

## Jawaban Non-Dataset

1. Outlier adalah data point yang menunjukkan pola yang berbeda dibanding dengan data point mayoritas.

Cara mengidentifikasi outlier:

- Menggunakan boxplot. Apabila data point terletak lebih dari kuartil 3 + 1.5 dikali interquartile range dan kuartil 1 - 1.5 dikali interquartile range, maka dapat dikatakan bahwa data point tersebut termasuk dalam outlier.
- Cook's measure. Cook's measure digunakan untuk menghitung perubahan estimasi dari parameter saat pengamatan. Dengan menggunakan perhitungan statistik oleh Cook's measure, hasilnya akan menunjukkan data yang sangat mempengaruhi hasil. Namun, data yang mempengaruhi hasil dapat dikategorikan menjadi 2 tipe yaitu influential point dan outlier. Influential point berarti data point yang sangat menginfluence parameter fit. Namun, influential point bisa jadi merupakan outlier. Maka dari itu, outlier tidak selalu buruk dan harus dihilangkan. Dengan perhitungan statistik, apabila  $D_i$  yaitu magnitude yang akan dihitung dengan perhitungan statistik lebih besar dari 1 dengan moderate  $n$  dan  $p$  yang tidak terlalu kecil,  $D_i$  dengan  $i$  adalah data point ke  $i$  merupakan influential point.
- DFFITS Measure. Measure ini digunakan untuk mengetahui pengaruh data point ke  $i$  terhadap regression model yang diamati dari fitted value. Dengan menggunakan perhitungan statistik, apabila nilai mutlak DFFITS di datapoint ke  $i$  melebihi nilai dari  $2 \sqrt{p/n}$ , maka DFFITS ke  $i$  tersebut dapat dikatakan sebagai outlier.

Cara mengatasi outlier:

- Menghapus outlier berdasarkan diagnostik outlier. Diagnostik outlier dapat dilakukan dengan melakukan perhitungan statistik dengan Cook's measure atau DFFITS measure.
- Menghapus outlier dengan menggunakan residual analysis. For example, dengan menggunakan plot dari scaled residual dengan fitted value untuk mengidentifikasi outlier.
- Retain outlier dengan menggunakan robust regression (untuk analisis regresi)

2. Korelasi adalah suatu statistik teori untuk menghitung hubungan antara 2 variabel. Dengan menggunakan pearson correlation coefficient, apabila  $r$  yaitu coefficient nya adalah 0, maka tidak ada korelasi antar 2 variabel tersebut. Apabila  $r$  nya positif, maka terdapat positive correlation antara 2 variabel tersebut. Positive correlation berarti apabila variabel  $a$  naik,

maka variable  $b$  juga naik. Sedangkan apabila  $r$  nya negatif, maka terdapat negative correlation antara 2 variabel tersebut. Negative correlation berarti apabila variable  $a$  naik, maka variable  $b$  turun dan sebaliknya. Apabila nilai mutlak  $r$  mendekati 1, maka korelasi antara 2 variabel tersebut semakin kuat. Korelasi berkaitan erat dengan regresi. Korelasi bertujuan untuk mengetahui derajat keeratan hubungan antar variabel. Sedangkan regresi adalah untuk prediksi variable dependent berdasarkan variabel-variabel independennya. Apabila terdapat korelasi antara beberapa variabel, maka dapat dilanjutkan dengan analisis regresi untuk memprediksi variabel dependen nya. variabel independen adalah variabel input dan asumsi yang dapat mempengaruhi variabel dependen nya. Variabel independen dalam regression analysis dapat berupa satu maupun lebih. Dari input dan asumsi variabel-variabel independen yang kita miliki, kita dapat memprediksi variabel dependen nya.

3. Machine Learning adalah sejenis AI (artificial intelligence) yang dapat membuat software yang mampu “belajar” dan secara perlahan meningkatkan akurasi dalam memprediksi hal-hal. Software yang menggunakan machine learning menggunakan data dari input maupun output masa lalu dan menganalisa data tersebut untuk dapat lebih akurat memprediksi output baru. Machine learning menggunakan sistem belajar melalui pengalaman. Oleh karena itu, dengan waktu yang cukup, sebuah AI dengan machine learning dapat berkembang dan mempelajari hal-hal baru. Tahapan kerja machine learning adalah memilah data, memodelkan data dan memvalidasi model, dan mengetes model dan pemakaian test model. Dengan machine learning, komputer-komputer sekarang hanya perlu diberikan input inisial berupa data, dan data tersebut akan diproses dan dipelajari oleh komputer untuk mengeluarkan output yang berguna, tanpa harus diprogram secara spesifik.

4. Secara dasar, deep learning dapat dilihat sebagai bagian dari machine learning, dan machine learning bagian dari artificial intelligence. Artificial intelligence berarti kepintaran buatan, atau kepintaran yang dimiliki oleh mesin. Artificial intelligence mencakup semua program atau software yang dapat berpikir dan berakal seperti manusia. Beberapa penerapan artificial intelligence adalah teknologi pengenalan suara/muka, mobil pengemudi otomatis, dan robot. Machine learning dan deep learning adalah 2 contoh penerapan artificial intelligence. Machine learning adalah algoritma yang dapat memungkinkan software/komputer untuk belajar dan berkembang dengan menggunakan input berupa data. Jika data yang diberikan cukup besar dan berguna, maka dengan machine learning komputer akan dapat memproses dan menganalisa data tersebut tanpa harus diprogram secara eksplisit.

