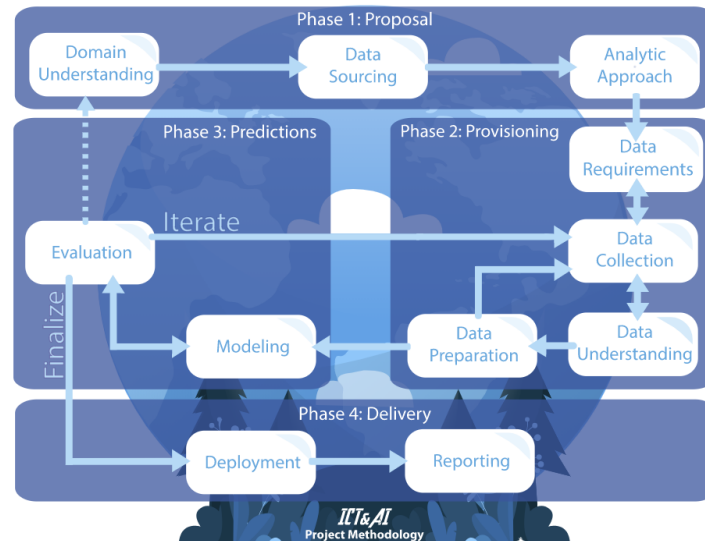
The project is currently in the proposing phase. Following this, the second phase will focus on collecting and setting up the data for necessary predictions. The third phase is aimed at modelling, evaluating, and collecting more data if required. Finally, the fourth phase will deal with the deployment and reporting of the result.

4.2. Research method

RAPP is applying one of the tools accepted by the Netherlands AI Coalition for impact analysis: TICT Tool. As for research and thinking method, the DOT framework will be used.
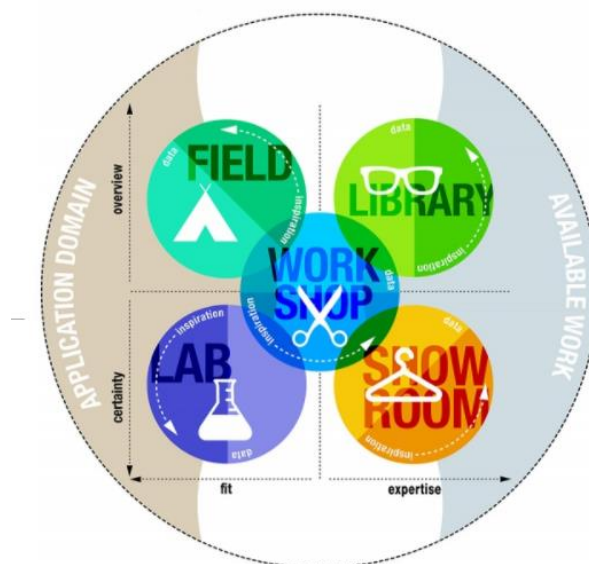
Beside answering the three aspects "what", "where", "how", there are also 5 strategies to use during research. Some major applications are as follow.

Activity	Method	Instrument	When
Understanding the situation	FIELD	Document analysis – comprehend similar projects online Problem analysis – Clearly set out issues to be investigated	Phase 1
Understanding the domain and related fields	FIELD	Domain modelling – based on experience and consultation with domain experts	Phase 1 & 2
Research tool and technology to be used	LIBRARY	Literature study – research and learn to develop scrapers; identify best approach for modelling.	Phase 1
Research available data	LIBRARY	Literature study – study the scraped data and find suitable anchoring	Phase 1 & 2
Designing and evaluating	WORKSHOP	Brainstorm – formulate action plan and deliverables Prototyping – improve MVP of model or product using feedbacks	Phase 1 Phase 3

4.3. Planning

The project time-constraint is 3 weeks, starting from 19/4 with 1-2 days reserved. It is necessary to finish the first three phases of the AI methodology and document the process report. As for deployment material, a suitable format should be selected such that the project can conclude on time. The list of deliverables is mentioned in section 1.2.

5. CONCLUSION

After analyzing the potential impacts of the product, no major concern that could serve as a setback to the direction RAPP is heading in could be found. As for the data, after building a scraper, collecting, and exploring, there appears to be enough relationships and opportunities for further enrichment and investigation. The next step would be to diversify the dataset with the trendiness of surrounding areas, using the potential help of Yelp's developer site (https://www.yelp.com/developers). From there, modelling using techniques such as Random Forest or Regression could be carried out. All in all, the conditions are suitable for the developer to continue with the project and finish within the constraint of 3 weeks.

6. REFERNCES

European Countries by population (2021). (n.d.). Worldometers. https://www.worldometers.info/population/countries-in-europe-by-population/

Top real estate websites in the Netherlands. (2021, April 16). FBW. https://foreignbuyerswatch.com/2020/09/03/top-real-estate-websites-in-the-netherlands/

Lalor, A. (2021, March 15). *Why is there a housing shortage in the Netherlands? The Dutch housing crisis explained*. DutchReview. https://dutchreview.com/expat/housing/why-is-there-a-housing-shortage-in-the-netherlands-the-dutch-housing-crisis-explained/