

# STREAMLIT DATASET WISATA JAWA BARAT

50421643\_Ilham Rizkyansyah\_Kelas C

## DATA UNDERSTANDING

### Memuat dataset

	id	kode_provinsi	nama_provinsi	kode_kabupaten_kota	nama_kabupaten_kota	jenis_odtw	jumlah
0	1	32	JAWA BARAT	3,201	KABUPATEN BOGOR	ALAM	
1	2	32	JAWA BARAT	3,201	KABUPATEN BOGOR	BUDAYA	
2	3	32	JAWA BARAT	3,201	KABUPATEN BOGOR	MINAT KHUSUS	
3	4	32	JAWA BARAT	3,202	KABUPATEN SUKABUMI	ALAM	
4	5	32	JAWA BARAT	3,202	KABUPATEN SUKABUMI	BUDAYA	
5	6	32	JAWA BARAT	3,202	KABUPATEN SUKABUMI	MINAT KHUSUS	
6	7	32	JAWA BARAT	3,203	KABUPATEN CIANJUR	ALAM	
7	8	32	JAWA BARAT	3,203	KABUPATEN CIANJUR	BUDAYA	
8	9	32	JAWA BARAT	3,203	KABUPATEN CIANJUR	MINAT KHUSUS	
9	10	32	JAWA BARAT	3,204	KABUPATEN BANDUNG	ALAM	

### Menampilkan informasi dasar tentang dataset

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 807 entries, 0 to 806
Data columns (total 9 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id                     807 non-null   int64
1   kode_provinsi          807 non-null   int64
2   nama_provinsi          807 non-null   object
3   kode_kabupaten_kota    807 non-null   int64
4   nama_kabupaten_kota    807 non-null   object
5   jenis_odtw             807 non-null   object
6   jumlah_odtw            807 non-null   int64
7   satuan                 807 non-null   object
8   tahun                 807 non-null   int64
dtypes: int64(5), object(4)
memory usage: 56.9+ KB
```

### Menampilkan 5 baris pertama dari Dataset

	id	kode_provinsi	nama_provinsi	kode_kabupaten_kota	nama_kabupaten_kota	jenis_odtw	jumlah
0	1	32	JAWA BARAT	3,201	KABUPATEN BOGOR	ALAM	
1	2	32	JAWA BARAT	3,201	KABUPATEN BOGOR	BUDAYA	
2	3	32	JAWA BARAT	3,201	KABUPATEN BOGOR	MINAT KHUSUS	
3	4	32	JAWA BARAT	3,202	KABUPATEN SUKABUMI	ALAM	

1	2	32	JAWA BARAT	3,201	KABUPATEN BOGOR	BUDAYA	
2	3	32	JAWA BARAT	3,201	KABUPATEN BOGOR	MINAT KHUSUS	
3	4	32	JAWA BARAT	3,202	KABUPATEN SUKABUMI	ALAM	
4	5	32	JAWA BARAT	3,202	KABUPATEN SUKABUMI	BUDAYA	

Menampilkan statistik deskriptif dari dataset

	id	kode_provinsi	nama_provinsi	kode_kabupaten_kota	nama_kabupaten_kota	jenis_odtw	jumlah_od
count	807	807	807	807	807	807	807
unique	None	None	1	None	27	4	None
top	None	None	JAWA BARAT	None	KABUPATEN BOGOR	ALAM	None
freq	None	None	807	None	30	269	None
mean	404	32	None	3,231.3829	None	None	23.18
std	33.1051	0	None	31.2686	None	None	37.45
min	1	32	None	3,201	None	None	
25%	202.5	32	None	3,207	None	None	
50%	404	32	None	3,214	None	None	
75%	605.5	32	None	3,273	None	None	

Cek nilai yang hilang (NaN) pada dataset:

	0
id	0
kode_p	0
nama_	0
kode_k	0
nama_	0
jenis_o	0
jumlah	0
satuan	0
tahun	0

## DATA PREPARATION

Menangani data dan mempersiapkannya untuk analisis.