Deep learning hampir sama dengan machine learning, tetapi lebih mendalam dan lebih kompleks. Dalam deep learning, algoritma yang digunakan menyerupai otak manusia. Deep learning menggunakan *artificial neural network* untuk mendapatkan data yang sangat besar. Dengan data ini performansi algoritma dapat terus berkembang. Salah satu perbedaan machine learning dan deep learning adalah perkembangan mereka. Algoritma deep learning akan terus berkembang semakin banyak data yang diberikan, sedangkan untuk machine learning perkembangannya akan memperlambat atau berhenti setelah data yang diberikan melampaui batas tertentu.

5. Interpretasi data adalah implementasi berbagai macam proses dimana data diolah atau dianalisa untuk dapat memperoleh sebuah konklusi yang berguna. Dengan interpretasi data, data-data yang sebelumnya terlihat tidak berguna dapat memberi berbagai macam informasi yang berguna.

Beberapa signifikansi interpretasi data adalah kita dapat membantu mengambil keputusan yang tepat, memprediksi trend - trend di masa depan, meningkatkan produktivitas kerja, memangkas ongkos, dan memprediksi dan menghindari masalah - masalah yang dapat timbul di masa depan. Dengan mengetahui data- data yang penting, sebuah perusahaan dapat menganalisa dan performa kerja dan tanggapan pelanggan. Informasi dari data tersebut dapat menunjukan hal-hal yang perlu diperbaiki dalam pengelolaan sebuah perusahaan. Sebuah perusahaan juga dapat memaksimalkan profit mereka, serta mengurangi ongkos-ongkos yang tidak penting. Terakhir data - data dapat memprediksi hal-hal di masa depan seperti trend-trend yang akan datang ataupun masalah yang dapat timbul. Dengan informasi ini perusahaan dapat mengambil keputusan yang diperlukan untuk menghindari timbulnya masalah tersebut dan juga memanfaatkan trend-trend tersebut.

Data interpretasi juga memiliki berbagai tantangan seperti korelasi disalah artikan sebagai sebab-akibat, terkadang data set yang didapatkan terlalu beragam dan irrelevant, dan pengolahan data dipengaruhi bias konfirmasi.

Data storytelling adalah langkah selanjutnya setelah kita menginterpretasi/mengolah data. Dengan data storytelling kita menyampaikan point-point penting data-data tersebut tidak hanya dengan graphic dan angka-angka, tapi melalui sebuah narasi, menjadi sesuatu yang lebih mudah dimengerti manusia. Setelah memahami data-data yang diberikan, adalah

saatnya untuk decision making, atau pengambilan keputusan. Data dapat sangat membantu dalam decision making, dengan data yang tepat, persentase untuk mengambil keputusan yang tepat untuk suatu hal akan meningkat. Keputusan – keputusan seperti mencari tempat lokasi membuka restoran baru, atau dimana harus mengiklankan sebuah produk, dengan data yang tepat kita dapat memutuskan keputusan yang paling optimal dan efisien.

# LAPORAN ANALISA DATA COVID - 19 DI INDONESIA

*Shenzhen Tech*



Disusun Oleh :

1. Glenys Charity Lion
2. Dharma Setiawan
3. Jefferson Tedjojuwono

The Chinese University of Hong Kong, Shenzhen

2021

## **BAB I : PENDAHULUAN**

### **1.1 Latar Belakang**

Covid-19 adalah penyakit menular yang disebabkan oleh jenis coronavirus yang baru ditemukan. Coronavirus sendiri merupakan suatu kelompok virus yang dapat menyebabkan penyakit pada hewan atau manusia. Beberapa jenis coronavirus diketahui menyebabkan infeksi saluran nafas pada manusia. Coronavirus merupakan jenis baru yang ditemukan menyebabkan penyakit Covid-19. Virus baru dan penyakit yang disebabkan ini tidak dikenal sebelum mulai nya wabah di Wuhan, Tiongkok, bulan Desember 2019. Covid-19 ini sekarang menjadi sebuah pandemi yang terjadi di banyak negara di seluruh dunia (WHO Indonesia, 2020).

Penyebaran Covid-19 di Indonesia dimulai didata pada tanggal 1 Maret 2020. Virus ini menyebar dengan sangat cepat ke seluruh provinsi di Indonesia. Jakarta, sebagai pusat ibukota, terkena dampak yang paling parah dari penyebaran virus ini. Pemerintah Indonesia telah melakukan tindakan untuk mencegah penyebaran virus ini salah satunya adalah PSBB (Pembatasan Sosial Berskala Besar).

Kenaikan Covid Indonesia dipicu oleh beberapa faktor. Salah satu faktor yang berpengaruh adalah libur panjang di Indonesia. Kenaikan ini tercatat mencapai puncak pada Januari 2021 setelah libur akhir tahun dan pada Juli 2021 setelah libur Idul Fitri. Saat ini, Indonesia mencatat kenaikan rekor kasus baru dan rekor kematian terbanyak hampir setiap hari.

Pada bulan Juli 2021, Indonesia secara resmi telah memasuki gelombang kedua Covid-19. Menurut Makki (2021), kondisi itu ditandai dengan kasus konfirmasi Covid-19 yang tembus 21.342 orang pada Minggu (27/6). Indonesia telah mengalami penurunan 244 persen dari puncak gelombang pertama.

