BSc (Hons) Computing (Part-time) 2022 Intake

Report on an Individual Database Development Project	
Module Title:	Data Science for Developers
Module Code:	5014CEM
Lecturer:	Johnny Kwong
Submission Deadline:	27 Dec 2022 at 23:59 pm (HKT)
Student Name:	Ho Chi Leung
Student ID:	227021109
Class (A/B):	Group B

CONTENTS

1.   Introduction

One of my close friends is now looking for a home in the Mong Kok or Tsim Sha Tsui neighborhoods in Hong Kong. The request of my friends is low prices, easy to sell and lease because my friends aim to invest. Within the two cities, I will provide some recommendations for neighborhoods that are a good fit for you based on criteria such as the cost of living, the level of safety, and the number of public facilities. As a point of reference for this writing, I will use the information compiled on the website that monitors changes in housing values. I want to look up information on local concerns on the website of the Hong Kong government, such as the facility of the Leisure and Cultural Services Department and other educational institutions, as well as information regarding safety and public services.

In addition, I will be using some tools to comply my needs. In data finding, I will find data from the websites through Python. It is because the tools can capture different information of the website and convert to CSV file format. Then, I will be analyzing this data through R studio. R studio is strong tool for data handling and data analyzing. It is because R studio support drawing function and easy to operate. It can generate different graph to present data. For example, bar chart, box plot, line graph, etc. Also, R studio can predict the trend of data base on valuable data.

2.   Background

Location 1: Mong Kok
South of Tsim Sha Tsui and closer to the peninsula's heart lies the neighborhood and cultural hub of Mong Kok. It is a poor neighborhood with some wealthy enclaves inside it, and on weekends, the streets of the Mong Kok areas are shut down for a huge night market. The night market's allure to passing visitors stems largely from the fact that it is not a specialty market but rather carries a wide variety of goods. Tourists may taste the Asian haggling tradition in the crowded marketplaces and then escape the madness by ducking into one of the many internationally famous chain shops. This region of Hong Kong has a very distinctive personality due to several factors, including its diverse mix of new and ancient buildings, flashing neon signs, a large number of retailers and restaurants, and the constant ebb and flow of inhabitants as well as tourists. When it comes to sightseeing, eating, and shopping, Mong Kok is a district that is enjoyable to explore at any time of day or night.

On each side of the road are creatures ready to pounce on you and take you off to an untimely death. Spending time in Mong Kok is like falling into a black hole since it demands your attention and resources. In Mong Kok, the usual structure is a multi-story building, and these buildings may be either ancient or new. On the ground level, you will find businesses and restaurants; on the higher floors; you will find offices and apartments. There is a vast range of people, some of whom are older than others, some are younger, some are more impolite than others, some are friendlier than others, some are more well-balanced, and so on. The commercial district of Mong Kok has a selection of ice cream flavors comparable to the variety available to a youngster.

Location 2: Tsim Sha Tsui

Tsim Sha Tsui is an urban region located in southern Kowloon in Hong Kong. The Yau Tsim Mong District is responsible for the area's administration and governance. Tsim Sha Tsui East is a section of land that was reclaimed from Hung Hom Bay and is located to the east of Tsim Sha Tsui. Austin Road encloses the neighborhood to the north, Hong Chong Road and Cheong Wan Road to the east, and Cheong Wan Road to the south. Tsim Sha Tsui is a cape on the Kowloon Peninsula that faces Central and points toward Victoria Harbour. This location is just across from Central. Before the year 1860, when Kowloon was officially handed over to the British Empire, this area was already home to several quaint villages. Tsim Sha Tsui is a Cantonese term that translates to "sharp sandspit." It was also known as Heung Po Tau, which means "port of the incense tree," when it served as a port. Tsim Sha Tsui is a significant tourist hotspot in Hong Kong, and as such, it is home to many upscale stores, bars, pubs, and restaurants geared toward serving visitors. The region is home to a significant number of Hong Kong's museums.

