

ANALISA DATA KUNJUNGAN WISATAWAN DKI JAKARTA 2020

UTS IBDA 3111 - PRAPEMROSESAN DAN REKAYASA DATA
MOODY ASYER - 191900154

OVERVIEW

PRESENTASI INI TERBAGI MENJADI BEBERAPA BAGIAN:

- Latar Belakang memilih studi kasus
- paparan studi kasus
- analisa kebutuhan data
- open data/data eksternal yang digunakan
- flow proses & visualisasi data
- kesimpulan

LATAR BELAKANG MEMILIH STUDI KASUS

Pandemi virus COVID-19 telah memberi dampak pada banyak bidang di bangsa ini, pandemi virus COVID-19 diawali pada tahun 2020 dan kita mengetahui bahwa **DKI Jakarta** sebagai ibu kota adalah salah satu bagian penting yang paling **terdampak oleh virus** ini. Salah satu bidang yang sangat utama bagi DKI Jakarta adalah **pelayanan publik pariwisata** dimana wisatawan biasanya **mengunjungi kota ataupun destinasi wisata provinsi/kota ini**. Studi kasus ini saya pilih untuk melihat/menganalisa data kunjungan wisatawan nusantara dan mancanegara ke Jakarta pada tahun 2020 dimana terjadi pandemi COVID-19 pada saat itu dan pandemi ini memberi dampak kepada bidang pariwisata ini (**melihat dampak dari COVID-19 terhadap kunjungan wisatawan**).

PAPARAN STUDI KASUS

Dataset awal yang akan digunakan adalah

- Data jumlah kunjungan wisatawan mancanegara ke destinasi wisata di provinsi DKI Jakarta dari bulan Januari hingga bulan Desember tahun 2020 (12 file, tiap bulan 1 file, tiap file terdiri dari 20 baris dan 4 kolom)
- Data jumlah kunjungan wisatawan nusantara dari bulan januari hingga bulan desember tahun 2020 (12 file, tiap bulan 1 file, tiap file terdiri dari 20 baris dan 4 kolom)
- Data kunjungan wisatawan mancanegara ke provinsi DKI Jakarta bulan januari tahun 2020 melalui 2 pintu masuk (1 file, terdiri dari 2 baris dan 4 kolom)

Nantinya semua dataset kunjungan wisatawan mancanegara akan digabung menjadi satu, demikian juga dengan dataset kunjungan wisatawan nusantara sehingga menampilkan data 1 tahun. Untuk kunjungan wisatawan mancanegara ke provinsi DKI karena hanya 1 bulan maka akan dilakukan pencarian data agar membuatnya menampilkan data 1 tahun juga.

ANALISA KEBUTUHAN DATA

Untuk melakukan studi kasus maka tentunya kita perlu memilih/mencari data yang kita perlukan, serta metode/cara/teknik yang digunakan untuk mengolah data (**prapemrosesan dan rekayasa pada data tersebut**).

Pada studi kasus kali ini **terdapat 40 file data awal** di folder pariwisata, namun hanya **25 file yang akan digunakan** karena 15 data lainnya ada yang terlalu banyak kesalahan ataupun ada yang tidak terlalu memiliki hubungan dengan studi kasus kita. **Data yang digunakan adalah dataset kunjungan wisatawan mancanegara dan nusantara ke destinasi wisata Jakarta pada tiap bulan tahun 2020(24 file), serta data kunjungan wisatawan mancanegara Jakarta bulan januari melalui 2 pintu masuk (1 file).** Selain itu, teknik prapemrosesan dan rekayasa data yang akan digunakan/diterapkan adalah **data cleaning(mendeteksi nilai yang hilang, mengisi/mengubah nilai data salah input, menambah/menggabungkan dataset jika dataset awal kurang, dan mengubah bentuk dataset)** dan **feature selection (ANOVA dan Pearson's).**

OPEN DATA/DATA EKSTERNAL YANG DIGUNAKAN

Karena dataset kunjungan wisatawan mancanegara ke Jakarta bulan Januari melalui 2 pintu masuk hanya ada 2 data dan hanya satu bulan, maka **diambil data eksternal** dari <https://data.jakarta.go.id/dataset/data-kunjungan-wisatawan-mancanegara-ke-provinsi-dki-jakarta-tahun-2020-berdasarkan-2-pintu-masuk> berupa **data kunjungan wisatawan mancanegara ke Jakarta pada bulan Februari-Desember melalui 2 pintu masuk**. Sehingga datanya nanti dapat digabungkan menjadi data setahun (bukan lagi 1 bulan).

FLOW PROCESS & VISUALISASI DATA

```
# Mengimport library yang dibutuhkan
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
from sklearn.feature_selection import SelectKBest
from sklearn.feature_selection import f_classif
from sklearn.feature_selection import f_regression
from matplotlib import pyplot
```

Mengimport library yang akan digunakan

Mendefinisikan fungsi untuk memanggil
suatu dataset

```
# Fungsi memanggil dataset
def load_dataset(filename):
    dataset = pd.read_csv(filename)
    return dataset
```

FLOW PROCESS & VISUALISASI DATA

```
# Memanggil dan melihat setiap dataset

# Data yang hanya terdiri dari 1 file
# Data kunjungan wisatawan mancanegara ke provinsi DKI Jakarta bulan januari tahun 2020 melalui 2 pintu masuk
kunjungan_mancanegara_ke_DKI_januari_2020 = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-kunjungan-wisatawan-mancanegara-ke-provinsi-dki-jakarta-bulan-januari-tahun-2020-melalui-2-pintu-masuk.csv')

# Data yang terdiri dari >1 file
# Data jumlah kunjungan wisatawan mancanegara ke destinasi wisata di provinsi DKI Jakarta dari bulan Januari hingga bulan Desember tahun 2020
kwm_januari = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-jumlah-kunjungan-wisatawan-mancanegara-ke-destinasi-wisata-di-provinsi-dki-jakarta-dari-bulan-januari-hingga-bulan-desember-tahun-2020-januari.csv')
kwm_februari = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-jumlah-kunjungan-wisatawan-mancanegara-ke-destinasi-wisata-di-provinsi-dki-jakarta-dari-bulan-januari-hingga-bulan-desember-tahun-2020-februari.csv')
kwm_maret = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-jumlah-kunjungan-wisatawan-mancanegara-ke-destinasi-wisata-di-provinsi-dki-jakarta-dari-bulan-januari-hingga-bulan-desember-tahun-2020-maret.csv')
kwm_april = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-jumlah-kunjungan-wisatawan-mancanegara-ke-destinasi-wisata-di-provinsi-dki-jakarta-dari-bulan-januari-hingga-bulan-desember-tahun-2020-april.csv')
kwm_mei = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-jumlah-kunjungan-wisatawan-mancanegara-ke-destinasi-wisata-di-provinsi-dki-jakarta-dari-bulan-januari-hingga-bulan-desember-tahun-2020-mei.csv')
kwm_juni = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-jumlah-kunjungan-wisatawan-mancanegara-ke-destinasi-wisata-di-provinsi-dki-jakarta-dari-bulan-januari-hingga-bulan-desember-tahun-2020-juni.csv')
kwm_juli = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-jumlah-kunjungan-wisatawan-mancanegara-ke-destinasi-wisata-di-provinsi-dki-jakarta-dari-bulan-januari-hingga-bulan-desember-tahun-2020-juli.csv')
kwm_agustus = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-jumlah-kunjungan-wisatawan-mancanegara-ke-destinasi-wisata-di-provinsi-dki-jakarta-dari-bulan-januari-hingga-bulan-desember-tahun-2020-agustus.csv')
kwm_september = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-jumlah-kunjungan-wisatawan-mancanegara-ke-destinasi-wisata-di-provinsi-dki-jakarta-dari-bulan-januari-hingga-bulan-desember-tahun-2020-september.csv')
kwm_oktober = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-jumlah-kunjungan-wisatawan-mancanegara-ke-destinasi-wisata-di-provinsi-dki-jakarta-dari-bulan-januari-hingga-bulan-desember-tahun-2020-oktober.csv')
kwm_november = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-jumlah-kunjungan-wisatawan-mancanegara-ke-destinasi-wisata-di-provinsi-dki-jakarta-dari-bulan-januari-hingga-bulan-desember-tahun-2020-november.csv')
kwm_desember = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-jumlah-kunjungan-wisatawan-mancanegara-ke-destinasi-wisata-di-provinsi-dki-jakarta-dari-bulan-januari-hingga-bulan-desember-tahun-2020-desember.csv')

# Data jumlah kunjungan wisatawan nusantara dari bulan januari hingga bulan desember tahun 2020
kwn_januari = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-jumlah-kunjungan-wisatawan-nusantara-bulan-januari-tahun-2020.csv')
kwn_februari = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-jumlah-kunjungan-wisatawan-nusantara-bulan-februari-tahun-2020.csv')
kwn_maret = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-jumlah-kunjungan-wisatawan-nusantara-bulan-maret-tahun-2020.csv')
kwn_april = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-jumlah-kunjungan-wisatawan-nusantara-bulan-april-tahun-2020.csv')
kwn_mei = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-jumlah-kunjungan-wisatawan-nusantara-bulan-mei-tahun-2020.csv')
kwn_juni = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-jumlah-kunjungan-wisatawan-nusantara-bulan-juni-tahun-2020.csv')
kwn_juli = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-jumlah-kunjungan-wisatawan-nusantara-bulan-juli-tahun-2020.csv')
kwn_agustus = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-jumlah-kunjungan-wisatawan-nusantara-bulan-agustus-tahun-2020.csv')
kwn_september = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-jumlah-kunjungan-wisatawan-nusantara-bulan-september-tahun-2020.csv')
kwn_oktober = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-jumlah-kunjungan-wisatawan-nusantara-bulan-oktober-tahun-2020.csv')
kwn_november = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-jumlah-kunjungan-wisatawan-nusantara-bulan-november-tahun-2020.csv')
kwn_desember = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-jumlah-kunjungan-wisatawan-nusantara-bulan-desember-tahun-2020.csv')
```

