The Identification of The Breach of Short-term Rental Regulations in Irish Rent Pressure Zones

Guowen Liu

A Dissertation

Presented to the University of Dublin, Trinity College

in partial fulfilment of the requirements for the degree of

Master of Science in Computer Science (Intelligent Systems)

Supervisor: Inmaculada Arnedillo-Sánchez

August 2022

Declaration

I, the undersigned, declare that this work has not previously been submitted as an exercise for a degree at this, or any other University, and that unless otherwise stated, is my own work.

_____

Guowen Liu

August 19, 2022

Permission to Lend and/or Copy

I, the undersigned, agree that Trinity College Library may lend or copy this thesis upon request.

_____

Guowen Liu

August 19, 2022

The Identification of The Breach of Short-term Rental Regulations in Irish Rent Pressure Zones

Guowen Liu, Master of Science in Computer Science

University of Dublin, Trinity College, 2022

Supervisor: Inmaculada Arnedillo-Sánchez

The housing crisis in Ireland has grown in recent years. In order to make a greater profit, many landlords are no longer renting out their homes under long term tenancies, but under short-term tenancies. The shift from long-term to short-term rentals has had a negative impact on the supply of private housing rentals. How to regulate those rentals that are in rent pressure zones with the highest and rising rents is becoming a tricky issue.

In this paper, we develop a breach identifier to check short-term rentals located in rent pressure zones with potential breaches using only publicly available data from Airbnb. First, we use the residual neural network to filter out outdoor photos which have an impact on identifying whether an owner has multiple rentals located in a rent pressure zone. Second, we use the Siamese neural network to compare the similarity of indoor photos to determine if multiple rental posts correspond to the same residence. Next, the Haversine algorithm is used to locate short-term rentals within a circle centered on the coordinate of a permit. Short-term rentals with a permit will not be restricted. Finally, we enhance the occupancy estimation model combined with sentiment analysis, which may provide higher accuracy.

Because Airbnb does not disclose accurate house coordinates and occupancy data, it is not possible to verify the accuracy of breach identifier. The accuracy of occupancy estimator cannot be verified either. It only provides an estimate within a reasonable range. Users should be somewhat skeptical of short-term rentals that are flagged as possible breaches.

Acknowledgments

I would like to express my sincere gratitude to my respected supervisor, Dr. Inmaculada Arnedillo-Sánchez, for her invaluable guidance and assistance throughout my thesis, for the professionalism and responsibility I felt at every meeting during the thesis, and to my family for their financial support and encouragement during my time at Trinity College.

GUOWEN LIU

University of Dublin, Trinity College
August 2022

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The housing crisis in Ireland has grown in recent years. In order to make a greater profit, many landlords are no longer renting out their homes under long term tenancies, but under short-term tenancies. Under Irish regulations, short-term tenancies are defined as those for less than 14 days. According to Airdna [AirDNA] and Residential Tenancies Board [26], the average monthly revenue from short-term rentals is greater than the revenue from long-term rentals as Table. 1.1, which has resulted in the number of available long-term rentals continues to decline, while lucrative short-term rentals are becoming more popular.

Table 1.1: The average monthly rents in the first quarter of 2022

Area	Long-term Average Monthly Rent	Short-term Average Monthly Rent
Dublin	€2183	€2622
Galway	€1413	€2874
Cork	€1453	€2097

The shift from long-term to short-term rentals has had a negative impact on the supply of private housing rentals. In the midst of the housing crisis, the number of houses used for tourism is higher than the number of houses used for private rentals in every county except Dublin. According to Inside Airbnb [Cox et al.], the number of rental properties on Airbnb[1] in the Irish capital alone has increased from around 1700 in 2016 to more than 7064 in 2022, and 97.2% of which are short-term rentals. This situation is even more serious in those popular tourist counties.

As a ripple effect, rental prices have increased significantly. According to Daft.ie [8],

---

[1]https://www.airbnb.ie

the rental market in Ireland has hit new highs as the number of available homes has fallen to historically low levels. Rents in the second quarter of 2022 rose by an average of 12.6% compared to the same period in 2021. This is the largest increase since Daft.ie began reporting in 2006 and an all-time low in the number of homes available for rent, with the scarcity of available homes in Ireland now unprecedented. The supply of rental properties has fallen by a staggering 97% compared to 2009. Only 716 homes were available for rent nationwide on August 1, down from nearly 2500 a year ago and a record low since 2006.

1.1   Motivation

"We need to crack down on the illegal short-term letting of rental homes to ensure these homes are returned to the rental system." Sinn Féin housing spokesperson Eoin Ó Broin said [11]. The government has approved measures to tighten regulation of the short-term rental market. Under the new regulations, short-term rental platforms such as Airbnb and landlords will be fined if they advertise non-compliant properties from 1 September 2022 [Citizens Information]. The regulations target areas where the accommodation shortage is most acute, requiring Airbnb to obtain planning permission before advertising on their platforms in rent pressure zones (see Section. 2.1).

During 2021, there were 286 short-term rentals advertised on Airbnb in Cork. In the same year Cork Council launched 94 investigations and issued a total of 68 warning letters. But there are still 11 counties that have not taken any action against these short-term rental owners on the popular website to check if they are in violation, making it a difficult issue to regulate short-term rentals in rent pressure zones [12]. The Office of the Planning Regulator has previously stated that how to regulate short-term rentals that are in rent pressure zones is becoming a tricky issue. As of August 2022 the Irish government has not yet provided an efficient way to investigate it, which motivated me to write this paper.

1.2   Objectives

This paper demonstrates how to identify short-term rentals that are located in rent pressure zones with potential violations using only publicly available data from Airbnb.

1. Using the residual neural network to identify and filter out outdoor landscape photos. These noisy photos have an impact on identifying whether an owner has multiple short-term rentals located in a rent pressure zone.

2. Using the Siamese neural network to compare the similarity of images to determine if the interior regions of any two houses are highly similar. There are cases in Airbnb

where the same home is posted multiple times or different rooms in the same home are posted multiple times.

3. Using the Haversine algorithm to locate short-term rentals within a circle with a radius of 150 meters, centered on the coordinate of a permit. Short-term rentals with a permit will not be restricted.

4. Estimating the occupancy of a house by its user reviews, combined with sentiment analysis.

Chapter 2

Background

2.1 The Rent Pressure Zone

A rent pressure zone (RPZ) is a designated area in Ireland where rents cannot exceed the general inflation rate, as recorded by the Harmonised Index of Consumer Prices, or 2% per annum pro rata. This applies to both new and existing tenancies (unless an exemption is being applied for). Rent pressure zones are located in the parts of the country with the highest and rising rents, and where households have the most difficulty finding affordable accommodation [17]. They are designed to moderate rent increases in these areas and create a stable and sustainable rental market, allowing owners and tenants to plan financially for their future.

For areas to be designated as rent pressure zones, the following criteria are used:

- The standardised average rent for the previous quarter must be higher than the appropriate reference standardised average rent for that quarter.

- Annual rental inflation in the area must have been 7% or higher in four of the last six quarters.

Ireland now has a total of three different standardised average rents and the assessment of which method is used is determined by the location of the area. Dublin is now compared to the national standardised average rent. Kildare, Meath, Wicklow are now compared to the national standardised rent excluding Dublin. The other areas are now compared to the standardised average rent outside of the Greater Dublin Area. As of July 2022, there are currently 48 areas set as rent pressure zones in six local authorities [16], these are:

- Dublin City Council

- South Dublin County Council

- Cork City Council

- Rathdown County Council

- Fingal County Council

- Kildare County Council

2.1.1 The Planning Permission

According to [9], if an owner is renting out an entire house or one of its rooms on a short-term basis, then they may need to apply for planning permission from the local authority to use it for short term rentals. The requirement for planning permission only applies to owners in rent pressure areas. Once planning permission is granted, they can use the entire principal residence for short-term rentals of more than 90 days, or rent out a second house for short-term rentals when they are not at home.

2.1.2 The New Regulation of Short-term Rentals

The following rules are the new regulation of short-term rentals [9]:

- Short-term letting is defined as the letting of a house or apartment, or part of a house or apartment, for any period not exceeding 14 days.

