

Bagian 1: Analisis Dataset

Laporan Analisis Covid-19 di Jakarta

Latar Belakang

Ini belum...

Jawaban Soal

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

```
In [6]: df.agg({"positif_harian":["max", np.median, "min", "mean"]})
Out[6]:
```

	positif_harian
max	14619.000000
median	884.000000
min	0.000000
mean	1356.835671

```
In [7]: df["positif_harian"].mode()[0]
Out[7]: 0
```

Gambar 1.1 Mencari Agg pada kolom positif_harian

Dari *syntax* yang ditampilkan pada Gambar 1.1 di atas. Dapat diketahui bahwa nilai *mean* atau rata-rata kasus positif harian di Jakarta adalah 1356 kasus, nilai *median* untuk kolom positif_harian adalah 884, dan *modus* (data yang paling sering muncul) pada kolom positif_harian adalah 0.

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

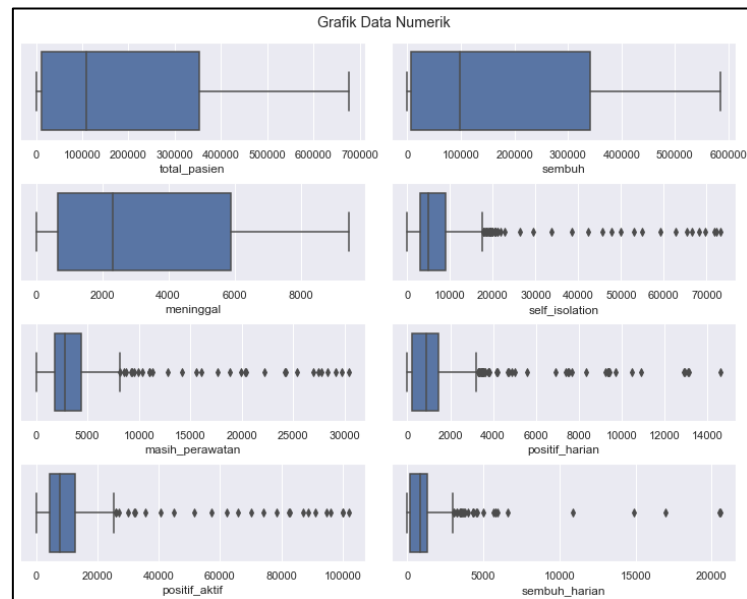
```
In [8]: df.agg({"positif_harian":["max", "min"]})
Out[8]:
```

	positif_harian
max	14619
min	0

Gambar 1.2 Nilai minimal dan maksimal kasus positif harian

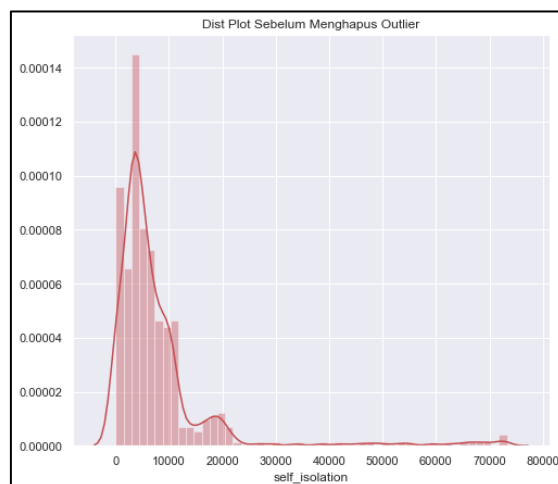
Dari *syntax* yang ditampilkan pada Gambar 1.2 di atas. Nilai maksimal untuk kasus positif harian di Jakarta adalah 14619 kasus, dan nilai minimal atau kasus terendah positif harian adalah 0 (*tidak ada kasus positif harian*).