One of the most popular tourist destinations and retail districts in Hong Kong is called Tsim Sha Tsui. It is a lively and bustling area that never stops moving due to the large concentration of hotels, retail businesses, and commercial offices that can be found there. Because of its popularity with high-end shopping tours, Tsim Sha Tsui has only a large number of high-end retail outlets offering jewelry and designer brands at first glance. This is likely because of the high prices of these items. On the other hand, it offers a wide variety of activities the whole family may enjoy together. There is something for everyone to do in Tsim Sha Tsui, whether going for a stroll in Kowloon Park, checking out some museums, shopping, or having a bite to eat.

3.   Data Found

These are data on residences, criminal activities, and public facilities. The data sets in the model come from various government departments and such organizations in Hong Kong as DATA.GOV.HK and Property.HK, the Hong Kong Transport Department, and the Hong Kong Census and Statistics Department. As the investigation progressed, I decided to concentrate my attention on the information obtained from listings in the Hong Kong neighborhoods of Mong Kok and Tsim Sha Tsui, which are, respectively, two of the most prosperous retail districts in the city. To gather information on topics such as building, criminal activity, neighborhood amenities, and school distribution, we relied heavily on the metadata made available by the Census and Statistics Department listings. Thanks to the information provided, I was able to acquire knowledge on various topics, including the cost of homes, the prevalence of crime, and the quality of the public schools in the area.

To organize and clean the data, I will be using R Studio. In my lab book, I will detail my steps to get the data cleaned up and ready for analysis. I will alter the NA values in the renamed columns to ensure that all of the columns corresponding to the various types of vehicles examined include values. I will use the summary command to get a gist of the information in our columns.

I will use Python to get data from the website for the existing house price index and property sales volume. For the tools library, there are using pandas, lxml, openpyxl, ssl of Python to create excel through the table of the website. The data have some requests for nearly five years and the location are MK and TST. The data includes transaction date, price, building price per sq. ft., and area. In that data, I will be using excel to filter some valuable data and remove some invalid data.

For crime data, I get an excel file about crime data from Hong Kong Police Force. Also, I will filter the data through the district, reported crime, and years.

For public facilities, I get the info on the facility from a website. I pick up some facility info about MK and TST. And then, I convert the data to a CSV file and a list of the table to analyze data.

Python coding for get data from website:

```python
import pandas as pd
import lxml
import openpyxl
import ssl
ssl._create_default_https_context = ssl._create_unverified_context

url = "https://852.house/zh/transactions?keywords=&districts%5B%5D=26&districts%5B%5D=31&usages%5B%5D=1&transacted_from=2020-01-01&transacted_to=2022-11-30"
full_table = pd.read_html(url)[0]
full_table.columns = ["transaction date", "purpose", "area", "Estate/Building (Street)","floor/unit", "final price", "Construction area (building price per sq. ft.)", "Salable Area (Usable Square Feet)", "category"]  #custome columns name

for x in range(2, 110):
    url = "https://852.house/zh/transactions?districts%5B0%5D=26&districts%5B1%5D=31&usages%5B0%5D=1&transacted_from=2020-01-01&transacted_to=2022-11-30&page=" + str(x)
    table = pd.read_html(url)[0]
    table.columns = ["transaction date", "purpose", "area", "Estate/Building (Street)","floor/unit", "final price", "Construction area (building price per sq. ft.)", "Salable Area (Usable Square Feet)", "category"]    # custom columns field name
    full_table = full_table.append(table, ignore_index=True)
full_table.to_excel("estate_transaction.xlsx")
```

Result:

transaction date purpose area Estate/Building (Street) floor/unit \ final price Construction area (building price per sq. ft.) \ Salable Area (Usable Square Feet) category

0 2022-11-30 住宅 尖沙咀 港景峰 2座 (廣東道188號) 38樓 A室 $1,860萬 1,253呎 ($14,844) 945呎 ($19,682) 買賣合約

1 2022-11-25 住宅 旺角 雋薈 (豉油街63號) 12樓 B室 $429萬 - 168呎 ($25,535) 買賣合約

2 2022-11-25 住宅 旺角 (山東街21-27號) 3樓 F室 $485萬 - - 買賣合約

3 2022-11-25 住宅 旺角 順景大廈 A座 (渡船街330號) 26樓 8室 $325萬 322呎 ($10,093) 250呎 ($13,000) 買賣合約

4 2022-11-25 住宅 尖沙咀 山林閣 (山林道10-12號) 8樓 A室 $500萬 759呎 ($6,587) 570呎 ($8,771) 樓契

... ...

