A logo for college computing

Description automatically generated

**Assessment Cover Page**

|  |  |
| --- | --- |
| *Student Full Name* | Derly Milded Montealegre Gonzalez |
| *Student Number* | 2024151 |
| *Module Title*  *Assessment Title* | Machine Learning for Business  CA1 |
| *Lecturer Name* | Dr. Muhammad Iqbal |
| *Assessment Due Date* | 27 October 2024 |
| *Date of Submission* | 27 October 2024 |

**Declaration**

By submitting this assessment, I confirm that I have read the CCT policy on academic misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source.

I declare it to be my own work and that all material from third parties has been appropriately referenced.

I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution.

Contents

[Introduction 2](#_Toc162568515)

[Objective 2](#_Toc162568516)

[Problem Definition 2](#_Toc162568517)

[Scope 2](#_Toc162568518)

[Data Sources 3](#_Toc162568519)

[Ethical Considerations 3](#_Toc162568520)

[References 4](#_Toc162568521)

PROJECTION OF FUTURE HOUSING RELOCATIONS IN THE CITY OF BEIJING

# Introduction

In this project, one of the most evident urban phenomena in the last decades in Beijing is being detected, this is the process of urban renewal of the city center, which CLOU has been managing to renovate inside the Capital Square Beijing and what they want to achieve with this is to reposition and improve the public space. Clou is managing through various processes to illuminate the office lobbies with a modern design language, introducing the concept "*The City Lantern*".

# Objective

In fact, what this project aims to achieve is to generate a notable impact on the real decline and abandonment that is being observed in houses and apartments. Additionally, in this Data it is observed that some houses were built since 1960 and through this it is intended to analyze the number of times these houses have been remodeled to date, according to the structure of the building.

## Problem Definition

Actually, the goal of this project is to be able to determine the perimeter areas that are most experiencing abandonment according to the year of construction and the number of renovations, or any type of repairs that have been made to the buildings in order to find the necessary results on what is the area of the buildings that is showing the greatest deterioration. In this way, measures can be taken to analyze what could be the possible causes that could be prevented so that the infrastructure looks more modern and more useful over the years.

### Scope

Regarding the methodology, according to the information observed it would be said that it is a segmentation methodology, which seeks to group the potential years of

### Data Sources

The database that will be worked on focuses on t

### Ethical Considerations

The project of future housing relocations in the city of Beijing does not have any problem associated with the data that is being taken since all the information will be

# 

# References

Housing price of Beijing from 2011 to 2017, fetching from <https://www.kaggle.com/datasets/ruiqurm/lianjia>

GidHub Link: C:\Users\monte\OneDrive\Documentos\GitHub\capstone-project-feb-2024-ft-derlyai

Assessment Task

El conjunto de datos debe tener un número decente de filas y columnas (por ejemplo, el tipo de variables puede ser categórico, continuo y discreto) después de la limpieza para justificar el uso del modelado de aprendizaje automático acercarse. El tipo de pregunta que debe formular para el proyecto dependerá de la persona elegida del conjunto de datos, como se mencionó anteriormente.

Las preguntas del proyecto podrían ser: (esta es una pequeña muestra sugerida de preguntas, otras preguntas pueden ser más adecuado a su proyecto)

– ¿Cómo medir la similitud o disimilitud entre diferentes clústeres?

– ¿Qué solución de clustering prefieres y por qué?

– Cómo analizar e investigar una tasa de inflación para un producto específico de la serie temporal proporcionada. ¿conjunto de datos?

You will present their findings and defend the results in the report (MS Doc). Your report should capture the following aspects that are relevant to your project investigations. All implementations are carried in Jupyter notebook and no Python files will be accepted.

Una introducción precisa, motivación, descripción del dominio del problema, objetivos del proyecto y la justificación del conjunto de datos elegido en las áreas mencionadas.