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



Gambar 1.3 BoxPlot data numerik

Berdasarkan Gambar 1.3 di atas, dari beberapa kolom yang ditampilkan dalam bentuk *boxplot*, ternyata terdapat beberapa kolom yang memiliki *outlier* diantaranya adalah kolom *self_isolation*, *masih_perawatan*, *positif_harian*, *positif_aktif*, dan kolom *sembuh_harian*. Oleh karena itu, kami memilih menganalisis *oulier* yang ada di kolom *self_isolation*. Grafik distribusi dari kolom *self_isolation* ditampilkan pada Gambar 1.4 di bawah ini.



Gambar 1.4 DistPlot kolom *self_isolation*

Untuk melihat batas yang dirasa merupakan outlier, maka kami menggunakan IQR untuk menghitungnya batas atas dan batas bawahnya. Rumus mencari nilai IQR dan batas-batasnya adalah sebagai berikut.

$$IQR = Q3 - Q1 \dots\dots\dots (1.1)$$

$$batas_{bawah} = Q1 + (1.5 * IQR) \dots\dots\dots (1.2)$$

$$batas_{atas} = Q3 + (1.5 * IQR) \dots\dots\dots (1.3)$$

Keterangan:

Q1 = kuartil 1 (25% data)

Q3 = kuartil 3 (75% data)

Sesuai dengan rumus di atas, kami melakukan perhitungan terhadap kolom self_isolation.

```
In [13]: q1 = df["self_isolation"].quantile(0.25)
q3 = df["self_isolation"].quantile(0.75)
IQR = q3 - q1
batas_bawah = q1 - (1.5 * IQR)
batas_atas = q3 + (1.5 * IQR)
print(f"nilai Q1 = {q1}\nnilai Q3 = {q3}\nnilai IQR = {IQR}\nnilai batas bawah = {batas_bawah}\nnilai batas atas = {batas_atas}")

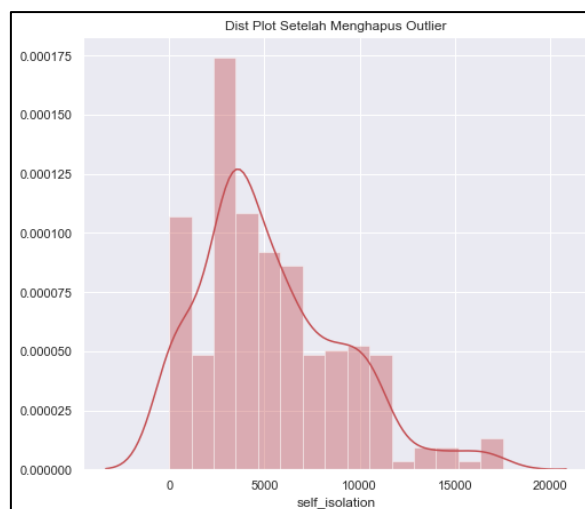
nilai Q1 = 2997.5
nilai Q3 = 8868.0
nilai IQR = 5870.5
Batas Bawah = -5808.25
Batas Atas = 17673.75

In [14]: df[(df["self_isolation"]>=batas_bawah) & (df["self_isolation"]<=batas_atas)]

Out[14]: max    17524
min         0
Name: self_isolation, dtype: int64
```