2720 2020-01-15 住宅　旺角 (廣東道1046號) 6樓 $290萬 －－ 樓契

2721 2020-01-15 住宅　旺角　廣富大廈 (甘霖街38號) 7樓 A室 $400萬 －－ 樓契

2722 2020-01-15 住宅　旺角 利民大樓 (通菜街110號) 7樓 A室 $450萬 －－ 樓契

2723 2020-01-14 住宅 尖沙咀 金冠大廈 (彌敦道66-70號) 5樓 H室 $690萬 －－ 樓契

[2724 rows x 9 columns]

Crime Data:

	Years	District	Report.Crime
1	2020	Reported crimes in Hong Kong	63232
2	2020	Yau Tsim District	5091
3	2020	Mong Kok Police District	4463
4	2021	Reported crimes in Hong Kong	64428
5	2021	Yau Tsim District	4393
6	2021	Mong Kok Police District	4239

Public facility Data:

|	Public.Facility	Count	District
1	Police Station	1	Mong Kok
2	Reporting Center	1	Mong Kok
3	Hong Kong Fire Services Department	1	Mong Kok
4	Hong Kong Government Offices	3	Mong Kok
5	Hong Kong Post Department	3	Mong Kok
6	Hong Kong Public Libraries	1	Mong Kok
7	Street Market	5	Mong Kok
8	Police Station	1	Tsim Sha Tsui
9	Reporting Center	2	Tsim Sha Tsui
10	Hong Kong Fire Services Department	2	Tsim Sha Tsui
11	Hong Kong Government Offices	0	Tsim Sha Tsui
12	Hong Kong Post Department	4	Tsim Sha Tsui
13	Hong Kong Public Libraries	1	Tsim Sha Tsui
14	Street Market	2	Tsim Sha Tsui

4. Exploratory Data Analysis

Existing House Price Index:
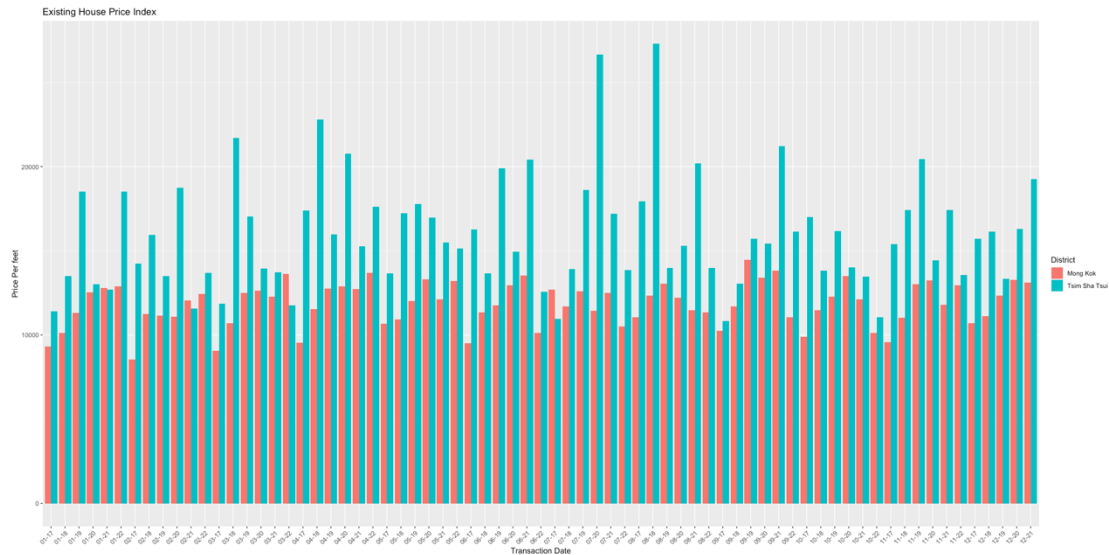

```
install.packages('tidyverse')
library(tidyverse)
hourse_price <- read.csv("price.csv")
ggplot(hourse_price, aes(x = Date, y = PricePerFeet, fill = District)) + geom_col(position = "dodge") +
theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) + ylab("Price Per feet") +
xlab("Transaction Date") + ggtitle("Existing House Price Index")
```

