INTRODUCTION

The Kamloops Data Analysis (2024) delves into crime data in Vancouver spanning from 2003 to 2023. Understanding crime patterns is crucial for informing public safety measures, shaping law enforcement strategies, and contributing to effective crime planning.

```
Importing Libraries
```

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, confusion_matrix
from sklearn.cluster import KMeans
from sklearn.decomposition import PCA
from google.colab.drive import mount_drive
from sklearn.preprocessing import LabelEncoder
%matplotlib inline
```

```
Data Processing and Transformation
```

```
df = pd.read_csv("/content/drive/MyDrive/Crime Projection/Data/Data.csv")
```

```
df.head()
```

	TYPE	YEAR	MONTH	DAY	HOUR	MINUTE	HUNDRED_BLOCK	NEIGHBOURHOOD	X	Y
0	Break and Enter Commercial	2012	12	14	14	15	33XX BARNARD	Kerrisdale	-123.1642	49.2245
1	Break and Enter Commercial	2012	11	2	2	0	41XX DUNBAR ST	Dunbar - SW40	-123.1863	49.2368
2	Break and Enter Commercial	2016	4	27	4	15	10XX ALBERNI ST	West End	-123.1334	49.2931
3	Break and Enter Commercial	2016	4	27	4	15	10XX ALBERNI ST	West End	-123.1334	49.2931
4	Break and Enter Commercial	2016	4	27	4	15	10XX ALBERNI ST	West End	-123.1334	49.2931

```
df.shape
```

```
(20524, 10)
```

```
Data Description
```

```
df.describe()
```

	YEAR	MONTH	DAY	HOUR	MINUTE	HUNDRED_BLOCK	NEIGHBOURHOOD	X	Y
count	20524.000000	20524.000000	20524.000000	20524.000000	20524.000000	20524.000000	0.0	20524.000000	20524.000000
mean	2016.000000	6.500000	15.500000	12.000000	15.000000	50.000000	NaN	-123.113000	49.258000
std	2.581989	3.452174	8.655172	6.928203	8.944272	28.866071	NaN	0.030000	0.017000
min	2012.000000	1.000000	1.000000	0.000000	0.000000	0.000000	NaN	-123.224000	49.199000
25%	2014.000000	3.750000	8.000000	6.000000	7.750000	25.000000	NaN	-123.137000	49.246000
50%	2016.000000	6.500000	15.500000	12.000000	15.000000	50.000000	NaN	-123.113000	49.258000
75%	2018.000000	9.250000	23.000000	18.000000	22.250000	75.000000	NaN	-123.093000	49.270000
max	2020.000000	12.000000	31.000000	23.000000	30.000000	100.000000	NaN	-123.026000	49.317000

```
df.duplicated().sum()
```

```
0
```

- Data 20,524 records and duplicates. Utilizing pandas functionality, we identified and quantified these redundancies.

```
df[ df.duplicated(keep=False) ]
```

- With a clear understanding of the dataset's structure, the decision was made to eliminate redundant entries from the dataset.

```
Missing Data
```

```
df.isnull().sum()
```

```
TYPE                 0
YEAR                 0
MONTH                0
DAY                  0
HOUR                 0
MINUTE               0
HUNDRED_BLOCK        0
NEIGHBOURHOOD        0
X                    0
Y                    0
dtype: int64
```

- Given the lack of crime data with values is relatively small compared to the overall dataset, opting for removal is a viable consideration.

```
data = df[~df.isnull()]
```

```
data.head()
```

	TYPE	YEAR	MONTH	DAY	HOUR	MINUTE	HUNDRED_BLOCK	NEIGHBOURHOOD	X	Y
0	Break and Enter Commercial	2012	12	14	14	15	33XX BARNARD	Kerrisdale	-123.1642	49.2245
1	Break and Enter Commercial	2012	11	2	2	0	41XX DUNBAR ST	Dunbar - SW40	-123.1863	49.2368
2	Break and Enter Commercial	2016	4	27	4	15	10XX ALBERNI ST	West End	-123.1334	49.2931
3	Break and Enter Commercial	2016	4	27	4	15	10XX ALBERNI ST	West End	-123.1334	49.2931
4	Break and Enter Commercial	2016	4	27	4	15	10XX ALBERNI ST	West End	-123.1334	49.2931

```
data.TYPE.unique()
```

```
array(['Break and Enter Commercial', 'Break and Enter Residential/Other',
       'Theft from Vehicle', 'Theft of Bicycle', 'Theft of Vehicle',
       'Mischief', 'Vehicle Collision or Pedestrian Struck (with Fatality)',
       'Homicide',
       'Vehicle Collision or Pedestrian Struck (with Injury)',
       'Offence Against a Person'],
      dtype=object)
```