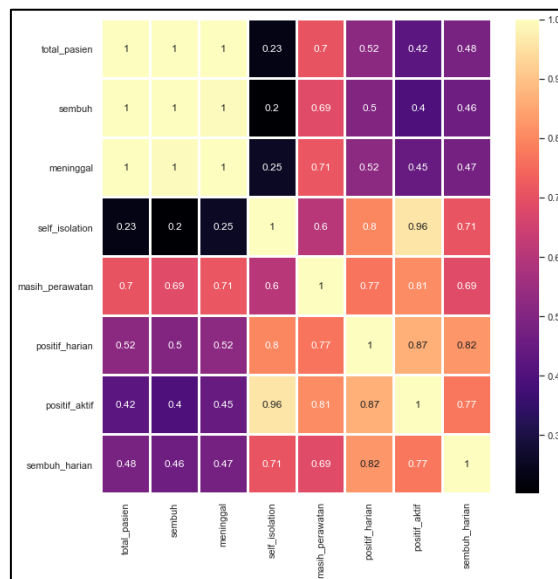
Gambar 1.5 Perhitungan IQR dan Batas

Hasil dari perhitungan yang ditampilkan pada Gambar 1.5 di atas adalah nilai IQR adalah 6870, batas atas adalah 17673, dan batas bawah adalah -5808 atau 0 karena tidak ada data yang negatif. Sedangkan data pada kolom self_isolation yang sudah dibersihkan *outlier* mempunyai rentang data dari 0 sampai 17524. Grafik distribusi kolom self_isolation setelah dihapus *outlier* dapat dilihat pada Gambar 1.6 di bawah ini.



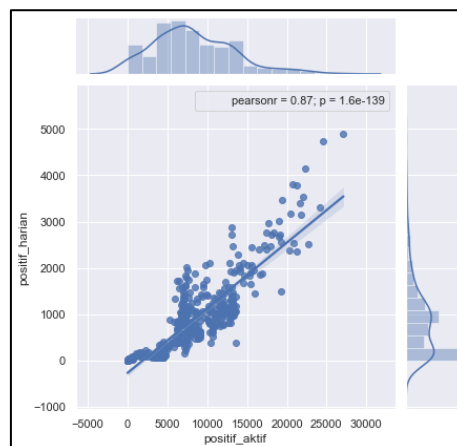
Gambar 1.6 DistPlot self_isolation setelah dihapus outlier

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 1.7 Heatmap data covid-19 Jakarta

Berdasarkan grafik heatmap yang dapat dilihat pada Gambar 1.7 di atas, diketahui bahwa kolom yang memiliki korelasi paling erat (*berdasarkan data*) adalah kolom *positif_harian* dan *positif_aktif*. Oleh karena itu, kami memilih kolom tersebut untuk kami analisis. Grafik untuk kedua kolom dapat dilihat pada gambar di bawah ini.



Gambar 1.8 JointPlot kolom *positif_harian* dan *positif_aktif*

Dari Gambar 1.8 di atas dapat dilihat bahwa pertambahan jumlah positif aktif selaras dengan bertambahnya jumlah positif harian yang ada. Dari grafik dapat diketahui bahwa korelasi dari kedua variabel tersebut mencapai 0.87 yang berarti memiliki korelasi yang kuat. Kesimpulan yang kami berikan dari analisis grafik di atas adalah kasus positif harian dan positif aktif yang meningkat secara bersamaan. Selama

jumlah kasus positif harian yang ada naik, maka kasus positif aktif juga akan naik. Dan juga menandakan pasien yang positif aktif belum banyak yang sembuh.

Hasil Analisis Tambahan

a. Problem Statement

Berdasarkan datasets yang ada dan analisis dari data tersebut, permasalahan yang ada adalah:

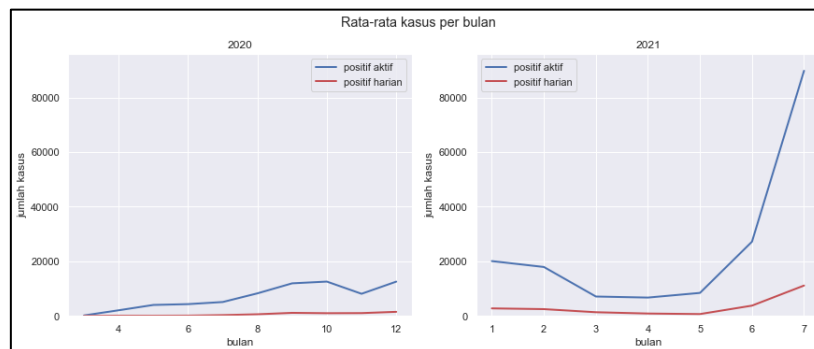
- Kasus positif aktif yang meningkat lagi dari pada tahun 2021 terutama pada bulan Juni dan Juli.

b. Hypothesis

Ini belum...