Memanggil
setiap
dataset

FLOW PROCESS & VISUALISASI DATA

Menggabungkan dataset kunjungan wisatawan (mancanegara/nusantara) dari bulan januari-desember

```
# Menggabungkan dataset kunjungan wisatawan mancanegara maupun nusantara ke destinasi wisata DKI Jakarta dari Januari-Desember 2020
data_kunjungan_wisatawan_mancanegara = pd.concat([kwm_januari, kwm_februari, kwm_maret, kwm_april, kwm_mei,
                                                  kwm_juni, kwm_juli, kwm_agustus, kwm_september, kwm_oktober,
                                                  kwm_november, kwm_desember], ignore_index=True)

data_kunjungan_wisatawan_nusantara = pd.concat([kwn_januari, kwn_februari, kwn_maret, kwn_april, kwn_mei,
                                                  kwn_juni, kwn_juli, kwn_agustus, kwn_september, kwn_oktober,
                                                  kwn_november, kwn_desember], ignore_index=True)
```

data_kunjungan_wisatawan_nusantara

	tahun	bulan	destinasi_wisata	jumlah
0	2020	1	Taman Impian Jaya Ancol	620327.000
1	2020	1	Taman Mini Indonesia Indah	272960.000
2	2020	1	Taman Margasatwa Ragunan	289748.000
3	2020	1	Monumen Nasional	24.325
4	2020	1	Museum Nasional	18145.000
5	2020	1	Museum Satria Mandala	2046.000
6	2020	1	Museum Sejarah Jakarta	66251.000
7	2020	1	Museum Tekstil	2658.000
8	2020	1	Museum Bahari	2087.000
9	2020	1	Museum Seni Rupa dan Keramik	12190.000
10	2020	1	Museum Wayang	21231.000
44	2020	1	Museum Ikon	584.000

data_kunjungan_wisatawan_mancanegara

	tahun	bulan	destinasi_wisata	jumlah
0	2020	1	Taman Impian Jaya Ancol	516
1	2020	1	Taman Mini Indonesia Indah	0
2	2020	1	Taman Margasatwa Ragunan	449
3	2020	1	Monumen Nasional	2070
4	2020	1	Museum Nasional	2877
5	2020	1	Museum Satria Mandala	0
6	2020	1	Museum Sejarah Jakarta	1831
7	2020	1	Museum Tekstil	312
8	2020	1	Museum Bahari	274
9	2020	1	Museum Seni Rupa dan Keramik	300
10	2020	1	Museum Wayang	865
44	2020	1	Museum Ikon	0

kunjungan_mancanegara_ke_DKI_januari_2020

	tahun	bulan	kedatangan_wisata_mancanegara	jumlah
0	2020	1	Bandara Soekarno-Hatta	173453
1	2020	1	Bandara Halim PK	161

FLOW PROCESS & VISUALISASI DATA

```
# memanggil dataset baru untuk ditambahkan
kunjungan_m mancanegara_ke DKI_februari_2020 = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-kunjungan-wisatawan-mancanegara-ke-DKI-februari-2020')
kunjungan_m mancanegara_ke DKI_maret_2020 = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-kunjungan-wisatawan-mancanegara-ke-DKI-maret-2020')
kunjungan_m mancanegara_ke DKI_april_2020 = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-kunjungan-wisatawan-mancanegara-ke-DKI-april-2020')
kunjungan_m mancanegara_ke DKI_mei_2020 = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-kunjungan-wisatawan-mancanegara-ke-DKI-mei-2020')
kunjungan_m mancanegara_ke DKI_juni_2020 = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-kunjungan-wisatawan-mancanegara-ke-DKI-juni-2020')
kunjungan_m mancanegara_ke DKI_juli_2020 = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-kunjungan-wisatawan-mancanegara-ke-DKI-juli-2020')
kunjungan_m mancanegara_ke DKI_agustus_2020 = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-kunjungan-wisatawan-mancanegara-ke-DKI-agustus-2020')
kunjungan_m mancanegara_ke DKI_september_2020 = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-kunjungan-wisatawan-mancanegara-ke-DKI-september-2020')
kunjungan_m mancanegara_ke DKI_oktober_2020 = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-kunjungan-wisatawan-mancanegara-ke-DKI-oktober-2020')
kunjungan_m mancanegara_ke DKI_november_2020 = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-kunjungan-wisatawan-mancanegara-ke-DKI-november-2020')
kunjungan_m mancanegara_ke DKI_desember_2020 = load_dataset('Public Service/DKI Jakarta/Pariwisata/data-kunjungan-wisatawan-mancanegara-ke-DKI-desember-2020')

# menggabungkan setiap dataset yang baru dengan yang awal(kunjungan_m mancanegara_ke DKI_januari_2020)
kunjungan_m mancanegara_2020 = pd.concat([kunjungan_m mancanegara_ke DKI_januari_2020,
                                           kunjungan_m mancanegara_ke DKI_februari_2020,
                                           kunjungan_m mancanegara_ke DKI_maret_2020,
                                           kunjungan_m mancanegara_ke DKI_april_2020,
                                           kunjungan_m mancanegara_ke DKI_mei_2020,
                                           kunjungan_m mancanegara_ke DKI_juni_2020,
                                           kunjungan_m mancanegara_ke DKI_juli_2020,
                                           kunjungan_m mancanegara_ke DKI_agustus_2020,
                                           kunjungan_m mancanegara_ke DKI_september_2020,
                                           kunjungan_m mancanegara_ke DKI_oktober_2020,
                                           kunjungan_m mancanegara_ke DKI_november_2020,
                                           kunjungan_m mancanegara_ke DKI_desember_2020], ignore_index=True)
```

