

Jawaban Non-Dataset

1. Jelaskan secara teori statistik mengenai *outlier* (pencilan), implikasinya dalam analisis data, serta bagaimana melakukan manajemen data terhadap kasus *outlier*.

Outlier dalam teori statistika adalah data yang berada jauh dari data-data lainnya (memiliki jarak yang *abnormal*). Pada umumnya faktor munculnya *outlier* bisa dikarenakan adanya kesalahan dalam pengumpulan data contohnya seperti kesalahan dalam *input* data ataupun adanya salah *response* dari partisipan. Dengan adanya *outlier* dapat memberikan dampak yang besar pada proposi data sehingga nantinya hasil proses pada data akan memberikan intepetasi yang menyimpang.

Cara memanajemen data ketika munculnya sebuah *outlier* ada banyak ragamnya namun semua tergantung pada jenis data dan penelitian yang dibuat. Pendekatan yang berbeda akan juga akan memberikan analisis yang berbeda, oleh karena itu pendekatan yang sesuai sangat dibutuhkan dalam manajemen data. Secara umum ada 3 metode yang dapat dilakukan dalam manajemen data ketika ada *outlier* yakni :

A. Trimming

Dengan metode ini artinya kita akan 'mengeluarkan' atau meng-*trim outlier* dari *data set* untuk dianalisa. Namun dari metode ini akan memberikan analisa yang kurang memadai sebab metode ini pendekatan yang kurang cocok bagi *outlier*.

B. Winsorization

Dengan metode ini artinya kita mengganti *outlier* dari *data set* dengan nilai persentil tertentu dari masing-masing ujung. Dengan metode ini memperbolehkan kita untuk memodifikasikan data tanpa menghapus *outlier*.

C. Robust Estimation Method

Pada metode ini kita akan menggunakan nilai estimasi dari *outliers* menggunakan *Robust Techniques*.

2. Jelaskan konsep dan prinsip korelasi, lalu sebisa mungkin kaitkan dengan dasar-dasar statistik serta implikasinya terhadap konsep/teori statistik lain

Korelasi adalah teknik dalam bidang statistika yang digunakan untuk mencari ada atau tidaknya hubungan antara dua variabel atau lebih yang sifatnya kuantitatif. Variabel-variabel tersebut dikatakan memiliki korelasi apabila grafik variabel mengalami arah perubahan yang sama.

Statistika dibagi menjadi 2 jenis yakni : statistika deskriptif dan statistika inferensial. Statistika deskriptif adalah jenis statistika yang digunakan untuk merangkum dan merepresentasikan data yang diterima. Statistika inferensial adalah jenis statistika yang digunakan untuk mencari prediksi atau menggambarkan kondisi populasi berdasarkan data *sample* yang diterima. Dalam statistika inferensial ada 2 kegiatan yang dilakukan yakni uji hipotesis dan estimasi. Dalam pengujian hipotesis pada umumnya akan menggunakan korelasi untuk melihat apakah adanya keterkaitan dari *sample* yang diteliti.

3. Sebutkan teori dasar *machine learning* yang kalian ketahui, lalu jelaskan dalam bahasa sederhana mengenai teori tersebut dan implikasinya.

a. Supervised Learning

Supervised Learning adalah proses pembelajaran dengan bantuan label pada data yang di-*input*. Pada *Supervised Learning* mesin akan diberi *input* data yang telah diberi label untuk dipelajari sehingga selanjutnya mesin dapat memberikan prediksi sesuai dengan data yang diberikan. Contoh simpelnya seperti ketika guru mengajarkan kita di sekolah. Guru diibaratkan sebagai *supervisor* yang memberikan data yang telah diberi label sedangkan murid adalah mesin yang diajari. *Supervised Learning* merupakan jenis teori *machine learning* yang sering digunakan dalam dunia AI untuk *training* data gambar. Pada *training* data gambar umumnya akan diberikan gambar dengan label kepada mesin (Contoh klasifikasi hewan seperti mamalia atau reptil) untuk dipelajari sehingga ketika lain waktu mesin diberikan gambar dapat memberikan prediksi yang tepat.

b. Unsupervised Learning

Unsupervised Learning adalah proses pembelajaran tanpa bantuan *labeling* data yang di-*input*. Pada *Unsupervised Learning* mesin akan diberikan *input* data tanpa *labelling* untuk dipelajari kemiripan antar data sehingga selanjutnya mesin dapat memberikan prediksi. Jika *Supervised Learning* diibaratkan seperti belajar di sekolah yang

dibimbing oleh guru maka *Unsupervised Learning* adalah belajar tanpa guru. *Unsupervised Learning* dapat diibaratkan seperti ketika bayi mempelajari sekelilingnya dengan memperhatikan.

Pada umumnya pada *Unsupervised Learning* dilakukan dalam 2 langkah untuk mendapatkan hasil yakni prediksi dan belajar. Contoh pada kasus 3 bunga iris yang memiliki kelopak bunga yang berbeda. Pada *Unsupervised Learning* awalnya akan dilakukan prediksi terlebih dahulu lalu mesin akan melakukan pembelajaran dari hasil prediksi untuk memperbaiki dirinya sendiri. Langkah ini akan terus berlanjut hingga mesin menemukan prediksi yang terbaik.

c. *Reinforcement Learning*

Proses pembelajaran dengan menggunakan *feedback* yang diberikan oleh lingkungan mesin. *Reinforcement Learning* juga dapat dikatakan sebagai pembelajaran '*Learning by doing*' teori ini diibaratkan seperti ketika bayi belajar berjalan pertama kalinya. Apabila *Supervised Learning* merupakan pembelajaran yang digunakan seperti guru - murid, *Unsupervised Learning* menggunakan *pattern* atau pola untuk menemukan jawaban maka *Reinforcement Learning* digunakan untuk melatih mesin ketika *skill* yang kita miliki juga belum memadai. Metode seperti ini akan menghasilkan *reward* berupa poin jika model yang diberikan makin baik dan memberikan poin minus apabila model yang diberikan menurun. Contoh aplikasi dari metode ini sering digunakan pada mesin pencari.

4. Menggunakan bahasa kalian sendiri, jelaskan kaitan antara *artificial intelligence*, *machine learning*, dan *deep learning*.

Artificial Intelligence atau kecerdasan buatan adalah ilmu dalam bidang komputer yang mempelajari dan mengembangkan algoritma agar mesin atau komputer dapat mengambil keputusan dan bereaksi layaknya manusia.

Machine Learning merupakan bagian dari AI dimana pada *Machine Learning* berfokus pada pembelajaran mesin menggunakan algoritma yang dibuat. Pada umumnya, *Machine Learning* digunakan agar komputer atau mesin dapat melakukan pembelajaran dengan sendirinya tanpa diberi instruksi dan dapat meningkatkan tingkat akurasi pada hasil prediksi. Ketika komputer atau mesin berhasil belajar pada segala situasi maka *engineer* tidak perlu lagi memberikan solusi atau jawaban pada mesin.

Deep Learning merupakan bagian dari *Machine Learning* yang terinspirasi dari sel-sel di otak manusia yang disebut neuron. Pada *Deep Learning* sistem dapat meniru kerja otak manusia dengan menggunakan ANN (*Artificial Neural Network*) sehingga dapat menyelesaikan permasalahan yang diberikan. *Deep learning* sendiri sering digunakan pada *Image Recognition* dan *Speech Recognition*.

5. Apakah yang kalian ketahui mengenai interpretasi data? Bagaimana signifikansi dan tantangannya? Bagaimana kaitan interpretasi data dengan data *story telling* dan *decision making*?

Interpretasi data adalah proses penyederhanaan data agar data dapat divisualisasikan menjadi grafik atau bentuk yang lebih mudah dipahami oleh orang awam. Jika kita ibaratkan dalam 3 langkah proses *input* - proses - *output*, data merupakan langkah *input*. Hal ini dikarenakan data merupakan informasi yang masih mentah yang harus kita olah agar mendapatkan *insight* atau informasi. Oleh karena itu dibutuhkanlah interpretasi data agar data mentah tersebut dapat diolah menjadi informasi yang berguna.

Namun tentu saja dalam melakukan interpretasi data banyak tantangan yang harus dihadapi salah satunya adalah ketika data set yang diberikan memiliki grafik yang distribusinya *skewed* ataupun memiliki banyak *outlier*. Hal ini menjadi tantangan bagi data engineer untuk mengolah data tersebut agar dapat diinterpretasikan sehingga menjadi mudah untuk dicerna.

Hasil dari proses interpretasi data akan menjadi informasi yang berguna bagi perusahaan atau organisasi yang berkaitan. Hal ini dikarenakan dengan data yang sudah divisualisasikan akan menjadi mudah untuk menarik kesimpulan atau informasi yang berguna. Contohnya kita melakukan kuesioner kepuasan pelanggan terhadap kualitas pelayanan di restoran tertentu dari data yang terkumpul dapat diolah menjadi informasi yang penting restoran.

Data *storytelling* adalah sebuah metodologi yang dilakukan untuk memberikan informasi bagi *audience* dengan menggunakan data. Tujuan dari data *storytelling* adalah agar *user* dapat mengerti arti dari data yang disajikan. Metodologi ini biasanya sering digunakan pada perusahaan untuk memberikan informasi yang kemudian berguna dalam *decision making* dalam perusahaan.