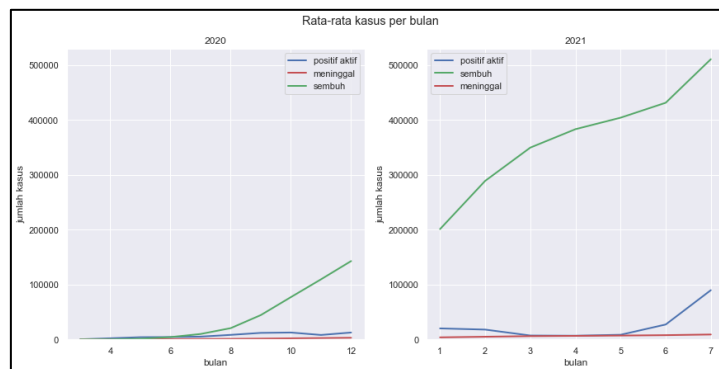
c. Exploratory Data Analysis

Untuk melakukan analisis kami melakukan beberapa plotting terhadap data. Gambar di bawah ini adalah gambar grafik perkembangan covid-19 di Jakarta menurut bulan, dan dibagi menurut tahun.



Gambar 1. 9 Perkembangan Covid-19 di Jakarta

Berdasarkan grafik pada gambar di atas dapat dikatakan bahwa kasus covid-19 di Jakarta mengalami kenaikan jumlah kasus yang sangat tinggi. Sedangkan tingkat kesembuhan dapat dilihat pada Gambar 1.10 di bawah ini.



Gambar 1. 10 Tingkat kesembuhan

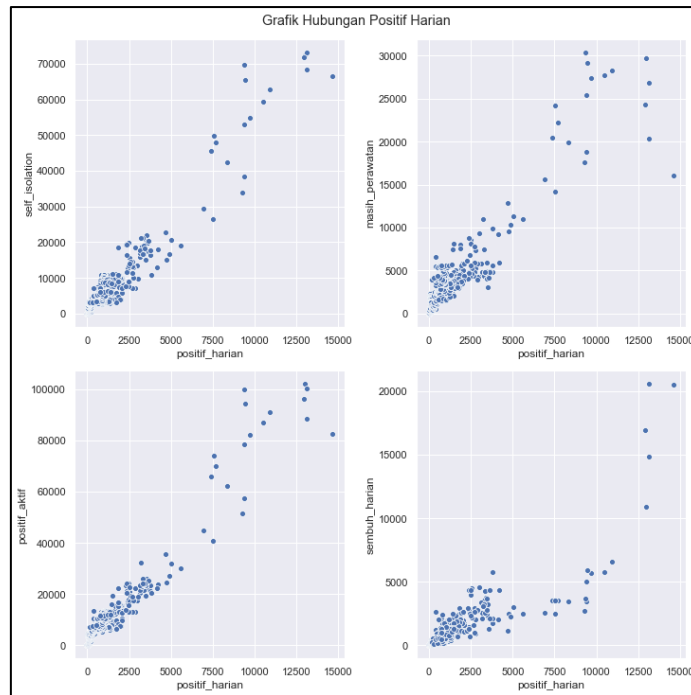
Dari gambar 1.10 di atas bisa dilihat bahwa ternyata jumlah kesembuhan yang ada ternyata sangat banyak, bahkan melebihi jumlah kasus aktif yang ada, karena jumlah kesembuh tersebut gabungan dari pasien positif aktif maupun kesembuhan dari pasien tanpa gejala maupun dengan gejala ringan lainnya. Data positif harian, positif aktif dan kesembuhan dalam bentuk tabel dapat dilihat pada Gambar 1.11 di bawah ini.