Mengisi nilai yang hilang dengan 0.

Dataset setelah menangani nilai yang hilang:

	id	kode_provinsi	nama_provinsi	kode_kabupaten_kota	nama_kabupaten_kota	jenis_odtw	jumlah
--	----	---------------	---------------	---------------------	---------------------	------------	--------

0	1	32	JAWA BARAT	3,201	KABUPATEN BOGOR	ALAM	
1	2	32	JAWA BARAT	3,201	KABUPATEN BOGOR	BUDAYA	
2	3	32	JAWA BARAT	3,201	KABUPATEN BOGOR	MINAT KHUSUS	
3	4	32	JAWA BARAT	3,202	KABUPATEN SUKABUMI	ALAM	
4	5	32	JAWA BARAT	3,202	KABUPATEN SUKABUMI	BUDAYA	

Encoding kolom kategorikal jika ada.

Dataset setelah encoding:

	id	kode_provinsi	nama_provinsi	kode_kabupaten_kota	nama_kabupaten_kota	jenis_odtw	jumlah
0	1	32	0	3,201	KABUPATEN BOGOR	ALAM	
1	2	32	0	3,201	KABUPATEN BOGOR	BUDAYA	
2	3	32	0	3,201	KABUPATEN BOGOR	MINAT KHUSUS	
3	4	32	0	3,202	KABUPATEN SUKABUMI	ALAM	
4	5	32	0	3,202	KABUPATEN SUKABUMI	BUDAYA	

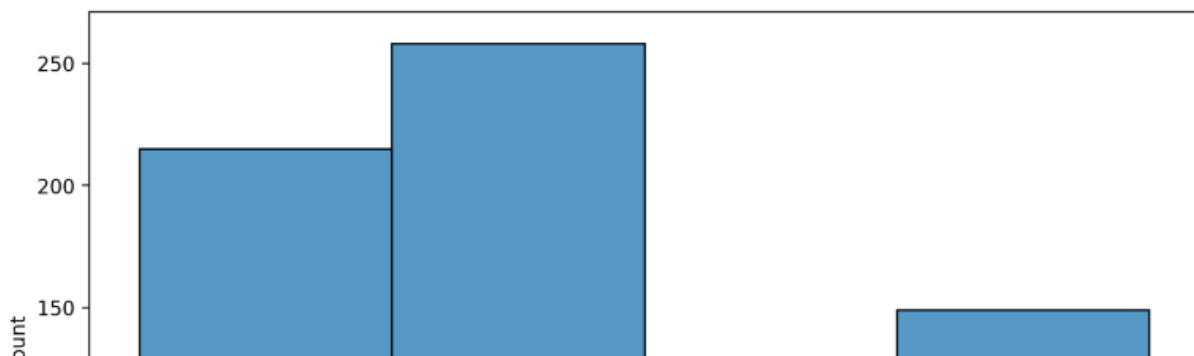
Trimming outliers dari jumlah obyek daya tarik wisata menggunakan metode IQR.