- The letting of one or more rooms in the owner's principal residence in a private home will be allowed without restriction.

- Home sharing will continue to be permissible on an unrestricted basis and be exempted from the new planning requirements. Where the 90 day threshold is exceeded, change of use planning permission will be required.

- If the house or flat is not a principal residence, the 90-day exemption does not apply and planning permission is required.

In the following cases, planning permission is not required [Housing Authority]:

- The property is not in a rent pressure zone.

- The property is in a rent pressure zone, but the rental period is 15 days or more at a time.

- The property already has planning permission to be used for tourism or short-term rental purposes.

- The property is used for corporate or executive rentals.

- The property is rented out under the rent-a-room scheme.

- One or more rooms are rented out in the owner's principal residence on a family share basis for short periods of time, the owner must also occupy the home at the same time.

- The entire principal private residence is rented out for short-term visitors for less than 90 days a year while the owner is temporarily away. The 90 days do not have to be consecutive.

- The property is purpose-built student accommodation.

2.2　The Residual Neural Network

Ever since deep neural networks made a splash on image recognition, the ones that have emerged since have moved toward deeper and deeper network layers. Intuitively, when increasing the depth of the network, the network can perform more complex feature extraction, so deeper models can achieve better results.

However, this was not the case, and it was found that the accuracy of the model did not always improve as the depth of the network increased. This problem was clearly not caused by overfitting. Not only did the test error become higher when the network was deepened, but its training error surprisingly became higher as well. This is because deeper networks are associated with the vanishing gradient problem, which hinders the convergence of the network. The phenomenon of deeper networks with reduced performance is called the degradation problem [28].

ResNet was proposed in 2015 and won first place in the ImageNet competition classification task [19]. Because it is both simple and practical, many methods have since been built on ResNet-50 or ResNet-101. It has been widely used in detection, segmentation and recognition.

2.3　The Siamese Neural Network

The Siamese neural network is a special type of neural network architecture. Unlike a model that learns to classify its inputs, a Siamese neural network learns how to differentiate between two inputs by learning their similarity [5]. It contains two or more identical sub-networks which have the same configuration i.e. the same parameters and weights. Parameter updates are performed jointly on both sub-networks as Fig. 2.1.
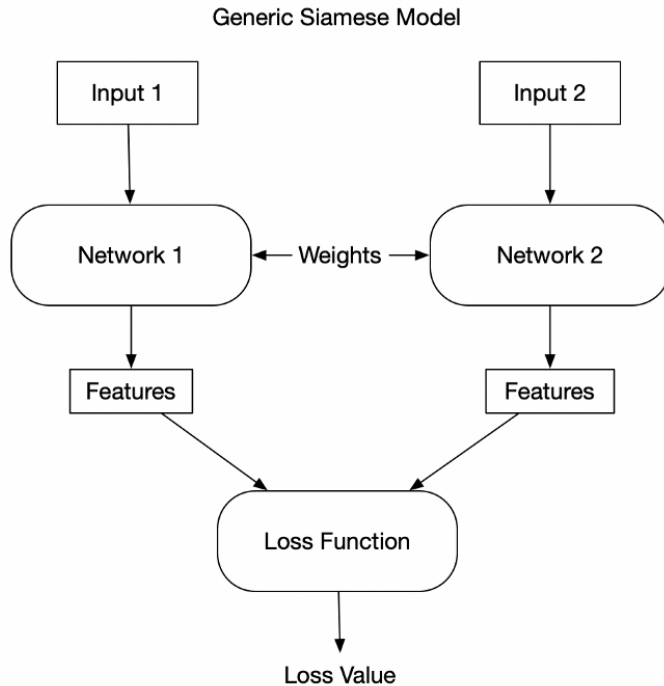
Figure 2.1: The Siamese neural network

Siamese neural networks show better performance in situations involving the discovery of similarity or relationship between two comparable things. For examples, it compares two sentences and calculates a score of how similar they are. Or for signature verification, it determines if two signatures come from the same person [4]. Typically, in such a task, two identical sub-networks are used to process the two inputs and another model will obtain their output and produce the final output.

2.4   Sentiment Analysis

Sentiment analysis is the identification of attitudes expressed in a piece of text through natural language processing. In particular, it determines whether an author's attitude towards a particular topic, product, etc. is positive, negative or neutral.

The explosion of modern sentiment analysis occurred around the mid-2000s and their focus was very often on product reviews. In 2002, Pang et al. studied movie reviews and classified them into positive and negative sentiments with the help of machine learning methods such as plain Bayesian, maximum entropy classification and support vector machines [24].

Now sentence-level sentiment classification has been widely and intensively studied.

The information contained in the text has been largely utilised. In order to further improve sentiment classification, some researchers have attempted to introduce information beyond text. In [31], Zhou et al. introduced user information to consider each individual's preferences and language habits, and proposed a neural population sentiment analysis model to solve the data sparsity problem. In [3], Aljebreen et al. studied the structure of tweets with URLs and proposed an algorithm solved the segmentation problem of tweets with URLs, which led to improved performance of sentiment analysis of Twitter texts.

2.5 The San Francisco Model

Inside Airbnb uses an occupancy estimation model, named the San Francisco model [Inside Airbnb]. This model can be used to estimate how often an Airbnb house is being rented out (Equation. 2.1), and also approximate its income.

$$Occupancy = \frac{Number\,of\,Reviews}{Review\,Rate} \times max(Average\,Nights,\ Min\,Nights) \qquad (2.1)$$

The review rate in Equation. 2.1 can be different numbers. Inside Airbnb set it to 50%, meaning half of the tenants comment on the house. For the average number of nights of stay, Airbnb reported that it was 5.5 nights in San Francisco. If the average number of nights in a city is not known then three nights should be used. Lastly, the minimum number of nights of stay can be set for a house by its owner.

Chapter 3

Data Preparation

3.1 Data Collection

3.1.1 Airbnb Data

The flow diagram of data collection is shown in Fig. 3.1. The publicly accessible home data (Fig. 3.2) and owner data (Fig. 3.3) were obtained using Airbnb APIs, including home type, reviews, home photos, total number of an owner's homes, etc. No private information was used.
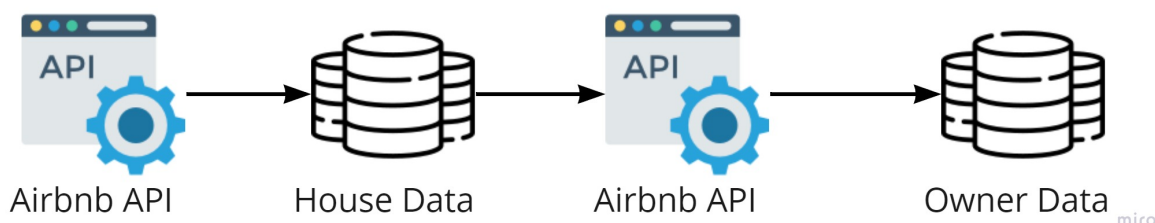
Figure 3.1: The Airbnb data collection

The location of home was randomized by Airbnb. In detail, the public geographic coordinate of a house in Airbnb data is a random location within a circle with a radius of 150 meters, centered on its real geographic coordinate. And all homes were anonymized by Airbnb. So even if two rooms come from the same building, it is impossible to distinguish them directly.

