

Introduction

Welcome to the Industrial Attachment (IA) program, an internship initiative initiated by HKIIT in collaboration with our esteemed IA partner, Greate (HK) Limited. This IA program focuses on providing students with valuable hands-on experience in the field of Big Data Analysis. We are proud to cooperate with Greate (HK) Limited, a renowned organization specializing in Big Data Analysis, whose expertise and guidance will be instrumental in making this IA program a success. The IA program will be conducted throughout the month of August, allowing students to fully immerse themselves in the project and accumulate a minimum of 100 working hours. Students will have the opportunity to select from a range of projects within the field of Big Data Analysis, tailoring their experience to align with their interests and learning objectives. Through this IA program, we aim to empower students, foster industry-academic collaboration, and contribute to the advancement of Big Data Analysis.

Bear in mind

Industrial Attachment (IA) Program Guideline

1. Group Formation and Project Selection:
 - I. Students are required to form groups of three or four and choose one project to work on collectively.
 - II. Detailed project information can be found in the attached appendix.
2. Attendance and Reporting Time:
 - I. Students must return to the campus promptly at 10:00 am and mark attendance daily.
3. Working Basic Informaiton:
 - I. The IA should start on **1st August 2025**, and terminate on **31st August 2025**, the **last day** depends on student working hours and performance. We would consider to early finish the IA if student have good discipline and performance during the IA.
 - II. The working hours for the IA program are from 9:30 am to 5:30 pm, Monday to Friday. You have to finish your work in **Room TYC511**.
 - III. Students are expected to adhere to the provided schedule in this document, ensuring punctuality for all activities.
4. Absence or Lateness Notification:
 - I. If a student plans to be absent or will be late, they must inform their workplace supervisor one day in advance.
5. Final Analytical Product:
 - I. The final deliverable for the IA program should be a Tableau workbook, showcasing the results of the data analysis conducted during the project.
6. Data and Information Resources:
 - I. All the necessary data required for the IA program has already been posted on Moodle.
 - II. Students should access the provided resources to gather the relevant data for their projects.
7. IA Workbook:
 - I. All students are required to complete the IA workbook, ensuring that all necessary information is filled in accurately.
8. Workplace Supervisors:
 - I. The IA program workplace supervisors are **Benson Lau** and **Blue Ho**. Students should reach out to them for guidance, support, and feedback throughout the program.

9. The following is the student list of this IA:

Name	Team	Case
NG Ping Yam	--	--
CHENG Kong Sang	1	Case 2021
CHEUNG Wang Kwong	1	Case 2021
KEUNG Pui Shing	1	Case 2021
CHENG Chun Hei	2	Case 2223
LUI Kwok Fung	2	Case 2223
NG Wang Hon	2	Case 2223
AQSA-AHMED	3	Case2324
LUI Wing Sing	3	Case2324
TSANG Man Chiu	3	Case2324
HE Xu	4	Case 2021
KUANG Chin Sing	4	Case 2021
LAM Ho Ching	4	Case 2021
GAN Zechun	5	Case 2526
ZHANG Shuang	5	Case 2526
ZHANG Weien	5	Case 2526

Please be sure to review this guideline thoroughly and reach out to your workplace supervisors or program coordinators if you have any further questions or need clarification. We wish you a successful and rewarding experience during the IA program!

Schedule

You have to follow the below schedules:

Date	To-Do list	Accumulated working hour
1 st August 2025	Form Group and Project selections Project start Project Briefing Meeting Fill in the information in the IA Handbook	8
4 th August 2025	Report the Project Planning Project On Progress	16
Date	To-Do list	Accumulated working hour
5 th – 8 th August 2024	Project On Progress Preliminary view of data Progress Report Meeting on 8 th or 9 th August	56
14 th – 18 th August 2024	Project On Progress Advanced analysis of the project (dashboarding) Progress Report Meeting on 15 th or 16 th August	96
21 th August 2024	Final Presentation Fill in the information in the IA Handbook and submit to Moodle submission box	104

Project Descriptions

There are six projects that are across various industries and data. All Project cases have to retrieve knowledge from the data and presentation to supervisor. The Final product of each project should be a Tableau Public workbook with a “.twbx” file extension.

