

# **Analisis Data Terkait Pandemi COVID-19 di Wilayah D.K.I. Jakarta**

Disusun untuk memenuhi prasyarat mengikuti Data Science Academy Compfest 13



**Disusun oleh:**

Tim Sarjana Korona:

Sebastian Cahyo Ardhi Iswara (1103174174)

Kenneth Angelo (1102184030)

Iga Narendra Pramawijaya (1101184256)

**FAKULTAS TEKNIK ELEKTRO**

**UNIVERSITAS TELKOM**

**BANDUNG**

**2021**

## A. Latar Belakang

Pandemi COVID-19 merupakan sebuah pandemi yang berdampak besar bagi kehidupan manusia. Dengan angka penularan yang tinggi berhasil melumpuhkan pergerakan ekonomi, transportasi, bahkan pariwisata. Hal ini tidak hanya dirasakan di Indonesia, namun juga di seluruh belahan dunia. Karena anomali yang terjadi secara mendadak dan berkepanjangan, banyak data menarik yang berhasil dikumpulkan. Dalam laporan ini, tim kami berusaha untuk melihat hubungan antara laju pergerakan penularan COVID-19 dengan variabel-variabel lain menggunakan data, baik itu data internal yang diberikan pada saat kompetisi maupun data eksternal. Diharapkan dengan diketahuinya hubungan antar variabel-variabel yang ada, dapat membantu optimalisasi penanganan pandemi COVID-19.

## B. Jawaban Soal

### a. Jawaban untuk soal dataset

1. Dari dataset yang disediakan, temukan nilai mean, median, dan modus dari positif COVID-19 harian Jakarta.

```
HASIL DARI NILAI YANG DICARI
Mean      : 1477.8455445544555
Median    : 895.0
Modus     : 0
```

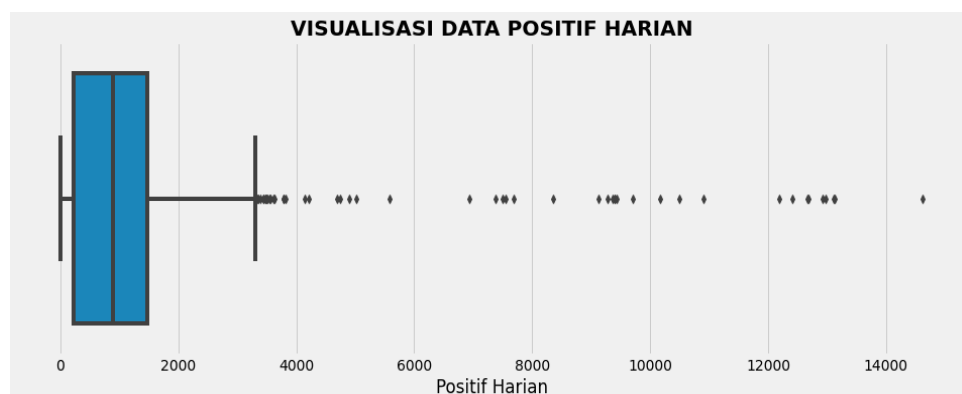
Gambar 1. Mean, Median, Modus positif Covid-19 harian Jakarta

2. Dari dataset yang disediakan, temukan nilai minimal dan maksimal dari positif COVID-19 harian Jakarta.

```
HASIL DARI NILAI YANG DICARI
Max       : 14619
Min       : 0
```