3.1.2 Geographic Information of Rent Pressure Zones

There are six local authority areas and 48 local electoral areas designated as rent pressure zones throughout Ireland as shown in Fig. 3.4. The local electoral area dataset [14]

URL	https://www.airbnb.com/rooms/52543856
Name	Coastal Apartment Cahore
ID	6687
Type	Private Room in House
Location	•
Country	Ireland
State	County Wexford
Min Nights	2
Max Nights	21
Reviews	•
Owner ID	5567
Photos	•

Longitude	52.56509
Latitude	-6.20175

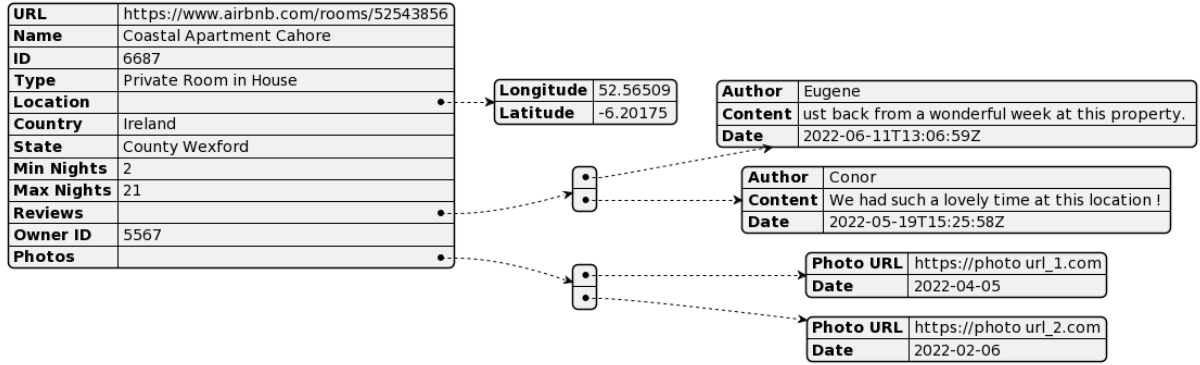| **Author**	Eugene
Content	ust back from a wonderful week at this property.
Date	2022-06-11T13:06:59Z

Author	Conor
Content	We had such a lovely time at this location !
Date	2022-05-19T15:25:58Z

Photo URL	https://photo url_1.com
Date	2022-04-05

Photo URL	https://photo url_2.com
Date	2022-02-06

Figure 3.2: The Airbnb house data

Name	Teresa
Owner ID	5567
Superhost	☑ true
Total Houses	1
Owner URL	https://www.airbnb.com.vn/users/show/id

Figure 3.3: The Airbnb owner data

contains geographical features as shown in Fig. 3.5, which helped identify whether a house was located within a rent pressure zone.

3.1.3 National Planning Application Dataset

The national planning application dataset [15] provides detailed information on all planning applications in Ireland and contains polygon coordinates of housing applications and descriptions of permission applications as shown in Fig. 3.6. This dataset was first filtered to obtain all entries related to short-term rental applications.

3.2 Data Preprocessing

3.2.1 Locating the Centroid of House

The national planning application dataset uses polygon coordinates to record the geographic location of a house. During this stage, the centroid coordinate was calculated and used to uniquely identify a house's location (Fig. 3.7) using Equation. 3.1.

$$(\overline{x}, \, \overline{y}) = \left(\frac{\sum_{i=1}^{n} A_i \overline{x}_i}{\sum_{i=1}^{n} A_i}, \, \frac{\sum_{i=1}^{n} A_i \overline{y}_i}{\sum_{i=1}^{n} A_i} \right) \tag{3.1}$$

10
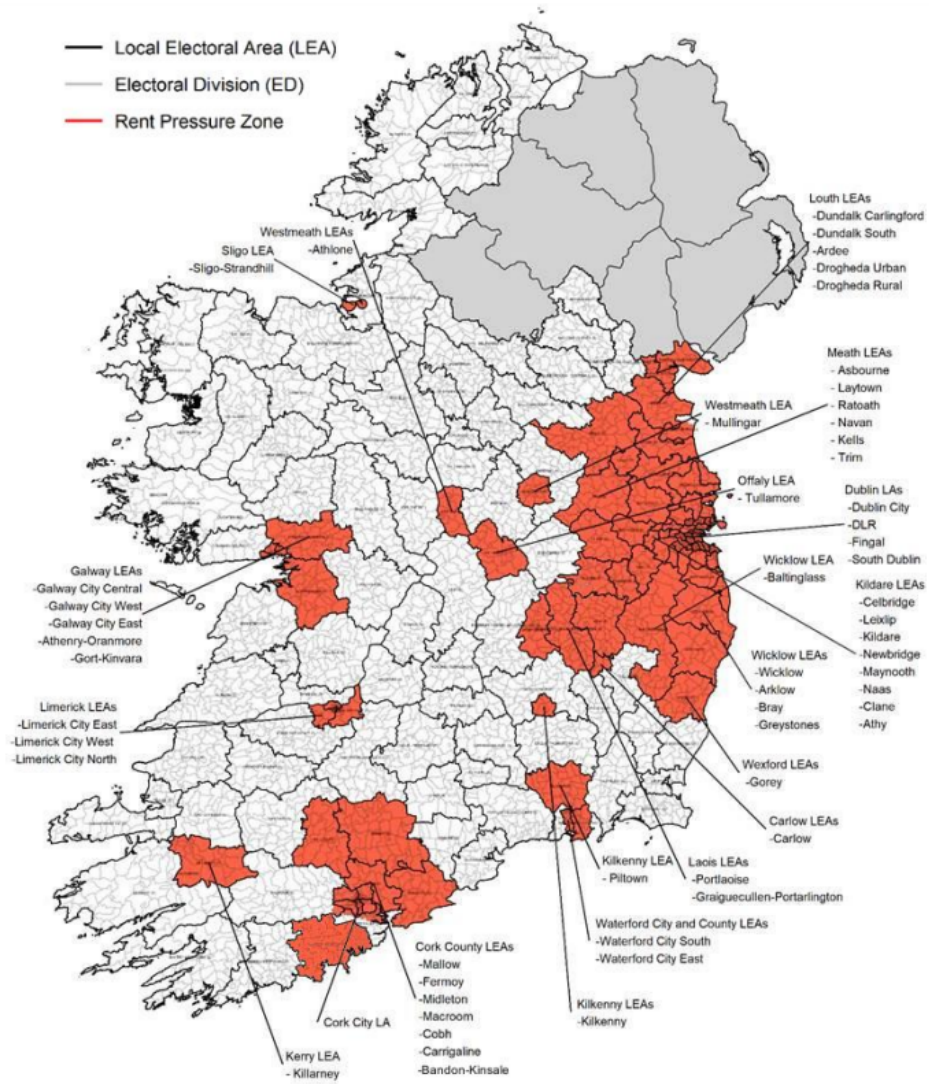
Figure 3.4: The map of rent pressure zones [16]

Local Electoral Area ID	13160403
Local Electoral Area	RATOATH LEA-7
County	MEATH
X-Centroid	696233.81
Y-Centroid	749103.41
Area	294980555.410
Object ID	1
Shape Area	294959510.061
Shape Length	107049.850
Geometry	•

Type	Polygon
Coordinates	•

-6.6982637900429
53.482236516507399

-6.69853412520148
53.482791700527102

-6.69431581798686
53.487216491747503
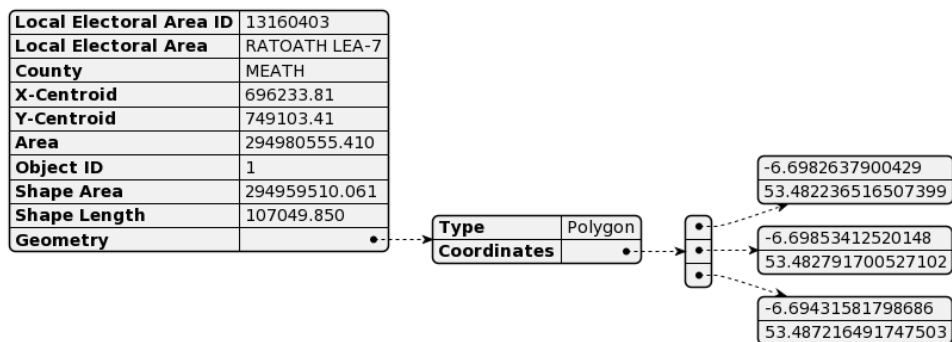
Figure 3.5: The local electoral area

11

Planning Authority	Meath County Council
Application Number	AA190968
Object ID	211173
Geometry	
URL	http://www.eplanning.ie/MeathCC/AppFileRefDetails/AA190968/0
Application Status	APPLICATION FINALISED
Decision	CONDITIONAL
Expiry Date	2024-11-04T00:00:00+00:00
size	Large
Development Address	Bellew , Rathfeigh , Co. Meath
Development Description	the refurbishment and conversion of a vernacular stone building to a house for short-term rental.

