Evaluation of Property Filtering Algorithms using Tags for a Property Rental Recommendation Application

The Thirteenth International Conference on Data Analytics
DATA ANALYTICS 2024
September 29, 2024 to October 03, 2024 - Venice, Italy

SEAN LYNCH, STEPHEN ANOKYE, OISÍN FITZPATRICK, FENNA KADIR, CHARLOTTE MARTIN, NISHANT KAPUR, BRENDAN TIERNEY, ANDREY CURLEY, DAMIAN GORDON

Contact: Brendan.Tierney@tudublin.ie

Project Team

MSc Project Team

Sean Lynch

Stephen Anokye

Oisín Fitzpatrick

Fenna Kadir

Charlotte Martin

Nishant Kapur

Academic Supervisors

Brendan Tierney

Andrey Curley

Damian Gordon

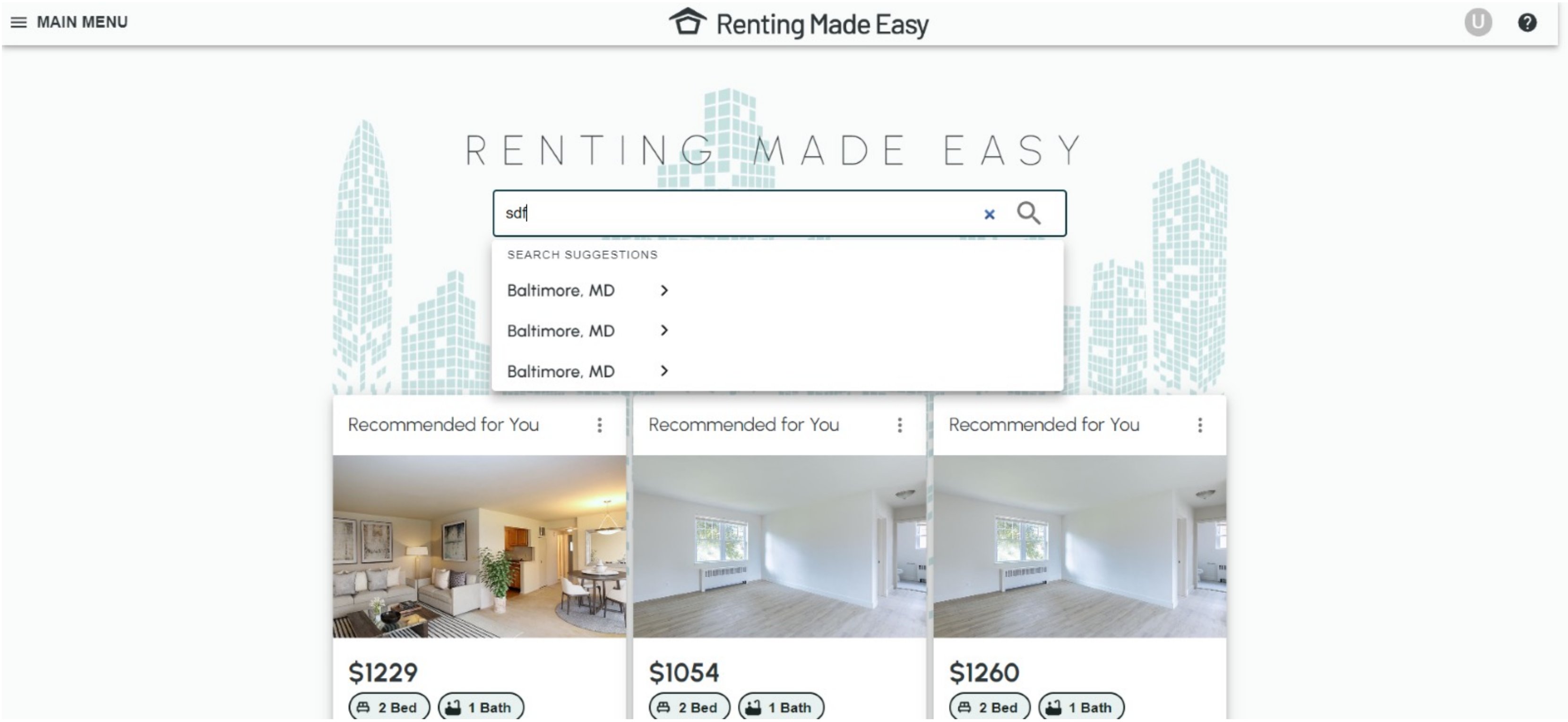
Project Overview

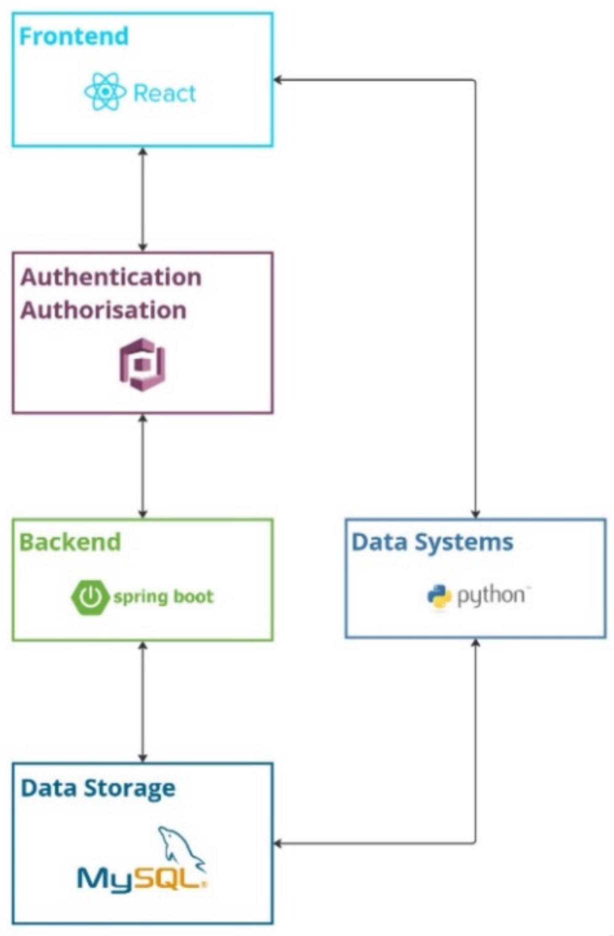
- Renting Made Easy (RME)
- provides a range of functionalities, including a comprehensive search experience, applying and managing rent applications, and property browsing
- Initially based on the Zillow data set, the application has been expanded to include data sets with location-based services and crime-related data
- A core feature is to provide a user-driven feature search based on property Tags.
- Property filtering algorithms were evaluated to determine which provide suitable properties to the end-users.
- These algorithms included k-Nearest Neighbors (kNN) and Collaborative filtering.
- Qualitative research was performed to assess the usefulness and accuracy of the filtering algorithms.