Gambar 2. Nilai Maksimal dan Minimal positif Covid-19 harian Jakarta

3. Dari dataset yang disediakan, temukan nilai-nilai outlier yang ada (menggunakan variabel yang kalian tentukan).



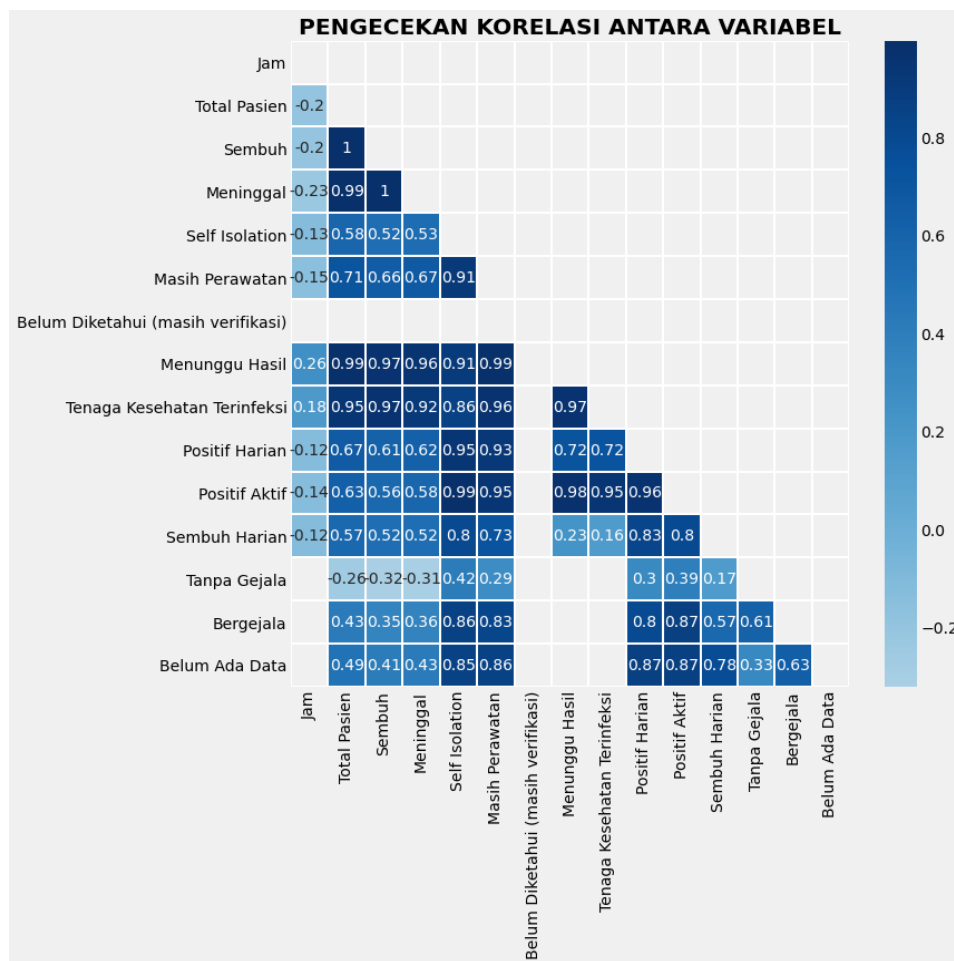
Gambar 3. Visualisasi data positif harian Jakarta

HASIL DARI NILAI YANG DICARI  
 Q1 : 219.0  
 Q3 : 1447.5  
 IQR : 1228.5  
 Qmin : -1623.75  
 Qmax : 2061.75

Gambar 4. Quartil data positif harian jakarta

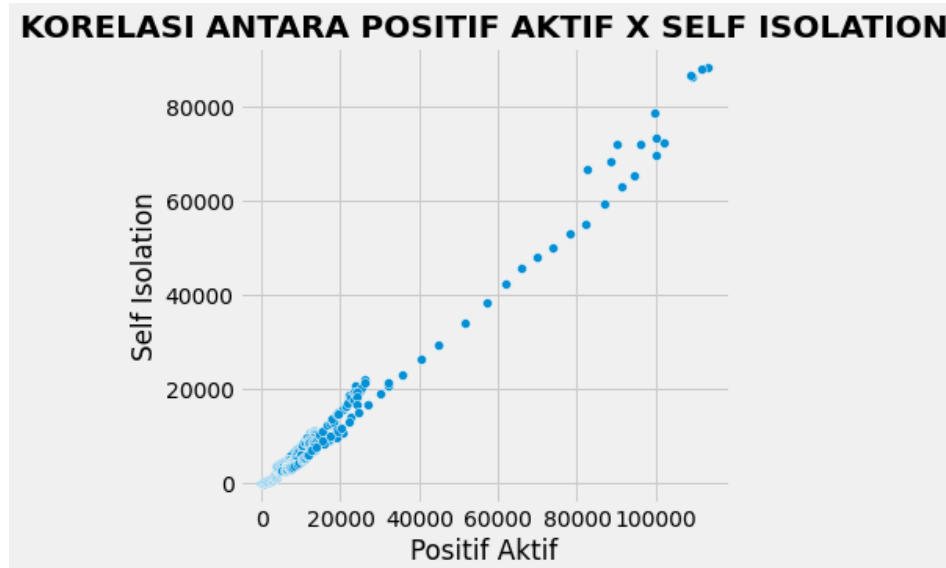
Setelah kami mencari pencilan menggunakan kuartil, ditemukan data positif harian diatas 3200 adalah pencilan (*outlier*)

4. Dari dataset yang disediakan, usulkan dua buah variabel dan berikan analisis korelasi antara kedua variabel tersebut. Jelaskan apa kesimpulan yang dapat diambil berdasarkan analisis kalian.



