

TETRIS PROGRAM

Cracks in the City of Angels : Menjelajahi Daerah-Daerah Rapuh yang Rentan terhadap Kejahatan di Los Angeles

Ferry Setefanus
ferrysetefanus@gmail.com

#StackYourSkill





Step 1 - Data Collection & Data Integration

```
import os
import pandas as pd
import datetime
from sqlalchemy import create_engine

def extract_data(url: str):

    csv_name = 'crime_data_la.csv'

    os.system(f"curl --insecure -o {csv_name} {url}")

    df = pd.read_csv(csv_name)

    return df
```

Capstone project ini menggunakan data yang tersedia pada website data.lacity.org yang merupakan dataset kejadian kriminal yang terjadi di kota Los Angeles pada tahun 2019 hingga sekarang. Untuk mendapatkan datasetnya, dibuatlah sebuah pipeline ETL (Extract, Transform, Load) menggunakan python untuk melakukan proses ETL secara otomatis ketika skrip python tersebut dijalankan. Gambar diatas merupakan skrip yang digunakan untuk mengekstrak dataset dari sumbernya.



Step 2 - Data Cleansing

```
ingest_data.py
def transform_data(df):
    transformed_df = df.copy()

    #mendrop kolom yang tidak perlukan karena berisikan informasi yang sudah tersedia di kolom lain dan berisikan ko
    transformed_df.drop(['DR_NO', 'AREA', 'Rpt Dist No', 'Part 1-2', 'Crm Cd',
                        'Mocodes', 'Premis Cd', 'Weapon Used Cd', 'Status', 'Crm Cd 1',
                        'Crm Cd 2', 'Crm Cd 3', 'Crm Cd 4', 'Cross Street'], axis = 1, inplace=True)

    #merename kolom agar lebih mudah dibaca
    transformed_df.rename(columns={'Date Rptd': 'date reported',
                                   'DATE OCC': 'date occurrence',
                                   'TIME OCC': 'time occurrence',
                                   'AREA NAME': 'area name',
                                   'Crm Cd Desc': 'crime description',
                                   'Vict Age': 'victim age',
                                   'Vict Sex': 'victim sex',
                                   'Vict Descent': 'victim descendant',
                                   'Premis Desc': 'premises description',
                                   'Weapon Desc': 'weapon description',
                                   'Status Desc': 'case status',
                                   'LOCATION': 'location',
                                   'LAT': 'lat',
                                   'LON': 'lon'},
                           inplace=True)

    #mengubah tipe data kolom date reported dan date occurrence menjadi datetime
    transformed_df['date reported'] = pd.to_datetime(transformed_df['date reported'], format="%m/%d/%Y %I:%M:%S %p")
```

Setelah dataset diekstrak, langkah selanjutnya adalah melakukan transformasi data untuk membersihkan dataset karena dataset tersebut kurang clean. Beberapa kolom didrop karena kolom tersebut mengandung informasi yang redundan (sudah tersedia di kolom lain) dan kolom yang tersisa direname agar mudah dibaca.



Step 2 - Data Cleansing

```
#mengubah tipe data kolom date reported dan date occurrence menjadi datetime
transformed_df['date reported'] = pd.to_datetime(transformed_df['date reported'], format="%m/%d/%Y %I:%M:%S %p")
transformed_df['date occurrence'] = pd.to_datetime(transformed_df['date occurrence'], format="%m/%d/%Y %I:%M:%S %p")

# melakukan transformasi terhadap kolom time occurrence
transformed_df['time occurrence'] = transformed_df['time occurrence'].astype(str)
transformed_df['time occurrence'] = transformed_df['time occurrence'].str[:-2].str.zfill(2) + ':' + transformed_df['time occurrence'].str[-2:]
transformed_df['time occurrence'] = pd.to_datetime(transformed_df['time occurrence'], format="%H:%M").dt.strftime("%H:%M")

#mengisi nilai kosong
transformed_df['victim sex'].fillna('unknown', inplace=True)
transformed_df['victim descendant'].fillna('unknown', inplace=True)
transformed_df['weapon description'].fillna('UNKNOWN WEAPON/OTHER WEAPON', inplace=True)
transformed_df['victim age'] = transformed_df['victim age'].abs()

transformed_df['location'] = transformed_df['location'].replace('\s+', ' ', regex=True).str.strip()
```

Beberapa kolom diubah tipe datanya agar memudahkan proses agregat nantinya, value yang tidak seragam pada kolom tertentu juga dilakukan cleansing agar value menjadi seragam untuk memudahkan proses analisis. Kolom yang memiliki nilai kosong diisi dengan 'unknown' karena informasi yang spesifik memang tidak diketahui untuk menghindari asumsi yang salah mengenai data yang tidak lengkap.