Memanggil dan menggabungkan dataset kunjungan wisatawan mancanegara ke daerah provinsi Jakarta dari bulan januari–desember

	tahun	bulan	kedatangan_wisata_mancanegara	jumlah
0	2020	1	Bandara Soekarno-Hatta	173453
1	2020	1	Bandara Halim PK	161
2	2020	2	Bandara Soekarno-Hatta	131032
3	2020	2	Bandara Halim PK	205
4	2020	3	Bandara Soekarno-Hatta	52631
5	2020	3	Bandara Halim PK	90
6	2020	4	Bandara Soekarno-Hatta	417
7	2020	4	Bandara Halim PK	7
8	2020	5	Bandara Soekarno-Hatta	395
9	2020	5	Bandara Halim PK	19
10	2020	6	Bandara Soekarno-Hatta	909
11	2020	6	Bandara Halim PK	15
12	2020	7	Bandara Soekarno-Hatta	3119
13	2020	7	Bandara Halim PK	27
14	2020	8	Bandara Soekarno-Hatta	4469
15	2020	8	Bandara Halim PK	18
16	2020	9	Bandara Soekarno-Hatta	7468
17	2020	9	Bandara Halim PK	60
18	2020	10	Bandara Soekarno-Hatta	10492
19	2020	10	Bandara Halim PK	37
20	2020	11	Bandara Soekarno-Hatta	14309
21	2020	11	Bandara Halim PK	56
22	2020	12	Bandara Soekarno-Hatta	21812
23	2020	12	Bandara Halim PK	48

FLOW PROCESS & VISUALISASI DATA

- Mengubah kesalahan input pada data jumlah kunjungan wisatawan nusantara maupun mancanegara dari bulan januari hingga bulan desember tahun 2020

kunjungan wisatawan nusantara:

- mengganti tipe data kolom 'jumlah' dari float menjadi integer
- nilai 'jumlah' pada data baris index ke 3, sehingga harus diubah ke kondisi data yang benar (24.325 -> 24325)
- kunjungan data destinasi wisata bulan ke 11 (museum joang -> museum joang 45)

kunjungan wisatawan mancanegara:

- nama destinasi museum joang -> museum joang 45

Mengubah/mengganti data

```
data_kunjungan_wisatawan_nusantara['jumlah'].replace(24.325, 24325, inplace=True)
data_kunjungan_wisatawan_nusantara['jumlah'] = data_kunjungan_wisatawan_nusantara['jumlah'].astype(int)
data_kunjungan_wisatawan_nusantara['destinasi_wisata'].replace('Museum Joang', 'Museum Joang 45', inplace=True)
data_kunjungan_wisatawan_mancanegara['destinasi_wisata'].replace('Museum Joang', 'Museum Joang 45', inplace=True)
```

data_kunjungan_wisatawan_nusantara

FLOW PROCESS & VISUALISASI DATA

```
# Memastikan/mencari apakah masih ada data nan
data_kunjungan_wisatawan_mancanegara.isna().sum()

tahun          0
bulan          0
destinasi_wisata  0
jumlah        0
dtype: int64

data_kunjungan_wisatawan_nusantara.isna().sum()

tahun          0
bulan          0
destinasi_wisata  0
jumlah        0
dtype: int64

kunjungan_mancanegara_2020.isna().sum()

tahun          0
bulan          0
kedatangan_wisata_mancanegara  0
jumlah        0
dtype: int64

# Memastikan/mencari apakah masih ada data 0
(data_kunjungan_wisatawan_mancanegara == 0).astype(int).sum()

tahun          0
bulan          0
destinasi_wisata  0
jumlah        144
dtype: int64

(data_kunjungan_wisatawan_nusantara == 0).astype(int).sum()

tahun          0
bulan          0
destinasi_wisata  0
jumlah         59
dtype: int64

(kunjungan_mancanegara_2020 == 0).astype(int).sum()

tahun          0
bulan          0
kedatangan_wisata_mancanegara  0
jumlah         0
dtype: int64
```

Mengecek missing values (0 dan NaN) untuk memastikan apakah masih ada data yang harus diubah/diisi

FLOW PROCESS & VISUALISASI DATA

Mengubah semua bentuk dataset yang ada agar lebih mudah dilihat dan dianalisis isinya

```
# Mengubah bentuk dataset  
data_kunjungan_wisatawan_nusantara = pd.pivot(data_kunjungan_wisatawan_nusantara, index='destinasi_wisata', columns='bulan', values='jumlah_kunjungan')  
data_kunjungan_wisatawan_nusantara
```

```
data_kunjungan_wisatawan_mancanegara = pd.pivot(data_kunjungan_wisatawan_mancanegara, index='destinasi_wisata', columns='bulan', values='jumlah_kunjungan')  
data_kunjungan_wisatawan_mancanegara
```

```
kunjungan_mancanegara_2020 = pd.pivot(kunjungan_mancanegara_2020, index='bulan', columns='kedatangan_wisata_mancanegara', values='jumlah_kunjungan')  
kunjungan_mancanegara_2020
```

FLOW PROCESS & VISUALISASI DATA

	destinasi_wisata	1	2	3	4	5	6	7	8	9	10	11	12
0	Kawasan Kota Tua	456902	587641	121516	0	0	2619	3925	3386	1837	8618	3354	0
1	Kepulauan Seribu	31041	63568	17948	46	0	8859	20360	37986	15270	14944	18301	22289
2	Monumen Nasional	24325	146042	50328	0	0	0	0	0	0	443034	0	0
3	Museum Bahari	2087	2102	1209	0	0	173	632	887	269	383	729	654
4	Museum Joang 45	581	1284	306	0	0	94	115	359	113	249	267	228
5	Museum Nasional	18145	27928	12820	0	0	173	583	233	105	65308	1451	1418
6	Museum Prasasti	1030	973	268	0	0	104	194	280	125	203	417	364
7	Museum Satria Mandala	2046	518	619	0	0	0	0	0	0	0	0	0
8	Museum Sejarah Jakarta	66251	59895	8947	0	0	982	2066	1857	769	2864	5610	4083
9	Museum Seni Rupa dan Keramik	12190	11800	2666	0	0	233	327	321	237	422	1003	627
10	Museum Tekstil	2658	3049	1206	0	0	190	132	162	80	96	161	203
11	Museum Wayang	21231	14728	2564	0	0	432	819	506	422	1390	2728	1788
12	PBB Setu babakan	35369	28848	11850	0	0	6057	1548	21997	4704	10372	17181	9883
13	Pelabuhan Sunda Kelapa	931	1895	1304	380	0	753	927	2535	2155	2232	2752	2296
14	Planetarium	9654	10931	1693	0	0	0	0	0	0	0	0	0
15	Rumah si Pitung	2048	1811	672	0	0	788	1722	2597	1018	1342	2316	1238
16	Taman Arkeologi Onrust	2200	3236	951	0	0	948	3711	4257	2969	2783	4186	1621
17	Taman Impian Jaya Ancol	620327	638729	358871	0	0	28175	85142	174988	71688	1978996	126744	479272
18	Taman Margasatwa Ragunan	289748	191179	82706	0	0	3837	9365	16570	6109	600470	21741	19012
19	Taman Mini Indonesia Indah	272960	317391	89617	0	0	28090	82696	134546	36977	962277	107810	105008

	bulan	Bandara Halim PK	Bandara Soekarno-Hatta
0	1	161	173453
1	2	205	131032
2	3	90	52631
3	4	7	417
4	5	19	395
5	6	15	909
6	7	27	3119
7	8	18	4469
8	9	60	7468
9	10	37	10492
10	11	56	14309
11	12	46	21812

[illegible]