Kaitan interpretasi data dengan data *story telling* adalah tanpa interpretasi data yang baik dan benar maka data *story telling* hampir sangat susah untuk dilakukan. Dengan

data yang mentah akan susah dimengerti oleh orang-orang yang *non-expert* dalam bidang data.

Kaitan interpretasi data dengan decision making adalah dengan visualisasi data yang simpel dan mudah dipahami kemudian dilanjutkan dengan *story telling* yang baik maka akan membantu *user / stakeholder* untuk mendapatkan informasi dibalik data yang disajikan. Data memiliki makna dan arti yang apabila kita olah dengan baik dapat memberikan *insight* yang baik bagi perusahaan sehingga informasi yang didapatkan ini dapat membantu *user* untuk mengambil keputusan yang tepat di lain waktu.

Contoh data penjualan boba. Di toko A menjual 2 jenis boba yaitu boba putih dan hitam. Dari data yang dikumpulkan terlihat bahwa penjualan boba hitam itu lebih tinggi dibandingkan boba putih. Dari data tersebut dilakukan interpretasi data menggunakan grafik agar mudah dipahami oleh penjual dan dilengkapi dengan data *story telling* dari grafik yang disajikan agar penjual memahami maksud dari grafik. Dari grafik dan hasil *story telling* penjual dapat menarik kesimpulan bahwa pembeli lebih menggemari boba hitam sehingga kedepannya penjual dapat memerlukan stock boba lebih banyak di boba hitam dibandingkan boba putih.

Laporan Penelitian

Korelasi Antara Jumlah Rapid Test Antigen, Antibodi & Vaksinasi terhadap Jumlah Kasus
Positif COVID – 19 di DKI Jakarta

Disusun Oleh:

AI-SAT

Ketua:

Steven Putra Tanoto

Anggota:

Kevin Yuslianto

Erlina

Latar Belakang

COVID – 19 adalah penyakit gangguan saluran pernapasan yang disebabkan oleh turunan Corona virus jenis baru. Karena tingkat penyebaran yang tinggi dan munculnya jenis mutasi baru pandemi COVID – 19 berhasil menaklukkan dunia di tahun 2020 hingga sekarang. Kasus COVID – 19 pertama kali ditemukan di Wu Han , Cina pada Desember 2019. Dalam beberapa bulan saja COVID – 19 berhasil menyebar ke berbagai benua termasuk benua Eropa, Afrika, Amerika dan Australia. Hal ini menyebabkan banyak kerugian dalam berbagai bidang oleh negara-negara yang terinfeksi COVID – 19 sehingga menyebabkan peningkatan pengangguran dan perlambatan ekonomi. Di Indonesia sendiri mengalami peningkatan pengangguran sebanyak 9,7 Juta orang selama masa pandemi COVID – 19 (Zamani, 2021).

Data adalah sekumpulan fakta yang dikumpulkan dari kejadian-kejadian di dunia sehari-hari. Karena data hidup di dunia keseharian manusia oleh karena itu data sangat dekat dengan manusia. Kita dapat menemukan data dimana saja di pasar, di sekolah hingga di internet. Hal inilah yang menyebabkan data menjadi kekuatan terbaru di era sekarang. Kekuatan yang dimaksud adalah melalui pengolahan data kita dapat mencari solusi dari berbagai masalah yang sebelumnya mungkin tidak dapat dipecahkan oleh metode lain. Melalui pengolahan data yang baik dan benar dapat menghasilkan informasi yang berguna. Pengolahan data seperti *data science* terbukti telah membantu dalam memberikan *insight* dan solusi dalam menyelesaikan masalah baik dalam pemerintahan maupun dalam bisnis.

Contoh salah satu solusi yang dapat dipecahkan melalui pengolahan data adalah pengendalian COVID – 19. Dengan pengolahan data yang baik dan benar kita dapat mendapatkan solusi dalam pengendalian COVID – 19 dan apa yang perlu diperbaiki dari sistem yang sudah ada sehingga dapat memaksimalkan sistem atau memperbaiki apabila terjadi kekurangan. Contoh melalui jumlah *Rapid Test Antigen* dan jumlah orang yang terinfeksi COVID – 19 kita dapat mencari relasi antara kedua factor apakah dengan meningkatnya *Rapid Test Antigen* dapat memengaruhi jumlah orang yang terinfeksi. Dengan begitu akan mempermudah pemerintah dalam *decision-making* saat pembuatan peraturan untuk mengendalikan penyebaran COVID – 19.

Jawaban Soal

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

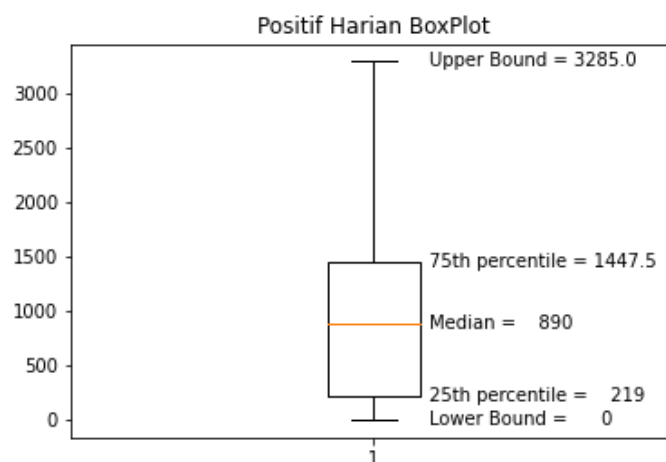
Mean	1445.359841
Median	890
Modus	0

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

Min	0
Max	14619

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

Disini kami menggunakan variabel yaitu Positif Harian, dimana kami mencari nilai *outlier* dari variabel positif harian pada dataset.



Gambar 1 Positif Harian Box Plot

Dengan menggunakan metode *Interquartile Range* (IQR) kita dapat mendeteksi *outlier* yang ada pada data dengan menggunakan rumus:

Q1 = 25 Persentil

Q2 = Median / 50 Persentil

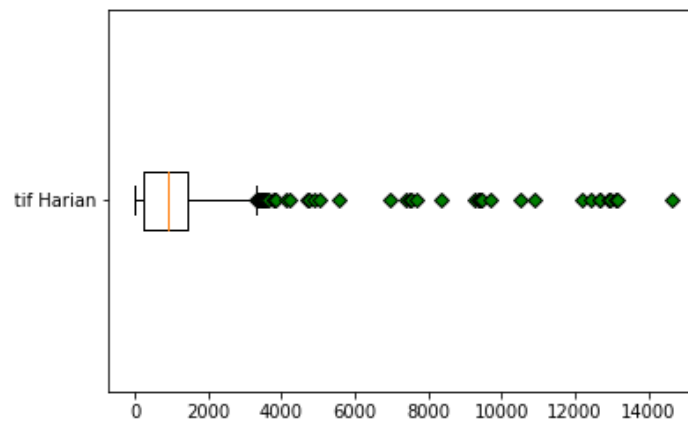
Q3 = 75 Persentil

IQR = Q3 - Q1

Lower bound = Q1 - (1.5 * IQR)

Upper bound = Q3 + (1.5 * IQR)

Sehingga *outlier* yang kami tentukan adalah yang lebih kecil dari *lower bound* dan lebih besar dari *upper bound*.



Gambar 2 Outlier Boxplot

Jumlah *Outlier*: 42

3309	3340	3362	3395
3437	3448	3474	3476
3491	3512	3536	3567
3614	3632	3786	3792
3810	4144	4213	4693
4737	4895	5014	5582
6934	7379	7505	7541
7680	8348	9271	9366
9394	9399	9439	9702

10485	10903	12182	12415
12667	12691	12920	12974
13112	13133	14619	

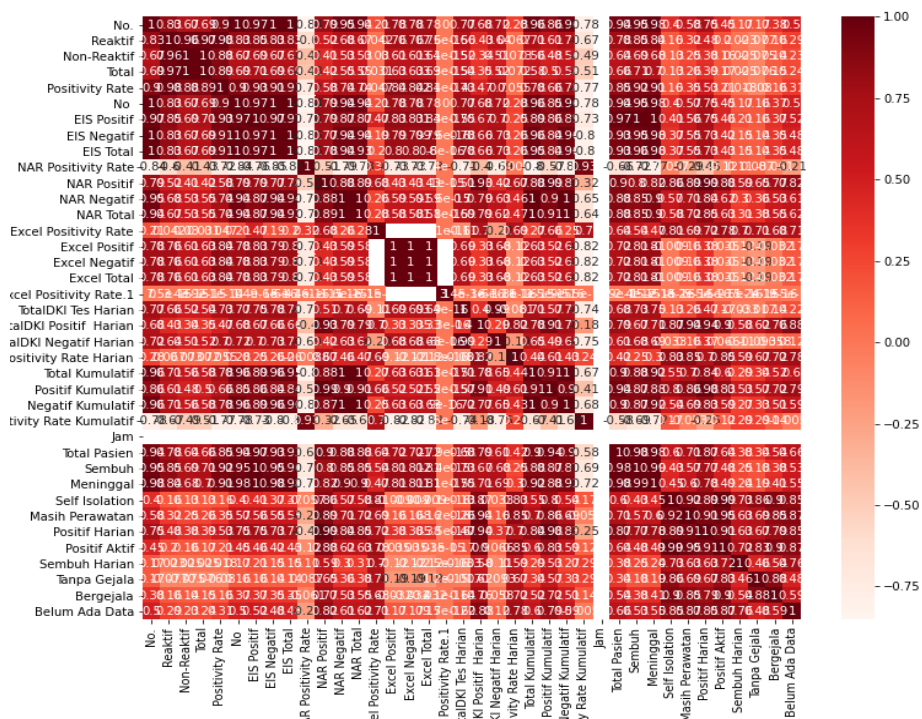
4. Dari dataset yang disediakan, usulkan dua buah variabel dan berikan analisis korelasi antara kedua variabel tersebut. Jelaskan apa kesimpulan yang dapat diambil berdasarkan analisis kalian.