TETRIS PROGRAM

Step 2 - Data Cleansing

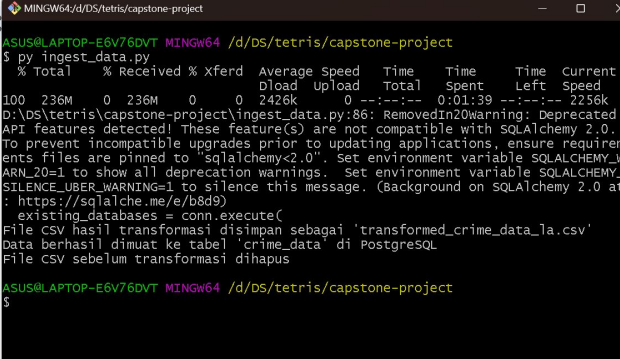
```
def load_data(df, table_name, connection_string):
    engine = create_engine(connection_string)
    database_name = "la_crime_db"

    # Mengecek apakah database sudah ada
    with engine.connect() as conn:
        existing_databases = conn.execute(
            "SELECT datname FROM pg_catalog.pg_database"
        )
        existing_databases = [db[0] for db in existing_databases]

    if database_name not in existing_databases:
        # Jika database belum ada, buat database baru
        conn.execute(f"COMMIT") # Mengakhiri transaksi sebelum membuat database
        conn.execute(f"CREATE DATABASE {database_name}")

    # Membuat koneksi baru ke database yang sudah ada atau baru saja dibuat
    connection_string_with_db = f'{connection_string}/{database_name}'
    engine_with_db = create_engine(connection_string_with_db)
    with engine_with_db.connect() as conn:
        df.to_sql(table_name, conn, index=False, if_exists='replace')

    print(f"Data berhasil dimuat ke tabel '{table_name}' di PostgreSQL")
```



```
ASUS@LAPTOP-E6V76DVT MINGW64 /d/ps/tetris/capstone-project
$ py ingest_data.py
% Total % Received % Xferd Average Speed Time Time Time Current
Dload Upload Total Spent Left Speed
100 236M 0 236M 0 0 2426k 0 --:--:-- 0:01:39 --:--:-- 2256k
D:\ps\tetris\capstone-project\ingest_data.py:86: RemovedIn20Warning: Deprecated
API features detected! These feature(s) are not compatible with SQLAlchemy 2.0.
To prevent incompatible upgrades prior to updating applications, ensure requirem
ents files are pinned to "sqlalchemy<2.0". Set environment variable SQLALCHEMY_W
ARN_20=1 to show all deprecation warnings. Set environment variable SQLALCHEMY_
SILENCE_USER_WARNING=1 to silence this message. (Background on SQLAlchemy 2.0 at
: https://sqlalche.me/e/b8d9)
existing_databases = conn.execute(
File CSV hasil transformasi disimpan sebagai 'transformed_crime_data_la.csv'
Data berhasil dimuat ke tabel 'crime_data' di PostgreSQL
File CSV sebelum transformasi dihapus
ASUS@LAPTOP-E6V76DVT MINGW64 /d/ps/tetris/capstone-project
$
```

Setelah proses transformasi selesai, selanjut dataset yang sudah clean di load ke postgresql untuk keperluan analisis menggunakan power bi. Dengan menjalankan skrip ingest_data.py ini, proses Extract Transform dan Load akan dijalankan secara otomatis dan siap untuk dilakukan analisis.



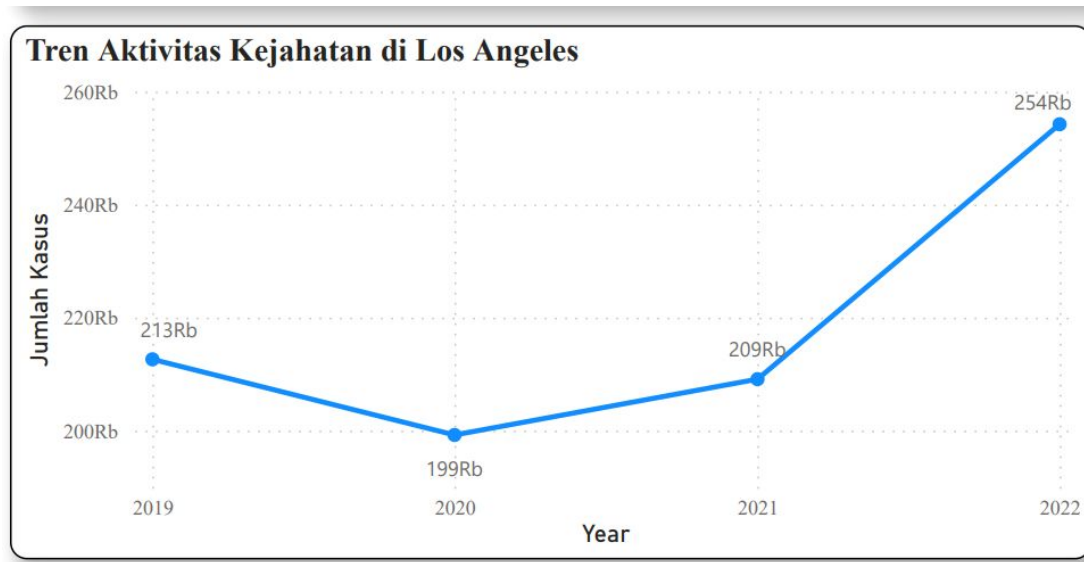
Step 2 - Data Cleansing

```
#mapping value
sex_mapping = {'F' : 'female', 'M' : 'male', 'X' : 'unknown', 'H' : 'female', 'N' : 'male', 'nan' : 'unknown'}
desc_mapping = {'A' : 'other asian', 'B' : 'black', 'C' : 'chinese',
                'D' : 'cambodian', 'F' : 'filipino', 'G' : 'guamanian',
                'H' : 'hispanic/latin/mexican', 'I' : 'american indian/alaskan native',
                'J' : 'japanese', 'K' : 'korean', 'L' : 'laotian', 'O' : 'other', 'P' : 'pacific islander',
                'S' : 'samoan', 'U' : 'hawaiian', 'V' : 'vietnamese', 'W' : 'white',
                'X' : 'unknown', 'Z' : 'asian indian',
                '-' : 'unknown',
                'nan' : 'unknown'
               }

transformed_df['victim sex'] = transformed_df['victim sex'].map(sex_mapping)
transformed_df['victim descendant'] = transformed_df['victim descendant'].map(desc_mapping)
```

