

# Analisis Data Boston Crime Tahun 2015-2019

Oleh: Gilbert Tjia/ 13417047

Rasa aman merupakan salah satu perasaan terpenting yang perlu untuk dimiliki oleh setiap orang. Berdasarkan penelitian, perasaan aman memiliki korelasi positif terhadap tingkat harapan hidup seseorang (Arora, et al., 2016). Apabila dikaitkan dengan kondisi lingkungan, lingkungan yang memiliki tingkat kriminalitas tinggi tentu tidak akan membuat para warganya merasa aman untuk tinggal di daerah tersebut.

Boston merupakan ibukota dari negara bagian Massachussets, sekaligus merupakan kota dengan tingkat populasi terbanyak ke-21 di Amerika Serikat dengan populasi tidak kurang dari 695.000 jiwa. Sebagai ibukota dengan jumlah populasi yang tidak sedikit, kota Boston menjadi suatu kota yang tidak dapat lepas dari tindak kriminalitas. Dilansir dari situs Alarms.org, didapatkan data bahwa dari sebanyak 150 kota yang terdapat di negara bagian Massachussets, kota Boston berada pada peringkat 130 untuk indeks keamanan kota di Amerika Serikat. Hal inilah yang menjadi dasar dilakukannya penelitian terhadap tingkat kejahatan di kota Boston selama 5 tahun terakhir.

Data kejahatan di kota Boston selama tahun 2015-2019 merupakan suatu *dataset* yang memberikan informasi mengenai setiap insiden yang dilaporkan kepada pihak kepolisian Boston. *Dataset* ini mengandung beberapa variabel berikut:

- |                        |                 |
|------------------------|-----------------|
| 1. Incident_Number     | 10. Month       |
| 2. Offense_Code        | 11. Day_Of_Week |
| 3. Offense_Code_Group  | 12. Hour        |
| 4. Offense_Description | 13. UCR_Part    |
| 5. District            | 14. Street      |
| 6. Reporting_Area      | 15. Lat         |
| 7. Shooting            | 16. Long        |
| 8. Occurred_On_Date    | 17. Location    |
| 9. Year                |                 |

## Data Preprocessing

Sebelum melakukan proses analisis terhadap data kejahatan di kota Boston selama 5 tahun terakhir, data mentah yang didapatkan sebelumnya perlu untuk dibersihkan terlebih dahulu melalui proses *data preprocessing*. Berikut merupakan langkah-langkah *data preprocessing* terhadap *dataset* Boston crime:

1. Memasukkan *dataset* Boston crime pada IDE R dengan query `read.csv()`, lalu melihat *summary* dari data tersebut

Berdasarkan hasil analisis awal terhadap *dataset* Boston crime, didapatkan beberapa *insight* berikut:

- Terdapat banyak kejahatan dengan Incident\_Number yang sama. Pada awalnya, diasumsikan bahwa Incident\_Number merupakan atribut unik yang tidak mungkin memiliki duplikat. Setelah dilakukan penelusuran lebih lanjut, didapat bahwa satu Incident\_Number memungkinkan untuk mengandung beberapa kejahatan yang

berbeda. Salah satu contohnya adalah kejadian dengan Incident\_Number I52071596 yang memiliki 20 duplikat data. Setelah ditelusuri lebih lanjut, ditemukan bahwa pada laporan kejadian tersebut terdapat 5 Offense\_Code\_Group berbeda yang terjadi (umumnya menunjukkan 5 kejadian / kejahatan berbeda yang terjadi). Masing-masing Offense\_Code\_Group juga berduplikasi sebanyak 3 kali. Hal ini kemungkinan menunjukkan bahwa terdapat 4 orang berbeda yang melaksanakan tindak kejahatan tersebut. Pada analisis ini, data yang akan digunakan hanyalah data 5 Offense\_Code\_Group berbeda yang terjadi pada insiden tersebut. Sedangkan 3 data duplikasi lainnya akan dibuang.

- Terdapat banyak *missing data*, hal ini akan terlihat lebih jelas pada tahapan ke-4 dari *data preprocessing*

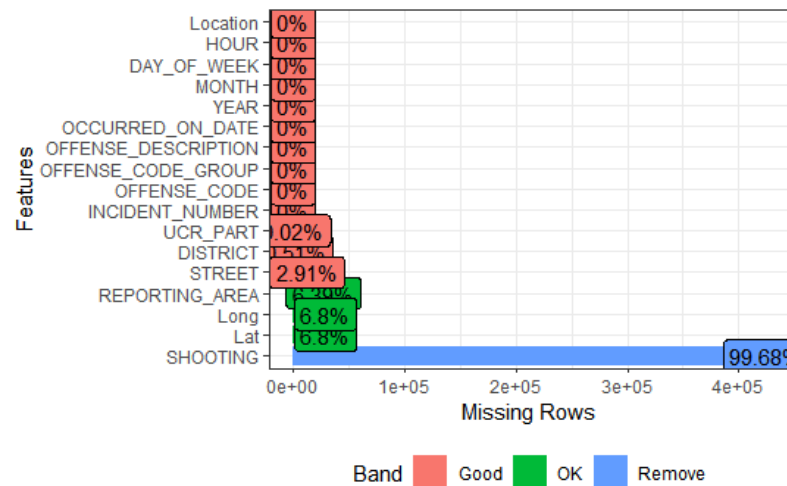
## 2. Menangani *duplicate* data dengan query `unique()`

Query `unique()` digunakan untuk menghapus data yang memiliki *values* setiap kolom yang persis sama. Hal ini dilakukan untuk mengatasi kasus yang sejenis dengan insiden nomor I52071596 yang telah dijelaskan pada bagian sebelumnya.