Type	Polygon
Coordinates	

-6.6982637900429
53.482236516507399

-6.69853412520148
53.482791700527102

-6.69431581798686
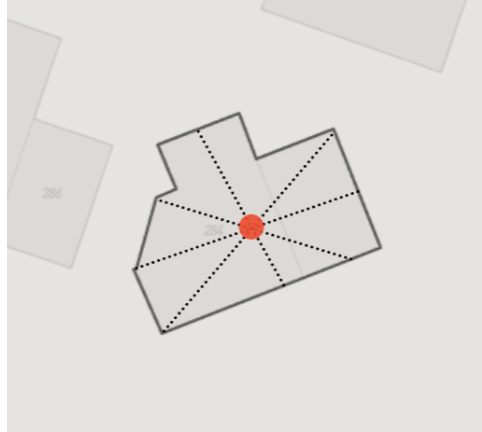53.487216491747503

Figure 3.6: The national planning application

Figure 3.7: The centroid of a house

3.2.2 Filtering Houses in Rent Pressure Zones

The PNPOLY algorithm was used to filter out houses that did not locate in a rent pressure zone. This algorithm was proposed by [Franklin] to solve the problem of determining whether a coordinate is inside a polygon for a geographic information management system.

Chapter 4

The Backend Design

Fig. 4.1 shows the components of the breach identifier. The yellow block is the breach identifier, which contains three components:

- A principal residence identifier containing two components:

  - An image classifier for indoor and outdoor. It deleted noisy outdoor photos before image similarity detection.

  - An image similarity detector used to determine whether photos from different posts come from the same residence or the same room.

- A permit finder. Under the rules, it is necessary to determine whether a house has a planning permission if it is not an owner's principal residence or its cumulative rental period within one year is more than 90 days. The permit finder checked if a house is within the scope of a valid short-term rental permit.

- An occupancy estimator. The rules require that when an owner entirely rents out their principal residence, it cannot exceed 90 days. Due to trade secrecy issues, Airbnb does not disclose the occupancy time of a home, so the only way to estimate it is by the number of reviews.

4.1   The House and Owner Models

The Airbnb house and owner data were modelled. A house object contained its ID, type, coordinates, reviews, photos, owner ID, etc, as shown in Fig. 4.2. An owner object contained its ID, a list of rented houses, etc, as shown in Fig. 4.3. Each owner object was associated with house objects using its owner ID.
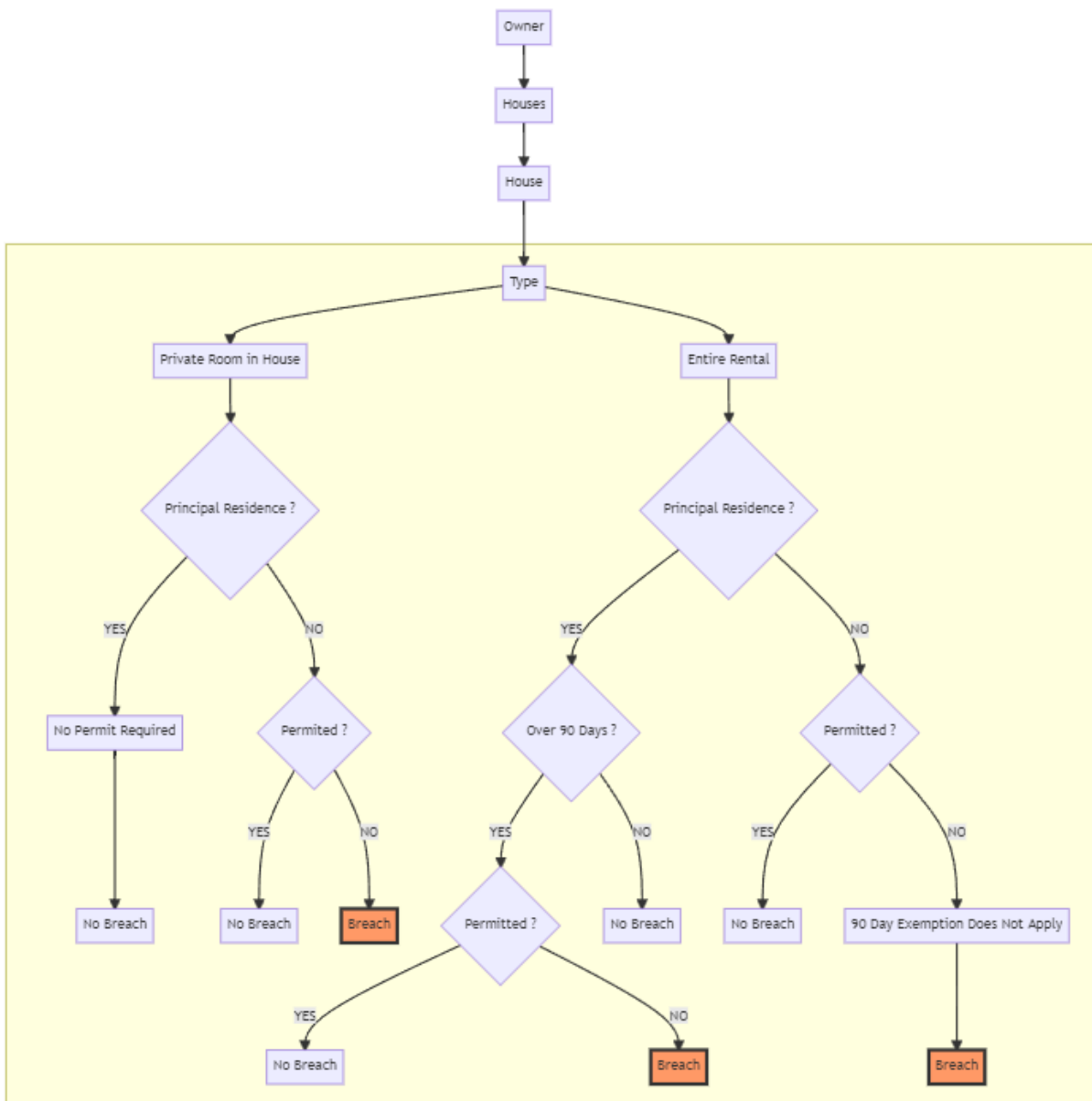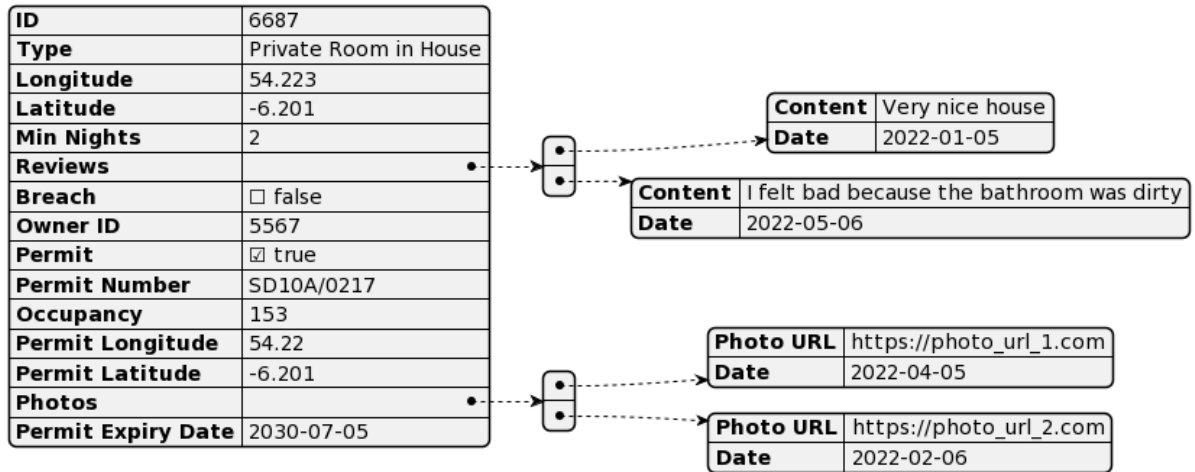
Figure 4.1: The system backend

