



Sri Lanka Institute of Information Technology

## **Rental Price Prediction in China**

Statement of Work Document

FDM Mini Project 2022

Submitted by:

IT20204716 – (Hemachandra K.M.T.A)

IT20149680 – (Athurugiriya A.A.A.G)

IT20203412 – (Dinoja N)

IT20123772 – (Pallawala P.K.B.D.S)

IT20123840 – (Sewwandi R.A.U.T)

15/10/2022

Table of Contents

Title	Page Number
<a href="#">Background</a>	3
<a href="#">Scope of work</a>	4
<a href="#">Activities</a>	6
<a href="#">Approach</a>	6
<a href="#">Deliverables</a>	7
<a href="#">Project Plan &amp; Timeline</a>	7
<a href="#">Assumptions</a>	8
<a href="#">Project team, roles and responsibilities</a>	8

## Background

We were advised to find out a real-world problem and propose a solution for it as our mini project. Data Mining and Machine Learning are the techniques that we can use. After discussing with the group members, we decided to take a rental price prediction problem which is very useful for property owners as well as the buyers.

The real-world scenario that we identified is that if a property owner who do not have a better knowledge about rental prices, wants to predict the price for his/her property, that person has to waste a lot of time to find the details of rental properties. Not only that, but also buyers also have to put in a lot of effort to find the budget of a property that suits their needs. Often a person is not aware of whether it is appropriate to cancel a booking. It can be a problem for buyers because they do not know the policies. Another problem that a property owner can have is, when a person wants to develop his/her rental business, it is difficult to decide which type of properties are best for rent.

By analyzing historical data of properties, we can mine these data and identify patterns that will finally conclude the rental price of a property as well as the cancellation policy. And also, we can predict which type of properties are best to expand the business.

This will help property owners to predict the price of their properties, and also the buyers to find the price of a property according to the features that they want. Buyers can predict the cancellation policy type of the property, so that they can get an idea about the risk of canceling a booked property. Property owners can get a better idea about which types of properties can gain a good profit.

The dataset we are going to use is Oyo Rental Price Prediction in China, found on a web site which is known as 'kaggle'. The data is related with China's largest hotel chain which is initially consisted mainly of budget accommodations.

The dataset will be used to create a model that can predict the price of a rental property according to their features, cancellation policy type of a particular property, as well as the most profitable property type, by splitting the dataset to training and testing datasets, and doing the necessary analysis.

**Dataset:** [https://www.kaggle.com/datasets/ramjasmaurya/oyo-rental-price-prediction-in-china?resource=download&select=rating\\_features.csv](https://www.kaggle.com/datasets/ramjasmaurya/oyo-rental-price-prediction-in-china?resource=download&select=rating_features.csv)

## Scope of work

This dataset contains 5834 rent outs done previously. There are 26 features collected as below,

1. accommodates (numeric)
2. amenities (categorical)
3. availability\_30 (numeric)
4. bathrooms (numeric)
5. bed\_type (categorical: 'Real bed, Airbed', 'Futon', 'Couch', 'Pull-out sofa')
6. bedrooms (numeric)
7. beds (numeric)
8. calculated\_host\_listings\_count (numeric)
9. guests\_included (numeric)
10. has\_availability (binary: t, f)
11. host\_is\_superhost (binary: t, f)
12. host\_listings\_count (numeric)
13. instant\_bookable (binary: t, f)
14. latitude(North) - (numeric)
15. longitude(East) - (numeric)
16. maximum\_nights (numeric)
17. number\_of\_reviews (numeric)
18. property\_type (categorical: 'Apartment', 'Bed & Breakfast', 'Boat', 'Bungalow', 'Cabin', 'Camper/RV', 'Chalet', 'Condominium', 'Earth House', 'House', 'Hut', 'Loft', 'Other', 'Tent', 'Tipi', 'Townhouse', 'Treehouse', 'Villa')
19. review\_scores\_checkin (numeric)
20. review\_scores\_communication (numeric)
21. review\_scores\_location (numeric)
22. review\_scores\_rating (numeric)
23. review\_scores\_value (numeric)
24. room\_type (categorical: 'Private room', 'Shared room', 'Entire home/apt')

### Desired Target Feature

25. price (numeric)
26. cancellation\_policy (categorical: 'strict', 'flexible', 'moderate')

We will analyze these data to:

- Predict the rental price
- Predict the cancellation policy
- Find top selling property types

When roughly considering the above features, we hope to remove some columns like `calculated_host_listings_count`, `host_is_superhost`, `host_listings_count`, `latitude`, `longitude`, `number_of_reviews`, `review_scores_checkin`, `review_scores_communication`, `review_scores_location`, `review_scores_rating`, `review_scores_value` since they may not affect the prediction. This will be concluded by doing the necessary data analysis by the Data Scientists of the project.