Korelasi adalah sebuah sebutan yang biasanya digunakan untuk merepresentasikan perhitungan statistik dari hubungan *linear* antara dua variabel. Korelasi juga bisa digunakan untuk mengukur dependensi dari kedua variabel. Ketika variabel yang tersedia banyak, maka diperlukan suatu metode untuk mendeteksi korelasi yang paling kuat, dengan cara menghitung semua nilai korelasi dan menyimpan nilai tersebut, hal ini bisa kita capai dengan menggunakan matriks data struktur, yang biasanya disebut *correlation matrix*.

Ada 2 cara untuk mendeteksi dependensi korelasi antar 2 variabel, yaitu:

- Correlation coefficient / Pearson correlation coefficient* yaitu sebuah metode dimana mengukur perbedaan nilai dari kedua variabel yang saling berkaitan.
- Spearman correlation coefficient* yaitu sebuah metode untuk mengukur korelasi dengan cara mengukur bagaimana variabel tersebut naik dan turun bergantung dengan variabel lain.

Disini kita menggunakan *pandas* untuk mendeteksi korelasi yang dimana merupakan perhitungan korelasi matriks $[-1,1]$ dimana 1 menunjukkan tingkat korelasi yang tinggi, dan -1 menunjukkan tingkat korelasi paling rendah. Dengan menggunakan `df.corr()` kita dapat mendeteksi korelasi yang dimana menggunakan metode *Pearson correlation coefficient*. Kita juga menggunakan *Seaborn* untuk melakukan visualisasi korelasi dengan menggunakan *heatmap*.



Gambar 3 Heatmap Korelasi

Kami memilih sebuah variabel yaitu positif harian untuk dijadikan acuan korelasi karena disini kami ingin mendeteksi apa yang menyebabkan angka positif harian bisa naik dengan cukup signifikan.

Dari data korelasi yang sudah terkumpul-kan, kita mengambil nilai korelasi yang lebih dari 0.95 untuk dianalisis.

NAR Positif	0.994080
Positif Kumulatif	0.980707
Positif Harian	1.000000

	NAR Positif	Positif Kumulatif
NAR Positif	1	0.992677
Positif Kumulatif	0.992677	1

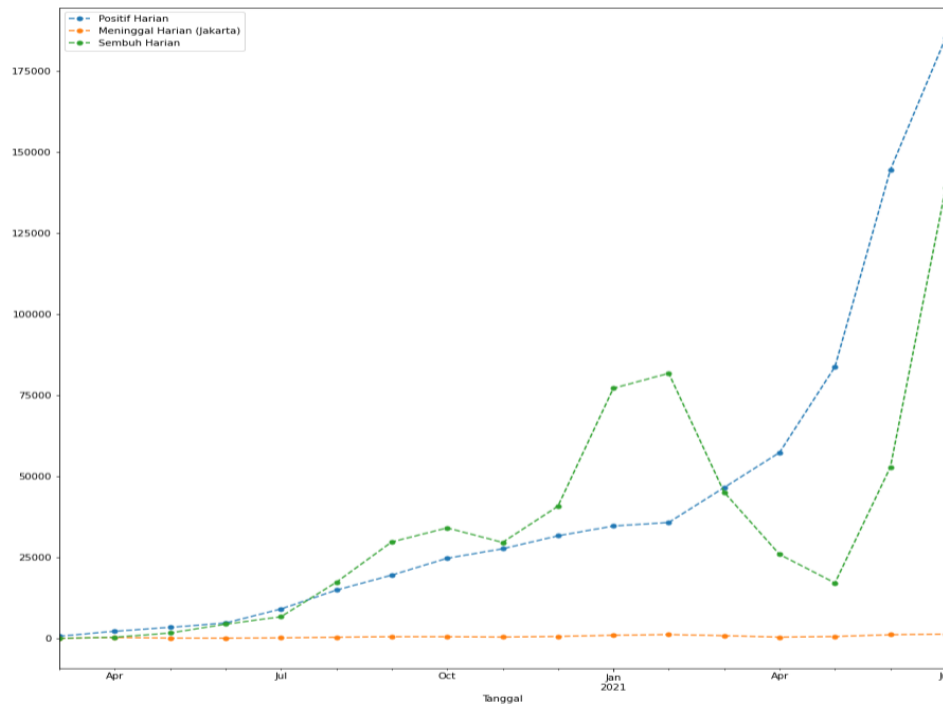
Dikarenakan positif korelatif dan NAR positif memiliki korelasi yang cukup tinggi, maka saya akan memilih salah satu untuk dijadikan variabel kedua untuk korelasi yaitu Positif Kumulatif yang berasal dari *dataset* untuk sheet RDTAg.

Sehingga berdasarkan analisa di atas maka kami menyimpulkan bahwa meningkatnya kasus positif harian dikarenakan meningkatnya juga testing antigen yang dilakukan. Sehingga pemerintah DKI Jakarta sudah melakukan *testing* dan *tracing* dengan baik dikarenakan sesuai dengan target yang diinginkan.

Hasil Analisis Tambahan

a. Problem Statement

Data yang diberikan merupakan data COVID-19 DKI-Jakarta dan Indonesia secara keseluruhan, dimana data tersebut berisi data positif, sembuh, meninggal, vaksin, pemakaman, tes antigen, tes antibodi. Problem yang ingin diselesaikan adalah mendeteksi penyebab utama data positif terus bertambah.



Gambar 4 Grafik Positif, Meninggal, Sembuh Harian Jakarta

Dari grafik diatas kita bisa lihat bahwa tingkat sembuh harian sempat lebih tinggi dari nilai positif harian pada bulan Agustus 2020 - Maret 2021 dan menurun drastis ketika ada-nya peningkatan kasus yang signifikan di April 2021 - Mei 2021. Menurut kami ini berkaitan dengan masuknya varian-varian baru yang masuk ke Indonesia sehingga menyebabkan kasus positif yang makin meningkat dan *recovery rate* varian baru yang cukup rendah.

b. Hipotesis

Kelompok kami membuat 2 hipotesis untuk dianalisis yaitu:

- H0: Angka positif terus meningkat dikarenakan jumlah testing RDTAg dan RDTAb yang terus meningkat.

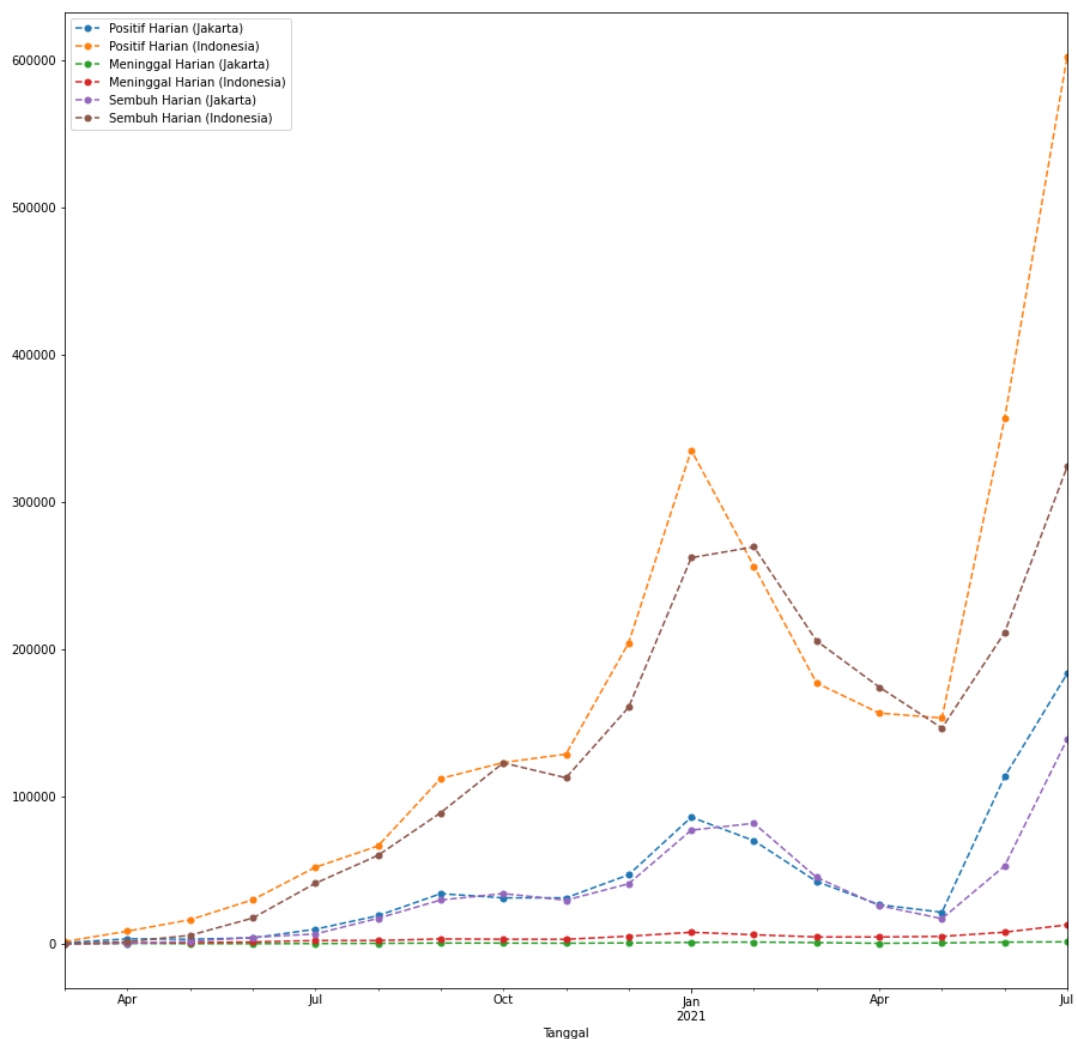
H1: Angka positif terus meningkat bukan dikarenakan jumlah testing RDTAg dan RDTAb yang terus meningkat.