MSc Team Project

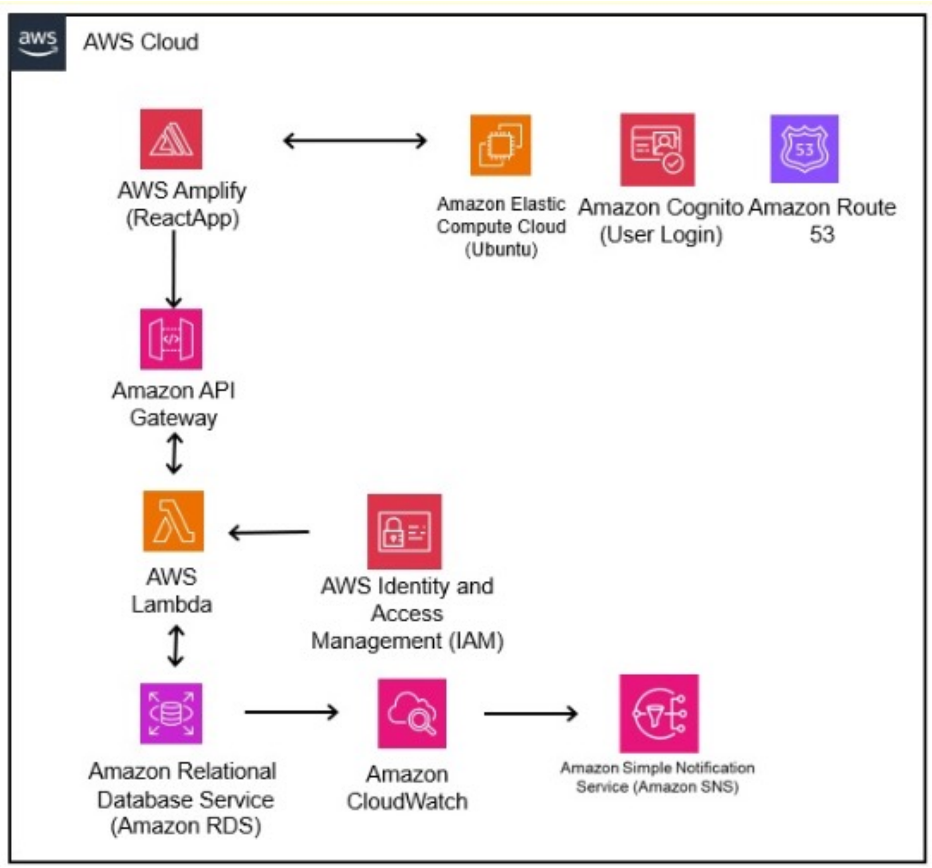
Coding	Front-End	Back-End
Middle-Tier	Load Balancing	Database
Research	Data Science	Software Engineering
Requirements Gathering	Market Testing	Project Management