## **BAB II : ISI**

### **2.1 Jawaban Soal**

(Data yang kami ambil adalah data dari tanggal 1 Maret 2020 hingga 3 Juli 2021)

1. Dari dataset yang disediakan, nilai mean, median, dan modus dari positif COVID-19 harian Jakarta adalah:
  - Mean: 1.163,489796 kasus positif
  - Median: 855 kasus positif
  - Modus: 0 kasus positif
2. Dari dataset yang disediakan, nilai maksimal dan minimal dari COVID-19 harian Jakarta adalah:
  - Nilai Maksimal: 9.702 kasus positif
  - Nilai Minimal: 0 kasus positif
3. Dari dataset yang disediakan, kita menggunakan analisis boxplot dan didapatkan nilai-nilai outlier yang ada sebagai berikut:
  - Meninggal (Indonesia) : 0
  - Sembuh (Indonesia) : 0
  - Dirawat (Indonesia) : 4
  - Positif (Indonesia) : 0
  - Meninggal (Jakarta) : 0
  - Sembuh (Jakarta) : 0
  - Self - Isolation (Jakarta) : 34
  - Dirawat (Jakarta) : 25
  - Positif (Jakarta) : 0
  - Positif Aktif (Jakarta) : 18
  - Meninggal Harian (Indonesia) : 14
  - Sembuh Harian (Indonesia) : 0
  - Dirawat Harian (Indonesia) : 75
  - Positif Harian (Indonesia) : 19
  - Meninggal Harian (Jakarta) : 23
  - Sembuh Harian (Jakarta) : 28
  - Self - Isolation Harian (Jakarta) : 68
  - Dirawat Harian (Jakarta) : 66
  - Positif Harian (Jakarta) : 38

- Positif Aktif Harian (Jakarta) : 83

4. Dari dataset yang disediakan, kami mengusulkan dua variabel yaitu positif Indonesia dan positif Jakarta untuk dianalisis. Kami menemukan bahwa kedua variabel tersebut berkorelasi positif dengan hasil analisis koefisien korelasi bernilai 0,999079. Koefisien korelasi yang bernilai positif mengindikasikan bahwa kenaikan kasus positif Jakarta berbanding lurus dengan kenaikan kasus positif di Indonesia secara keseluruhan dan nilai yang mendekati 1 menandakan korelasi yang ditunjukkan sangat kuat. Jika digambar di grafik kartesius dengan kasus positif Indonesia sebagai sumbu x dan kasus positif Jakarta sebagai sumbu y, maka pada penambahan variabel x akan ada penambahan pada variabel y.

Kesimpulan dari hasil analisis diatas adalah Jakarta sebagai ibukota dengan tingkat kepadatan penduduk terpadat di Indonesia sangat berhubungan dengan kenaikan kasus positif di Indonesia.

## 2.2 Hasil Analisis Tambahan

5. Dari dataset yang disediakan, dapat kita analisis sebagai berikut:

- a. Problem Statement

- i. Variabel mana sajakah yang memiliki korelasi yang kuat antara satu sama lain?

- b. Hypothesis

Dari dataset yang tersedia, semua data menunjukkan korelasi positif dan variabel-variabel dataset di Jakarta dan Indonesia memiliki korelasi yang kuat.

- c. Exploratory Data Analysis

Dari dataset yang tersedia, kami melakukan exploratory data analysis dengan menggunakan data visualisasi. Beberapa data visualisasi yang kami gunakan adalah boxplot, scatterplot, dan heatmap.

- Boxplot adalah suatu cara menganalisis dalam statistik dimana kita dapat menggambarkan suatu data berdasarkan 5 hal penting seperti nilai observasi terkecil (minimum), kuartil terendah atau kuartil pertama (Q1) yang memotong 25% dari data terendah, median (Q2) atau nilai pertengahan, kuartil tertinggi atau kuartil ketiga (Q3), yang memotong 25% dari



data tertinggi, nilai observasi terbesar (nilai maksimal). Dalam menggunakan boxplot, kita dapat mendapatkan nilai-nilai outlier dari observasi.

Dalam penggunaannya, kita dapat mengenali Interquartile Range. Interquartile Range (IQR) adalah jarak antara kuartil atas dan kuartil bawah. Rumus IQR adalah  $Q3 - Q1$ .

Bagian tubuh boxplot terdiri dari sebuah kotak dan dua buah garis yang terletak di atas dan di bawah. Pada bagian kotak tersebut terdapat sebuah garis tambahan yang kita kenal sebagai median. Semua data yang berada di luar kotak dan garis tersebut dinamakan outlier. Dalam boxplot, outlier dilambangkan dengan titik- titik.

Penggunaan boxplot dapat digunakan untuk menunjukkan bagaimana data tersebar. Data dapat tersebar secara positif, negatif, dan simetris.

- Scatterplot adalah cara menganalisis data dengan menggunakan koordinat kartesian untuk membandingkan 2 data. Scatterplot sangat berguna untuk analisis data karena dengan menggunakan scatterplot akan lebih terlihat korelasi yang ada di antara dua jenis data. Jika tidak ada korelasi di antara dua data, maka poin-poin yang ada di scatterplot akan tampil acak. Jika terdapat sebuah korelasi, maka poin-poin di scatterplot akan mendekati sebuah garis.