## 3. Mengubah seluruh blank cells menjadi NA

Pada data mentah Boston *crime*, terdapat sebagian *missing values* yang bernilai NA dan sebagian lagi hanya berupa *blank space*. Seluruh *blank space* akan diubah menjadi NA agar memudahkan tahapan selanjutnya dari *data preprocessing*.

## 4. Melakukan *plotting missing values* sebelum melalui proses *data cleaning*



Berdasarkan hasil *plotting missing values* tersebut, terlihat bahwa *missing values* terbanyak terdapat pada kolom Shooting. Namun, *missing values* pada kolom Shooting sebenarnya bukan merupakan data yang hilang, hanya saja ketentuan dari proses input data menentukan bahwa jika terjadi peristiwa penembakan maka kolom Shooting akan diisi dengan Y, sedangkan bila tidak terjadi maka sel akan dikosongkan.

Selain kolom Shooting, kolom lainnya yang memiliki *missing values* yang cukup signifikan adalah kolom Latitude, Longitude, dan Reporting\_Area. Nilai Latitude dan Longitude akan

diimputasi pada tahap ke-8, sedangkan kolom Reporting\_Area akan dihilangkan. Terakhir, terdapat pula *missing values* pada kolom Street, District, dan UCR\_PART. Ketiga variabel ini memiliki jumlah *missing values* yang cukup kecil (< 3 % total data). Sehingga, baris yang mengandung *missing values* pada salah satu dari ketiga variabel ini akan dihilangkan.

## 5. Melakukan input N pada sel NA pada kolom Shooting

Tahap ini dilakukan agar data kosong pada kolom Shooting tidak dianggap sebagai *missing values*.

## 6. Menghapus kolom Reporting\_Area

Kolom Reporting\_Area akan dihapus karena tidak akan digunakan untuk analisis lebih lanjut.

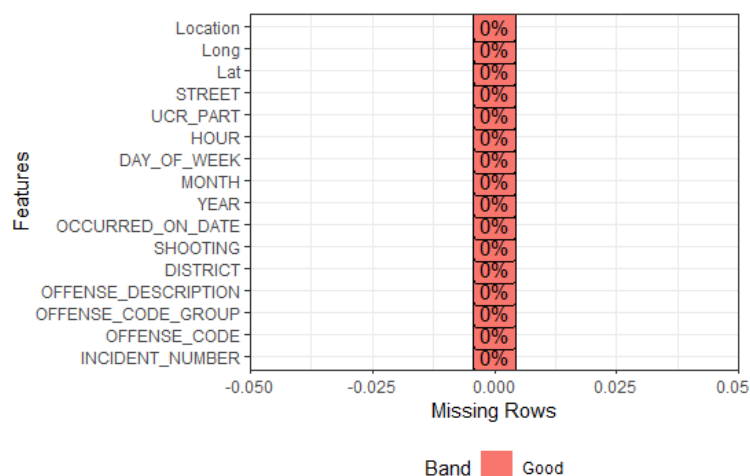
## 7. Menghapus data dengan nilai DISTRICT, UCR\_PART, dan STREET NA

Ketiga variabel tersebut memiliki *missing values* dalam jumlah yang tidak signifikan, sehingga *row data* yang mengandung *missing values* untuk ketiga variabel tersebut dapat dihilangkan.

## 8. Mencari nilai Latitude dan Longitude rata-rata dari setiap District sebagai dasar penggantian *missing values*

Dasar penggantian nilai Latitude dan Longitude pada analisis ini adalah rata-rata dari District di mana Latitude dan Longitude data tersebut bernilai NA. Hal ini dapat diterima karena kejadian yang terjadi pada District yang sama seharusnya memiliki Latitude dan Longitude yang tidak jauh berbeda. Pencarian rata-rata tiap District dilakukan menggunakan query **aggregate()**, setelah itu dimasukkan ke dalam sel *missing values* data yang bersangkutan.

