Institute of Technology of Cambodia

Department of Applied Mathematics and Statistics (AMS)

Project guideline: Condo for Sale in Cambodia

Lecturer: Mr. CHAN Sophal

Subject: Programming for Data Science

Group's members	ID
SAO Samarth	e20200084
SET Sophy	e20201576
SANG Rithpork	e20200232
SOK Sreyseey	e20201226
SONG Phalla	e20200439

2022~2023

Contents

1. Project Goal:

The goal of this project is to analyze the Condo for Sale Market in Cambodia which condo is the most expansive and where it is, and identify the factors that affect the price of condos. This information can be used by buyers and sellers to make informed decisions about the condo market.

2. Project Objectives:

The specific objectives of this project are to:

- Collect data on Condo for Sale in Cambodia by using web scraping technique

- Analyze the data to identify the factors that affect to predict condo prices.

- Develop a model to predict condo prices.

- Communicate the findings of the project to stakeholders.

3. Project Scope:

The scope of this project includes the following:

- Data collection: The data for this project will be scrapped from **Khmer24** website.

- Data preprocessing: The data will be cleaned by checking missing value, duplicate value, and outlier, and then do EDA.

- Data analysis: The data will be analyzed using statistical and machine learning techniques.

- Model development: A model will be developed to predict condo prices.

4. Project Timeline:

The project will be completed in four phases:

- Phase 1: Data collection (Scraped data)(week 2, week 3 and week 4 of April)

- Phase 2: Exploratory Data Analysis (EDA) (During SEA Games of May)

- Phase 3: Data Analysis (June and week 1 of July)

- Phase 3: Model development (June and week 1 of July)

- Phase 4: Project Development (week 2, week 3 of July)

5. Project Resources:

The following resources needed to complete the project:

- Data: The data for this project was scraped from **Khmer24** website by using beautiful soup and regular expression technique.

- Software: The software that will be used for data analysis and model. development included **Python, seaborn, scikit-learn**.

6. Project members:

The project members consist of:

- Project manager: The project manager will be responsible for overseeing the project and ensuring that it is completed on time and all members work together.

- Data analyst: The data analyst will be responsible for collecting and analyzing the data.

- Model developer: The model developer will be responsible for developing the model to predict condo prices.

- Project Developer: Check every part which has been finished and then discuss to find which part should be adjusted or changed to improve our project accuracy.

7. Project problems:

The following problems have been identified for this project:

- The dataset was hard to scrap (Since data is unstructured data, so we had to use regular expression to help in the process of scraping in order to extract some string text that we want like feature Sub Location, Bedrooms, Bathrooms, and Floors).

- The dataset is incomplete or inaccurate (Have many missing value).

8. Project Mitigation Strategies:

The following mitigation strategies have been developed to address the risks identified for this project:

- The data will be carefully screened for completeness and accuracy.

- The data analysis will be performed by a qualified data analyst.

- The model will be tested on a holdout dataset to ensure that it is accurate.

9. Project Approvals:

The project will be approved by the following stakeholders:

- The project manager

- The data analyst

- The model developer

- The project development specialist