Project Overview

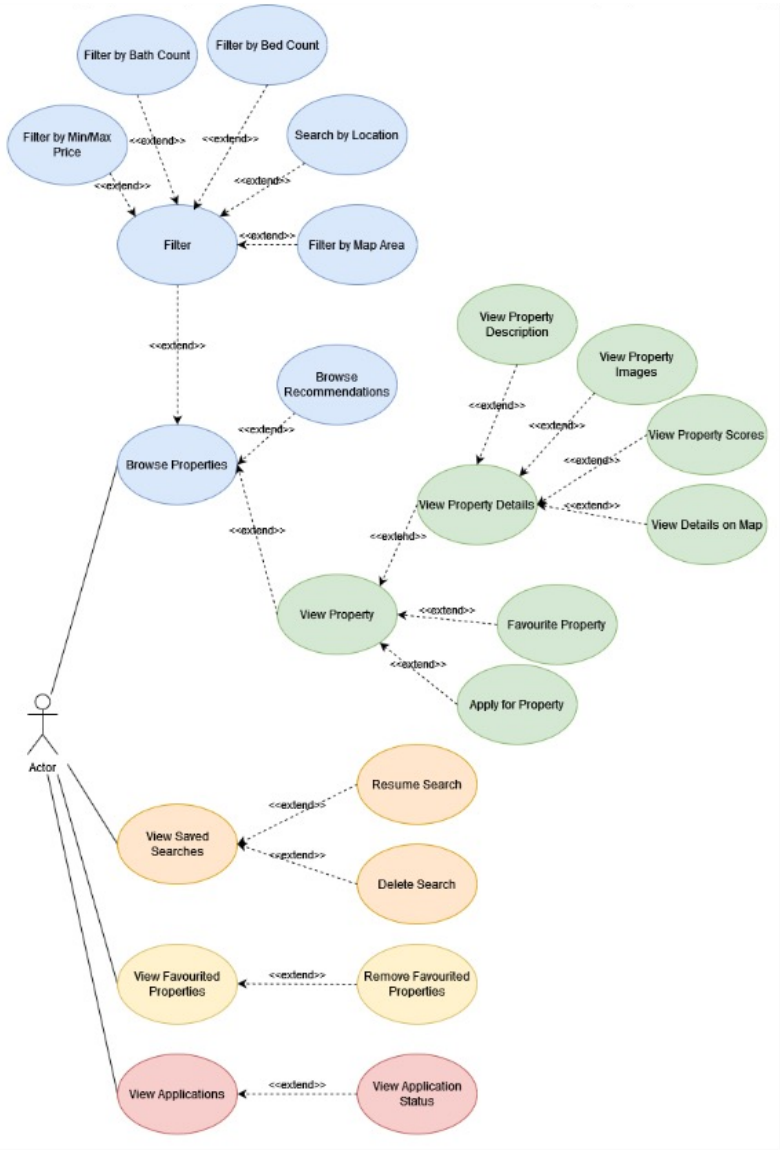




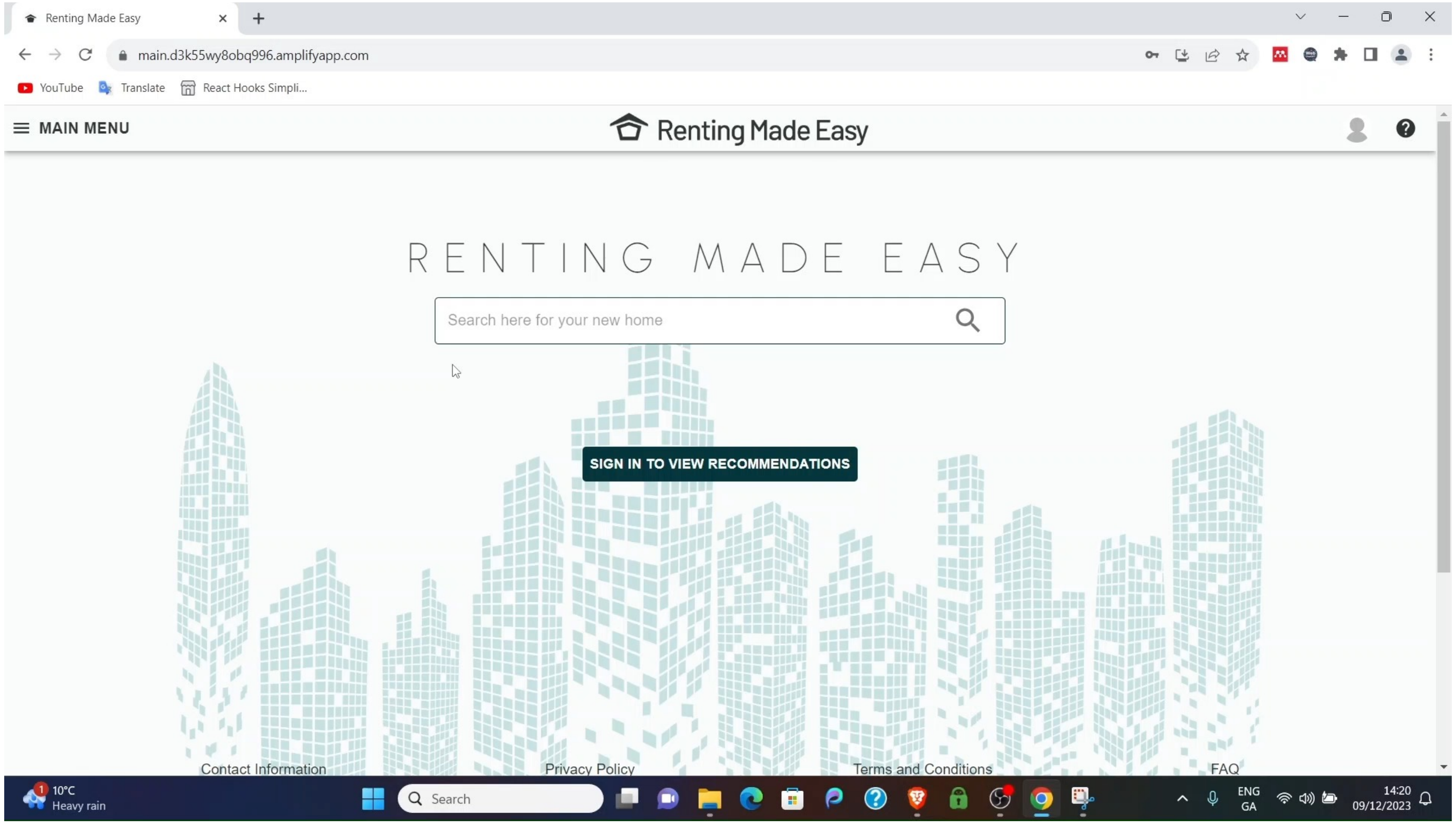
Frontend
React
Authentication
Authorisation
Backend
spring boot
Data Systems
python
Data Storage
MySQL