To Conduct the project smoothly, you may have some questions for your starting point, and the following question is only an **example**:

1. Comparison between two regions of the video?
2. What is the audience interested in?
3. What is Youtube used for?
4. Is virtual Youtube popular in the western world?

Project Case 2324

The data is about LA Crime Data in the past 13 years. All groups and individuals use the provided data. To Conduct the project smoothly, you may have some questions for your starting point, and the following question is only an **example**:

1. What are the Crime Change in these 10 Years?
2. A Safety Instruction for tourists?
3. Which area is avoid?
4. What are the circumstances of crime in LA?

Project Case 2425

The data is about the transaction data of British Property. All groups and individuals use the provided data and finish the following tasks: Data preparation, Mandatory task, Advanced technique task, and Finding demonstrations. All tasks and demonstrations are required to utilise the features of Tableau Public and R for data pre-processing.

To Conduct the project smoothly, The project topic should not be limited to the following suggestions but should included in the theme of “**When Britain parallel with Pandemic**”:

1. Which type of property is popular in the UK during the Pandemic?
2. How is the pandemic affecting property transactions?
3. Which kind of property will become the next trend?

Project Case 2324 Full Document

Project Overview

You are required to form a group of 2~ 3 classmates or even work individually. Each group or individual will receive one set of data. All groups or individuals have to process the dataset and generate some findings. The data is about LA Crime Data in the past 13 years. All groups and individuals use the provided data.

To Conduct the project smoothly, you may have some questions for your starting point, and the following question is only an **example**:

1. What are the Crime Change in these 10 Years?
2. A Safety Instruction for tourists?
3. Which area is avoid to visit?
4. What are the circumstances of crime in LA?

Dataset Description

Youtube video record

This is open data from the data organisation Kaggle.com. The data contain the details of each crime case from 2010 to 2023. The following is a sample of the dataset.

DR_NO	Date Rptd	DATE OCC	TIME OC	AREA	Nr Rpt	Dist	Part 1-2	Crn Cd	Crn Cd L1	Mocodes	Vict Age	Vict Sex	Vict Desc
10304468		1/8/2020 0:00	1/8/2020 0:00	2230	3	Southwest	377	2	624	BATTER	0444 0913	36 F	B
190101086		1/2/2020 0:00	1/1/2020 0:00	330	1	Central	163	2	624	BATTER	0416 1822	25 M	H
200110444	04/14/2020 12:00:00 AM	02/13/2020 12:00	1200	1	Central	155	2	845	SEX OFF	1501	0	X	X
191501505		1/1/2020 0:00	1/1/2020 0:00	1730	15	N Hollyw	1543	2	745	VANDAI	0329 1402	76 F	W
191921269		1/1/2020 0:00	1/1/2020 0:00	415	19	Mission	1998	2	740	VANDAI	329	31 X	X
200100501		1/2/2020 0:00	1/1/2020 0:00	30	1	Central	163	1	121	RAPE	FC0413 1822	25 F	H
200100502		1/2/2020 0:00	1/2/2020 0:00	1315	1	Central	161	1	442	SHOPLIF	1402 2004	23 M	H
200100504		1/4/2020 0:00	1/4/2020 0:00	40	1	Central	155	2	946	OTHER	N1402 0392	0	X
200100507		1/4/2020 0:00	1/4/2020 0:00	200	1	Central	101	1	341	THEFT-G	1822 0344	23 M	B
201710201	05/19/2020 12:00:00 AM	05/26/2020 12:00	1925	17	Devonshir	1708	1	341	THEFT-G	1300 0202	0	X	X
200100509		1/4/2020 0:00	1/4/2020 0:00	2200	1	Central	192	1	330	BURGLA	1822 1414	29 M	A
200100510		1/5/2020 0:00	1/5/2020 0:00	955	1	Central	111	2	990	CRIMIN	0421 0906	35 M	O
200100514		1/5/2020 0:00	1/5/2020 0:00	1355	1	Central	162	1	341	THEFT-G	1822 0344	41 M	A
200100515		1/7/2020 0:00	1/7/2020 0:00	1638	1	Central	162	1	648	ARSON	1402 1501	0	X
200100520		1/8/2020 0:00	1/8/2020 0:00	1805	1	Central	128	1	442	SHOPLIF	0325 1402	24 F	H
211916029	11/25/2021 12:00:00 AM	11/30/2020 12:00	730	19	Mission	1916	2	626	INTIMAT	12000 1814	24 F	H	
201116159	11/29/2020 12:00:00 AM	11/28/2020 12:00	2018	11	Northeast	1124	2	626	INTIMAT	10400 0416	34 F	H	
200506268	02/22/2020 12:00:00 AM	02/22/2020 12:00	1900	5	Harbor	511	1	440	THEFT	P10319 0344	29 F	W	
210916801	11/22/2021 12:00:00 AM	11/19/2020 12:00	1200	9	Van Nuys	932	2	354	THEFT	O 1501 1822	46 M	B	
200100535	01/14/2020 12:00:00 AM	01/14/2020 12:00	1330	1	Central	132	1	210	ROBBER	0416 0411	66 M	B	
200914517	9/10/2020 0:00	9/9/2020 0:00	1735	9	Van Nuys	909	2	354	THEFT	O 0377 1822	40 M	O	