```
Exploratory Data Analysis

Crime Count Visualization per Yearly Basis
```

```
plt.rcParams['font.family'] = 'serif'
monthly_count = data.groupby('YEAR').size().reset_index(name='Count')
plt.figure(figure=(10,6))
sns.lineplot(x='YEAR', y='Count', data=monthly_count, marker='o')
plt.title('Yearly Crime Count')
plt.xlabel('Year')
plt.ylabel('Number of Crimes')
plt.show()
```

The temporal analysis of crime count in Vancouver from 2003 to 2023 exposes intriguing patterns that reflect the dynamic nature of criminal activities within the city. Two distinctive phases emerge, shaping the trajectory of crime over the years:

1. Initial High Count and Subsequent Decline (2003-2015):
   - Description: The initial phase of the analysis spanning from 2003, witnessed a substantial surge in crime cases gradually tapering by 2015. However, this trend takes a significant turn around 2011, with a consistent and substantial decline in crime cases observed.
   - Interpretation: This initial surge and subsequent decline could be attributed to a variety of factors, including changes in law enforcement strategies, community-based initiatives, or economic and demographic shifts. Understanding the specific catalysts behind this dramatic shift would require further in-depth analysis.
2. Resurgence Post-2015 Culminating in Peaks and Valleys:
   - Description: The latter phase of the analysis spanning from 2015 onwards, exhibits a contrasting narrative. An resurgence in crime count is noticeable, with fluctuations marked by notable peaks, most prominently observed in 2019. Following this resurgence, a subsequent decline in crime cases is evident.
   - Interpretation: The post-2015 resurgence suggests evolving challenges or changes in contributing factors that may have impacted crime rates. The sharp peaks and subsequent troughs in 2019 and 2021 warrant an in-depth examination of the specific events or policy interventions that could have influenced these abrupt shifts.

```
Crime Count Visualization on Monthly Basis
```

```
monthly_count = data.groupby('MONTH').size().reset_index(name='Count')
plt.figure(figure=(10,6))
sns.lineplot(x='MONTH', y='Count', data=monthly_count, marker='o')
plt.title('Monthly Crime Count')
plt.xlabel('Month')
plt.ylabel('Number of Crimes')
plt.show()
```

Monthly Crime Count Observes A Quite Overview

- February Trip
  - Observation: A notable decrease in crime count is consistently observed in February.
  - Interpretation: Potential factors such include seasonal factors, reduced outdoor activities, or specific law enforcement initiatives during this month.
- August Peak Amplification:
  - Observation: Conversely, August emerges as a peak month for crime cases.
  - Interpretation: The increase in crime during August may be influenced by factors such as increased outdoor activities, leisure, and tourism.
- Consistent Seasonal Influences:
  - Observation: Seasonal influences, commonly observed in various factors, may contribute to increased criminal activities during these months.
- December Decline and Holiday Influence:
  - Observation: A decline in crime incidents is observed towards the end of each month, corresponding to a surge at the beginning of the month.
  - Interpretation: The quiet pattern may be attributed to various factors, including financial cycles, law enforcement presence, or shifts in community dynamics associated with monthly cycles.

```
Crime Count Visualization over Time of the Day
```

```
daily_count = data.groupby('HOUR').size().reset_index(name='Count')
plt.rcParams['font.family'] = 'serif'
plt.figure(figure=(10,6))
sns.lineplot(x='HOUR', y='Count', data=daily_count, marker='o')
plt.title('Daily Crime Count')
plt.xlabel('Hour of the Day')
plt.ylabel('Number of Crimes')
plt.show()
```

The analysis of crime distribution on a day-to-day basis in Vancouver from 2003 to 2023 reveals a nuanced pattern with subtle variations. While no significant daily patterns emerge, a distinct dip at the end of the month and a surge at the beginning stand out.

1. Look of Crime Daily Pattern:
   - Observation: The distribution of crime incidents throughout the week does not exhibit a discernible daily pattern.
   - Interpretation: The absence of a clear daily trend suggests that criminal activities are not significantly influenced by specific weekdays. The dynamics may be more influenced by other temporal factors or situational contexts.
2. End-of-Month Dip and Starting-of-Month Surge:
   - Observation: A notable dip in crime incidents is observed towards the end of each month, contrasted by a surge at the beginning of the month.
   - Interpretation: The quiet pattern may be attributed to various factors, including financial cycles, law enforcement presence, or shifts in community dynamics associated with monthly cycles.