aws
AWS Cloud
AWS Amplify (ReactApp)
Amazon Elastic Compute Cloud (Ubuntu)
Amazon Cognito (User Login)
Amazon Route 53
Amazon API Gateway
AWS Lambda
AWS Identity and Access Management (IAM)
Amazon Relational Database Service (Amazon RDS)
Amazon CloudWatch
Amazon Simple Notification Service (Amazon SNS)



Filter by Bath Count
Filter by Bed Count
Filter by Min/Max Price
Search by Location
Filter
Filter by Map Area
<<extend>>
Browse Recommendations
Browse Properties
View Property Description
View Property Images
View Property Scores
View Property Details
View Details on Map
View Property
Favourite Property
Apply for Property
Actor
Resume Search
View Saved Searches
Delete Search
View Favourited Properties
Remove Favourited Properties
View Applications
View Application Status

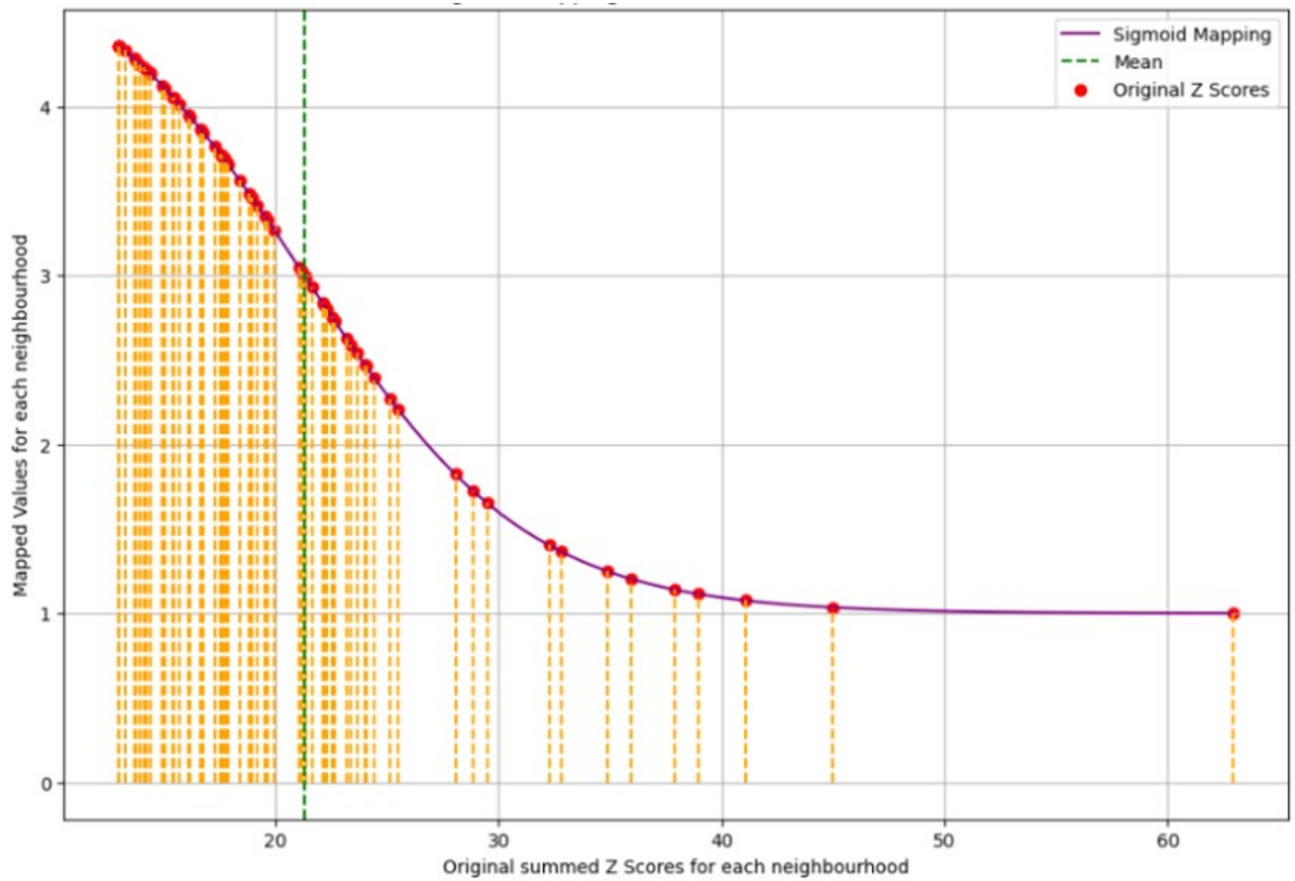


- Demo RME: https://youtu.be/EnSPlGhYU-s

Property Scores

- Property scores were generated for each property and displayed as values out of five to users. These scores provided a general overview of the area of each property. This first score displays the crime safety rating based on the neighbourhood of the property.
- Adjustments or normalization techniques should be applied to balance the impact of different crime categories, allowing for a more accurate and fair representation of crime levels.
- To solve this problem a z-score standardisation was implemented. This generates a different score for each category in each neighbourhood.
- These scores were described in terms of their relationship to the mean, where their values are measured in terms of standard deviations from the mean.
- The z-scores were mapped between one and five using sigmoid transformation.
- A sigmoid function was used over other mapping techniques following experimentation, including Min-max normalization, Winsorized min-max normalization, and Winsorized linear transformation.

Z Score for each Neighorhood



kNN

- A kNN model was incorporated within the system, including the steps for data pretreatment and feature selection.
- The kNN model was trained using the pre-processed and encoded dataset to select attributes closest in similarity based on their features. The training process entailed the following:
  - Instantiating the kNN model using the NearestNeighbors class from the scikit-learn library.
  - Applying the model to the dataset that has been both normalised and encoded.
  - Evaluating the model using a certain attribute and obtaining the closest suggestions.

Collaborative Filter

- The user collaborative filter system only incorporated click data. Where a click data point represents a user clicking on a property card.
- The model took a user as input and created a pattern for them. This pattern was represented as a vector with the number of properties in the database in it as its length and the number of clicks for each as the property's value.
- It took all of the other vectors for the other users in the database and compared the vectors to one another, looking for users with similar interactions.
- This was measured by finding the Cosine of the angles with values closest to zero between each vector.

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}},$$

FIGURE 4. COSINE SIMILARITY FUNCTION

Collaborative Filter

- The ten most similar users to the input user were retrieved and a new data set was made.
- This data set contained all the combined clicks for each property between all the recommended users. The properties were sorted in descending order by click count, and the first ten were displayed to the user. These were the ten most interacted properties between all recommended users.
- Once the Cosine similarity model was deployed to a lambda function, the data was stored temporality in the front-end. When the lambda function is called, the user can observe the recommendations presented to them.
- These iterations were hybrid models that combined geocoordinate data as well as property data.