¿Qué algoritmos de agrupación consideraría para la segmentación y por qué? Implementar las soluciones y discutir las diferencias entre la puntuación de la silueta y el índice de Davies-Bouldin en el contexto de la agrupamiento. Compare los resultados obtenidos de dos algoritmos de agrupamiento de los conjunto de datos.

Implemente sus soluciones y analice qué ideas puede obtener de la exploración inicial de la

¿Datos de series temporales basados en los temas proporcionados? Describa las tendencias, estacionalidades o anomalías Observado. ¿Cómo se determinaron los parámetros apropiados (p, d, q) para el modelo ARIMA? Evalúe el rendimiento del modelo ARIMA en la previsión de valores futuros, destacando las fortalezas y limitaciones en función del conjunto de datos elegido.

Interpretar y justificar los resultados en función de la especificación del problema o de los objetivos del proyecto mediante el uso de Visualizaciones adecuadas. Comentarios y descripción del código Python, las conclusiones del proyecto deben se especificará en el informe, así como en Jupyter Notebook. Las citas y referencias deben estar en el estilo de Harvard. Grabe un vídeo para explicar y justificar el informe y el código del cuaderno de Jupyter Notebook y la duración del video no debe ser superior a 7 minutos.

Nota: Puede elegir dos conjuntos de datos diferentes para la tarea (ii) o la tarea (iii) por separado o un conjunto de datos para ambas tareas

**Submission Requirements**

Se deben cargar todos los archivos (MS Word, Jupyter Notebook, archivos de datos y cualquier documento de respaldo)

por separado en Moodle.

● Detallar claramente el número de palabras utilizadas en el informe.

● Número de palabras en el informe (1250 palabras +/-10%) excluyendo diagramas, código, referencias, citas

y títulos.

● Debe utilizar el enlace de Github classroom proporcionado (https://classroom.github.com/a/gGcqo8Qz) para

control de versiones. Debes tener al menos 5 confirmaciones en Github antes del envío.

● La rúbrica se proporciona para el desglose detallado de las notas al final de este CA1.

● No se permite el uso de voz de agente de IA en la grabación de video.

● Utilice Harvard Referencing cuando cite material de terceros

● Ser obra propia del alumno.

● Incluir la portada de la evaluación del CCT.

● Presentarse antes de la fecha límite especificada o estar sujeto a multas por presentación tardía

● Debe especificarse claramente el número de palabras utilizadas después de cada sección del informe.

**Acceptable Use of AI for Assignment at CCT**

El uso de herramientas de IA generativa (por ejemplo, ChatGPT, Dall-e, etc.) está permitido en este

Asignación de las siguientes actividades:

○ Lluvia de ideas y refinamiento de sus ideas;

○ Afinar sus preguntas de investigación;

○ Encontrar información sobre su tema;

○ Redactar un esquema para organizar sus pensamientos; y

○ Revisión gramatical y estilo.

● El uso de herramientas de IA generativa no está permitido en este curso para lo siguiente

Actividades:

○ Hacerse pasar por usted en el contexto del aula

○ Completar el trabajo en grupo que su grupo le ha asignado

○ Escribir un borrador de una tarea de escritura

○ Escribir oraciones, párrafos o trabajos completos para completar las tareas de la clase.

● Usted es responsable de la información que envíe en función de una consulta de IA. Su uso de

Las herramientas de IA deben estar debidamente documentadas y citadas.

● Cualquier tarea que se descubra que ha utilizado herramientas de IA generativa de forma no autorizada

estará sujeto a los procedimientos disciplinarios de la universidad como se describe en el Manual de Control de Calidad.

Se podria optar por rellenar los valores faltantes con la media o mediana, pero el enfoque kNN tiende a dar mejores resultados, especialmente cuando las características tienen relaciones no triviales. Por esta razon se decide utilizer un metodo no supervisado.

<https://www.kaggle.com/code/shtrausslearning/bayesian-regression-house-price-prediction>

<https://github.com/derlyai/CA1-Machine-Learning-for-Business.git>