

Kelompok 7

Anggota Kelompok:

1. Faiz Bayu Erlangga (2311231)
2. Marco Henrik Abineno (2301093)
3. Muhammad Alfi Fariz (2311174)
4. Naufal Dzaki Ibrahim (2309815)
5. Qalam Noer Fazrian (2304746)

Kelas: 3KOMC2

Mata Kuliah: IK505 - Data Mining and Warehouse

Dosen Pengampu: Dr. Yudi Wibisono, S.T., M.T.

Pendahuluan

Kesehatan masyarakat merupakan salah satu indikator penting dalam menilai kualitas hidup di suatu wilayah. Kota Bandung, sebagai salah satu kota metropolitan di Indonesia, menghadapi berbagai tantangan dalam mengelola kesehatan masyarakatnya. Berbagai penyakit menular seperti diare, demam berdarah dengue (DBD), kaki gajah, TBC, dan AIDS, serta penyakit tidak menular seperti hipertensi, masih menjadi masalah kesehatan utama yang memerlukan perhatian serius. Selain itu, penyakit yang kurang umum namun tetap berdampak, seperti hemolisisida, juga muncul sebagai ancaman bagi kesehatan masyarakat.

Penyakit-penyakit menular, seperti DBD dan TBC, memiliki potensi penyebaran yang cepat, terutama di wilayah dengan tingkat kepadatan penduduk yang tinggi dan sanitasi yang kurang memadai. Di sisi lain, penyakit tidak menular seperti hipertensi sering kali diabaikan karena sifatnya yang asimtomatik, namun bisa menyebabkan komplikasi serius seperti penyakit jantung dan stroke. Untuk itu, pemahaman yang lebih mendalam mengenai prevalensi, faktor risiko, dan distribusi penyakit ini menjadi sangat penting dalam upaya pengendalian dan pencegahan.

Jumlah fasilitas kesehatan (faskes) dan data demografi penduduk di Kota Bandung juga merupakan komponen vital dalam analisis ini. Informasi ini membantu dalam memahami kapasitas sistem kesehatan untuk menangani berbagai penyakit serta memastikan bahwa intervensi yang dilakukan dapat menjangkau seluruh masyarakat dengan efektif. Dengan memperhatikan jumlah dan distribusi faskes, dapat diketahui apakah akses terhadap layanan kesehatan sudah merata dan mencukupi kebutuhan penduduk yang terus meningkat dari tahun ke tahun.

Import Library

```
In [313]: # display
%matplotlib inline
# numerik
import numpy as np
# analisis
import pandas as pd
# plot
import matplotlib.pyplot as plt
# visualisasi
import seaborn as sns

from tabulate import tabulate
```

Membaca Data file CSV

```
In [316]: #Tabel Beberapa Penyakit
df_aids = pd.read_csv("jmlh_kss_br_ds_brdsrkn_klmpk_mr_d_kt_bndng.csv")
df_dbd = pd.read_csv("jmlh_kss_dmm_brdrh_dng_dbd_mrnt_psksms_d_kt_bndng.csv")
df_hpt = pd.read_csv("jmlh_pndrt_hprtns_brs_15_thn_brdsrkn_pmrkn_psksms_d_kt_bndng.csv")
df_kk_gjh = pd.read_csv("jmlh_psn_flrss_kk_gjh_mrnt_psksms_d_kt_bndng.csv")
df_diare = pd.read_csv("jumlah_pasien_penderita_diare_dilayani_puskesmas_kota_bandung.csv")
df_hemo = pd.read_csv("jumlah_kunjungan_pasien_hemodialisa_di_rsud_kota_bandung.csv")
df_tbc = pd.read_csv("jmlh_kss_br_tbrklss_pr_brdsrkn_fslts_kshtn_d_kt_bndng.csv")
df_mal = pd.read_csv("jmlh_kss_pstf_mlr_brdsrkn_fslts_kshtn_d_kt_bndng.csv")

#Tabel Pendukung
df_penduduk = pd.read_csv("jumlah_penduduk_kota_bandung_berdasarkan_kecamatan.csv")
df_faskes = pd.read_csv("jumlah_fasilitas_kesehatan_di_kota_bandung_menurut_kepemilikan.csv")
```

Menampilkan Dataframe 5 Baris Teratas

Ini merupakan dataframe penderita penyakit DBD di Kota Bandung

In [320..

df_dbd.head(5)

Out[320..

	id	kode_provinsi	nama_provinsi	bps_kode_kabupaten_kota	bps_nama_kabupaten_kota	bps_kode_kecamatan	bps_nama_kecan
0	1	32	JAWA BARAT	3273	KOTA BANDUNG	3273130	ARCAM
1	2	32	JAWA BARAT	3273	KOTA BANDUNG	3273130	ARCAM
2	3	32	JAWA BARAT	3273	KOTA BANDUNG	3273150	KIARACOND
3	4	32	JAWA BARAT	3273	KOTA BANDUNG	3273150	KIARACOND
4	5	32	JAWA BARAT	3273	KOTA BANDUNG	3273020	BABAKAN CIP/

Ini merupakan dataframe penderita penyakit Aids di Kota Bandung

In [323..

df_aids.head(5)

Out[323..

	id	kode_provinsi	nama_provinsi	bps_kode_kabupaten_kota	bps_nama_kabupaten_kota	kelompok_umur	jumlah_kasus_baru	
0	1	32	JAWA BARAT	3273	KOTA BANDUNG	(<=4)	5	C
1	2	32	JAWA BARAT	3273	KOTA BANDUNG	(5-14)	0	C
2	3	32	JAWA BARAT	3273	KOTA BANDUNG	(15-19)	4	C
3	4	32	JAWA BARAT	3273	KOTA BANDUNG	(20-24)	31	C
4	5	32	JAWA BARAT	3273	KOTA BANDUNG	(25-49)	149	C

Ini merupakan dataframe penderita penyakit Hipertensi di Kota Bandung

In [326..

df_hpt.head(5)

Out[326..

	id	kode_provinsi	nama_provinsi	bps_kode_kabupaten_kota	bps_nama_kabupaten_kota	bps_kode_kecamatan	bps_nama_kecan
0	1	32	JAWA BARAT	3273	KOTA BANDUNG	3273180	A
1	2	32	JAWA BARAT	3273	KOTA BANDUNG	3273180	A
2	3	32	JAWA BARAT	3273	KOTA BANDUNG	3273141	ANTA
3	4	32	JAWA BARAT	3273	KOTA BANDUNG	3273141	ANTA
4	5	32	JAWA BARAT	3273	KOTA BANDUNG	3273141	ANTA

Ini merupakan dataframe penderita penyakit Kaki Gajah di Kota Bandung

In [329..

df_kk_gjh.head(5)

Out [329..

	id	kode_provinsi	nama_provinsi	bps_kode_kabupaten_kota	bps_nama_kabupaten_kota	bps_kode_kecamatan	bps_nama_kecan
0	1	32	JAWA BARAT	3273	KOTA BANDUNG	3273160	BATUNUNC
1	2	32	JAWA BARAT	3273	KOTA BANDUNG	3273160	BATUNUNC
2	3	32	JAWA BARAT	3273	KOTA BANDUNG	3273160	BATUNUNC
3	4	32	JAWA BARAT	3273	KOTA BANDUNG	3273160	BATUNUNC
4	5	32	JAWA BARAT	3273	KOTA BANDUNG	3273160	BATUNUNC

Ini merupakan dataframe penderita penyakit Diare di Kota Bandung

In [332..

df_diare.head(5)

Out [332..

	id	kode_provinsi	nama_provinsi	bps_kode_kabupaten_kota	bps_nama_kabupaten_kota	bps_kode_kecamatan	bps_nama_kecan
0	1	32	JAWA BARAT	3273	KOTA BANDUNG	3273180	A
1	2	32	JAWA BARAT	3273	KOTA BANDUNG	3273141	ANTA
2	3	32	JAWA BARAT	3273	KOTA BANDUNG	3273130	ARCAM
3	4	32	JAWA BARAT	3273	KOTA BANDUNG	3273050	ASTANAAN
4	5	32	JAWA BARAT	3273	KOTA BANDUNG	3273020	BABAKAN CIP/

Ini merupakan dataframe penderita penyakit Hemodialisa di Kota Bandung

In [335..

df_hemo.head(5)

Out [335..

	id	kode_provinsi	nama_provinsi	bps_kode_kabupaten_kota	bps_nama_kabupaten_kota	bulan	jumlah_pasien	satuan	ta
0	1	32	JAWA BARAT	3273	KOTA BANDUNG	JANUARI	209	ORANG	2
1	2	32	JAWA BARAT	3273	KOTA BANDUNG	FEBRUARI	224	ORANG	2
2	3	32	JAWA BARAT	3273	KOTA BANDUNG	MARET	281	ORANG	2
3	4	32	JAWA BARAT	3273	KOTA BANDUNG	APRIL	304	ORANG	2
4	5	32	JAWA BARAT	3273	KOTA BANDUNG	MEI	314	ORANG	2

Ini merupakan dataframe penderita penyakit TuberColosis di Kota Bandung

In [338..

df_tbc.head(5)

Out [338..

	id	kode_provinsi	nama_provinsi	bps_kode_kabupaten_kota	bps_nama_kabupaten_kota	fasilitas_kesehatan	jenis_kelamin	juml
0	1	32	JAWA BARAT	3273	KOTA BANDUNG	UPT CIBUNTU	LAKI LAKI	
1	2	32	JAWA BARAT	3273	KOTA BANDUNG	UPT CIBUNTU	PEREMPUAN	
2	3	32	JAWA BARAT	3273	KOTA BANDUNG	UPT CARINGIN	LAKI LAKI	
3	4	32	JAWA BARAT	3273	KOTA BANDUNG	UPT CARINGIN	PEREMPUAN	
4	5	32	JAWA BARAT	3273	KOTA BANDUNG	UPT CETARIP	LAKI LAKI	

Ini merupakan dataframe penderita penyakit Malaria di Kota Bandung

In [341..

df_mal.head(5)

Out[341...

	id	kode_provinsi	nama_provinsi	bps_kode_kabupaten_kota	bps_nama_kabupaten_kota	fasilitas_kesehatan	jumlah_positif_mala
0	1	32	JAWA BARAT	3273	KOTA BANDUNG	RS ADVENT	
1	2	32	JAWA BARAT	3273	KOTA BANDUNG	RS AL ISLAM	
2	3	32	JAWA BARAT	3273	KOTA BANDUNG	RS DR. SALAMUN	
3	4	32	JAWA BARAT	3273	KOTA BANDUNG	RS HERMINA ARCAMANIK	
4	5	32	JAWA BARAT	3273	KOTA BANDUNG	RS HERMINA PASTEUR	

Ini merupakan dataframe Penduduk di Kota Bandung

In [344...

df_penduduk.head(5)

Out[344...

	id	kode_provinsi	nama_provinsi	bps_kode_kabupaten_kota	bps_nama_kabupaten_kota	bps_kode_kecamatan	bps_nama_kecan
0	1	32	JAWA BARAT	3273	KOTA BANDUNG	3273180	A
1	2	32	JAWA BARAT	3273	KOTA BANDUNG	3273180	A
2	3	32	JAWA BARAT	3273	KOTA BANDUNG	3273141	ANTA
3	4	32	JAWA BARAT	3273	KOTA BANDUNG	3273141	ANTA
4	5	32	JAWA BARAT	3273	KOTA BANDUNG	3273130	ARCAM

Ini merupakan dataframe fasilitas kesehatan di Kota Bandung

In [347...

df_faskes.head(5)

Out[347...

	id	kode_provinsi	nama_provinsi	bps_kode_kabupaten_kota	bps_nama_kabupaten_kota	fasilitas_kesehatan	pemilik	jumlah
0	1	32	JAWA BARAT	3273	KOTA BANDUNG	RUMAH SAKIT UMUM	KEMENKES	
1	2	32	JAWA BARAT	3273	KOTA BANDUNG	RUMAH SAKIT UMUM	PEM KAB/KOTA	
2	3	32	JAWA BARAT	3273	KOTA BANDUNG	RUMAH SAKIT UMUM	TNI/POLRI	
3	4	32	JAWA BARAT	3273	KOTA BANDUNG	RUMAH SAKIT UMUM	SWASTA	
4	5	32	JAWA BARAT	3273	KOTA BANDUNG	RUMAH SAKIT KHUSUS	KEMENKES	

Membersihkan atribut data yang tidak diperlukan (Praproses)

Dilakukan untuk membersihkan data dengan menghapus kolom yang tidak diperlukan, sehingga hanya kolom yang relevan dengan analisis yang tersisa.

In [350...

df_aids = df_aids.drop(['id', 'satuan', 'kelompok_umur', 'bps_kode_kabupaten_kota', 'kode_provinsi', 'nama_prov

In [352...

df_dbd = df_dbd.drop(['id', 'satuan', 'kode_provinsi', 'bps_kode_kabupaten_kota', 'kode_provinsi', 'nama_provin

In [354...

df_hpt = df_hpt.drop(['id', 'satuan', 'kode_provinsi', 'bps_kode_kabupaten_kota', 'kode_provinsi', 'nama_provin

In [356...

df_kk_gjh = df_kk_gjh.drop(['id', 'satuan', 'kode_provinsi', 'bps_kode_kabupaten_kota', 'kode_provinsi', 'nama_j

In [358...

df_diare = df_diare.drop(['id', 'satuan', 'kode_provinsi', 'bps_kode_kabupaten_kota', 'kode_provinsi', 'nama_pr

In [360...

df_hemo = df_hemo.drop(['id', 'kode_provinsi', 'nama_provinsi', 'bps_kode_kabupaten_kota', 'bps_nama_kabupaten_l

In [362...

df_tbc = df_tbc.drop(['id', 'kode_provinsi', 'nama_provinsi', 'bps_kode_kabupaten_kota', 'bps_nama_kabupaten_ko

In [364...

df_mal = df_mal.drop(['id', 'kode_provinsi', 'nama_provinsi', 'bps_kode_kabupaten_kota', 'bps_nama_kabupaten_ko

In [366...

df_penduduk = df_penduduk.drop(['id', 'kode_provinsi', 'nama_provinsi', 'bps_kode_kabupaten_kota', 'bps_nama_kal

In [368...

df_faskes = df_faskes.drop(['id', 'kode_provinsi', 'nama_provinsi', 'bps_kode_kabupaten_kota', 'bps_nama_kabupa