FLOW PROCESS & VISUALISASI DATA

```
# Menambah kolom/membuat fitur baru tambahan untuk melakukan seleksi fitur
data_kunjungan_wisatawan_nusantara['total'] = 0
data_kunjungan_wisatawan_nusantara['total'] = data_kunjungan_wisatawan_nusantara.iloc[:, 1:].sum(axis=1)
data_kunjungan_wisatawan_nusantara['total'] = data_kunjungan_wisatawan_nusantara['total'].astype(int)

data_kunjungan_wisatawan_nusantara['kategori'] = 0
for i in range(0, len(data_kunjungan_wisatawan_nusantara)):
    if 0 <= data_kunjungan_wisatawan_nusantara['total'][i] <= 1000000:
        data_kunjungan_wisatawan_nusantara['kategori'][i] = "Sedikit"
    elif 1000000 < data_kunjungan_wisatawan_nusantara['total'][i] <= np.inf:
        data_kunjungan_wisatawan_nusantara['kategori'][i] = "Cukup Banyak"

data_kunjungan_wisatawan_nusantara
```

Menambah
kolom/membuat
fitur baru

```
data_kunjungan_wisatawan_mancanegara['total'] = 0
data_kunjungan_wisatawan_mancanegara['total'] = data_kunjungan_wisatawan_mancanegara.iloc[:, 1:].sum(axis=1)
data_kunjungan_wisatawan_mancanegara['total'] = data_kunjungan_wisatawan_mancanegara['total'].astype(int)

data_kunjungan_wisatawan_mancanegara['kategori'] = 0
for i in range(0, len(data_kunjungan_wisatawan_mancanegara)):
    if 0 <= data_kunjungan_wisatawan_mancanegara['total'][i] <= 1000:
        data_kunjungan_wisatawan_mancanegara['kategori'][i] = "Sedikit"
    elif 1000 < data_kunjungan_wisatawan_mancanegara['total'][i] <= np.inf:
        data_kunjungan_wisatawan_mancanegara['kategori'][i] = "Cukup Banyak"

data_kunjungan_wisatawan_mancanegara
```

```
kunjungan_mancanegara_2020['total'] = 0
kunjungan_mancanegara_2020['total'] = kunjungan_mancanegara_2020.iloc[:, 1:].sum(axis=1)
kunjungan_mancanegara_2020['total'] = kunjungan_mancanegara_2020['total'].astype(int)

kunjungan_mancanegara_2020['kategori'] = 0
for i in range(0, len(kunjungan_mancanegara_2020)):
    if 0 <= kunjungan_mancanegara_2020['total'][i] <= 1000:
        kunjungan_mancanegara_2020['kategori'][i] = "Sedikit"
    elif 1000 < kunjungan_mancanegara_2020['total'][i] <= np.inf:
        kunjungan_mancanegara_2020['kategori'][i] = "Cukup Banyak"

kunjungan_mancanegara_2020
```

FLOW PROCESS & VISUALISASI DATA

	destinasi_wisata	1	2	3	4	5	6	7	8	9	10	11	12	total	kategori
0	Kawasan Kota Tua	456902	587641	121516	0	0	2619	3925	3386	1837	8618	3354	0	1189798	Cukup Banyak
1	Kepulauan Seribu	31041	63568	17948	46	0	8859	20360	37986	15270	14944	18301	22289	250612	Sedikit
2	Monumen Nasional	24325	146042	50328	0	0	0	0	0	0	443034	0	0	663729	Sedikit
3	Museum Bahari	2087	2102	1209	0	0	173	632	887	269	383	729	654	9125	Sedikit
4	Museum Joang 45	581	1284	306	0	0	94	115	359	113	249	267	228	3596	Sedikit
5	Museum Nasional	18145	27928	12820	0	0	173	583	233	105	65308	1451	1418	128164	Sedikit
6	Museum Prasasti	1030	973	268	0	0	104	194	280	125	203	417	364	3958	Sedikit
7	Museum Satria Mandala	2046	518	619	0	0	0	0	0	0	0	0	0	3183	Sedikit
8	Museum Sejarah Jakarta	66251	59895	8947	0	0	982	2066	1857	769	2864	5610	4083	153324	Sedikit
9	Museum Seni Rupa dan Keramik	12190	11800	2666	0	0	233	327	321	237	422	1003	627	29826	Sedikit
10	Museum Tekstil	2658	3049	1206	0	0	190	132	162	80	96	161	203	7937	Sedikit
11	Museum Wayang	21231	14728	2564	0	0	432	819	506	422	1390	2728	1788	46608	Sedikit
12	PBB Setu babakan	35369	28848	11850	0	0	6057	1548	21997	4704	10372	17181	9883	147809	Sedikit
13	Pelabuhan Sunda Kelapa	931	1895	1304	380	0	753	927	2535	2155	2232	2752	2296	18160	Sedikit
14	Planetarium	9654	10931	1693	0	0	0	0	0	0	0	0	0	22278	Sedikit
15	Rumah si Pitung	2048	1811	672	0	0	788	1722	2597	1018	1342	2316	1238	15552	Sedikit
16	Taman Arkeologi Onrust	2200	3236	951	0	0	948	3711	4257	2969	2783	4186	1621	26862	Sedikit
17	Taman Impian Jaya Ancol	620327	638729	358871	0	0	28175	85142	174988	71688	1978996	126744	479272	4562932	Cukup Banyak
18	Taman Margasatwa Ragunan	289748	191179	82706	0	0	3837	9365	16570	6109	600470	21741	19012	1240737	Cukup Banyak
19	Taman Mini Indonesia Indah	272960	317391	89617	0	0	28090	82696	134546	36977	962277	107810	105008	2137372	Cukup Banyak

FLOW PROCESS & VISUALISASI DATA

	destinasi_wisata	1	2	3	4	5	6	7	8	9	10	11	12	total	kategori
0	Kawasan Kota Tua	6009	7123	2144	0	0	73	88	125	41	139	45	0	15787	Cukup Banyak
1	Kepulauan Seribu	1352	800	419	0	0	48	221	209	87	126	138	214	3614	Cukup Banyak
2	Monumen Nasional	2070	1689	580	0	0	0	0	0	0	0	0	0	4339	Cukup Banyak
3	Museum Bahari	274	149	55	0	0	1	13	5	2	3	17	30	549	Sedikit
4	Museum Joang 45	0	15	18	0	0	0	0	0	0	0	0	0	33	Sedikit
5	Museum Nasional	2877	1877	546	0	0	7	13	1	0	12	39	25	5397	Cukup Banyak
6	Museum Prasasti	50	12	18	0	0	0	0	0	0	0	0	0	80	Sedikit
7	Museum Satria Mandala	0	0	0	0	0	0	0	0	0	0	0	0	0	Sedikit
8	Museum Sejarah Jakarta	1831	1755	389	0	0	3	0	0	2	2	0	0	3982	Cukup Banyak
9	Museum Seni Rupa dan Keramik	300	234	142	0	0	5	2	10	2	1	6	0	702	Sedikit
10	Museum Tekstil	312	121	73	0	0	5	0	0	0	0	2	0	513	Sedikit
11	Museum Wayang	865	1006	648	0	0	18	11	28	4	11	47	0	2638	Cukup Banyak
12	PBB Setu babakan	2	10	0	0	0	0	0	0	0	0	0	0	12	Sedikit
13	Pelabuhan Sunda Kelapa	115	61	0	0	0	0	0	0	0	0	4	0	180	Sedikit
14	Planetarium	0	0	0	0	0	0	0	0	0	0	0	0	0	Sedikit
15	Rumah si Pitung	0	0	0	0	0	0	1	0	0	0	0	0	1	Sedikit
16	Taman Arkeologi Onrust	1	4	0	0	0	0	0	0	0	0	4	0	9	Sedikit
17	Taman Impian Jaya Ancol	516	395	165	0	0	0	0	0	0	0	0	0	1076	Cukup Banyak
18	Taman Margasatwa Ragunan	449	372	82	0	0	3	7	0	43	27	40	31	1054	Cukup Banyak
19	Taman Mini Indonesia Indah	0	0	0	0	0	0	0	0	0	0	0	0	0	Sedikit

	bulan	Bandara Halim PK	Bandara Soekarno-Hatta	total	kategori
0	1	161	173453	173614	Cukup Banyak
1	2	205	131032	131237	Cukup Banyak
2	3	90	52631	52721	Cukup Banyak
3	4	7	417	424	Sedikit
4	5	19	395	414	Sedikit
5	6	15	909	924	Sedikit
6	7	27	3119	3146	Cukup Banyak
7	8	18	4469	4487	Cukup Banyak
8	9	60	7468	7528	Cukup Banyak
9	10	37	10492	10529	Cukup Banyak
10	11	56	14309	14365	Cukup Banyak
11	12	46	21812	21858	Cukup Banyak