## 9. Melakukan *plotting missing values* setelah melalui proses *data cleaning*

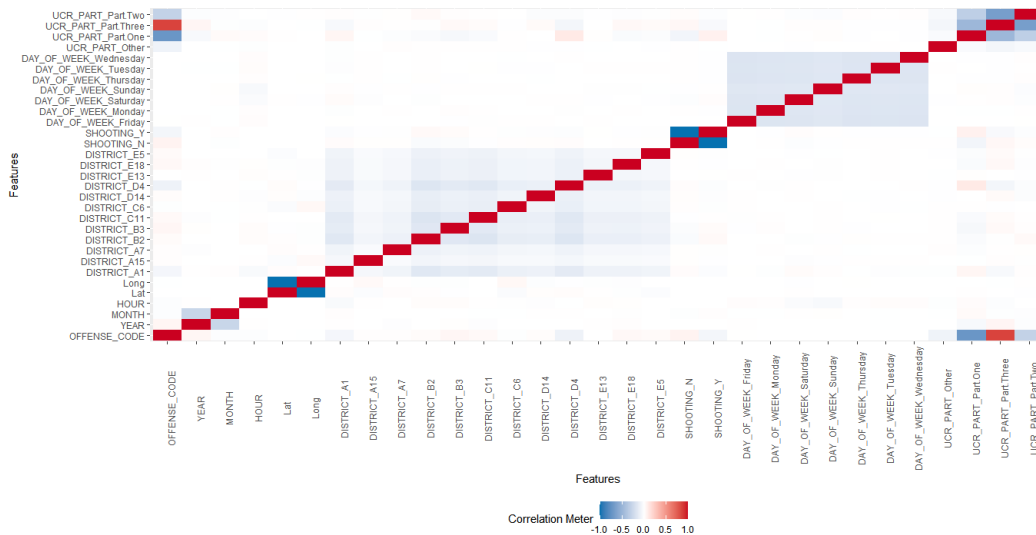


Setelah melalui tahapan *data preprocessing*, terlihat bahwa seluruh kolom dalam *dataset* tidak lagi mengandung *missing values*. Hal ini menunjukkan bahwa dataset tersebut telah siap untuk digunakan dalam proses analisis selanjutnya

## Exploratory Data Analysis

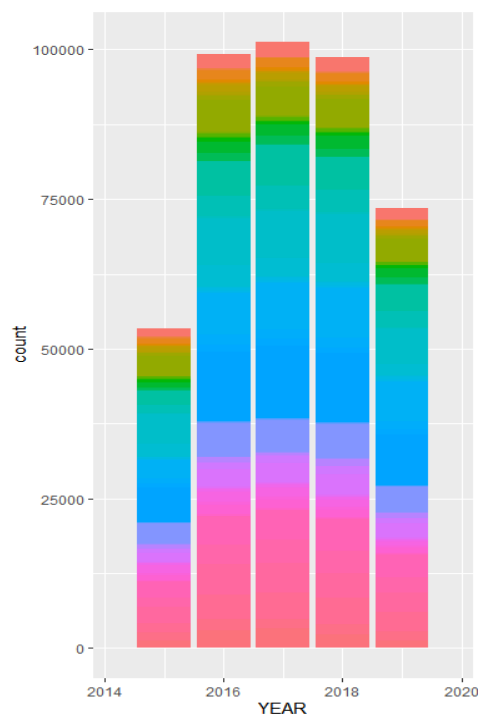
Tahapan selanjutnya adalah tahap pencarian *insight* berdasarkan dataset yang ada. Pada tahap ini akan dilakukan proses visualisasi data untuk membantu menemukan pola yang sebelumnya sulit untuk terlihat dari dataset.

### ➤ Melihat korelasi antar variabel dalam dataset



Hasil *plotting* korelasi antar variabel dalam dataset tidak menunjukkan adanya hubungan khusus ataupun menghasilkan *insight* yang menarik.

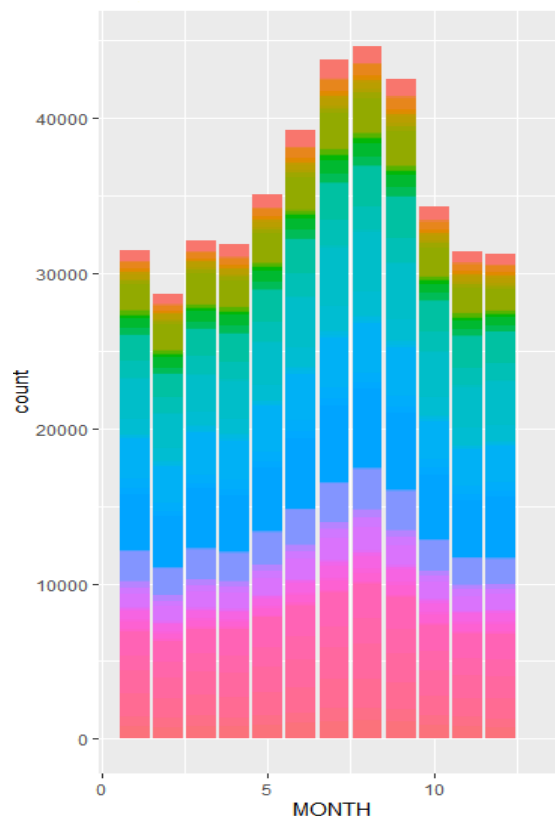
### ➤ Tren tindak kriminalitas selama 5 tahun terakhir