Data Description

The following is the Data Description from the Open Data website, and this information may help you summarise the data dictionary for this project.

Data item	Explanation (where appropriate)
DR_NO	Official file number
DateRptd	Date of the Crime Reported
DATE OCC	The date of the Crime occurred
TIME OCC	Time of The crime occurred in 24 24-hour military time format
AREA NAME	The 21 Geographic Areas or Patrol Divisions are also given a name designation that references a landmark or the surrounding community that it is responsible
Crm Cd Desc	Description of the Crime
Vict Age	Age of the Victim
Weapon Desc	Weapon Used
LAT	Latitude of the Crime
LON	Longitude of the Crime

Task

Data Preparation

You will receive 13 Years of Crime data for your project, you have to ETL (using **R** or **Python**) before pouring the data into Tableau:

Stage 1

1. The dataset originally divided into two timelines, 2010 to 2019 and 2020 to now(2023). You must combine them to single dataset file.
2. The Date Data are messy, For example the 12:00:00 AM in the Report Date and occurred Date was useless, you have to extract the date from the date string.

Before

Date Rptd
1/8/2020 0:00
1/2/2020 0:00
04/14/2020 12:00:00 AM
1/1/2020 0:00
1/1/2020 0:00

After

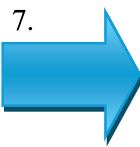
Date.Rptd	DATE.OCC
20/2/2010	20/2/2010
13/9/2010	12/9/2010
9/8/2010	9/8/2010

3. Convert the occurred time military hour to standard time format and combine with occurred date

4. Before

DATE OCC	TIME OC
1/8/2020 0:00	2230
1/1/2020 0:00	330
02/13/2020 12:00	1200

5.



6. After

DATE.OCC	TIME.OC
20/2/2010	13:50
12/9/2010	0:45
9/8/2010	15:15+

8.

Occ_DateTime
20/2/2010 13:50
12/9/2010 0:45
9/8/2010 15:15

4. Fill "No weapon Recorded" to all missing value in Weapon Description

Before

Weapon Desc
STRONG-ARM (HANDS, FI
UNKNOWN WEAPON/OTH
ILLING (APARTMENT, DUF
STORE
UNKNOWN WEAPON/OTH
ORE

After

Weapon.Desc
No Weapon Recorded
No Weapon Recorded
No Weapon Recorded

Stage 2 (Open end)