Selain menunjukkan ada tidaknya korelasi antara dua data, scatterplot juga bisa menjelaskan jenis korelasi yang terdapat. Jika nilai sumbu y meningkat seiring nilai naiknya nilai sumbu x, atau nilai sumbu y berbanding lurus dengan nilai sumbu x, maka korelasinya positif. Jika nilai sumbu y meningkat seiring dengan turunnya nilai sumbu x, atau nilai sumbu y berbanding terbalik dengan nilai sumbu x, maka korelasinya negatif. Jika poin-poin tidak mengikuti kedua kriteria tersebut, maka korelasinya 0, atau tidak ada korelasi. Walaupun scatterplot sangat berguna untuk mencari tahu korelasi antara dua data,

scatterplot tidak dapat menunjukkan apakah kedua data tersebut memiliki hubungan sebab-akibat (causation).

- Heatmap adalah cara merepresentasikan data dengan cara membedakan nilai berdasarkan warna. Dalam heatmap, terdapat data dalam bentuk tabel atau matriks dan setiap nilai data diberi warna yang berbeda tergantung nilai data tersebut. Nilai - nilai yang tinggi biasanya diberi warna yang lebih gelap daripada nilai - nilai yang rendah. Heatmap sangat berguna untuk menganalisis data karena penyajiannya yang lebih mudah dimengerti. Dengan heatmap data-data yang bernilai tinggi dapat lebih mudah ditemukan karena mereka akan memiliki warna yang lebih gelap dibandingkan data lainnya.

#### d. Initial Finding

- i. Ditemukan dari dataset bahwa variabel-variabel yang memiliki korelasi kuat (koefisien korelasi  $\Rightarrow 0,8$ ) adalah :

- Meninggal (Indonesia) & Sembuh (Indonesia)
- Meninggal (Indonesia) & Dirawat (Indonesia)
- Meninggal (Indonesia) & Positif (Indonesia)
- Meninggal (Indonesia) & Meninggal (Indonesia)
- Meninggal (Indonesia) & Sembuh (Jakarta)
- Meninggal (Indonesia) & Positif (Jakarta)
- Sembuh (Indonesia) & Dirawat (Indonesia)
- Sembuh (Indonesia) & Positif (Indonesia)
- Sembuh (Indonesia) & Meninggal (Jakarta)
- Sembuh (Indonesia) & Sembuh (Jakarta)
- Sembuh (Indonesia) & Positif (Jakarta)
- Dirawat (Indonesia) & Positif (Indonesia)
- Dirawat (Indonesia) & Meninggal (Jakarta)
- Dirawat (Indonesia) & Sembuh (Jakarta)
- Dirawat (Indonesia) & Positif (Jakarta)
- Dirawat (Indonesia) & Meninggal Harian (Indonesia)
- Dirawat (Indonesia) & Sembuh Harian (Indonesia)

- Dirawat (Indonesia) & Positif Harian (Indonesia)
- Dirawat (Indonesia) & Sembuh Harian (Indonesia)
- Positif (Indonesia) & Meninggal (Jakarta)
- Positif (Indonesia) & Sembuh (Jakarta)
- Positif (Indonesia) & Positif (Jakarta)
- Meninggal (Jakarta) & Sembuh (Jakarta)
- Meninggal (Jakarta) & Positif (Jakarta)
- Sembuh (Jakarta) & Positif (Jakarta)
- Self - Isolation (Jakarta) & Dirawat (Jakarta)
- Self - Isolation (Jakarta) & Positif Aktif (Jakarta)
- Self - Isolation (Jakarta) & Positif Harian (Indonesia)
- Self - Isolation (Jakarta) & Positif Harian (Jakarta)
- Dirawat (Jakarta) & Positif Aktif (Jakarta)
- Dirawat (Jakarta) & Positif Harian (Indonesia)
- Dirawat (Jakarta) & Positif Harian (Jakarta)
- Positif Aktif (Jakarta) & Meninggal Harian (Jakarta)
- Positif Aktif (Jakarta) & Positif Harian (Indonesia)
- Positif Aktif (Jakarta) & Positif Harian (Jakarta)
- Meninggal Harian (Indonesia) & Sembuh Harian (Indonesia)
- Meninggal Harian (Indonesia) & Positif Harian (Indonesia)
- Meninggal Harian (Indonesia) & Positif Harian (Jakarta)
- Sembuh Harian (Indonesia) & Positif Harian (Indonesia)
- Sembuh Harian (Indonesia) & Sembuh Harian (Jakarta)
- Dirawat Harian (Indonesia) & Positif Aktif Harian (Jakarta)
- Positif Harian (Indonesia) & Sembuh Harian (Jakarta)
- Positif Harian (Indonesia) & Positif Harian (Jakarta)
- Self - Isolation Harian (Jakarta) & Positif Harian (Jakarta)

ii. Ditemukan dari dataset bahwa variabel-variabel yang memiliki korelasi rendah (koefisien korelasi  $< 0,8$ ) adalah:

- Meninggal (Indonesia) & Self-Isolation (Jakarta)
- Meninggal (Indonesia) & Dirawat (Jakarta)
- Meninggal (Indonesia) & Positif Aktif (Jakarta)
- Meninggal (Indonesia) & Meninggal Harian (Indonesia)
- Meninggal (Indonesia) & Sembuh Harian (Indonesia)