Mengubah Tipe Data dari Int64 Menjadi Int32

```
In [371.. df_aids['jumlah_kasus_baru'] = df_aids['jumlah_kasus_baru'].astype('int32')
df_aids['tahun'] = df_aids['tahun'].astype('int32')

In [373.. df_dbd['jumlah_kasus'] = df_dbd['jumlah_kasus'].astype('int32')
df_dbd['tahun'] = df_dbd['tahun'].astype('int32')

In [375.. df_hpt['jumlah_penderita_hipertensi'] = df_hpt['jumlah_penderita_hipertensi'].astype('int32')
df_hpt['tahun'] = df_hpt['tahun'].astype('int32')

In [377.. df_kk_gjh['jumlah_pasien'] = df_kk_gjh['jumlah_pasien'].astype('int32')
df_kk_gjh['tahun'] = df_kk_gjh['tahun'].astype('int32')

In [379.. df_diare['jumlah_pasien'] = df_diare['jumlah_pasien'].astype('int32')
df_diare['tahun'] = df_diare['tahun'].astype('int32')

In [381.. df_hemo['jumlah_pasien'] = df_hemo['jumlah_pasien'].astype('int32')
df_hemo['tahun'] = df_hemo['tahun'].astype('int32')

In [383.. df_tbc['jumlah_kasus_baru'] = df_tbc['jumlah_kasus_baru'].astype('int32')
df_tbc['tahun'] = df_tbc['tahun'].astype('int32')

In [385.. df_penduduk['jumlah_penduduk'] = df_penduduk['jumlah_penduduk'].astype('int32')
df_penduduk['tahun'] = df_penduduk['tahun'].astype('int32')

In [387.. df_mal['jumlah_positif_malaria'] = df_mal['jumlah_positif_malaria'].astype('int32')
df_mal['tahun'] = df_mal['tahun'].astype('int32')
```

Menyaring data hanya dari tahun 2019 - 2022, jika jumlah kasus adalah 0 maka baris data dihapus dan melakukan sorting dari tahun terkecil ke tahun yang besar

```
In [390.. con1 = (df_aids['tahun'] >= 2019) & (df_aids['tahun'] <= 2022)
df_aids = df_aids[con1].reset_index(drop=True)
df_aids = df_aids[df_aids['jumlah_kasus_baru'] != 0]
df_aids.to_csv('filtered_data.csv', index=False)
if 'tahun' in df_aids.columns:
    df_aids.sort_values(by='tahun', ascending=True, inplace=True)

In [392.. con1 = (df_dbd['tahun'] >= 2019) & (df_dbd['tahun'] <= 2022)
df_dbd = df_dbd[con1].reset_index(drop=True)
df_dbd = df_dbd[df_dbd['jumlah_kasus'] != 0]
df_dbd.to_csv('filtered_data.csv', index=False)
if 'tahun' in df_aids.columns:
    df_dbd.sort_values(by='tahun', ascending=True, inplace=True)

In [394.. con1 = (df_hpt['tahun'] >= 2019) & (df_hpt['tahun'] <= 2022)
df_hpt = df_hpt[con1].reset_index(drop=True)
df_hpt = df_hpt[df_hpt['jumlah_penderita_hipertensi'] != 0]
df_hpt.to_csv('filtered_data.csv', index=False)

In [396.. con1 = (df_kk_gjh['tahun'] >= 2019) & (df_kk_gjh['tahun'] <= 2022)
df_kk_gjh = df_kk_gjh[con1].reset_index(drop=True)
df_kk_gjh = df_kk_gjh[df_kk_gjh['jumlah_pasien'] != 0]
df_kk_gjh.to_csv('filtered_data.csv', index=False)
if 'tahun' in df_aids.columns:
    df_kk_gjh.sort_values(by='tahun', ascending=True, inplace=True)

In [398.. con1 = (df_diare['tahun'] >= 2019) & (df_diare['tahun'] <= 2022)
df_diare = df_diare[con1].reset_index(drop=True)
df_diare = df_diare[df_diare['jumlah_pasien'] != 0]
df_diare.to_csv('filtered_data.csv', index=False)
if 'tahun' in df_aids.columns:
    df_diare.sort_values(by='tahun', ascending=True, inplace=True)

In [400.. con1 = (df_hemo['tahun'] >= 2019) & (df_hemo['tahun'] <= 2022)
df_hemo = df_hemo[con1].reset_index(drop=True)
df_hemo.to_csv('filtered_data.csv', index=False)
if 'tahun' in df_aids.columns:
    df_hemo.sort_values(by='tahun', ascending=True, inplace=True)

In [402.. con1 = (df_tbc['tahun'] >= 2019) & (df_tbc['tahun'] <= 2022)
df_tbc = df_tbc[con1].reset_index(drop=True)
df_tbc.to_csv('filtered_data.csv', index=False)
if 'tahun' in df_aids.columns:
    df_tbc.sort_values(by='tahun', ascending=True, inplace=True)
```

```
In [404.. con1 = (df_mal['tahun'] >= 2019) & (df_mal['tahun'] <= 2022)
df_mal = df_mal[con1].reset_index(drop=True)
df_mal.to_csv('filtered_data.csv', index=False)
if 'tahun' in df_mal.columns:
    df_mal.sort_values(by='tahun', ascending=True, inplace=True)
```

```
In [406.. con1 = (df_penduduk['tahun'] >= 2019) & (df_penduduk['tahun'] <= 2022)
df_penduduk = df_penduduk[con1].reset_index(drop=True)
df_penduduk.to_csv('filtered_data.csv', index=False)
if 'tahun' in df_penduduk.columns:
    df_penduduk.sort_values(by='tahun', ascending=True, inplace=True)
```

```
In [408.. con1 = (df_faskes['tahun'] >= 2019) & (df_faskes['tahun'] <= 2022)
df_faskes = df_faskes[con1].reset_index(drop=True)
df_faskes = df_faskes[df_faskes['jumlah_faskes'] != '-']
df_faskes.to_csv('filtered_data.csv', index=False)
if 'tahun' in df_faskes.columns:
    df_faskes.sort_values(by='tahun', ascending=True, inplace=True)
```

Mengganti nama atribut menjadi "jumlah_kasus_nama penyakit"

```
In [411.. df_aids = df_aids.rename(index=str, columns={"jumlah_kasus_baru": "jumlah_kasus_aids"})
```

```
In [413.. df_dbd = df_dbd.rename(index=str, columns={"jumlah_kasus": "jumlah_kasus_dbd"})
```

```
In [415.. df_hpt = df_hpt.rename(index=str, columns={"jumlah_penderita_hipertensi": "jumlah_kasus_hipertensi"})
```

```
In [417.. df_kk_gjh = df_kk_gjh.rename(index=str, columns={"jumlah_pasien": "jumlah_kasus_kaki_gajah"})
```

```
In [419.. df_diare = df_diare.rename(index=str, columns={"jumlah_pasien": "jumlah_kasus_diare"})
```

```
In [421.. df_hemo = df_hemo.rename(index=str, columns={"jumlah_pasien": "jumlah_kasus_hemodialisa"})
```

```
In [423.. df_tbc = df_tbc.rename(index=str, columns={"jumlah_kasus_baru": "jumlah_kasus_tuberculosis_paru"})
```

```
In [425.. df_mal = df_mal.rename(index=str, columns={"jumlah_positif_malaria": "jumlah_kasus_malaria"})
```

Kita merubah tipe data dari INT64 ke INT32 karena alokasi untuk INT64 membutuhkan alokasi memori yang sangat besar dibanding INT32

```
In [428.. df_aids.info()

<class 'pandas.core.frame.DataFrame'>
Index: 22 entries, 4 to 38
Data columns (total 2 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   jumlah_kasus_aids  22 non-null     int32
1   tahun            22 non-null     int32
dtypes: int32(2)
memory usage: 352.0+ bytes
```

```
In [430.. df_dbd.info()

<class 'pandas.core.frame.DataFrame'>
Index: 639 entries, 0 to 639
Data columns (total 2 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   jumlah_kasus_dbd  639 non-null     int32
1   tahun            639 non-null     int32
dtypes: int32(2)
memory usage: 10.0+ KB
```

```
In [432.. df_hpt.info()

<class 'pandas.core.frame.DataFrame'>
Index: 318 entries, 0 to 319
Data columns (total 2 columns):
#   Column          Non-Null Count  Dtype
---  ---
0   jumlah_kasus_hipertensi  318 non-null     int32
1   tahun            318 non-null     int32
dtypes: int32(2)
memory usage: 5.0+ KB
```

```
In [434.. df_kk_gjh.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 91 entries, 42 to 1388
Data columns (total 2 columns):
#   Column                Non-Null Count  Dtype
---  -
0   jumlah_kasus_kaki_gajah  91 non-null    int32
1   tahun                   91 non-null    int32
dtypes: int32(2)
memory usage: 1.4+ KB
```

In [436.. df_diare.info()

```
<class 'pandas.core.frame.DataFrame'>
Index: 322 entries, 0 to 321
Data columns (total 2 columns):
#   Column                Non-Null Count  Dtype
---  -
0   jumlah_kasus_diare    322 non-null    int32
1   tahun                 322 non-null    int32
dtypes: int32(2)
memory usage: 5.0+ KB
```

In [438.. df_hemo.info()

```
<class 'pandas.core.frame.DataFrame'>
Index: 48 entries, 0 to 47
Data columns (total 2 columns):
#   Column                Non-Null Count  Dtype
---  -
0   jumlah_kasus_hemodialisa  48 non-null    int32
1   tahun                 48 non-null    int32
dtypes: int32(2)
memory usage: 768.0+ bytes
```

In [440.. df_tbc.info()

```
<class 'pandas.core.frame.DataFrame'>
Index: 642 entries, 0 to 641
Data columns (total 2 columns):
#   Column                Non-Null Count  Dtype
---  -
0   jumlah_kasus_tuberculosis_paru  642 non-null    int32
1   tahun                 642 non-null    int32
dtypes: int32(2)
memory usage: 10.0+ KB
```

In [442.. df_mal.info()

```
<class 'pandas.core.frame.DataFrame'>
Index: 42 entries, 0 to 41
Data columns (total 2 columns):
#   Column                Non-Null Count  Dtype
---  -
0   jumlah_kasus_malaria    42 non-null    int32
1   tahun                   42 non-null    int32
dtypes: int32(2)
memory usage: 672.0+ bytes
```

Grouping atribut tahun dengan jumlah kasus penyakitnya, kemudian dijumlahkan menurut tahunnya, lalu tabel-tabel tersebut dimerged menjadi 1 tabel berdasarkan tahun

```
In [445.. df_aids_grouped = df_aids.groupby('tahun')['jumlah_kasus_aids'].sum().reset_index()
df_dbd_grouped = df_dbd.groupby('tahun')['jumlah_kasus_dbd'].sum().reset_index()
df_hpt_grouped = df_hpt.groupby('tahun')['jumlah_kasus_hipertensi'].sum().reset_index()
df_kk_gjh_grouped = df_kk_gjh.groupby('tahun')['jumlah_kasus_kaki_gajah'].sum().reset_index()
df_diare_grouped = df_diare.groupby('tahun')['jumlah_kasus_diare'].sum().reset_index()
df_hemo_grouped = df_hemo.groupby('tahun')['jumlah_kasus_hemodialisa'].sum().reset_index()
df_tbc_grouped = df_tbc.groupby('tahun')['jumlah_kasus_tuberculosis_paru'].sum().reset_index()
df_mal_grouped = df_mal.groupby('tahun')['jumlah_kasus_malaria'].sum().reset_index()

# Kemudian jika ingin menggabungkan semua dataframe menjadi satu:
df_merged1 = pd.merge(df_aids_grouped, df_dbd_grouped, on='tahun', how='outer', suffixes=('_aids', '_dbd'))
df_merged2 = pd.merge(df_merged1, df_hpt_grouped, on='tahun', how='outer', suffixes=('', '_hpt'))
df_merged3 = pd.merge(df_merged2, df_diare_grouped, on='tahun', how='outer', suffixes=('', '_diare'))
df_merged4 = pd.merge(df_merged3, df_kk_gjh_grouped, on='tahun', how='outer', suffixes=('', '_kk_gjh'))
df_merged5 = pd.merge(df_merged4, df_hemo_grouped, on='tahun', how='outer', suffixes=('', '_hemo'))
df_merged6 = pd.merge(df_merged5, df_tbc_grouped, on='tahun', how='outer', suffixes=('', '_tbc'))
df_merged7 = pd.merge(df_merged6, df_mal_grouped, on='tahun', how='outer', suffixes=('', '_mal'))
```

```
In [456.. # Ambil kolom yang diinginkan, sesuaikan dengan kolom yang sudah digabungkan
df_merged = df_merged7.loc[:, ['tahun', 'jumlah_kasus_aids', 'jumlah_kasus_dbd', 'jumlah_kasus_hipertensi', 'ju

# Tampilkan 5 baris pertama
df_merged.head(5)
```

Out[456...	tahun	jumlah_kasus_aids	jumlah_kasus_dbd	jumlah_kasus_hipertensi	jumlah_kasus_diare	jumlah_kasus_kaki_gajah	jumlah_kasus_malaria
0	2019	113	4424	722927	51267	42	
1	2020	67	2790	698686	30954	29	
2	2021	70	3743	696372	15647	20	
3	2022	31	5205	706048	26523	15	

```
In [458... # Ambil kolom yang diinginkan, sesuaikan dengan kolom yang sudah digabungkan
df_merged = df_merged7.loc[:, ['tahun', 'jumlah_kasus_tuberculosis_paru', 'jumlah_kasus_malaria']]

# Tampilkan 5 baris pertama
df_merged.head(5)
```

Out[458...	tahun	jumlah_kasus_tuberculosis_paru	jumlah_kasus_malaria
0	2019	7491	24
1	2020	2440	8
2	2021	1991	23
3	2022	2643	40