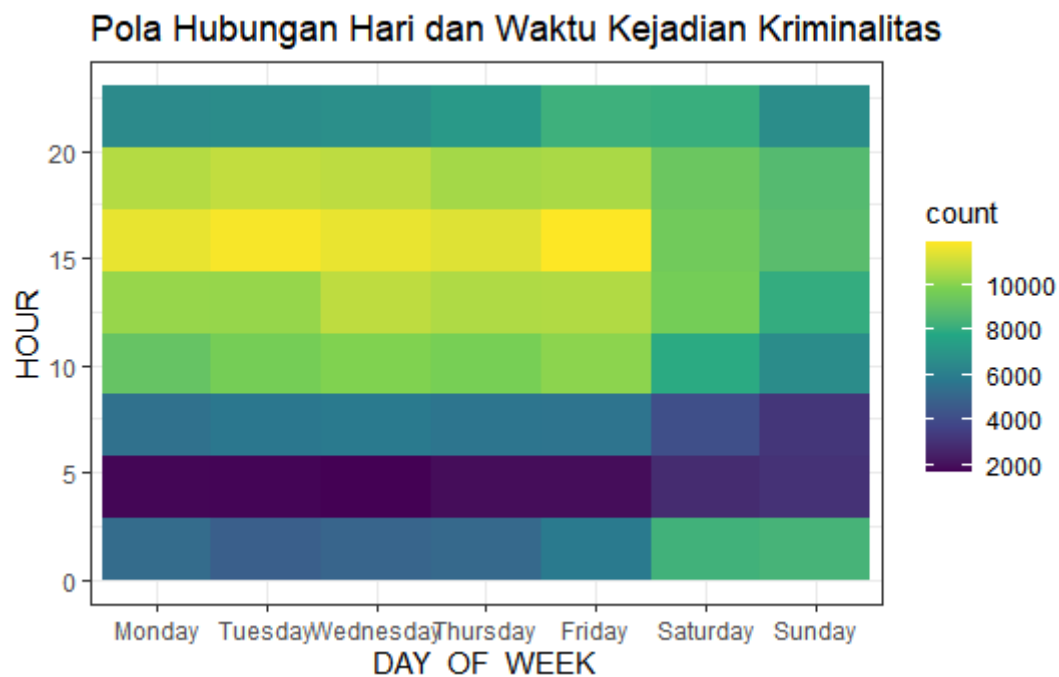
Berdasarkan data tingkat kejahatan selama 5 tahun terakhir, terlihat bahwa tingkat kriminalitas memuncak pada tahun 2017. Namun, pola kenaikan drastis pada tahun 2015 menuju ke tahun 2016 dan penurunan drastis dari tahun 2018 menuju 2019 bukan menunjukkan pola peningkatan / penurunan kejahatan. Alasan kenaikan dan penurunan tersebut adalah karena data yang diambil pada tahun 2015 hanya mencakup data yang dimulai dari bulan Juni hingga Desember 2015, sedangkan untuk data tahun 2019 baru mencakup data dari Januari hingga September 2019. Apabila melihat tahun-tahun dengan data lengkap dari bulan Januari hingga Desember, terlihat bahwa tingkat kriminalitas per tahunnya cukup stabil dan tidak terlihat pola peningkatan / penurunan yang signifikan.

#### ➤ Tren kriminalitas berdasarkan bulan kejadian



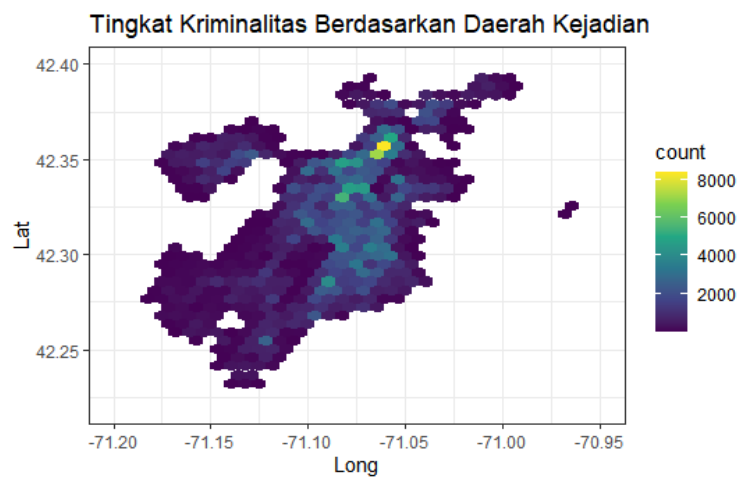
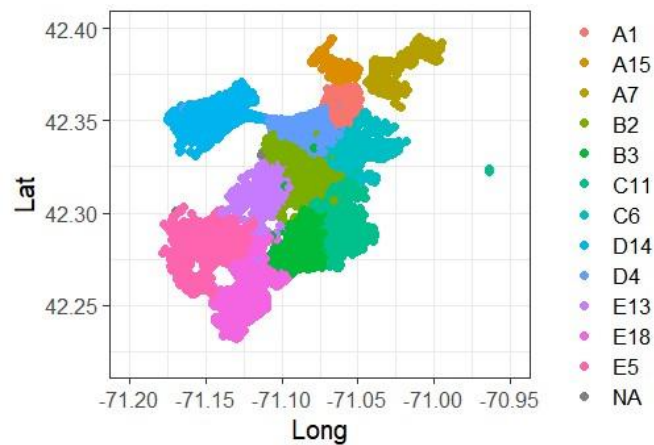
Pada *barplot* tersebut, terlihat suatu pola bahwa tindak kriminalitas mengalami pola kenaikan menjelang pertengahan tahun, terutama pada bulan Juli dan Agustus, kemudian kembali mengalami penurunan menjelang akhir tahun. Penjelasan yang mungkin mengenai tren ini adalah menjelang liburan akhir tahun, tingkat kriminalitas mengalami penurunan karena banyak orang yang sedang menghabiskan waktu bersama keluarganya pada periode-periode tersebut. Selain itu, terlihat pula bahwa tingkat kriminalitas paling rendah terjadi pada bulan Februari. Hal ini kemungkinan besar disebabkan karena jumlah hari pada bulan Februari yang lebih sedikit dibandingkan dengan bulan-bulan lainnya.