ID	6687
Type	Private Room in House
Longitude	54.223
Latitude	-6.201
Min Nights	2
Reviews	●
Breach	☐ false
Owner ID	5567
Permit	☑ true
Permit Number	SD10A/0217
Occupancy	153
Permit Longitude	54.22
Permit Latitude	-6.201
Photos	●
Permit Expiry Date	2030-07-05

Content	Very nice house
Date	2022-01-05

Content	I felt bad because the bathroom was dirty
Date	2022-05-06

Photo URL	https://photo_url_1.com
Date	2022-04-05

Photo URL	https://photo_url_2.com
Date	2022-02-06

Figure 4.2: The house object

Total Homes	4
Owner ID	5567
Breach	☑ true
Houses in Rent Pressure Zones	●
Breach Houses	●

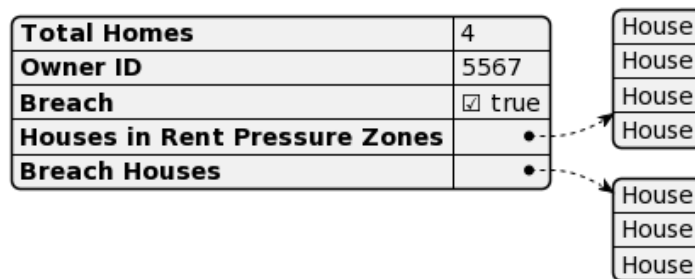| House |
|---|
| House |
| House |
| House |

House
House
House

Figure 4.3: The owner object

4.2 The Principal Residence Identifier

The principal residence identifier was based on the rules for short-term rentals within rent pressure zones. If a short-term rented house is not an owner's principal residence, then they need to obtain the relevant planning permission to be allowed to legally rent it out. The flow chart is shown in Fig. 4.4, where an owner's all house photos were filtered through a residual neural network based image classifier to remove outdoor photos, then a Siamese network was used to compare the similarity of indoor photos to remove duplicate houses.
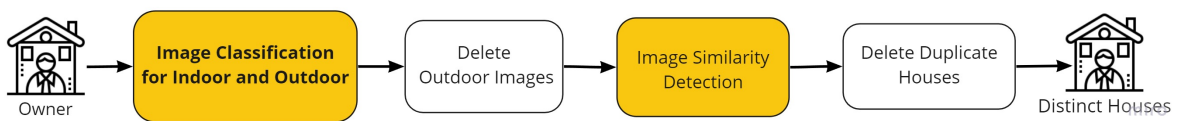

Figure 4.4: The principal residence identifier

4.2.1 Image Classification for Indoor and Outdoor

Some posts on Airbnb contain many outdoor photos such as beautiful landscapes because owners want to use them to attract more visitors. These noisy data can affect the subsequent identification of whether multiple short-term rentals correspond to the same house posted by an owner. For example, two posts contained the same landscape photos in Fig. 4.5, but they came from different houses. So it was necessary to remove these noisy images.

The image filter used the residual neural network ResNet-18 to classify whether an image was indoor or outdoor and deleted those landscape photos. The training data for ResNet-18 was the MIT Indoor Scenes dataset from Kaggle [AHMAD] and its accuracy was 75%. It could effectively filter out outdoor photos.

4.2.2 Image Similarity Detection

There were two legitimate scenarios that the breach identifier might incorrectly assume that an owner has multiple homes in a rent pressure zone, which is not allowed.

- Owners are allowed to rent out multiple rooms in their principal residence without any restriction. They can post two or more rooms in their principal residence on Airbnb.

Figure 4.5: The outdoor images in two posts

- Owners may post the same room multiple times in Airbnb in order to give their home more exposure. In Fig. 4.6, these two posts correspond to the same room but the photos were taken from different angles.

Figure 4.6: Two posts correspond to the same room

To address these two issues, photos from different posts needed to be checked to see whether they came from the same principal residence or the same room. If a room was posted multiple times, there must be many similar or the same photos. If multiple rooms were from the same residence, their photos probably contained the identical common areas such as the kitchen, hallway and bathroom. For example, three posts in Fig. 4.7 were from different rooms in the same residence. It is clear that they had the identical common areas with a very high degree of similarity.

In most cases, two images will not be fully identical even if they have the same scene or object due to differences in size, scale, rotation and skew. So the comparison cannot be made directly using pixel-to-pixel methods. The problem shifted from identifying pixel similarity to object similarity.

`sentence-transformers` library was used in this project, which provides a way to compute a dense vector representation of an image and find identical images [25]. It encodes all images into vector space and then find high density regions corresponding to areas where the images are similar. For example, when using image search, a photo is entered and converted into a set of vectors. `sentence-transformers` compares it with other images in the database and finds the image with the highest similarity (the closest distance). When two photos from one post had greater than 95% similarity to the photos
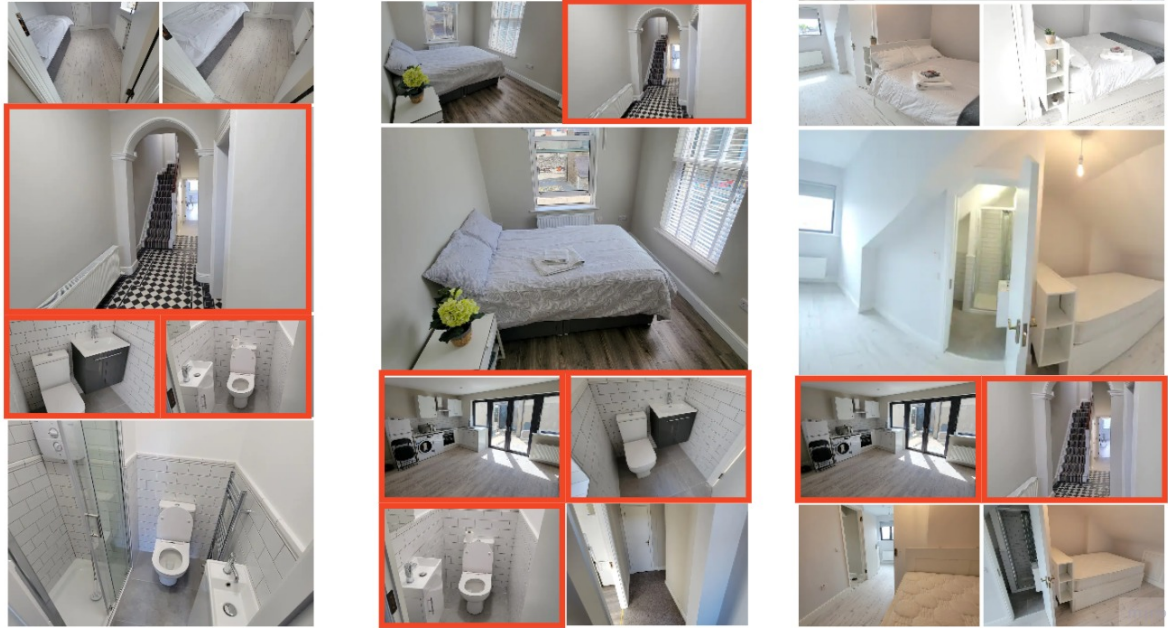
Figure 4.7: Three posts correspond to the same residence

from another post, or one photo had 100% similarity to another, then two posts were considered to be associated with the same room. So the breach identifier would not assume that its owner had violated the rules.

4.3   The Permit Finder

As the real geographic coordinate of a house had been anonymized by Airbnb, replaced with a random location within a circle with a radius of 150 meters (Fig. 4.8), house coordinates were unreliable. It is not possible to know whether a permit is available for a short-term rental house by conventional coordinate comparison. The flow chart of permit finder is shown in Fig. 4.9. Because of the anonymization issue, when there were multiple houses around a permit, they would all be considered to have the same available permit.

The Haversine algorithm was used to determine whether a location A was within a circle with a radius of 150 meters, centered on another location B [27].