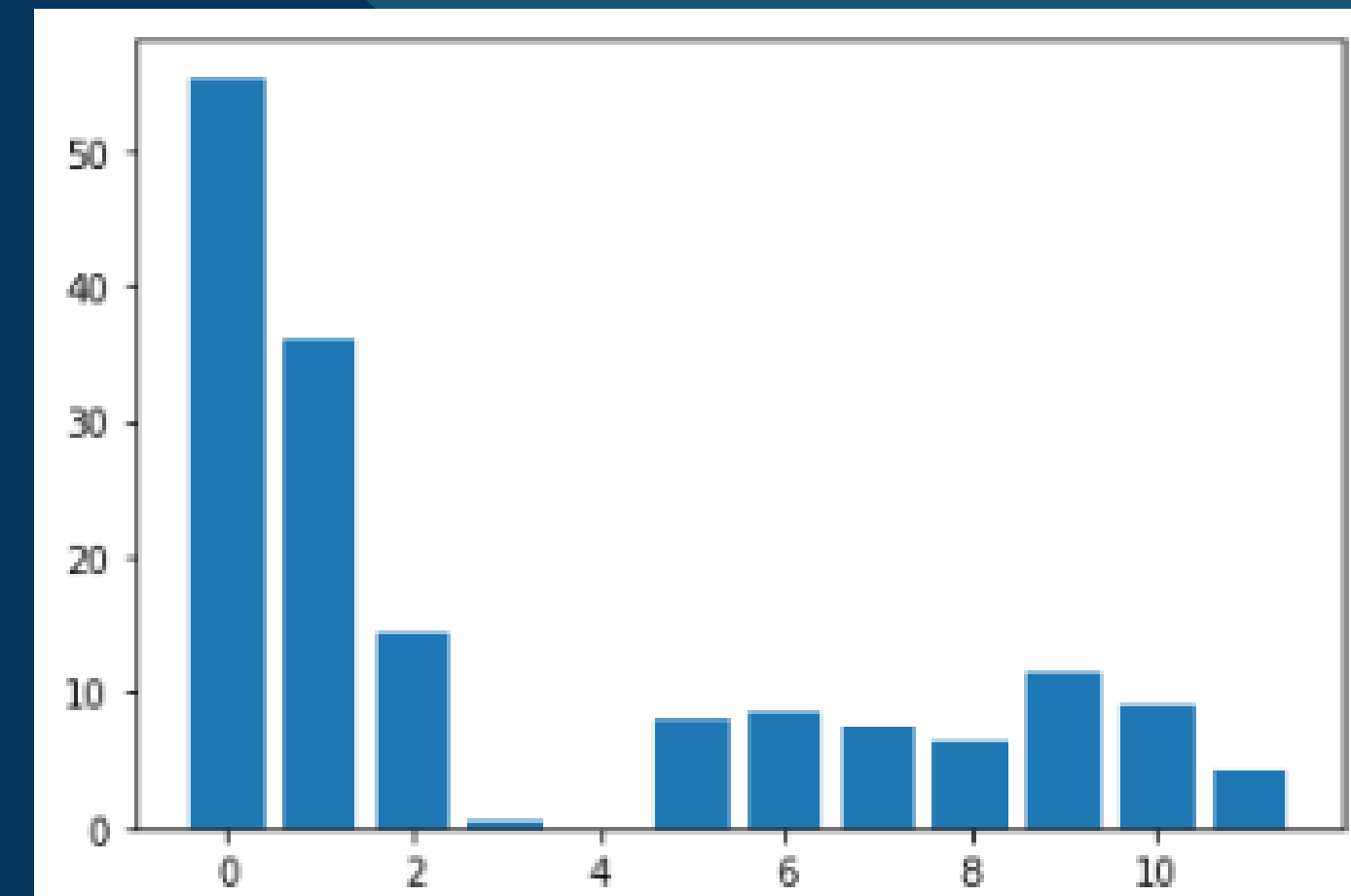
FLOW PROCESS & VISUALISASI DATA

```
# Fungsi seleksi fitur
def anova_select_features(X_train, y_train, X_test):
    # configure to select all features
    fs = SelectKBest(score_func=f_classif, k="all")
    # learn relationship from training data
    fs.fit(X_train, y_train)
    # transform train input data
    X_train_fs = fs.transform(X_train)
    # transform test input data
    X_test_fs = fs.transform(X_test)
    return X_train_fs, X_test_fs, fs

# Load dataset
X, y = data_kunjungan_wisatawan_nusantara.iloc[:,1:13], data_kunjungan_wisatawan_nusantara.iloc[:, -1]
# Membagi data menjadi train dan test
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.33, random_state=1)
# Proses feature selection
X_train_fs, X_test_fs, fs = anova_select_features(X_train, y_train, X_test)
# Nilai/scores untuk features
for i in range(len(fs.scores_)):
    print('Feature %d: %f' % (i, fs.scores_[i]))
# plot nilai features
pyplot.bar([i for i in range(len(fs.scores_))], fs.scores_)
pyplot.show()
```

```
Feature 0: 55.363419
Feature 1: 35.925461
Feature 2: 14.293306
Feature 3: 0.540137
Feature 4: nan
Feature 5: 8.090654
Feature 6: 8.547724
Feature 7: 7.317955
Feature 8: 6.358130
Feature 9: 11.388943
Feature 10: 9.120028
Feature 11: 4.312662
```