- Meninggal (Indonesia) & Dirawat Harian (Indonesia)
- Meninggal (Indonesia) & Positif Harian (Indonesia)
- Meninggal (Indonesia) & Meninggal Harian (Jakarta)
- Meninggal (Indonesia) & Sembuh Harian (Jakarta)
- Meninggal (Indonesia) & Self Isolation Harian (Jakarta)
- Meninggal (Indonesia) & Dirawat Harian (Jakarta)
- Meninggal (Indonesia) & Positif Harian (Jakarta)
- Meninggal (Indonesia) & Positif Aktif Harian (Jakarta)
- Sembuh (Indonesia) & Self-Isolation (Jakarta)
- Sembuh (Indonesia) & Dirawat (Jakarta)
- Sembuh (Indonesia) & Positif Aktif (Jakarta)
- Sembuh (Indonesia) & Meninggal Harian (Indonesia)
- Sembuh (Indonesia) & Sembuh Harian (Indonesia)
- Sembuh (Indonesia) & Dirawat Harian (Indonesia)
- Sembuh (Indonesia) & Positif Harian (Indonesia)
- Sembuh (Indonesia) & Meninggal Harian (Jakarta)
- Sembuh (Indonesia) & Sembuh Harian (Jakarta)
- Sembuh (Indonesia) & Self-Isolation (Jakarta)
- Sembuh (Indonesia) & Dirawat Harian (Jakarta)
- Sembuh (Indonesia) & Positif Harian (Jakarta)
- Sembuh (Indonesia) & Positif Aktif Harian (Jakarta)
- Dirawat (Indonesia) & Self-Isolation (Jakarta)
- Dirawat (Indonesia) & Dirawat (Jakarta)
- Dirawat (Indonesia) & Positif Aktif (Jakarta)
- Dirawat (Indonesia) & Dirawat Harian (Indonesia)
- Dirawat (Indonesia) & Meninggal Harian (Jakarta)
- Dirawat (Indonesia) & Self-Isolation Harian (Jakarta)
- Dirawat (Indonesia) & Dirawat Harian (Jakarta)
- Dirawat (Indonesia) & Positif Harian (Jakarta)
- Dirawat (Indonesia) & Positif Aktif Harian (Jakarta)
- Positif (Indonesia) & Self-Isolation (Jakarta)
- Positif (Indonesia) & Dirawat (Jakarta)
- Positif (Indonesia) & Positif Aktif (Jakarta)
- Positif (Indonesia) & Meninggal Harian (Indonesia)

- Positif (Indonesia) & Sembuh Harian (Indonesia)
- Positif (Indonesia) & Dirawat Harian (Indonesia)
- Positif (Indonesia) & Positif Harian (Indonesia)
- Positif (Indonesia) & Meninggal Harian (Jakarta)
- Positif (Indonesia) & Sembuh Harian (Jakarta)
- Positif (Indonesia) & Self-Isolation Harian (Jakarta)
- Positif (Indonesia) & Dirawat Harian (Jakarta)
- Positif (Indonesia) & Positif Harian (Jakarta)
- Positif (Indonesia) & Positif Aktif Harian (Jakarta)
- Meninggal (Jakarta) & Self-Isolation (Jakarta)
- Meninggal (Jakarta) & Dirawat (Jakarta)
- Meninggal (Jakarta) & Positif Aktif (Jakarta)
- Meninggal (Jakarta) & Meninggal Harian (Indonesia)
- Meninggal (Jakarta) & Sembuh Harian (Indonesia)
- Meninggal (Jakarta) & Dirawat Harian (Indonesia)
- Meninggal (Jakarta) & Positif Harian (Indonesia)
- Meninggal (Jakarta) & Meninggal Harian (Jakarta)
- Meninggal (Jakarta) & Sembuh Harian (Jakarta)
- Meninggal (Jakarta) & Self-Isolation Harian (Jakarta)
- Meninggal (Jakarta) & Dirawat Harian (Jakarta)
- Meninggal (Jakarta) & Positif Harian (Jakarta)
- Meninggal (Jakarta) & Positif Aktif Harian (Jakarta)
- Sembuh (Jakarta) & Self-Isolation (Jakarta)
- Sembuh (Jakarta) & Dirawat (Jakarta)
- Sembuh (Jakarta) & Positif Aktif (Jakarta)
- Sembuh (Jakarta) & Meninggal Harian (Indonesia)
- Sembuh (Jakarta) & Sembuh Harian (Indonesia)
- Sembuh (Jakarta) & Dirawat Harian (Indonesia)
- Sembuh (Jakarta) & Positif Harian (Indonesia)
- Sembuh (Jakarta) & Meninggal Harian (Jakarta)
- Sembuh (Jakarta) & Sembuh Harian (Jakarta)
- Sembuh (Jakarta) & Self-Isolation Harian (Jakarta)
- Sembuh (Jakarta) & Dirawat Harian (Jakarta)
- Sembuh (Jakarta) & Positif Harian (Jakarta)