Visualisasi Data

```
In [452... # Misalkan df_merged7 adalah DataFrame yang sudah ada sebelumnya
# Hapus entri yang memiliki tahun desimal
df_merged7 = df_merged7[df_merged7['tahun'].isin([2019, 2020, 2021, 2022])]

# Set style seaborn
sns.set(style="whitegrid")

# Buat list penyakit dengan kolom yang sesuai pada df_merged7
penyakit = [
    ('Kasus AIDS', 'jumlah_kasus_aids'),
    ('Kasus DBD', 'jumlah_kasus_dbd'),
    ('Kasus Hipertensi', 'jumlah_kasus_hipertensi'),
    ('Kasus Kaki Gajah', 'jumlah_kasus_kaki_gajah'),
    ('Kasus Diare', 'jumlah_kasus_diare'),
    ('Kasus Hemodialisa', 'jumlah_kasus_hemodialisa'),
    ('Kasus TBC', 'jumlah_kasus_tuberculosis_paru'),
    ('Kasus Malaria', 'jumlah_kasus_malaria') # Tambahkan Malaria jika perlu
]

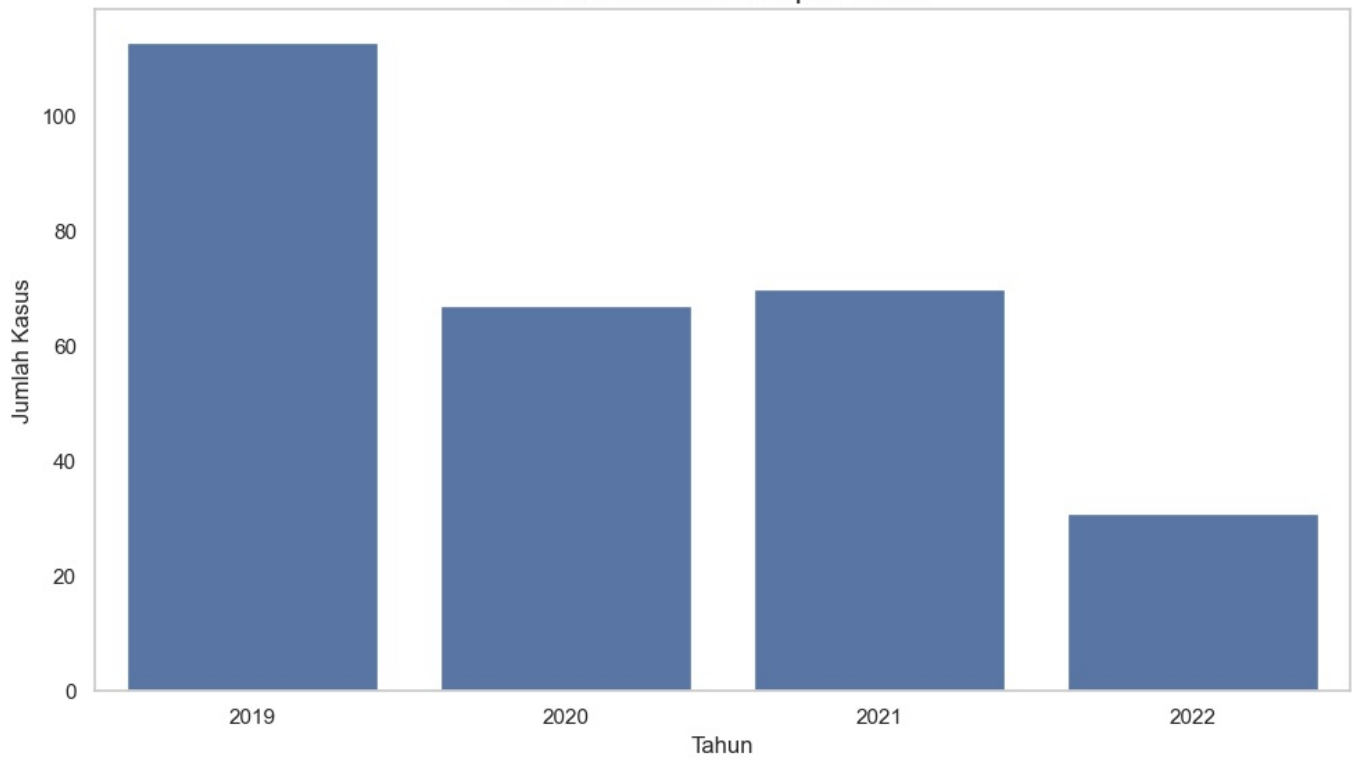
# Buat plot untuk setiap penyakit
for label, kolom in penyakit:
    plt.figure(figsize=(10, 6)) # Set ukuran figure untuk setiap plot
    sns.barplot(data=df_merged7, x='tahun', y=kolom) # Hapus parameter palette

    # Menambah judul dan label sumbu
    plt.title(f'Jumlah {label} per Tahun', fontsize=16)
    plt.xlabel('Tahun', fontsize=12)
    plt.ylabel('Jumlah Kasus', fontsize=12)

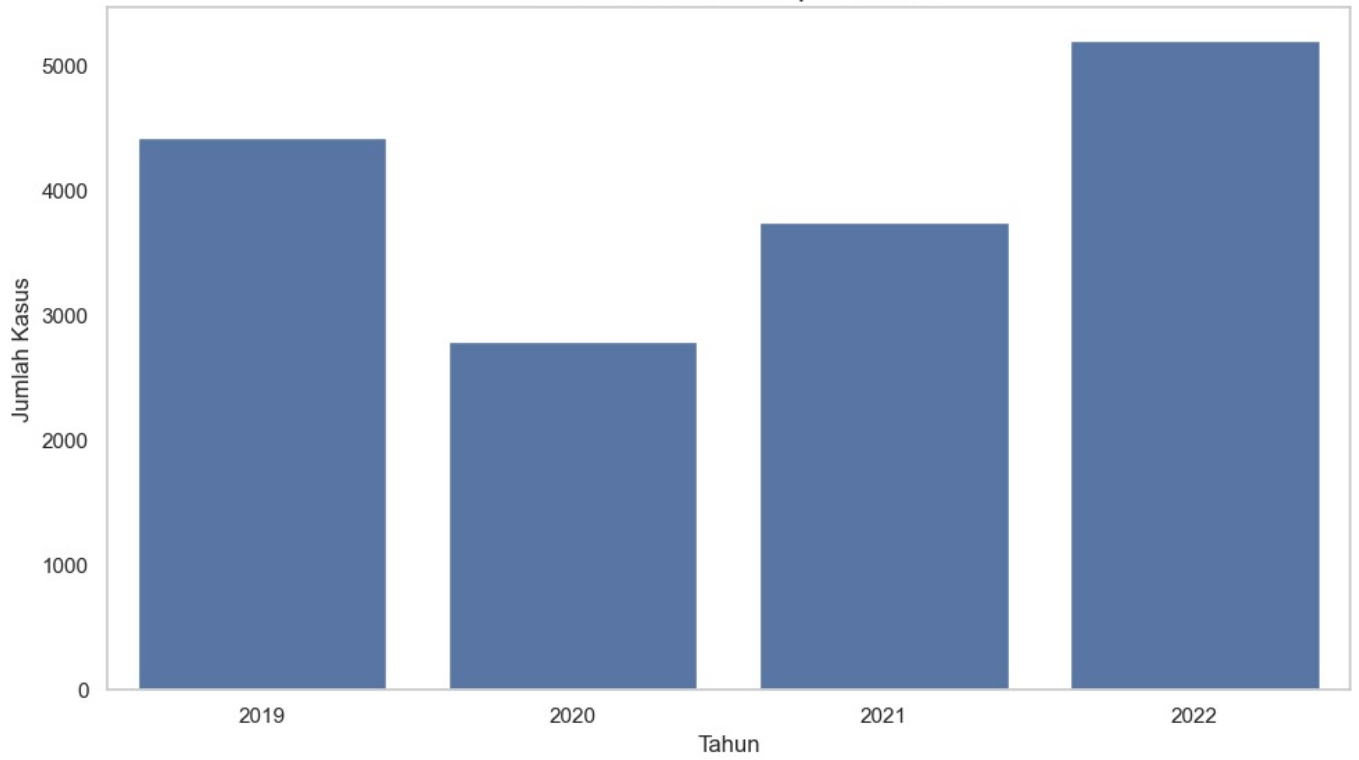
    # Menampilkan grid
    plt.grid(axis='y')

    # Tampilkan plot
    plt.xticks(rotation=0) # Rotasi label tahun untuk kejelasan
    plt.tight_layout() # Mengatur layout agar lebih rapi
    plt.show()
```

