

MUHAMMAD TARMIZI BARIQ

51422161

2IA11

MUHAMMAD TARMIZI BARIQ.ipynb

File Edit View Insert Runtime Tools Help

Overview Dataset

"Data Komponen Beton".

Data set ini berisikan beberapa kategori sebagai berikut

- cement (kg)
- slag (blast furnace slag, kg)
- ash (fly ash, kg)
- water (kg)
- superplastic (superplasticizer, kg)
- coarseagg (coarse aggregate, kg)
- fineagg (fine aggregate, kg)
- age (days, 1-365)
- strength (Concrete compressive strength, MPa)

NAMA : MUHAMMAD TARMIZI BARIQ

KELAS : 2IA11

NPM : 51422161

Exploratory Data Analysis (EDA)

Instruksi Tugas Praktikum FIKTI dan FTI

- Buat univariate dan bivariate analysis untuk data set di atas
- Jelaskan kesimpulan yang bisa di ambil dari analysis yg anda lakukan

Hot weather

completed at 2:07 PM

MUHAMMAD TARMIZI BARIQ.ipynb

File Edit View Insert Runtime Tools Help All changes saved

1. Pengambilan Data:

1.1 Import packages

```
[14] import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
import warnings
warnings.filterwarnings("ignore")
from scipy.stats import ttest_ind, ttest_rel
from scipy import stats
```

1.2 Import data

```
[15] data = pd.read_csv("https://raw.githubusercontent.com/supasonix/ATA-praktikum-01/main/concrete.csv")
data.head(10)
```

	cement	slag	ash	water	superplastic	coarseagg	fineagg	age	strength
0	141.3	212.0	0.0	203.5	0.0	971.5	748.5	28	29.89
1	168.9	42.2	124.3	198.3	10.8	1080.8	796.2	14	23.51
2	230.0	0.0	95.7	187.4	3.5	956.9	861.2	28	29.22
3	266.0	114.0	0.0	228.0	0.0	932.0	670.0	28	45.85
4	154.8	183.4	0.0	193.3	9.1	1047.4	696.7	28	18.29
5	255.0	0.0	0.0	192.0	0.0	899.8	945.0	90	21.86
6	166.8	250.2	0.0	203.5	0.0	975.6	692.6	7	15.75
7	251.4	0.0	118.3	188.5	6.4	1028.4	757.7	56	36.64
8	298.0	0.0	0.0	192.0	0.0	1085.0	765.0	28	21.65

Starting background recording
Microphone recording is on.
Press **ESC** - **G** for Xbox Game Bar