- Sembuh (Jakarta) & Positif Aktif Harian (Jakarta)
- Self-Isolation (Jakarta) & Positif (Jakarta)
- Self-Isolation (Jakarta) & Meninggal Harian (Indonesia)
- Self-Isolation (Jakarta) & Sembuh Harian (Indonesia)
- Self-Isolation (Jakarta) & Dirawat Harian (Indonesia)
- Self-Isolation (Jakarta) & Meninggal Harian (Indonesia)
- Self-Isolation (Jakarta) & Sembuh Harian (Indonesia)
- Self-Isolation (Jakarta) & Self-Isolation Harian (Jakarta)
- Self-Isolation (Jakarta) & Dirawat Harian (Jakarta)
- Self-Isolation (Jakarta) & Positif Aktif Harian (Jakarta)
- Dirawat (Jakarta) & Positif (Jakarta)
- Dirawat (Jakarta) & Meninggal Harian (Indonesia)
- Dirawat (Jakarta) & Sembuh Harian (Indonesia)
- Dirawat (Jakarta) & Dirawat Harian (Indonesia)
- Dirawat (Jakarta) & Meninggal Harian (Jakarta)
- Dirawat (Jakarta) & Sembuh Harian (Jakarta)
- Dirawat (Jakarta) & Self-Isolation (Jakarta)
- Dirawat (Jakarta) & Dirawat Harian (Jakarta)
- Dirawat (Jakarta) & Positif Aktif Harian (Jakarta)
- Positif (Jakarta) & Positif Aktif (Jakarta)
- Positif (Jakarta) & Meninggal Harian (Indonesia)
- Positif (Jakarta) & Sembuh Harian (Indonesia)
- Positif (Jakarta) & Dirawat Harian (Indonesia)
- Positif (Jakarta) & Positif Harian (Indonesia)
- Positif (Jakarta) & Meninggal Harian (Jakarta)
- Positif (Jakarta) & Sembuh Harian (Jakarta)
- Positif (Jakarta) & Self-Isolation Harian (Jakarta)
- Positif (Jakarta) & Dirawat Harian (Jakarta)
- Positif (Jakarta) & Positif Harian (Jakarta)
- Positif (Jakarta) & Positif Aktif Harian (Jakarta)
- Positif Aktif (Jakarta) & Sembuh Harian (Indonesia)
- Positif Aktif (Jakarta) & Dirawat Harian (Indonesia)
- Positif Aktif (Jakarta) & Meninggal Harian (Jakarta)
- Positif Aktif (Jakarta) & Sembuh Harian (Jakarta)

- Positif Aktif (Jakarta) & Self-Isolation Harian (Jakarta)
- Positif Aktif (Jakarta) & Dirawat Harian (Jakarta)
- Positif Aktif (Jakarta) & Positif Aktif Harian (Jakarta)
- Meninggal Harian (Indonesia) & Dirawat Harian (Indonesia)
- Meninggal Harian (Indonesia) & Meninggal Harian (Jakarta)
- Meninggal Harian (Indonesia) & Sembuh Harian (Jakarta)
- Meninggal Harian (Indonesia) & Self-Isolation Harian (Jakarta)
- Meninggal Harian (Indonesia) & Dirawat Harian (Jakarta)
- Meninggal Harian (Indonesia) & Positif Aktif Harian (Jakarta)
- Sembuh Harian (Indonesia) & Dirawat Harian (Indonesia)
- Sembuh Harian (Indonesia) & Meninggal Harian (Jakarta)
- Sembuh Harian (Indonesia) & Self-Isolation Harian (Jakarta)
- Sembuh Harian (Indonesia) & Dirawat Harian (Jakarta)
- Sembuh Harian (Indonesia) & Positif Harian (Jakarta)
- Sembuh Harian (Indonesia) & Positif Aktif Harian (Jakarta)
- Dirawat Harian (Indonesia) & Positif Harian (Indonesia)
- Dirawat Harian (Indonesia) & Meninggal Harian (Jakarta)
- Dirawat Harian (Indonesia) & Self-Isolation Harian (Jakarta)
- Dirawat Harian (Indonesia) & Sembuh Harian (Jakarta)
- Dirawat Harian (Indonesia) & Dirawat Harian (Jakarta)
- Dirawat Harian (Indonesia) & Positif Harian (Jakarta)
- Positif Harian (Indonesia) & Meninggal Harian (Jakarta)
- Positif Harian (Indonesia) & Self-Isolation Harian (Jakarta)
- Positif Harian (Indonesia) & Dirawat Harian (Jakarta)
- Positif Harian (Indonesia) & Positif Harian (Jakarta)
- Positif Harian (Indonesia) & Positif Aktif Harian (Jakarta)
- Meninggal Harian (Jakarta) & Sembuh Harian (Jakarta)
- Meninggal Harian (Jakarta) & Self-Isolation Harian (Jakarta)
- Meninggal Harian (Jakarta) & Dirawat Harian (Jakarta)
- Meninggal Harian (Jakarta) & Positif Harian (Jakarta)
- Meninggal Harian (Jakarta) & Positif Aktif Harian (Jakarta)
- Sembuh Harian (Jakarta) & Self-Isolation Harian (Jakarta)
- Sembuh Harian (Jakarta) & Dirawat Harian (Jakarta)
- Sembuh Harian (Jakarta) & Positif Aktif Harian (Jakarta)