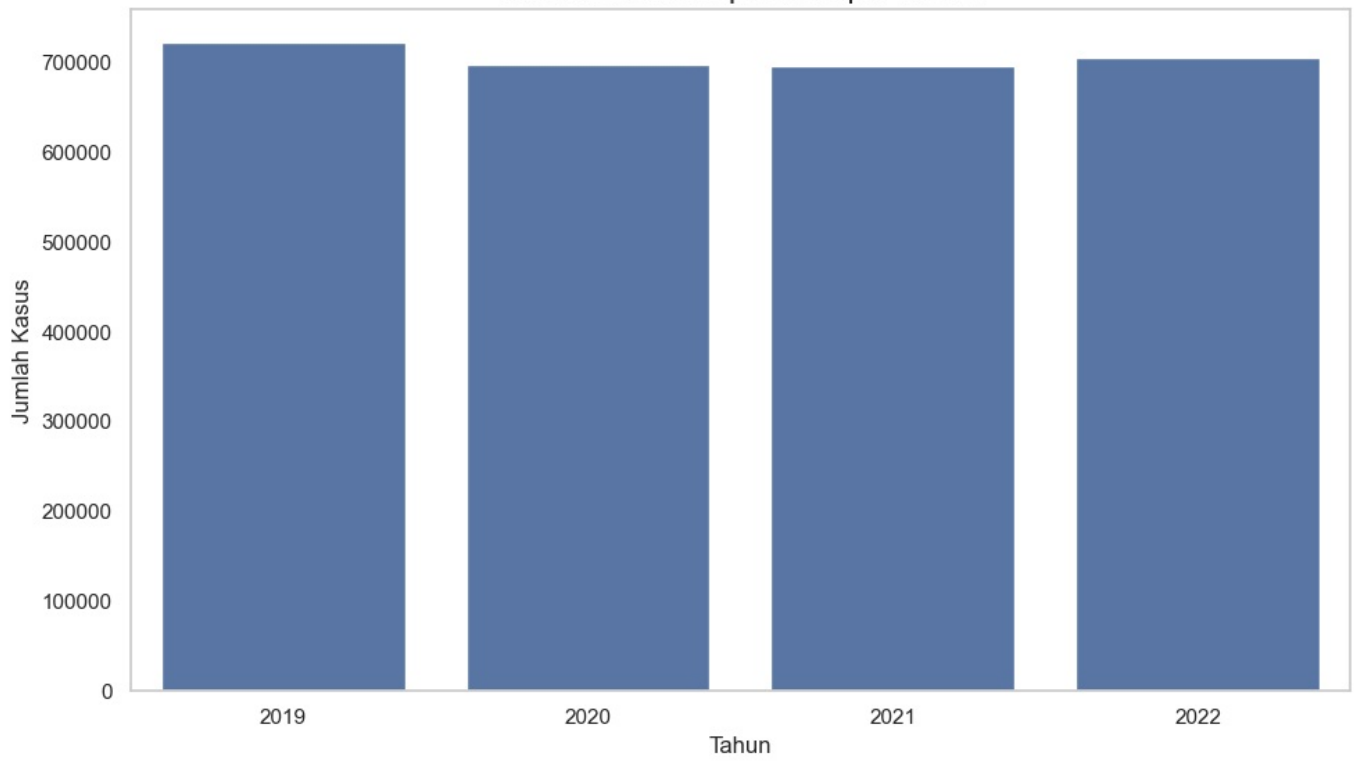

Jumlah Kasus AIDS per Tahun



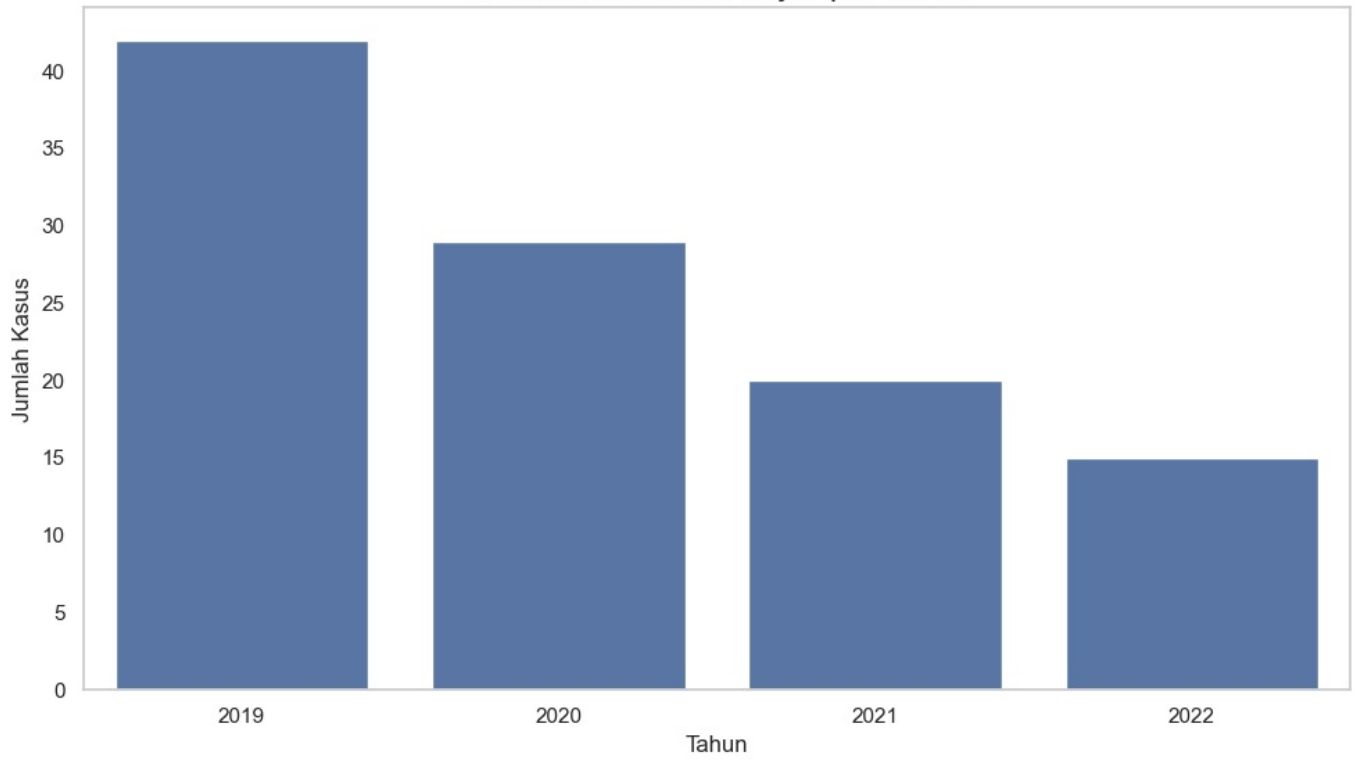
Jumlah Kasus DBD per Tahun



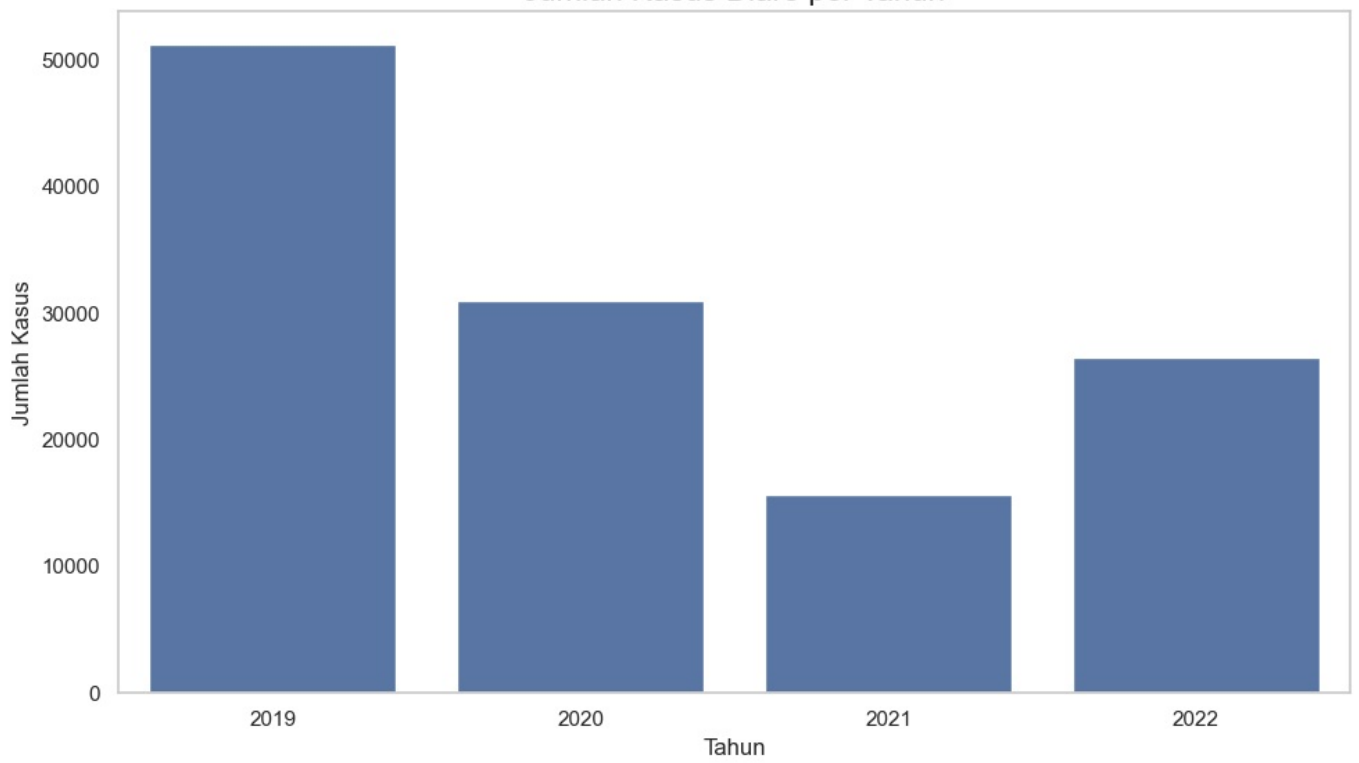
Jumlah Kasus Hipertensi per Tahun



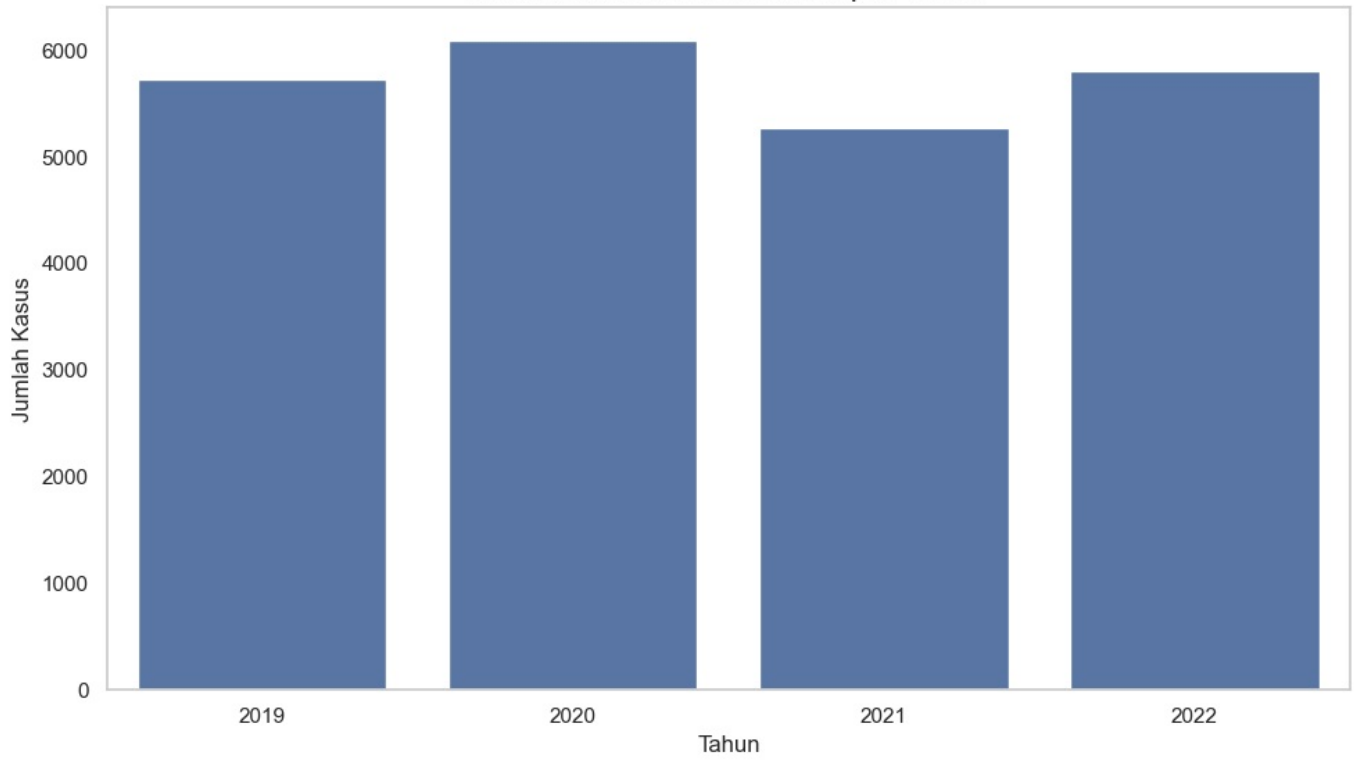
Jumlah Kasus Kaki Gajah per Tahun



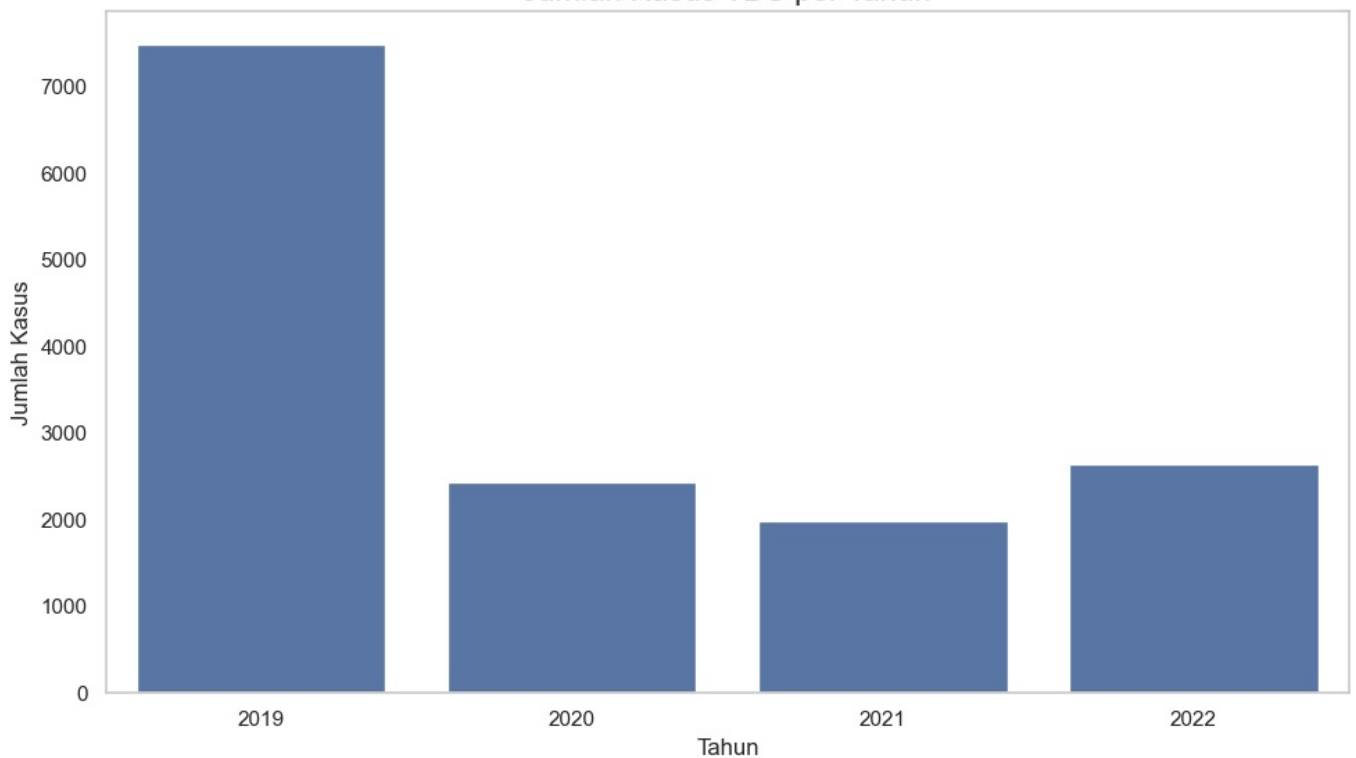
Jumlah Kasus Diare per Tahun



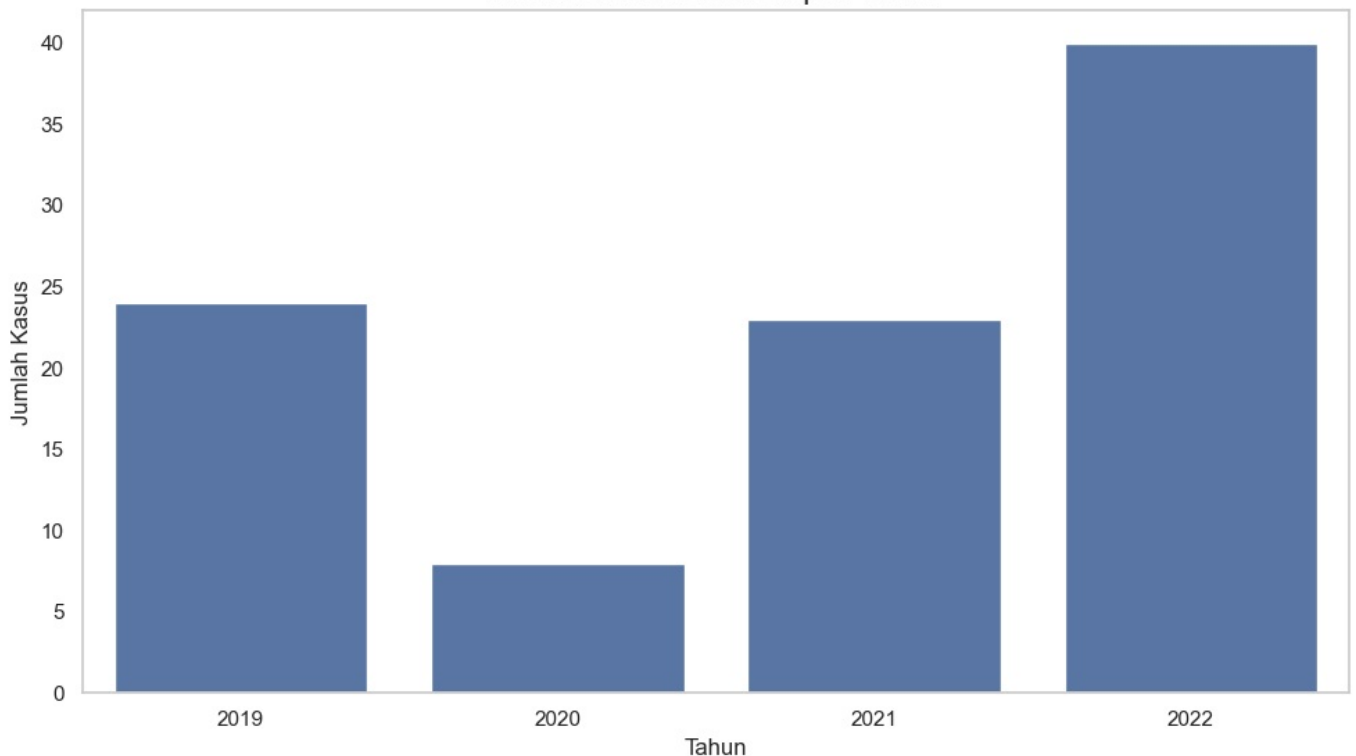
Jumlah Kasus Hemodialisa per Tahun



Jumlah Kasus TBC per Tahun



Jumlah Kasus Malaria per Tahun



AIDS:

Jika kita amati data yang divisualisasikan di grafik berdasarkan tahunnya, penyakit aids cenderung mengalami tren penurunan dari tahun 2019 hingga tahun 2022. Pada tahun 2019 terdapat sekitar 113 kasus aids di kota bandung ini, tahun 2020 terdapat sekitar 67 kasus, tahun 2021 terdapat 70 kasus dan pada akhir tahun 2022 telah menurun hingga mencapai angka 31 kasus saja. Mulai dari tahun 2019-2020 terjadi tren penurunan hingga 59,2%, kemudian pada tahun 2020-2021 ada kenaikan sedikit tren penyakit sebesar 4,3%, tahun 2021-2022 terjadi lagi tren penurunan sebesar 44,2% yang artinya pasien penderita penyakit AIDS dalam kurun waktu 4 tahun tersebut cenderung mengalami penurunan dan juga jumlah puskesmas serta rumah sakit yang ada pada data tahun 2022 bisa menjadi salah satu tolak ukur dan factor yang memengaruhi tren penurunan ini.

DBD:

Melihat dari data yang disajikan dalam grafik, tren penyakit DBD cenderung mengalami tren kenaikan dari tahun 2019 hingga tahun 2022 walaupun sempat ada penurunan pada tahun 2019-2020. Pada tahun 2019 terdapat sekitar 4424 kasus, tahun 2020 terdapat 2790 kasus, tahun 2021 terdapat 3743 kasus, dan tahun 2022 terdapat 5205 kasus. Mulai dari tahun 2019-2020 terjadi penurunan yang cukup signifikan sebesar 63%, lalu pada tahun 2020-2021 mulai terjadi tren kenaikan kasus hingga 34%, dan pada tahun 2021-2022 terdapat

kenaikan kasus sebesar 39% artinya pasien penderita penyakit DBD dalam kurun waktu 4 tahun tersebut cenderung mengalami kenaikan kasus dan juga jumlah puskesmas serta rumah sakit yang ada pada data tahun 2022 bisa menjadi tolak ukur dan factor yang memengaruhi tren kenaikan ini.

Hipertensi:

Jika kita amati data yang divisualisasikan di grafik berdasarkan tahunnya, penyakit aids cenderung mengalami tren penurunan dari tahun 2019 hingga tahun 2021 meskipun pada tahun 2022 terjadi kenaikan tidak signifikan. Pada tahun 2019 terdapat sekitar 722927 kasus hipertensi di kota bandung ini, tahun 2020 terdapat sekitar 698686 kasus, tahun 2021 terdapat 696372 kasus dan pada akhir tahun 2022 naik hingga mencapai angka 706048 kasus. Mulai dari tahun 2019-2020 terjadi tren penurunan hingga 3,4%, kemudian pada tahun 2020-2021 turun sebesar 0,34%, tahun 2021-2022 terjadi tren kenaikan sebesar 1,38% yang artinya pasien penderita penyakit Hipertensi dalam kurun waktu 4 tahun tersebut cenderung mengalami penurunan dan juga jumlah puskesmas serta rumah sakit yang ada pada data tahun 2022 bisa menjadi tolak ukur dan factor yang memengaruhi tren penurunan ini.