Gambar 5. Visualisasi Korelasi antar variabel

Dataset menyediakan banyak variabel untuk di analisis. Untuk menentukan dua variabel mana yang akan di analisis, kami memetakan korelasi antara masing-masing variabel dengan *heatmap*. Tim kami memilih variabel jumlah kasus positif aktif dengan variabel jumlah isolasi mandiri.



Gambar 6. Grafik hubungan antara positif aktif dengan isolasi mandiri

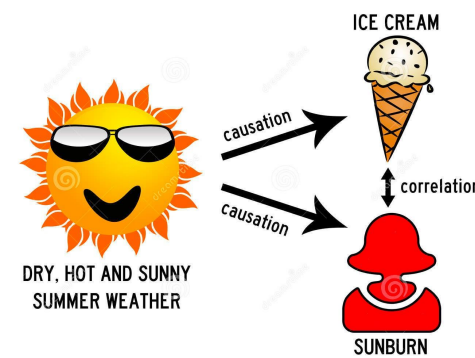
Seperti yang terlihat pada Gambar 6. Korelasi antara kedua variabel saling berbanding lurus. Semakin tinggi kasus positif yang aktif, semakin tinggi juga jumlah isolasi mandiri.

#### **b. Jawaban untuk soal non-dataset**

1. Jelaskan secara teori statistik mengenai outlier (pencilan), implikasinya dalam analisis data, serta bagaimana melakukan manajemen data terhadap kasus outlier.
  - a. Dalam *Engineering Statistics Handbook* (2012), *Outlier* adalah data yang letaknya berada pada jarak yang abnormal dari data-data lain. Sedangkan menurut Moore and McCabe dalam John (, *Outlier* adalah data yang terletak diluar pola keseluruhan dari sebuah distribusi. Sehingga outlier adalah data yang nilainya memiliki perbedaan yang signifikan dengan mayoritas data yang ada. Cara menentukan apakah sebuah data adalah outlier ada dua, yaitu berdasarkan standar deviasi dan kuartil.

- b. Apabila kita ingin memakai data yang ada untuk membuat model machine learning, ada baiknya data outlier dibuang sehingga model yang dibuat nanti memiliki performa yang baik dan tidak terpengaruh dengan adanya *outlier*.
2. Jelaskan konsep dan prinsip korelasi, lalu sebisa mungkin kaitkan dengan dasar-dasar statistik serta implikasinya terhadap konsep/teori statistik lain.

Korelasi adalah sebuah teknik dalam statistika untuk menentukan kuat pengaruh satu variabel dengan variabel lain. Korelasi antara dua variabel semakin berbanding lurus secara linear (semakin besar variabel A, semakin besar variabel B) bila nilai korelasi mendekati satu. Bila nilai korelasi semakin mendekati negatif satu, maka hubungan antara dua variabel akan semakin berbanding terbalik secara linear (semakin besar variabel A, semakin kecil variabel B).



Gambar 7. Penggambaran korelasi dan kausalitas.

Korelasi memang menggambarkan hubungan antara satu variabel dengan variabel lain, namun tidak selalu menggambarkan hubungan yang terjadi karena keterkaitan suatu hal atau kausalitas. Semisal bila dilihat dari korelasi, kenaikan kasus kulit terbakar (*sunburn*) berbanding lurus dengan penjualan es krim. Semakin banyak es krim yang terjual, maka semakin tinggi juga kasus kulit terbakar. Namun dibalik korelasi tersebut ada variabel lain yang juga berperan penting yaitu suhu dan tingkat konsentrasi sinar UV. Semakin tinggi suhu harian, semakin tinggi juga penjualan es krim. Hal itu dapat kita logikakan dengan sederhana. Begitu juga dengan konsentrasi sinar UV harian,

