

EVALUASI TENGAH SEMESTER EXPLORATORY DATA ANALYSIS PROJECT REPORT

Judul Proyek: Analisis Kriminalitas Kota Los Angeles

Kelompok: 8

Anggota: Johanes Efata Putra Diaz Prasetyo (24083010016), Erlina Ayunda Cahyanti (24083010027), Alifah Zuhrah Ulyya S. (24083010033), Desi Nofitasari (24083010058), Baihaqi Nur Muhammad (24083010111).

1. Executive Summary

1.1 Ringkasan tujuan, data sumber, dan 3–5 temuan utama

Tujuan dari proyek ini adalah untuk menganalisis pola kriminalitas di Kota Los Angeles dengan pendekatan eksploratif, sehingga dapat memberikan wawasan yang berguna bagi pihak berwenang dan masyarakat dalam merancang strategi pencegahan kejahatan yang efektif. Analisis ini berfokus pada identifikasi tren berdasarkan waktu, lokasi, jenis kejahatan, serta karakteristik korban. Sumber data yang digunakan berasal dari catatan kriminalitas publik di Kota Los Angeles, mencakup laporan kejahatan dari berbagai distrik, jenis kejahatan, waktu terjadinya, informasi demografis korban, serta persebaran sumber daya keamanan. Data ini memberikan gambaran menyeluruh mengenai distribusi kriminalitas di seluruh kota dan faktor-faktor yang mempengaruhi tingkat kejahatan.

Hasil analisis menunjukkan beberapa temuan utama:

1. Pola waktu kejahatan: Aktivitas kriminal cenderung meningkat pada siang dan malam hari, dengan puncak risiko terjadi pada hari Jumat hingga Minggu. Hal ini mencerminkan adanya hubungan antara pola aktivitas masyarakat dengan waktu terjadinya kejahatan, sehingga periode rawan tersebut harus menjadi fokus utama dalam pengawasan dan patroli keamanan.
2. Konsentrasi lokasi: Daerah padat penduduk seperti Pacific dan Southeast memiliki tingkat kriminalitas yang lebih tinggi, menunjukkan bahwa kepadatan dan dinamika sosial mempengaruhi risiko kejahatan.
3. Jenis kejahatan dominan: Kejahatan yang paling sering terjadi meliputi pencurian kendaraan, perampokan, dan serangan fisik mencerminkan kombinasi antara kejahatan kekerasan dan properti. Hal ini juga dapat menunjukkan sistem keamanan di Los Angeles masih belum maksimal.
4. Kelompok korban rentan: Individu berusia produktif, terutama 25–30 tahun, menjadi kelompok yang paling terdampak, menandakan hubungan antara mobilitas masyarakat dan risiko kriminalitas.
5. Rasio polisi : Dari hasil analisis, rasio polisi menunjukkan bahwa persebaran petugas di daerah rawan kejahatan kurang merata. Hal ini mendukung temuan terkait jenis kejahatan yang menyoroti pentingnya penguatan sistem keamanan di area tersebut.
6. Implikasi strategis: Temuan ini menekankan pentingnya strategi keamanan berbasis waktu dan lokasi, termasuk peningkatan patroli di jam rawan dan daerah dengan tingkat kejahatan tinggi, serta edukasi masyarakat untuk meningkatkan kewaspadaan terhadap potensi ancaman kriminal di lingkungan perkotaan.
7. Secara keseluruhan, analisis ini menyediakan dasar yang kuat bagi perencanaan pencegahan kejahatan dan pengambilan keputusan berbasis data, sehingga keamanan dan kualitas hidup masyarakat Los Angeles dapat lebih terjaga.

1.2 Rekomendasi praktis

1. Peningkatan Pengawasan di Jam Rawan: Penempatan petugas kepolisian dan patroli tambahan perlu difokuskan pada sore hingga malam hari, terutama di distrik dengan Z-score waktu positif tinggi seperti *77th Street* dan *Central*.
2. Edukasi Masyarakat dan Pencegahan Partisipatif: Program kesadaran publik dapat difokuskan pada usia 20–35 tahun, dengan kampanye pencegahan pencurian kendaraan dan keamanan pribadi di area publik.
3. Integrasi Data dan Monitoring Berkelanjutan: Disarankan untuk menggabungkan data kriminalitas LAPD dengan data sosial-ekonomi (pendapatan, pendidikan, kepadatan penduduk) agar analisis lanjutan dapat mengidentifikasi faktor penyebab kriminalitas secara lebih mendalam.
4. Optimalisasi Distribusi Personel Kepolisian: Berdasarkan temuan *police staffing data*, perlu dilakukan redistribusi sumber daya manusia kepolisian ke wilayah dengan tingkat kasus tinggi dan variasi kejahatan yang kompleks.

2. Pendahuluan

2.1 Latar belakang & masalah bisnis/riset

Los Angeles merupakan wilayah yang terdapat di Amerika Serikat yang memiliki lebih dari 10 juta penduduk dengan latar belakang etnis dan budaya yang beragam. Sebagai kota yang besar dengan dinamika sosial ekonomi yang kompleks, Los Angeles menghadapi berbagai tantangan, salah satunya yaitu tingginya tingkat kriminalitas. Berdasarkan Laporan Los Angeles Times (2016,30 Desember) kriminalitas di Los Angeles mengalami peningkatan di hampir semua kategori utama dari tahun 2015 hingga 2016. Oleh karena itu daerah ini sangat menarik untuk dianalisis lebih lanjut.

2.2 Tujuan EDA dan ruang lingkup

Penelitian ini bertujuan untuk melakukan Analisis Data Eksploratif (Exploratory Data Analysis / EDA) guna memperoleh pemahaman mendalam mengenai karakteristik data kriminalitas di Los Angeles. Melalui EDA, peneliti berupaya mengidentifikasi pola distribusi kejahatan berdasarkan lokasi, jenis kejahatan, dan waktu kejadian, serta mengetahui jenis kejahatan yang paling dominan dan area dengan tingkat kriminalitas tertinggi. Analisis ini juga bertujuan untuk mendeteksi adanya anomali atau pola tak biasa yang muncul dalam data, sehingga dapat memberikan gambaran awal yang lebih jelas mengenai kondisi kriminalitas di wilayah tersebut.

3. Konsep Dasar EDA & Perbandingan

3.1 Data Science & Peran EDA

Exploratory Data Analysis (EDA) merupakan tahap awal dalam analisis data yang bertujuan memahami karakteristik dan struktur dasar data sebelum dilakukan pemodelan lanjutan. Konsep ini diperkenalkan oleh John Tukey sebagai pendekatan eksploratif untuk menemukan pola, mendeteksi anomali, dan memeriksa asumsi tanpa bergantung pada model statistik tertentu. Melalui EDA, peneliti berusaha memperoleh pemahaman mendalam mengenai data melalui statistik deskriptif dan visualisasi. Dalam *data science*, EDA menjadi tahap penting karena menjembatani data mentah dengan proses analisis lanjutan seperti regresi atau *machine learning*.

3.2 Pentingnya EDA untuk Kriminalitas Perkotaan

EDA berperan penting dalam memahami karakteristik dan kualitas dataset kriminalitas Los Angeles dengan mengidentifikasi data kosong, duplikasi, serta nilai ekstrim agar analisis yang dilakukan lebih akurat dan

representatif. Melalui EDA, dapat diketahui distribusi jenis kejahatan, pola waktu kejadian, sebaran lokasi berdasarkan distrik, serta karakteristik korban seperti usia dan jenis kelamin. Visualisasi seperti heatmap, scatter plot, dan box plot membantu mengungkap pola kriminalitas di berbagai wilayah, menemukan anomali, serta melihat keterkaitan antara variabel seperti area kejadian dan jenis kejahatan. Hasil eksplorasi ini menjadi dasar untuk menentukan pendekatan analisis yang tepat dan mendukung pengambilan keputusan berbasis data dalam memahami dinamika kriminalitas di Los Angeles.

3.3 Langkah-langkah EDA

Tahapan EDA untuk data kriminalitas Los Angeles dimulai dengan integrasi data kriminalitas dan demografi antar distrik agar formatnya konsisten. Selanjutnya dilakukan pembersihan data dengan menghapus duplikat, memperbaiki nilai kosong, dan menyesuaikan tipe data. Setelah itu, analisis univariat dilakukan untuk melihat distribusi tiap variabel seperti jumlah kasus, populasi, dan pendapatan, kemudian dilanjutkan analisis bivariat atau multivariat untuk memeriksa hubungan antar variabel. Langkah berikutnya adalah analisis spasial untuk memetakan persebaran kejahatan dan mengidentifikasi wilayah dengan resiko tinggi, serta analisis waktu bila tersedia data tahunan atau bulanan guna melihat tren kriminalitas. Tahap akhir berupa formulasi insight dan hipotesis awal yang menggambarkan pola hubungan sosial-ekonomi dengan kejahatan, diikuti iterasi bila ditemukan pola baru agar analisis lebih mendalam dan akurat.

3.4 EDA vs Analisis Klasikal vs Bayesian

Dalam penelitian kriminalitas perkotaan, EDA digunakan untuk memahami pola dan struktur data tanpa asumsi kuat agar peneliti dapat menemukan tren dan hubungan potensial secara bebas. Pendekatan ini tepat ketika peneliti belum memiliki hipotesis spesifik atau ingin menilai kualitas data terlebih dahulu. Sebaliknya, analisis klasik (frequentist) digunakan untuk menguji hipotesis tertentu, seperti pengaruh pendapatan terhadap tingkat kejahatan, dengan asumsi distribusi data yang jelas. Sementara analisis Bayesian lebih sesuai bila terdapat pengetahuan awal misalnya data kriminalitas tahun sebelumnya yang dapat diperbarui dengan informasi baru.

3.5 Software/Pustaka

Proses analisis data dilakukan menggunakan pustaka utama Pandas, NumPy, Matplotlib, dan SciPy. Pustaka Pandas digunakan untuk membaca, membersihkan, serta memanipulasi data tabular dalam format DataFrame. NumPy berperan dalam perhitungan numerik dan operasi matematis seperti transformasi dan statistik dasar. Matplotlib digunakan untuk menghasilkan berbagai visualisasi data seperti histogram, scatter plot, dan heatmap guna mempermudah interpretasi pola dan tren. Sementara itu, SciPy dimanfaatkan untuk analisis statistik lanjutan seperti deteksi outlier menggunakan metode *z-score*.

4. Data & Metodologi

4.1 Sumber Data

Penelitian ini menggunakan tiga kelompok utama data publik terkait kriminalitas di Los Angeles. (1) *Crime Data from 2020 to Present* dari [Los Angeles Open Data Portal](#), disediakan oleh LAPD dengan lisensi *CC0 1.0 Universal*, berisi laporan kejahatan individu beserta metadata lokasi, waktu, dan jenis kejahatan. (2) *Police Staffing Data* diperoleh dari dua sumber: [NextRequest LA City Public Records Platform](#), yang memuat data distribusi personel polisi dengan lisensi non-eksklusif untuk penggunaan publik, serta [WatchTheWatchers.net](#) dari *Stop LAPD Spying Coalition*, yang mengkompilasi informasi publik tentang personel LAPD dari dokumen resmi pemerintah (domain publik). (3) *Census Data by Council District* dari [data.lacity.org](#) berisi data demografi penduduk berdasarkan distrik dengan lisensi *CC0 1.0 Universal*. (4) Selain itu, digunakan juga *City Council Districts (Adopted 2021)* dari [GeoHub LA City](#) untuk mengaitkan koordinat lokasi ke dalam wilayah administratif distrik.