Kaki Gajah:

Jika kita amati data yang divisualisasikan di grafik berdasarkan tahunnya, penyakit kaki gajah selalu mengalami tren penurunan selama 4 tahun dari tahun 2019 hingga tahun 2022. Pada tahun 2019 terdapat 42 kasus, sedangkan pada tahun 2020 terdapat 29 kasus, pada tahun 2021 terdapat 20 kasus dan pada tahun 2022 terdapat 15 kasus. Mulai dari tahun 2019-2020 terjadi penurunan sebesar 30.95%, pada tahun 2020-2021 terjadi penurunan juga hingga 31.03%, dan pada tahun 2021-2022 terjadi penurunan sebesar 25.00% yang artinya pasien penderita penyakit Diare dalam kurun waktu 4 tahun tersebut cenderung mengalami penurunan dan juga jumlah puskesmas serta rumah sakit yang ada pada data tahun 2022 bisa menjadi salah satu tolak ukur dan factor yang memengaruhi tren penurunan ini.

Diare:

Jika kita amati data yang divisualisasikan di grafik berdasarkan tahunnya, penyakit diare cenderung mengalami tren penurunan dari tahun 2019 hingga tahun 2021 meskipun pada tahun 2022 terjadi tren kenaikan. Pada tahun 2019 terdapat 51267 kasus, sedangkan pada tahun 2020 terdapat 30954 kasus, pada tahun 2021 terdapat 15647 kasus dan pada tahun 2022 terdapat 26523 kasus. Mulai dari tahun 2019-2020 terjadi penurunan sebesar 39.62%, pada tahun 2020-2021 terjadi penurunan juga hingga 49.36%, dan pada tahun 2022 terjadi kenaikan sebesar 69.22% yang artinya pasien penderita penyakit Diare dalam kurun waktu 4 tahun tersebut cenderung mengalami penurunan dan juga jumlah puskesmas serta rumah sakit yang ada pada data tahun 2022 bisa menjadi salah satu tolak ukur dan factor yang memengaruhi tren penurunan ini.

Hemodialisa:

Jika kita amati data yang divisualisasikan di grafik berdasarkan tahunnya, penyakit Hemodialisa cenderung fluktuatif sepanjang tahun 2019-2022. Mulai dari tahun 2019-2020 terjadi kenaikan tren penyakit mencapai 6,35%, kemudian pada tahun 2020-2021 turun sebesar 13,56% dan pada tahun 2021-2022 naik sebesar 10,32% yang artinya pasien penderita penyakit Hemodialisa dalam kurun waktu 4 tahun tersebut fluktuatif dan juga jumlah puskesmas serta rumah sakit yang ada pada data tahun 2022 bisa menjadi tolak ukur dan faktor yang memengaruhi ke fluktuatifan data ini, walaupun tidak sepenuhnya adalah factor dari rumah sakit dan puskesmas yang ada di kota bandung.

TBC:

Jika kita amati data yang divisualisasikan di grafik berdasarkan tahunnya, penyakit TBC cenderung mengalami tren penurunan dari tahun 2019 hingga tahun 2021 meskipun pada tahun 2022 terjadi tren kenaikan. Pada tahun 2019 terdapat 7491 kasus, sedangkan pada tahun 2020 terdapat 2440 kasus, pada tahun 2021 terdapat 1991 kasus dan pada tahun 2022 terdapat 2643 kasus. Mulai dari tahun 2019-2020 terjadi penurunan sebesar 67,42%, pada tahun 2020-2021 terjadi penurunan juga hingga 18,40%, dan pada tahun 2022 terjadi kenaikan sebesar 32,75% yang artinya pasien penderita penyakit TBC dalam kurun waktu 4 tahun tersebut cenderung mengalami penurunan dan juga jumlah puskesmas serta rumah sakit yang ada pada data tahun 2022 bisa menjadi salah satu tolak ukur dan factor yang memengaruhi tren penurunan ini.

NORMALISASI PENYAKIT DENGAN JUMLAH PENDUDUK DI KOTA BANDUNG

In [145]

```
# Menjumlahkan Masing Masing Kasus Menurut Tahunnya
df_penduduk_per_tahun = df_penduduk.groupby('tahun')['jumlah_penduduk'].sum().reset_index()
df_mal_per_tahun = df_mal.groupby('tahun')['jumlah_kasus_malaria'].sum().reset_index()
df_tbc_per_tahun = df_tbc.groupby('tahun')['jumlah_kasus_tuberculosis_paru'].sum().reset_index()
df_hemo_per_tahun = df_hemo.groupby('tahun')['jumlah_kasus_hemodialisa'].sum().reset_index()
df_diare_per_tahun = df_diare.groupby('tahun')['jumlah_kasus_diare'].sum().reset_index()
df_kk_gjh_per_tahun = df_kk_gjh.groupby('tahun')['jumlah_kasus_kaki_gajah'].sum().reset_index()
df_hpt_per_tahun = df_hpt.groupby('tahun')['jumlah_kasus_hipertensi'].sum().reset_index()
df_dbd_per_tahun = df_dbd.groupby('tahun')['jumlah_kasus_dbd'].sum().reset_index()
df_aids_per_tahun = df_aids.groupby('tahun')['jumlah_kasus_aids'].sum().reset_index()
```

```

# Menggabungkan data penduduk dan penyakit berdasarkan tahun
df_merged_mal = pd.merge(df_mal_per_tahun, df_penduduk_per_tahun, on='tahun')
df_merged_tbc = pd.merge(df_tbc_per_tahun, df_penduduk_per_tahun, on='tahun')
df_merged_hemo = pd.merge(df_hemo_per_tahun, df_penduduk_per_tahun, on='tahun')
df_merged_diare = pd.merge(df_diare_per_tahun, df_penduduk_per_tahun, on='tahun')
df_merged_kk_gjh = pd.merge(df_kk_gjh_per_tahun, df_penduduk_per_tahun, on='tahun')
df_merged_hpt = pd.merge(df_hpt_per_tahun, df_penduduk_per_tahun, on='tahun')
df_merged_dbd = pd.merge(df_dbd_per_tahun, df_penduduk_per_tahun, on='tahun')
df_merged_aids = pd.merge(df_aids_per_tahun, df_penduduk_per_tahun, on='tahun')

# Normalisasi jumlah penderita penyakit dengan jumlah penduduk per tahun
df_merged_mal['persentase_penderita'] = (df_merged_mal['jumlah_kasus_malaria'] / df_merged_mal['jumlah_penduduk'])
df_merged_hpt['persentase_penderita'] = (df_merged_hpt['jumlah_kasus_hipertensi'] / df_merged_hpt['jumlah_penduduk'])
df_merged_hemo['persentase_penderita'] = (df_merged_hemo['jumlah_kasus_hemodialisa'] / df_merged_hemo['jumlah_penduduk'])
df_merged_diare['persentase_penderita'] = (df_merged_diare['jumlah_kasus_diare'] / df_merged_diare['jumlah_penduduk'])
df_merged_kk_gjh['persentase_penderita'] = (df_merged_kk_gjh['jumlah_kasus_kaki_gajah'] / df_merged_kk_gjh['jumlah_penduduk'])
df_merged_dbd['persentase_penderita'] = (df_merged_dbd['jumlah_kasus_dbd'] / df_merged_dbd['jumlah_penduduk'])
df_merged_aids['persentase_penderita'] = (df_merged_aids['jumlah_kasus_aids'] / df_merged_aids['jumlah_penduduk'])
df_merged_tbc['persentase_penderita'] = (df_merged_tbc['jumlah_kasus_tuberculosis_paru'] / df_merged_tbc['jumlah_penduduk'])

# Membuat bar chart untuk persentase penderita penyakit per tahun

# Malaria
sizes = df_merged_mal['persentase_penderita'] # Persentase kecil dari total penduduk
labels = df_merged_mal['tahun']
colors = ['#ff9999', '#66b3ff', '#99ff99', '#ffcc99']
# Membuat bar chart untuk persentase penderita per tahun
plt.figure(figsize=(10,6))
plt.bar(df_merged_mal['tahun'], df_merged_mal['persentase_penderita'], color=colors)
plt.title('Persentase Penduduk Kota Bandung yang Terkena Penyakit Malaria per Tahun')
plt.xlabel('Tahun')
plt.ylabel('Persentase (%)')
plt.show()

# Hipertensi
sizes = df_merged_hpt['persentase_penderita'] # Persentase kecil dari total penduduk
labels = df_merged_hpt['tahun']
colors = ['#ff9999', '#66b3ff', '#99ff99', '#ffcc99']
# Membuat bar chart untuk persentase penderita per tahun
plt.figure(figsize=(10,6))
plt.bar(df_merged_hpt['tahun'], df_merged_hpt['persentase_penderita'], color=colors)
plt.title('Persentase Penduduk Kota Bandung yang Terkena Penyakit Hipertensi per Tahun')
plt.xlabel('Tahun')
plt.ylabel('Persentase (%)')
plt.show()

# Hemodialisa
sizes = df_merged_hemo['persentase_penderita'] # Persentase kecil dari total penduduk
labels = df_merged_hemo['tahun']
colors = ['#ff9999', '#66b3ff', '#99ff99', '#ffcc99']
# Membuat bar chart untuk persentase penderita per tahun
plt.figure(figsize=(10,6))
plt.bar(df_merged_hemo['tahun'], df_merged_hemo['persentase_penderita'], color=colors)
plt.title('Persentase Penduduk Kota Bandung yang Terkena Penyakit Hemodialisa per Tahun')
plt.xlabel('Tahun')
plt.ylabel('Persentase (%)')
plt.show()

# Diare
sizes = df_merged_diare['persentase_penderita'] # Persentase kecil dari total penduduk
labels = df_merged_diare['tahun']
colors = ['#ff9999', '#66b3ff', '#99ff99', '#ffcc99']
# Membuat bar chart untuk persentase penderita per tahun
plt.figure(figsize=(10,6))
plt.bar(df_merged_diare['tahun'], df_merged_diare['persentase_penderita'], color=colors)
plt.title('Persentase Penduduk Kota Bandung yang Terkena Penyakit Diare per Tahun')
plt.xlabel('Tahun')
plt.ylabel('Persentase (%)')
plt.show()

# Kaki Gajah
sizes = df_merged_kk_gjh['persentase_penderita'] # Persentase kecil dari total penduduk
labels = df_merged_kk_gjh['tahun']
colors = ['#ff9999', '#66b3ff', '#99ff99', '#ffcc99']
# Membuat bar chart untuk persentase penderita per tahun
plt.figure(figsize=(10,6))
plt.bar(df_merged_kk_gjh['tahun'], df_merged_kk_gjh['persentase_penderita'], color=colors)
plt.title('Persentase Penduduk Kota Bandung yang Terkena Penyakit Kaki Gajah per Tahun')
plt.xlabel('Tahun')
plt.ylabel('Persentase (%)')
plt.show()

# DBD

