

2. Dengan mengubah kategori kecil, sedang, dan besar menjadi 3, 2, 1 untuk kolom workload dan Data requirement dan 1, 2, 3 untuk selainya secara berurutan dan mengalikannya terhadap bobot pada tiap kolom maka akan didapatkan tabel sebagai berikut:

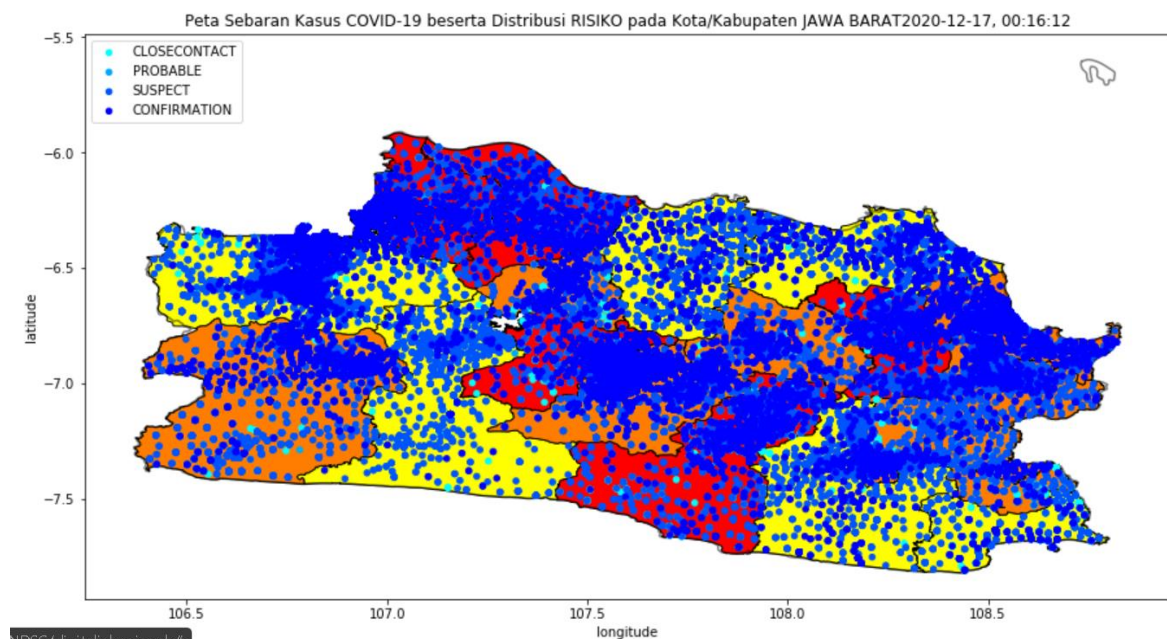
Opsi	Potential impact	Workload: technical	Workload: non-technical	Workload: monitoring	Data requirement	Relevance to stakeholders needs	Sustainability	Val
A	3	2	2	2	1	2	3	2.10
B	2	2	1	2	2	1	1	1.65
C	3	3	1	2	1	2	2	2.05
D	1	1	2	3	1	3	1	1.60
E	2	1	3	2	3	2	2	2.15

Dengan Bobot,

Potential impact	Workload: technical	Workload: non-technical	Workload: monitoring	Data requirement	Relevance to stakeholders needs	Sustainability
0	0.2	0.15	0.1	0.1	0.2	0.15

Nilai Val merupakan nilai rata-rata berbobot yang dapat diinterpretasikan sebagai tingkat kesukaan tiap opsi secara umum. Berdasarkan nilai Val yang didapatkan dapat disimpulkan bahwa Opsi terbaik adalah opsi E.

4. Gambar di bawah ini merupakan peta distribusi RISIKO pada kota/kabupaten di Jawa Barat yang beserta titik persebaran kasus. Data status RISIKO tiap kota dan kabupaten Jawa Barat didapatkan dari <https://covid19.go.id/peta-risiko>, data sebaran kasus COVID-19 Jawa Barat didapatkan dari API https://covid19-public.digitalservice.id/api/v1/sebaran_v2/jabar, dan data batas provinsi, kota, dan kabupaten didapatkan dari Open Data Jawa Barat. Kabupaten dan Kota yang berwarna merah merupakan lokasi dengan RISIKO TINGGI, berwarna oranye memiliki status RISIKO SEDANG, dan yang berwarna kuning memiliki status RISIKO RENDAH.



Dari Gambar diatas dapat terlihat bahwa daerah pada zona merah memiliki jumlah kasus yang padat.

5. Nama Tabel adalah sample_data

a) Lokasi dengan kemacetan level 5

```
SELECT a.id FROM sample_data AS a
WHERE a.level = 5;
```

b) Banyak kemacetan level 5 sebulan terakhir

```
SELECT COUNT(a.id) AS Total FROM sample_data AS a
WHERE a.level = 5 AND
DATEDIFF(a.date, CONVERT(varchar, GETDATE(), 104)) < 30;
```

c) 3 Lokasi kecepatan terendah dari jam 15-18

```
SELECT a.id T
FROM sample_data AS a
WHERE a.hour BETWEEN 15 AND 18
ORDER BY a.avg_speed_kmh ASC
LIMIT 3;
```

d) Temukan Anomali

```
SELECT hour_level, COUNT(hour_level)
FROM sample_data AS a
GROUP BY concat_ws('_', a.hour, a.level) AS hour_level
ORDER BY COUNT(hour_level) ASC;
LIMIT 3;
```

6.

- a) Univariate, bivariate, dan multivariate dibedakan berdasarkan jumlah variabel atau karakteristik dari sample yang akan dianalisis.
- b) Normality merupakan test untuk sampel yang berdistribusi normal. T-test merupakan test untuk membandingkan 2 sample dengan distribusi normal dengan jumlah sampel kecil, Chi-square test merupakan test signifikansi antara dua kategori.
- c) Cara menemukan data duplikat adalah dengan melakukan komparasi data dengan masing-masing elemennya kemudian akan didapatkan array boolean yang kemudian dijumlahkan,

jika elemen memiliki jumlah yang >1 maka elemen tersebut merupakan data duplikat. Cara mengatasi data duplikat adalah dengan menghilangkan data tersebut.

- d) Peranan data analyst dalam pengembangan produk digital adalah untuk menganalisis pola yang terdapat pada dataset, dan menentukan variabel yang berpengaruh pada tujuan penelitian.