1. Select all necessary columns for your project

Before

AREA	AREA_N	Rpt Dist	Part 1-2	Crm Cd	Crm Cd 1	Mocodes	Vict Age	Vict Sex	Vict Deso	Premis Cd	Premis De	Weapon U	Weapon Desc	Status	Status Des	Crm Cd 1	Crm Cd 2	Crm Cd 3	Crm Cd 4	LOC
2230	3	Southwest	377	2	624	BATTER 0444 0913	36 F	B	501 SINGLE I	400	STRONG-ARM (HANDS, F/AO	Adult Oth	624							1100 V
330	1	Central	163	2	624	BATTER 0416 1822	25 M	H	102 SIDEWAI	500	UNKNOWN WEAPON/OTHIC	Invest Cor	624							700 S
1200	1	Central	155	2	845	SEX OFF	1501	0 X	X	726	POLICE FACILITY	AA	Adult Arm	845						200 E
1730	15	N Hollyw	1543	2	745	VANDAI 0329 1402	76 F	W	502 MULTI-UNIT DWELLING (APARTMENT, DUF	IC	Invest Cor	745	998							5400
415	19	Mission	1998	2	740	VANDAI	329	31 X	X	409	BEAUTY SUPPLY STORE	IC	Invest Cor	740						14400
30	1	Central	163	1	121	RAPE, F(0413 1822	25 F	H	735	NIGHT C	500	UNKNOWN WEAPON/OTHIC	Invest Cor	121	998					700 S
1315	1	Central	161	1	442	SHOPLIF 1402 2004	23 M	H	404	DEPARTMENT STORE	IC	Invest Cor	442	998						700 S
40	1	Central	155	2	946	OTHER N 1402 0392	0 X	X	726	POLICE FACILITY	IC	Invest Cor	946	998						200 E
200	1	Central	101	1	341	THEFT-G 1822 0344	23 M	B	502	MULTI-UNIT DWELLING (APARTMENT, DUF	IC	Invest Cor	341	998						700
1925	17	Devonshir	1708	1	341	THEFT-G 1300 0302	0 X	X	203	OTHER BUSINESS	AO	Adult Oth	341							11900
2200	1	Central	192	1	330	BURGLA 1822 1414	29 M	A	101	STREET	306	ROCK/THROWN OBJECT	IC	Invest Cor	330					15TH
955	1	Central	111	2	930	CKIMIN/0421 0906	35 M	O	108	PARKIN	511	VERBAL THREAT	IC	Invest Cor	930					800 N
1355	1	Central	162	1	341	THEFT-G 1822 0344	41 M	A	503	HOTEL	500	UNKNOWN WEAPON/OTHIC	AA	Adult Arm	341					800 S
1638	1	Central	162	1	648	ARSON 1402 1301	0 X	X	404	DEPART	500	UNKNOWN WEAPON/OTHIC	Invest Cor	648	998					700 W
1805	11	Central	128	1	442	SHOPLIF 0325 1402	24 F	H	252	COFFEE SHOP (STARBUCKS, COFFEE BEAN, IC	Invest Cor	442								100 S



After

DR_NO	Date.Rptd	DATE.OCC	TIME.OCC	AREA_N	Crm.Cd	Vict.Age	Vict.Sex	Vict.Deso	Premis.De	Weapon.Desc	Status.Des	LAT	LO
1307355	20/2/2010	20/2/2010	13:50	Newton	VIOLATI	48 M	H	SINGLE I	No Weapon	Recorder	Adult Arr	33.9825	-
11401303	13/9/2010	12/9/2010	0:45	Pacific	VANDAI	0 M	W	STREET	No Weapon	Recorder	Invest Cor	33.9599	-1
70309629	9/8/2010	9/8/2010	15:15	Newton	OTHER M	0 M	H	ALLEY	No Weapon	Recorder	Invest Cor	34.0224	-1
90631215	5/1/2010	5/1/2010	1:50	Hollywood	VIOLATI	47 F	W	STREET	HAND GUN		Invest Cor	34.1016	-
1E+08	3/1/2010	2/1/2010	21:00	Central	RAPE, A	47 F	H	ALLEY	STRONG-ARM (H/	Invest Cor	34.0387	-1	
1E+08	5/1/2010	4/1/2010	16:50	Central	SHOPLIF	23 M	B	DEPARTIN	No Weapon	Recorder	Adult Arr	34.048	-1
1E+08	8/1/2010	7/1/2010	20:05	Central	BURGLA	46 M	H	STREET	No Weapon	Recorder	Invest Cor	34.0389	-1
1E+08	9/1/2010	8/1/2010	21:00	Central	ASSAUL	51 M	B	OTHER F	UNKNOWN WEAF	Adult Arr	34.0435	-1	
1E+08	9/1/2010	9/1/2010	2:30	Central	ASSAUL	30 M	H	PARKIN	(STRONG-ARM (H/	Invest Cor	34.045	-1	
1E+08	9/1/2010	6/1/2010	21:00	Central	THEFT-G	55 M	W	OTHER F	No Weapon	Recorder	Invest Cor	34.0538	-1
1E+08	14/1/2010	14/1/2010	14:45	Central	BATTER	38 F	B	STREET	STRONG-ARM (H/	Invest Cor	34.064	-1	