4.2 Data Dictionary

4.2.1 Data Kriminal Kota Los Angeles

Tabel 1. Penjelasan Data *Dictionary* untuk Dataset Crime

Nama Kolom	Tipe Data	Skala	Contoh Nilai	Deskripsi
Dr No	Integer	Nominal	201904032	Nomor arsip resmi LAPD, terdiri dari tahun, kode area, dan 5 digit nomor unik.
Date Rptd	Objek	-	2020 Jan 02 12:00:00 AM	Tanggal kejadian dilaporkan ke polisi
Date OCC	Objek	-	2020 Jan 01 12:00:00 AM	Tanggal kejadian tindak kriminal terjadi
Time OCC	Integer	-	2135	Waktu kejadian dalam format 24 jam militer
Area	Integer	Nominal	19	Kode numerik area geografis tempat kejadian, dari 1 sampai 21
Area Name	Objek	-	Mission	Nama divisi patroli dari area tempat kejadian.
Rpt Dist No	Integer	Nominal	1924	Kode distrik pelaporan 4 digit di dalam area
Part 1-2	Integer	Ordinal	1	Kategori kejahatan: 1 untuk serius, 2 untuk kurang serius
Crm Cd	Integer	Nominal	761	Kode kejahatan utama yang dilakukan.
Crm Cd Desc	Objek	-	BRANDISH WEAPON	Deskripsi dari kode kejahatan.
Mocodes	Objek	Nominal	1822 3006 3028 3039 3032 3037 3101 4019 3602	Kode modus operandi pelaku, bisa lebih dari satu dalam satu kejadian.
Vict Age	Integer	Rasio	41	Usia korban
Vict Sex	Objek	-	M	Jenis kelamin korban
Vict Descent	Objek	-	H	Kode etnis korban
Premis Cd	Float	Nominal	101.0	Kode tempat kejadian
Premis Desc	Objek	-	STREET	Deskripsi dari kode tempat kejadian
Weapon Used Cd	Float	Nominal	200.0	Kode jenis senjata yang digunakan pelaku
Weapon Desc	Objek	-	KNIFE WITH BLADE 6INCHES OR LESS	Deskripsi dari kode senjata yang digunakan.
Status	Objek	-	AA	Kode status kasus
Status Desc	Objek	-	Adult Arrest	Deskripsi dari kode status kasus
Crm Cd 1	Float	Nominal	761.0	Kode kejahatan utama yang paling serius
Crm Cd 2	Float	Nominal	930.0	Kode kejahatan tambahan yang kurang serius dari Crm Cd 1

Crm Cd 3	Float	Nominal	997.0	Kode kejahatan tambahan kedua
Crm Cd 4	Float	Nominal	998.0	Kode kejahatan tambahan ketiga
Location	Objek	-	ASTORIA ST	Alamat lokasi kejadian, dibulatkan ke blok terdekat untuk menjaga anonimitas.
Cross Street	Objek	-	SAN FERNANDO RD	Nama jalan yang bersilangan dengan lokasi kejadian.
Lat	Float	Interval	34.2949	Koordinat lintang lokasi kejadian.
Lon	Float	Interval	-118.4571	Koordinat bujur lokasi kejadian.

4.2.2 Data Kepolisian

Tabel 2. Penjelasan Data *Dictionary* untuk Dataset Kepolisian

Nama Kolom	Tipe Data	Skala	Contoh Nilai	Deskripsi
Name	Objek	-	ABAD, MIGUEL A	Nama lengkap petugas
SerialNo	Integer	Nominal	42963	Nomor identifikasi unik petugas.
Area	Objek	-	NEWT	Nama area tempat petugas tersebut ditugaskan.
...

4.2.3 Data sensus Penduduk Los Angeles

Tabel 3. Penjelasan Data *Dictionary* untuk Dataset Penduduk

Nama Kolom	Tipe Data	Skala	Contoh Nilai	Deskripsi
Council District	Objek	-	1 - Gilbert Cedillo	Nama atau nomor distrik dewan perwakilan daerah
Pop2010	Float	Rasio	245216.22	Total populasi di distrik pada tahun 2010.
White pop	Float	Rasio	90857	Jumlah populasi berkulit putih di distrik.
Black pop	Float	Rasio	7759.13	Jumlah populasi berkulit hitam di distrik.
Ameri es pop	Float	Rasio	2991.73	Jumlah populasi penduduk asli Amerika di distrik.
Other pop	Float	Rasio	88016.11	Jumlah populasi etnis lain yang tidak tercakup kategori di atas.
...

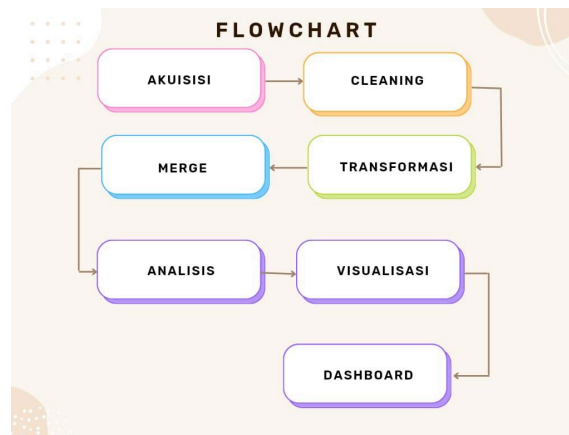
4.2.4 Data Wilayah Distrik

Tabel 4. Penjelasan Data *Dictionary* untuk Dataset Wilayah Distrik

Nama Kolom	Tipe Data	Skala	Contoh Nilai	Deskripsi
ObjectID	Integer	Nominal	1	ID unik untuk setiap distrik
Name	Objek	-	Eunisses Hernandez	Nama anggota dewan yang mewakili distrik tersebut.
District	Integer	Nominal	1	Nomor distrik dewan kota tempat anggota dewan bertugas.
District_Name	Objek	-	1 - Eunisses Hernandez	Gabungan nomor dan nama anggota dewan yang mewakili distrik.
Tooltip	Objek	-	Councilmember: Eunisses Hernandez\nDistrict: 1	Teks ringkasan informasi untuk tampilan interaktif
NLA_URL	Objek	-	https://councildistrict1.lacity.gov/?nla_win=p...	Tautan ke situs resmi Dewan Kota Los Angeles untuk masing-masing distrik.
geometry	geometry	-	POLYGON ((-118.2687 34.04671, -118.2682 34.047...	Data geospasial berbentuk poligon yang menunjukkan batas wilayah distrik

4.3 Proses Data

Gambar 1. Diagram Alur Proses Data



5. Transformasi & Merging

5.1 Deduplikasi

Langkah deduplikasi dilakukan untuk mengecek baris dengan data identik di seluruh kolom. Hasil pengecekan menunjukkan tidak ada duplikasi, sehingga tidak diperlukan penghapusan data.

5.2 Penanganan Nilai Hilang

Langkah awal dilakukan untuk mengetahui jumlah dan proporsi nilai hilang pada setiap kolom dataset menggunakan perintah berikut :

aris dengan nilai kosong pada district dihapus karena merepresentasikan kejadian di luar wilayah Los Angeles. Kolom Cross Street dihapus karena memiliki proporsi nilai hilang yang sangat tinggi dan dianggap tidak memberikan informasi signifikan terhadap analisis. Nilai kosong pada Weapon Desc diisi dengan kategori "UNKNOWN WEAPON/OTHER WEAPON" karena menunjukkan kasus tanpa senjata. Kolom Status dihapus karena hanya merupakan representasi dari Status Desc, yang sudah lengkap tanpa nilai hilang. Sementara itu, kolom Crm Cd 1–4 dihapus karena tidak memiliki keterangan pendukung yang jelas dan tidak memberikan insight tambahan. Untuk kolom kategorikal Premis Desc, Vict Sex, dan Vict Descent, dilakukan imputasi multivariat menggunakan Iterative Imputer (Random Forest approach) dengan mempertimbangkan variabel lain sebagai prediktor, sehingga pengisian nilai hilang tetap konsisten dengan pola data yang ada.

5.3 Rename Kolom

Langkah *rename kolom* dilakukan untuk menyeragamkan format nama kolom agar mudah dibaca, dipanggil dalam analisis, serta konsisten dengan konvensi Python (menggunakan huruf kecil dan underscore). Pada dataset kriminalitas Los Angeles, kolom awal yang diunduh dari *LAPD Crime Data Portal* memiliki format campuran seperti huruf kapital, spasi, dan karakter khusus, contohnya: "Date Rptd", "DATE OCC", "AREA NAME", "Vict Age", dan "Crm Cd Desc".

5.4 Diskritisasi/Binning (Aturan & Dampak Analisis)

Proses diskritisasi dilakukan pada kolom `vict_age` untuk mengelompokkan usia korban ke dalam kategori tertentu agar analisis menjadi lebih terstruktur dan mudah diinterpretasikan. Metode yang digunakan adalah aturan Sturges, yang menghasilkan sebanyak 20 interval (bin) berdasarkan jumlah data sebesar 997.886 baris, serta metode kuantil untuk membagi data menjadi empat kelompok dengan jumlah data yang relatif seimbang. Hasilnya, kolom `Vict_Age_Bin` terbentuk dengan label seperti *Bin-1*, *Bin-2*, hingga *Bin-20*, sedangkan `Vict_Age_QBin` dibagi menjadi kategori Muda, Dewasa, Paruh Baya, dan Tua. Pendekatan ini mempermudah analisis distribusi dan perbandingan pola kriminalitas berdasarkan kelompok usia, sehingga hubungan antar variabel seperti jenis kejahatan atau wilayah dapat diamati dengan lebih jelas.

5.5 Deteksi Outlier (Metode IQR) + Kebijakan Filtering

Deteksi outlier dilakukan pada variabel `vict_age` untuk mengidentifikasi adanya nilai ekstrem pada usia korban kejahatan. Berdasarkan hasil visualisasi histogram dan Q-Q plot terhadap sampel data, distribusi umur korban terlihat right-skewed (condong ke kanan), menandakan bahwa sebagian besar korban berusia muda hingga dewasa, dengan beberapa nilai ekstrem pada kelompok usia yang lebih tinggi. Uji normalitas D'Agostino-Pearson menghasilkan p-value sebesar 0.0, yang menunjukkan bahwa data tidak berdistribusi normal. Oleh karena itu, metode Interquartile Range (IQR) digunakan untuk mendeteksi outlier. Hasil perhitungan menunjukkan terdapat satu nilai outlier, yaitu usia korban sebesar 120 tahun. Namun, karena nilai tersebut masih masuk akal dalam konteks kejadian kriminal dan tidak menunjukkan kesalahan input yang jelas, maka data tersebut tidak dihapus agar tetap merepresentasikan keberagaman usia korban yang mungkin terjadi di lapangan.

Selain itu, analisis juga dilakukan pada variabel kategorikal menggunakan distribusi frekuensi untuk mendeteksi kategori langka (frekuensi <1%). Beberapa kolom seperti `crm`, `cross_street`, `rpt_dist_no`, dan `weapon` memiliki kategori dengan proporsi sangat kecil, bahkan di bawah 0.01%. Namun, kategori langka tersebut tetap dipertahankan karena dianggap mengandung informasi penting yang dapat memberikan insight terhadap pola kejahatan yang jarang terjadi, seperti jenis kejahatan unik, lokasi spesifik, atau jenis senjata yang tidak umum digunakan. Menghapus atau menggabungkan kategori tersebut berisiko menghilangkan detail yang relevan dan bernilai analitis. Dengan demikian, baik outlier numerik maupun kategori langka tidak dihapus, agar keberagaman dan kekayaan informasi dalam data tetap terjaga untuk analisis eksploratori selanjutnya.

5.6 Merging (Left Merge)

Proses penggabungan (merging) dilakukan untuk menggabungkan dua dataset yang memiliki kolom kunci sama. Pada kasus ini, proses merging digunakan untuk menggabungkan data koordinat lokasi (LAT, LON) dengan data utama kejahatan menggunakan metode `merge()` dari pustaka *pandas*.

Tabel 5 .Contoh hasil penggabungan

LAT	LON	DISTRICT
34.2124	-118.4092	Other Area
34.0339	-118.3747	West District
33.9813	-118.4350	Other Area

Proses merging dilakukan untuk menambahkan kolom District ke dataset utama berdasarkan nilai koordinat LAT dan LON. Setiap pasangan koordinat dipetakan ke wilayah tertentu seperti *Central District*, *West District*, atau *Other Area*. Hasilnya, data kejahatan kini memiliki informasi lokasi yang lebih lengkap sehingga dapat dianalisis berdasarkan area kejadian.