➤ **Tren kriminalitas berdasarkan hari dan jam kejadian**

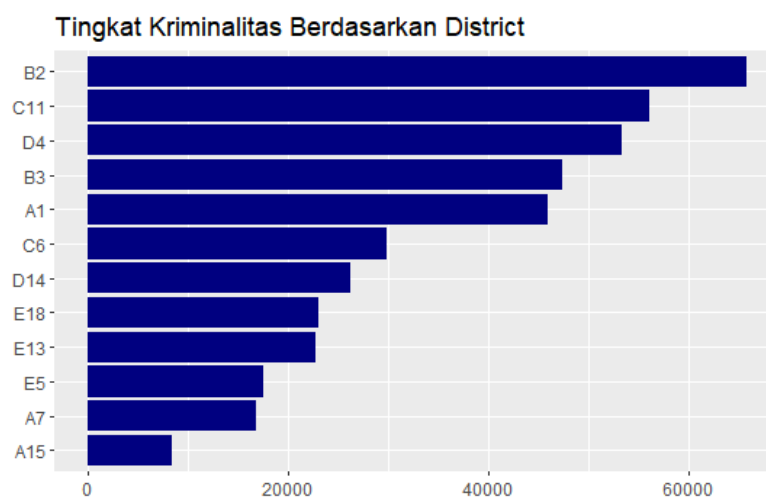


Berdasarkan *heatmap* tersebut, terlihat bahwa tindak kriminal di kota Boston paling rawan terjadi pada hari biasa (Senin-Jumat) sekitar pukul 15.00-18.00. Hal ini mungkin berbeda dengan dugaan kebanyakan orang bahwa kejadian kriminal paling sering terjadi pada malam hingga dini hari. Pada *heatmap* tersebut juga dapat terlihat bahwa terjadi perbedaan pola waktu kriminalitas pada hari Sabtu dan Minggu dibandingkan dengan hari lainnya. Pada kedua hari ini, tindak kriminal paling banyak terjadi mulai pukul 20.00 ke atas. Namun, jumlah tindak kriminal pada hari Sabtu dan Minggu secara umum lebih sedikit jika dibandingkan dengan jumlah tindak kriminal dari hari Senin-Jumat.

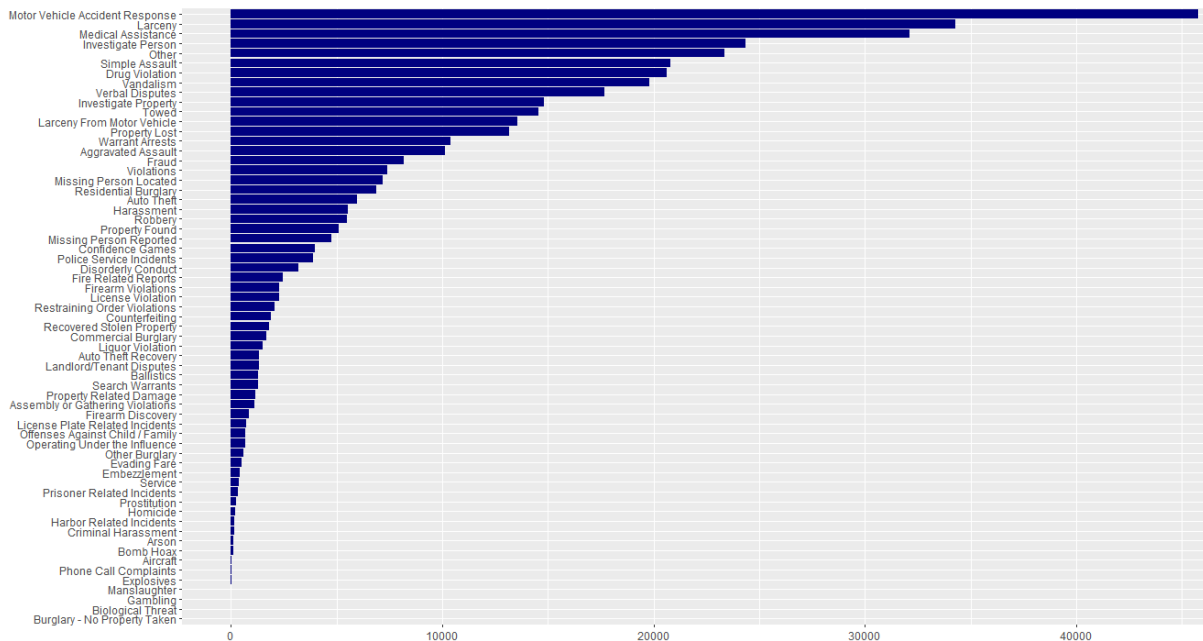
➤ **Daerah yang paling rawan terhadap tindak kriminalitas di kota Boston**