ii. H0: Vaksin berpengaruh dalam menurunkan angka positif COVID-19.

H1: Vaksin tidak berpengaruh dalam menurunkan angka positif COVID-19.

c. EDA (*Exploratory Data Analysis*)

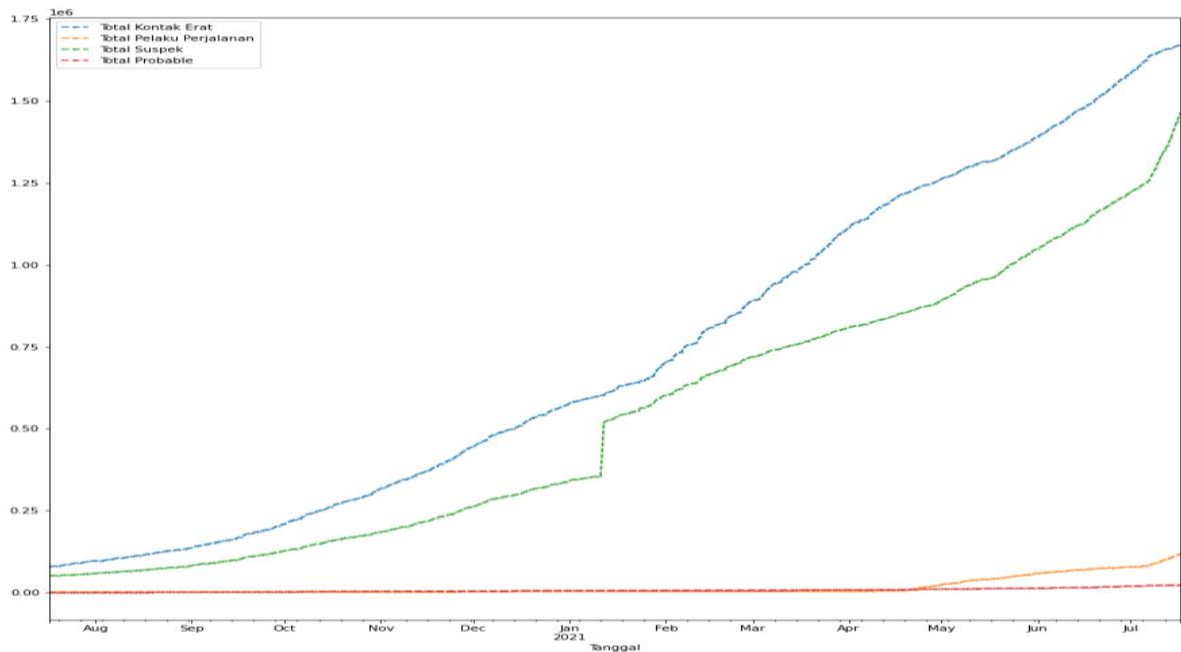
Sebelum memulai membuktikan hipotesis mari kita lihat data positif, sembuh, meninggal, total pasien di Indonesia dan DKI Jakarta, agar kita lebih memahami analisa yang akan dilakukan:



Gambar 5 Grafik Total Pasien, Sembuh, Positif, Meninggal Indonesia dan Jakarta

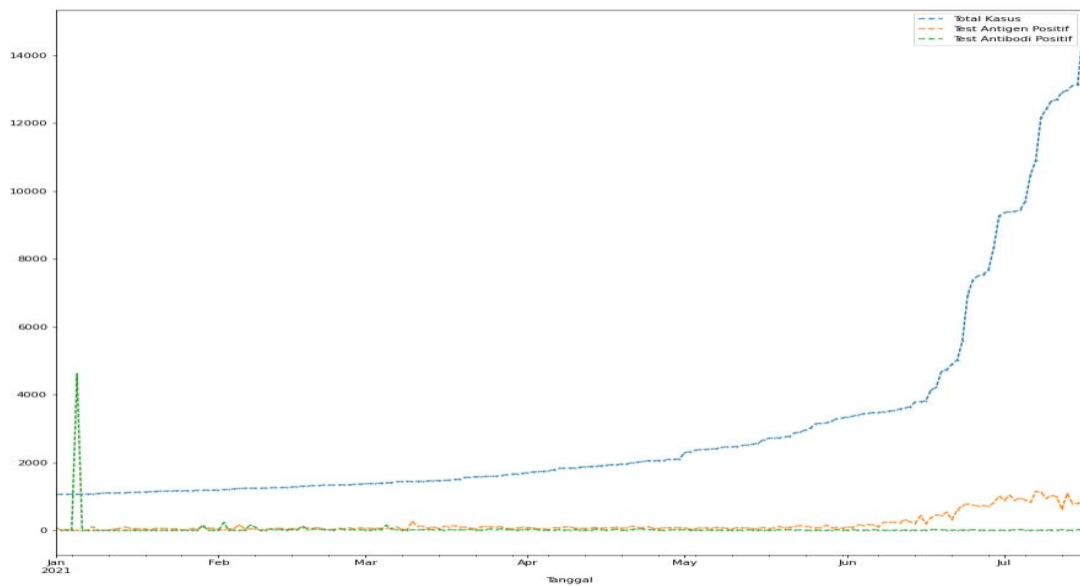
Kita bisa lihat dari grafik diatas bahwa Jakarta memiliki pola peningkatan kasus positif, sembuh, meninggal yang hampir mirip dengan kasus Nasional. Dari analisis

kami DKI Jakarta bahkan memberikan 3% kontribusi kasus positif terhadap kasus Nasional pada bulan Juli 2021. Diimbangi dengan kontribusi sebesar 2,7% kontribusi kasus sembuh terhadap kasus Nasional pada bulan Juli 2021. Sehingga disini bisa kita simpulkan bahwa kasus positif Nasional bukanlah penyebab meningkatnya kasus positif DKI Jakarta karena kenaikan antara 2 variabel tersebut linear dan hampir sama polanya.



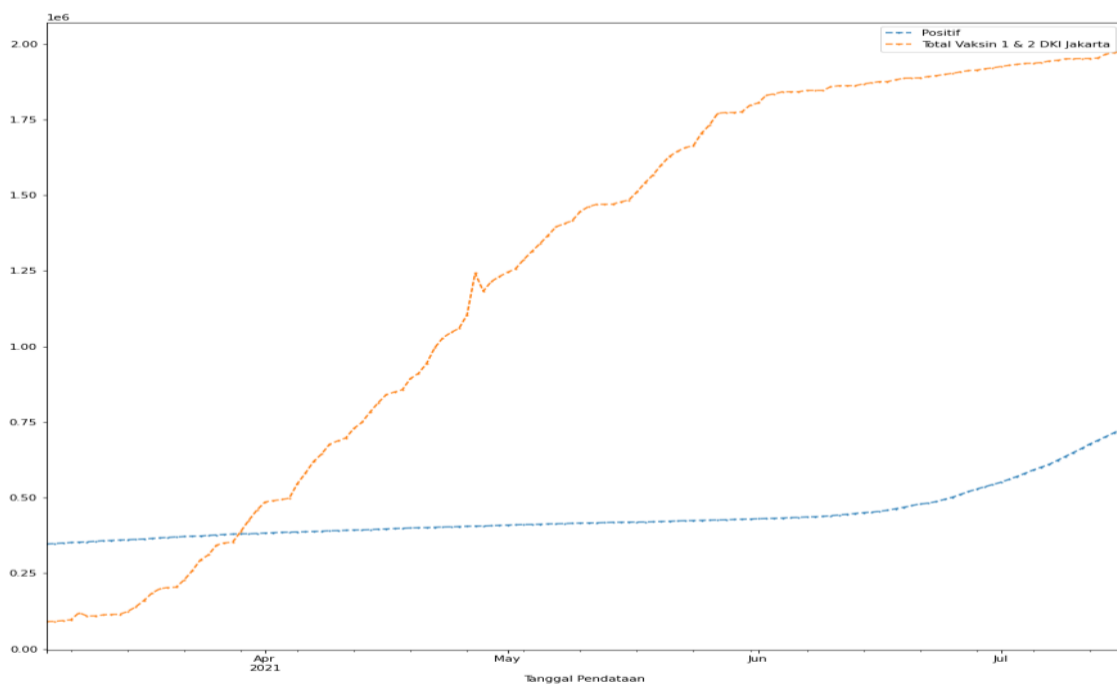
Gambar 6 Grafik Kontak Erat, Pelaku Perjalanan, Suspek, Probable

Dari gambar tersebut bisa dilihat bahwa mayoritas kasus positif berasal dari kontak erat dan suspek lingkungan sekitar, disini bisa dilihat bahwa pelaku perjalanan dan probable memberikan nilai yang rendah karena sudah adanya pencegahan diawal seperti *rapid test* dan PCR ketika ingin melakukan perjalanan.



