Subject: VT5014CEM

Coursework Title: Report on an Individual Data Science Project

Lecturer: Fung Kai Tat

Student: Hung Yiu Wing

Student ID: 247018266

**找不到目錄項目。**

# Introduction

When buying a home, various factors such as affordability, safety, and access to quality education play a vital role in determining the suitability of a location. In this study, we aim to help friends find suitable areas in the New Territories of Hong Kong, especially New Territories East (NTE) and New Territories West (NTW). We will consider community data to provide a comprehensive assessment of each area.

## Problem Definition

There are several administrative districts in New Territories East and New Territories West, each with its own characteristics. The regions being considered are as follows:

**New Territories East**

* Sha Tin
* Tai Po
* Sai Kung
* North District

**New Territories West**

* Tuen Mun
* Yuen Long
* Kwai Chung
* Kwai Tsing
* Islands District

Our goal is to analyze and compare these areas based on available data so that we can recommend the most suitable areas for homebuyers.

## Dataset Description

The dataset used in this analysis comes from reliable sources, including housing price websites and the Hong Kong Government Open Data Portal. These datasets cover a wide range of community and region-specific characteristics necessary to assess liveability and suitability. The data is summarized as follows:

1. House Prices and Location: House price data for each region, providing insights into affordability.
2. Crime: Crime statistics for each area are used to measure how safe the area is.
3. School distribution: Information on the number and quality of schools within each region to assess educational opportunities.
4. Traffic accidents: This data reflects the overall road safety conditions.
5. Transportation: Information on transportation facilities and accessibility in various regions.
6. Community facilities:

The use and distribution of cultural activities of fire stations, hospitals, libraries, post offices, police stations and shopping centers, museums, showcase recreational and cultural opportunities.

1. Regional Income: The average income level data of each region provides the economic status of the region.

## Justification of data applicability

These datasets are well suited for this task because they cover key factors that influence homebuyers’ decisions, including financial (price), safety (crime rates, traffic accidents), and quality of life indicators (schools, community amenities). By combining these datasets, we can fully understand the characteristics of each region and make data-driven recommendations. The data sources are open to ensure transparency and reproducibility. For example:

* The housing price information comes from a website that provides insights into Hong Kong real estate prices.
* Crime, school, transport and community facilities information comes from the Hong Kong government's open data portal, which provides the latest and most detailed regional characteristics information.

## Target

This study will involve the following steps:

1. Clean the dataset and integrate it into a unified data model.
2. Perform exploratory data analysis (EDA) to identify patterns and relationships among variables.
3. Investigate the statistical relationship between housing prices, crime rates, and school quality.
4. Build a simple recommendation system to rank areas based on their suitability for homebuyers.

At the end of this analysis, we aim to provide actionable recommendations for NTE and NTW regions that balance affordability, safety, and access to quality education while also considering other community factors that enhance livability.