

LAPORAN DATA MINING

**ANALISA DATA KRIMINALITAS KOTA LOS ANGELES
TAHUN 2012 HINGGA 2016**



DISUSUN OLEH :

**Aaron Dwi Caesar(26415026)
Randy Nugraha T. (26415051)
Harry Ariyanto Putra (26415133)
Andy Sukanto Kan(26415147)**

**PROGRAM STUDI
TEKNIK INFORMATIKA DAN SISTEM INFORMASI BISNIS
UNIVERSITAS KRISTEN PETRA
SURABAYA**

BAB I

PENDAHULUAN

I. Judul Percobaan

Judul laporan untuk proyek data mining adalah Analisa Terhadap Data Kriminalitas Kota *Los Angeles* Dari Tahun 2013 Hingga 2016

II. Latar Belakang

Keamanan merupakan salah satu prioritas dalam hidup sehari - hari sehingga kemampuan untuk memprediksi munculnya atau terjadinya suatu kejahatan merupakan hal yang sangat dibutuhkan . Tindak kejahatan semakin berkembang setiap harinya , sehingga penanganan tindak kejahatan tersebut juga wajib mengalami peningkatan yang signifikan . Mengingat suatu kejahatan pasti terjadi bukan karena kebetulan semata , tetapi karena kondisi yang memungkinkan , Kemampuan untuk memprediksi kondisi penyebab yang mengarah pada kegiatan kejahatan tersebut dan dapat memprediksi jenis kejahatan yang akan terjadi memungkinkan peningkatan terhadap pencegahan kejahatan tersebut. Tujuan dari penelitian ini adalah untuk menganalisis kumpulan data kejahatan - kejahatan di Los Angeles dengan dataset iklim Los Angeles untuk memperoleh pola dari *dataset* yang digunakan tersebut, yang dapat membantu memprediksi kejahatan yang akan terjadi. Misalnya dengan menghadirkan suatu bagan informasi ke departemen kepolisian , sehingga mereka bisa mengetahui jenis , dimana , dan kapan kejahatan itu akan terjadi dengan menggunakan variabel dan bagan yang diberikan.

III. Rumusan Masalah

Dengan membahas mengenai topik ini , kami berharap bisa menjawab:

- a. Apakah data tindakan criminal memiliki kecenderungan/pola tertentu?
- b. Apa hubungan antara waktu pelaporan dan waktu kejadian criminal?
- c. Apa hubungan antara tindakan kriminal dan waktu kejadian, serta lokasi dan kondisi cuaca kejadian?

IV. Tujuan

Tujuan utama analisa ini adalah untuk menghasilkan bagan/model terbaik untuk analisis kejahatan di Los Angeles .Selain itu , tujuan dari proyek ini adalah untuk mengungkap karakteristik dan relasi yang tersembunyi dalam kumpulan data terutama melalui tipe kejahatan, lokasi dan cuaca dengan menciptakan suatu model yang dapat mengklasifikasikan terjadinya suatu kejahatan dengan kondisi tertentu.

BAB II KAJIAN PUSTAKA

I. Kajian Teori

Gaussian Naïve Bayes Algorithm(Bayesian classification)

Bayesian classification adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. Bayesian classification didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan decision tree dan neural network. Bayesian classification terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar. (Kusrini,2009). Teorema Bayes memiliki bentuk umum sebagai berikut :

$$P(X|H)=P(H|X) P(X) P(H)$$

X = Data dengan class yang belum diketahui

H = Hipotesis data X merupakan suatu class spesifik

$P(H|X)$ = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)

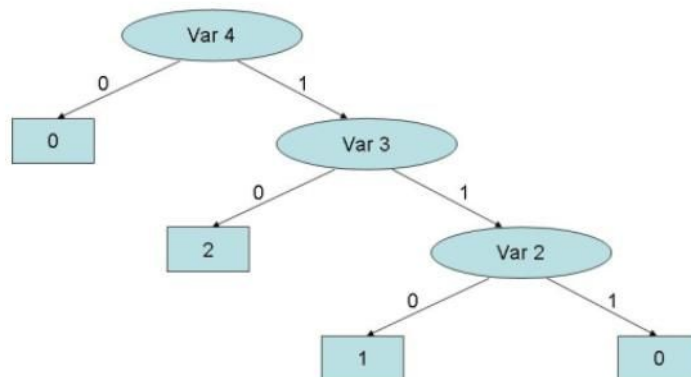
$P(H)$ = Probabilitas hipotesis H (prior prob.)

$P(X|H)$ = Probabilitas X berdasarkan kondisi tersebut

$P(X)$ = Probabilitas dari X

Decision Tree

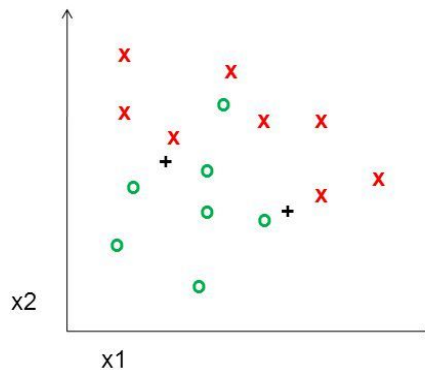
Decision Tree (Pohon Keputusan) adalah pohon dimana setiap cabangnya menunjukkan pilihan diantara sejumlah alternatif pilihan yang ada, dan setiapdaunnya menunjukkan keputusan yang dipilih. Decision tree biasa digunakan untuk mendapatkan informasi untuk tujuan pengambilan sebuah keputusan. Decision tree dimulai dengan sebuah root node(titik awal) yang dipakai oleh user untuk mengambil tindakan. Dari node root ini,user memecahnya sesuai dengan algoritma decision tree. Hasil akhirnya adalah sebuah decision tree dengan setiap cabangnya menunjukkan kemungkinan skenario dari keputusan yang diambil serta hasilnya