```

```

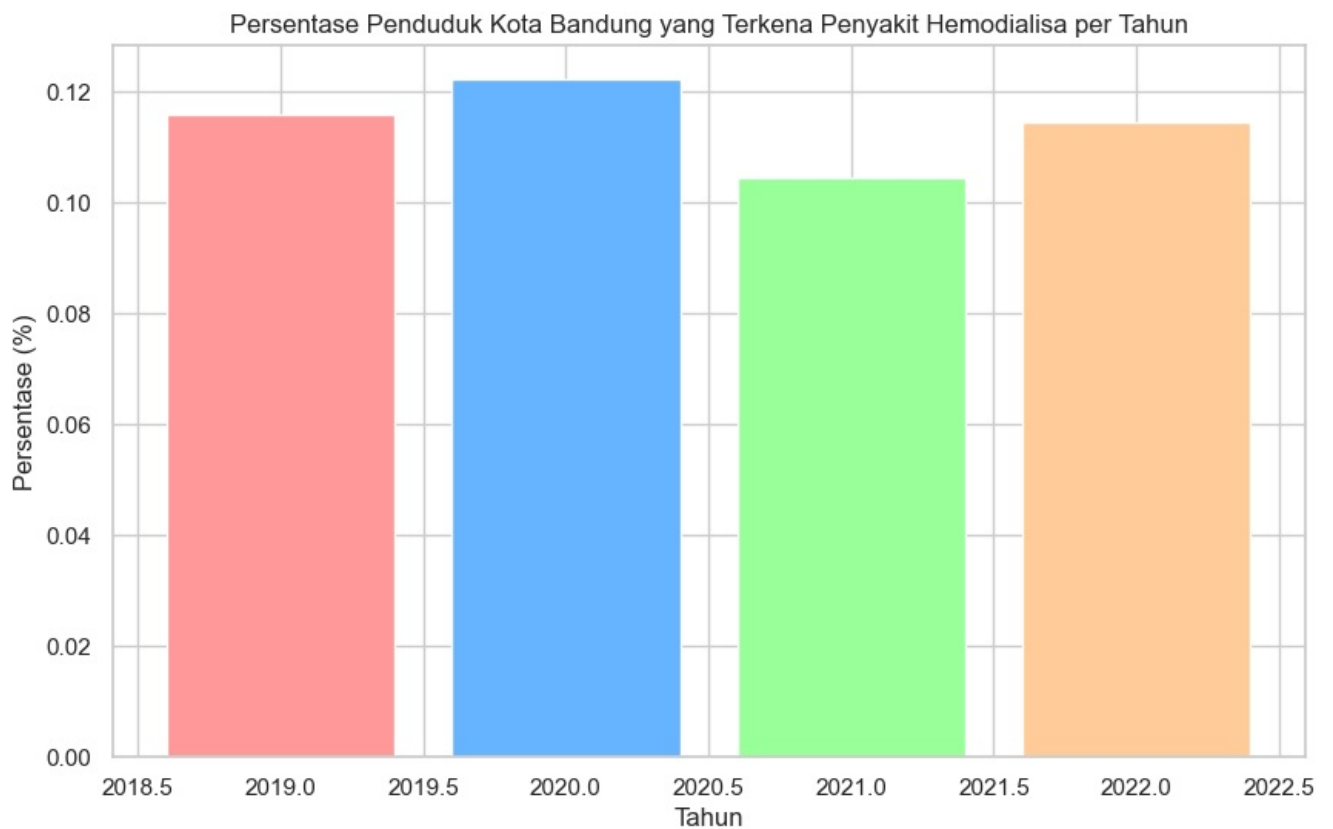
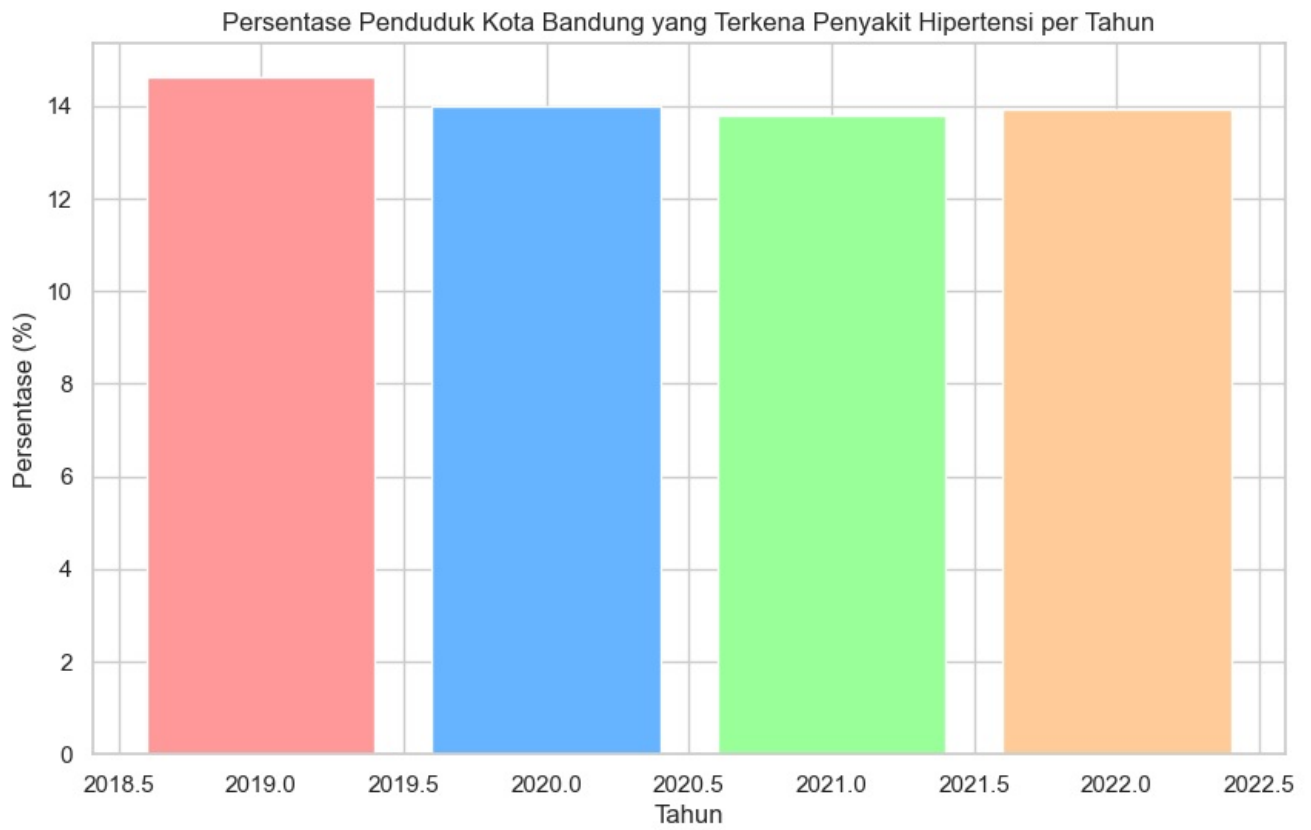
sizes = df_merged_dbd['persentase_penderita'] # Persentase kecil dari total penduduk
labels = df_merged_dbd['tahun']
colors = ['#ff9999', '#66b3ff', '#99ff99', '#ffcc99']
# Membuat bar chart untuk persentase penderita per tahun
plt.figure(figsize=(10,6))
plt.bar(df_merged_dbd['tahun'], df_merged_dbd['persentase_penderita'], color=colors)
plt.title('Persentase Penduduk Kota Bandung yang Terkena Penyakit DBD per Tahun')
plt.xlabel('Tahun')
plt.ylabel('Persentase (%)')
plt.show()

# AIDS
sizes = df_merged_aids['persentase_penderita'] # Persentase kecil dari total penduduk
labels = df_merged_aids['tahun']
colors = ['#ff9999', '#66b3ff', '#99ff99', '#ffcc99']
# Membuat bar chart untuk persentase penderita per tahun
plt.figure(figsize=(10,6))
plt.bar(df_merged_aids['tahun'], df_merged_aids['persentase_penderita'], color=colors)
plt.title('Persentase Penduduk Kota Bandung yang Terkena Penyakit AIDS per Tahun')
plt.xlabel('Tahun')
plt.ylabel('Persentase (%)')
plt.show()

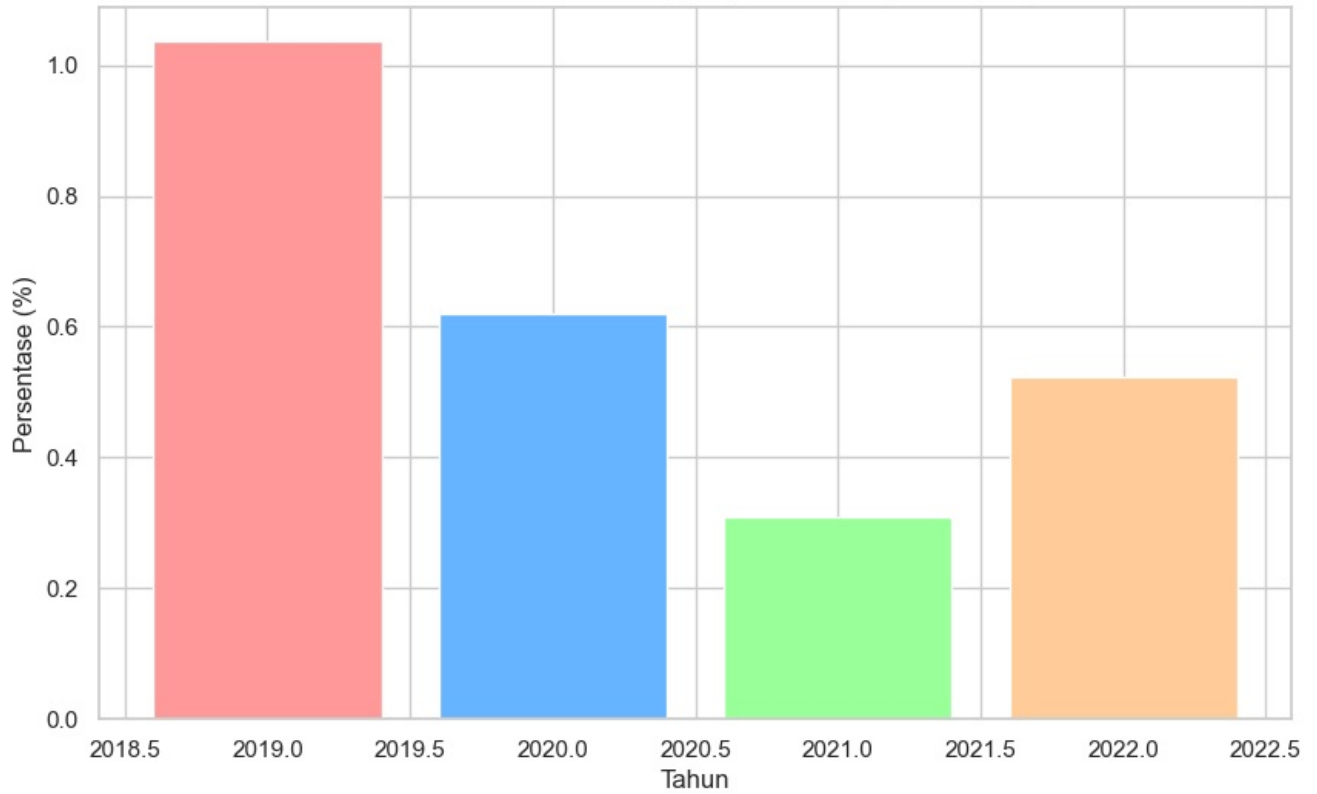
# TBC
sizes = df_merged_tbc['persentase_penderita'] # Persentase kecil dari total penduduk
labels = df_merged_tbc['tahun']
colors = ['#ff9999', '#66b3ff', '#99ff99', '#ffcc99']
# Membuat bar chart untuk persentase penderita per tahun
plt.figure(figsize=(10,6))
plt.bar(df_merged_tbc['tahun'], df_merged_tbc['persentase_penderita'], color=colors)
plt.title('Persentase Penduduk Kota Bandung yang Terkena Penyakit TBC per Tahun')
plt.xlabel('Tahun')
plt.ylabel('Persentase (%)')
plt.show()

```

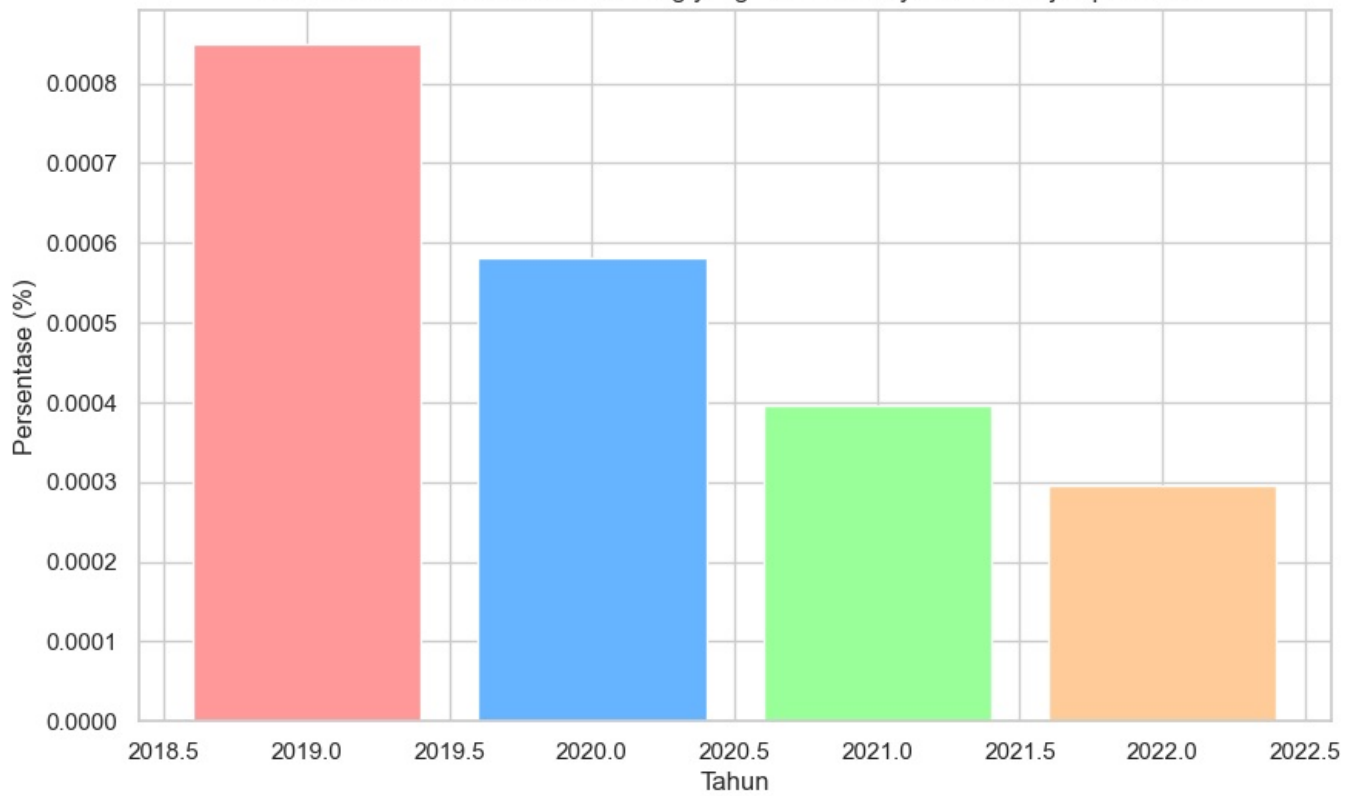


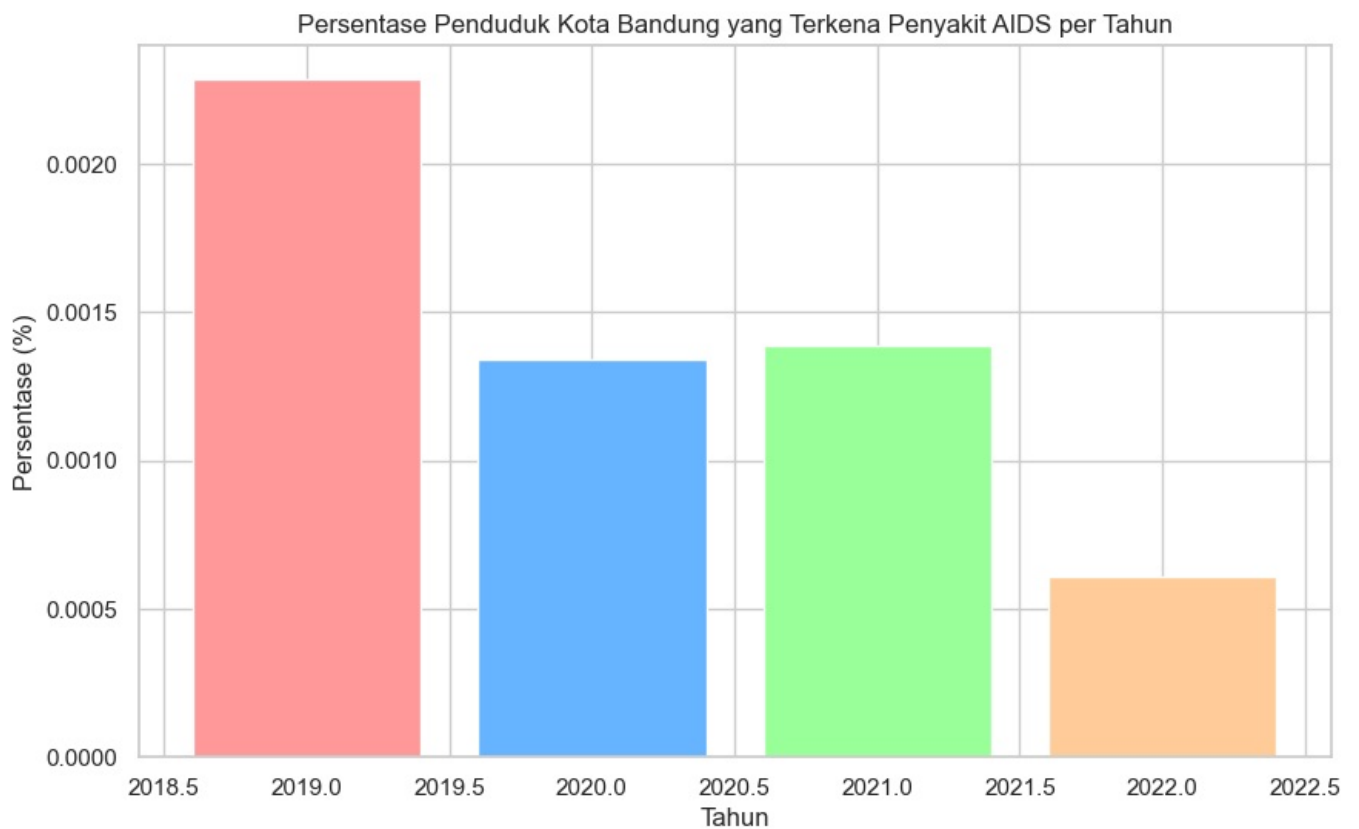
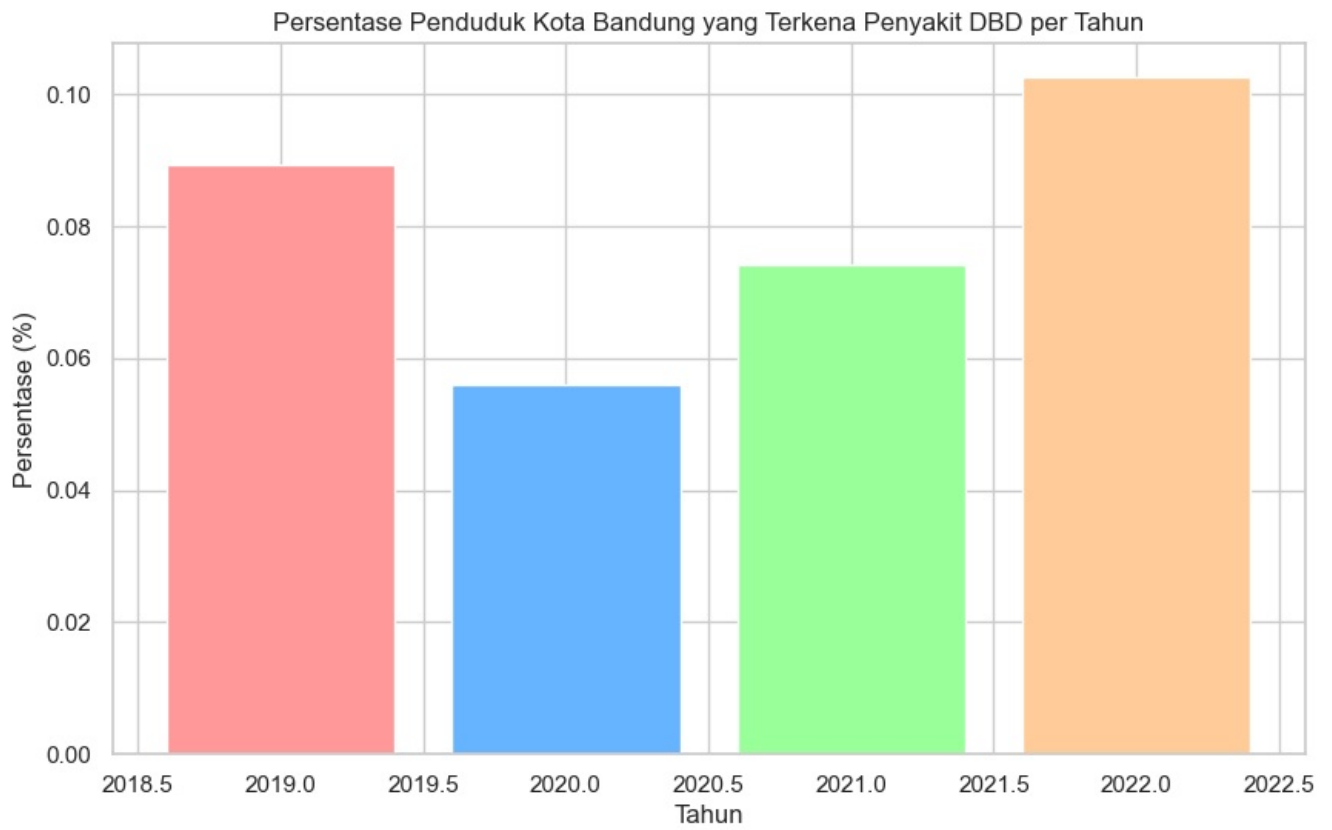