From the graph, we can see that MK's property price was low. Also, The price is more
stable than TST. That means, This is conducive to the preservation of property prices,
and will not be bound by the impact of the general environment and economy. We can
find that the property prices in TST are not stable, and they are expensive. In
comparison, the highest price was twice as expensive as the lowest price. For
investment, there are greater risks.

Property Sales Volume
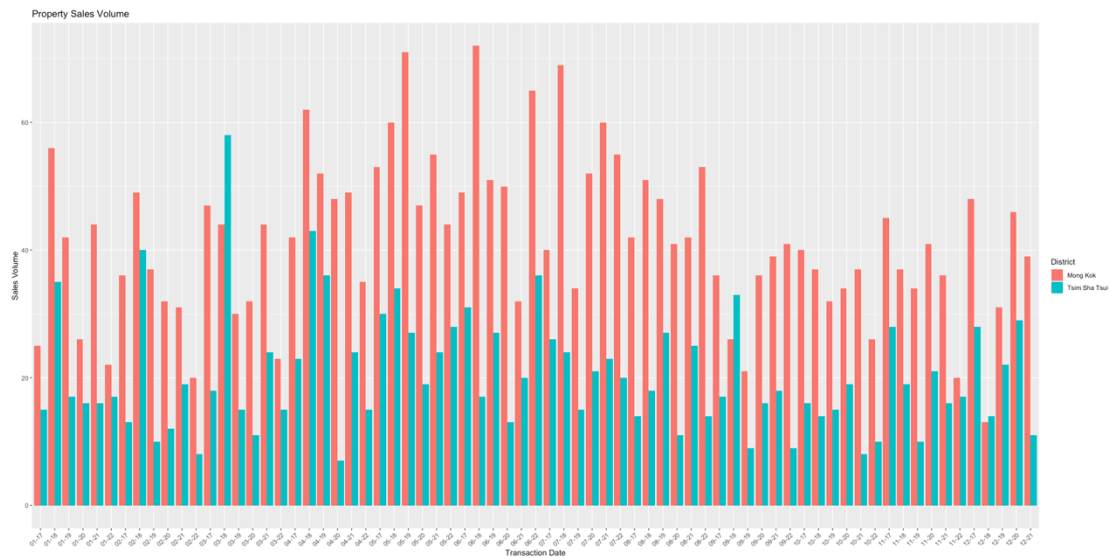

```
install.packages('tidyverse')
library(tidyverse)
hourse_price <- read.csv("price.csv")
ggplot(hourse_price, aes(x = Date, y = SalesVolume, fill = District)) + geom_col(position = "dodge") +
theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1)) + ylab("Sales Volume") +
xlab("Transaction Date") + ggtitle("Property Sales Volume")
```