- Self-Isolation Harian (Jakarta) & Dirawat Harian (Jakarta)
- Self-Isolation Harian (Jakarta) & Positif Harian (Jakarta)
- Dirawat Harian (Jakarta) & Positif Harian (Jakarta)
- Dirawat Harian (Jakarta) & Positif Aktif Harian (Jakarta)
- Positif Harian (Jakarta) & Positif Aktif Harian (Jakarta)

e. Deep Dive Analysis

Dari dataset yang telah diberikan, kami melakukan beberapa analisa untuk mencari jawaban dari setiap soal yang ditanyakan. Untuk mencari nilai mean, median, minimum dan maksimum, kami menggunakan function `describe` di python. Fungsi dari function `describe` pada python adalah untuk memberikan detail statistik mengenai variabel data frame yang memiliki nilai numerikal. Detail statistik yang bisa didapatkan adalah nilai count, rata-rata, standar deviasi, minimum, kuartil satu, kuartil dua, kuartil tiga, dan nilai maksimum. Untuk mencari nilai modus dari suatu dataset, kami menggunakan function `mode`. Function `mode` digunakan untuk mencari modus dari setiap variabel yang ada di data frame tersebut yang memiliki nilai numerikal.

Untuk mencari nilai-nilai outlier dari setiap variabel, kami menggunakan pendekatan melalui analisa boxplot. Agar visualisasi data terlihat rapi, kami membagi variabel-variabel yang ada menjadi dua kategori yaitu kategori `small` dan `large`. `Small` memiliki arti dataset yang memiliki value kurang dari 60.000. Sedangkan `large` memiliki arti dataset yang memiliki value lebih dari 60.000. Dengan menggunakan library dari `matplotlib`, boxplot dapat divisualisasikan dengan menggunakan function `df.boxplot()`. Kami menambahkan beberapa feature untuk memperjelas hasil dari visualisasi. Dengan memperbesar figure dari visualisasi tersebut menggunakan function `figsize`, maka akan memperbesar resolusinya dengan satuan inch. Selain itu, kami juga menggunakan function `rot` untuk merotasi keterangan dari setiap boxplot nya agar terlihat rapi di visualisasi tersebut. Kami juga menambahkan title pada grafik tersebut untuk memperjelas grafik yang digambarkan. Sama hal nya dengan `large value`, kami menggunakan function yang sama untuk memvisualisasinya.

Untuk mengetahui total outlier tiap variabelnya, kami menggunakan *statistical approach* dari boxplot. Kami mencari tahu nilai dari kuartil 1 dan



kuartil 3 dari setiap variabelnya dengan menggunakan function `quantile` dan memberikan value yang tepat dari setiap quantile. Kuartil satu merupakan data yang terletak pada posisi 25% dari keseluruhan data. Sedangkan kuartil 3 terletak pada posisi ke 75% dari keseluruhan data. Maka dari itu, function dari kuartil 1, kami menggunakan function `df.quantile(0.25)`. Sama halnya dengan kuartil 1. Lalu, dengan analisa statistik melalui boxplot, kami perlu mencari tahu nilai dari *interquartile range* (IQR) dengan cara mengurangi hasil dari kuartil 3 dengan kuartil satu. Outlier dengan menggunakan approach boxplot didapatkan apabila nilai dari data tersebut melebihi kuartil 3 ditambah dengan 1.5 dikali IQR atau kuartil 1 dikurangi 1.5 dikali IQR. Nilai-nilai tersebut merupakan outlier. Maka dari itu, kami menggunakan function `((df < (Q1 - 1.5 * IQR)) | (df > (Q3 + 1.5 * IQR))).sum()`. Function `sum` adalah untuk menjumlah ada berapa banyak angka yang termasuk dalam outlier. Function `sum` disini fungsinya mirip dengan function `count` pada umumnya. Kami menggunakan function `sum`, karena hasil output dari code yang kami extract sebelum `sum` akan memiliki output `true` atau `false`. `True` berarti 1 dan `false` berarti 0. Karena `true` berarti data tersebut merupakan nilai outlier, maka function `sum` digunakan untuk menghitung total outlier yang ada di dataframe tersebut.

Dari data yang diberikan, kami juga dapat mengetahui korelasi antara satu variabel dengan variabel lainnya. Untuk mencari tahu korelasi, kami menggunakan dua analisis yaitu yang pertama dengan melakukan analisis visualisasi heatmap dan yang kedua adalah dengan melakukan analisis korelasi menggunakan *pearson coefficient*. Dengan menggunakan analisa heatmap, kami dapat secara langsung mengetahui angka korelasi dari setiap variabel yang ada. Kami menggunakan library `matplotlib` dan `seaborn` dari python untuk memvisualisasikan heatmap tersebut. Function `plt.figure(figsize=(14,7))` kami gunakan untuk memperjelas hasil dari heatmap agar memiliki ukuran yang lebih besar dari default yang disediakan. Setelah itu, kami menggunakan function `sns.heatmap(data=df.corr(), annot = True)`. Maksud dari kode yang kami jelaskan adalah memvisualisasikan heatmap dengan menggunakan data yaitu dengan mencari *correlation coefficient* dari data yang ada. Function `df.corr()` kami gunakan untuk mencari koefisien korelasi. Lalu `annot True` berarti kami memunculkan angka korelasi pada