K-Nearest Neighbour(KNN)

K-Nearest Neighbor (K-NN) adalah suatu metode yang menggunakan algoritma supervised dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada K-NN. Tujuan dari algoritma ini adalah mengklasifikasi objek baru berdasarkan atribut dan sampel latih. pengklasifikasian tidak menggunakan model apapun untuk dicocokkan dan hanya berdasarkan pada memori. Diberikan titik uji, akan ditemukan sejumlah K objek (titik training) yang paling dekat dengan titik uji. Klasifikasi menggunakan voting terbanyak di antara klasifikasi dari K objek. Algoritma K-NN menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari sample uji yang baru. Dekat atau jauhnya tetangga biasanya dihitung berdasarkan jarak Euclidian.

K-nearest neighbor classifier



II. Input Proses Output

A. Input

Input berupa data laporan kejahatan yang terjadi di *Los Angeles* dari tahun 2013 hingga 2016. Dimana terdapat beberapa *field* sebagai berikut :

- a) Date.Rptd : Date at which the event was reported :DateTime
- b) DR.NO :The Division of records; Case number, etc.: Numeric
- c) DATE.OCC :Date the crime occurred:DateTime
- d) TIME.OCC:Time Occurrence:DateTime
- e) AREA :Numeric Area location in Los Angeles
- f) AREA.NAME:The name of the area in Los Angeles:String
- g) RD:reporting department:Numeric
- h) Crm.Cd :The numeric code for the crime:DateTime
- i) CrmCd.Desc The crime description :String
- j) Status : The crime status :String
- k) Status.Desc :The crime status description :String
- l) LOCATION :Direct street location:String
- m) Cross.Street: Main cross street:Numeric
- n) Location.1:Longitude Latitude:Numeric
- o) Dan beberapa fitur /kolom lain yang akan ditambahkan di bab berikutnya

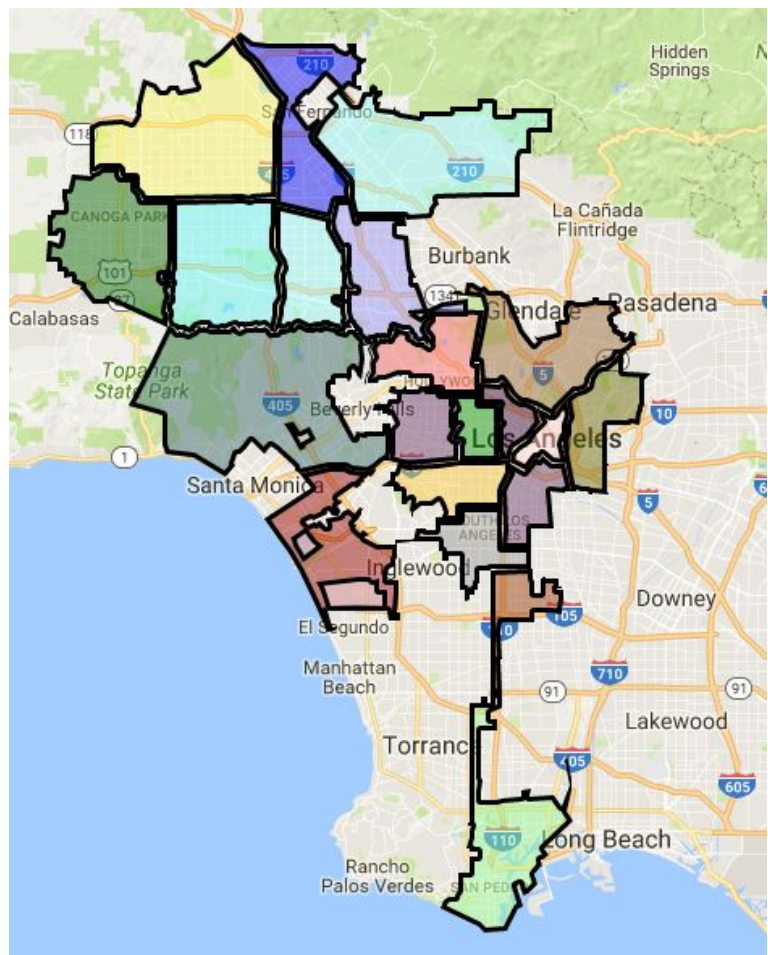
B. Proses

Kami menggunakan data sebagai berikut:

Feature Name	feature Description	Variable
Date.Rptd	Date at which the event was reported	DateTime
DR.NO	The Division of records; Case number, etc.	Numeric
DATE.OCC	Date the crime occurred	DateTime
TIME.OCC	Time Occurrence	DateTime
AREA	Numeric Area location in Los Angeles	Numeric
AREA.NAME	The name of the area in Los Angeles	String
RD	reporting department	Numeric
Crm.Cd	The numeric code for the crime	DateTime
CrmCd.Desc	The crime description	String

Status	The crime status	String
Status.Desc	The crime status description	String
LOCATION	Direct street location	String
Cross.Street	Main cross street	Numeric
Location.1	Longitude Latitude	Numeric
Dan beberapa fitur /kolom lain yang akan ditambahkan di bab berikutnya		

- 77th St. Division
- Central Division
- Devonshire Division
- Foothill Division
- Harbor Division
- Hollenbeck Division
- Hollywood Division
- Mission Division
- Newton Div
- North Hollywood Division
- Universal City, not LAPD
- NE Division
- Olympic Division
- Pacific Division
- Non LAPD Airport
- Non LAPD marina Del Rey
- Rampart Division
- SE Division
- SW Division
- Topanga Division
- Van Nuys Division
- West LA Division
- West LA area
- West Valley Division
- Wilshire Division



C. Output

Prediksi suatu kejahatan akan terjadi pada suatu kondisi tertentu(wilayah ,waktu, hari, bulan,cuaca).