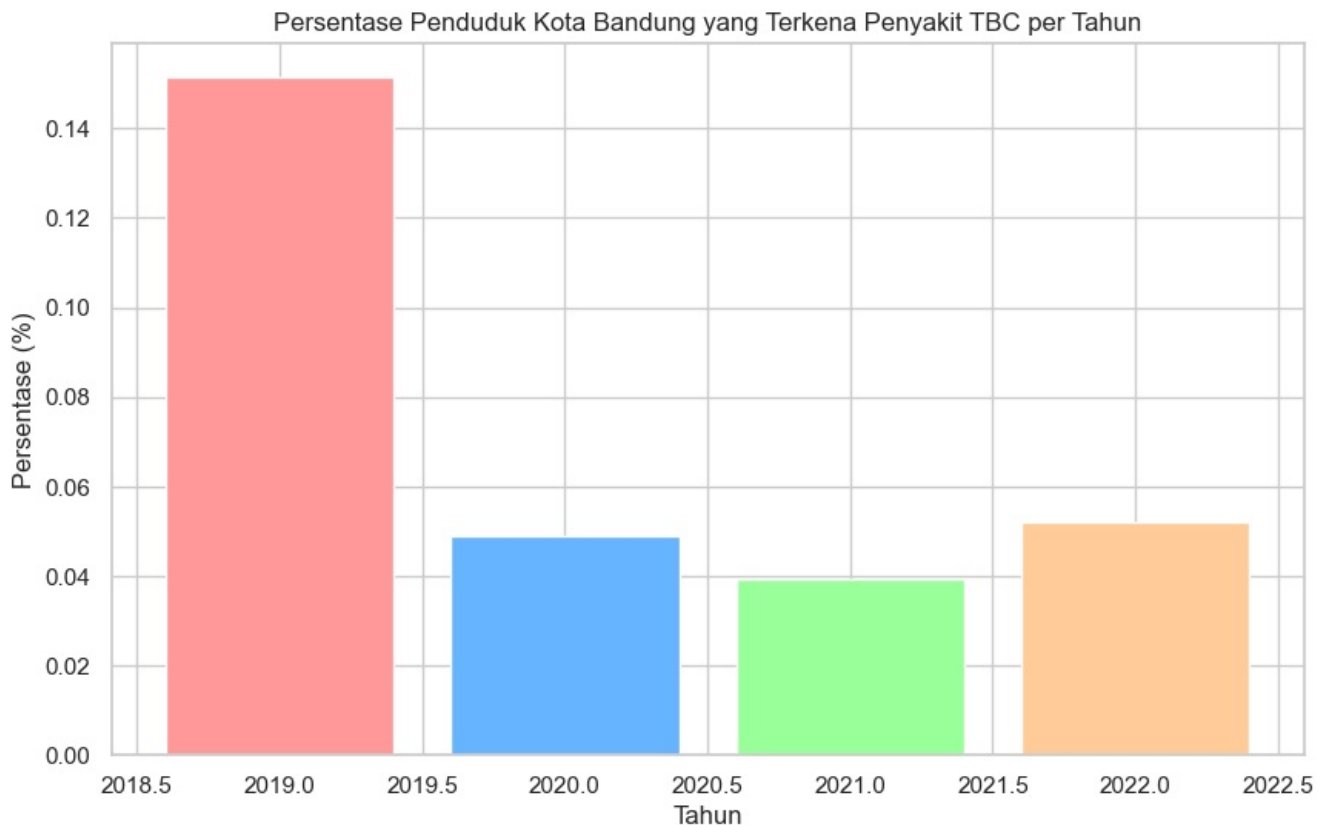
Persentase Penduduk Kota Bandung yang Terkena Penyakit Diare per Tahun



Persentase Penduduk Kota Bandung yang Terkena Penyakit Kaki Gajah per Tahun







Data yang kami visualisasikan diatas bertujuan untuk melihat persentase masing masing penyakit dibandingkan dengan jumlah penduduk Kota Bandung dari tahun 2019-2022. Dapat dilihat bahwa persentase penyakit Malaria cenderung naik tiap tahunnya, hanya pada tahun 2020 yang turun kemudian tahun selanjutnya naik lagi namun jumlah persentase nya masih tergolong sangat sedikit.

Penyakit Hipertensi memiliki persentase yang menurun walaupun turunnya tidak terlalu signifikan tetapi penyakit hipertensi memiliki jumlah persentase yang besar diantara semua data penyakit yang kita proses.

Kemudian untuk penyakit hemodialisa persentase nya cenderung fluktuatif dengan kenaikan dan penurunannya tidak terlalu signifikan. Lalu penyakit diare memiliki persentase penurunan yang cukup signifikan, hanya tahun 2022 saja yang persentasenya bertambah.

Penyakit yang selanjutnya adalah penyakit kaki gajah yang memiliki kecenderungan persentasenya menurun dalam rentang 4 tahun tersebut.

Penyakit DBD memiliki kecenderungan naik persentasenya walaupun pada tahun ke 2020 sempat terjadi penurunan. Penyakit AIDS cenderung turun persentase tiap tahunnya walaupun ada kenaikan tidak signifikan pada tahun 2021.