Beberapa kolom juga dilakukan mapping ulang terhadap value yang ada pada kolom tersebut agar value - value tersebut lebih bermakna dan sesuai untuk analisis yang akan dilakukan. Dengan mapping ini, format value menjadi mudah dipahami, dianalisis dan digunakan dalam proses penggalian insight.

Step 3 - Data Exploration & Data Visualisation



Dalam empat tahun terakhir, Los Angeles telah menghadapi tantangan serius terkait aktivitas kriminal. Data statistik menunjukkan **tren yang perlu diperhatikan**. Pada tahun 2019, jumlah kasus mencapai 212.618, menunjukkan **tingkat kejahatan yang signifikan**.

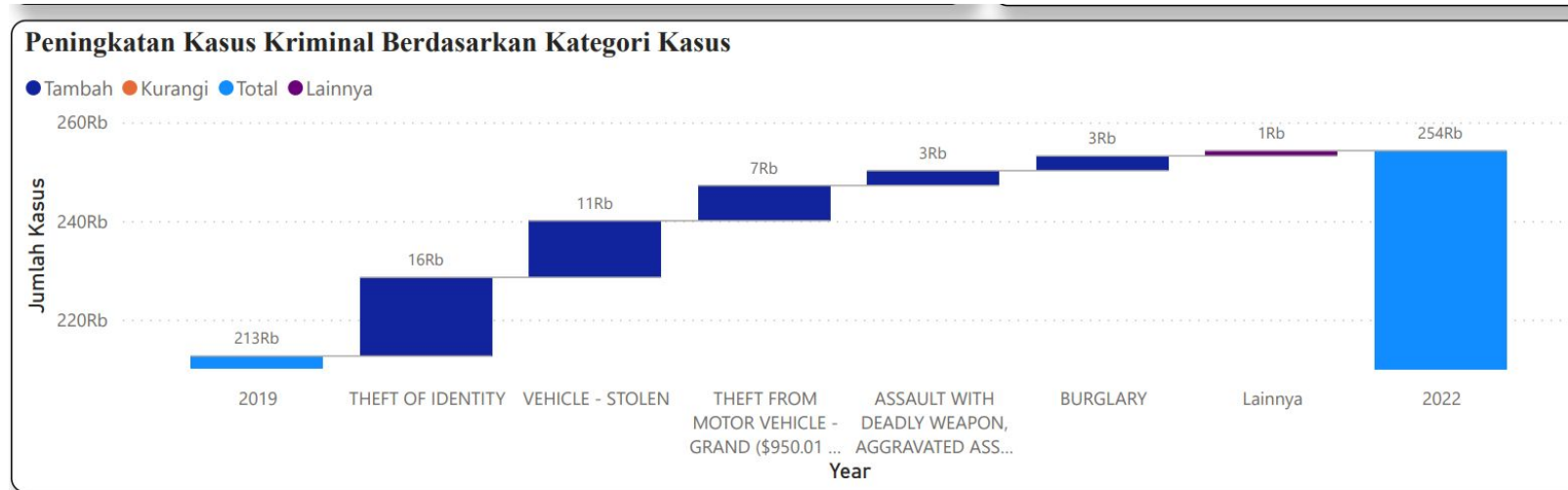
Namun, angka tersebut **mengalami penurunan** pada tahun 2020 menjadi 199.220, mungkin disebabkan oleh berbagai faktor seperti **situasi pandemi dan langkah-langkah pengendalian kejahatan yang diterapkan**. Namun, penurunan ini hanya sementara, karena pada tahun 2021 angka kasus **kembali meningkat** menjadi 209.102. Tahun 2022 mencatat **angka tertinggi** dengan 254.251 kasus, menunjukkan adanya peningkatan yang signifikan dibandingkan tahun-tahun sebelumnya.

Tahapan eksplorasi data dan visualisasi dilakukan secara bertahap dengan analisis univariat dan bivariat. Ini membantu dalam memperoleh insight yang berharga dari data yang ada. Analisis univariat memeriksa setiap variabel secara terpisah, sedangkan analisis bivariat memeriksa hubungan antara dua variabel secara simultan. Dengan pendekatan ini, kita dapat mengidentifikasi pola, korelasi, dan perbedaan yang signifikan. Hasilnya memberikan informasi berharga untuk pengambilan keputusan yang lebih baik dan perencanaan strategi.