Gambar 7 Grafik Total Kasus & Antigen & Antibodi Harian

Dari grafik diatas menunjukan bahwa meningkatnya kasus positif harian di DKI Jakarta tidak terpengaruhi dari jumlah *testing* RDTAg (*Antigen*) & RDTAb (*Antibodi*). Sehingga H_0 akan ditolak dan H_1 akan diterima untuk hipotesis pertama.



Gambar 8 Grafik Vaksinasi Kedua DKI Terhadap Positif Jakarta

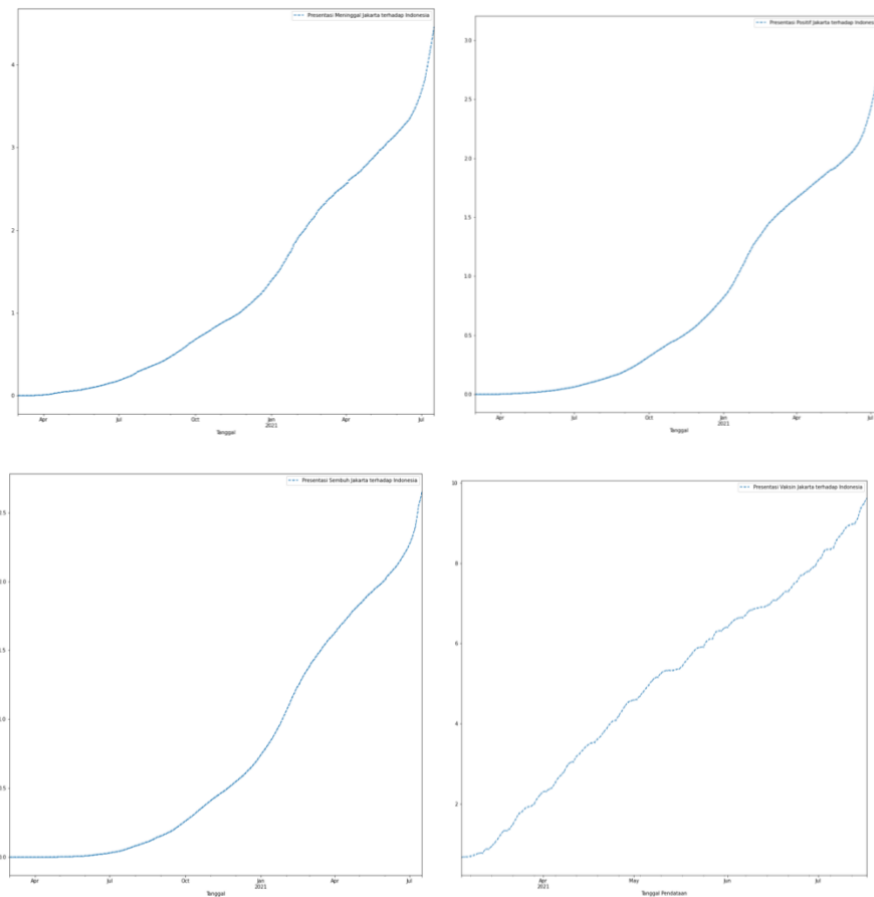
Dari grafik di atas menunjukkan bahwa meningkatnya angka vaksinasi tidak menurunkan angka positif harian DKI Jakarta sehingga H0 akan ditolak dan H1 akan diterima untuk hipotesis kedua.

d. Initial Findings

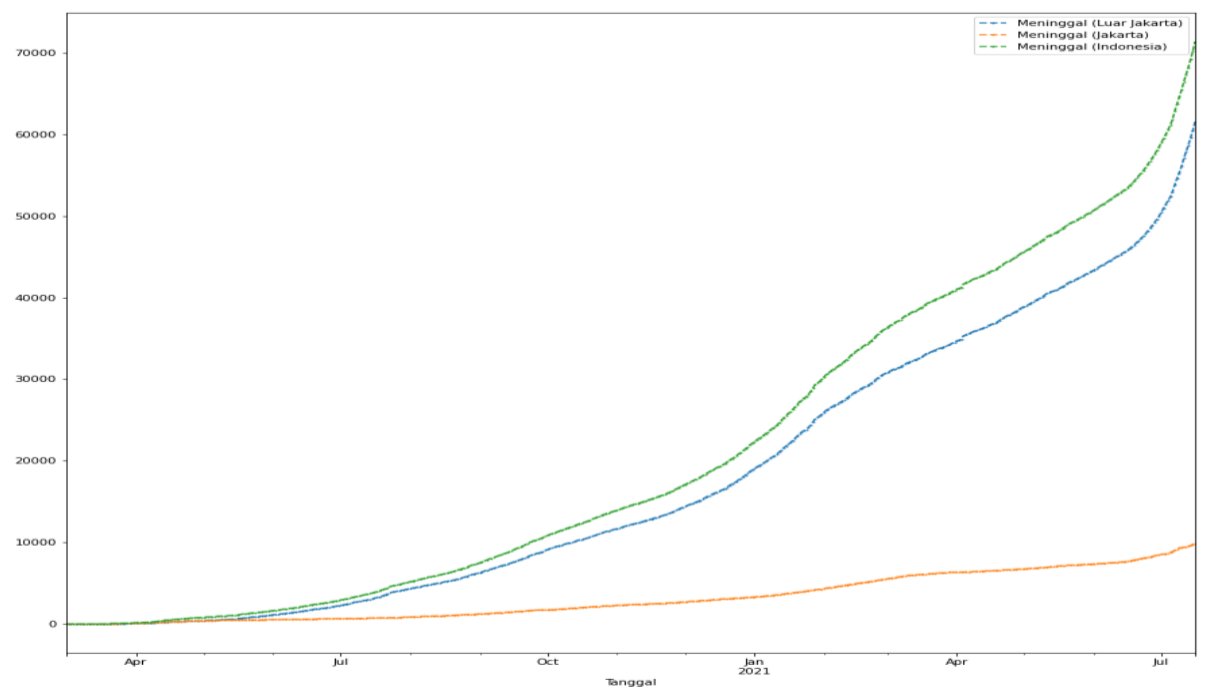
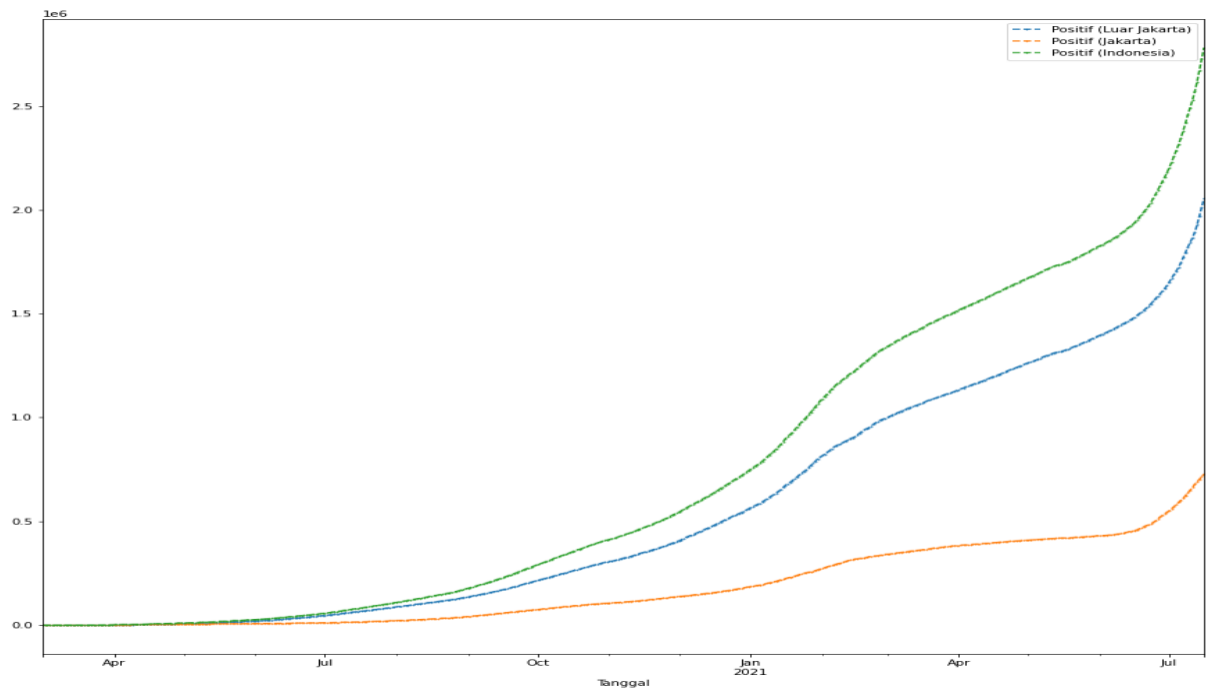
Setelah melakukan pembuktian hipotesis maka disimpulkan bahwa dengan meningkatnya kasus positif DKI Jakarta tidak serta merta dikarenakan angka testing yang makin meningkat dikarenakan banyaknya kasus positif bukanlah berasal dari hasil testing, kami menafsirkan bahwa banyak kasus positif tersebut berasal dari kontak dekat dengan orang yang terpapar oleh COVID19. Melalui analisis ini kami juga menemukan bahwa program vaksinasi yang dilakukan oleh pemerintah belum sepenuhnya bisa memberikan *herd immunity* di masyarakat, dikarenakan dari data yang terlihat bahwa meningkatnya kasus positif tidak terpengaruh oleh vaksinasi. Sehingga diperlukan analisa lebih lanjut dengan meneliti persebaran vaksinasi, penyebab penyebaran secara cepat apakah dikarenakan adanya isolasi mandiri yang menyebabkan keluarga atau kontak terdekat menjadi positif COVID19.