By using the longitude and latitude of two locations, it can calculate their distances on the surface of the Earth. The formula is Equation. 4.1 where R is the radius of the Earth which is 6371 kilometers.
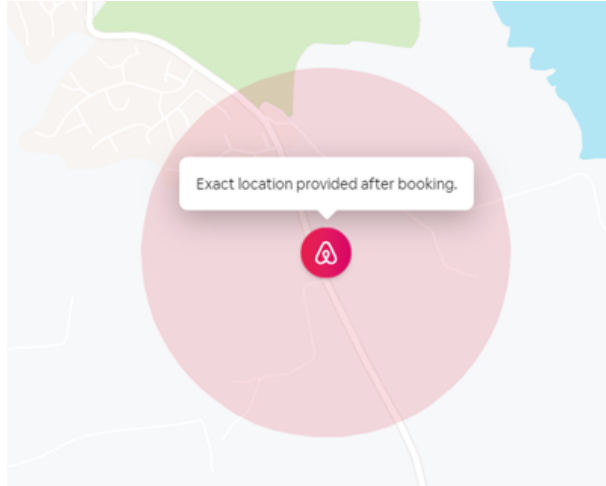
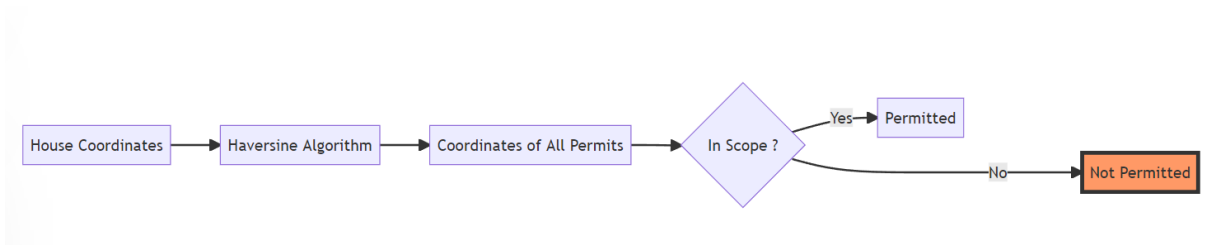Figure 4.8: The anonymous geographic coordinate of a house

Figure 4.9: The permit finder

$$\Delta Lat = Latitude_2 - Latitude_1$$

$$\Delta Lon = Longitude_2 - Longitude_1$$

$$Distance = 2 \cdot R \cdot \arcsin\left(\sqrt{\sin^2\left(\frac{\Delta Lat}{2}\right) + \cos(Latitude_2) \cdot \cos(Latitude_1) \cdot \sin^2\left(\frac{\Delta Lon}{2}\right)}\right)$$

$$(4.1)$$

4.4 The Occupancy Estimator

The occupancy estimator was based on the rule that in some cases short-term rentals are not allowed to accumulate more than 90 days within a year. One of the biggest issues with Airbnb is whether owners are renting out residential properties as hotels on a permanent basis, rather than occasional sharing. On Airbnb, the occupancy time of a house is unpublished data. But it can be estimated by the number of user reviews.

In [30], Wu et al. found that the sentiment of hotel reviews affected its occupancy. In this project, an enhancement for Inside Airbnb upgraded San Francisco model was

proposed. Instead of using a fixed review conversion rate, the enhanced model used a rate of 50% as a base, and then had a sentiment bias of 10% increment or reduction depending on the sentiment score of house reviews. Better reviews resulted in more occupancy time. The formula is Equation. 4.2, where the average number of nights of stay for guests was 4.6, taken from public Airbnb data, and the minimum number of nights of stay could be set for a house by its owner. Finally, the review rate was set as 50%.

$$Occupancy = \frac{Number\,of\,Reviews}{Review\,Rate + Sentiment\,Bias} \times max(Average\,Nights,\;Min\,Nights)$$

$$(4.2)$$

The flow chart of sentiment bias calculation is shown in Fig. 4.10. `TextBlob` was introduced which is a Python library for natural language processing [Loria]. `TextBlob` mainly uses its pre-trained built-in classifiers to perform sentiment analysis and can obtain a corresponding sentiment score for a comment.

1. The reviews of a house were filtered to keep only the reviews within one past year.

2. Reviews might be written in non-English languages and DeepL APIs[1] were used to automatically translate those reviews to English.

3. `TextBlob` analyzer calculated an average sentiment score for all reviews. A sentiment score was between $[-1,\,1]$ where -1 defined extremely negative emotion and 1 defined extremely positive emotion.

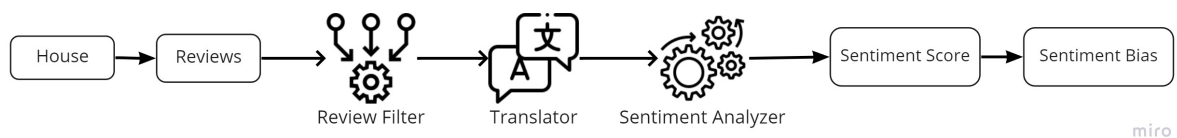4. With the sentiment score, a sentiment bias could be produced using Equation. 4.3.

Figure 4.10: The calculation of sentiment bias

$$Sentiment\,Bias = Polarity \times 0.1 \qquad\qquad (4.3)$$

Chapter 5

The Frontend Design

Considering that the breach identifier needs to keep running in the background for a long time, so the traditional Browser/Server architecture was used as Fig. 5.1.
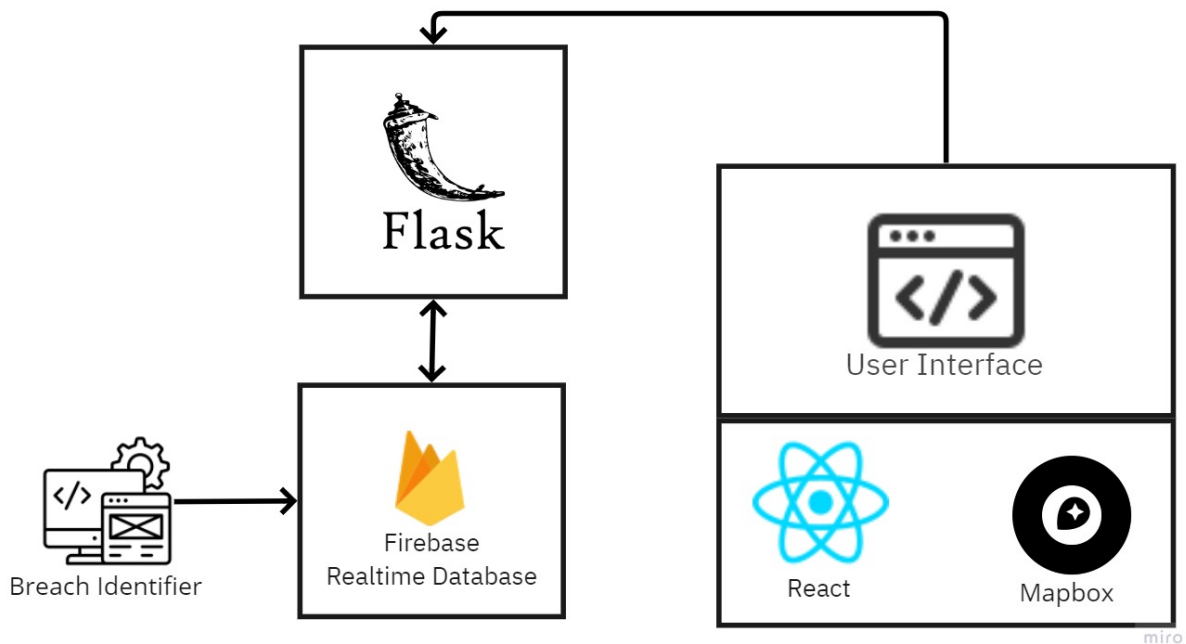

Figure 5.1: The frontend design

5.1 The Flask Server

`Flask` is a lightweight and customizable website framework in Python [18], providing a rich set of basic components for the web development. It was used to build a backend server that provides APIs for data transfer from backend to frontend. As shown in Fig. 5.2,

four API endpoints were designed in this project. In practice, different HTTP requests might invoke the same processing logic, they could be identified by a class of URLs.