Step 3 - Insight Analysis

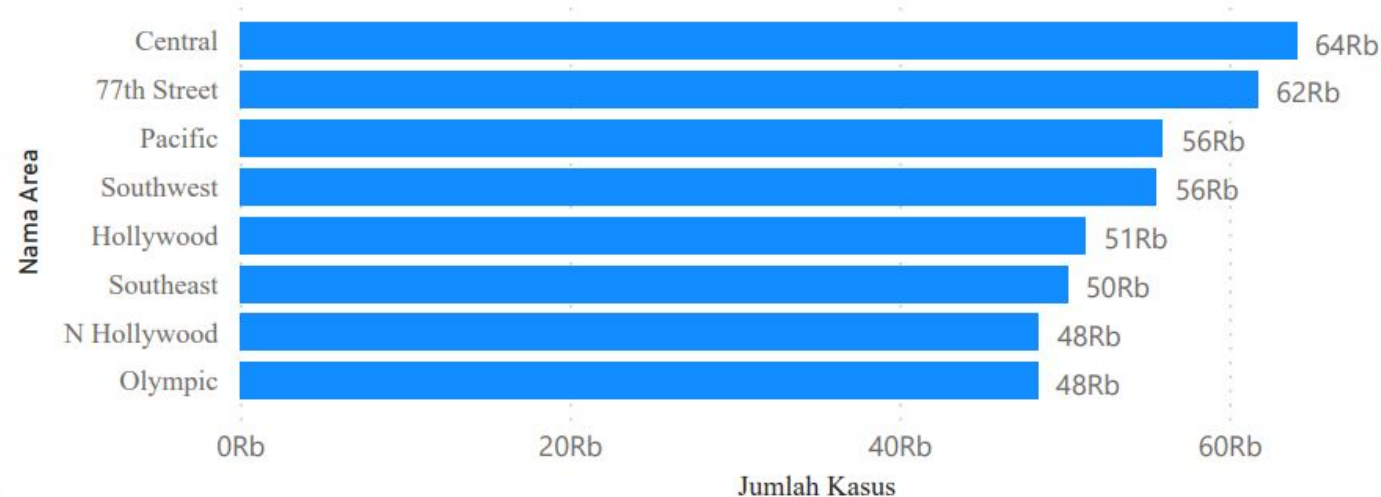
Peningkatan kasus kriminal **berkategori pencurian identitas** dan **pencurian kendaraan** di Los Angeles telah menjadi **sorotan utama** dalam 3 tahun terakhir. Data statistik di bawah ini menunjukkan tren yang mencemaskan, dengan angka kasus yang meningkat **secara signifikan** dari tahun 2021 hingga tahun 2022. Fenomena ini mengindikasikan adanya tantangan serius terkait **keamanan identitas** dan **kendaraan individu** di kota ini.





Step 3 - Insight Analysis

Jumlah Kasus Kriminal per Area



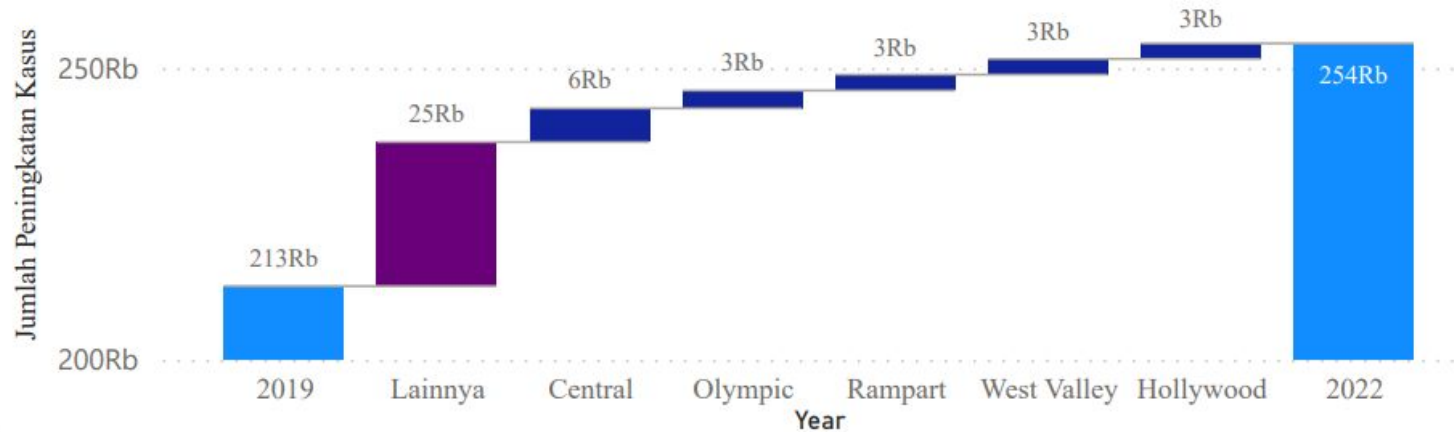
Di seluruh 21 distrik, jumlah kasus berkisar dari 32.820 hingga 64.058. Dengan jumlah kasus sebanyak 64.058, **Distrik Central memiliki jumlah kasus tertinggi** dan 95,18% **lebih tinggi daripada Foothill**, yang memiliki jumlah kasus terendah yaitu sebanyak 32.820. Dengan angka ini, distrik Central menyumbang **6,61% dari total kasus yang ada.**