Regarding the property sales volume, the transaction volume in TST is much lower
than that in MK. Obviously, the number of buyers in TST is down due to high
property prices, while the transaction volume in MK is very considerable, almost
twice that of TST for a while. This means that residences in MK are very popular.
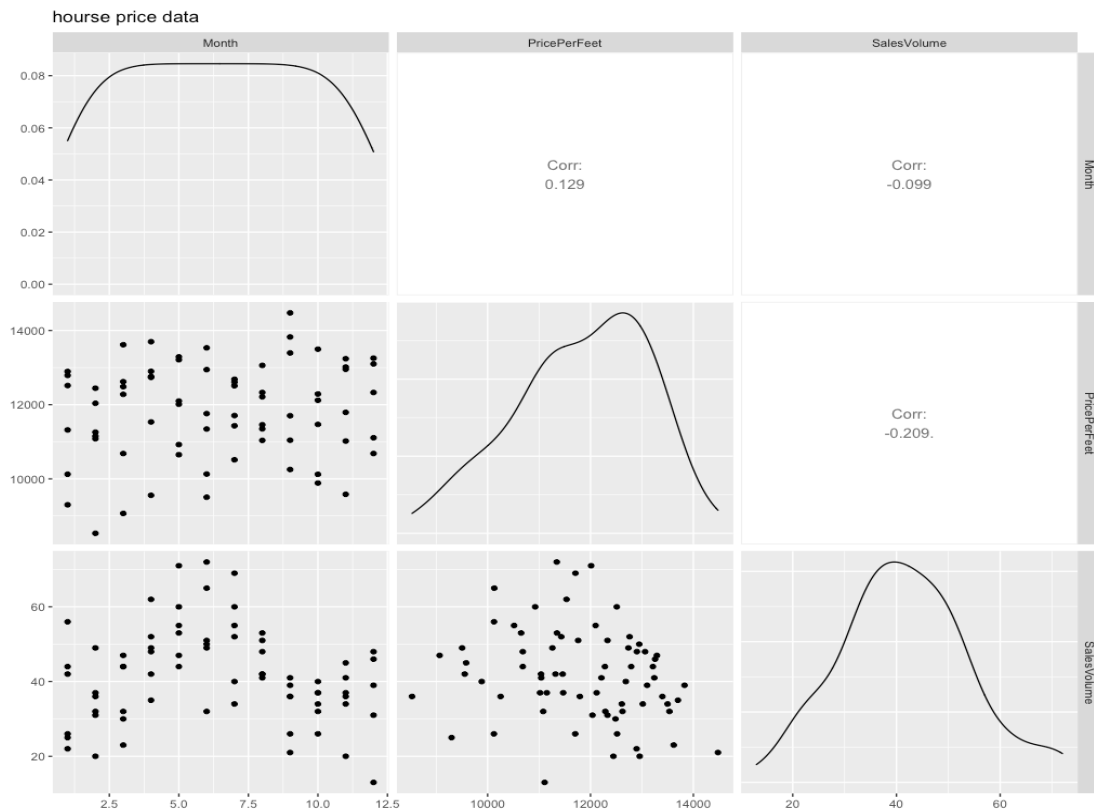
hourse price data

Call: lm(formula = PricePerFeet ~ SalesVolume, data = hourse_price)

Residuals:

Min	1Q	Median	3Q	Max
-5184	-1816	-423	1141	12150

Coefficients:

	Estimate Std.	Error	t value	Pr(>|t|)
(Intercept)	16820.95	564.89	29.778	< 2e-16 ***
SalesVolume	-93.54	16.33	-5.729	5.94e-08 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2970 on 140 degrees of freedom

Multiple R-squared: 0.1899, Adjusted R-squared: 0.1841

F-statistic: 32.82 on 1 and 140 DF,   p-value: 5.943e-08

Coding:

```
install.packages('tidyverse')
library(tidyverse)
hourse_price <- read.csv("price_lm.csv")
mk_price <- hourse_price %>%
    filter(District == "Mong Kok") %>%
    select(Month, PricePerFeet, SalesVolume)
```

```
install.packages('GGally')
library(GGally)
ggpairs(data=mk_price, columns=1:3, title="hourse price data")
model_1 <- lm(PricePerFeet ~ SalesVolume, data = hourse_price)
summary(model_1)
```

For this graph, we can find that the rise and fall of property prices have nothing to do with the transaction volume. In addition, the month may be little effect on property prices. So, we cannot predict the trend of price based on sales volume because the model is unreliable. In the summary function, multiple R-squared and p-value show that the model is not good.