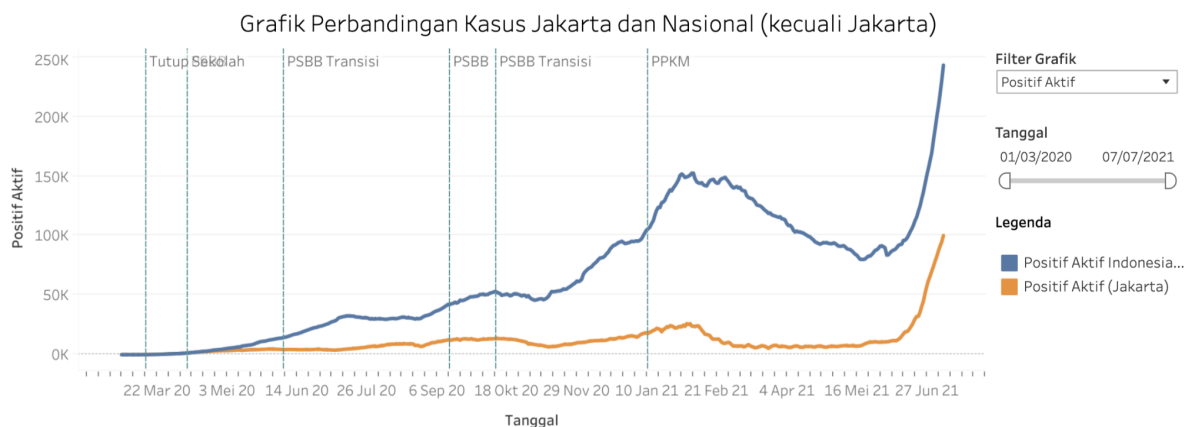
semakin tinggi suhu pada hari tersebut, konsentrasi sinar UV akan semakin tinggi dan dapat menyebabkan *sunburn* (Sulistya, 2020).

3. Sebutkan teori dasar machine learning yang kalian ketahui, lalu jelaskan dalam bahasa sederhana mengenai teori tersebut dan implikasinya.
  - a. Menurut Tom Mitchell (dalam Sahil, 2018), pembelajaran mesin adalah “Sebuah komputer di program untuk belajar dari pengalaman (data)  $E$  yang berkaitan dengan beberapa tugas (class)  $T$  dan ukuran performansi  $P$  saat melaksanakan tugas  $T$ , saat diukur performa  $P$  meningkat dengan pengalaman  $E$ ”.
  - b. Machine learning adalah suatu sistem yang dibuat untuk menyelesaikan permasalahan dengan cara mempelajari data-data yang berkaitan agar mendapatkan pengalaman guna menyelesaikan permasalahan tersebut.
  - c. Pada machine learning, terdapat konsep *garbage in garbage out*, yaitu sebaik apapun model algoritma yang digunakan, apabila data yang digunakan itu kualitasnya buruk maka model yang dihasilkan akan linear dengan kualitas data yang diberikan. Sehingga penting untuk seorang *machine learning engineer* untuk memahami data yang diperlukan guna membuat model yang baik.
4. Menggunakan bahasa kalian sendiri, jelaskan kaitan antara *artificial intelligence*, *machine learning*, dan *deep learning*.
  - a. *Artificial Intelligence* adalah kemampuan program untuk belajar dan berpikir seperti manusia sehingga dapat menyelesaikan suatu permasalahan tertentu.
  - b. *Machine Learning* adalah sebuah algoritma yang memiliki kemampuan untuk belajar tanpa perlu di program secara eksplisit dengan cara dilatih menggunakan data yang bersangkutan dengan permasalahan yang ingin diselesaikan.

- c. *Deep Learning* adalah bagian dari machine learning yang dimana algoritma ini berbentuk seperti jaringan syaraf tiruan yang dapat beradaptasi dan belajar dari data dalam jumlah yang besar.
5. Apakah yang kalian ketahui mengenai interpretasi data? Bagaimana signifikansi dan tantangannya? Bagaimana kaitan interpretasi data dengan *data story telling* dan *decision making*?
- a. Interpretasi data adalah suatu metode yang digunakan untuk menyampaikan arti atau temuan yang terdapat pada data, pada umumnya seorang *data analyst* akan melakukan interpretasi data dengan cara visualisasi sehingga data dapat dengan lebih mudah untuk dipahami oleh *stakeholder*. Proses visualisasi ini sangatlah penting dikarenakan kita sebagai manusia akan lebih mudah untuk memahami sesuatu berdasarkan visual.

