

TAMPLATE CAPSTONE PROJECT MARI BISNIS

NURUL FIKRIYATIL ANIQOH

DAI 006

1. Ambillah dataset MariBisnis di

<https://1drv.ms/u/s!AsD9wYmLMhwamk2qqOFAqKqFwo-Gg?e=177SLV;>

2. Buka dan kaji apa saja yang dapat diambil dari dataset tersebut;

Menggunakan google collab untuk melakukan loading dan cheking data

Berikut dokumentasinya

- Melihat file csv

```
import pandas as pd
import numpy as np

[ ] #membaca file csv
house_df = pd.read_csv('/content/MariBisnis.csv')
house_df.head()
```

	id	date	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	view	condition
0	7129300520	20141013T000000	221900.0	3	1.00	1180	5650	1.0	0	0	3
1	6414100192	20141209T000000	538000.0	3	2.25	2570	7242	2.0	0	0	3
2	5631500400	20150225T000000	180000.0	2	1.00	770	10000	1.0	0	0	3
3	2487200875	20141209T000000	604000.0	4	3.00	1960	5000	1.0	0	0	5
4	1954400510	20150218T000000	510000.0	3	2.00	1680	8080	1.0	0	0	3

Pada data maribisnis terdapat 21 kolom dan 21613 baris

ID :Identifikasi

Date : Tanggal terjual

Price : Harga penjualan

Bedroom : Jumlah kamar tidur

Bathrooms : Jumlah kamar mandi

sqft_liv : Ukuran ruang tamu dalam kaki persegi

sqft_lot : Ukuran kavling dalam kaki persegi

floors :Nomor lantai

waterfront : '1' jika properti memiliki tepi laut, '0' jika tidak.

View : Indeks dari 0 hingga 4 tentang seberapa bagus pemandangan properti itu

Condition : Kondisi rumah, peringkat 1 sampai 5

Grade : Klasifikasi menurut kualitas konstruksi yang mengacu pada jenis bahan yang digunakan dan kualitas pengerjaan. Bangunan dengan kualitas yang lebih baik (kelas yang lebih tinggi) lebih mahal untuk dibangun per unit ukuran dan memiliki nilai yang lebih tinggi.

sqft_above : Kaki persegi di atas tanah

sqft_basmt : Kaki persegi di bawah tanah

thn_built : Tahun dibangun

thn_renov : Tahun direnovasi. '0' jika tidak pernah direnovasi

zipcode : 5 digit kode pos

lat : Garis Lintang

long :Garis bujur

sqft_liv15 : Ukuran rata-rata ruang hidup perumahan interior untuk 15 rumah terdekat, dalam kaki persegi

sqft_lot15 : Rata-rata ukuran kavling untuk 15 rumah terdekat, dalam kaki persegi

- Melihat ringkasan data

```
[ ] #melihat ringkasan data
house_df.describe()
```

	id	price	bedrooms	bathrooms	sqft_living	sqft_lot	floors	waterfront	view	condition
count	2.161300e+04	2.161300e+04	21613.000000	21613.000000	21613.000000	2.161300e+04	21613.000000	21613.000000	21613.000000	21613.000000
mean	4.580302e+09	5.400881e+05	3.370842	2.114757	2079.899736	1.510697e+04	1.494309	0.007542	0.234303	3.409430
std	2.876566e+09	3.671272e+05	0.930062	0.770163	918.440897	4.142051e+04	0.539989	0.086517	0.766318	0.650743
min	1.000102e+06	7.500000e+04	0.000000	0.000000	290.000000	5.200000e+02	1.000000	0.000000	0.000000	1.000000
25%	2.123049e+09	3.219500e+05	3.000000	1.750000	1427.000000	5.040000e+03	1.000000	0.000000	0.000000	3.000000
50%	3.904930e+09	4.500000e+05	3.000000	2.250000	1910.000000	7.618000e+03	1.500000	0.000000	0.000000	3.000000
75%	7.308900e+09	6.450000e+05	4.000000	2.500000	2550.000000	1.068800e+04	2.000000	0.000000	0.000000	4.000000
max	9.900000e+09	7.700000e+06	33.000000	8.000000	13540.000000	1.651359e+06	3.500000	1.000000	4.000000	5.000000