completed at 2:07 PM

```

MUHAMMAD TARMIDZI BARIQ.ipynb
File Edit View Insert Runtime Tools Help All changes saved
+ Code + Text
[6] 8 296.0 0.0 0.0 192.0 0.0 1085.0 765.0 28 21.65
9 155.0 184.0 143.0 194.0 9.0 880.0 699.0 28 28.99
+ Code + Text

2. EDA

2.1. Univariate analysis

2.1.1 [Component Name]

I. Distribusi

[16] # Buat diagram di sini

menampilkan jumlah kolom pada shape 1 dan 0

[17] print("jumlah kolom pada data - ", data.shape[1])
print("jumlah kolom pada data - ", data.shape[0])

jumlah kolom pada data = 9
jumlah baris pada data = 1030

menampilkan data describe

[18] data.describe()

count      cement      slag      ash      water      superplastic      coarseagg      fineagg      age      strength
count  1030.000000  1030.000000  1030.000000  1030.000000  1030.000000  1030.000000  1030.000000  1030.000000  1030.000000
mean      281.167864    73.895825    54.188350    181.567282    6.204660    972.918932    773.580485    45.662136    35.817961
std       104.506364    86.279342    63.997004    21.354219    5.973841    77.739954    80.173980    63.169912    16.705742
min       102.000000    0.000000    0.000000    121.800000    0.000000    801.000000    994.000000    1.000000    2.330000
25%      192.375000    0.000000    0.000000    164.900000    0.000000    832.000000    730.950000    7.000000    23.710000
50%      272.900000    22.000000    0.000000    185.000000    6.400000    968.000000    778.500000    28.000000    34.445000
75%      350.000000    142.950000    116.300000    192.000000    10.200000    1029.400000    824.000000    56.000000    46.135000
max       540.000000    359.400000    200.100000    247.000000    32.200000    1145.000000    992.600000    365.000000    82.600000

menampilkan info dari data

[19] data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1030 entries, 0 to 1029
Data columns (total 9 columns):
 #   Column        Non-Null Count  Dtype  
---  --
 0   cement        1030 non-null   float64
 1   slag          1030 non-null   float64
 2   ash           1030 non-null   float64
 3   water         1030 non-null   float64
 4   superplastic  1030 non-null   float64
 5   coarseagg     1030 non-null   float64
 6   fineagg       1030 non-null   float64
 7   age           1030 non-null   int64  
 8   strength      1030 non-null   float64
dtypes: float64(8), int64(1)
memory usage: 72.5 KB

menampilkan data jumlah
```

```

MUHAMMAD TARMIDZI BARIQ.ipynb
File Edit View Insert Runtime Tools Help All changes saved
+ Code + Text
[16] data.describe()