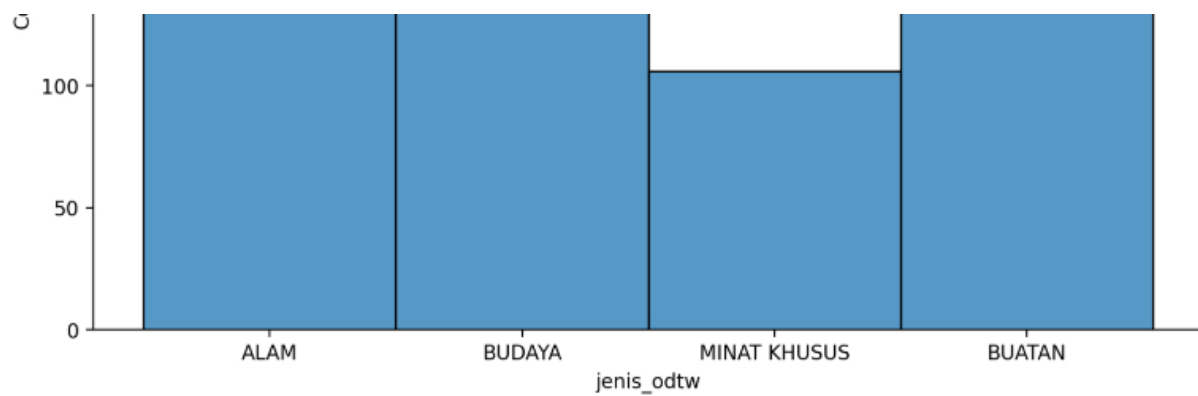
Dataset setelah trimming outliers:

	id	kode_provinsi	nama_provinsi	kode_kabupaten_kota	jumlah_odtw	tahun
count	728	728	728	728	728	728
mean	387.3159	32	0	3,233.7761	13.4821	2,018.2885
std	234.3886	0	0	31.8691	15.7131	2.8655
min	1	32	0	3,201	0	2,014
25%	183.75	32	0	3,208	2	2,016
50%	374.5	32	0	3,215	7	2,018
75%	590.25	32	0	3,273	19	2,021
max	807	32	0	3,279	65	2,023

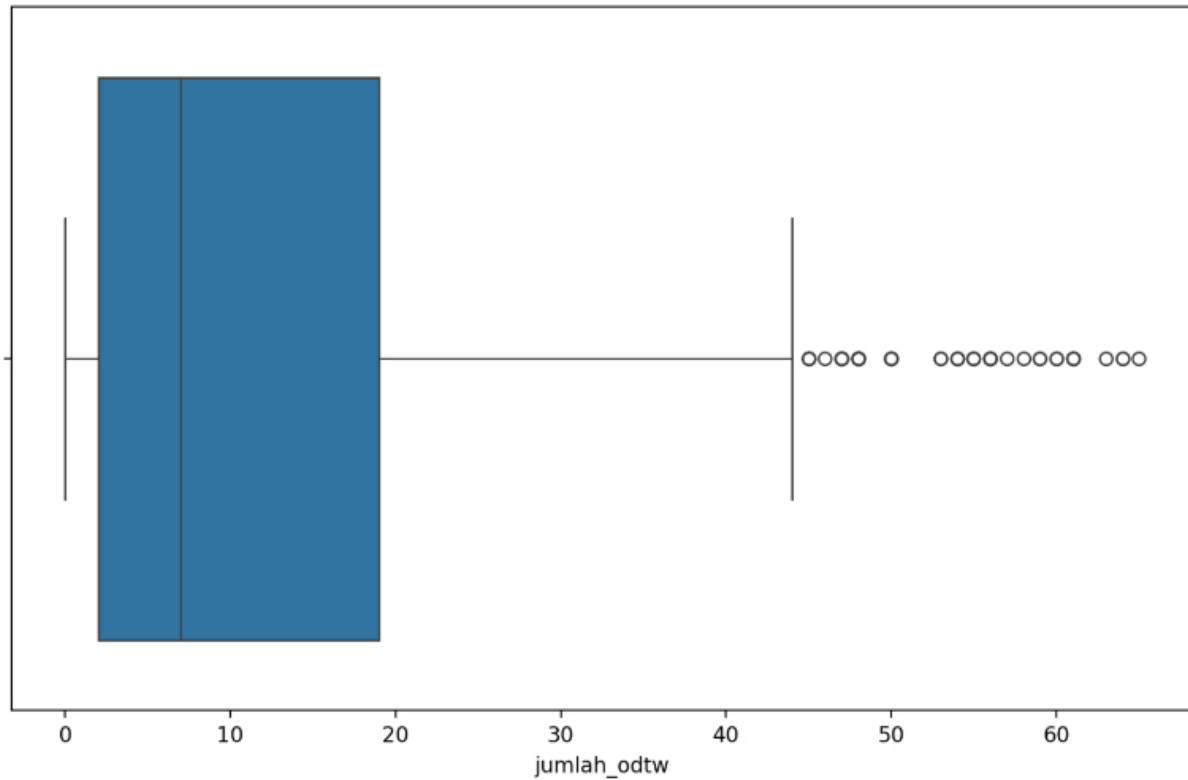
## DATA VISUALIZATION

Visualisasi data jumlah obyek daya tarik wisata berdasarkan jenis.