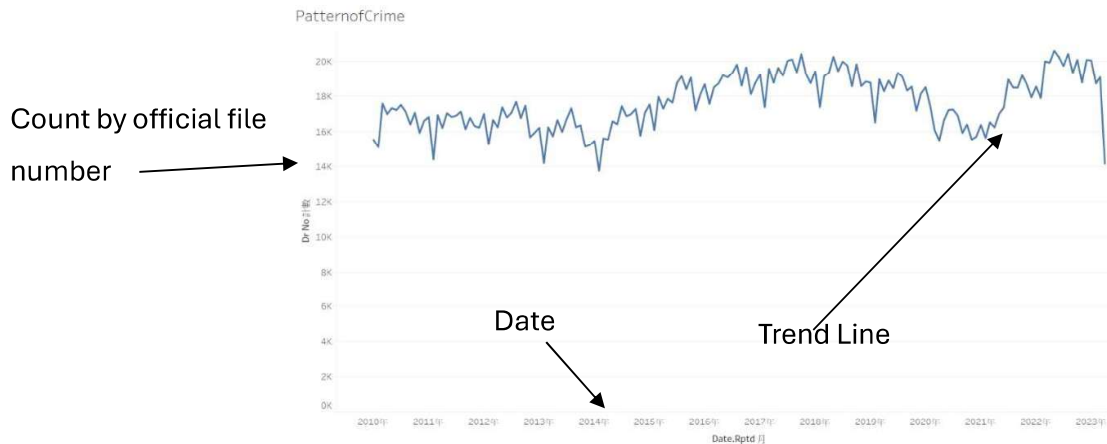
Stage 3

1. Save your cleared data in CSV or Excel format named "GroupName_Crime_Data".

Mandatory task

Visualize the following requirements, the below plotting is only for your reference.

1. The time series trend line on the number of Crimes among years



2. Proportion of Victim gender.

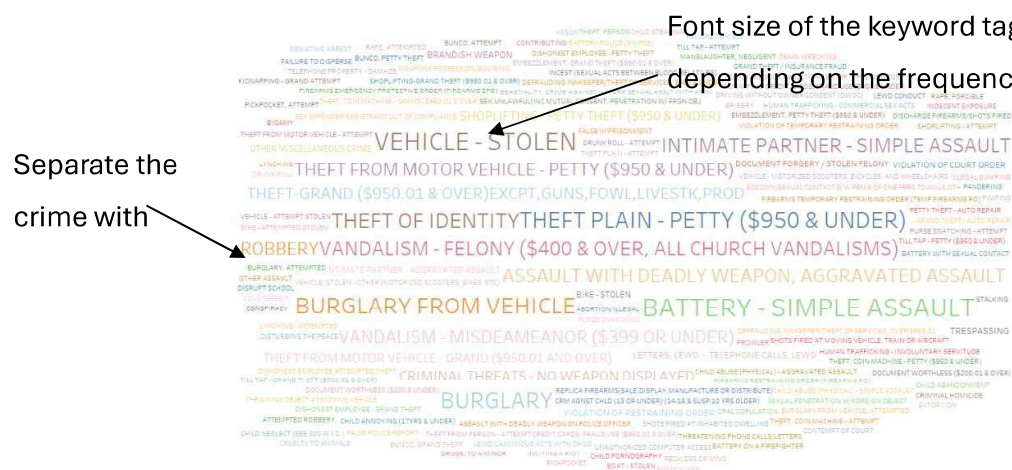
Follow the below table to convert the gender code to the description by Tableau calculated field