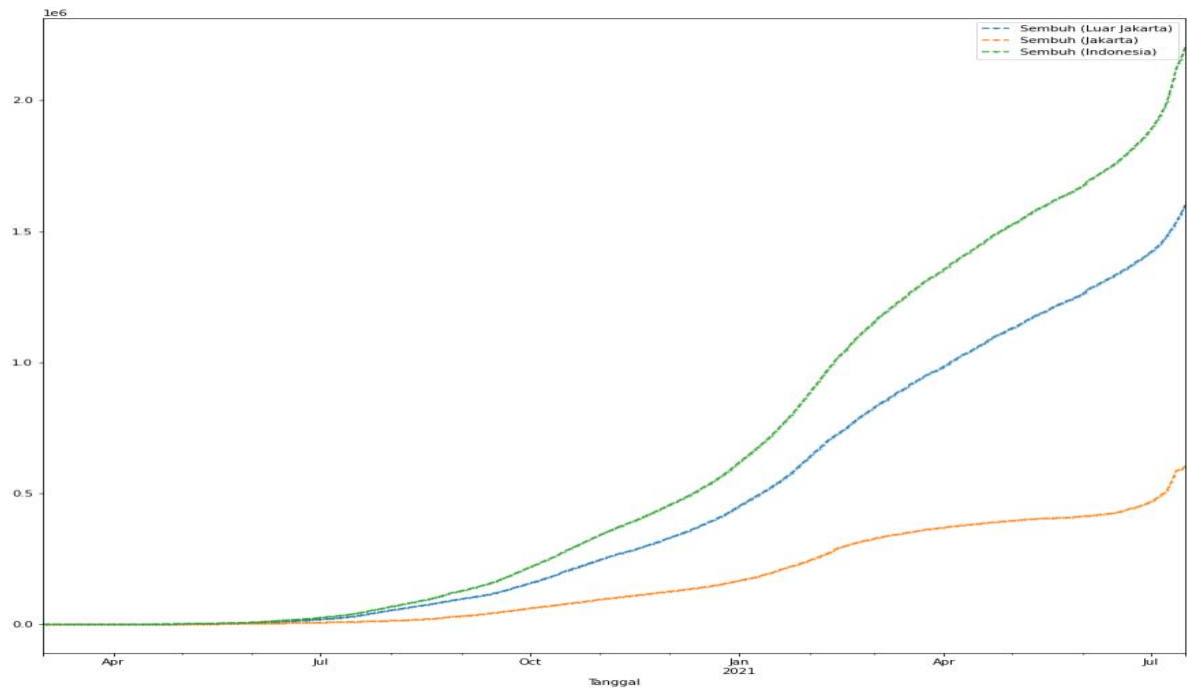
e. Deep Dive Analysis

Kita akan menganalisis grafik persentase positif, sembuh, meninggal dan vaksinasi DKI Jakarta terhadap Indonesia.

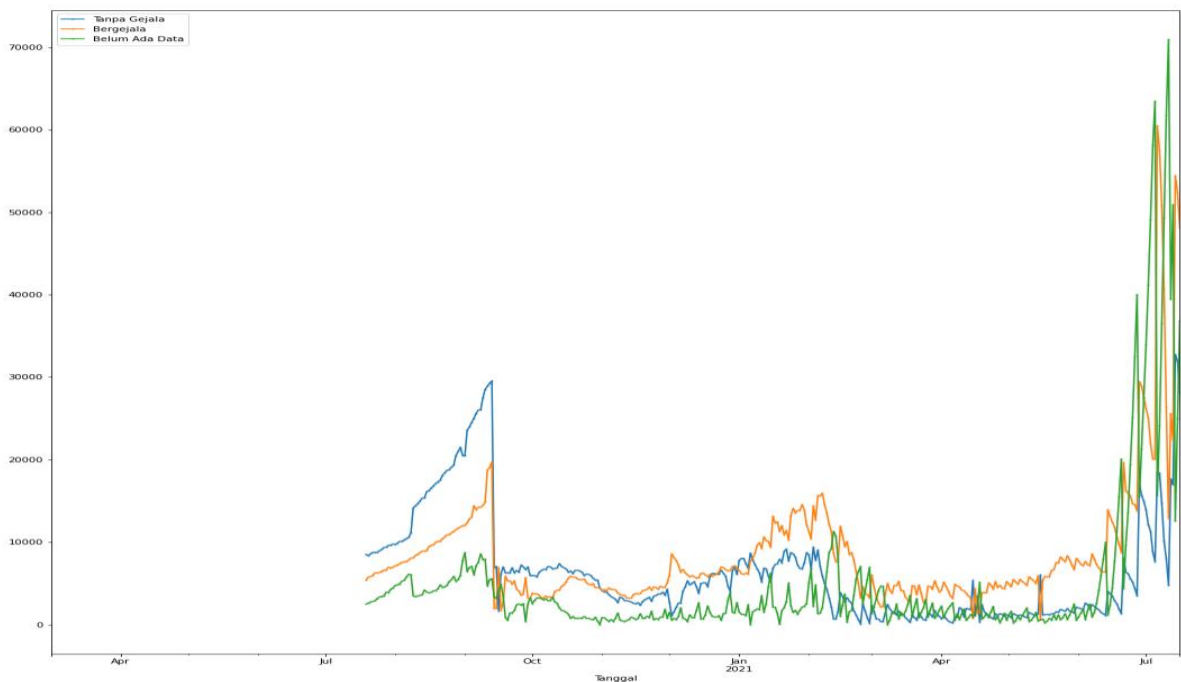


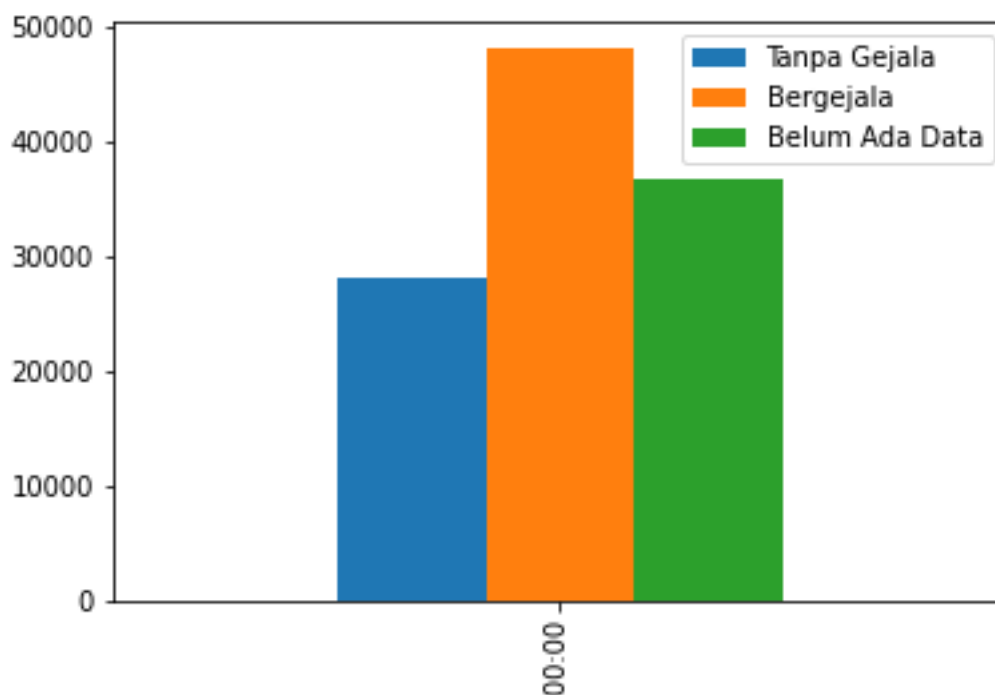
Dari keempat grafik diatas bisa dilihat bahwa DKI Jakarta memberikan kontribusi 4% terhadap kasus meninggal Nasional, 3% terhadap kasus positif Nasional, 2,5% terhadap kasus sembuh Nasional, 10% terhadap vaksinasi Nasional. Dari hasil tersebut kita bisa menyimpulkan bahwa kasus meninggal dan Positif di DKI Jakarta sangat lah memprihatinkan dikarenakan mencapai 4% dan 3% dari kasus Nasional, dengan angka vaksinasi yang tinggi harusnya DKI Jakarta bisa menekan kasus meninggal dan positif karena warga DKI Jakarta sudah 10% mendapatkan vaksinasi hingga vaksinasi ke-2.