Visualisasi distribusi jumlah ODTW.



#### MODELING

Penerapan Model Linear Regression untuk prediksi jumlah obyek daya tarik wisata.

Evaluation Metrics:

Mean Absolute Error (MAE): 11.319836409081589

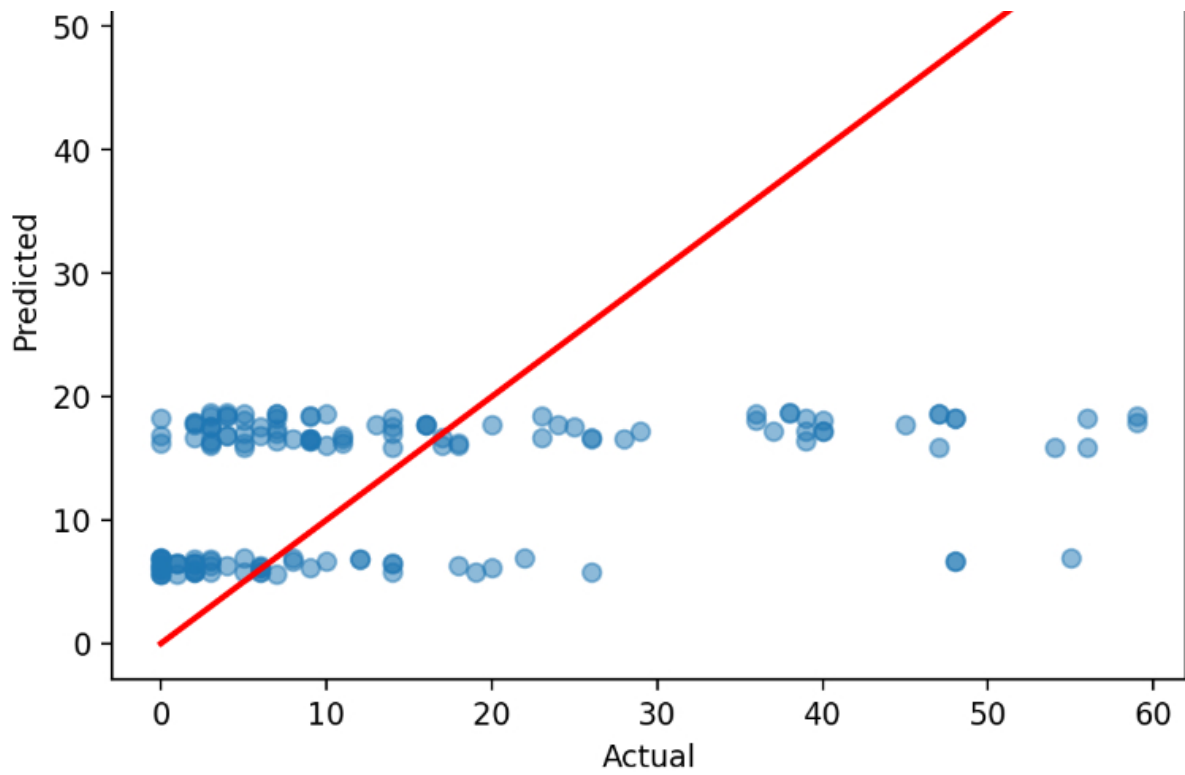
Mean Squared Error (MSE): 227.56585283998422

Root Mean Squared Error (RMSE): 15.08528597143535

R2 Score: 0.1129428947312121

#### Actual vs Predicted



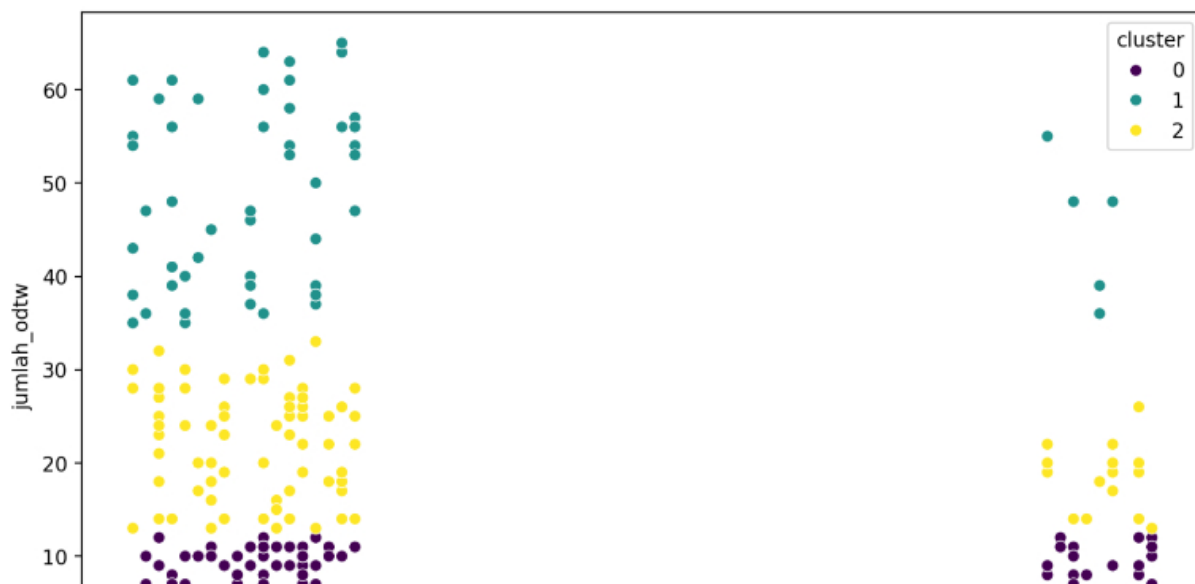