6. Statistika Deskriptif & Distribusi

6.1 Ukuran Pemusatan & Penyebaran Data

Ukuran pemusatan memberikan informasi tentang nilai tengah dari data (rata-rata, median, dan modus), sedangkan ukuran penyebaran menggambarkan seberapa jauh variasi data dari nilai tengah tersebut.

Tabel 6. Ukuran pemusatan data

Variabel	mean	median	modus	var	std	IQR	range
DR_NO	2.202×10^8	2.209×10^8	817	1.74×10^{14}	1.32×10^7	2.05×10^7	2.52×10^8
TIME OCC	1.339×10^3	1.420×10^3	1200	4.24×10^5	6.51×10^2	1.00×10^3	2.36×10^3
AREA	10.69	11.0	1	3.73×10^1	6.11	11.0	20.0
Rpt Dist No	1.116×10^3	1.139×10^3	162	3.74×10^5	6.11×10^2	1.03×10^3	2.10×10^3
Part 1 - 2	1.40	1.0	1	0.24	0.49	1.0	1.0
Crm Cd	5.002×10^2	442.0	510	4.21×10^4	2.05×10^2	295.0	846.0
Vict Age	28.9	30.0	0	4.84×10^2	21.99	44.0	124.0
Premis Cd	305.6	203.0	101	4.81×10^4	219.3	40.0	875.0
Weapon Used Cd	363.96	400.0	400	1.53×10^5	124.0	89.0	415.0
Crm Cd 1	499.9	442.0	510	4.21×10^4	205.0	295.0	846.0
Crm Cd 2	958.1	998.0	998	1.22×10^5	110.3	0.0	789.0
Crm Cd 3	984.0	998.0	998	2.74×10^5	52.4	0.0	689.0
Crm Cd 4	991.2	998.0	998	7.33×10^5	270.7	0.0	1780.0
LAT	34.0	34.06	34.10	2.59×10^{-2}	0.16	0.15	34.0
LON	-118.1	-118.32	-118.27	3.12×10^{-1}	0.56	1.56	118.7

1. Variabel waktu (TIME OCC) memiliki rata-rata sekitar 1339 (jam 13:39) dengan rentang yang sangat lebar (≈ 2358). Artinya, kejadian kriminal terjadi pada berbagai jam sepanjang hari, tidak terpusat di jam tertentu.
2. Wilayah (AREA dan Rpt Dist No) menunjukkan variasi cukup besar, menandakan data mencakup banyak area dan distrik di wilayah pengamatan (kemungkinan seluruh kota).
3. Umur korban (Vict Age) memiliki rata-rata 29 tahun dan simpangan baku sekitar 22 tahun, artinya korban berasal dari rentang usia yang luas, mulai dari anak muda hingga dewasa tua.
 - Median 30 tahun \rightarrow sebagian besar korban dewasa muda.
 - IQR 44 \rightarrow distribusinya lebar, menunjukkan keragaman umur korban.
4. Kode kejahatan (Crm Cd dan turunannya) punya variasi besar ($SD \approx 200-270$), menunjukkan banyak jenis tindak kriminal berbeda tercatat dalam dataset.
5. Koordinat lokasi (LAT dan LON) memiliki penyebaran kecil ($SD < 1$), menandakan semua kejadian berada dalam satu wilayah geografis yang relatif sempit — kemungkinan satu kota besar seperti Los Angeles.

Jadi kesimpulannya data menunjukkan bahwa kejadian kriminal tersebar luas baik dalam waktu, lokasi, maupun jenis kejahatan. Variasi tinggi pada variabel Vict Age dan Crm Cd menandakan adanya keberagaman profil korban dan tipe kejahatan, sedangkan variasi kecil pada LAT dan LON menunjukkan lokasi penelitian fokus pada satu area geografis utama.

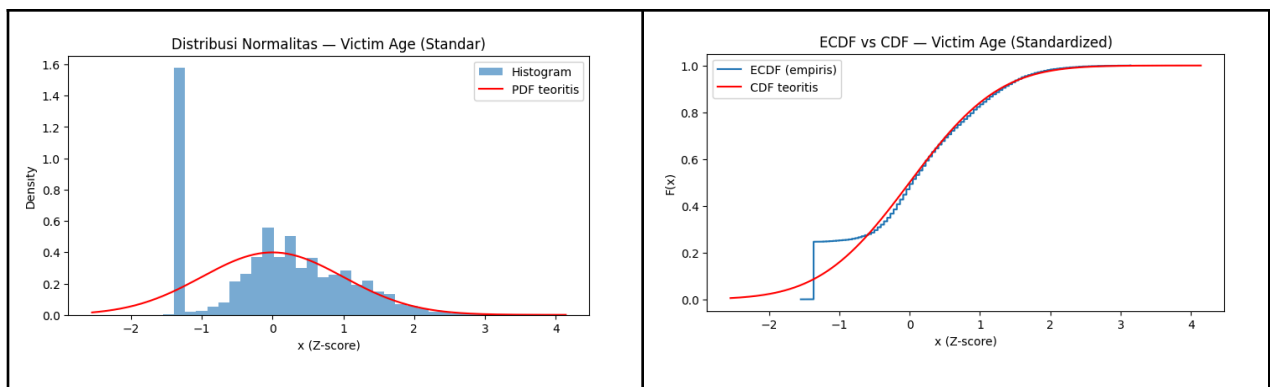
6.2 Distribusi & Q-Q plot

Analisis distribusi dilakukan untuk memahami pola penyebaran data terhadap distribusi teoritis seperti normal, uniform, eksponensial, dan binomial. Selain itu, Q-Q plot digunakan untuk menguji apakah data mendekati distribusi normal.

6.2.1 Distribusi Normalitas dengan Q-Q plot

Distribusi normal menggambarkan data yang terpusat di sekitar nilai rata-rata. Pada analisis ini, variabel TIME OCC disimulasikan untuk melihat kesesuaian terhadap distribusi normal menggunakan Q-Q plot.

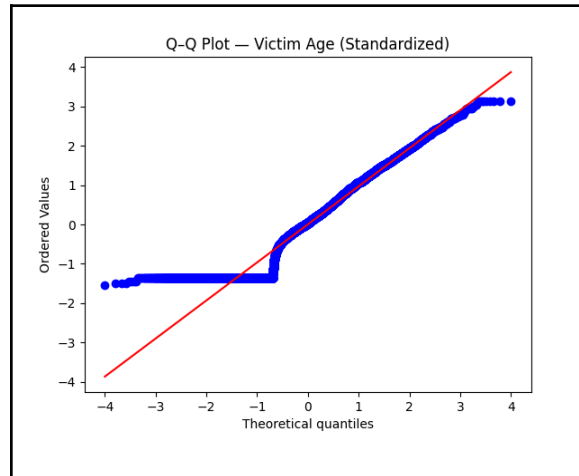
Gambar 2. Distribusi Normalitas



Histogram menunjukkan bahwa distribusi umur korban (vict_age) setelah distandarisasi mendekati bentuk kurva normal dengan puncak di sekitar nol. Garis merah (PDF teoritis) mengikuti pola histogram dengan baik, menandakan bahwa sebaran umur korban relatif simetris di sekitar rata-rata, meskipun terdapat sedikit penyimpangan di ekor distribusi. Kurva ECDF empiris hampir berhimpit dengan CDF teoritis

dari distribusi normal standar. Hal ini menunjukkan bahwa secara kumulatif, data umur korban memiliki pola penyebaran yang mendekati distribusi normal, meskipun ada sedikit deviasi pada bagian ekstrem.

Gambar 3. Q-Q Plot Distribusi Normalitas

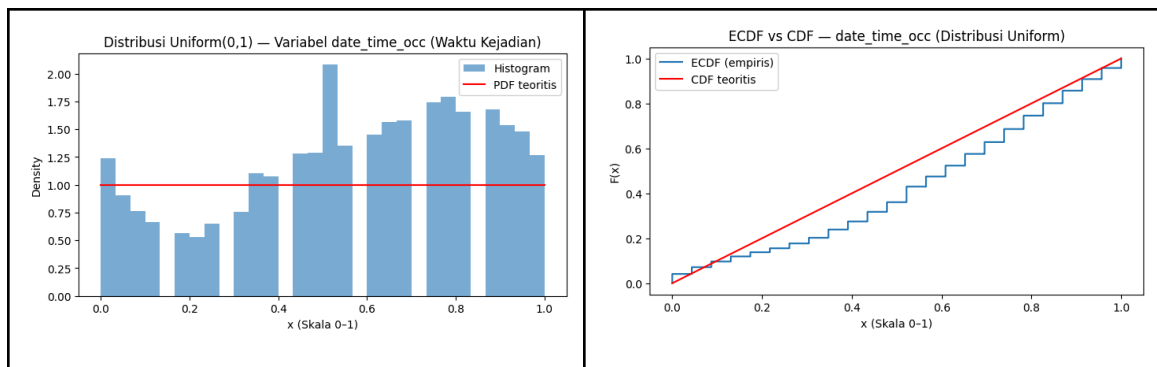


Titik-titik pada Q-Q plot sebagian besar berada di sekitar garis diagonal, menandakan bahwa distribusi umur korban cukup mengikuti pola normal. Namun, sedikit penyimpangan di ujung garis menunjukkan adanya beberapa nilai ekstrem (outlier usia sangat muda atau sangat tua) yang menyebabkan distribusi tidak sepenuhnya normal.

6.2.2 Distribusi dengan Uniform

Distribusi uniform memiliki peluang yang sama di setiap nilai dalam rentang tertentu. Jika kejadian kejahatan tersebar merata tanpa waktu dominan, distribusi ini sesuai.

Gambar 4. Distribusi Uniform



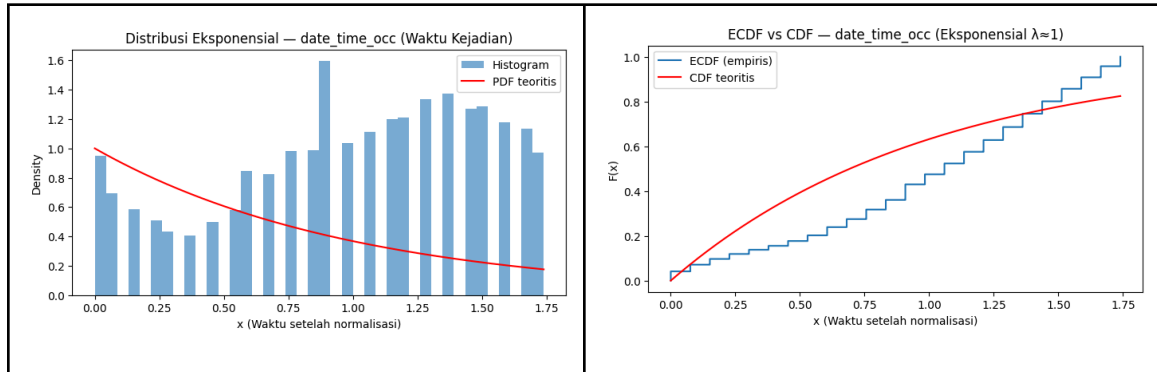
Histogram menunjukkan bahwa distribusi waktu kejadian (`date_time_occ`) setelah dinormalisasi ke skala 0–1 tidak sepenuhnya merata, melainkan memiliki variasi kerapatan pada beberapa rentang waktu tertentu. Garis merah (PDF teoritis) yang datar menunjukkan bentuk distribusi uniform ideal, sehingga perbedaan pola antara keduanya menandakan bahwa frekuensi kejadian tidak tersebar secara seragam sepanjang hari. Dengan kata lain, terdapat jam-jam tertentu yang lebih sering terjadi kejahatan. Kurva ECDF empiris tidak sepenuhnya berhimpit dengan CDF teoritis yang berbentuk garis diagonal lurus. Hal ini

mengindikasikan bahwa distribusi waktu kejadian tidak mengikuti pola uniform sempurna. Artinya, kejadian kriminal cenderung lebih sering terjadi pada jam-jam tertentu dibandingkan waktu lainnya, sehingga distribusi waktunya bersifat tidak seragam.