We will remove the rows containing null values in the data preprocessing level. The dataset will be split into training dataset and testing dataset. Regression, classification and clustering algorithms will be used to train the dataset and build the model. The testing dataset will be used to validate the model and check the accuracy of the predictions. Finally, the best fit model will be selected. Then a web application will be developed for the users to easily get the predictions by entering the related features.

## Activities

We identified a problem that is related to rental industry, then found a dataset of Oyo rental in Shenzhen, China from Kaggle. And we are going to perform some data pre-processing technique to transform the raw data into a useful and efficient format. And also, we are cleaning the dataset so that it is easy for us to come up with solutions. Then we are defining the entities and using models such as multiple linear regression, decision tree and clustering models to train the dataset. Our next plan is to design a user friendly front-end and back-end and connect them so that property owners can decide the price of their properties based on the facilities and other necessary things of their property. Also, the buyers can decide which one is the best to rent.

## Approach

In this mini project, our goal is to analyze Oyo rental price prediction data (Selected from Kaggle website) using machine learning techniques and algorithms to forecast whether the price of a rental in the future will be changed than its price on a given day, based on historical data while providing an in-depth understanding of the models being used. We will employ the data mining methods, data preprocessing methods such as data cleansing, data transformation, data conversion and feature selection And Classification, clustering and regression techniques will be used to get highest accuracy of the predictions. We desire to utilize Streamlit for the Front-end development while python will be used for back-end development and analyze the data.

([https://www.kaggle.com/datasets/ramjasmaurya/oyo-rental-price-prediction-in-china?resource=download&select=rating\\_features.csv](https://www.kaggle.com/datasets/ramjasmaurya/oyo-rental-price-prediction-in-china?resource=download&select=rating_features.csv) )

## Deliverables

The outcome of this project will predict the rental price of a property according to the features that user enters. This will help for property owners to predict the price of a rental property and also buyers can find the prices of properties that they hope to buy.

This project will predict the cancellation policy type of a rental property, so that buyers can get a rough idea about it. This helps to predict the most rented property type. Property owners can develop their business according to this prediction.

## Project Plan & Timeline

[illegible]

## Assumptions

- The columns that were removed do not affect the prediction of the final output.
- The audience of the Final Product is someone with the proper knowledge to understand and get a clear idea of the system outcome.
- The users have the values for the required inputs for the system to predict the values.
- The dataset is being updated annually as mentioned in the source.

## Project team, roles and responsibilities

Registration Number	Roles	Responsibilities
IT20204716	Data Scientist Data Engineer UI/UX Engineer	<ul style="list-style-type: none"><li>• Data cleaning and preprocessing.</li><li>• Building the models.</li><li>• Testing and Evaluating.</li><li>• Design UI/UX and integrating the project</li><li>• Creating the final report</li></ul>
IT20149680	Data Scientist Data Engineer UI/UX Engineer	<ul style="list-style-type: none"><li>• Data cleaning and preprocessing.</li><li>• Building the models.</li><li>• Testing and Evaluating.</li><li>• Design UI/UX and integrating the project</li><li>• Creating the final report</li></ul>
IT20203412	Data Scientist Data Engineer UI/UX Engineer	<ul style="list-style-type: none"><li>• Data cleaning and preprocessing.</li><li>• Building the models.</li><li>• Testing and Evaluating.</li><li>• Design UI/UX and integrating the project</li><li>• Creating the final report</li></ul>
IT20123772	Data Scientist Data Engineer Data Analyst	<ul style="list-style-type: none"><li>• Data cleaning and preprocessing.</li><li>• Building the models.</li><li>• Testing and evaluating.</li><li>• Visualize the results.</li><li>• Creating the final report</li></ul>
IT20123840	Data Scientist Data Engineer Data Analyst	<ul style="list-style-type: none"><li>• Data cleaning and preprocessing.</li><li>• Building the models.</li><li>• Testing and evaluating.</li><li>• Visualize the results.</li><li>• Creating the final report</li></ul>