Step 3 - Insight Analysis

Jumlah Peningkatan Kasus berdasarkan Distrik

● Tambah ● Kurangi ● Total ● Lainnya

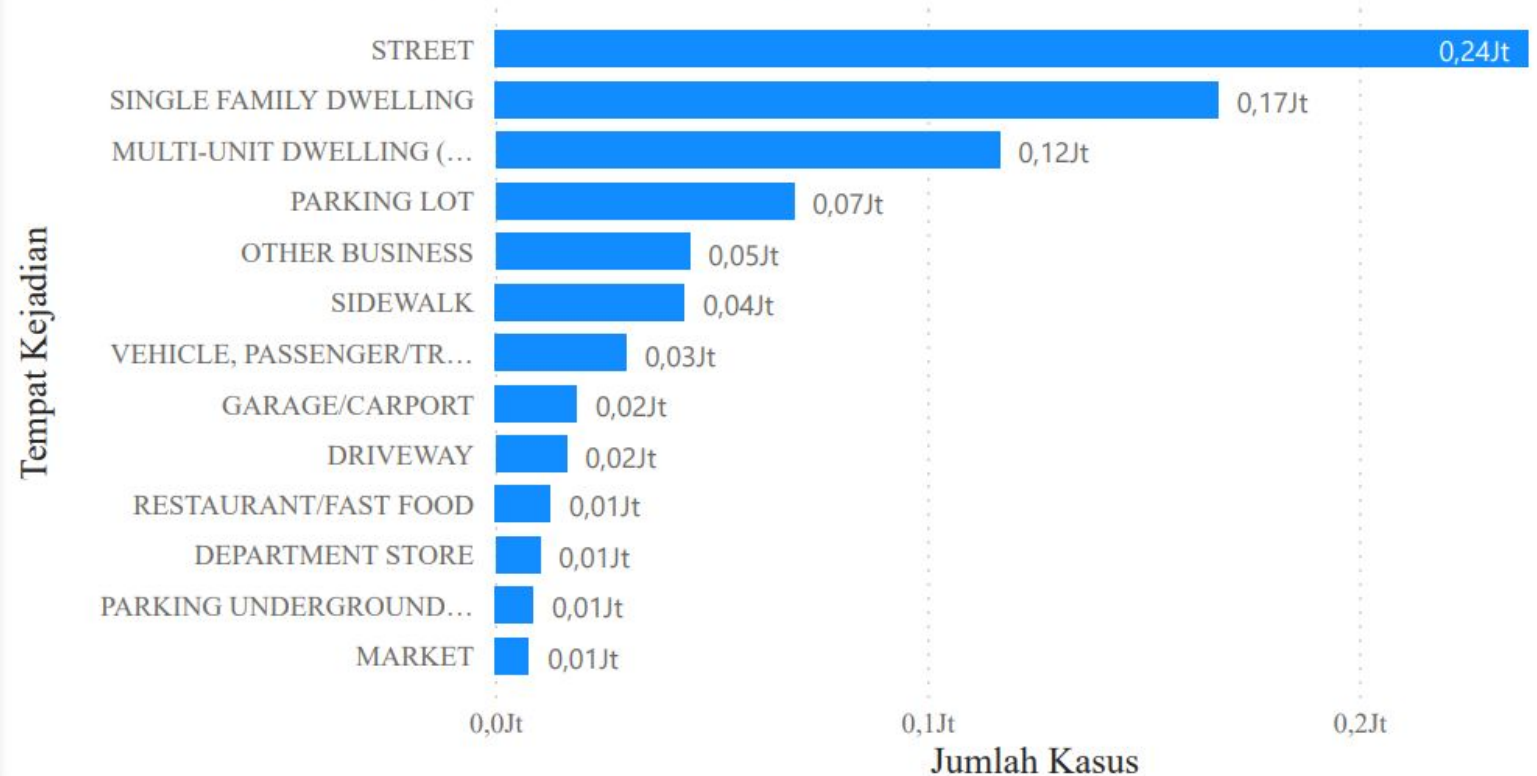


Distrik **Central** juga menjadi area yang paling banyak mengalami peningkatan kasus kriminal pada tahun 2019 hingga tahun 2022 dengan kategori **pencurian dari kendaraan** yang paling banyak mengalami peningkatan. Sedangkan untuk distrik lainnya, **Olympic** dengan kategori **pencurian kendaran** dan **Rampart** dengan kategori **pencurian identitas** menjadi distrik yang paling banyak mengalami peningkatan setelah distrik central.