- Melihat tipe data

```
#melihat informasi dataset
house_df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 21613 entries, 0 to 21612
Data columns (total 21 columns):
#   Column                Non-Null Count  Dtype  
---  --
0   id                     21613 non-null  int64  
1   date                   21613 non-null  object  
2   price                  21613 non-null  float64 
3   bedrooms               21613 non-null  int64  
4   bathrooms              21613 non-null  float64 
5   sqft_living            21613 non-null  int64  
6   sqft_lot               21613 non-null  int64  
7   floors                 21613 non-null  float64 
8   waterfront             21613 non-null  int64  
9   view                   21613 non-null  int64  
10  condition              21613 non-null  int64  
11  grade                  21613 non-null  int64  
12  sqft_above             21613 non-null  int64  
13  sqft_basement          21613 non-null  int64  
14  yr_built               21613 non-null  int64  
15  yr_renovated           21613 non-null  int64  
16  zipcode                21613 non-null  int64  
17  lat                    21613 non-null  float64 
18  long                   21613 non-null  float64 
19  sqft_living15          21613 non-null  int64  
20  sqft_lot15             21613 non-null  int64  
dtypes: float64(5), int64(15), object(1)
memory usage: 3.5+ MB
```

- Mengecek missing value

```
#melihat missing value
house_df.isnull().sum()

id          0
date        0
price       0
bedrooms    0
bathrooms   0
sqft_living  0
sqft_lot    0
floors      0
waterfront  0
view        0
condition   0
grade       0
sqft_above  0
sqft_basement 0
yr_built    0
yr_renovated 0
zipcode     0
lat         0
long        0
sqft_living15 0
sqft_lot15  0
dtype: int64
```

- Mengecek korelasi

```
[ ] #melihat korelasi antar price dengan setiap kolom
house_df.corr()["price"].sort_values(ascending=False)
```

```
price          1.000000
sqft_living    0.702035
grade          0.667434
sqft_above     0.605567
sqft_living15  0.585379
bathrooms      0.525138
view           0.397293
sqft_basement  0.323816
bedrooms       0.308350
lat            0.307003
waterfront     0.266369
floors         0.256794
yr_renovated   0.126434
sqft_lot       0.089661
sqft_lot15     0.082447
yr_built       0.054012
condition      0.036362
long           0.021626
id             -0.016762
zipcode        -0.053203
Name: price, dtype: float64
```

3. Lakukan pre-processing dan susun sebuah model data yang akan digunakan
4. Buatlah model Machine Learning seperti regresi, klasifikasi atau clustering;
Langkah-langkah membuat model regresi machine learning :
 1. Membuat *workspace machine learning* di azure
 2. Membuat *compute instance*
 3. Mengunggah data yang akan dibuat model *machine learning* regresi
 4. Membuat designer *machine learning* regresi
 5. Mengunduh data *score model*
 6. Membuat *inference real time* dan mencoba memuat data manual dengan format csv
 7. Mendeploy model yang telah dibuat
 8. Mengetes model yang telah dibuat di bagian *endpoint*
 9. Membuat script model di notebook dan memasukkan *key* dan *endpoint*
5. Merancang *report* berdasarkan dataset yang dipilih;
 1. Buka power BI desktop
 2. Pilih *get data* untuk mengunggah data yang akan di buat report. Data yang akan dibuatkan *report* ada dua yaitu data asli penjualan rumah dan data prediksi harga rumah yang didapatkan dari model regresi yang telah dibuat sebelumnya.
 3. Melakukan transformasi data yang akan divisualisasikan

4. Memilihh kolom mana aja yang akan divisualisasikan
5. Membuat analisis
6. Merancang dasbor berbasis desktop yang dapat digunakan pada saat rapat bisnis