BAB III

HASIL DAN PEMBAHASAN

A. Analisa

Kami melakukan analisa data yang diperoleh dari kaggle.com , yaitu data berupa laporan kasus kejahatan yang terjadi di Los Angeles dari tahun 2013 hingga 2016 , yang akan diproses melalui library(pandas) dan dieksekusi melalui aplikasi jupyter , dengan basis bahasa pemrograman Python dan menggunakan algoritma Naïve bayes , *decision tree* , serta *K-nearest Neighbor* dalam klasifikasinya untuk menemukan algoritma terbaik untuk data set yang kami miliki.

B. Deskripsi data

Kami mengkategorikan data yang ada menjadi beberapa jenis seperti berikut :

Jenis-jenis kejadian kriminal:

Arson : Melakukan kerusakan secara sengaja, atau mencoba untuk merusak, sebuah properti ataupun properti personal dengan api atau alat pembakar.

Assault : Tindakan serangan dari satu orang ke orang lain yang melawan hukum.

Embezzlement: Penyalahgunaan kekuasaan untuk kepentingan pribadi

Fraud / Forgery : melakukan pengubahan, imitasi, membuat sebuah kopi tanpa izin, dengan tujuan untuk melakukan penipuan. atau pembelian, penjualan atau kepemilikan benda tersebut dengan tujuan untuk menipu.

Disorderly Conduct: Mengganggu ketenangan umum, merusak moralitas masyarakat

Drug/Narcotic: Tindakan memiliki , menggunakan , maupun menjual zat-zat kimia yang berbahaya

Extortion/Blackmail: Memperoleh aset tunai , properti , atau benda berharga lainnya , melalui ancaman berupa kekerasan , penyalahgunaan kekuasaan , perusakan nama baik , atau melalui cara lain yang serupa.

Family Offense: Tindakan mengancam atau kekerasan terhadap anggota keluarga baik secara fisik maupun verbal

Homicide : Tindakan penghilangan nyawa suatu individu oleh seseorang atau suatu kelompok yang bersifat disengaja

Kidnapping: Tindakan penculikan

Property:secara tidak sah memasuki tanah, tempat tinggal, atau properti nyata lainnya,dan dengan sengaja merusak

Prostitution: Kegiatan seksual untuk profit/uang

Sex Offense :Setiap tindakan seksual ditujukan terhadap orang lain, secara paksa dan / atau melawan kehendak orang tersebut.

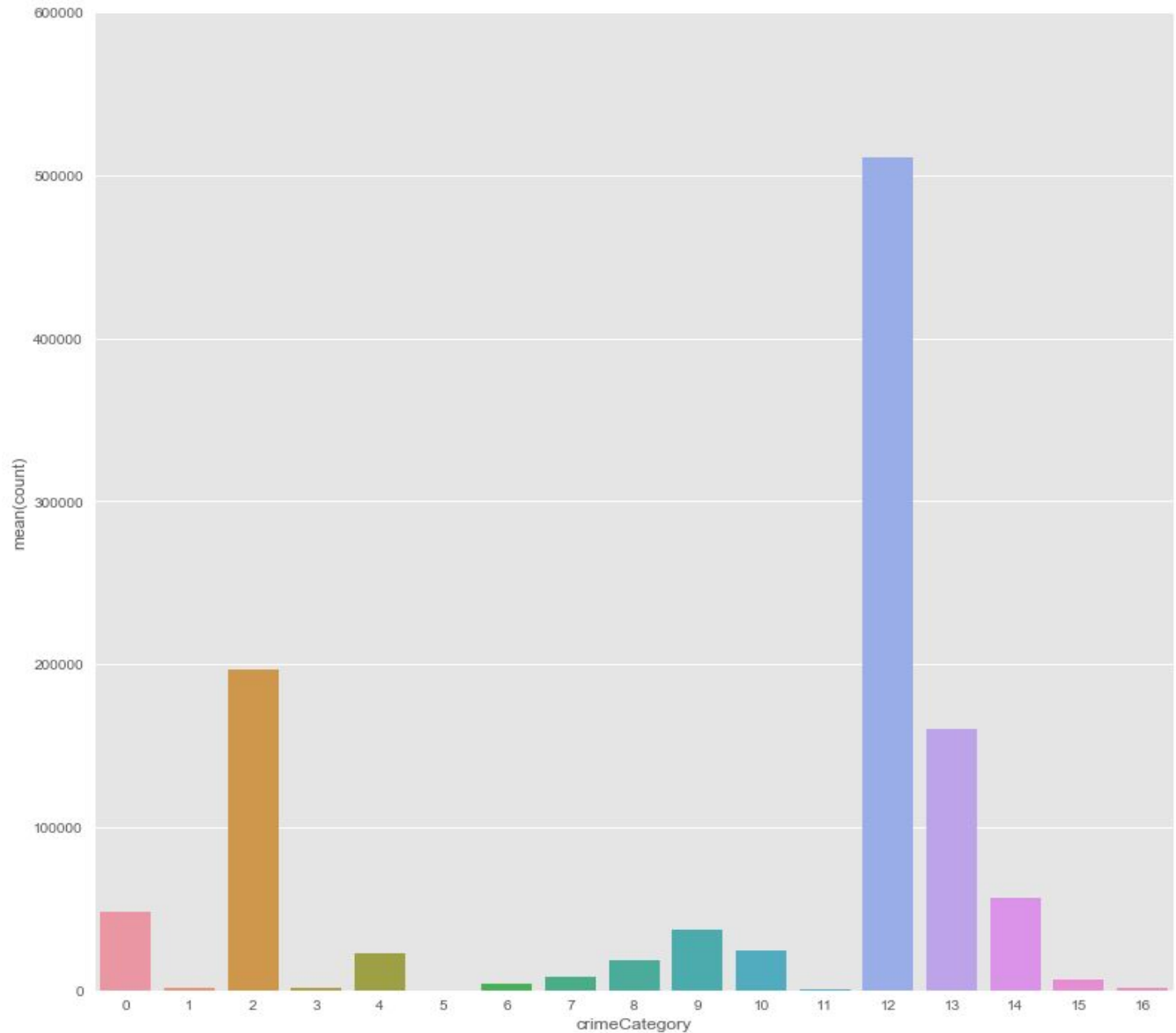
Theft :Tindakan pengambilan suatu benda ,properti ,atau hal lain serupa dan memiliki nilai.