Terakhir untuk penyakit TBC pada tahun 2019 memiliki persentase yang tinggi daripada tahun tahun berikutnya, namun pada tahun 2020 turun signifikan dan tahun selanjutnya turun lagi, pada akhirnya pada tahun 2022 kembali naik lagi.

```
In [454]: # Menghitung jumlah penduduk per tahun
df_penduduk_per_tahun = df_penduduk.groupby('tahun')['jumlah_penduduk'].sum().reset_index()

# Menghitung jumlah kasus per tahun
df_mal_per_tahun = df_mal.groupby('tahun')['jumlah_kasus_malaria'].sum().reset_index()
df_tbc_per_tahun = df_tbc.groupby('tahun')['jumlah_kasus_tuberculosis_paru'].sum().reset_index()
df_hemo_per_tahun = df_hemo.groupby('tahun')['jumlah_kasus_hemodialisa'].sum().reset_index()
df_diare_per_tahun = df_diare.groupby('tahun')['jumlah_kasus_diare'].sum().reset_index()
df_kk_gjh_per_tahun = df_kk_gjh.groupby('tahun')['jumlah_kasus_kaki_gajah'].sum().reset_index()
df_hpt_per_tahun = df_hpt.groupby('tahun')['jumlah_kasus_hipertensi'].sum().reset_index()
df_dbd_per_tahun = df_dbd.groupby('tahun')['jumlah_kasus_dbd'].sum().reset_index()
df_aids_per_tahun = df_aids.groupby('tahun')['jumlah_kasus_aids'].sum().reset_index()

# Menggabungkan semua DataFrame berdasarkan tahun
df_merged = df_penduduk_per_tahun.copy()
df_merged = df_merged.merge(df_mal_per_tahun, on='tahun')
df_merged = df_merged.merge(df_aids_per_tahun, on='tahun')
df_merged = df_merged.merge(df_kk_gjh_per_tahun, on='tahun')
df_merged = df_merged.merge(df_tbc_per_tahun, on='tahun')
df_merged = df_merged.merge(df_hemo_per_tahun, on='tahun')
df_merged = df_merged.merge(df_diare_per_tahun, on='tahun')
df_merged = df_merged.merge(df_dbd_per_tahun, on='tahun')
df_merged = df_merged.merge(df_hpt_per_tahun, on='tahun')

# Normalisasi untuk mendapatkan persentase
df_merged['persentase_malaria'] = (df_merged['jumlah_kasus_malaria'] / df_merged['jumlah_penduduk']) * 100
```

```

df_merged['persentase_aids'] = (df_merged['jumlah_kasus_aids'] / df_merged['jumlah_penduduk']) * 100
df_merged['persentase_kk_gjh'] = (df_merged['jumlah_kasus_kaki_gajah'] / df_merged['jumlah_penduduk']) * 100
df_merged['persentase_tbc'] = (df_merged['jumlah_kasus_tubercolosis_paru'] / df_merged['jumlah_penduduk']) * 100
df_merged['persentase_hemo'] = (df_merged['jumlah_kasus_hemodialisa'] / df_merged['jumlah_penduduk']) * 100
df_merged['persentase_diare'] = (df_merged['jumlah_kasus_diare'] / df_merged['jumlah_penduduk']) * 100
df_merged['persentase_dbd'] = (df_merged['jumlah_kasus_dbd'] / df_merged['jumlah_penduduk']) * 100
df_merged['persentase_hpt'] = (df_merged['jumlah_kasus_hipertensi'] / df_merged['jumlah_penduduk']) * 100 # No.

# Membuat grafik untuk semua penyakit dalam satu grafik
bar_width = 0.1
index = np.arange(len(df_merged['tahun']))

plt.figure(figsize=(14, 6))
plt.bar(index, df_merged['persentase_malaria'], width=bar_width, label='Malaria', color='tab:olive')
plt.bar(index + bar_width, df_merged['persentase_aids'], width=bar_width, label='AIDS', color='tab:blue')
plt.bar(index + 2 * bar_width, df_merged['persentase_kk_gjh'], width=bar_width, label='Kaki Gajah', color='tab:red')
plt.bar(index + 3 * bar_width, df_merged['persentase_tbc'], width=bar_width, label='TBC', color='tab:red')
plt.bar(index + 4 * bar_width, df_merged['persentase_hemo'], width=bar_width, label='Hemodialisa', color='tab:purple')
plt.bar(index + 5 * bar_width, df_merged['persentase_diare'], width=bar_width, label='Diare', color='tab:cyan')
plt.bar(index + 6 * bar_width, df_merged['persentase_dbd'], width=bar_width, label='DBD', color='tab:orange')
plt.bar(index + 7 * bar_width, df_merged['persentase_hpt'], width=bar_width, label='Hipertensi', color='tab:pink')
plt.title('Persentase Penduduk Kota Bandung yang Terkena Penyakit')
plt.xlabel('Tahun')
plt.ylabel('Persentase (%)')
plt.xticks(index + bar_width, df_merged['tahun'], rotation=0)
plt.legend()
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()

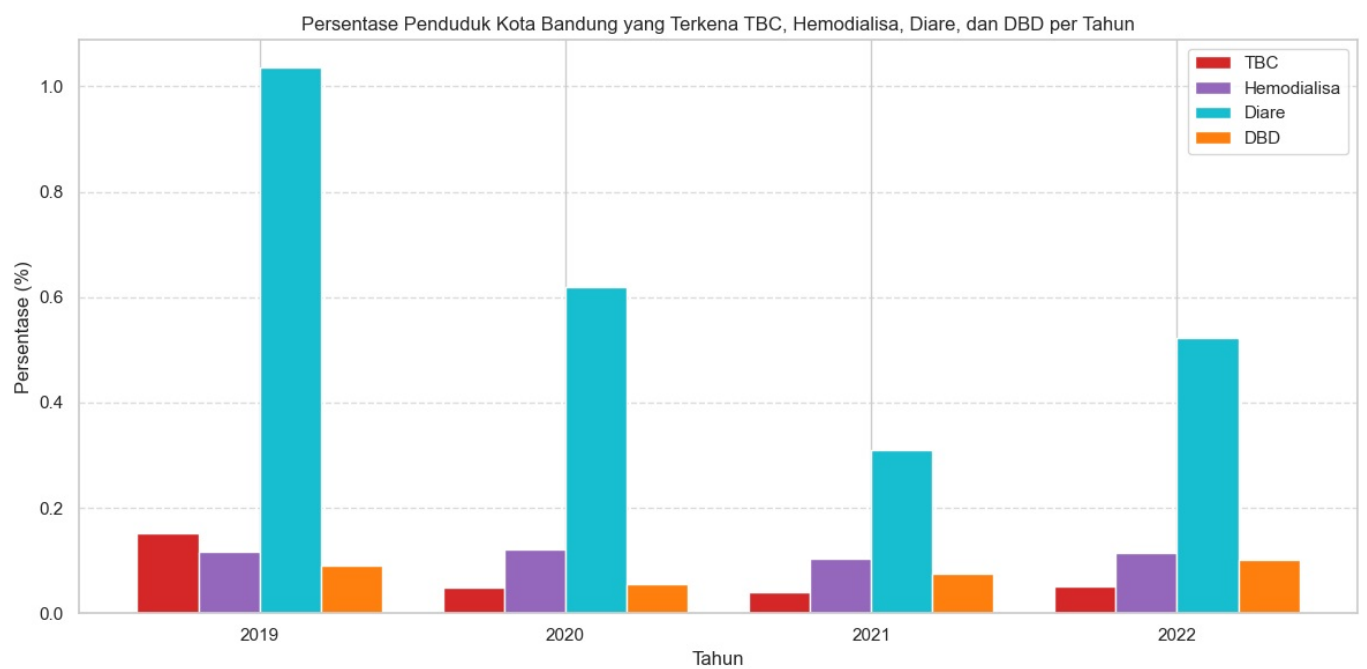
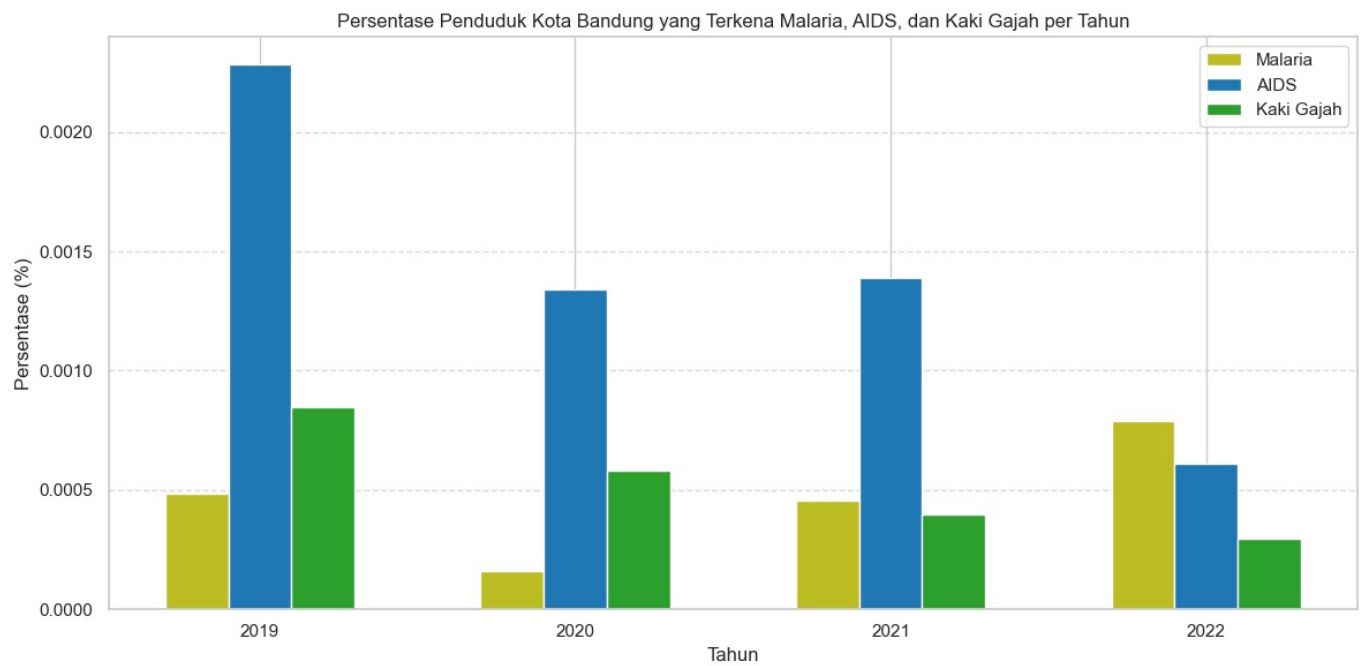
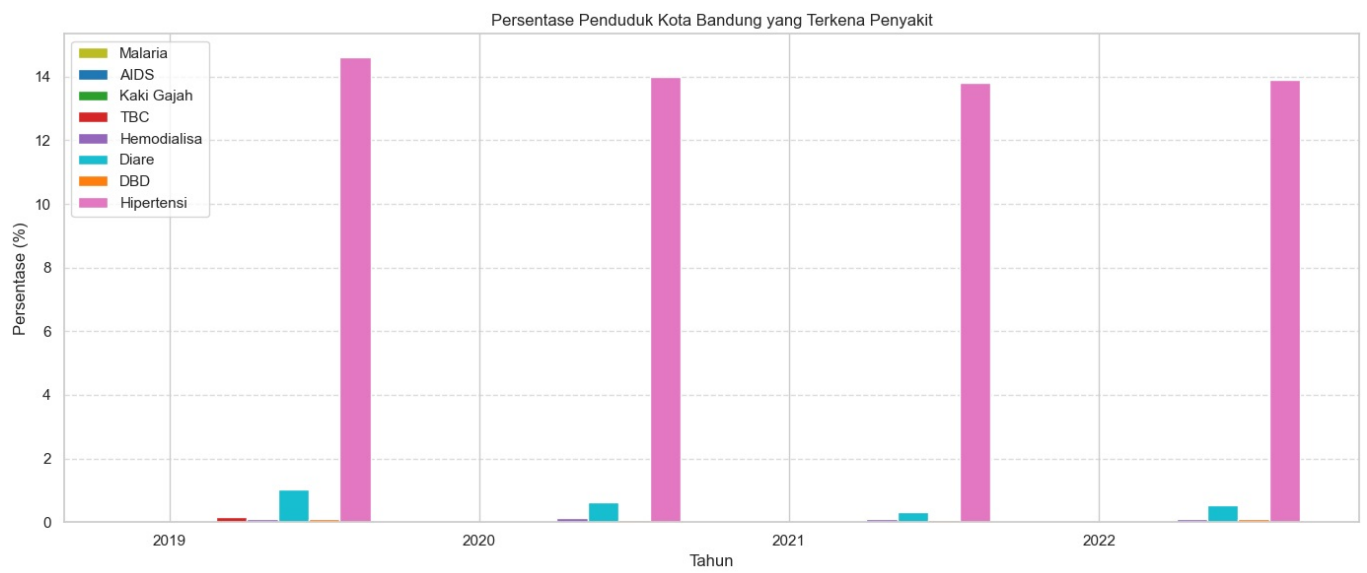
# Membuat grafik untuk Malaria, AIDS, dan Kaki Gajah
bar_width = 0.2
index = np.arange(len(df_merged['tahun']))

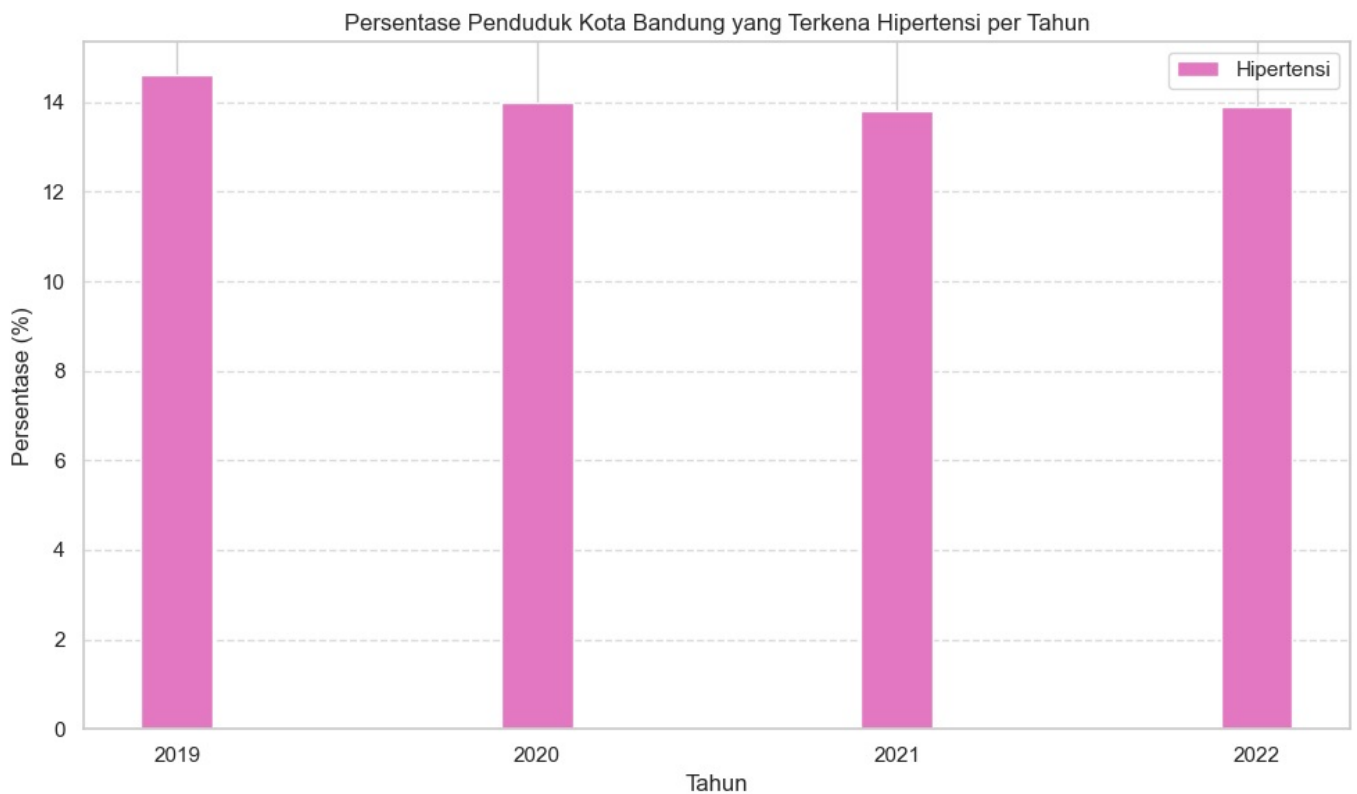
plt.figure(figsize=(12, 6))
plt.bar(index, df_merged['persentase_malaria'], width=bar_width, label='Malaria', color='tab:olive')
plt.bar(index + bar_width, df_merged['persentase_aids'], width=bar_width, label='AIDS', color='tab:blue')
plt.bar(index + 2 * bar_width, df_merged['persentase_kk_gjh'], width=bar_width, label='Kaki Gajah', color='tab:red')
plt.title('Persentase Penduduk Kota Bandung yang Terkena Malaria, AIDS, dan Kaki Gajah per Tahun')
plt.xlabel('Tahun')
plt.ylabel('Persentase (%)')
plt.xticks(index + bar_width, df_merged['tahun'], rotation=0)
plt.legend()
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()

# Membuat grafik untuk TBC, Hemodialisa, Diare, dan DBD
plt.figure(figsize=(12, 6))
plt.bar(index, df_merged['persentase_tbc'], width=bar_width, label='TBC', color='tab:red')
plt.bar(index + bar_width, df_merged['persentase_hemo'], width=bar_width, label='Hemodialisa', color='tab:purple')
plt.bar(index + 2 * bar_width, df_merged['persentase_diare'], width=bar_width, label='Diare', color='tab:cyan')
plt.bar(index + 3 * bar_width, df_merged['persentase_dbd'], width=bar_width, label='DBD', color='tab:orange')
plt.title('Persentase Penduduk Kota Bandung yang Terkena TBC, Hemodialisa, Diare, dan DBD per Tahun')
plt.xlabel('Tahun')
plt.ylabel('Persentase (%)')
plt.xticks(index + 1.5 * bar_width, df_merged['tahun'], rotation=0)
plt.legend()
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()

# Membuat grafik untuk Hipertensi
plt.figure(figsize=(10, 6))
plt.bar(index, df_merged['persentase_hpt'], width=bar_width, color='tab:pink', label='Hipertensi')
plt.title('Persentase Penduduk Kota Bandung yang Terkena Hipertensi per Tahun')
plt.xlabel('Tahun')
plt.ylabel('Persentase (%)')
plt.xticks(index, df_merged['tahun'], rotation=0)
plt.legend()
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()
plt.show()

```





Grafik diatas kami kelompokkan berdasarkan penyakit penyakit dengan tingkatan jumlah penderita dari yang paling banyak hingga yang paling jarang.Kelompok kelompok tersebut yaitu penyakit dengan tingkat jumlah penderita rendah (Malaria, Aids, Kaki Gajah), tingkat jumlah penderita sedang (TBC, Hemodialisa, Diare dan DBD), dan tingkat jumlah penderita tinggi (Hipertensi)

Grafik diatas dibuat dengan tujuan untuk melihat perbandingan masing masing persentase jumlah penderita penyakit di kota Bandung tahun 2019-2022 yang mana mempermudah untuk memberikan gambaran umum tentang tren persentase penyakit yang menjangkit warga di kota bandung dalam 4 tahun tersebut.

Untuk diagram pertama kita membandingkan semua penyakit yang ada di dataset,tetapi ada data penyakit yang terlalu tinggi dan ada juga yang terlalu sedikit sehingga ada beberapa data yang tidak terlihat karena rentang yang terlalu jauh tersebut

Untuk diagram kedua kita membandingkan 3 penyakit (Malaria, AIDS, Kaki Gajah) yang memiliki rentang persennya tidak terlalu beda jauh, yaitu rentang 0,0000-0,0025 agar terlihat jelas perbandingan di rentang tersebut.

Untuk diagram ketiga kita membandingkan 4 penyakit (TBC, Hemodialisa, Diare, DBD) yang memiliki rentang persennya tidak terlalu beda jauh juga,yaitu rentangnya sekitar 0,0-1,1 agar terlihat jelas perbandingan di rentang tersebut.

Untuk diagram keempat kita membuat hipertensi memiliki tabel sendiri dikarenakan rentangnya yang terlalu extreme dibandingkan dengan rentang penyakit lainnya.

Dengan memisahkan penyakit ke dalam beberapa grafik, kita dapat lebih mudah menganalisis tren dan perbandingan antar penyakit..

Kesimpulan

Data Data yang divisualisasikan oleh kelompok kami bertujuan untuk memberi informasi kepada pembaca tentang tren penyakit selama beberapa tahun kebelakang antara tahun 2019-2022. Kami menyediakan data data ini juga sebagai pengingat kepada pembaca akan pentingnya menjaga kesehatan walaupun dalam data tersebut ada tren penurunan dan kenaikan penyakit yang beragam,juga kami menyediakan data puskesmas beserta data rumah sakit untuk memperlihatkan kepada para pembaca dan bisa menjadi salah satu faktor yang memengaruhi terjadinya tren kenaikan atau penurunan penyakit yang terjadi di kota bandung selama kurun waktu 4 tahun dari tahun 2019 hingga tahun 2022 walaupun masih banyak factor penentu lain yang mempengaruhinya.