count      cement      slag      ash      water      superplastic      coarseagg      fineagg      age      strength
count  1030.000000  1030.000000  1030.000000  1030.000000  1030.000000  1030.000000  1030.000000  1030.000000  1030.000000
mean      281.167864    73.895825    54.188350    181.567282    6.204660    972.918932    773.580485    45.662136    35.817961
std       104.506364    86.279342    63.997004    21.354219    5.973841    77.739954    80.173980    63.169912    16.705742
min       102.000000    0.000000    0.000000    121.800000    0.000000    801.000000    994.000000    1.000000    2.330000
25%      192.375000    0.000000    0.000000    164.900000    0.000000    832.000000    730.950000    7.000000    23.710000
50%      272.900000    22.000000    0.000000    185.000000    6.400000    968.000000    778.500000    28.000000    34.445000
75%      350.000000    142.950000    116.300000    192.000000    10.200000    1029.400000    824.000000    56.000000    46.135000
max       540.000000    359.400000    200.100000    247.000000    32.200000    1145.000000    992.600000    365.000000    82.600000

menampilkan info dari data

[19] data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1030 entries, 0 to 1029
Data columns (total 9 columns):
 #   Column        Non-Null Count  Dtype  
---  --
 0   cement        1030 non-null   float64
 1   slag          1030 non-null   float64
 2   ash           1030 non-null   float64
 3   water         1030 non-null   float64
 4   superplastic  1030 non-null   float64
 5   coarseagg     1030 non-null   float64
 6   fineagg       1030 non-null   float64
 7   age           1030 non-null   int64  
 8   strength      1030 non-null   float64
dtypes: float64(8), int64(1)
memory usage: 72.5 KB

menampilkan data jumlah
```

```

MUHAMMAD TARMIDZI BARIQ.ipynb
File Edit View Insert Runtime Tools Help All changes saved
+ Code + Text
[20] data.isnull().sum()

cement      0
slag        0
ash          0
water       0
superplastic  0
coarseagg   0
fineagg     0
age         0
strength    0
dtype: int64

menduplicate data dan di jumlah kan ada brp

[21] data.duplicated().sum()

25

menghapus data duplicate

[22] data.drop_duplicates(inplace=True)

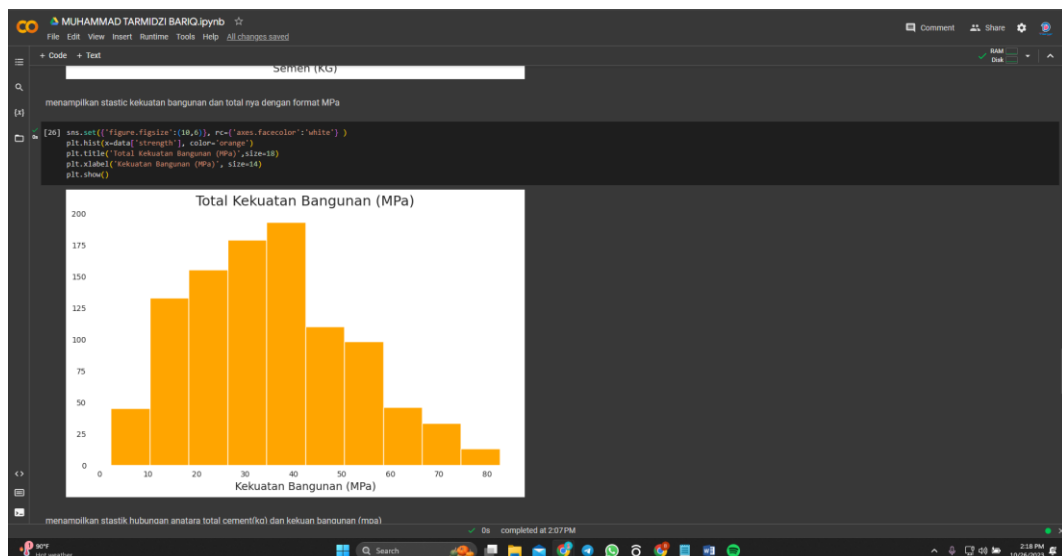
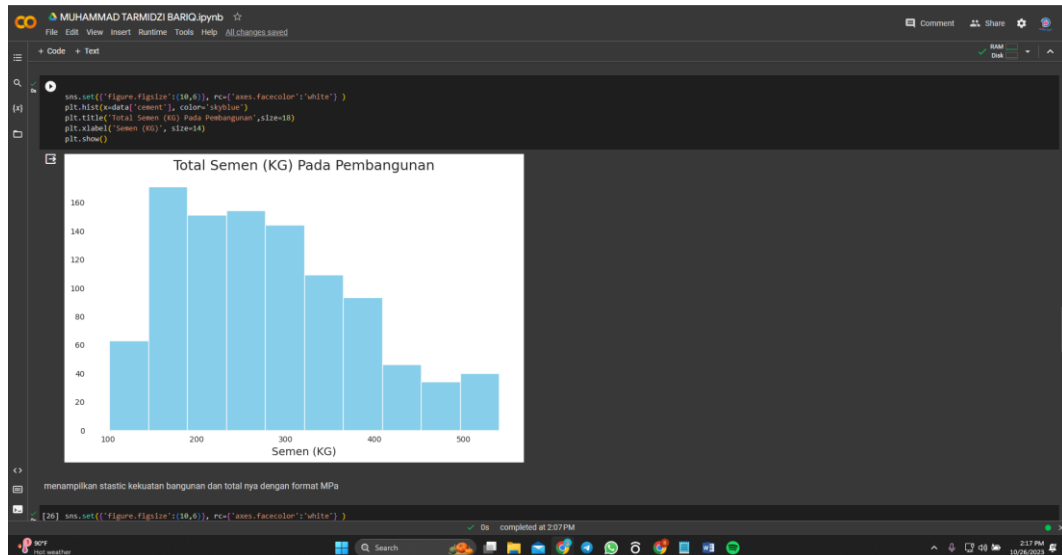
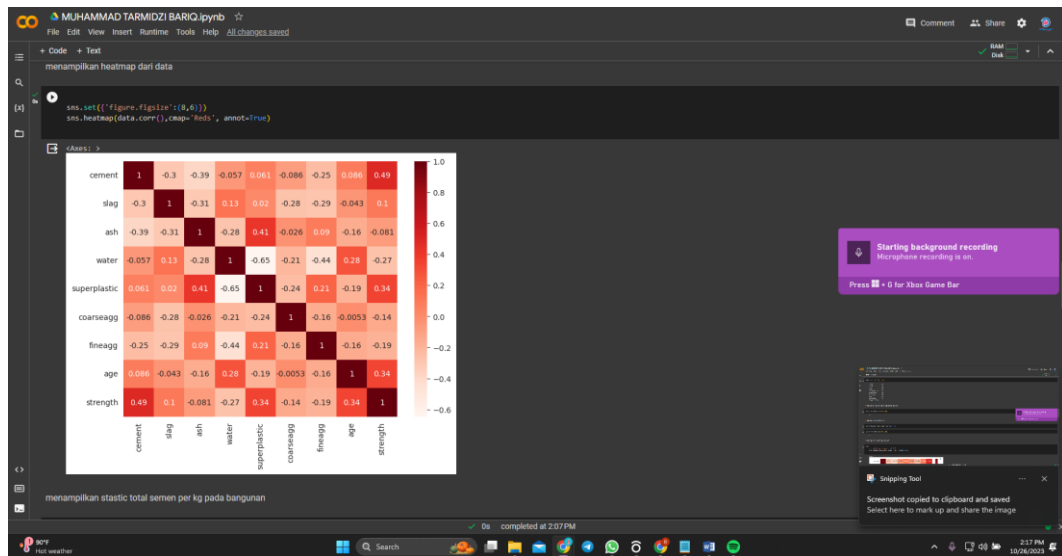
[23] data.duplicated().sum()

0

menampilkan heatmap dari data

[24] sns.set('figure.figsize',(8,6))
sns.heatmap(data.corr(),cmap='magma',annot=True)

<Figure>
cment  1    0.3  -0.39  -0.057  0.061  -0.086  -0.25  0.086  0.48  1.0
```