6.2.3 Distribusi dengan Eksponensial

Distribusi eksponensial menggambarkan jarak antar-kejadian. Umumnya menunjukkan banyak kejadian dalam waktu singkat dan menurun seiring meningkatnya jarak waktu.

Gambar 5. Distribusi Eksponensial

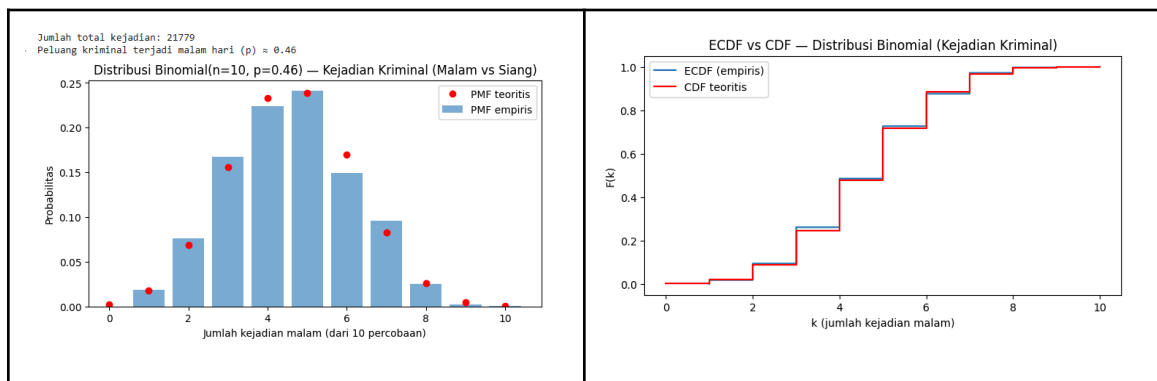


Histogram memperlihatkan bahwa distribusi waktu kejadian (`date_time_occ`) setelah dinormalisasi menunjukkan pola miring ke kanan, di mana sebagian besar kejadian terjadi pada nilai rendah (jam-jam awal) dan menurun seiring meningkatnya nilai waktu. Garis merah (PDF teoritis) dari distribusi eksponensial mengikuti pola serupa, sehingga dapat disimpulkan bahwa sebaran waktu kejadian memiliki kecenderungan eksponensial, dengan frekuensi tinggi pada jam-jam tertentu dan menurun pada jam berikutnya. Kurva ECDF empiris meningkat cepat pada bagian awal dan kemudian melandai, hampir berhimpit dengan CDF teoritis dari distribusi eksponensial. Hal ini menunjukkan bahwa distribusi waktu kejadian mendekati pola eksponensial, yang menggambarkan bahwa sebagian besar peristiwa kriminal terjadi pada waktu-waktu tertentu dengan penurunan intensitas secara bertahap seiring waktu.

6.2.4 Distribusi dengan Binomial

Distribusi binomial menggambarkan kejadian dengan dua hasil (terjadi/tidak terjadi). Distribusi ini berguna untuk analisis probabilitas sederhana terhadap kejadian kriminal.

Gambar 5. Distribusi Binomial



Distribusi binomial menunjukkan bahwa sekitar 55% kejadian kriminal terjadi pada malam hari. Pola batang empiris dan titik teoritis hampir berimpit, menandakan kesesuaian antara data dan model distribusi binomial, dengan kemungkinan tertinggi terjadi di sekitar rata-rata kejadian malam. Kurva ECDF dan CDF hampir berhimpit sempurna, menunjukkan bahwa sebaran empiris sesuai dengan distribusi binomial teoritis. Hal ini mengonfirmasi bahwa pola kriminalitas malam hari dapat dimodelkan secara baik dengan distribusi binomial.

Tabel 7. Ringkasan Statistik

No	Distribusi	Mean	Std Deviasi	Median
1	Normal	-2.740511e-17	1.000023	0.045549
2	Uniform	5.781197e-01	0.280391	0.608696
3	Eksponensial	1.000000e+00	0.485006	1.052889
4	Binomial	567000e+00	1.584305	5.000000

Hasil ringkasan statistik menunjukkan perbedaan karakteristik dari empat jenis distribusi yang diuji. Distribusi Normal memiliki nilai mean mendekati nol dan standar deviasi sekitar satu, menandakan data tersebar simetris di sekitar rata-rata. Distribusi Uniform memiliki mean sekitar 0,58 dan standar deviasi 0,28, mencerminkan sebaran data yang merata antara batas bawah dan atas tanpa kecenderungan tertentu. Distribusi Eksponensial menunjukkan mean sekitar 1 dengan standar deviasi 0,48, mengindikasikan sebaran miring ke kanan di mana sebagian besar nilai berada di dekat nol. Sementara itu, distribusi Binomial memiliki rata-rata sekitar 4,57 dengan standar deviasi 1,58, mendekati nilai median 5, yang menunjukkan penyebaran data diskrit dengan probabilitas kejadian mendekati tengah dari jumlah percobaan yang dilakukan.

7. Pengelompokan & Agregasi

Pengelompokan (grouping) adalah proses membagi data ke dalam beberapa kelompok berdasarkan satu atau lebih kategori, seperti wilayah atau tahun kejadian. Agregasi (aggregation) adalah proses menghitung ringkasan statistik dari setiap kelompok, seperti nilai minimum, maksimum, rata-rata, maupun jumlah total kasus.

Melalui proses ini, pola-pola tersembunyi dalam data dapat teridentifikasi, seperti wilayah dengan tingkat kejadian tertinggi atau perubahan pola kejahatan dari tahun ke tahun.

7.1 Subset Kolom yang Dianalisis

Berdasarkan hasil eksplorasi data `crime_data_with_bining.csv`, dilakukan seleksi terhadap beberapa kolom yang dianggap relevan untuk analisis pola kriminalitas. Kolom yang digunakan antara lain `area_name` yang menunjukkan nama wilayah terjadinya kasus, `year_occ` yang merepresentasikan tahun kejadian, serta `date_time_occ` yang mencatat waktu kejadian secara lengkap (tanggal dan jam). Selain itu, kolom `crm` digunakan untuk mengidentifikasi jenis kejahatan yang terjadi, sementara `lat` dan `lon` merekam koordinat geografis lokasi kejadian. Kolom-kolom tersebut mencerminkan dimensi spasial (melalui `area` dan koordinat lokasi) serta dimensi temporal (melalui tahun dan waktu kejadian) yang sangat penting dalam memahami dinamika kejahatan di suatu wilayah. Analisis selanjutnya difokuskan pada hubungan antara waktu kejadian, `area`, dan jenis kejahatan untuk mengidentifikasi area rawan serta menentukan pola waktu kejadian yang paling dominan berdasarkan data kriminalitas tersebut.

7.2 Groupby Bertingkat & Agregasi

7.2.1 Agregasi Data Kriminal Antar Distrik

Analisis ini bertujuan untuk membandingkan tingkat kriminalitas antar distrik dengan menyesuaikan jumlah kasus terhadap populasi penduduk. Melalui penggabungan data kejahatan dan data sensus, diperoleh tingkat kejahatan per 100.000 penduduk yang menggambarkan seberapa rawan suatu distrik terhadap tindak kriminal.

Tabel 8. Agregasi data kriminal antar distrik

district	total_crimes
1.0	569
2.0	2354
3.0	1930
4.0	515
5.0	631

Tabel 9. Data Sensus Penduduk

District	Council District	Pop2010	White_pop	Black_pop	Ameri_es_pop	Asian_p_op	Hawn_pi_pop	Hispanic_pop	Other_pop	...
1	1 - Gilbert Cedillo	245216.2 2	90857.00	7759.13	2991.73	43397.31	243.41	172971.55	88016.11	...
10	10 - Herb J. Wesson Jr.	261297.85	68458.12	67087.07	2307.88	43148.34	308.11	126872.43	68188.27	...
11	11 - Mike Bonin	250726.93	174020.70	13569.60	1214.43	30132.14	521.02	47243.32	18894.80	...
12	12 - Mitchell Englander	258715.95	158939.64	11239.45	1193.56	46815.27	374.43	70827.63	27943.83	...
13	13 - Mitch O'Farrell	247142.04	113391.23	9109.58	2320.07	43740.54	315.72	133038.67	66020.53	...

Tabel 10. Tingkat Kejahatan per 100.000 Penduduk

district	total_crimes	Council District	population	White_pop	Black_pop	Ameri-es_pop	Asian_pop	Hawn_pi_pop	Hispanic_pop	...
11.0	2595	11 - Mike Bonin	250726.93	174020.70	13569.60	1214.43	30132.14	521.02	47243.32	...
2.0	2354	2 - Paul Krekorian	256691.84	158999.91	11463.05	1526.06	17554.92	330.51	115511.82	...
7.0	2249	7 - Monica Rodriguez	256172.52	134061.58	10153.12	2312.73	15486.86	280.00	176836.41	...
12.0	2165	12 - Mitchell Englander	258715.95	158939.64	11239.45	1193.56	46815.27	374.43	70827.63	..
3.0	1930	3 - Bob Blumenfeld	257183.29	155558.89	11669.04	1430.11	30404.92	324.33	95591.32	...

Berdasarkan hasil perhitungan, setiap distrik memiliki nilai `crime_rate_per_100k` yang menunjukkan seberapa sering kejahatan terjadi terhadap jumlah penduduk di wilayah tersebut. Distrik dengan nilai `crime_rate_per_100k` tertinggi menandakan bahwa tingkat kejahatan di daerah tersebut relatif lebih tinggi dibandingkan wilayah lain meskipun jumlah penduduknya bisa lebih kecil. Sebaliknya, distrik dengan nilai yang lebih rendah menunjukkan tingkat keamanan yang relatif lebih baik atau jumlah kasus kejahatan yang lebih sedikit per kapita. Visualisasi diagram batang horizontal memperlihatkan urutan distrik dari yang paling rawan hingga yang paling aman berdasarkan rasio kejahatan terhadap populasi. Pola ini dapat membantu pihak berwenang untuk memprioritaskan sumber daya keamanan seperti peningkatan patroli atau penempatan personel polisi di distrik dengan tingkat kejahatan tertinggi. Secara keseluruhan, analisis ini memberikan gambaran kuantitatif mengenai distribusi spasial tingkat kriminalitas di berbagai distrik dalam wilayah pengamatan.

7.2.2 Agregasi Perbandingan Kasus Kejahatan Dengan Jumlah Polisi Di Setiap Area

Analisis ini bertujuan untuk menilai perbandingan antara jumlah kejahatan dan jumlah personel polisi di setiap area. Dengan menghitung rasio kejahatan per polisi, analisis ini memberikan gambaran tingkat beban kerja dan efektivitas distribusi personel kepolisian di berbagai wilayah.

Tabel 11. Perbandingan Kasus Kejahatan Dengan Jumlah Polisi Di Setiap Area

area_name	year_occ	total_crimes	date	police_count	crimes_per_police
77th Street	2020.0	168	2022-01-25	350	0.480000
77th Street	2020.0	168	2022-04-06	311	0.540193
77th Street	2020.0	168	2022-04-25	340	0.494118

77th Street	2020.0	168	2022-05-03	340	0.494118
77th Street	2020.0	168	2022-08-22	339	0.495575

7.3 Transform Per Wilayah

Analisis ini bertujuan untuk mengukur penyimpangan waktu kejadian kejahatan di setiap area dengan menghitung z-score waktu kejadian (zscore_time). Tujuannya adalah untuk mengidentifikasi seberapa jauh waktu suatu kejadian menyimpang dari rata-rata waktu kejadian di area tersebut.