Berdasarkan hasil kedua pemetaan spasial tersebut, terlihat bahwa daerah tengah dari kota Boston merupakan daerah yang lebih rawan terhadap tindak kriminalitas. Hal tersebut utamanya disebabkan karena daerah tengah tersebut merupakan daerah *downtown* dari kota Boston, yaitu daerah terpadat dan tersibuk dari kota Boston. Berikut akan dilakukan *plotting* tingkat kriminalitas tiap distrik untuk memperjelas proses analisa data:



## ➤ Analisis tindakan kriminal yang paling banyak terjadi di kota Boston

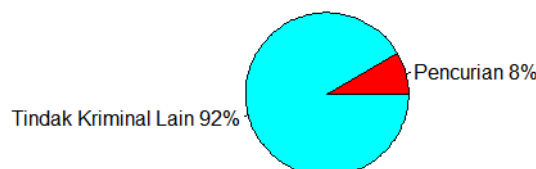


Berdasarkan diagram tersebut, berikut merupakan 10 insiden yang paling sering terjadi di kota Boston selama 5 tahun terakhir:

- |                                    |                          |
|------------------------------------|--------------------------|
| 1. Motor Vehicle Accident Response | 6. Simple Assault        |
| 2. Larceny                         | 7. Drug Violation        |
| 3. Medical Assistance              | 8. Vandalism             |
| 4. Investigate Person              | 9. Verbal Disputes       |
| 5. Other                           | 10. Investigate Property |

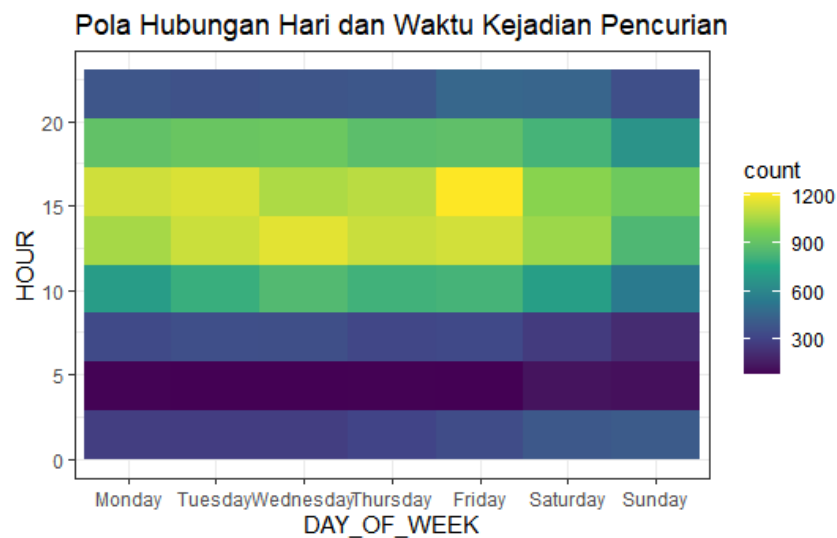
Dari ke-10 insiden tersebut, hanya 4 insiden yang merupakan tindak kriminal. Ke-4 insiden tersebut adalah *larceny* (pencurian), *simple assault* (penyerangan), *drug violation* (penyalahgunaan narkoba), *vandalism* (perusakan terhadap properti). Analisis selanjutnya akan difokuskan pada tindak pencurian yang terjadi di kota Boston karena tindak pencurian merupakan tindak kriminal dengan jumlah terbanyak.