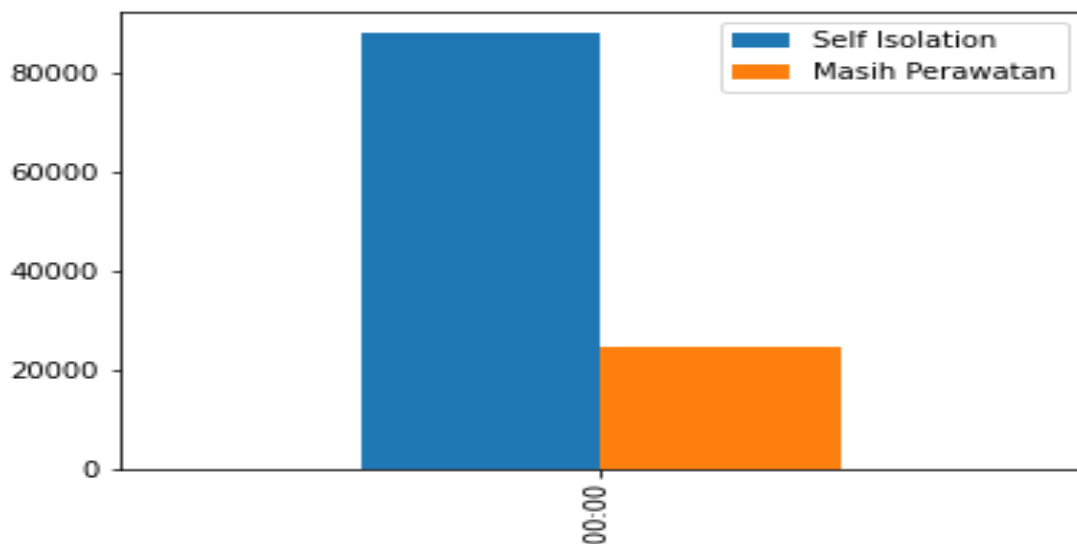
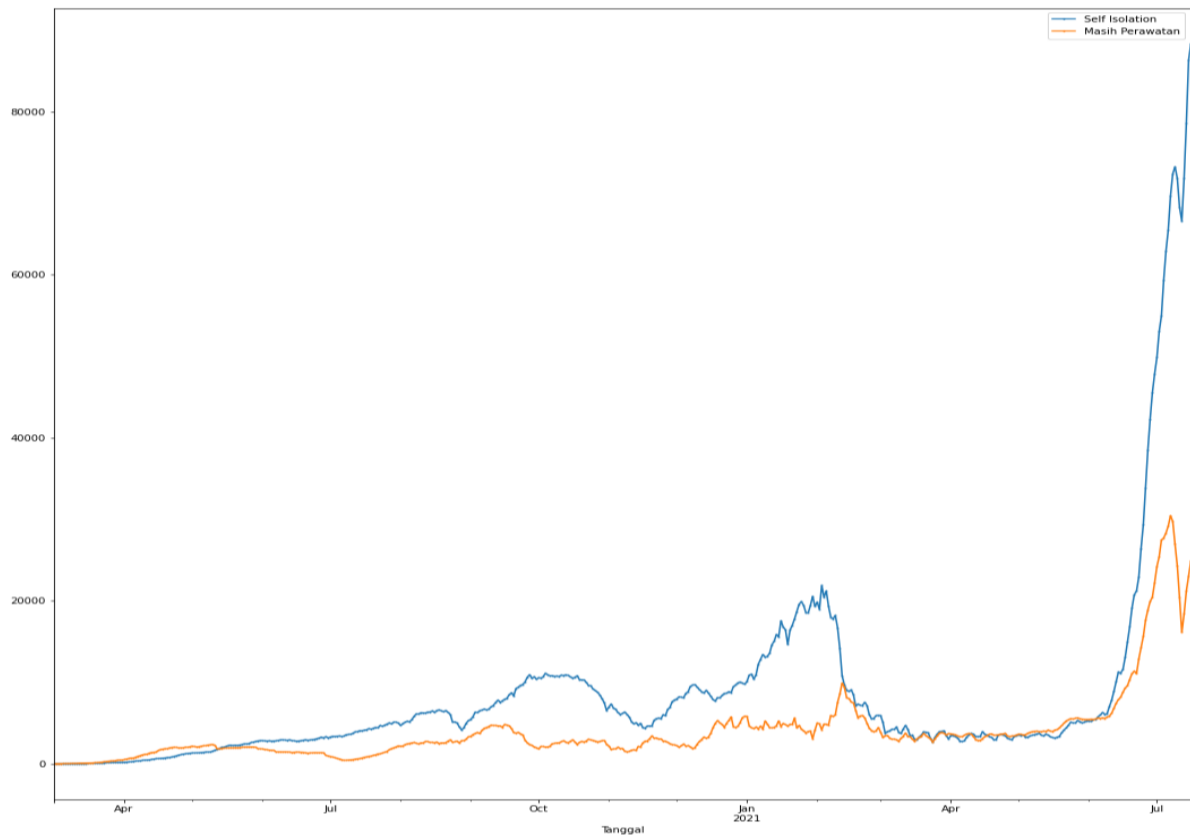
Dari keempat grafik diatas bisa dilihat bahwa terbukti bahwa jumlah meningkat, kasus positif, sembuh, vaksinasi Jakarta memberi dampak yang cukup besar juga terhadap data Nasional sehingga jika kita ingin menekan angka penyebaran COVID19 kita harus memulainya dari DKI Jakarta, dikarenakan *traffic* bisnis semua berpusat di DKI Jakarta.





Gambar 9 Perbandingan Tanpa Gejala, Bergejala, dan Belum Ada Data

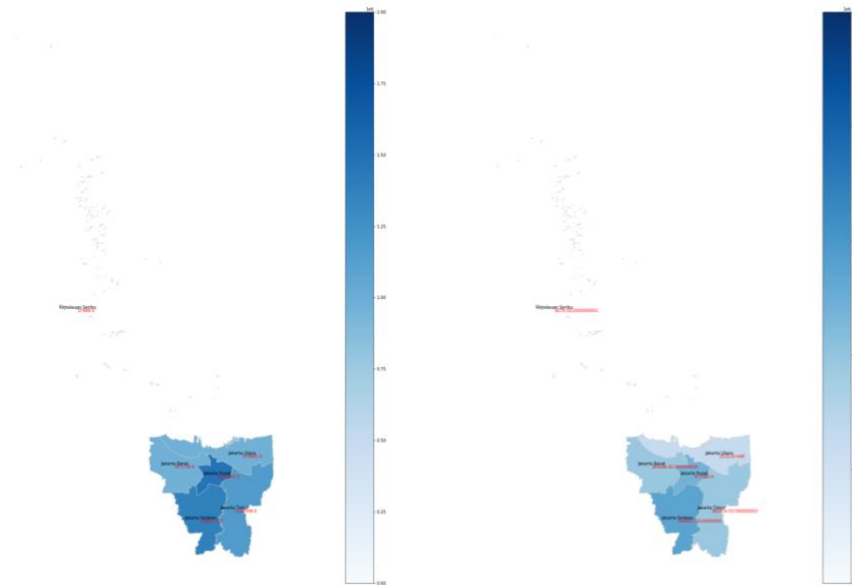
Dari grafik diatas bisa kita lihat bahwa pada tahun 2020 kebanyakan kasus COVID19 merupakan kasus tidak bergejala, tetapi pada awal tahun 2021 kebanyakan kasus positif COVID19 merupakan kasus yang bergejala. Sehingga ini bukan penyebab utama merebaknya kasus COVID19 di DKI Jakarta. Tetapi dari data tersebut bisa dilihat bahwa kesadaran masyarakat begitu rendah dengan banyaknya yang bergejala harusnya masyarakat bisa sadar dan memakai masker atau menjaga jarak dari sekitar sehingga tidak menyebar ke anggota terdekat atau kontak terdekat kita.