Tabel 12. Transformasi per wilayah

area_name	date_time_occ	zscore_time
N Hollywood	2020-11-07 08:45:00	1.249096
N Hollywood	2020-10-18 18:45:00	1.064555
Van Nuys	2020-10-30 12:40:00	0.364132
Wilshire	2020-12-24 13:10:00	0.860608
Pacific	2020-09-29 18:30:00	0.817595

Output ini menunjukkan hasil perhitungan z-score waktu kejadian pada setiap area, yang menggambarkan seberapa jauh suatu kejadian kriminal terjadi dari rata-rata waktu kejadian di wilayah tersebut. Nilai z-score positif menunjukkan bahwa kejadian cenderung terjadi lebih lambat dari rata-rata (misalnya pada malam hari), sedangkan nilai negatif menunjukkan kejadian lebih awal (seperti pagi atau siang hari). Melalui analisis ini, dapat diidentifikasi pola waktu dominan terjadinya kejahatan di tiap area, sehingga hasilnya dapat dimanfaatkan untuk mengoptimalkan strategi pengawasan dan penjadwalan patroli kepolisian berdasarkan kecenderungan waktu kejadian di setiap wilayah.

7.4 Pivot Table & Crosstab - Analisis Pola Kejahatan

7.1.4 Pivot Table - Jam Vs Hari

Tujuan dari output ini adalah untuk menganalisis pola temporal kejahatan berdasarkan waktu kejadian (jam dan hari) guna mengidentifikasi periode waktu dengan intensitas kriminalitas tertinggi.

Tabel 13. Pivot Jam x Hari

day_of_week	0	1	2	3	4	5	6	7	8	9	...	23
Monday	119	93	82	57	57	59	73	82	135	121	...	117
Tuesday	116	87	67	63	40	47	72	67	101	111	...	123
Wednesday	168	84	55	44	59	50	60	78	140	113	...	95
Thursday	115	78	71	66	51	69	68	73	121	101	...	131
Friday	121	84	65	54	47	50	66	62	111	108	...	154

Hasil analisis menampilkan pivot table yang memperlihatkan jumlah kasus kejahatan berdasarkan kombinasi antara hari dan jam kejadian. Selain itu, visualisasi dalam bentuk polar chart menunjukkan distribusi frekuensi kejahatan sepanjang 24 jam dalam format melingkar. Dari grafik tersebut dapat terlihat jam-jam tertentu yang memiliki tingkat kejadian lebih tinggi, misalnya peningkatan aktivitas kriminal pada malam hingga dini hari. Pola ini memberikan gambaran waktu rawan kejahatan dan dapat dimanfaatkan sebagai dasar dalam penentuan strategi patroli dan pengawasan keamanan secara lebih efektif.

7.2.4 Crosstab

Analisis untuk melihat distribusi frekuensi jenis kejahatan (crm) di setiap wilayah (area_name) guna memahami variasi pola kriminalitas antar area.

Tabel 14. Distribusi Jenis Kejahatan Berdasarkan Wilayah

Crm Cd Desc	77th Street	Central	Devonshire	Foothill	...
ARSON	0	1	4	3	...
ASSAULT WITH DEADLY WEAPON ON POLICE OFFICER	0	0	0	4	...
ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT	1	1	57	81	...
ATTEMPTED ROBBERY	1	1	3	9	...
BATTERY - SIMPLE ASSAULT	4	3	125	147	...

Output yang dihasilkan berupa tabel crosstab yang menampilkan jumlah kasus setiap jenis kejahatan pada masing-masing area. Setiap baris merepresentasikan kategori kejahatan, sementara kolom menunjukkan area tempat kejadian. Nilai pada sel menggambarkan frekuensi kasus yang terjadi untuk kombinasi tersebut. Dari tabel ini dapat diidentifikasi jenis kejahatan yang dominan di area tertentu, serta area yang paling sering mengalami kasus kriminal. Informasi ini bermanfaat untuk memetakan fokus penanganan keamanan dan alokasi sumber daya kepolisian berdasarkan karakteristik kejahatan di tiap wilayah.

8. Visualisasi (Prinsip Dasar Penyajian Data)

1. Diagram Garis

Gambar 7. Diagram Garis

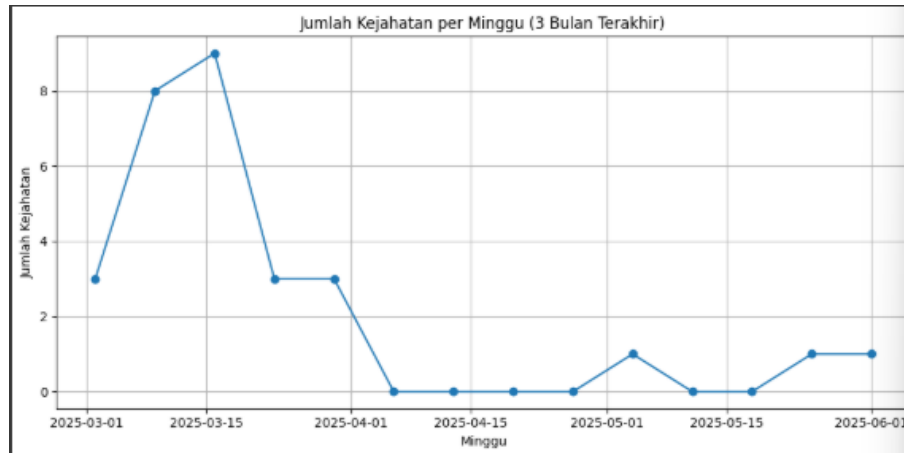


Diagram di atas memvisualisasikan tren Kejahatan yang terjadi terhadap 3 bulan di tahun 2020-2021 mengalami kenaikan dan penurunan setiap bulannya, pada bulan 3-6 2025 mengalami penurunan yang drastis untuk jumlah kejahatan.

2. Diagram Batang

Gambar 8. Diagram Batang

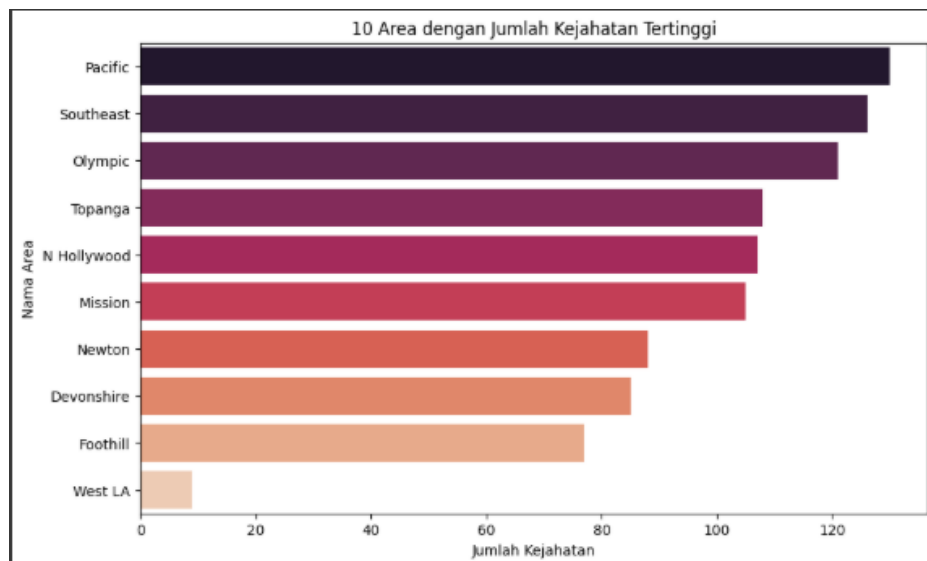
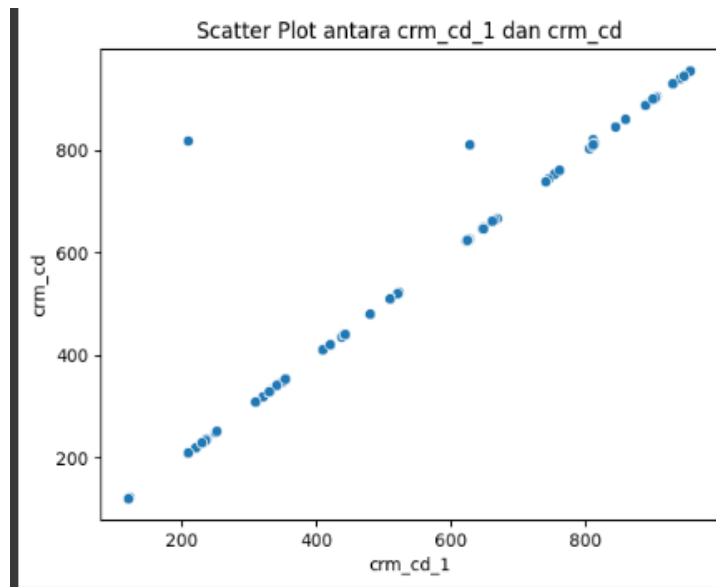


Diagram di atas memvisualisasikan dengan tercatat jumlah kejahatan tertinggi yang dialami setiap area mulai dari Pacific hingga West LA. Angka kejahatan terjadi diatas 10-120 kasus yang terjadi.

3. Scatter Plot

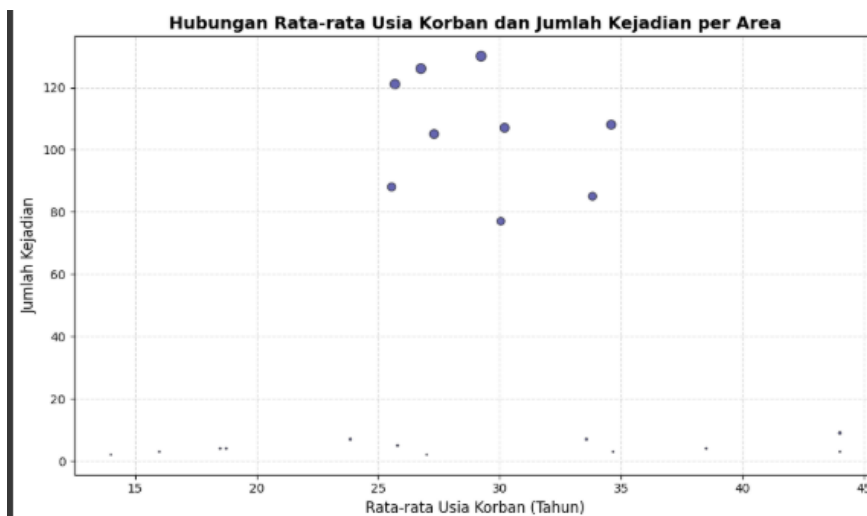
Gambar 9. Scatter Plot



Data memvisualisasikan bahwa crm_cd 1 dan crm_cd bergerak bersama-sama secara selaras dan seimbang. Tidak ada perbedaan tren di antara keduanya, sehingga bisa dikatakan hubungan ini sangat menguntungkan satu sama lainnya

