Mini Project

Fundamentals of Data Mining - IT3051

Statement of Work Document

Group No : 15

Group Members :

Student No	Name
IT21029868	B.D.A.D.Hettiarachchi
IT20708276	U.D.K.Navaratne
IT21088582	A.N. Elvitigala
IT21015212	V.L.K .Tennekoon

Table of Content

1. Introduction

Dubai is not just a city, but it's a country of its own that has everything you need to experience life to its fullest. Living in Dubai is like living in the heart of everything where you can have mind-blowing experiences every day.

Dubai's flourishing economy has led to a leap in its real estate market. Choosing the properties which range from apartments and houses to villas, lands and commercial spaces is not restricted to investors and businessman only, many individuals prefer to have a permanent residence in Dubai. Freehold property policies, tax-free policies, a diversified economy and the assurance of overall safety and security are the reasons why someone benefits from buying properties in Dubai.

Buying a property in Dubai is a wise decision on both commercial and financial fronts. Dubai property prices vary based on the type of property, its location and facilities. It has properties that come in all shapes and sizes and will mostly match the budgets of those interested in purchasing them.

The purpose of this report is to identify the factors which influence the property prices of Dubai to develop a predictive model to facilitate investors in determining prices for a property according to the requirements which the investors specify.

2. Description of the Issue

Dubai is considered to be one of the fastest-growing cities in the world. Yet it should be noted that the Dubai real estate market remains fairly priced compared to other major cities like London, Honk Kong, Munich, and Frankfurt. These major cities according to the UBS Real Estate Bubble Index released for the year 2020 are overvalued and the bubble can burst at any point sinking the prices drastically (UBS Chief Investment Office, 2020).

High market valuations and uncertain short term outlooks are bringing the longer-term trajectory of city housing to focus. Yet what needs to be noted is the fact that Dubai is considered to be a real estate investment paradise because of its high rental yields. Properties in Dubai provide an average yield of around 7% annually which is significantly high when compared to New York which has an annual yield of around 2.9%, London (2.7%), Singapore (2.5%), or Hong Kong (2.4%) (FinSMEs, 2020). Another point to note is that Dubai has no property tax.

Therefore a predictive model for property prices in Dubai can assist any investors and sellers in making decisions related to applying for a mortgage or in balancing their investment portfolio. Because all investors should have a piece of Dubai in their investment basket.

3. Description of the Dataset

Over the last two decades, Dubai has emerged as one of the most popular destinations for real estate investments owing to the gorgeous skyscrapers, unlimited retail and dining options, low crime rate, excellent transportation routes, favorable return on investments that the city has on offer. The dataset “Dubai Property” available in Kaggle was used to fit a predictive model using advanced analysis after performing exploratory data analysis in the first step. Creating a more effective data product where property buyers and also sellers can easily and quickly find the price of a property with desired requirements would be the main objective of the project and thus best fit model for the data would be identified in this step.

In this study the response variable is price.

4. Scope of Work

The project consists of 5 layers, each briefly explained as follows.

1. **User Interface Layer**

The frontend of the system which acts as in intermediate between the user and the underlying application. The interface provides the user with functions to input data and configure parameters, to select and configure the model and to enter other necessary analytical requirements. It is a simple and user-friendly provision, developed by optimizing user convenience.

2. **Data Pre-processing and Data Transformation Layer**

This layer is primarily tasked with data preprocessing and data cleaning, which includes the handling of missing data, removal of duplicate data, outlier detection and correction of inconsistencies within the data. Through these methods all incomplete, inaccurate, and irrelevant data is dealt with, in a precise manner. At its core, the process involves converting and organizing raw data into a suitable format for future analytical requirements.

3. **Data Mining Layer**

The process of exploring the data to gain insights into its structure and characteristics, followed by the recognition and understanding of patterns and relationships between data by applying data mining techniques, including statistical analysis, clustering and association rule mining.

4. **Model Building and Evaluation Layer**

This layer consists of selecting an appropriate model by considering the scalability and availability of different models, tuning of hyperparameters to optimize performance, training the model, testing and evaluating the model and iterating the processes to minimize the loss function. The model will be able to predict the housing prices according to the given set of features with high accuracy rates.

5. **Data Visualization Layer**

This layer is connected to the UI layer in order to represent the final analysed data using graphical means such as histograms, bar charts, scatter plots etc. This allows the users to view results in a less complex, and visually appealing manner and get a clear understanding of the analysed data.

5. Activities

Select a dataset that contains a business issue that can be resolved by applying machine learning techniques.

i. Pre-process data
ii. Pick various models.
iii. Train each model

Our goal is to build a machine learning model which is capable of predicting the price of a property based on other features .

iv. Verify the model's accuracy (Evaluation).
v. Make predictions by providing inputs.
vi. The development of a user interface.

6. Approach

First and foremost, a dataset was found and prepared to ensure that the data is clean as we approached the model building from scratch. The dataset was then analyzed to facilitate the outcomes mentioned under deliverables in order to ensure the best deliverables are delivered.