Crime Data
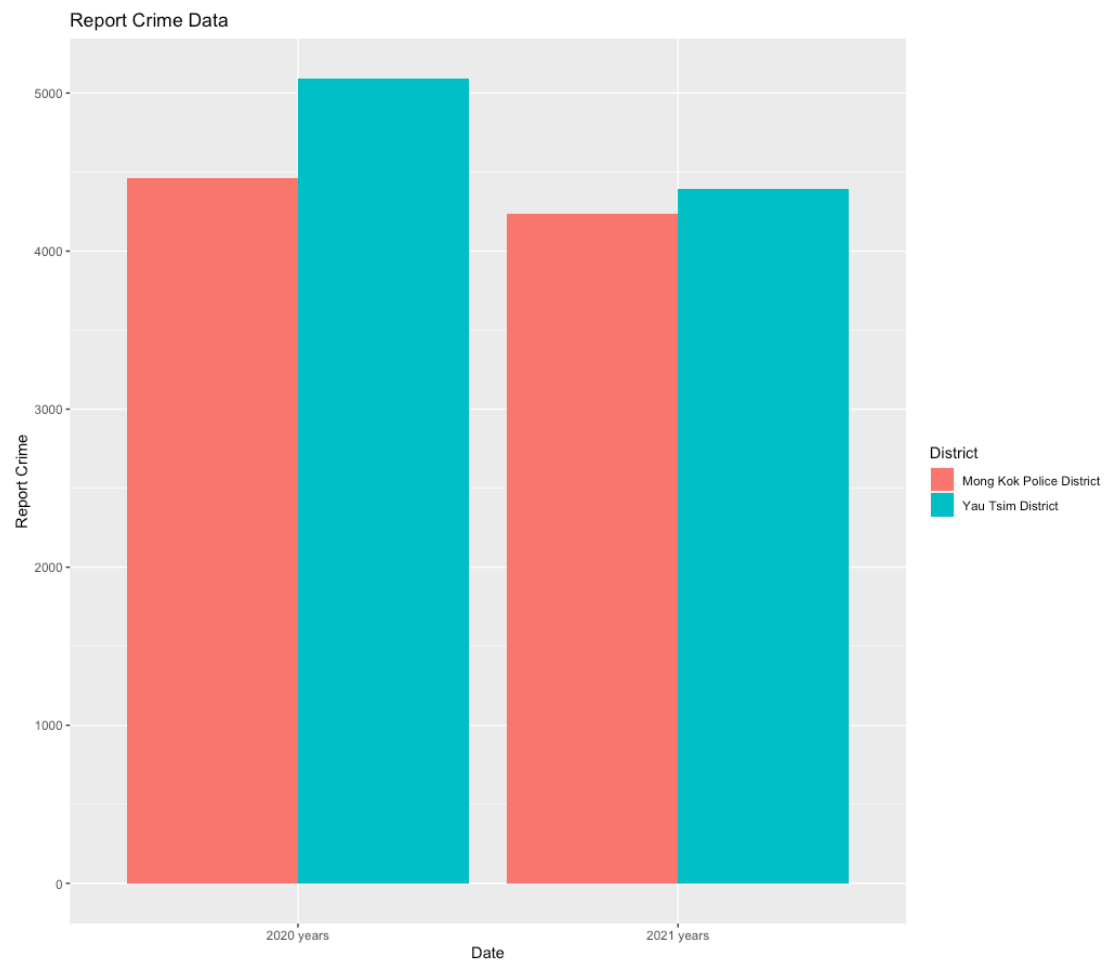

```
install.packages('tidyverse')
library(tidyverse)
crime <- read.csv("crime_data.csv")
ggplot(crime, aes(x = Years, y = ReportCrime, fill = District)) + geom_col(position = "dodge") +
ylab("Report Crime") + xlab("Date") + ggtitle("Report Crime Date")
```

About the crime data, I was using report crime data because report crime numbers can show the need of people in the district. The crime data shows that the number of reported crimes is on a downward trend, indicating that public security in the two districts are getting better. In compare, the number of MK is 5% less than TST. The show that the security of MK district is better than TST.