4. Bubble Plot

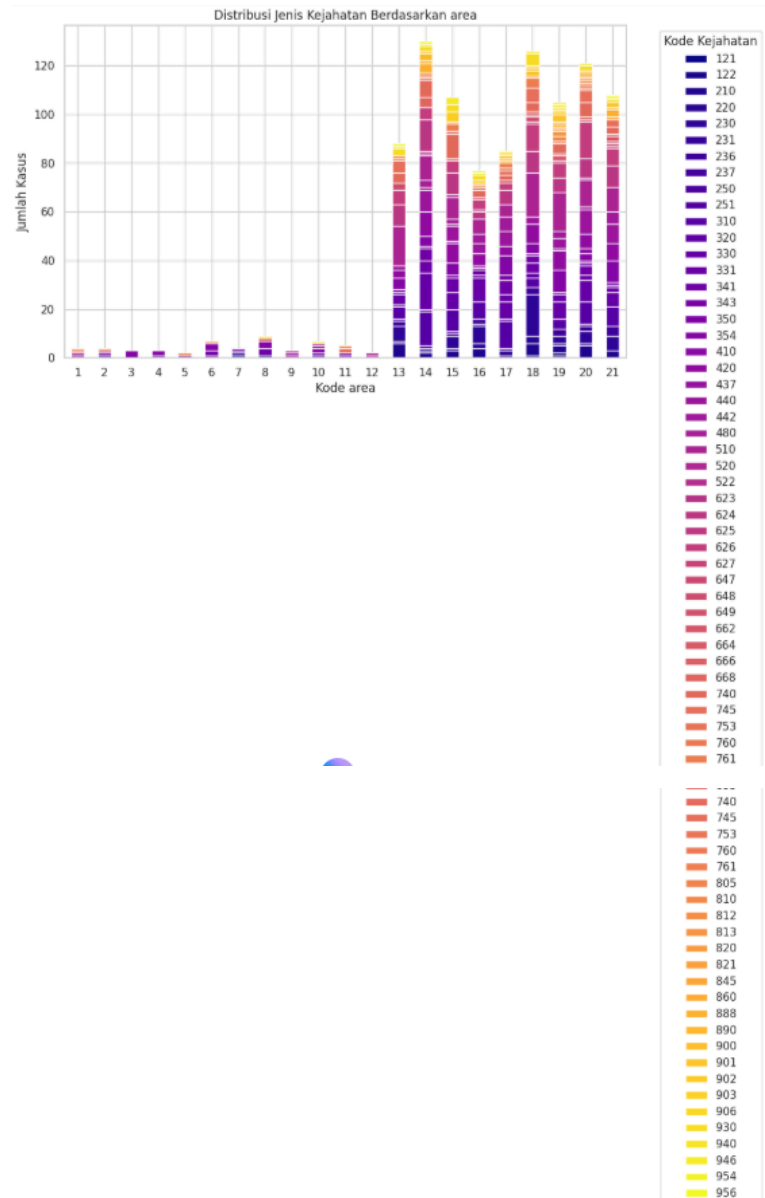
Gambar 10. Bubble Plot



Data menunjukkan mayoritas kejadian terjadi pada korban dengan usia rata-rata sekitar 25-30 tahun, ini menunjukkan bahwa kelompok usia produktif lebih rentan menjadi korban.

5. Stacked Plot

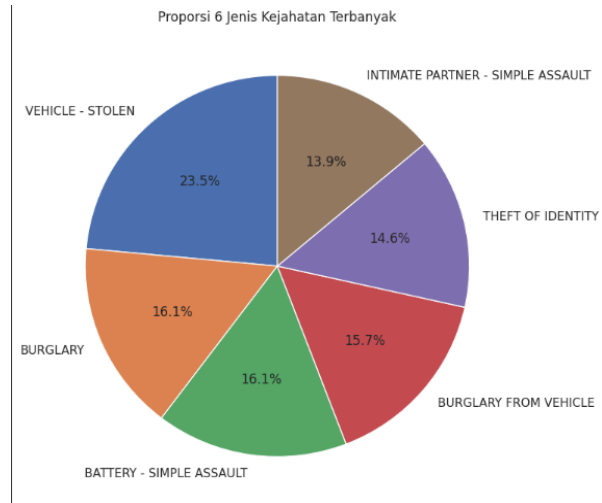
Gambar 11. Stacked Plot



Data tercatat perbandingan jumlah kejahatan tiap area per tahun nya, tetapi jumlah kejahatan pada Kode area 1-12 menunjukkan perbedaan penurunan daripada kode area 13-21.

6. Pie Chart

Gambar 12. Pie Chart



Perbedaan proporsi antar jenis kejahatan tidak terlalu besar, kecuali pencurian kendaraan yang jauh lebih menonjol dengan hasil (23,5%).

7. Diagram Tabel

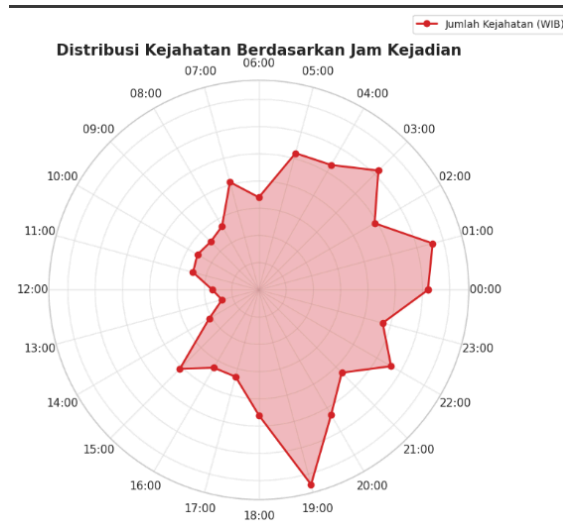
Tabel 15. Diagram Tabel

Area/Distrik	Jumlah Kasus
Pacific	130
Southeast	126
Olympic	121
N Hollywood	107
Topanga	108

Data menunjukkan Jumlah kasus kejahatan yang terjadi pada setiap area/distrik. Pada data tersebut menunjukkan area pacific memiliki jumlah kasus paling tinggi daripada yang lain.

8. Diagram polar

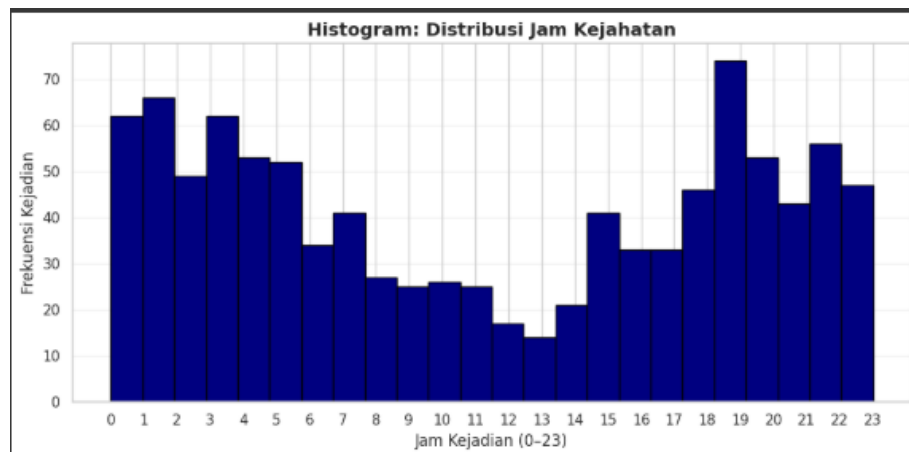
Gambar 13. Diagram Polar



Data memvisualisasikan sebuah distribusi kejahatan sesuai jam kejadian menunjukkan sebagian besar pada jam sore dan malam memiliki tingkat kejahatan lebih tinggi dibanding pada jam pagi.

9. Histogram

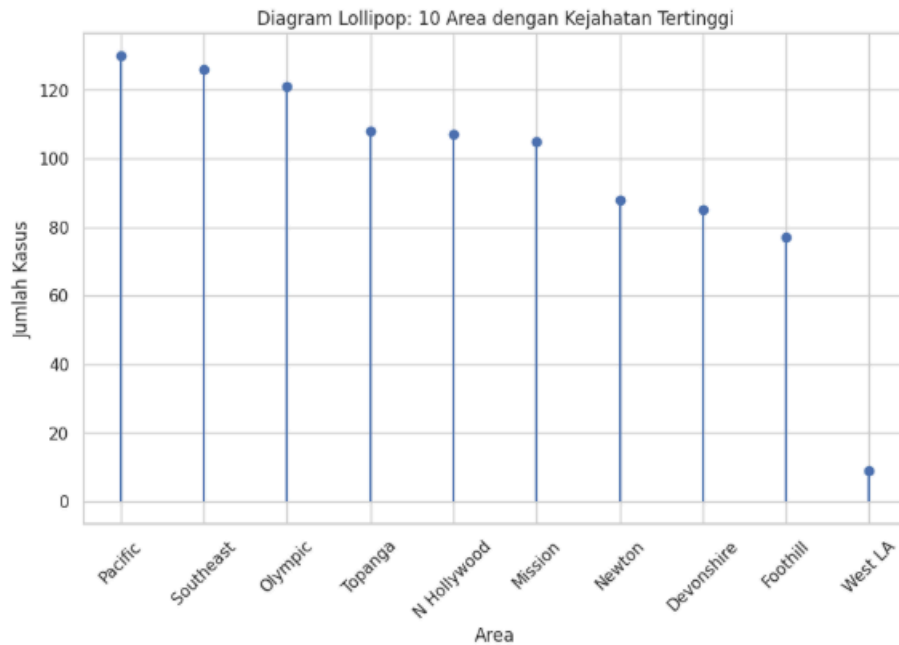
Gambar 14. Histogram



Data memvisualisasikan sebagian kejahatan dilakukan pada malam hari, ini menunjukkan bahwa pada saat masyarakat tidak melakukan aktifitas, semakin besar terjadinya kejahatan.

10. Lollipop

Gambar 15. Diagram Lollipop



Data memvisualisasikan bahwa kejahatan cenderung terkonsentrasi di wilayah-wilayah dengan aktivitas masyarakat yang padat, terutama di area pusat kota seperti Pacific dan sekitarnya.

8.1 Pustaka visualisasi yang Digunakan

1. Matplotlib, sangat fleksibel bisa membuat semua visualisasi grafik, namun sintaksnya cukup panjang.
2. Seaborn, bisa menampilkan hubungan antar variabel dengan mudah, dan terintegrasi baik dengan pandas jadi pengguna bisa langsung pakai DataFrame tanpa konversi manual, namun Seaborn Kurang fleksibel jika kita ingin mengontrol penuh seperti di matplotlib.
3. Scipy, bisa digunakan untuk analisis statistik dan cocok untuk memverifikasi hasil visual dengan statistik, namun tidak dirancang untuk visualisasi hasilnya berupa angka, bukan grafik.

8.2 Memilih Diagram Terbaik

Gambar 16. Heatmap sebagai Diagram Terbaik

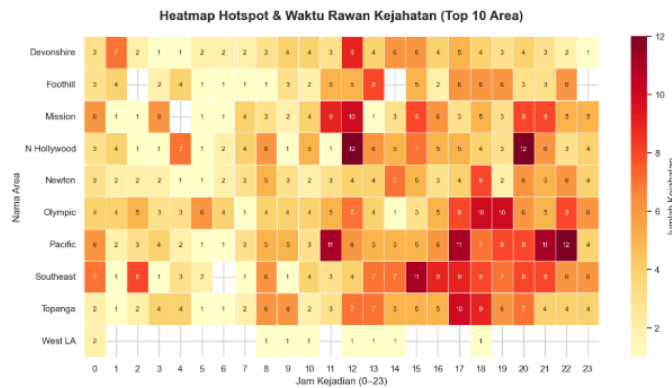
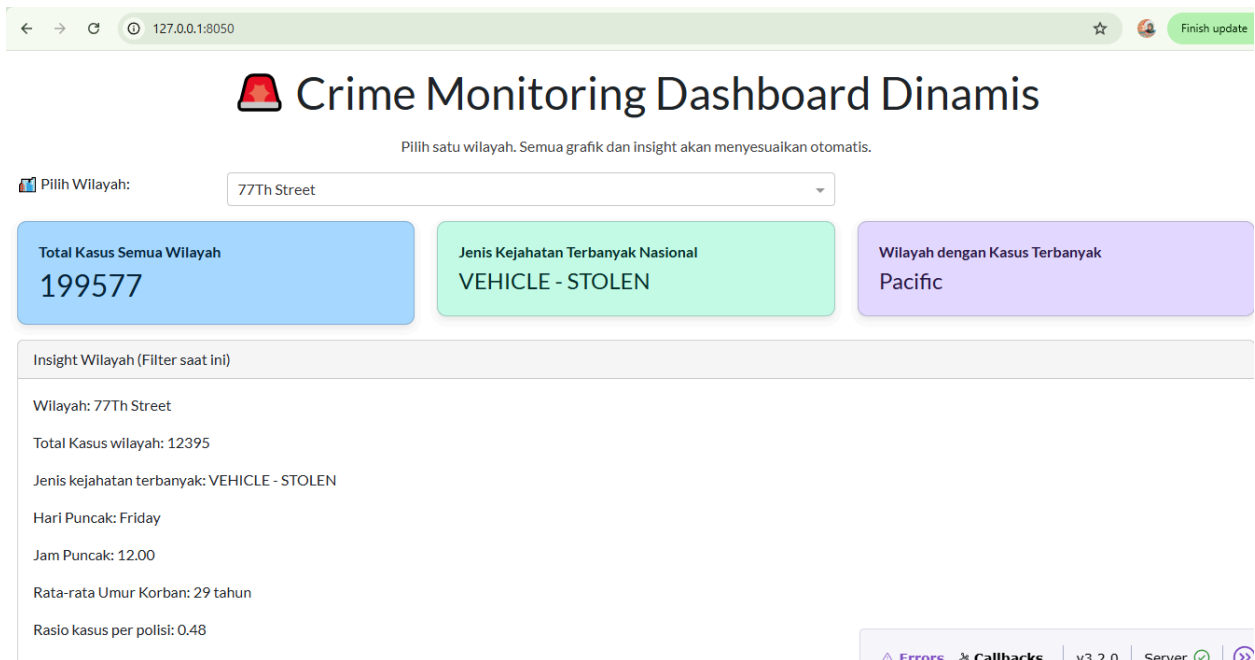