- Dataset – Dubai Properties
  - **i. Data Pre-processing:**
    - After going through the dataset, we analyzed that it did not contain any missing values in any columns
    - However, we figured out we may need to replace some categorical columns to numeric columns

- Selected columns of the data set will be used for the analysis.
- The selected data set will be pre-processed with the steps like data cleansing, integration, reduction, data transformation, and discretization.
- The pre-processed data set will be divided into two sets, testing data set and training data set.

ii. Building the model:

- The techniques that we will be using:
  → Linear Regression
- Algorithms that will be used:
  → Neural Networks, Decision Tree based methods.
- Language we will be using:
  → Python

iii. Analyzing:

- Testing dataset will be used to validate the model by measuring the accuracy and other metrics of the model.

iv. Building UI/UX :

- Method willing to be used: Heroku using FLASK with Python

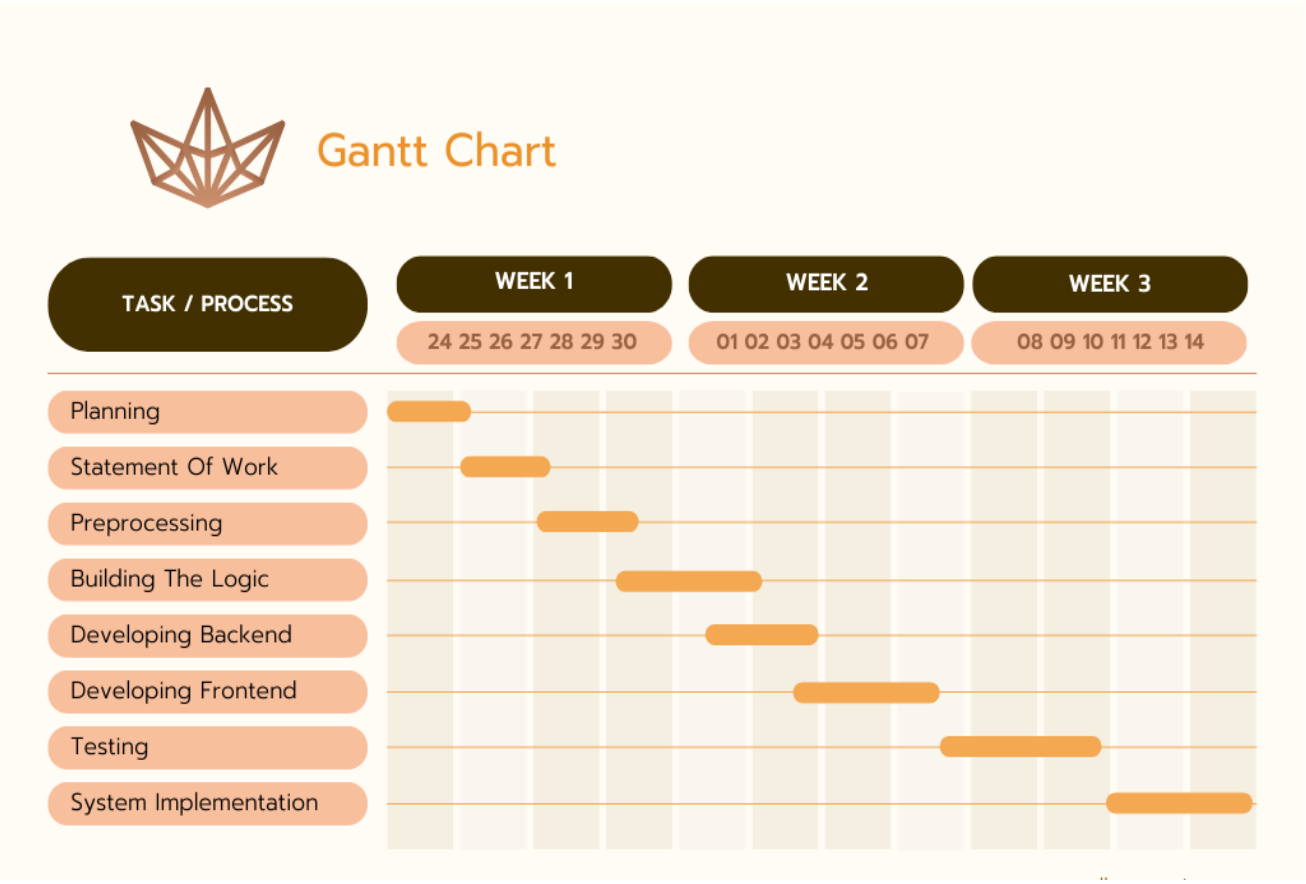
7. Deliverables

Predicting the price of a property in Dubai based on other factors is our goal. The column price in the dataset could be used to train the model to predict the price. Based on this the consumer can assume which property fits their budget.

8. Assumptions

- The source data is accurate throughout.
- Except for the features included in the dataset, none of the data has been impacted by outside factors.
- Outliers are eliminated in the data preprocessing step because they may have an impact on normal data.

9. Project Plan and Grantt Chart



10.Project Team and Responsibilities

IT Number	Name	Responsibilities
IT21029868	B.D.A.D.Hettiarachchi	• Implementing the model • Building and designing the UI • Analyzing & visualization of data • Testing the data • Documentation
IT20708276	U.D.K.Navaratne	• Implementing the model • Analyzing & visualization of data • Building and designing the UI • Testing the data • Documentation
IT21088582	A.N. Elvitigala	• Implementing the model • Analyzing & visualization of data • Building and designing the UI • Testing the data • Documentation
IT21015212	V.L.K .Tennekoon	• Implementing the model • Analyzing & visualization of data • Building and designing the UI • Testing the data • Documentation