- Database update subscription.

- Map data access.

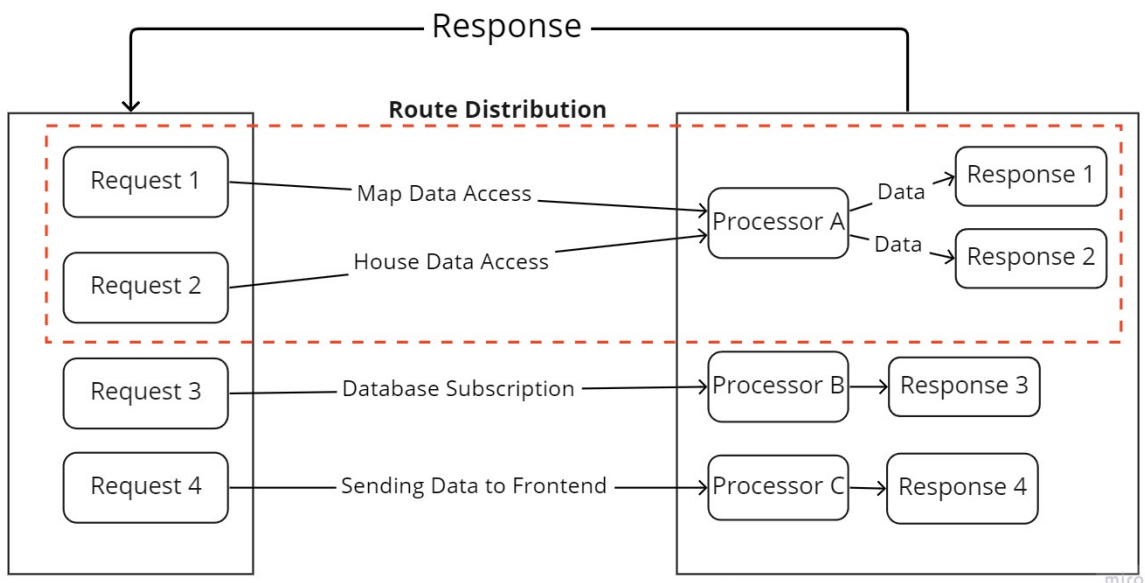- House data access.

- Sending data to frontend.

Figure 5.2: The route distribution of the server

5.2 The Firebase Realtime Database

`Flask` does not specify objects such as databases and template engines, so users can choose their own databases as needed. The Firebase realtime database used in this project was a database hosted in the cloud [Google]. The data were stored using `JSON` format and synchronized in real time to each connected client. The Flask server used a real time listener for the Firebase database. When the breach identifier finished parsing and depositing the data, Firebase would automatically notify the server to obtain the latest data. Additionally, the Firebase database has offline functionality so that even if the network connection is temporarily disrupted, the application can still work.

5.3 Mapbox

`Mapbox` is a JavaScript library that renders interactive vector tile maps and raster tile maps using WebGL [Mapbox]. WebGL rendering brings high performance and `Mapbox` is capable of rendering a large number of map elements with smooth interaction and animation effects.

Fig. 5.3 and Fig. 5.4 show short-term rentals in Galway and entire Ireland respectively.

- The red marker represents that a rental is probable already in breach.

- The yellow marker represents a rented house is an owner's principal residence and its rental period is greater than 70 days, which means it is about to be in breach (up to 90 days).

- The black marker represents a permitted location, where the red area is a circle with a radius of 150 meters, centered on the permitted location. All houses in this area are considered to have a permit.
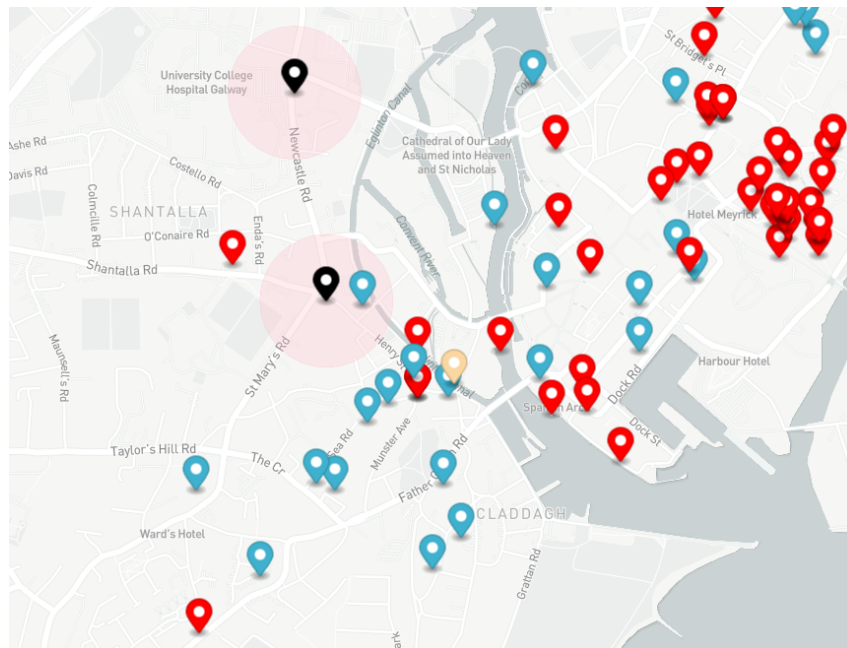

Figure 5.3: Short-term rentals in Galway

When a user clicks on a marker, the corresponding information will be displayed. Clicking on a black marker will show its expiry date and permit description, etc. While clicking on a red marker will show its house type, occupancy and the total number of houses owned by its owner, etc, as shown in Fig. 5.5.
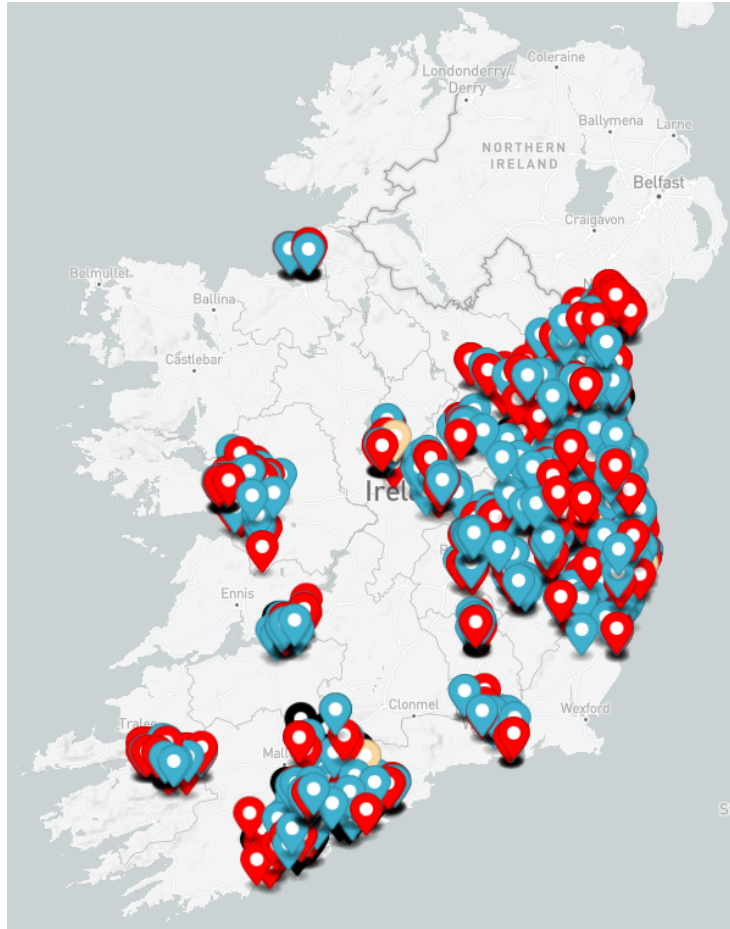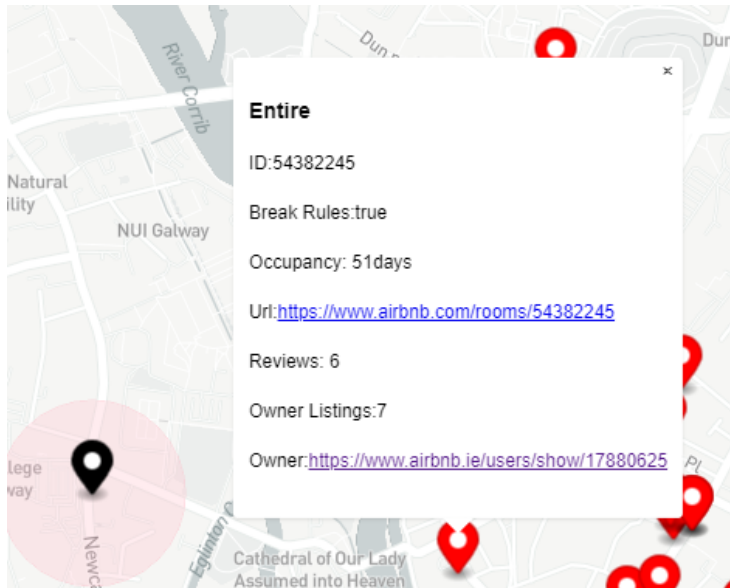
24