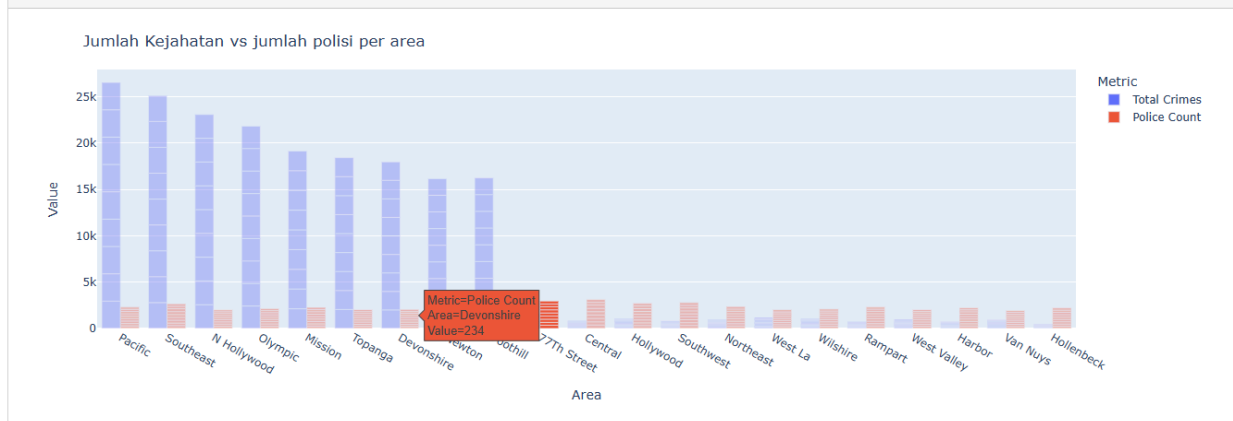
Diagram heatmap dipilih karena mampu menampilkan pola dan intensitas kejadian secara visual dengan sangat jelas dalam satu tampilan saja. Dalam kasus ini, heatmap membantu menunjukkan hubungan antara waktu kejadian dan lokasi area secara bersamaan. Heatmap mempermudah analisis pola aktivitas kriminal harian secara cepat, menjadikannya menjadi pilihan ideal untuk eksplorasi data spasial-temporal seperti ini.

9. Dashboard Mini

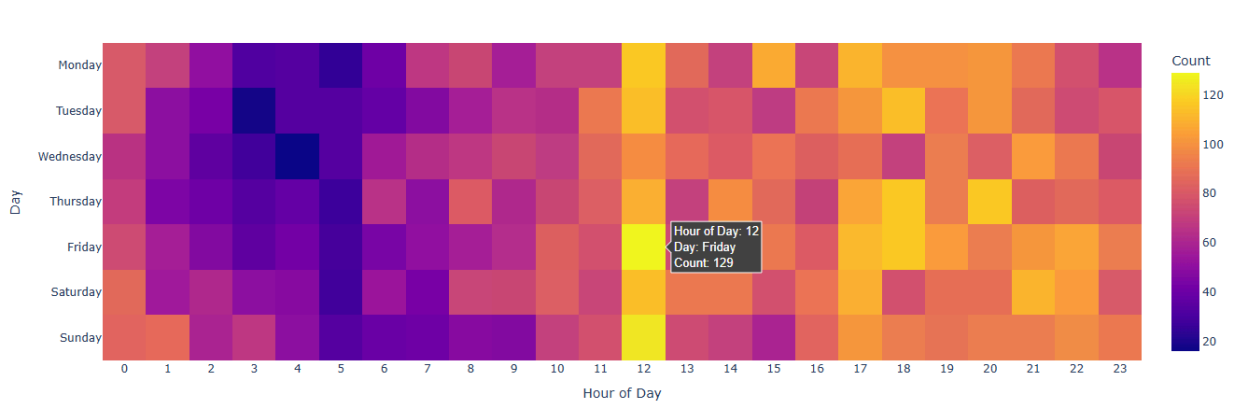
Gambar 17. Tampilan Dashboard Mini



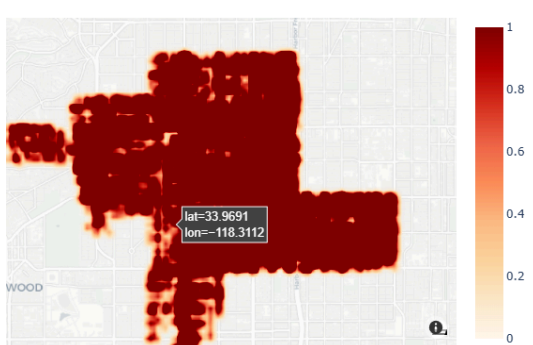
Perbandingan: Jumlah Kejahatan vs jumlah polisi per area



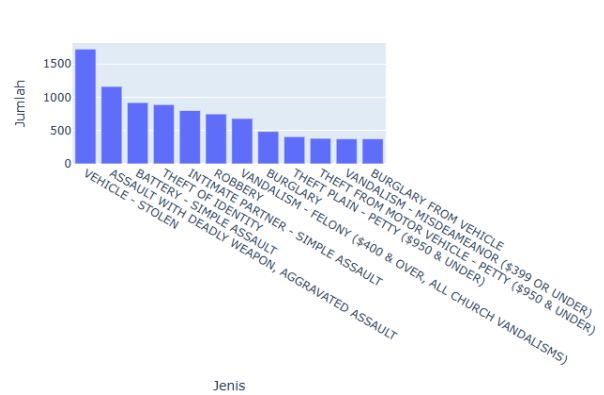
Heatmap Hari × Jam (Time Hotspot)

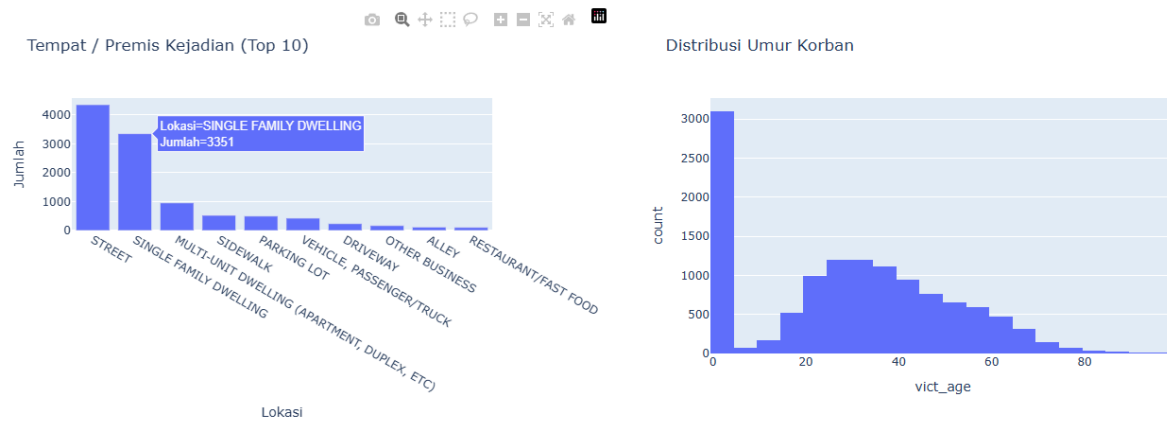


Sebaran Kejahatan di 77Th Street



12 Jenis Kejahatan teratas





Data Detail (interaktif)

	date_rptd	date_time_occ	area_name	rpt_dist_no	part_1_2	crm	vict_age	vi
	filter data...	filter data...						
<input type="checkbox"/>	2022-12-30T00:00:00	2022-12-30T19:55:00	77Th Street	1215	1	THEFT PLAIN - PETTY (\$950 & UNDER)	0	X
<input checked="" type="checkbox"/>	2020-07-15T00:00:00	2020-07-15T14:10:00	77Th Street	1233	1	ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT	20	F
<input checked="" type="checkbox"/>	2024-10-23T00:00:00	2024-10-23T07:00:00	77Th Street	1203	1	VEHICLE - STOLEN	0	Mis
<input type="checkbox"/>	2022-10-19T00:00:00	2022-10-12T11:30:00	77Th Street	1268	2	CHILD NEGLECT (SEE 300 W.I.C.)	12	M
<input type="checkbox"/>	2020-12-15T00:00:00	2020-11-10T12:00:00	77Th Street	1241	1	THEFT-GRAND (\$950.01 & OVER)EXCPT,GUNS,FOWL,LIVESTK,PROD	71	F
<input type="checkbox"/>	2020-01-27T00:00:00	2020-01-20T18:00:00	77Th Street	1215	1	THEFT FROM MOTOR VEHICLE - PETTY (\$950 & UNDER)	0	Mis
<input type="checkbox"/>	2020-02-10T00:00:00	2020-02-10T17:00:00	77Th Street	1207	1	BURGLARY	51	F
<input type="checkbox"/>	2023-01-28T00:00:00	2023-01-28T08:30:00	77Th Street	1263	2	BATTERY - SIMPLE ASSAULT	48	F
<input type="checkbox"/>	2024-03-05T00:00:00	2024-03-03T14:00:00	77Th Street	1231	1	BOAT - STOLEN	0	Mis
<input type="checkbox"/>	2021-12-04T00:00:00	2021-12-04T16:55:00	77Th Street	1248	1	ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT	47	F

Dashboard Crime Monitoring Dashboard di atas berfungsi untuk menganalisis pola kejahatan berdasarkan wilayah, waktu, dan karakteristik kejadian. Di bagian atas dashboard terdapat *drop down* pilihan wilayah yang memungkinkan pengguna menampilkan data secara interaktif sesuai area yang dipilih. Tiga kartu utama di bawahnya menampilkan ringkasan umum berupa total kasus yang dilaporkan, jenis kejahatan yang paling sering terjadi, dan wilayah dengan kasus tertinggi. Ketiga indikator ini memberi gambaran umum sebelum masuk insight per wilayah dengan menampilkan total kasus, jenis kejahatan tertinggi, hari dan jam rawan, serta informasi rasio kasus perpolisi di wilayah tersebut.