Step 3 - Insight Analysis

Tren yang menarik terungkap dari analisis data, bahwa kasus kejahatan yang paling sering terjadi adalah di lokasi **jalan umum**. Dalam kategori ini, kasus kejahatan mencakup sebesar **24,70%** dari total jumlah kasus yang tercatat. Fenomena ini menyoroti pentingnya keamanan di lingkungan publik dan memperkuat kebutuhan untuk meningkatkan pengawasan dan langkah-langkah pencegahan di jalan-jalan umum. Tingginya jumlah kasus kejahatan di lokasi ini menunjukkan risiko yang lebih tinggi bagi masyarakat yang beraktivitas di ruang publik.

Jumlah Kasus yang Terjadi Berdasarkan Tempat Kejadian

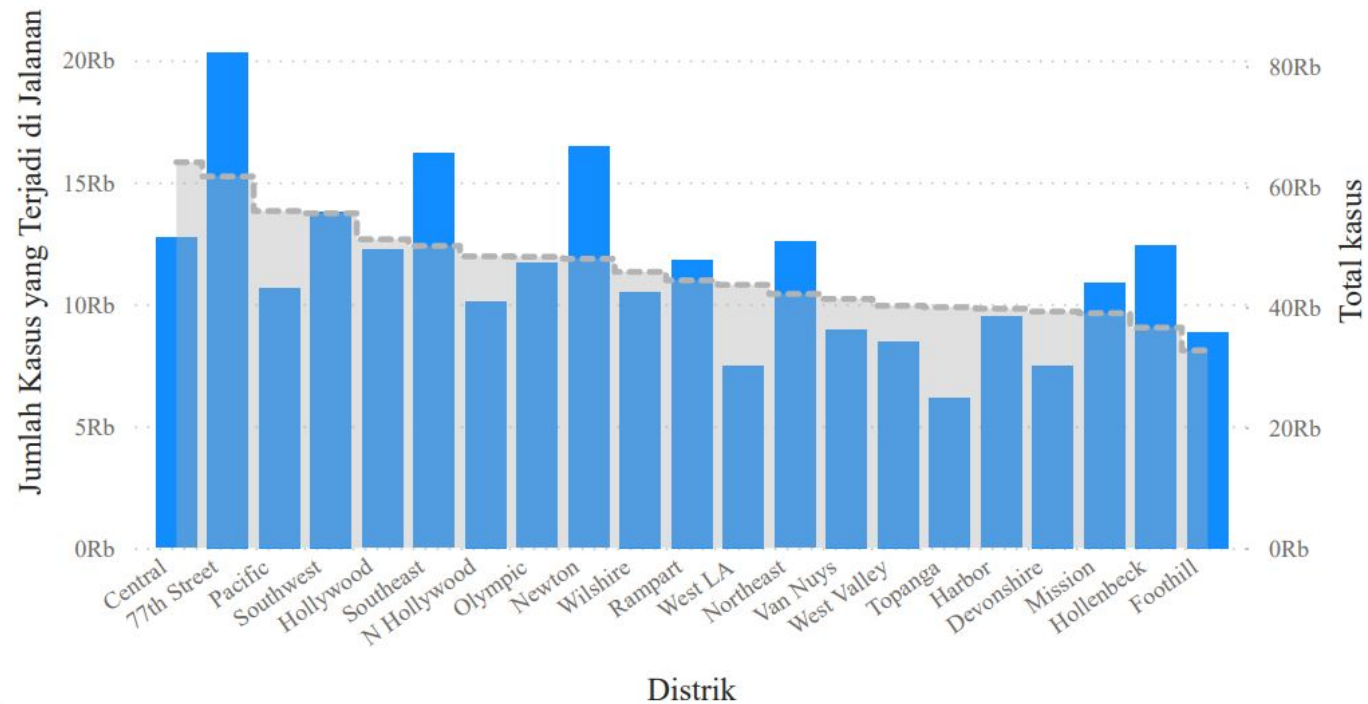




Step 3 - Insight Analysis

Jumlah Kasus Kriminal yang Terjadi di Jalan Umum

● Jumlah Kasus yang Terjadi di Jalanan ● Total kasus

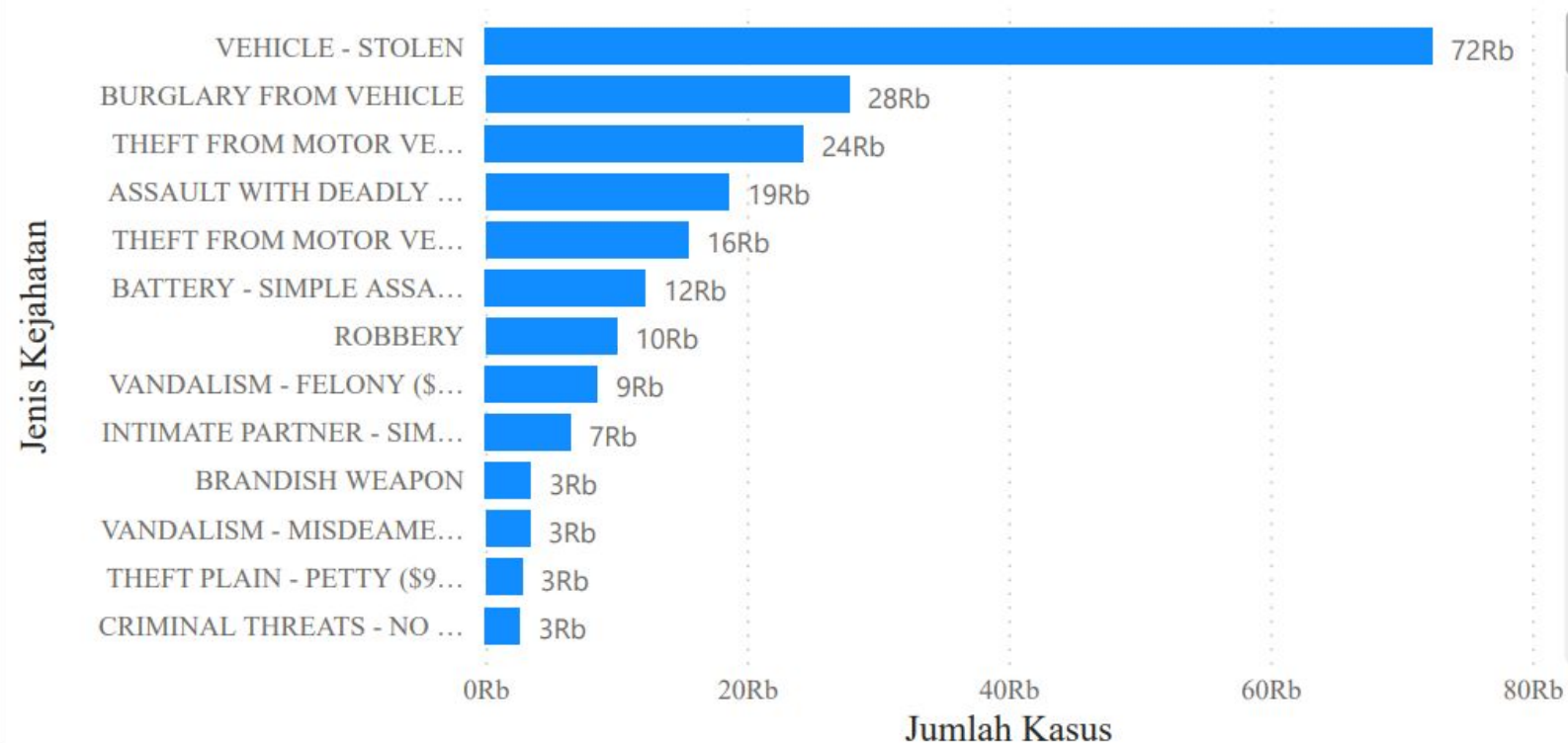


Melalui analisis grafik yang disajikan, terlihat bahwa Distrik **77th Street**, **Southeast**, dan **Newton** adalah distrik-distrik yang memiliki **tingkat kejahatan yang tinggi terjadi di jalan raya**. Ketiga distrik ini secara signifikan menyumbang lebih dari **30% dari total kasus kejahatan yang terjadi di jalan raya**. Peningkatan kasus kejahatan di jalan raya di distrik-distrik ini dapat menimbulkan kekhawatiran akan keamanan dan keselamatan masyarakat yang tinggal, bekerja, atau berkunjung ke daerah-daerah tersebut. Adanya konsentrasi kasus kejahatan di jalan raya di distrik-distrik ini mengindikasikan **adanya potensi risiko yang perlu diperhatikan dan ditangani dengan serius**.



Step 3 - Insight Analysis

Jenis Kejahatan yang Terjadi di Jalan Umum



Data menunjukkan bahwa **pencurian kendaraan** menjadi kasus yang paling dominan terjadi di jalan umum, dengan menyumbang sebesar **30.26%** dari total kasus kejahatan. Hal ini menggambarkan betapa pentingnya keamanan kendaraan di lingkungan publik. **Pencurian barang berharga dari kendaraan, pencurian dari pengguna sepeda motor, dan penyerangan menggunakan senjata mematikan** juga menjadi jenis kejahatan yang sering terjadi di jalan umum.

Pencurian barang berharga dari kendaraan menjadi ancaman yang signifikan, di mana pelaku kejahatan