Tanggal	Positif Harian
2021-06-27	9394
2021-06-28	8348
2021-06-29	7379
2021-06-30	7680
2021-07-01	7541
2021-07-02	9399
2021-07-03	9702
2021-07-04	10485
2021-07-05	10903
2021-07-06	9439

Gambar 8. data tanpa visualisasi, susah untuk melihat pola di dalamnya.



Gambar 9. Dengan visualisasi data, kita dapat memahami pola dengan lebih mudah.

Kaitan interpretasi data dengan *data story telling* adalah dalam melakukan *data storytelling*, dibutuhkan informasi yang bermanfaat, memiliki tujuan dan cerita yang jelas, dan juga terdapat visualisasi yang menarik sehingga orang yang melihat akan dengan cepat memahami data yang ada.

Sedangkan kaitan dari interpretasi data dengan decision making adalah dengan adanya interpretasi data yang informatif dan mencantumkan juga rekomendasi aksi selanjutnya yang dapat diambil, orang yang melihat interpretasi data tersebut dapat dengan yakin mengambil keputusan dikarenakan telah melihat situasi yang ada dari data dan rekomendasi keputusan yang perlu diambil.

## **C. Analisis Tambahan**

### *1. Problem Statement*

- Hubungan antara kasus positif harian COVID-19 di Jakarta dengan bertambahnya jumlah orang yang melakukan isolasi mandiri.
- seberapa signifikan pengaruh kasus COVID-19 di Jakarta dengan keparahan kasus keseluruhan di Indonesia
- pengaruh vaksinasi dengan penekanan angka kenaikan covid-19 di Indonesia
- pengaruh pandemi COVID-19 terhadap mobilitas masyarakat Indonesia

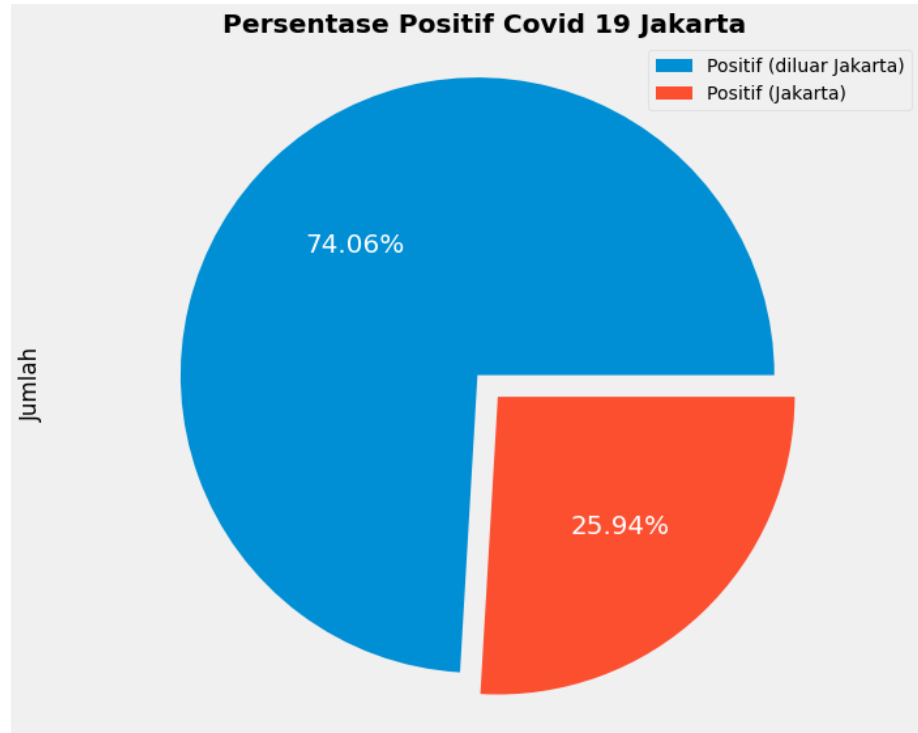
### *2. Hipotesis*

- Semakin banyak orang yang didiagnosa positif COVID-19 maka semakin banyak jumlah orang yang melakukan isolasi mandiri.
- Kasus positif aktif di Jakarta bertambah setiap harinya khususnya setelah arus mudik
- Vaksinasi tidak mempengaruhi penekanan angka kenaikan kasus positif Covid-19 di Indonesia
- Angka kenaikan kasus positif Covid-19 melambung tinggi setelah varian delta muncul