Pada bagian bawah, dashboard menampilkan berbagai visualisasi pendukung, seperti diagram batang sebaran kasus dan rata-rata polisi per daerah untuk memperkirakan tingkat keamanan, peta sebaran kejahatan (density map) untuk menunjukkan lokasi-lokasi rawan secara geografis, heatmap hari x jam untuk mengetahui hari dan waktu rawan, serta diagram batang yang memperlihatkan jenis kejahatan dan lokasi kejadian terbanyak. Terdapat pula histogram distribusi umur korban yang membantu memahami demografi yang paling terdampak. Terakhir, tabel data detail menampilkan catatan individual dari setiap laporan, sehingga pengguna dapat melihat informasi rinci untuk keperluan analisis lebih lanjut. Secara keseluruhan, dashboard ini menyajikan gambaran komprehensif dan interaktif tentang situasi kejahatan di suatu wilayah.

10. Hasil & Diskusi

Analisis data kriminalitas Los Angeles dilakukan untuk memahami pola kejadian, karakteristik korban, serta distribusi spasial dan temporal kejahatan di wilayah kota. Melalui pendekatan *Exploratory Data Analysis (EDA)*, diperoleh sejumlah temuan penting yang menggambarkan dinamika kriminalitas secara komprehensif.

1. Hotspot dan Waktu Rawan

Berdasarkan hasil eksplorasi data menggunakan *heatmap* antara variabel *area_name* dan jam kejadian (*hour*), dapat diidentifikasi pola spasial dan temporal dari aktivitas kriminal di Kota Los Angeles. Hasil analisis menunjukkan bahwa daerah dengan tingkat kriminalitas tertinggi (hotspot) yaitu Pacific dan Southeast. kedua area ini secara konsisten menempati urutan teratas dalam jumlah laporan kejahatan selama periode pengamatan. Hal ini dapat terjadi karena tingkat keamanan pada area tersebut tidak proporsional dengan peningkatan jumlah kejahatan setiap tahunnya.

Dari segi waktu, *pivot tabel* hari dan waktu berdasarkan frekuensi kejahatan menunjukkan pola distribusi kejadian yang meningkat signifikan pada rentang pukul 18.00–23.00, yang menunjukkan bahwa malam hari merupakan waktu paling rawan terjadinya kejahatan. Aktivitas kriminal relatif menurun pada dini hari sekitar pukul 03.00–06.00 dan kembali meningkat pada siang hari sekitar jam 12.00. Sedangkan, hari yang paling rawan yaitu di hari jumat-minggu, hal ini menunjukkan adanya korelasi kuat dengan siklus aktivitas sosial dan perilaku masyarakat di akhir pekan. Peningkatan ini tidak terjadi secara kebetulan, melainkan dipengaruhi oleh beberapa faktor yang terkait dengan perubahan pola hidup dan interaksi sosial. Secara keseluruhan, hasil analisis ini menunjukkan bahwa kejahatan di Los Angeles cenderung terkonsentrasi di area padat penduduk dan paling sering terjadi pada malam hari. Informasi ini dapat digunakan sebagai dasar dalam penentuan prioritas patroli polisi dan alokasi sumber daya keamanan publik.

2. Jenis Kejahatan Dominan

Berdasarkan *crosstab* antara *Crm Cd Desc* dan *AREA NAME*, ditemukan bahwa jenis kejahatan yang paling sering terjadi adalah pencurian kendaraan dan perampokan, menunjukkan perlunya pengawasan ketat di area parkir dan jalan raya. Temuan ini juga didukung dengan sistem keamanan atau rasio polisi per daerah yang tidak seimbang dengan tingkat kejahatan. Sedangkan, kejahatan lain seperti kasus kekerasan fisik menunjukkan adanya tekanan sosial dan interaksi antarindividu yang kompleks di wilayah urban.

3. Profil Korban dan Demografi

Analisis terhadap variabel *Vict Age* menunjukkan bahwa mayoritas korban berada pada rentang usia 25–30 tahun, yang merupakan kelompok usia produktif. Distribusi usia yang lebar ($IQR = 44$, $SD \approx 22$) menandakan bahwa korban kejahatan berasal dari berbagai kalangan, termasuk anak muda dan lansia. Temuan ini menunjukkan bahwa kelompok usia produktif lebih rentan menjadi korban karena tingkat mobilitas dan aktivitas luar ruang yang lebih tinggi. Analisis ini juga menegaskan pentingnya pendekatan pencegahan yang mempertimbangkan faktor demografis dan gaya hidup masyarakat.

11. Kesimpulan

Hasil analisis eksploratif terhadap data kriminalitas di Kota Los Angeles menunjukkan bahwa pola kejahatan dipengaruhi oleh faktor waktu, lokasi, dan jenis tindak kriminal. Temuan menunjukkan kejahatan memiliki pola spasial dan temporal yang konsisten. Hotspot kriminalitas berada di area Pacific dan Southeast, sementara kejahatan paling

sering terjadi pada malam hari pukul 18.00–23.00 dan meningkat pada hari Jumat hingga Minggu. Pola ini menunjukkan korelasi antara aktivitas sosial masyarakat dan risiko kriminalitas, sehingga informasi ini dapat dijadikan dasar untuk perencanaan patroli polisi dan alokasi sumber daya keamanan publik secara lebih efisien.

Jenis kejahatan yang paling dominan adalah pencurian kendaraan, perampokan diikuti kekerasan fisik, dengan mayoritas korban berada pada usia produktif 25–30 tahun. Secara keseluruhan, temuan ini menekankan pentingnya penerapan strategi keamanan yang memperhatikan waktu dan lokasi misalnya, meningkatkan patroli di jam-jam rawan dan di daerah dengan tingkat kejahatan tinggi serta perlunya edukasi masyarakat agar lebih waspada terhadap potensi ancaman kriminal di lingkungan perkotaan. Pendekatan ini diharapkan dapat mengurangi risiko kriminalitas sekaligus meningkatkan keamanan publik secara efektif.

12. Lampiran

POTONGAN KODE PENTING

Penanganan Nilai Hilang

```
df.isna().sum().sort_values(ascending=False)
Output :
Jumlah nilai hilang per kolom sebelum imputasi:
```

Mocodes	151619	Vict Sex	144644	Vict Descent	144656
Premis Cd	16	Premis Desc	588	Weapon Used Cd	677744
Weapon Desc	677744	Status	1	Crm Cd 1	11
Crm Cd 2	935831	Crm Cd 3	1002677	Crm Cd 4	1004927
Cross Street	850755	DISTRICT	6535		

```
# Hapus baris dengan DISTRICT kosong
df = df.dropna(subset=['DISTRICT'])
# Hapus kolom dengan proporsi missing tinggi
df = df.drop(columns=['CROSS STREET'])
# Isi nilai kosong pada kolom Weapon Desc
df['Weapon Desc'].fillna('UNKNOWN WEAPON', inplace=True)
# Hapus baris dengan status kosong
df = df.dropna(subset=['STATUS'])
Output :
Jumlah baris sebelum: 998456 Jumlah baris sesudah: 998456
Baris dihapus (karena DISTRICT NaN): 0
Weapon Desc
UNKNOWN WEAPON/OTHER WEAPON      709527
STRONG-ARM (HANDS, FIST, FEET OR BODILY FORCE)  173444
VERBAL THREAT                      23722
HAND GUN                          20096
SEMI-AUTOMATIC PISTOL             7217
```

CONTOH SEBELUM/SESUDAH TRANSFORMASI/MERGE

Tabel 16. Dataset 1 Crime

Nama Kolom	Tipe Data	Skala	Contoh Nilai
Dr No	Integer	Nominal	201904032
Date Rptd	Objek	-	2020 Jan 02 12:00:00 AM
Date OCC	Objek	-	2020 Jan 01 12:00:00 AM
Time OCC	Integer	-	2135
Area	Integer	Nominal	19
Area Name	Objek	-	Mission
Rpt Dist No	Integer	Nominal	1924
Part 1-2	Integer	Ordinal	1
Crm Cd	Integer	Nominal	761
Crm Cd Desc	Objek	-	BRANDISH WEAPON
Mocodes	Objek	Nominal	1822 3006 3028 3039 3032 3037 3101 4019 3602
Vict Age	Integer	Rasio	41
...

Tabel 17. Dataset 2 Kepolisian

Nama Kolom	Tipe Data	Skala	Contoh Nilai
Name	Objek	-	ABAD, MIGUEL A
SerialNo	Integer	Nominal	42963
Area	Objek	-	NEWT
...

Tabel 18. Dataset 3 Sensus Penduduk

Nama Kolom	Tipe Data	Skala	Contoh Nilai
Council District	Objek	-	1 - Gilbert Cedillo
Pop2010	Float	Rasio	245216.22
White pop	Float	Rasio	90857
Black pop	Float	Rasio	7759.13
...

Tabel 19. Hasil Merge Dataset Crime dan Sensus Penduduk

District	Total Crimes	Council District	Population	White_pop	Black_pop	..
11.0	2595	11 - Mike Bonin	250,726.93	174,020.70	13,569.60	..
2.0	2354	2 - Paul Krekorian	256,691.84	158,999.91	11,463.05	...
7.0	2249	7 - Monica Rodriguez	256,172.52	134,061.58	10,153.12	...

12.0	2165	12 - Mitchell Englander	258,715.95	158,939.64	11,239.45	...
3.0	1930	3 - Bob Blumenfield	257,183.29	155,558.89	11,669.04	...
...

Tabel 20. Hasil Merge Dataset Crime dan Kepolisian

area_name	year_occ	total_crimes	date	police_count	crimes_per_police
77th Street	2020.0	168	2022-01-25	350	0.480000
77th Street	2020.0	168	2022-04-06	311	0.540193
77th Street	2020.0	168	2022-04-25	340	0.494118
77th Street	2020.0	168	2022-05-03	340	0.494118
77th Street	2020.0	168	2022-08-22	339	0.495575
...

DAFTAR GAMBAR

Gambar 1. Diagram Alur Proses Data.....	6
Gambar 2. Distribusi Normalitas.....	9
Gambar 3. Q-Q Plot Distribusi Normalitas.....	10
Gambar 4. Distribusi Uniform.....	10
Gambar 5. Distribusi Eksponensial.....	11
Gambar 5. Distribusi Binomial.....	11
Gambar 7. Diagram Garis.....	17
Gambar 8. Diagram Batang.....	17
Gambar 9. Scatter Plot.....	18
Gambar 10. Bubble Plot.....	18
Gambar 11. Stacked Plot.....	19
Gambar 12. Pie Chart.....	20
Gambar 13. Diagram Polar.....	21
Gambar 14. Histogram.....	21
Gambar 15. Diagram Lollipop.....	22
Gambar 16.Heatmap sebagai Diagram Terbaik.....	23
Gambar 17. Tampilan Dashboard Mini.....	23

DAFTAR TABEL

Tabel 1. Penjelasan Data Dictionary untuk Dataset Crime.....	4
Tabel 2. Penjelasan Data Dictionary untuk Dataset Kepolisian.....	5
Tabel 3. Penjelasan Data Dictionary untuk Dataset Penduduk.....	5
Tabel 4. Penjelasan Data Dictionary untuk Dataset Wilayah Distrik.....	5
Tabel 5 .Contoh hasil penggabungan.....	7
Tabel 6. Ukuran pemusatan data.....	8
Tabel 8. Agregasi data kriminal antar distrik.....	13
Tabel 9. Data Sensus Penduduk.....	13
Tabel 10. Tingkat Kejahatan per 100.000 Penduduk.....	14
Tabel 11. Perbandingan Kasus Kejahatan Dengan Jumlah Polisi Di Setiap Area.....	14
Tabel 12. Transformasi per wilayah.....	15
Tabel 13. Pivot Jam x Hari.....	15
Tabel 14. Distribusi Jenis Kejahatan Berdasarkan Wilayah.....	16
Tabel 15. Diagram Tabel.....	20
Tabel 16.Dataset 1 Crime.....	28
Tabel 17. Dataset 2 Kepolisian.....	28
Tabel 18. Dataset 3 Sensus Penduduk.....	28
Tabel 19. Hasil Merge Dataset Crime dan Sensus Penduduk.....	28
Tabel 20. Hasil Merge Dataset Crime dan Kepolisian.....	29