Code	Description
F	Female
H	Hybrid
M	Male
N	No tell
X,-,Null	unknown

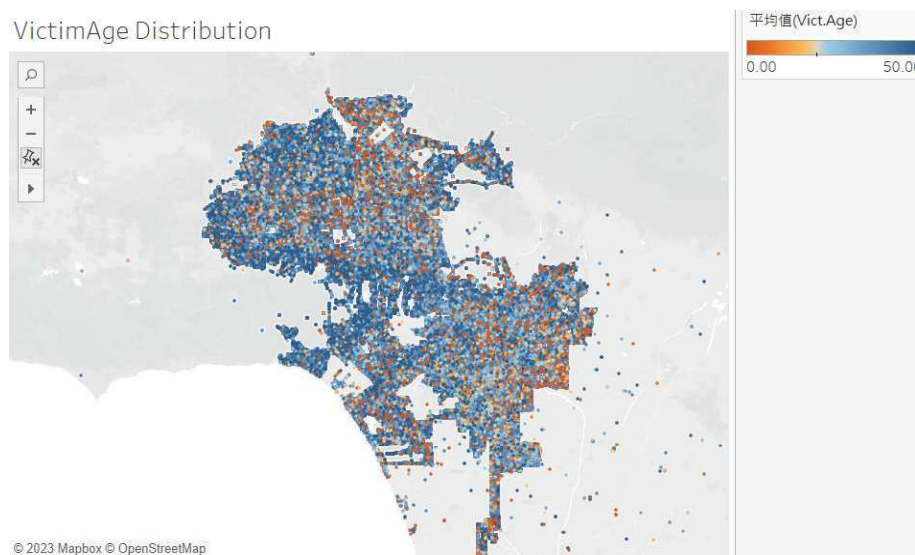


3. Crime Keyword Cloud

Font size of the keyword tags is depending on the frequency of



4. Crime Map (coloured by Age)

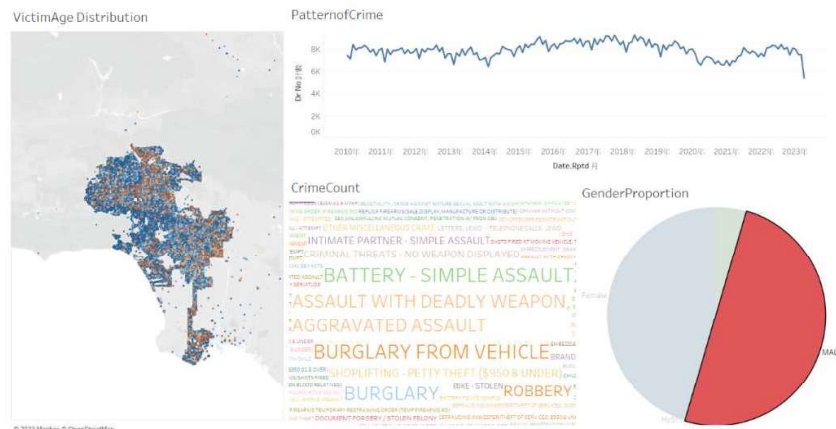


- Crime count by weekday (additional mark for visualization average count by using tableau only)



Advanced technique task

This task is a freestyle from your idea. First, you have to create dashboards for your project purpose. Remember that you must complete one of the dashboards by using visualised plottings from the last section and drilling some valuable findings. Furthermore, you still have to plot some graphs not mentioned in the previous section and consolidate them as a story-telling dashboard. Below is a dashboard for your reference. You may follow the guideline of Steve Few's informative dashboards for designing a good dashboard. Remember, a good dashboard is not an ornament; it should be a tool for assisting your findings. PLEASE DON'T USE A SINGLE TYPE OF GRAPH FOR YOUR DASHBOARD.



Finding demonstrations

You should present your idea/findings through the story mode in 15 mins. No PowerPoint or other presentation software is allowed. The presentation should make use of the following Tableau features:

1. Story mode
2. Dashboard
3. Filtering data
4. Drill-in data
5. Story-telling
6. Presentation Mode
7. Single Graph sheet

Furthermore, you also have prepared a report for our record. You may find the report template in Moodle. The Report have to include the following parts:

1. Introduction of project
2. Data Dictionary
3. Findings
4. Suggestions
5. Reference (if any)