Gambar 10 Perbandingan Self Insolation dan Masih Perawatan

Dari grafik di atas bisa dilihat bahwa jumlah isolasi mandiri meningkat drastis mulai bulan Juni 2021 disertai dengan kasus positif yang meningkat drastis, di sini menurut analisis tim kami penyebab utama merebaknya kasus positif bisa disebabkan oleh tingginya isolasi mandiri di rumah yang tidak menerapkan protokol kesehatan,

sehingga menyebabkan anggota rumah dan lingkungan sekitar terpapar, apalagi ketika ada lansia di rumah yang dapat menyebabkan lansia terserang COVID19 dan meninggal. Sehingga diperlukannya pendisiplinan protokol isolasi mandiri oleh pemerintah untuk menekan angka positif COVID19 di DKI Jakarta.



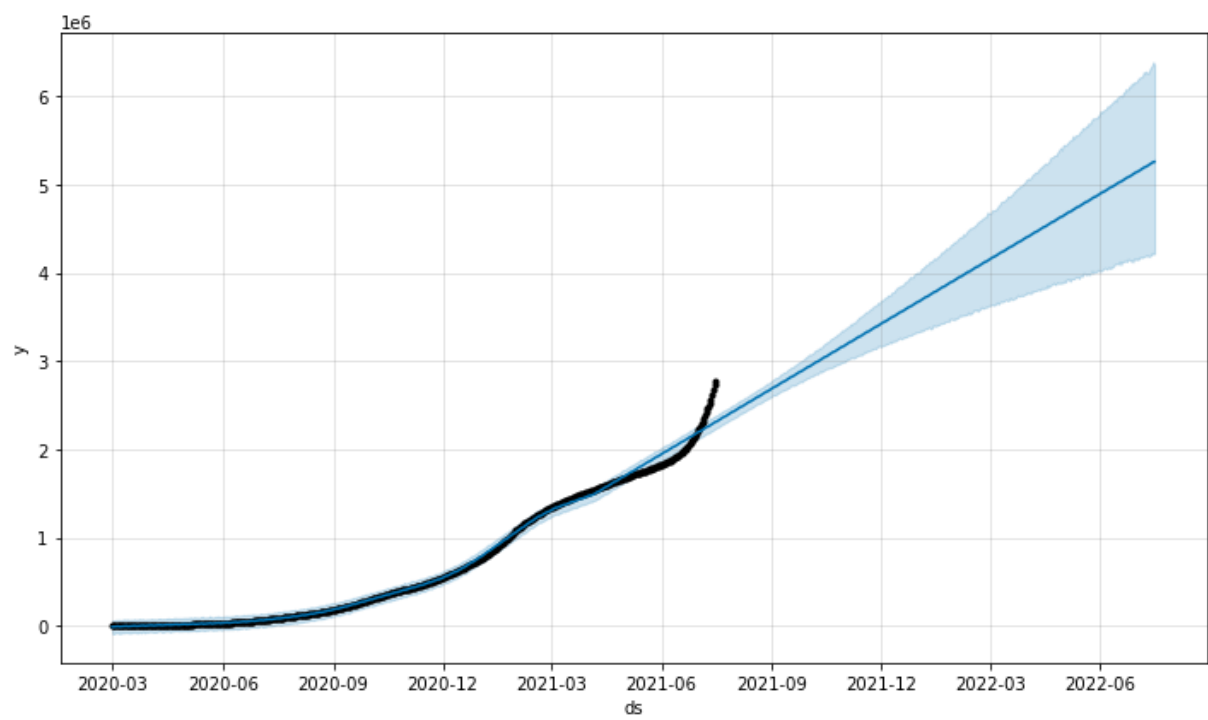
Gambar 11 Grafik Persebaran Vaksinasi ke-1 di Kota DKI Jakarta

Gambar 12 Grafik Persebaran Vaksinasi ke-2 di Kota DKI Jakarta

Grafik gambar 11 adalah grafik persebaran pelaksanaan vaksinasi ke-1 di kota-kota yang berada di DKI Jakarta. Sedangkan, grafik gambar 12 adalah grafik persebaran pelaksanaan vaksinasi ke-2 di kota-kota yang berada di DKI Jakarta.

Dari grafik tersebut bisa dilihat bahwa pembagian vaksinasi belum merata dan hanya terfokus ke Jakarta Timur, Jakarta Pusat, Jakarta Selatan, sedangkan untuk Jakarta Utara masih belum setengah tervaksinasi. Sehingga menurut kami diperlukan adanya pemerataan vaksinasi di seluruh wilayah DKI Jakarta agar *herd immunity* dapat terjadi dan penyebaran COVID19 akan lebih terkontrol pada seluruh daerah karena tidak berfokus pada satu daerah yang vaksinasi nya belum terlaksana sesuai target.

Prediksi kenaikan angka positif COVID19 pada DKI Jakarta jika tidak ada perubahan drastis yang dilakukan oleh pemerintah:



Kesimpulan

Dari hasil penelitian, kami menemukan bahwa DKI Jakarta sangat lah berperan penting dalam pertumbuhan kasus positif, sembuh, meninggal, vaksinasi secara Nasional. Lalu kami menganalisis penyebabnya, apakah ini disebabkan oleh tingginya *testing* yang dilakukan oleh pemerintah DKI Jakarta yang hasilnya bukan itu penyebabnya, dan apakah vaksinasi dapat menekan kenaikan angka positif COVID19 dan hasilnya juga menunjukkan bahwa vaksinasi tidak dapat menekan kenaikan angka positif COVID19, lalu kami melakukan analisa ulang pada beberapa aspek lain dan menemukan bahwa kemungkinan kenaikan angka positif COVID19 dikarenakan isolasi mandiri yang meningkat drastis bersamaan dengan kenaikan angka positif COVID19, menurut kami protokol pelaksanaan isolasi mandiri oleh orang yang terinfeksi oleh COVID19 masih kurang, dikarenakan kemungkinan adanya pelanggaran protokol kesehatan oleh orang yang terinfeksi oleh COVID19 sehingga orang yang kontak dekat dengan orang yang terinfeksi juga ikut terinfeksi oleh COVID19. Sehingga perlu adanya pembuatan protokol isolasi yang lebih terstruktur. Lalu, kami juga menemukan tingkat *testing* yang masih rendah pelaksanaan, karena jumlah *testing* tidak berimbang dengan angka kenaikan angka positif COVID19 harian, sehingga pemerintah DKI Jakarta menurut kami harus menaikkan jumlah *testing* dan melakukan *tracing* secara cermat. Faktor lain yang menurut kami menyebabkan kenaikan kasus COVID19 adalah vaksinasi yang belum merata pada beberapa kota di DKI Jakarta, sehingga diperlukan pemerataan vaksin agar rumah sakit di DKI Jakarta tidak akan tumbang karena kasus positif yang berfokus pada kota yang tingkat vaksinasi-nya masih rendah.

Daftar Pustaka

- Ajitesh, K. (2020, September 29). *Correlation Concepts, Matrix & Heatmap using Seaborn*. Retrieved from vitalflux: <https://vitalflux.com/correlation-heatmap-with-seaborn-pandas/>
- Badr, W. (2019, Maret 5). *towardsdatascience*. Retrieved from 5 Ways to Detect Outliers/Anomalies That Every Data Scientist Should Know (Python Code): 5 Ways to Detect Outliers/Anomalies That Every Data Scientist Should Know (Python Code)
- Kwan, W. (2021, Juni 4). *Perbedaan Statistika Deskriptif dan Inferensial*. Retrieved from algorit.ma: <https://algorit.ma/blog/statistika-deskriptif-inferensial/>
- Lutes, J. (2020, Desember 3). *Correlation Is Simple With Seaborn And Pandas*. Retrieved from towardsdatascience: <https://towardsdatascience.com/correlation-is-simple-with-seaborn-and-pandas-28c28e92701e>
- Mufti, S. A. (2020, Juli 28). *Spatial Data Solution for City Planning in Indonesia: Understanding The GeoDataFrame*. Retrieved from towardsdatascience: <https://towardsdatascience.com/spatial-data-solution-for-city-planning-in-indonesia-understanding-the-geodataframe-f50d58e6c9f2>
- Muluberhan-Berhe, N. (2020, Juni 3). *COVID-19 Sentiment Analysis Series, Part 1: Getting Started*. Retrieved from towardsdatascience: <https://towardsdatascience.com/covid-19-sentiment-analysis-series-part-1-getting-started-70892122630c>
- Muluberhan-Berhe, N. (2020, Juni 10). *COVID-19 Sentiment Analysis Series, Part 2: Hypothesis Testing*. Retrieved from towardsdatascience: <https://towardsdatascience.com/covid-19-sentiment-analysis-series-part-2-hypothesis-testing-3db511875165>
- R, V. (2018, September 12). *Feature selection — Correlation and P-value*. Retrieved from towardsdatascience: <https://towardsdatascience.com/feature-selection-correlation-and-p-value-da8921bfb3cf>
- Scott, N. (2019, Agustus 19). *Defining A Data Science Problem*. Retrieved from towardsdatascience: <https://towardsdatascience.com/defining-a-data-science-problem-4cbf15a2a461>

Lampiran

Source Code Jupyter Notebook: <https://github.com/StevenTannn/Compfest-DS-Covid>