Evaluation

- The recommendation system for both the kNN and User Collaborative Filtering models were evaluated using user feedback obtained from surveys.
- The recommendation systems gave recommendations based on content and user interactions.
- The recommendation engine's performance is based on user sentiment toward their recommendations.
- The surveys allowed for qualitative feedback that enabled the models to be adjusted based on user preferences. The evaluation of the recommendation system aims to determine:
  - The suitability score for each feature in the recommended properties (based on user feedback).
  - The overall suitability score of the recommended properties (based on user feedback).

Evaluation

- Due to the interconnecting nature of the recommendation and tag systems the evaluation of both was joined into a single questionnaire in conjunction with the scoring system.
- As such, the methodology and evaluation metrics are similar for the property tags and the recommendation system.
- Property tags are designed to increase the usability of RME. They also confirm the recommended properties, for example, the user's favourite properties with 'secure' tags, and therefore secure properties should be recommended to the user.
- The two recommendation models were made available at different intervals with approximately half of the participants testing each model.
- Users did not know which recommendation system they were testing. The evaluation aimed to explore the tag contribution to the recommended properties, testing kNN and Cosine recommendation models.

Evaluation

- Kendall’s Tau which is used to measure the correlation coefficient between each of the suitability ratings of the kNN features and the overall suitability rating of the kNN model.
- The decision to measure the correlation between participant kNN feature suitability and participant kNN overall suitability scores was based on the actual values of the variables without assuming any specific underlying distribution.

TABLE 1. KENDALL’S TAU CORRELATION COEFFICIENT OF KNN FEATURES

Feature	P-value	Kendall's Tau	Correlation Strength	Failed to reject:
Available amenities	p = 0.191	r = 0.29	Weak correlation	H (0)
Nearby personal care	p = 0.069	r = 0.41	Moderate correlation	H (0)
Nearby banks	p = 0.049	r = 0.46	Moderation correlation	H (1)
Price	p = 0.007	r = 0.60	Strong correlation	H (1)
Nearby emergency services	p = 0.028	r = 0.50	Moderate correlation	H (1)
Nearby public transportation	p = 0.208	r = 0.29	Weak correlation	H (0)
Nearby leisure activities	p = 0.009	r = 0.59	Moderate correlation	H (1)
Nearby retail	p = 0.014	r = 0.56	Moderate correlation	H (1)
Nearby gyms	p = 0.060	r = 0.44	Moderate correlation	H (0)
Area safety	p = 0.210	r = 0.28	Weak correlation	H (0)
Number of bathrooms	p = 0.042	r = 0.46	Moderate correlation	H (1)
Number of bedrooms	p = 0.267	r = 0.25	Weak correlation	H (0)

Conclusion

- The RME application was built to provide a better user experience for the end-user.
- A key component of this application was the recommendation system which included a combination of a Property Scoring System, Property Tag selection and filtering algorithms.
- The filtering algorithms evaluated included kNN recommendation system and User Collaborative Filtering using Cosine Similarity.
- These different approaches were evaluated using subject market experts. These were selected from various industry-related roles including property rental agents, estate agents, property owners and renters in the 20-35 year old range.
- The outcomes from the initial testing demonstrated positive outcomes and feedback from the end-users with particular feedback pointing to the usefulness of the Tagging system and the inclusion of their preferences.
- The evaluation also revealed a relationship between the input features and the overall suitability score that was consistent but not constant.
- Future work will further develop the recommendation engines. This will involve expanding the dataset by incorporating data from different cities or geographic regions and increasing the number of subject matter experts to ensure a broader and more comprehensive analysis.

Q&A

If you have any Questions about the paper or MSc Team Project, you can email them to

Brendan.Tierney@tudublin.ie