Traffic: Kasus yang terjadi berhubungan dengan lalu lintas , baik itu tabrak lari , kecepatan melewati batas , dan lainnya.

Weapon Law violation:Pelanggaran hukum atau peraturan yang melarang pembuatan, penjualan, pembelian, pengangkutan, kepemilikan, menyembunyian, atau penggunaan senjata api, instrumen pemotongan, bahan peledak, alat pembakar, atau senjata mematikan lainnya

C. Pembahasan

Data dicleaning dengan membuang data bernilai null untuk setiap atribut, kecuali Location.1 karena terlalu banyak yang null, serta untuk tiap pelaporan yang terlebih dahulu daripada kejadiannya . Berikut merupakan grafik dari relasi yang kami dapatkan:



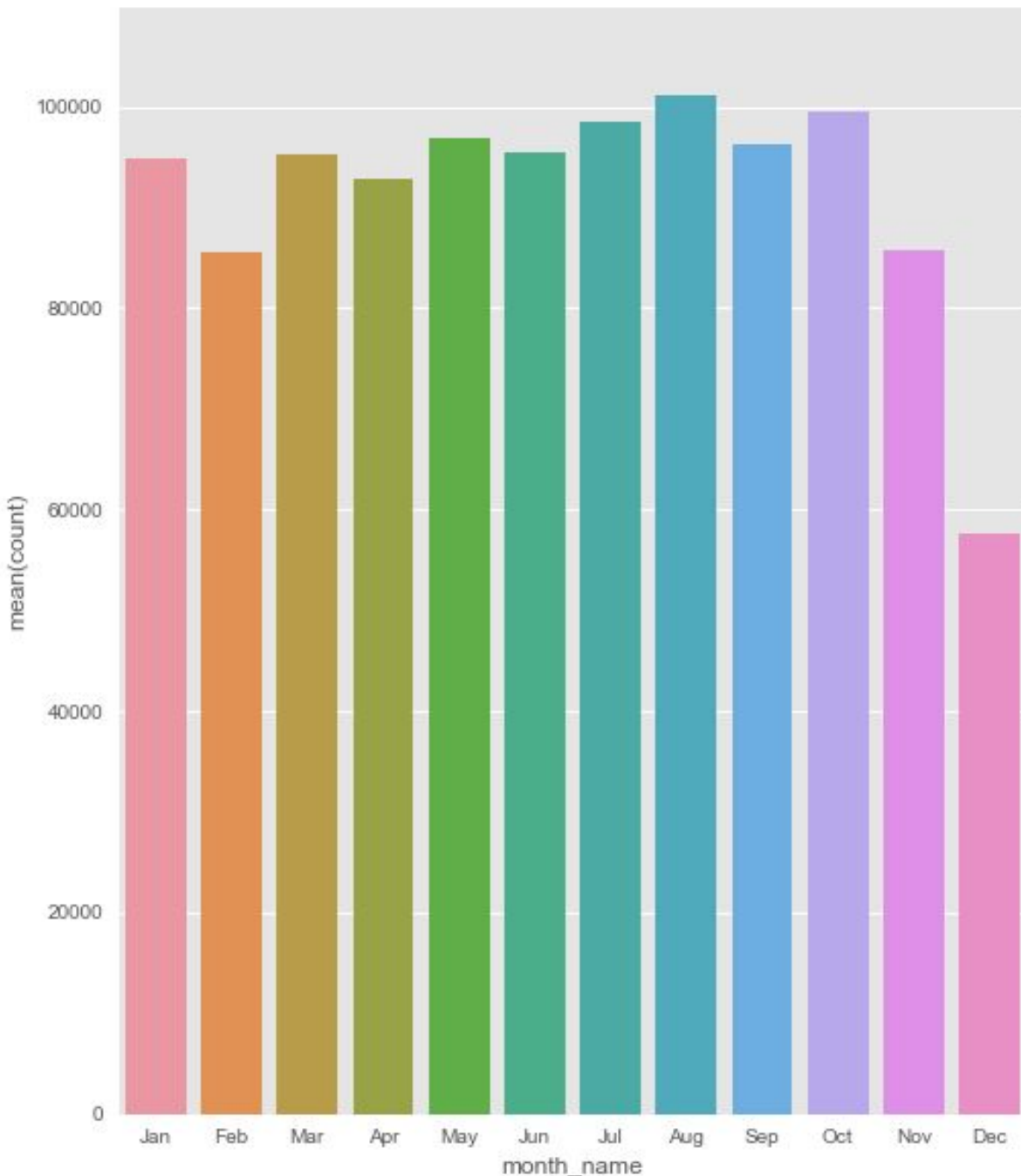
Keterangan:

- | | |
|-----------------------|------------------------|
| 0. Misc | 6. Embezzlement |
| 1. Arson | 7. Family Offense |
| 2. Assault | 8. Fraud/Forgery |
| 3. Homicide | 9. Blackmail/Extortion |
| 4. Disorderly Conduct | 10. Sex Offense |
| 5. Drug/Narcotic | 11. Prostitution |

12. Theft.
13. Traffic

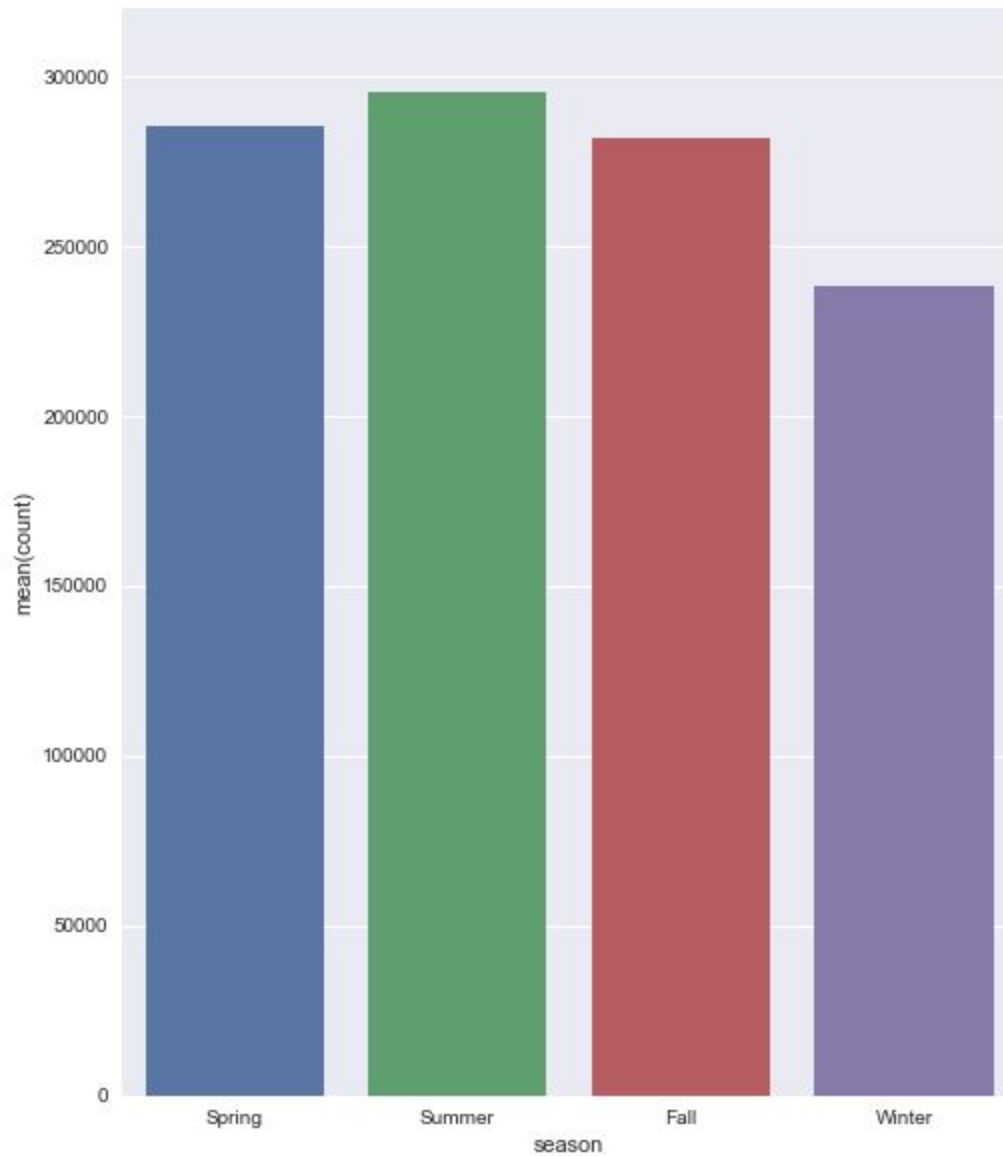
14. Property
15. Weapon Law violation
16. Kidnapping

Grafik diatas merupakan jumlah kejahatan selama 5 tahun terakhir yang disusun berdasarkan tipenya. Dapat dilihat kasus *Theft* merupakan kasus dengan jumlah terbanyak , yang lalu diikuti dengan *assault* dan *traffic*?