Public facilities

Facilities	MK	TST
Police Station	1	1
Reporting Center	1	2
Hong Kong Fire Services Department	1	2
Hong Kong Government Offices	3	0
Hong Kong Post Department	3	4
Hong Kong Public Libraries	1	1
Street Market	5	2

The table shows that the reporting center of TST is one more than MK, and it is easier for people to seek help from the police. In addition, the number of fire stations also shows that it is easier for firefighters to rush to the scene to solve the disaster in the TST district when a fire breaks out. But, TST district does not have any Hong Kong government offices, and the street market is less than MK district. Therefore, it is more troublesome for people to apply for some passport or ID card. MK has more street markets and is more suitable for a family to live in because many street markets make it easier for people to buy cheaper products.

5. Recommendation

Item	Factors	Rating	MK	TST
A	House Price Index	*0.6*	✓	
B	Property Sales Volume	*0.2*	✓	
C	Crime Trend	*0.1*	✓	
D	Public facilities	*0.1*	✓	
		1		

First of all, I suggest to buy a flat in MK district based on the above analysis item. As the analysis, property price is the most significant factor affecting the recommendation. Sales volume is next, and the crime rate and public facilities. The level of property prices is the key to attracting the public to buy, and the volatility of property prices affects the appreciation potential of the area, but it is important that stable volatility is needed for investors. And the affordable price gives people enough ability to afford it. This can be reflected by the sales volume. People are more willing to buy buildings in the MK district. Compared with the transaction volume of TST district. In some cases, MK district is twice that of TST district. The trading volume also reflects active buying and selling, which means that rushing to cash out funds or invest is good. In addition, MK district is a good choice whether it is for residential or rental. The low crime rate indicates the quality of public security of area high than TST district. The facilities are also relatively complete so it can meet the needs of residents in the MK district. Overall, I suggest my friends buy a flat in MK district.

6.   Conclusion

In data analysis, I have followed the house price index, property sales volume, crime data, and facilities to analyze both districts. We can find out which one is better through graphs of analysis from R studio. Finally, MK district was better than TST district because MK district can match the requirements of my friends.

In addition, we cannot predict the trend of price because data finding problem. It leads we cannot analyze the potential of the district. In future, we can find enough data to analyze the trend of price to give better suggestion.

7. Reflection

In this data research report, I learned how to find the data. We can use different ways to get data from websites and companies. For example, download from the website, capture data through Python, etc. Also, I know how to handle missing data and filter some valuable data to match our needs. In addition, I used R studio tools to help me analyze the data. I can generate some graphs to let people easily read and understand data through tools. It can help me handle some jobs in data analysis. I enjoy this data research report because it can know more information about Hong Kong, house prices, and crime data through research progress.

8. Reference

852.House (n.d.). Transaction trend. Retrieved November 30, 2022, from https://852.house/zh/transactions?districts%5B0%5D=26&districts%5B1%5D=31& usages%5B0%5D=1&transacted_from=2020-01-01&transacted_to=2022-11-30

Crime statistics in detail. Crime Statistics in Detail | Hong Kong Police Force. (n.d.). Retrieved November 30, 2022, from https://www.police.gov.hk/ppp_en/09_statistics/csd.html

Keung, H.-M. (2010). Other amenities in yau-tsim-mong-district. Retrieved November 30, 2022, from http://ihouse.hkedcity.net/~hm1203/links/hk-ytm-amenities.htm

Mong Kok: District Page: Midland Realty. midland. (n.d.). Retrieved November 30, 2022, from https://www.midland.com.hk/en/district/Kowloon-Mongkok-130ND20023

Tsim Sha Tsui: District Page: Midland Realty. midland. (n.d.). Retrieved November 30, 2022, from https://www.midland.com.hk/en/district/Kowloon-Tsim-Sha-Tsui-130ND20016