Seleksi fitur
ANOVA



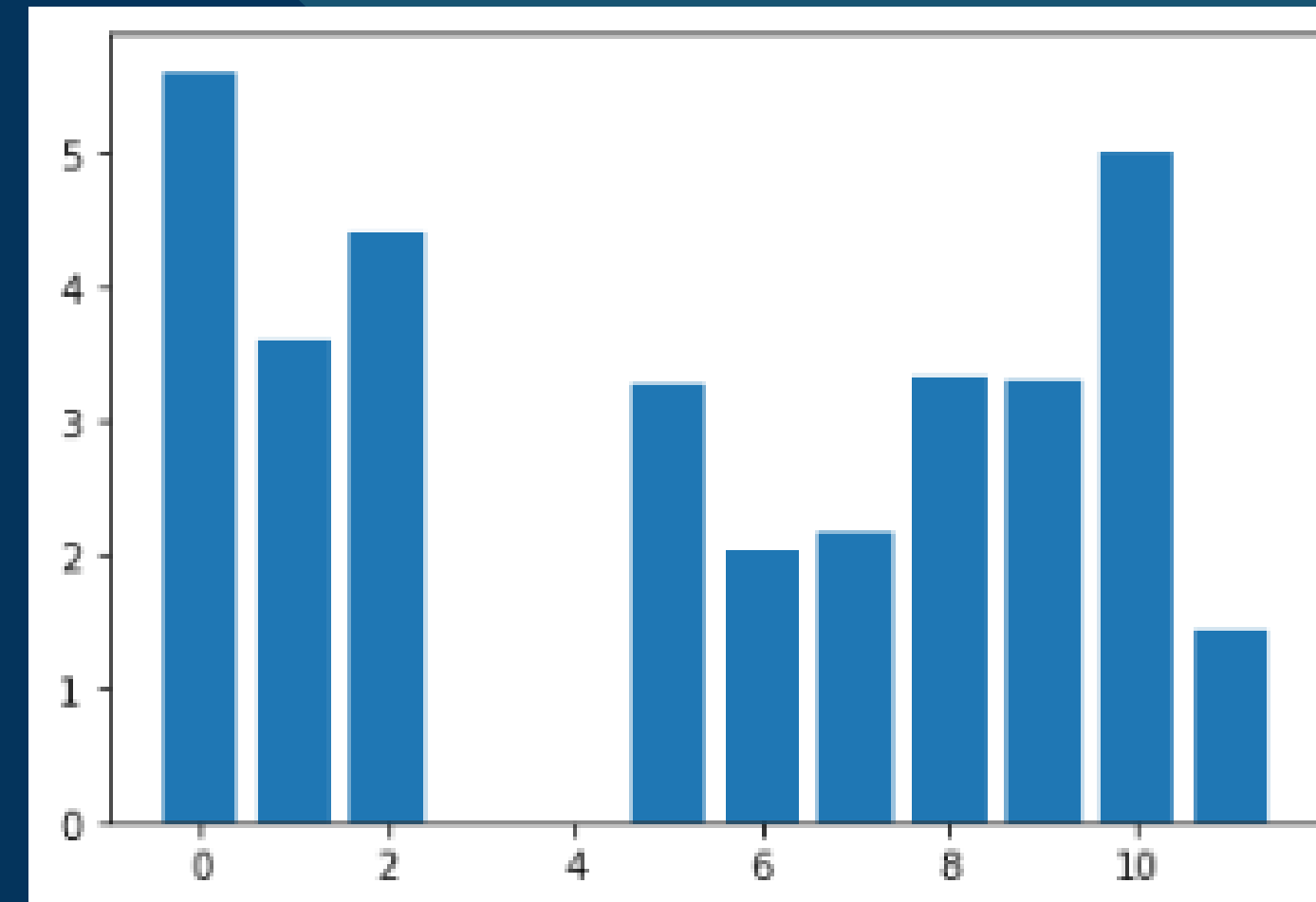
Kolom dengan nilai tertinggi
adalah bulan Januari

FLOW PROCESS & VISUALISASI DATA

Seleksi fitur ANOVA

```
# Feature Selection untuk data kunjungan wisatawan mancanegara ke destinasi wisata DKI Jakarta 2020
# Load dataset
X, y = data_kunjungan_wisatawan_mancanegara.iloc[:,1:13], data_kunjungan_wisatawan_mancanegara.iloc[:, -1]
# Membagi data menjadi train dan test
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.33, random_state=1)
# Proses feature selection
X_train_fs, X_test_fs, fs = select_features(X_train, y_train, X_test)
# Nilai/scores untuk features
for i in range(len(fs.scores_)):
    print('Feature %d: %f' % (i, fs.scores_[i]))
# plot nilai features
pyplot.bar([i for i in range(len(fs.scores_))], fs.scores_)
pyplot.show()
```

```
Feature 0: 5.609697
Feature 1: 3.594560
Feature 2: 4.415278
Feature 3: nan
Feature 4: nan
Feature 5: 3.266108
Feature 6: 2.032120
Feature 7: 2.165029
Feature 8: 3.328990
Feature 9: 3.312536
Feature 10: 5.011606
Feature 11: 1.430235
```



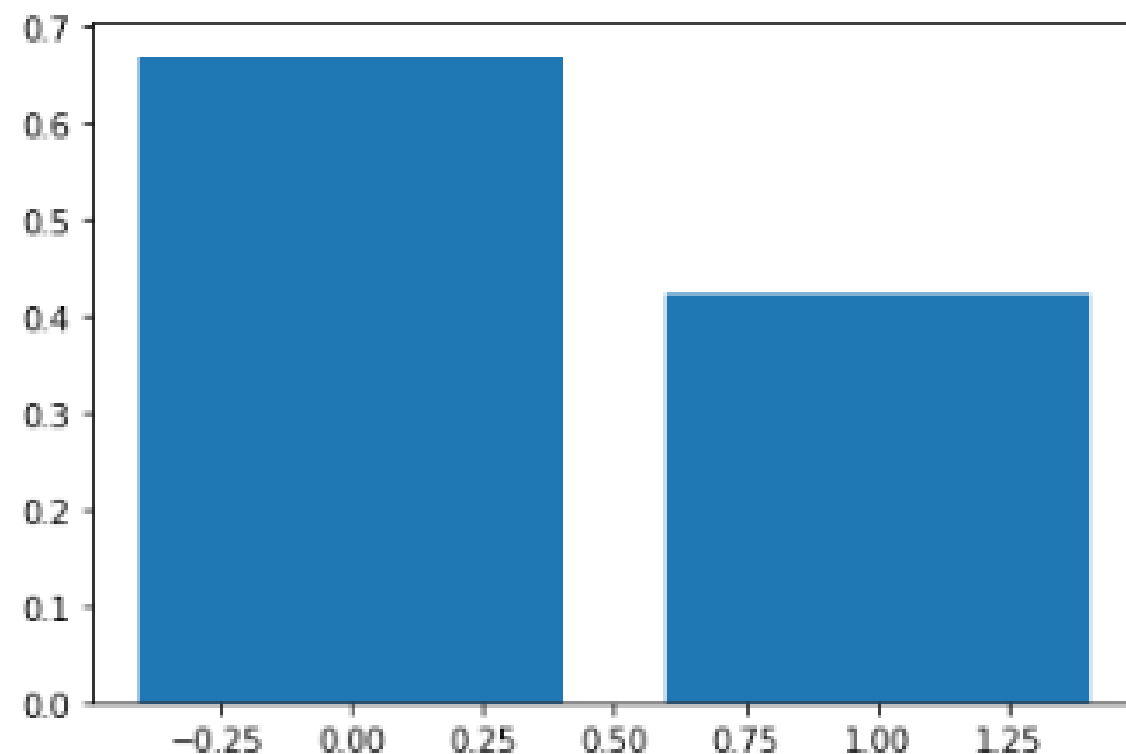
Kolom dengan nilai tertinggi
adalah bulan Januari

FLOW PROCESS & VISUALISASI DATA

```
# Feature Selection untuk data kunjungan wisatawan mancanegara ke provinsi DKI Jakarta tahun 2020

# Load dataset
X, y = kunjungan_mancanegara_2020.iloc[:,1:3], kunjungan_mancanegara_2020.iloc[:, -1]
# Membagi data menjadi train dan test
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.33, random_state=1)
# Proses feature selection
X_train_fs, X_test_fs, fs = select_features(X_train, y_train, X_test)
# Nilai/scores untuk features
for i in range(len(fs.scores_)):
    print('Feature %d: %f' % (i, fs.scores_[i]))
# plot nilai features
pyplot.bar([i for i in range(len(fs.scores_))], fs.scores_)
pyplot.show()
```

Feature 0: 0.670002
Feature 1: 0.422872



Seleksi fitur
ANOVA

Kolom hanya ada 2, yaitu
Bandara Halim P K dan
Bandara Soekarno-Hatta.
Tertinggi -> Bandara
Halim P K