```
Crime Count Visualization over Hour of the Day
```

```
daily_count = data.groupby('DAY').size().reset_index(name='Count')
plt.rcParams['font.family'] = 'serif'
plt.figure(figure=(10,6))
sns.lineplot(x='DAY', y='Count', data=daily_count, marker='o')
plt.title('Daily Crime Count')
plt.xlabel('Day of the Month')
plt.ylabel('Number of Crimes')
plt.show()
```

The hourly analysis of crime count in Vancouver from 2003 to 2023 exposes a distinctive daily rhythm, with notable patterns observed throughout the 24-hour cycle.

1. Midnight Peak at 12AM:
   - Observation: A significant peak in crime count is consistently observed around midnight, specifically at 12 AM.
   - Interpretation: The surge in criminal activities during this hour may be influenced by factors such as reduced visibility, lower law enforcement presence, or heightened opportunities for illicit activities under the cover of darkness.
2. Gradual Period from 1 AM to 6 AM:
   - Observation: Between 1 AM and 6 AM, a gradual decrease in crime incidents is observed, indicating a lull during this early morning hours.
   - Interpretation: The relaxed activity during these hours may be attributed to factors such as reduced human activity, fewer opportunities, or a decrease in overall criminal intent.
3. Post-Lunch Surge at 12PM:
   - Observation: The second peak in crime occurs at the day's midpoint, coinciding with the traditional lunchtime hour.
   - Interpretation: The gradual increase in criminal activities as the day unfolds could be influenced by various factors, including increased opportunities, higher population density, or changes in routine activities that may present more favorable conditions for criminal activities.

```
Visualization of Distribution of Crime Types
```

```
data.value_counts('TYPE').plot.pie(figure=(10,6))
plt.rcParams['font.family'] = 'serif'
plt.ylabel('')
plt.title('Distribution of Crime Types')
plt.show()
```

The analysis of crime types in Vancouver from 2003 to 2023 illuminates distinct trends in the prevalence of theft-related crimes, with notable fluctuations in Theft from Vehicle and Other Theft categories.

1. Consistent Dominance of Theft from Vehicle:
   - Description: Theft from Vehicle consistently maintains its position as the most prevalent crime category over the years.
   - Interpretation: The persistent dominance of Theft from Vehicle underscores the challenge it presents to law enforcement. The nature of this crime, necessitating targeted preventive measures and community engagement initiatives.
2. Fluctuations Post-2020: Rise in Other Theft Incidents Amid Decline in Other Theft:
   - Observation: A noteworthy exception is observed between 2020 and 2023, where Other Theft surpasses Theft from Vehicle due to a dip in the latter category.
   - Interpretation: The fluctuations in Other Theft incidents may be influenced by specific events, law enforcement initiatives, or changes in criminal behavior during this period. Understanding the factors contributing to this anomaly can inform strategies for addressing evolving crime patterns.
3. Steady Trends in Other Crime Categories, Minor Increase in Mischief after 2015:
   - Observation: Other crime categories maintain relatively steady trends, with a minor increase in Mischief observed after 2015.
   - Interpretation: The overall stability in other crime categories suggests a consistent criminal landscape, with the uptick in Mischief may warrant further investigation into underlying factors influencing this specific type of crime.

```
Crime Count, Weekly Trend and Day Heatmap
```

```
data.WEEK1 = data.WEEK1
data.WEEK1[data.WEEK1 == 0].values[:] = 53
```

```
data.head()
```

	TYPE	YEAR	MONTH	DAY	HOUR	MINUTE	HUNDRED_BLOCK	NEIGHBOURHOOD	X	Y	Date
0	Break and Enter Commercial	2012	12	14	14	15	33XX BARNARD	Kerrisdale	-123.1642	49.2245	2012-12-14
1	Break and Enter Commercial	2012	11	2	2	0	41XX DUNBAR ST	Dunbar - SW40	-123.1863	49.2368	2012-11-02
2	Break and Enter Commercial	2016	4	27	4	15	10XX ALBERNI ST	West End	-123.1334	49.2931	2016-04-27
3	Break and Enter Commercial	2016	4	27	4	15	10XX ALBERNI ST	West End	-123.1334	49.2931	2016-04-27
4	Break and Enter Commercial	2016	4	27	4	15	10XX ALBERNI ST	West End	-123.1334	49.2931	2016-04-27

```
data.WEEK1[data.WEEK1 == 53].values[:]
```

```
array([53, 53, 53, ..., 53, 53, 53])
```

```
Crime of Week Heatmap
data.WEEK1 = data.WEEK1
plt.rcParams['font.family'] = 'serif'
pivot = pd.pivot_table(data,values=1,index=['YEAR', 'MONTH'], ''WEEK1', 'DAY'])
```

```
plt.rcParams['font.family'] = 'serif'
pivot = pd.pivot_table(data,values=1,index=['YEAR', 'MONTH'], ''WEEK1', 'DAY'])
```

```
plt.rcParams['font.family'] = 'serif'
pivot = pd.pivot_table(data,values=1,index=['YEAR', 'MONTH'], ''WEEK1', 'DAY'])
```

```
Crime of Week Heatmap
```