

## Exploratory Data Analysis

---

Melakukan eksplorasi dan persiapan data untuk membangun model. Data yang dipakai dari

<https://opendata.jabarprov.go.id/id/dataset?organization=badan-penanggulangan-bencana-daerah>

```
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import seaborn as sns
%matplotlib inline
import matplotlib.cm as cm

df1 = pd.read_csv('angin_topan_puting_beliung.csv')
df2 = pd.read_csv('banjir.csv')
df3 = pd.read_csv('gelombang_pasang.csv')
df4 = pd.read_csv('gempa_bumi.csv')
df5 = pd.read_csv('kebakaran.csv')
df6 = pd.read_csv('kebakaran_bangunan.csv')
df7 = pd.read_csv('kebakaran_lahan.csv')
df8 = pd.read_csv('tanah_longsor.csv')

df6.isnull().any()

id                False
kode_provinsi     False
nama_provinsi     False
kode_kabupaten_kota False
nama_kabupaten_kota False
jumlah_kebakaran_bangunan False
satuan            False
tahun            False
dtype: bool

key_columns=['id','kode_provinsi','nama_provinsi','kode_kabupaten_kota',
            'nama_kabupaten_kota','satuan','tahun']

df_1 = pd.merge(df1,df2, how='left', on=key_columns)
df_2 = pd.merge(df_1, df3, how='left',on=key_columns)
df_3 = pd.merge(df_2, df4, how='left',on=key_columns)
df_4 = pd.merge(df_3, df5, how='left',on=key_columns)
df_5 = pd.merge(df_4, df6, how='left',on=key_columns)
df_6 = pd.merge(df_5, df7, how='left',on=key_columns)
df = pd.merge(df_6, df8, how='left',on=key_columns)

df.to_csv('bencana.csv', index=False)
```

```
data = pd.read_csv("bencana.csv")
data.head()
```

	id	kode_provinsi	...	jumlah_kebakaran_lahan	jumlah_tanah_longsor
0	1	32	...	NaN	0
1	2	32	...	NaN	48
2	3	32	...	NaN	9
3	4	32	...	NaN	12
4	5	32	...	NaN	36

[5 rows x 15 columns]

```
data.columns
```

```
Index(['id', 'kode_provinsi', 'nama_provinsi', 'kode_kabupaten_kota',  
      'nama_kabupaten_kota', 'jumlah_puting_beliung', 'satuan',  
      'tahun',  
      'jumlah_banjir', 'jumlah_gelombang_pasang',  
      'jumlah_gempa_bumi',  
      'jumlah_kebakaran', 'jumlah_kebakaran_bangunan',  
      'jumlah_kebakaran_lahan', 'jumlah_tanah_longsor'],  
      dtype='object')
```

```
data.describe().transpose()
```

	count	mean	...	75%	max
id	241.0	121.000000	...	181.0	241.0
kode_provinsi	241.0	32.000000	...	32.0	32.0
kode_kabupaten_kota	241.0	3231.443983	...	3273.0	3279.0
jumlah_puting_beliung	241.0	9.568465	...	12.0	183.0
tahun	241.0	2016.029046	...	2018.0	2020.0
jumlah_banjir	241.0	6.717842	...	8.0	71.0
jumlah_gelombang_pasang	241.0	0.087137	...	0.0	4.0
jumlah_gempa_bumi	241.0	1.941909	...	1.0	70.0
jumlah_kebakaran	187.0	11.545455	...	16.0	93.0
jumlah_kebakaran_bangunan	0.0	NaN	...	NaN	NaN
jumlah_kebakaran_lahan	0.0	NaN	...	NaN	NaN
jumlah_tanah_longsor	241.0	17.738589	...	25.0	133.0

```
[12 rows x 8 columns]
```

```
data.rename(index=str, columns={'jumlah_banjir': 'banjir',
'jumlah_gelombang_pasang' : 'pasang',
                                'jumlah_gempa_bumi' : 'gempa',
'jumlah_kebakaran' : 'kebakaran',
                                'jumlah_kebakaran_lahan' :
'kebakaran_lahan', 'jumlah_kebakaran_bangunan' :
                                'kebakaran_bangunan',
'jumlah_tanah_longsor' : 'longsor', 'jumlah_puting_beliung' :
'puting_beliung'}, inplace =True)
```

```
data = data.drop(['id'], axis=1).head()
data
```

	kode_provinsi	nama_provinsi	...	kebakaran_lahan	longsor
0	32	JAWA BARAT	...	NaN	0
1	32	JAWA BARAT	...	NaN	48
2	32	JAWA BARAT	...	NaN	9
3	32	JAWA BARAT	...	NaN	12
4	32	JAWA BARAT	...	NaN	36

```
[5 rows x 14 columns]
```

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
Index: 5 entries, 0 to 4
```

```
Data columns (total 14 columns):
```

#	Column	Non-Null Count	Dtype
0	kode_provinsi	5 non-null	int64
1	nama_provinsi	5 non-null	object
2	kode_kabupaten_kota	5 non-null	int64
3	nama_kabupaten_kota	5 non-null	object
4	puting_beliung	5 non-null	int64
5	satuan	5 non-null	object
6	tahun	5 non-null	int64
7	banjir	5 non-null	int64
8	pasang	5 non-null	int64
9	gempa	5 non-null	int64
10	kebakaran	5 non-null	float64
11	kebakaran_bangunan	0 non-null	float64
12	kebakaran_lahan	0 non-null	float64
13	longsor	5 non-null	int64

```
dtypes: float64(3), int64(8), object(3)
```

```
memory usage: 600.0+ bytes
```

```
data.isnull().sum()
```

kode_provinsi	0
nama_provinsi	0
kode_kabupaten_kota	0
nama_kabupaten_kota	0
puting_beliung	0
satuan	0
tahun	0
banjir	0
pasang	0
gempa	0
kebakaran	0
kebakaran_bangunan	5
kebakaran_lahan	5

```
longsor          0
dtype: int64
```

Terlihat bahwa, pada kolom "kebakaran\_bangunan" dan "kebakaran\_lahan" terdapat nilai Null