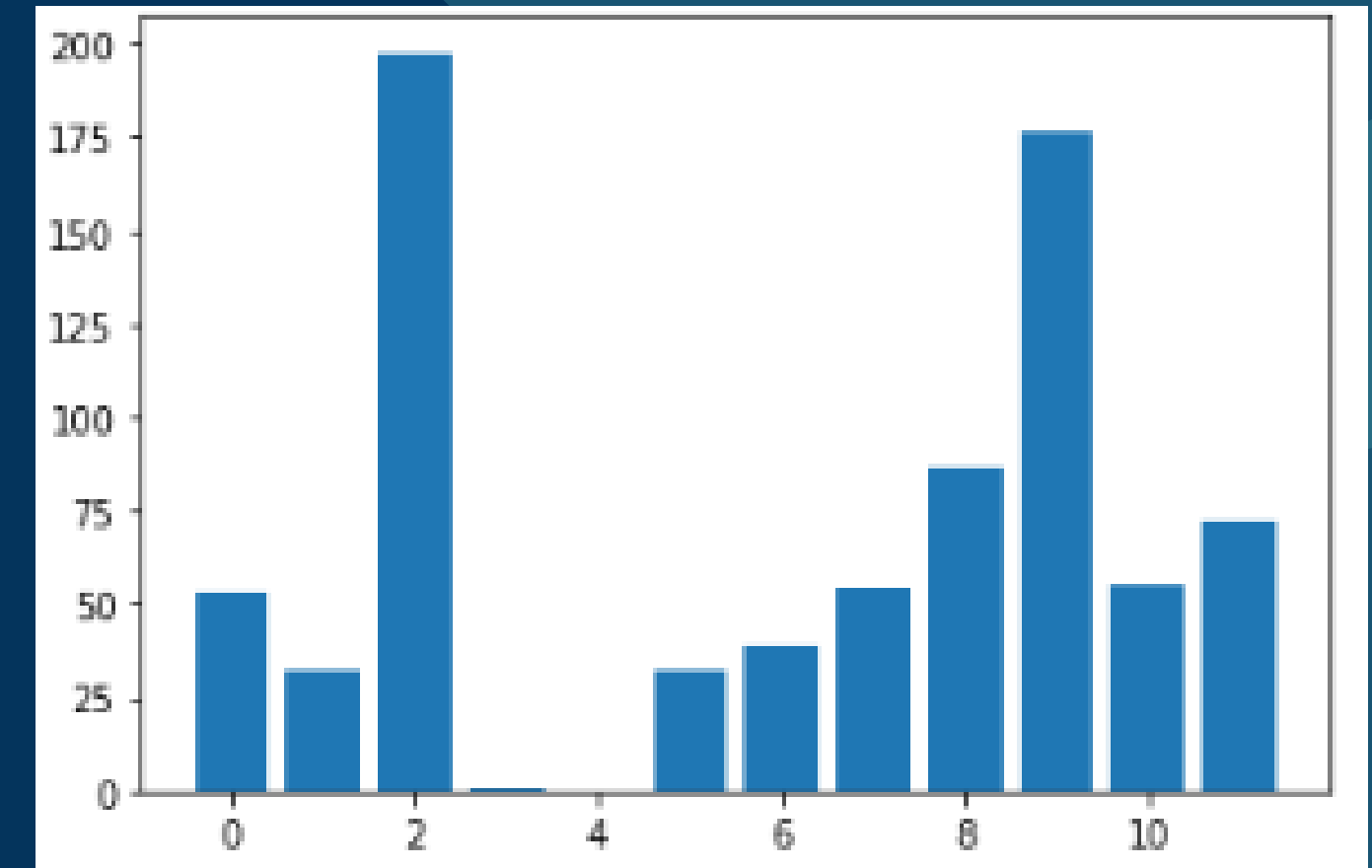
FLOW PROCESS & VISUALISASI DATA

```
# Feature Selection untuk data kunjungan wisatawan nusantara ke destinasi wisata DKI Jakarta 2020
# Fungsi seleksi fitur
def pearson_select_features(X_train, y_train, X_test):
    """# configure to select all features
    """fs = SelectKBest(score_func=f_regression, k="all")
    """# Learn relationship from training data
    """fs.fit(X_train, y_train)
    """# transform train input data
    """X_train_fs = fs.transform(X_train)
    """# transform test input data
    """X_test_fs = fs.transform(X_test)
    """return X_train_fs, X_test_fs, fs

# Load dataset
X, y = data_kunjungan_wisatawan_nusantara.iloc[:,1:13], data_kunjungan_wisatawan_nusantara.iloc[:, -2]
# Membagi data menjadi train dan test
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.33, random_state=1)
# Proses feature selection
X_train_fs, X_test_fs, fs = pearson_select_features(X_train, y_train, X_test)
# Nilai/scores untuk features
for i in range(len(fs.scores_)):
    print('Feature %d: %f' % (i, fs.scores_[i]))
# plot nilai features
pyplot.bar([i for i in range(len(fs.scores_))], fs.scores_)
pyplot.show()
```

```
Feature 0: 52.928810
Feature 1: 32.052603
Feature 2: 197.691959
Feature 3: 0.386885
Feature 4: nan
Feature 5: 32.730175
Feature 6: 38.500880
Feature 7: 53.752085
Feature 8: 87.067463
Feature 9: 176.063251
Feature 10: 55.049971
Feature 11: 72.144347
```