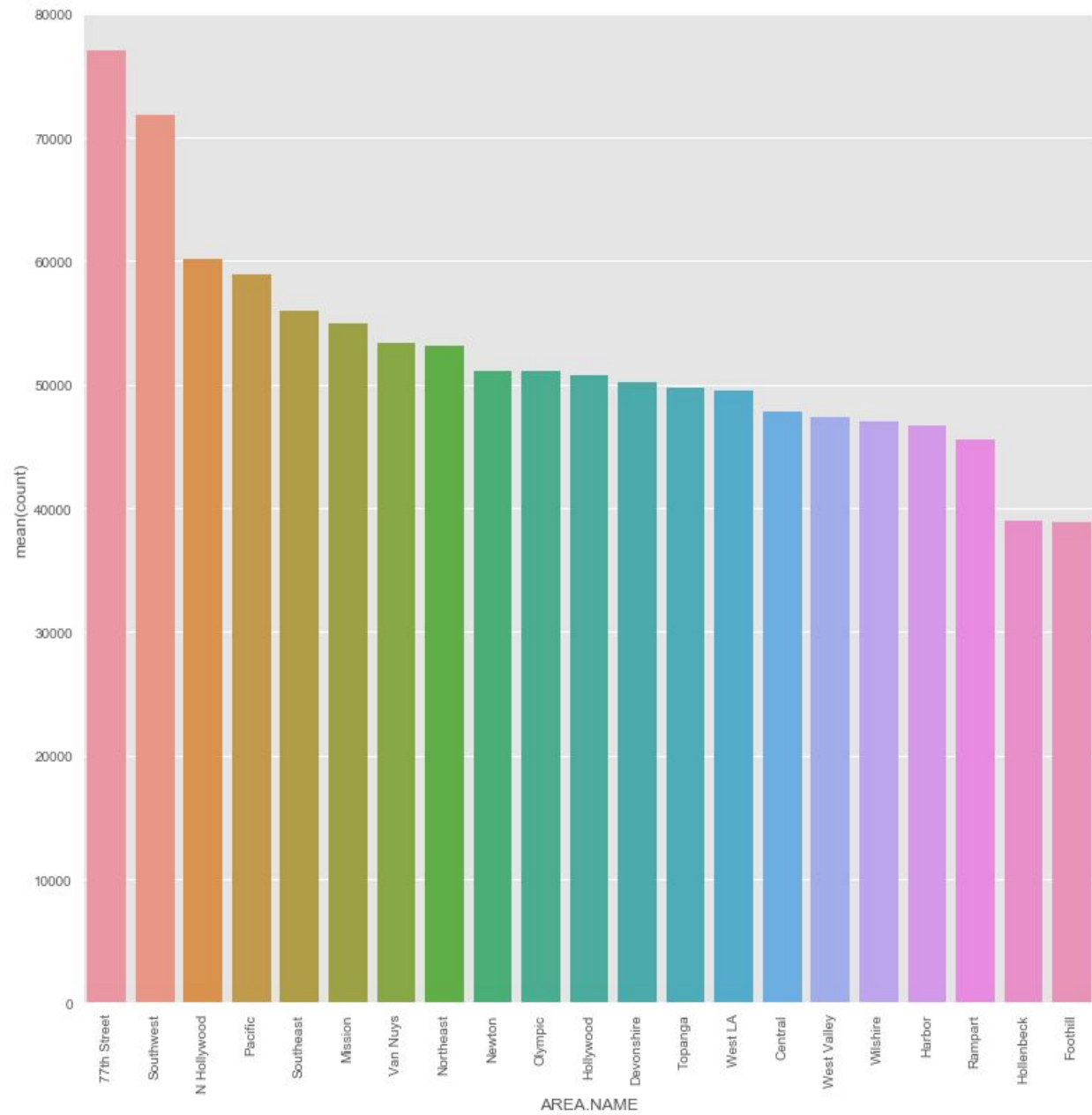


Grafik diatas menunjukkan jumlah kejahatan yang terjadi secara keseluruhan dari awal bulan januari hingga ke akhir bulan Desember .Dapat dilihat pada Desember , jumlah