[32]:

		positif_harian	positif_aktif	sembuh
year	month			
2021	7	11132.750000	89767.166667	510709.666667
	6	3780.300000	27204.466667	431337.766667
	1	2773.645161	20088.290323	201050.580645
	2	2500.607143	17931.821429	288952.428571
2020	12	1512.064516	12564.516129	142770.096774
2021	3	1365.161290	7115.096774	349573.387097
2020	9	1135.300000	11944.100000	44344.766667
	11	1042.133333	8118.000000	109508.166667
	10	1007.387097	12582.064516	77101.516129
2021	4	885.500000	6732.966667	383145.233333
	5	691.580645	8441.483871	404089.129032
2020	8	616.387097	8281.451613	20555.741935
	7	320.161290	5108.161290	9914.677419
	6	133.466667	4332.333333	4229.966667
	4	113.233333	2084.466667	194.600000
	5	101.096774	4033.225806	1236.645161
	3	23.903226	188.774194	12.806452

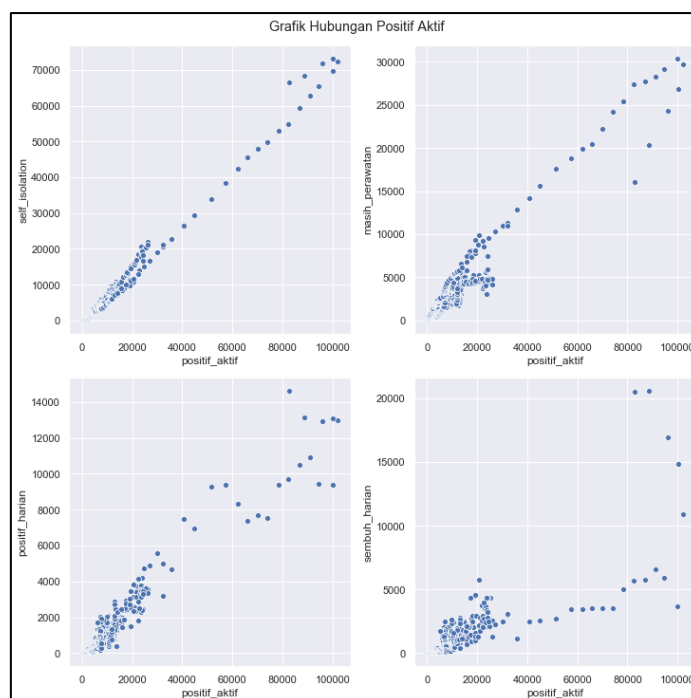
Gambar 1. 11 *Data Dibagi per tahun dan bulan*

Sedangkan kemungkinan korelasi data yang ada dari data positif harian ditampilkan pada Gambar 1.12 di bawah ini.



Gambar 1. 12 Grafik data positif harian

Dari grafik di atas terlihat hampir semua data yang ada memiliki korelasi yang baik, yaitu semakin banyak nilai kasus positif harian (*dalam sumbu x*) maka semakin banyak pula nilai yang ditampilkan pada sumbu y. Mulai dari jumlah data *self_isolation* sampai data kesembuhan yang bertambah walaupun sempat ada kenaikan data positif harian namun kesembuhan tidak mengalami pertambahan. Untuk data yang berkorelasi dengan *positif_aktif* ditampilkan pada Gambar 1.13 di bawah ini.



Gambar 1. 13 Grafik kasus positif harian

Dari Gambar 1.13 di atas terlihat bahwa data positif aktif semakin naik, maka data lain (*ditampilkan dalam sumbu y*) juga ikut naik. Walaupun dalam data kesembuhan terdapat data positif aktif yang mengalami kenaikan namun data kesembuhan tidak naik.

d. Initial Findings

Ini belum...

e. Deep Dive Analysis

f. Conclusion and Recommendation

Kesimpulan

Daftar Pustaka