### 3. Exploratory Data Analysis

- Seberapa signifikan pengaruh kasus COVID-19 di Jakarta dengan keparahan kasus keseluruhan di Indonesia:



Gambar 10. Persentase Positif Covid-19 di Jakarta dengan Indonesia

- Pengaruh vaksinasi dengan penekanan angka kenaikan covid-19 di DKI Jakarta

	Vaksin 2	Positif Harian
Vaksin 2	1.000000	0.654995
Positif Harian	0.654995	1.000000

Gambar 11. Korelasi hubungan antara masyarakat tervaksinasi dengan positif harian Jakarta

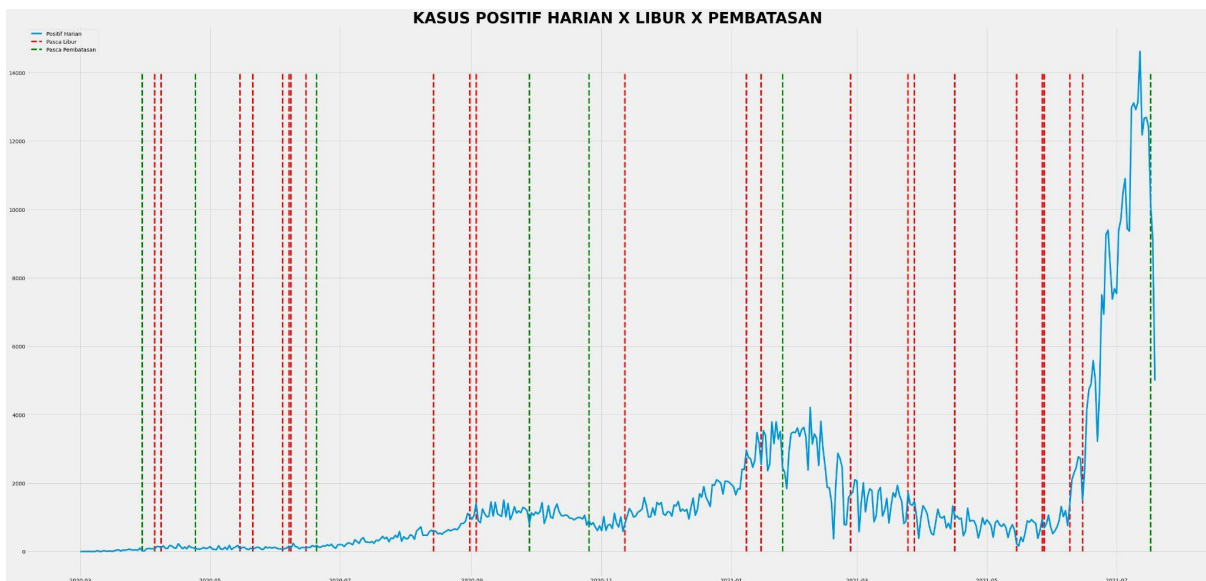
### 4. Initial Findings

- Sampai tanggal 17 Juli 2021 jumlah positif harian maksimal Jakarta mencapai 14619 kasus dengan rata-rata positif harian 1445 kasus
- Jakarta menyumbang 25.94% dari total keseluruhan kasus positif Covid-19 di Indonesia

- Vaksinasi tidak serta-merta menurunkan angka kenaikan positif COVID-19. Untuk menekan angka kenaikan positif COVID-19 tetap perlu menjalankan protokol kesehatan yang tepat.