Seleksi fitur
Pearson's

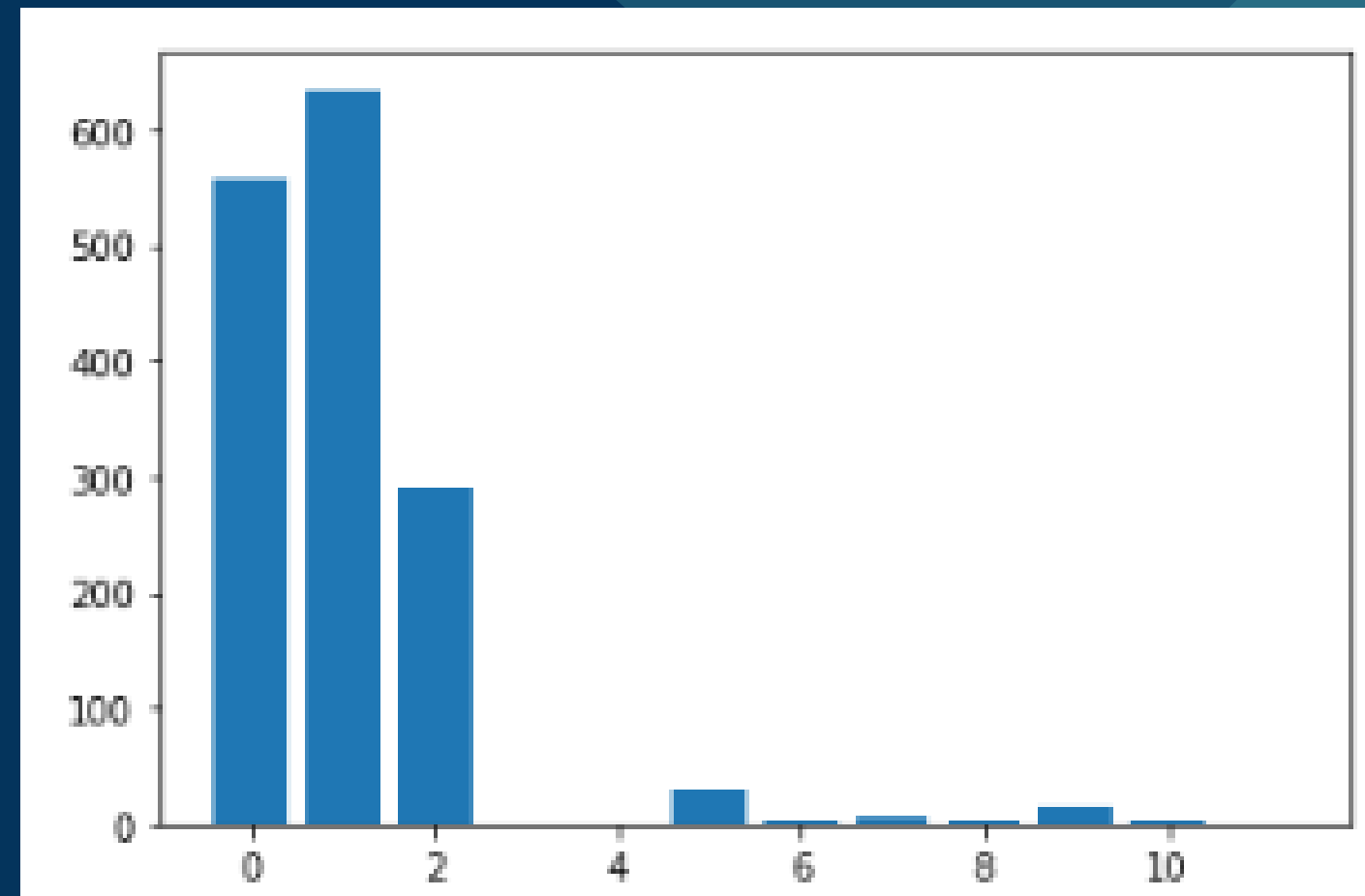


FLOW PROCESS & VISUALISASI DATA

```
# Feature Selection untuk data kunjungan wisatawan mancanegara ke destinasi wisata DKI Jakarta 2020
# Load dataset
X, y = data_kunjungan_wisatawan_mancanegara.iloc[:,1:13], data_kunjungan_wisatawan_mancanegara.iloc[:, -2]
# Membagi data menjadi train dan test
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.33, random_state=1)
# Proses feature selection
X_train_fs, X_test_fs, fs = pearson_select_features(X_train, y_train, X_test)
# Nilai/scores untuk features
for i in range(len(fs.scores_)):
    print('Feature %d: %f' % (i, fs.scores_[i]))
# plot nilai features
pyplot.bar([i for i in range(len(fs.scores_))], fs.scores_)
pyplot.show()
```

```
Feature 0: 558.027053
Feature 1: 634.459842
Feature 2: 290.803192
Feature 3: nan
Feature 4: nan
Feature 5: 31.295039
Feature 6: 2.382055
Feature 7: 4.684205
Feature 8: 2.037115
Feature 9: 15.537049
Feature 10: 1.675614
Feature 11: 0.059703
```

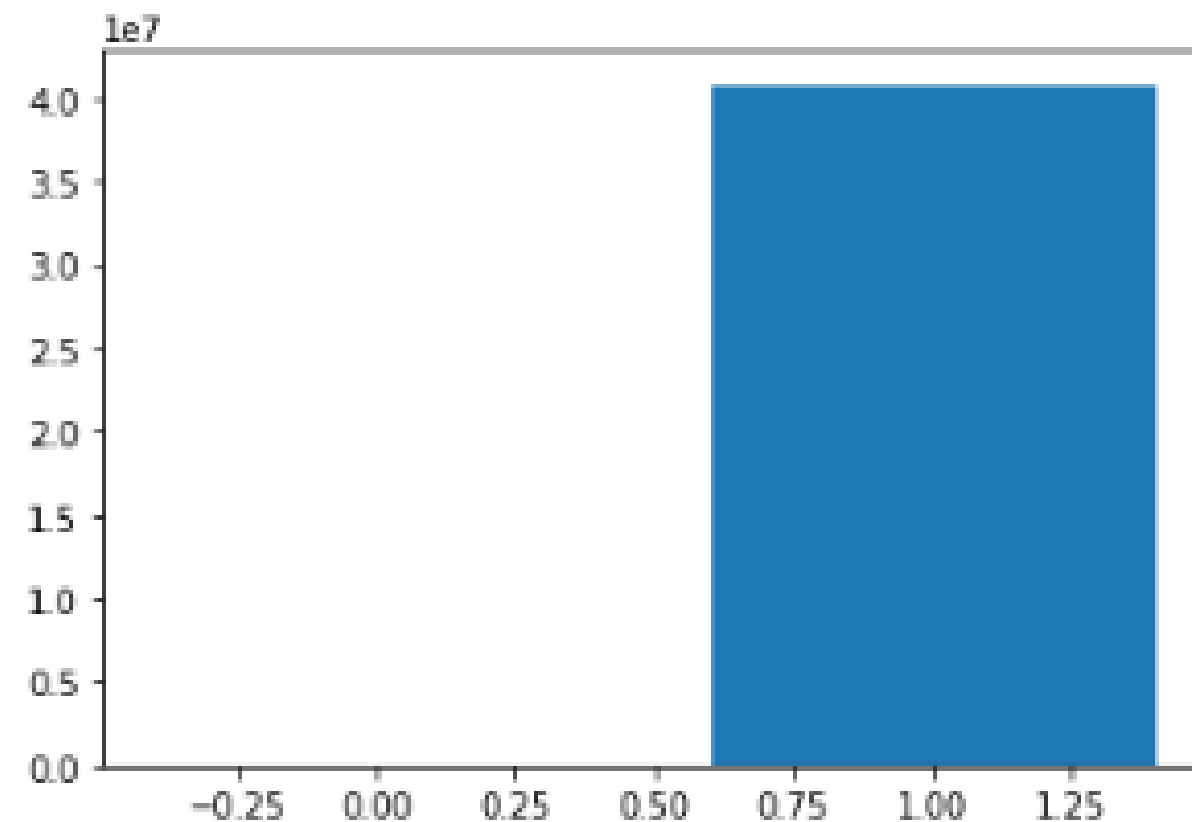
Seleksi fitur
Pearson's



FLOW PROCESS & VISUALISASI DATA

```
# Feature Selection untuk data kunjungan wisatawan mancanegara ke provinsi DKI Jakarta tahun 2020
# Load dataset
X, y = kunjungan_mancanegara_2020.iloc[:,1:3], kunjungan_mancanegara_2020.iloc[:, -2]
# Membagi data menjadi train dan test
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.33, random_state=1)
# Proses feature selection
X_train_fs, X_test_fs, fs = pearson_select_features(X_train, y_train, X_test)
# Nilai/scores untuk features
for i in range(len(fs.scores_)):
    print('Feature %d: %f' % (i, fs.scores_[i]))
# plot nilai features
pyplot.bar([i for i in range(len(fs.scores_))], fs.scores_)
pyplot.show()
```

```
Feature 0: 39.197176
Feature 1: 40826340.905416
```



Seleksi fitur
Pearson's

KESIMPULAN

Pandemi COVID-19 sangat berdampak dan berpengaruh pada bidang pariwisata bangsa kita pada tahun 2020, dalam kasus ini adalah kunjungan wisatawan baik itu mancanegara maupun nusantara ke destinasi wisata dan kunjungan wisatawan mancanegara ke bandara Jakarta. Dari hasil seleksi fitur, Bulan 1, 2, dan 3 memiliki nilai yang tinggi cukup jauh dibanding bulan 4-12, ini karena kasus COVID-19 menjadi isu/naik ke skala nasional mulai pada bulan 4 dan seterusnya, situasi dan kondisi pandemi semakin baik setiap bulannya membuat data kunjungan wisatawan naik perlahan-lahan, untuk data kunjungan nusantara sendiri sempat naik besar karena adanya pelanggaran aturan, namun setelah itu banyaknya kerumunan dan pelanggaran/kelalaian justru menjatuhkan kembali secara besar angka kunjungan wisatawan nusantara ke destinasi wisata Jakarta yang kemudian naik lagi tapi dengan perlahan-lahan. Untuk kunjungan wisatawan mancanegara ke bandara Jakarta yang memiliki nilai tertinggi adalah jika menggunakan seleksi fitur ANOVA adalah Bandara Halim P K, sedangkan jika menggunakan seleksi fitur Pearson' adalah Bandara Soekarno-Hatta, tapi untuk perbandingan nilai Soekarno-Hatta dan Halim P K sebenarnya nilai Soekarno-Hatta jauh lebih tinggi