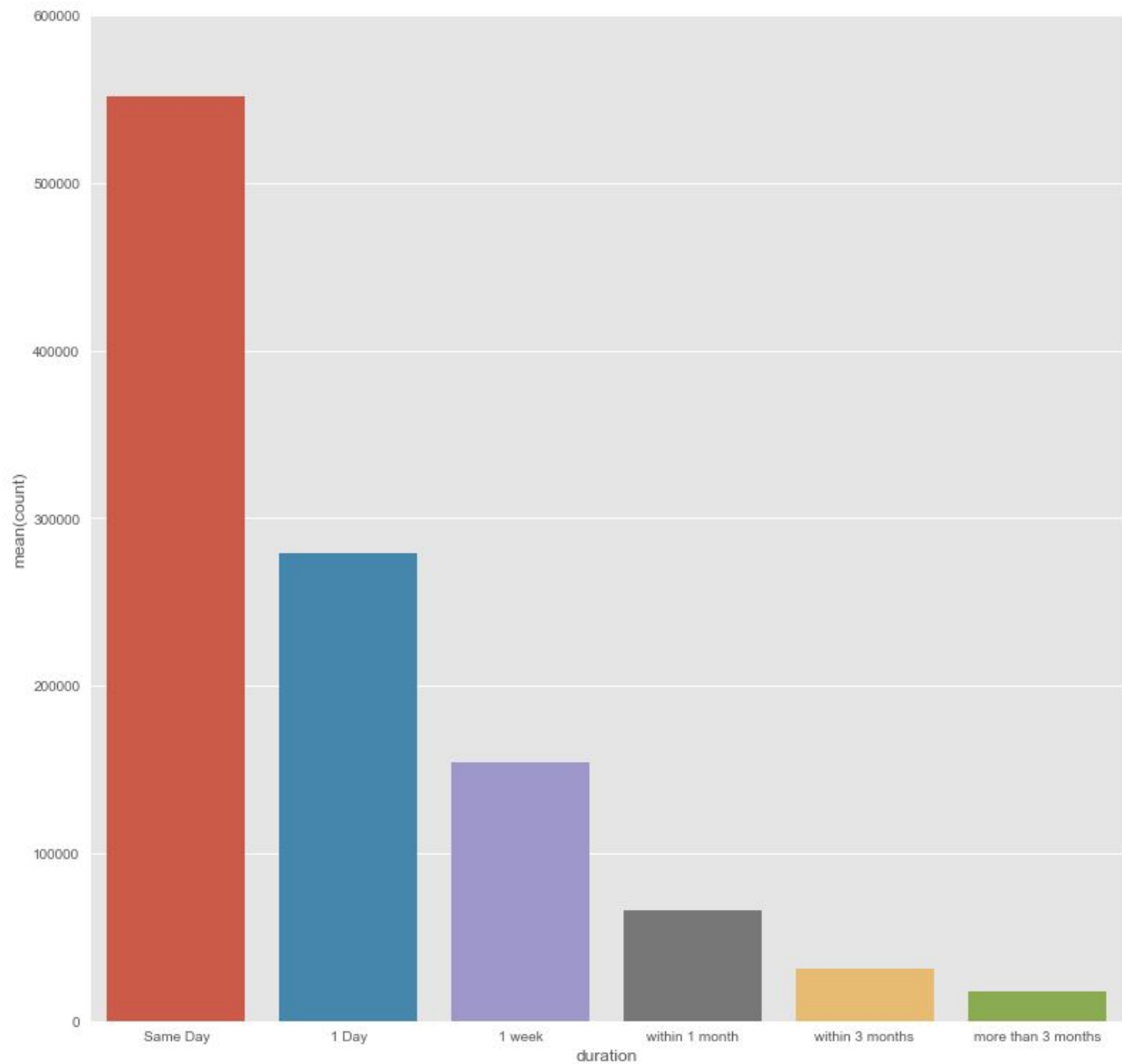
kejahatan jauh lebih rendah dari bulan-bulan sebelumnya



Grafik diatas menunjukkan jumlah kejahatan yang terjadi disusun berdasarkan musimnya. Dapat dilihat bahwa pada musim dingin (*winter*) , jumlah kejahatan lebih rendah dibanding musim-musim lainnya.



Grafik diatas menunjukkan jumlah kejahatan yang terjadi dari tahun 2012 hingga 2016 diurutkan berdasarkan areanya. Dari grafik diatas , dapat dilihat bahwa pada *77th street* , kejahatan terjadi lebih banyak dibanding area lainnya.



Grafik diatas menunjukkan jumlah kejahatan yang dilaporkan dalam jangka waktu tertentu. Dari grafik diatas , dapat dilihat kejahatan yang dilaporkan pada hari yang sama jauh lebih banyak dari jangka waktu lainnya . Berikut merupakan detail dari kejahatan - kejahatan yang dilaporkan pada durasi tertentu:

- Same Day

	CrnCd.Desc	count
0	TRAFFIC DR #	112361
1	BATTERY - SIMPLE ASSAULT	60671
2	SPOUSAL(COHAB) ABUSE - SIMPLE ASSAULT	37985
3	ASSAULT WITH DEADLY WEAPON, AGGRAVATED ASSAULT	30250
4	BURGLARY	28408
5	ROBBERY	26591