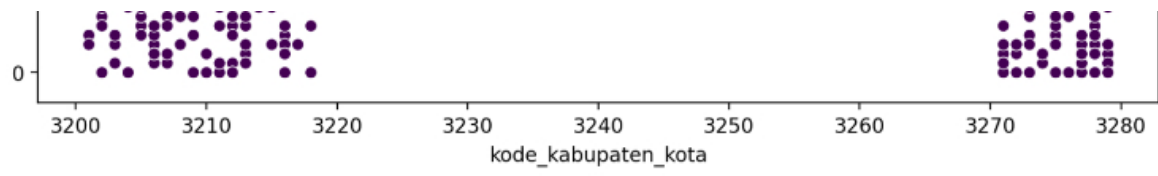
#### K-MEANS CLUSTERING

Clustering menggunakan K-Means berdasarkan jumlah ODTW.

Hasil clustering:

	id	kode_provinsi	nama_provinsi	kode_kabupaten_kota	nama_kabupaten_kota	jenis_odtw	jumlah
0	1	32	0	3,201	KABUPATEN BOGOR	ALAM	
1	2	32	0	3,201	KABUPATEN BOGOR	BUDAYA	
2	3	32	0	3,201	KABUPATEN BOGOR	MINAT KHUSUS	
3	4	32	0	3,202	KABUPATEN SUKABUMI	ALAM	
4	5	32	0	3,202	KABUPATEN SUKABUMI	BUDAYA	





run success