## Proporsi Pencurian Dibanding Tindak Kriminal Lain

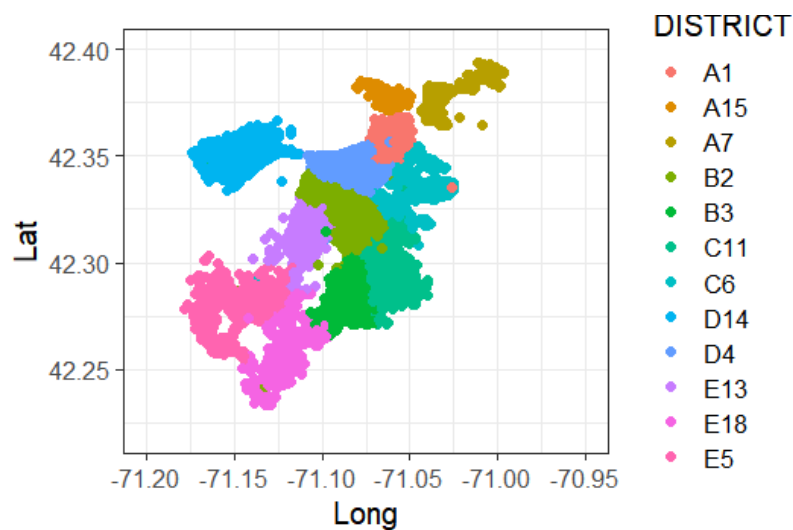


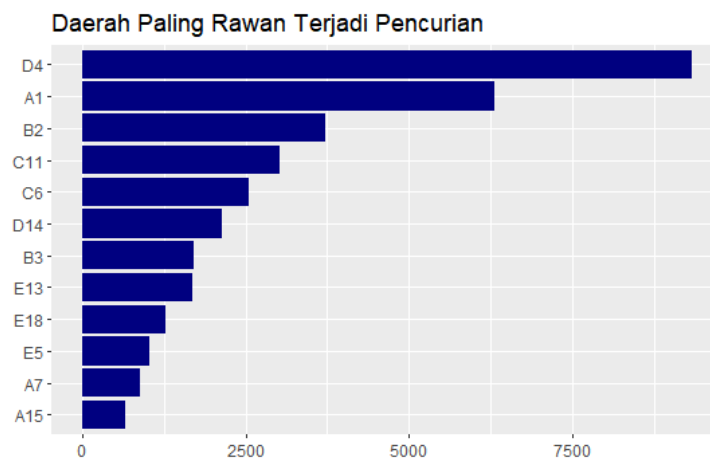
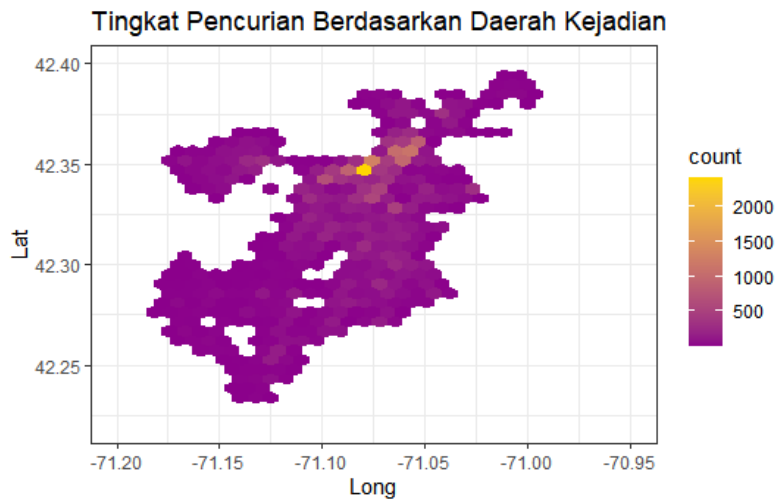


Berdasarkan *pie chart* tersebut, terlihat bahwa sebanyak 8 % dari seluruh tindak kriminalitas di kota Boston selama 5 tahun terakhir merupakan tindak pencurian. Selanjutnya, akan dianalisis lebih jauh mengenai waktu dan hari paling rawan terhadap tindak pencurian.