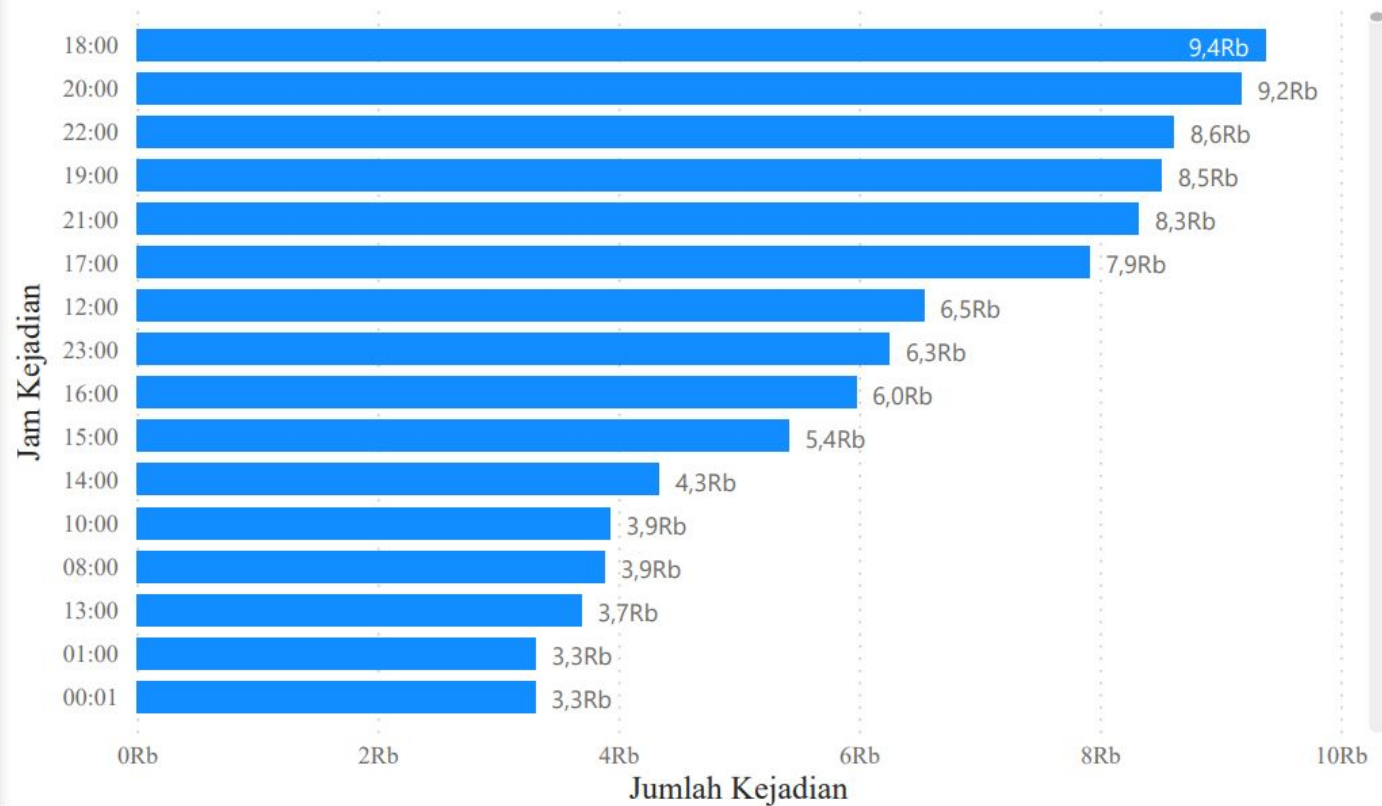


Step 3 - Insight Analysis

Grafik yang disajikan menunjukkan adanya pola yang jelas terkait waktu kejadian kasus kriminal di jalan umum. **Pukul 18:00 menjadi waktu dengan jumlah kasus terbanyak, diikuti oleh pukul 20:00 dan 22:00.** Rentang waktu antara pukul 18:00 hingga 23:00 menjadi periode yang **paling rawan** terjadinya kasus kriminal di jalan umum.

Penting untuk memahami pola ini agar masyarakat dapat **meningkatkan kewaspadaan mereka saat berada di luar rumah pada rentang waktu tersebut.** Pukul 18:00, yang seringkali merupakan waktu pulang kerja atau saat aktivitas di jalan umum mencapai puncak, menjadi momen yang rentan terhadap tindakan kriminal.

Jumlah Kasus yang terjadi di Jalan Umum Berdasarkan Jam Waktu Kejadian





Step 3 - Insight Analysis

Rekomendasi Berdasarkan Insight yang didapat

- Kesadaran dan Pengawasan Tingkatkan kesadaran akan situasi sekitar dan tingkatkan pengawasan pribadi saat berada di jalan umum. Perhatikan sekeliling Anda, sebisa mungkin hindari daerah yang dikenal rawan (77th Street, Southeast, Newton), dan waspada terhadap tanda-tanda kejadian mencurigakan.
- Perhatikan rentang waktu antara pukul 18:00 hingga 23:00 yang menjadi periode dengan jumlah kasus kriminalitas yang tinggi. Selama jam-jam ini, tingkatkan kewaspadaan dan perhatian ekstra terhadap lingkungan sekitar.
- Tingkatkan kerjasama dengan komunitas sekitar, kepolisian, dan pihak berwenang. Melaporkan kejadian mencurigakan dan berbagi informasi akan membantu dalam upaya mencegah kejahatan dan meningkatkan keamanan di jalan umum.
- Gunakan teknologi dan inovasi dalam upaya pencegahan dan penanggulangan kejahatan di jalan umum. Misalnya, penggunaan aplikasi keamanan, sistem pemantauan CCTV yang canggih, atau pengembangan alat-alat keamanan pintar untuk membantu masyarakat dalam melindungi diri mereka sendiri.

DΦLab

AYO #STACKYOURSKILL SEKARANG

dan Persiapkan Diri Menjadi Praktisi Data!