### 5. Deep Dive Analysis

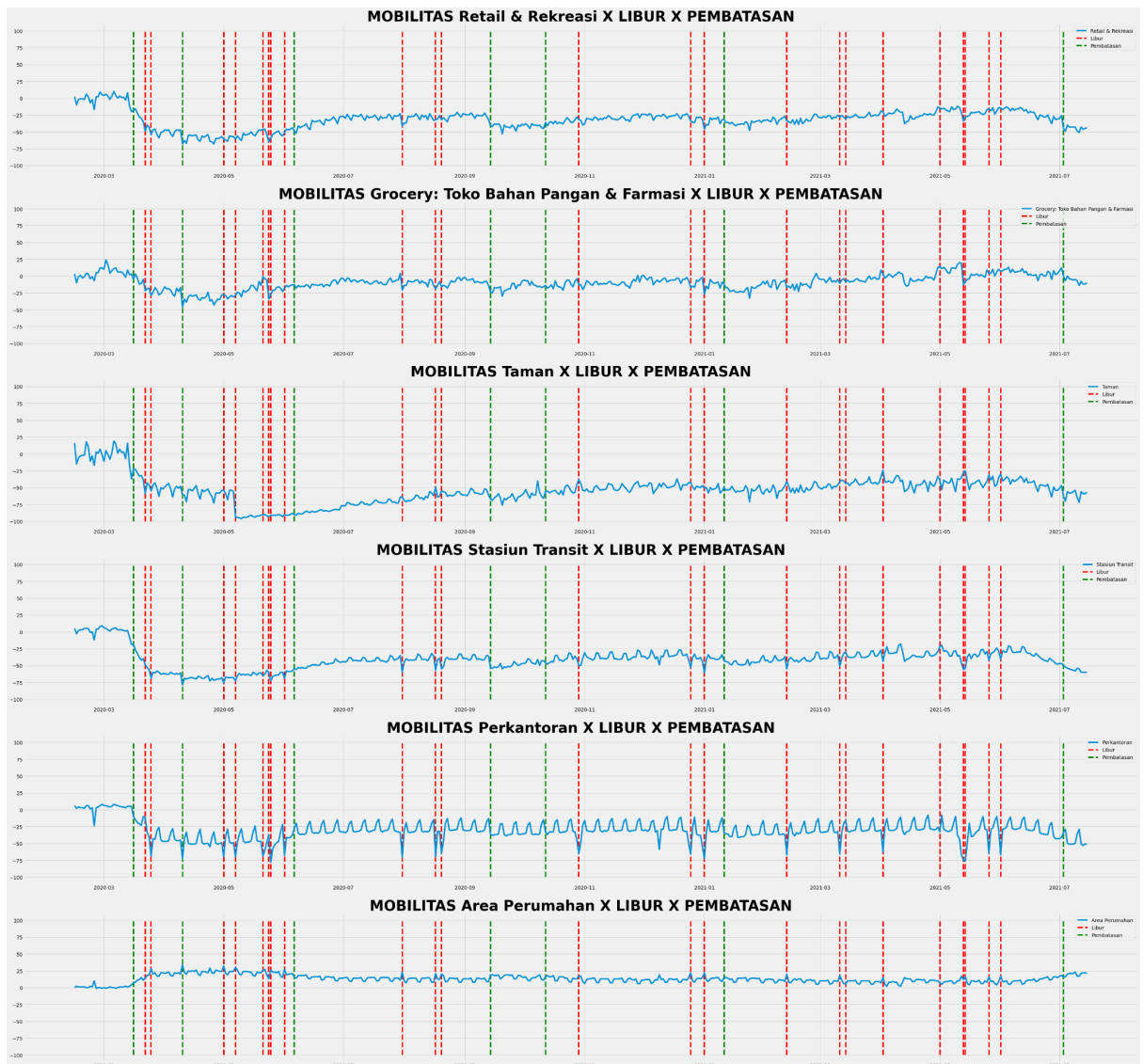
Selanjutnya untuk analisis mendalam, tim kami mencari pengaruh PPKM dan hari raya di Indonesia terhadap kenaikan positif COVID-19.