- 1 day

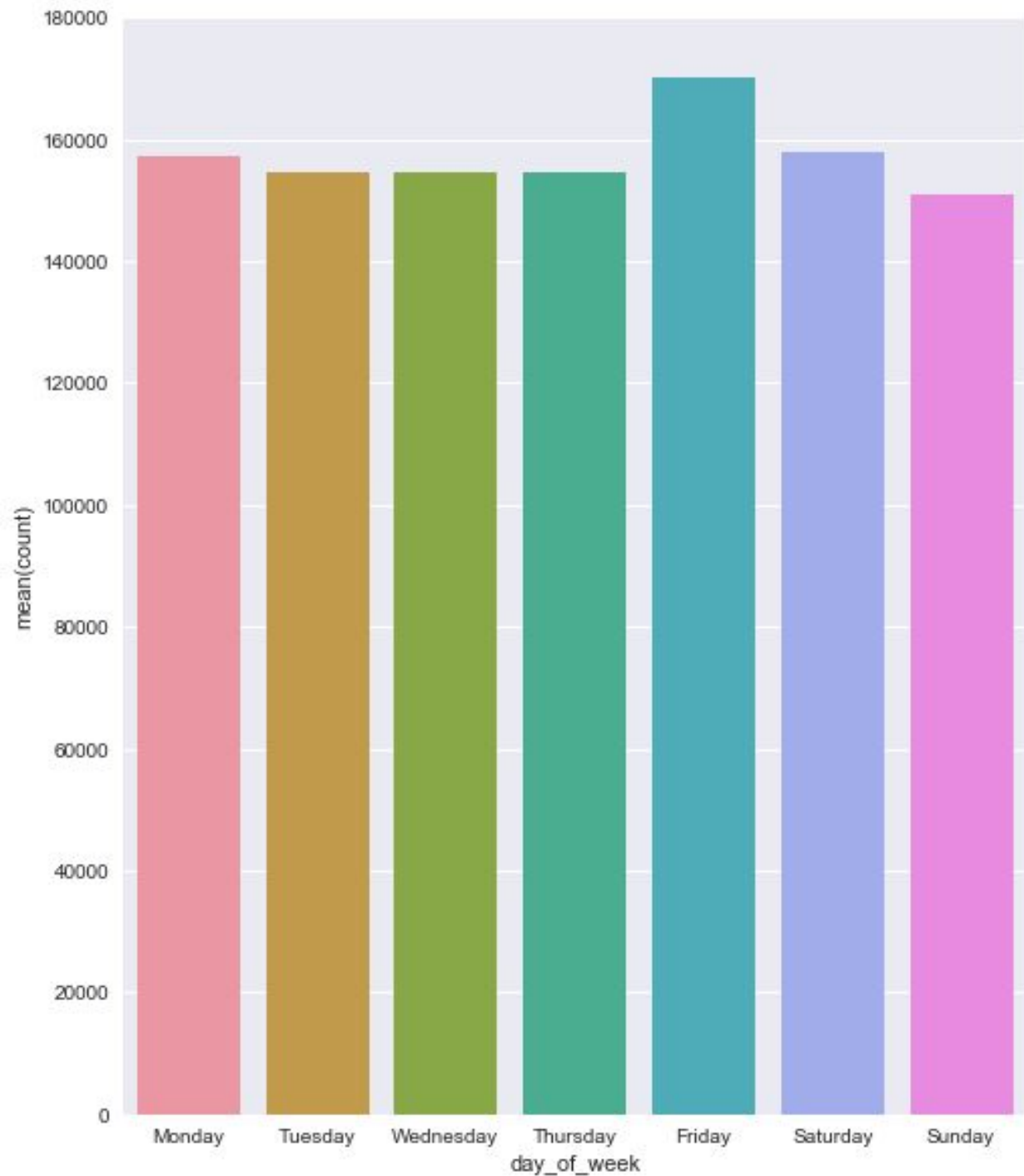
- 1 week

- Within 1 month

	CrmCd.Desc	count
0	THEFT OF IDENTITY	16921
1	THEFT PLAIN - PETTY (\$950 & UNDER)	5250
2	BURGLARY	3885
3	THEFT-GRAND (\$950.01 & OVER)EXCPT,GUNS,FOWL,LI...	3687
4	DOCUMENT FORGERY / STOLEN FELONY	3072
5	TRAFFIC DR #	2611

- Within 3 Months

- More than 3 Months



Grafik diatas menunjukkan jumlah rata-rata kejahatan yang terjadi tiap harinya . Dapat dilihat bahwa jumlah kejahatan pada hari jumat mengalami peningkatan yang cukup signifikan dibandingkan pada hari sebelumnya , dan turun secara perlahan ke hari Sabtu dan Minggu .

C. Hasil Pengamatan

Dari hasil pengamatan kami, kami mendapatkan informasi bahwa :

1. Naive Bayes dan Decision Tree memberikan hasil yang nyaris sama, dan lebih baik daripada metode K-Nearest Neighbor untuk dataset kami
2. Kejahatan yang paling banyak adalah pencurian/Theft dan traffic
3. Cuaca tidak terlalu berpengaruh terhadap hasil data
4. Tindakan Kriminal bertipe cybercrime dilaporkan dalam jangka waktu yang relatif lebih lama dari waktu kejadian dibanding dengan kejahatan yang terjadi secara langsung(fisik)

D. Kelemahan Metode

Metode yang kami gunakan juga memiliki beberapa kelemahan:

1. Pengelompokan tipe kriminal masih manual, bukan melalui text-mining, sehingga mungkin ada human-error
2. Pengambilan data cuaca kurang akurat karena banyak lokasi kosong dalam data, serta tidak mengambil data disaat tindakan kriminal benar-benar terjadi, melainkan data per hari di setiap daerah
3. Jumlah training set yang terlalu besar menyebabkan overfitting, menyebabkan akurasi algoritma menurun
4. Gaussian Naive Bayes didasarkan pada independensi yang tinggi tiap atribut, namun sebenarnya tiap atribut di data kami memiliki hubungan meskipun tidak langsung
5. Beberapa crime, misalnya Identity Theft dan Bribery tidak dipengaruhi tempat dan waktu secara langsung. Meskipun jumlahnya sedikit, namun bisa menjadi noise di data kami

BAB IV PENUTUP

A. Kesimpulan

Setelah melakukan pengujian dan analisis, maka dapat diperoleh kesimpulan sebagai berikut :

- a. Ada pola yang terbentuk dari data-data kejahatan yang digunakan , misalnya pada musim dingin kejahatan lebih rendah.
- b. Ada, setelah hasil analisis dapat dilihat bahwa jenis kriminal yang bertipe *cybercrime* memiliki jarak waktu pelaporan yang relatif lama jika dibandingkan dengan waktu kejadian , dan kejahatan yang dilaporkan pada hari yang sama cenderung merupakan kejahatan yang dapat dilihat secara langsung
- c. Penggunaan metode K-nearest neighbour , Decision tree , dan Naive Bayes Algorithm dapat memprediksi kategori kejahatan yang terjadi dengan akurasi sekitar 46%.

DAFTAR PUSTAKA

- [1]<http://www.tsijournals.com/articles/detecting-crime-types-using-classification-algorithms.pdf>
- [2]<https://cseweb.ucsd.edu/classes/wi17/cse258-a/reports/a086.pdf>
- [3]http://dspace.bracu.ac.bd/xmlui/bitstream/handle/10361/8197/15141009_CSE.pdf?sequence=1&isAllowed=y
- [4]<https://darksy.net/dev>
- [5]<https://www.kaggle.com/kingburrito666/los-angeles-crime>