heatmap tersebut. Karena banyaknya data yang ada, kami menyimpulkan hubungan antara 2 variabel yang memiliki relasi yang kuat adalah variabel-variabel yang memiliki koefisien korelasi lebih dari 0.8. Selain itu, kami kategorikan kedalam korelasi lemah. Selain analisa heatmap, kami juga melakukan analisis statistik. Kami menggunakan analisis pearson correlation coefficient tanpa menggunakan visualisasi heatmap. Dari analisa kami, beberapa variabel memiliki korelasi yang berbeda dengan hasil dari heatmap. Namun, nilai dari kedua analisis tersebut tidaklah berbeda jauh. Maka dari itu, kami menggunakan analisa heatmap untuk menjawab pertanyaan mengenai analisis korelasi. Selain itu, kami juga melakukan pengecekan dengan melakukan visualisasi regplot. Regplot disini berarti regression plot. Data akan dibentuk berdasarkan sumbu x dan y yang diinput, kemudian akan terbentuk garis regresi secara otomatis dari data x dan y yang ada. Apabila garis regresi menyerupai  $x=y$ , maka dapat dikatakan kedua variabel tersebut memiliki korelasi positif. Apabila garis korelasi semakin landai, maka dapat dikatakan bahwa korelasi antara kedua variabel tersebut tidak cukup kuat.

Untuk melihat pertumbuhan covid 19 di Indonesia dan di Jakarta, kami melakukan analisis visualisasi dengan menggunakan line plot. Sebelum melakukan visualisasi, kami check data terlebih dahulu apakah ada kejanggalan. Setelah kami analisa dengan menggunakan function tail, kami melihat bahwa terdapat kejanggalan pada kolom ke 490. Di kolom ke 490, terlihat bahwa terdapat data dengan null value dan 0, maka dari itu, kami drop value di row ke 490. Setelah itu kami melakukan visualisasi dengan menggunakan function plot. Dari dua visualisasi yang kita temukan, terlihat bahwa kenaikan di Indonesia juga diikuti dengan kenaikan di Jakarta. Dari sini, dapat disimpulkan bahwa positif harian di Indonesia berbanding lurus dengan positif harian di Jakarta.

#### f. Conclusion and Recommendation

##### Kesimpulan

Dari dataset yang telah dianalisis, dapat diambil kesimpulan sebagai berikut. Penggunaan teknik analisis yang kita ambil adalah boxplot, scatterplot, dan heatmap. Data-data covid di Jakarta berkorelasi kuat terhadap data-data covid di Indonesia. Apabila jumlah peningkatan covid bertambah dari hari

sebelumnya di Jakarta, maka pertambahan di Indonesia akan lebih besar dari hari sebelumnya juga. Selain itu, kami juga melihat bahwa terdapat korelasi positif antara meninggal (Jakarta) dan sembuh (Jakarta). Disini dapat diartikan bahwa banyaknya orang yang meninggal berbanding lurus dengan banyaknya orang yang sembuh.

#### Saran

Dari data yang telah dianalisis, diharapkan peran serta masyarakat, guna mencegah berkembangnya Covid-19 di tengah-tengah masyarakat. Masyarakat perlu menjaga jarak dan melakukan protokol kesehatan. Untuk pihak pemerintah dapat lebih meningkatkan program pengetatan dan membubarkan massa yang berkumpul. Pemerintah juga perlu untuk mempercepat proses vaksinasi untuk memperkecil kemungkinan orang meninggal karena Covid-19.

### **BAB III : Penutup**

#### **3.1 Kesimpulan**

Dari pembahasan diatas, diketahui bahwa Covid-19 sangat cepat menyebar. Di Indonesia, kasus Covid-19 mencapai puncak-puncaknya pada bulan Desember sampai Januari dan Juni sampai Juli. Kasus Covid-19 di Jakarta berbanding lurus dengan kasus Covid-19 di Indonesia. Hal ini dibuktikan dengan beberapa korelasi yang kuat seperti Positif Harian (Jakarta) dan Positif Harian (Indonesia). Teknik-teknik yang kita lakukan adalah analisis boxplot, scatter plot, dan heatmap. Setelah melakukan analisis menyeluruh, masyarakat diminta untuk terus menjalankan protokol kesehatan dan menjaga kesehatan. Pemerintah juga diharapkan untuk mempercepat proses vaksinasi dan memperketat mobilisasi masyarakat.

## Daftar Pustaka

Layanan Darurat Covid-19. (2021). *Pantauan Covid-19 Hari Ini*.

<https://corona.jakarta.go.id/id>

Makki, S. (Juni, 2021). *Satgas Sebut Indonesia Masuk Gelombang Kedua Covid-19*. CNN

Indonesia.[https://www.cnnindonesia.com/nasional/20210630080610-20-](https://www.cnnindonesia.com/nasional/20210630080610-20-661120/satgas-sebut-indonesia-masuk-gelombang-kedua-covid-19)

[661120/satgas-sebut-indonesia-masuk-gelombang-kedua-covid-19](https://www.cnnindonesia.com/nasional/20210630080610-20-661120/satgas-sebut-indonesia-masuk-gelombang-kedua-covid-19)

WHO Indonesia (2020). *QA for public*. [https://www.who.int/indonesia/news/novel-](https://www.who.int/indonesia/news/novel-coronavirus/qa/qa-for-public)

[coronavirus/qa/qa-for-public](https://www.who.int/indonesia/news/novel-coronavirus/qa/qa-for-public)

## Lampiran

[Code Shenzhen Tech](#)