Figure 5.4: Short-term rentals in Ireland

Figure 5.5: The details of a rental

Chapter 6

Evaluation

There were 1713 short-term rentals located in rent pressure zones in collected data. 845 of them were identified as having potential breaches, and 21 principal houses with occupancy greater than 70 days were about to breach the rule. The overall breach rate was 50.67% However, the reality can be worse, as a permit only works for one house in the real situation. But the anonymous coordinates of houses might cause many houses to be located within a certain available permit at the same time, resulting in a situation where one permit was used for multiple houses.

Because Airbnb does not disclose accurate house coordinates and occupancy data, it is not possible to verify the accuracy of breach identifier. The accuracy of occupancy estimator cannot be verified either. It only provides an estimate within a reasonable range. Users should be somewhat skeptical of short-term rentals that are flagged as possible breaches.

Chapter 7

Conclusion

In summary, this project achieved the following contributions and benefits to the current regulation of the Irish short-term rental market:

- It provided a viable method to identify potential short-term rentals that may be in violation and their owners when no accurate data is available.

- It provided a more efficient way to monitor the short-term rental market, rather than relying on traditional user reporting.

- It provided an image identifier that can determine whether multiple rental posts correspond to the same room or residence by calculating the similarity of room photos. This model may be applied to other areas. For instance, some second-hand goods websites may prohibit users from posting duplicate items.

- It enhanced the occupancy estimation model combined with sentiment analysis, which may provide higher accuracy. The enhanced model may also be used for hotel occupancy estimation, etc.

There are two main points about the future work:

- With the increasing concern of the Irish government about the housing crisis, the possibility of obtaining accurate data exists. Then this project will be able to be verified for its accuracy and validity.

- The Haversine algorithm provides a simpler and faster calculation, but it is not as accurate as the Vincenty algorithm [29], which can be used to replace the Haversine algorithm in the future. The higher accuracy is necessary.

Bibliography

[AHMAD] AHMAD, M. Mit indoor scenes. Available at `https://www.kaggle.com/datasets/itsahmad/indoor-scenes-cvpr-2019` (2019).

[AirDNA] AirDNA. Airdna rental data. Available at `https://www.airdna.co/vacation-rental-data/app/ie/default/dublin/overview` (2022/08/18).

[3] Aljebreen, A., Meng, W., and Dragut, E. (2021). Segmentation of tweets with urls and its applications to sentiment analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 12480–12488.

[4] Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., and Shah, R. (1993). Signature verification using a" siamese" time delay neural network. *Advances in neural information processing systems*, 6.

[5] Chicco, D. (2021). Siamese neural networks: An overview. *Artificial Neural Networks*, pages 73–94.

[Citizens Information] Citizens Information. Renting your property for short-term lets. Available at `https://www.citizensinformation.ie/en/housing/owning_a_home/home_owners/renting_your_property_for_shortterm_lets.html` (2022/08/11).

[Cox et al.] Cox, M., Morris, J., and Higgins, T. Inside airbnb dublin dataset. Available at `http://insideairbnb.com/dublin` (2022/08/18).

[8] Daft.ie (2022). The daft.ie rental price report an analysis of recent trends in the irish rental market 2022 q2. Technical report, DUBLIN, 6 Ireland.

[9] Department of Housing, Local Government and Heritage (2022/03/01). PLANNING AND DEVELOPMENT REGULATIONS 2001 – 2022. Technical report.

[Franklin] Franklin, W. R. Pnpoly - point inclusion in polygon test. Available at `https://wrfranklin.org/Research/Short_Notes/pnpoly.html#otherwork`.

[11] Gataveckaite, G. Airbnb crackdown: Sinn fein bill to turn screw on short term rentals. Available at `https://www.independent.ie/irish-news/airbnb-crackdown-sinn-fein-bill-to-turn-screw-on-short-term-rentals-41598813.html` (2022/04/28).

[12] Gataveckaite, G. Revealed: The counties that are clamping down on airbnb rentals... and the ones that aren't. Available at `https://www.independent.ie/irish-news/revealed-the-counties-that-are-clamping-down-on-airbnb-rentals-and-the-ones-that-arent-41595034.html` (2022/04/28).

[Google] Google. Firebase realtime database. Available at `https://firebase.google.com/products/realtime-database` (2022).

[14] Government of Ireland. local electoral areas osi national statutory boundaries generalised-20m1. Available at `https://data.gov.ie/dataset/local-electoral-areas-osi-national-statutory-boundaries-generalised-20m1`.

[15] Government of Ireland. National planning applications. Available at `https://data.gov.ie/dataset/national-planning-applications` (2022).

[16] Government of Ireland (2020/04/30). Designated rent pressure zones in ireland. Technical report.

[17] Government of Ireland (2021). Residential tenancies(amendment) act 2021 what this means for landlords and tenants. Technical report.

[18] Grinberg, M. (2018). *Flask web development: developing web applications with python.* " O'Reilly Media, Inc.".

[19] He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

[Housing Authority] Housing Authority. Planning permission. Available at `https://www.citizensinformation.ie/en/housing/planning_permission/planning_permission_general.html` (2022/08/11).

[Inside Airbnb] Inside Airbnb. San francisco model. Available at `http://insideairbnb.com/data-assumptions` (2022).

[Loria] Loria, S. Text blob. Available at `https://textblob.readthedocs.io/en/dev/` (2021).

[Mapbox] Mapbox. Mapbox gl js. Available at `https://docs.mapbox.com/mapbox-gl-js/api/` (2022).

[24] Pang, B., Lee, L., and Vaithyanathan, S. (2002). Thumbs up? sentiment classification using machine learning techniques. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, pages 79–86. Association for Computational Linguistics.

[25] Reimers, N. and Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

[26] Residential Tenancies Board (2022). Residential Tenancies Board Publishes Q1 2022 Rent Index. Technical report, Residential Tenancies Board, Po Box 47, Clonakilty, County Cork.

[27] Robusto, C. C. (1957). The cosine-haversine formula. *The American Mathematical Monthly*, 64(1):38–40.

[28] Roy, P., Ghosh, S., Bhattacharya, S., and Pal, U. (2018). Effects of degradations on deep neural network architectures. *arXiv preprint arXiv:1807.10108*.

[29] Wang, J., Sun, Y., and Dai, Y. (2022). Inverse geodetic problem for long distance based on improved vincenty's formula. *Journal of Applied Geodesy*.

[30] Wu, D. C., Zhong, S., Qiu, R. T., and Wu, J. (2022). Are customer reviews just reviews? hotel forecasting using sentiment analysis. *Tourism Economics*, 28(3):795–816.

[31] Zhou, D., Zhang, M., Zhang, L., and He, Y. (2021). A neural group-wise sentiment analysis model with data sparsity awareness. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 14594–14601.