Gambar 12. Grafik hubungan positif harian, hari raya, dan PPKM

Pada *reference line* berwarna hijau menunjukkan 14 hari setelah terjadinya pembatasan sosial. *Reference line* warna merah menunjukkan 14 hari setelah hari raya di Indonesia. Dari data tersebut kita melihat sebuah fakta bahwa pembatasan sosial yang terjadi di D.K.I. Jakarta berhasil menurunkan angka kasus positif harian COVID-19 namun hanya bersifat sementara. Berbanding terbalik dengan pembatasan sosial, libur hari raya merupakan pemicu terbesar dalam menaikkan kasus positif harian COVID-19. Turunnya kasus harian akibat pembatasan sosial bisa kembali meningkat di sekitaran hari raya. Hal ini mungkin dikarenakan pada libur hari raya, orang Indonesia cenderung untuk berkumpul dengan anggota keluarga, dan biasanya tidak menaati protokol kesehatan. Kemungkinan tersebut diperkuat dari contoh kenaikan kasus 14 hari pasca idul fitri. Masyarakat cenderung nekat untuk berpergian

dan berkumpul dengan anggota keluarga. Pernyataan ini didukung dengan data mobilitas masyarakat Jakarta selama pandemi COVID-19 (Google, 2021) .



Gambar 13. Grafik pengaruh libur nasional dan PPKM terhadap mobilitas masyarakat selama masa pandemi.

Seperti grafik sebelumnya, *reference line* berwarna hijau menunjukkan batas terjadinya pembatasan sosial. *Reference line* warna merah menunjukkan batas hari raya. Grafik di atas menunjukkan peningkatan mobilitas bahkan setelah pembatasan sosial, khususnya di sekitaran hari raya.

#### 6. *Conclusion and Recommendation*

Sampai saat ini angka kenaikan kasus positif COVID-19 masih terus bertambah setiap harinya bahkan semakin parah setelah masuknya varian delta. Meskipun vaksinasi sudah diberikan tetapi masih belum cukup untuk menekan angka kenaikan kasus positif COVID-19 yang ada. Hal yang harus dilakukan oleh warga negara Indonesia khususnya warga D.K.I. Jakarta yang menyumbang 25.94% dari total keseluruhan angka kenaikan kasus positif COVID-19 adalah untuk menerapkan protokol kesehatan semaksimal mungkin, karena vaksinasi tidak menjamin seseorang kebal terhadap COVID-19.

#### **D. Kesimpulan**

Pandemi COVID-19 masih terus ada bahkan kasus positif meningkat setiap harinya walaupun vaksinasi dan PPKM sudah dijalankan tetapi protokol kesehatan khususnya pembatasan bepergian masih belum dijalankan dengan maksimal dan masyarakat masih meremehkan Pandemi COVID-19 ini. Agar pandemi COVID-19 ini cepat berakhir diperlukan kerjasama antara pemerintah dan masyarakatnya tidak hanya salah satu pihak, pemerintah dalam menjalankan PPKM dan memberikan bantuan sosial terhadap masyarakatnya dan rakyat yang mengikuti aturan dengan baik dan saling mengingatkan untuk selalu melakukan protokol kesehatan sehingga pandemi COVID-19 dapat cepat berakhir.

## DAFTAR PUSTAKA

NIST, U.S. Department of Commerce. 2012, *Engineering Statistics Handbook*.  
<https://www.itl.nist.gov/div898/handbook/prc/section1/prc16.htm>  
diakses pada tanggal 10 Juli 2021

Renze, John. 2021. "Outlier." dari *MathWorld--A Wolfram Web Resource*, oleh Eric W. Weisstein. <https://mathworld.wolfram.com/Outlier.html>  
diakses pada tanggal 10 Juli 2021

Sulistiya, Prima. 2020. *Beda Korelasi dan Kausalitas, Agar Tak Salah Ambil Kesimpulan*. Tersedia pada  
<https://mojok.co/prm/komen/versus/beda-korelasi-dan-kausalitas-agar-tak-salah-ambil-kesimpulan/>  
diakses pada tanggal 10 Juli 2021

Sahil, Verma. 2018. *Machine Learning Notes 1*. Tersedia pada  
<https://hackernoon.com/machine-learning-notes-1-c496591f49b8>  
diakses pada tanggal 10 Juli 2021

Google. 2021. *COVID-19 Community Mobility Reports*. Tersedia pada  
<https://www.google.com/covid19/mobility/>  
diakses pada tanggal 11 Juli 2021

## LAMPIRAN

*Notebook* yang kami gunakan untuk mengerjakan soal ini tersedia pada:

<https://colab.research.google.com/drive/1TsrFIkm7Vb6dxIOjByCD5kY5W0AHno89?usp=sharing>