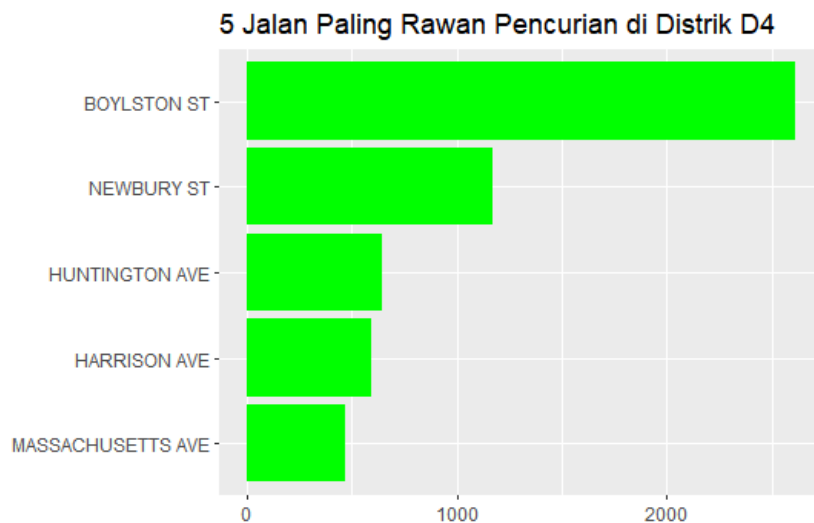


Pola pemetaan hubungan antara waktu dan kejadian pencurian menghasilkan *heatmap* yang mirip dengan pola pada *heatmap* tindak kriminalitas secara umum. Namun, khusus untuk tindak pencurian terlihat bahwa kejadian pencurian meningkat secara cukup drastis pada hari Jumat sekitar pukul 15.00-17.00. Selanjutnya, akan dilakukan analisis terhadap data spasial terkait tindak pencurian di kota Boston:





Berdasarkan analisis spasial terhadap lokasi paling rawan terjadinya tindak pencurian di kota Boston, didapatkan data bahwa tindak pencurian paling banyak terjadi pada distrik D4. Hal ini selanjutnya dibuktikan melalui plot tindak pencurian berdasarkan *district*. Selain itu, terlihat bahwa perbedaan jumlah pencurian di daerah tersebut jauh lebih tinggi dibandingkan dengan tingkat pencurian di daerah lain di kota Boston. Kemudian, dilakukan penelusuran lebih lanjut mengenai beberapa jalan pada distrik D4 yang paling rawan terhadap pencurian.



Berdasarkan hasil analisis, terlihat bahwa jalan Boylon ST merupakan jalan yang paling rawan terhadap pencurian. Dalam menangani hal ini, salah satu langkah yang dapat diambil oleh pemerintah kota Boston adalah untuk memperketat keamanan pada 5 jalan tersebut untuk mengurangi tindak pencurian yang terjadi. Hal ini dapat dilakukan misalnya melalui penambahan CCTV di jalan tersebut, serta menambah personil kepolisian terutama pada hari Jumat pukul 15.00-17.00.

## Kesimpulan dan Saran untuk Policymaker serta Stakeholder Terkait

Berdasarkan hasil *exploratory data analysis* yang telah dilakukan terhadap dataset “Boston Crime 2015-2019”, didapatkan beberapa kesimpulan sebagai berikut:

1. Tindak kejahatan di kota Boston meningkat terutama saat bulan Juli dan Agustus. Hal ini dapat menjadi catatan penting bagi aparat kepolisian kota Boston untuk memperketat penjagaan pada bulan-bulan tersebut. Selain itu, data ini juga dapat bermanfaat bagi para pemilik toko dan pihak-pihak lain yang mungkin menjadi target kejahatan untuk lebih berhati-hati.
2. Tingkat kriminalitas di kota Boston mengalami kenaikan pada hari kerja (Senin-Jumat), terutama pada sore hari (pukul 15.00-18.00) dan penurunan pada akhir pekan (Sabtu dan Minggu). Hampir sama dengan poin sebelumnya, informasi ini juga dapat digunakan oleh pihak kepolisian dan masyarakat untuk lebih waspada pada hari dan waktu tertentu.
3. Distrik B2 merupakan distrik yang paling rawan terhadap tindak kejahatan. Setelah melakukan penelusuran lebih lanjut, distrik B2 merupakan kode untuk daerah Roxbury, salah satu daerah yang memang terkenal rawan terjadi tindak kriminal. Berdasarkan situs AreaVibes.com, daerah Roxbury memiliki *overall crime rate* sebesar 85 % lebih tinggi dibandingkan rata-rata *crime rate* Amerika Serikat.
4. Tindak pencurian merupakan tindak kejahatan yang paling sering terjadi di kota Boston. Kejadian pencurian paling sering terjadi pada distrik B4, terutama di jalan Boylston ST. Data ini dapat digunakan oleh pemerintah daerah kota Boston untuk memperkuat keamanan di daerah tersebut. Hal ini sesuai dengan *pareto principle* yang memfokuskan *resource* untuk menangani 20% dari total permasalahan dengan frekuensi terbanyak.
5. Peningkatan keamanan di kota Boston dapat dilakukan dengan penambahan CCTV pada jalan-jalan yang rawan terhadap tindak kejahatan, serta mengarahkan aparat keamanan pada waktu dan tempat yang memiliki resiko tinggi terjadinya tindak kejahatan.

## Daftar Pustaka

Areavibes. (n.d.). Roxbury, Boston, MA Crime. Retrieved March 8, 2020, from <https://www.areavibes.com/boston-ma/roxbury/crime/>

Arora, A., Spatz, E., Herrin, J., Riley, C., Roy, B., Kell, K., ... Krumholz, H. M. (2016). Population Well-Being Measures Help Explain Geographic Disparities In Life Expectancy At The County Level. *Health Affairs*, 35(11), 2075–2082. doi: 10.1377/hlthaff.2016.0715

Safest Cities in Massachusetts – 2020. (n.d.). Retrieved March 8, 2020, from <https://www.alarms.org/safest-cities-in-massachusetts/>