MUHAMMAD TARMIDZI BARIQ.ipynb

menampilkan statistik hubungan antara total cement(kg) dan kekuatan bangunan (mpa)

ii. Deskripsi

Jelaskan diagram di atas

2.2.Bivariate analysis

2.2.1 [Component Name]

i. Boxplot

[27] # Buat diagram di sini

ii. Deskripsi

Jelaskan diagram di atas

iii. Uji Hipotesis

Two sampled T-test - Uji Independent Samples t Test atau 2-sample t-test membandingkan rata-rata dua kelompok independen untuk menentukan apakah ada bukti statistik bahwa rata-rata populasi terkait berbeda secara signifikan.

Tentukan null dan alternate hypothesis.

Secara umum, null hypothesis dan menyatakan bahwa kedua populasi yang diuji tidak memiliki perbedaan yang signifikan secara statistik. alternate hypothesis akan menyatakan bahwa ada perbedaan.

Starting background recording
Microphone recording is on.
Press **⌘ + G** for Xbox Game Bar

Screenshot copied to clipboard and saved
Select here to mark up and share the image

MUHAMMAD TARMIDZI BARIQ.ipynb

iii. Uji Hipotesis

Two sampled T-test - Uji Independent Samples t Test atau 2-sample t-test membandingkan rata-rata dua kelompok independen untuk menentukan apakah ada bukti statistik bahwa rata-rata populasi terkait berbeda secara signifikan.

Tentukan null dan alternate hypothesis.

Secara umum, null hypothesis dan menyatakan bahwa kedua populasi yang diuji tidak memiliki perbedaan yang signifikan secara statistik. alternate hypothesis akan menyatakan bahwa ada perbedaan.

Dalam contoh ini kita dapat mengatakan bahwa:

- Null Hypothesis - tidak ada perbedaan Rerata usia kelompok pendapatan >50k dan kelompok pendapatan <=50k.
- Alternate Hypothesis - ada perbedaan Rerata usia kelompok pendapatan >50k dan kelompok pendapatan <=50k.

[28] # Buat uji hipotesis untuk menentukan null atau alternate hypothesis

```
sns.set(style='figure.figsize':(10,6)), rc={'axes.facecolor':'white'})
plt.scatter(x=data['cement'], y=data['strength'], color='purple')
plt.title('Hubungan antara Total Cement (KG) dan Kekuatan Bangunan (MPa)',size=18)
plt.xlabel('Semen (KG)',size=14)
plt.ylabel('Kekuatan (MPa)',size=14)
plt.show()
```

Hubungan antara Total Cement(KG) dan Kekuatan Bangunan (MPa)

MUHAMMAD TARMIDZI BARIQ.ipynb

[28] # Buat uji hipotesis untuk menentukan null atau alternate hypothesis

```
sns.set(style='figure.figsize':(10,6)), rc={'axes.facecolor':'white'})
plt.scatter(x=data['cement'], y=data['strength'], color='purple')
plt.title('Hubungan antara Total Cement (KG) dan Kekuatan Bangunan (MPa)',size=18)
plt.xlabel('Semen (KG)',size=14)
plt.ylabel('Kekuatan (MPa)',size=14)
plt.show()
```

Hubungan antara Total Cement(KG) dan Kekuatan Bangunan (MPa)

