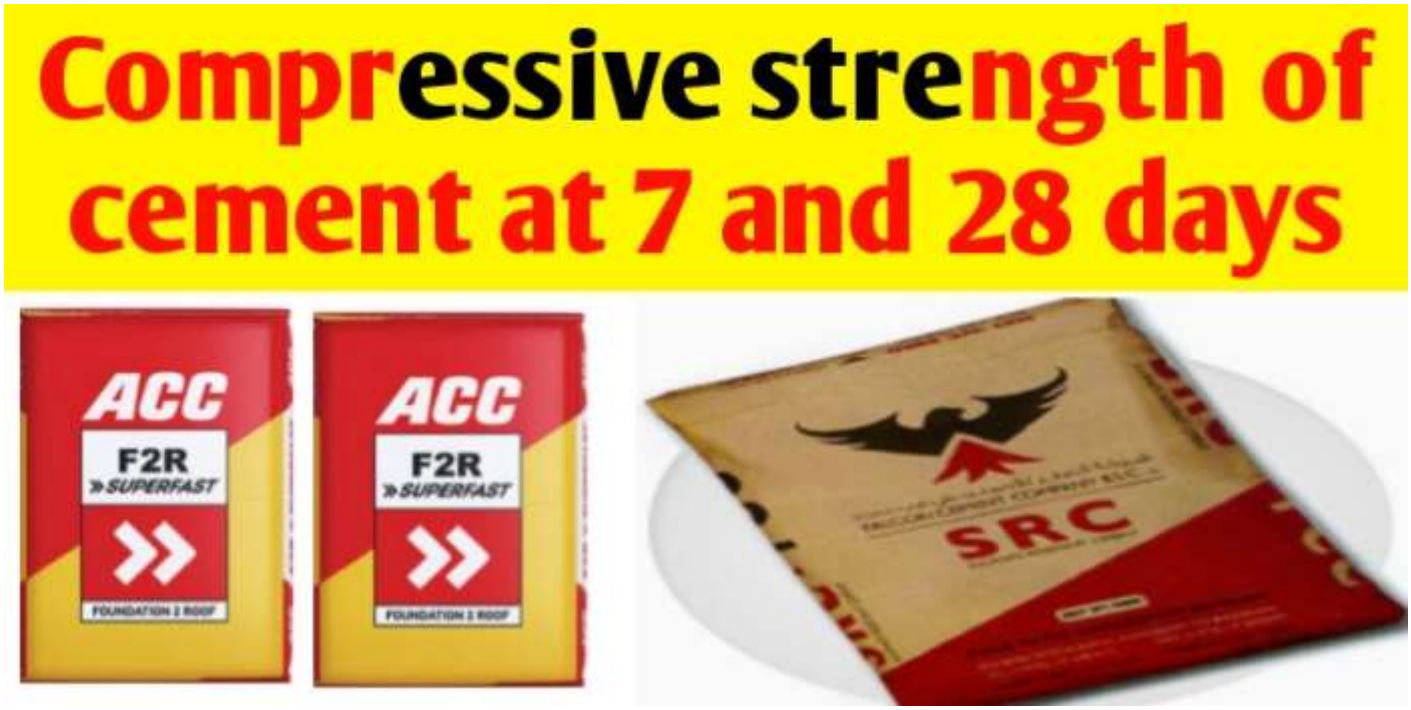
⭐ Regression Model to Predict Cement Compressive Strength




```
# import library
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns


# import data
cement = pd.read_csv('https://github.com/ybifoundation/Dataset/raw/main/Concrete%20Compres


# view data
cement.head()
```

	Cement (kg in a m^3 mixture)	Blast Furnace Slag (kg in a m^3 mixture)	Fly Ash (kg in a m^3 mixture)	Water (kg in a m^3 mixture)	Superplasticizer (kg in a m^3 mixture)	Coarse Aggregate (kg in a m^3 mixture)	Fine Aggregate (kg in a m^3 mixture)	(d
0	540.0	0.0	0.0	162.0	2.5	1040.0	676.0	

```
# info of data
cement.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1030 entries, 0 to 1029
Data columns (total 9 columns):
 #   Column                                          Non-Null Count  Dtype  
---  ------                                          --------------  -----  
 0   Cement (kg in a m^3 mixture)                    1030 non-null   float64
 1   Blast Furnace Slag (kg in a m^3 mixture)        1030 non-null   float64
 2   Fly Ash (kg in a m^3 mixture)                   1030 non-null   float64
 3   Water (kg in a m^3 mixture)                     1030 non-null   float64
 4   Superplasticizer (kg in a m^3 mixture)          1030 non-null   float64
 5   Coarse Aggregate (kg in a m^3 mixture)          1030 non-null   float64
 6   Fine Aggregate (kg in a m^3 mixture)            1030 non-null   float64
 7   Age (day)                                       1030 non-null   int64  
 8   Concrete Compressive Strength(MPa, megapascals) 1030 non-null   float64
dtypes: float64(8), int64(1)
memory usage: 72.5 KB
```

```
# summary statistics
cement.describe()
```

Blast Coars

```
# check for missing value
cement.isna().sum()
```

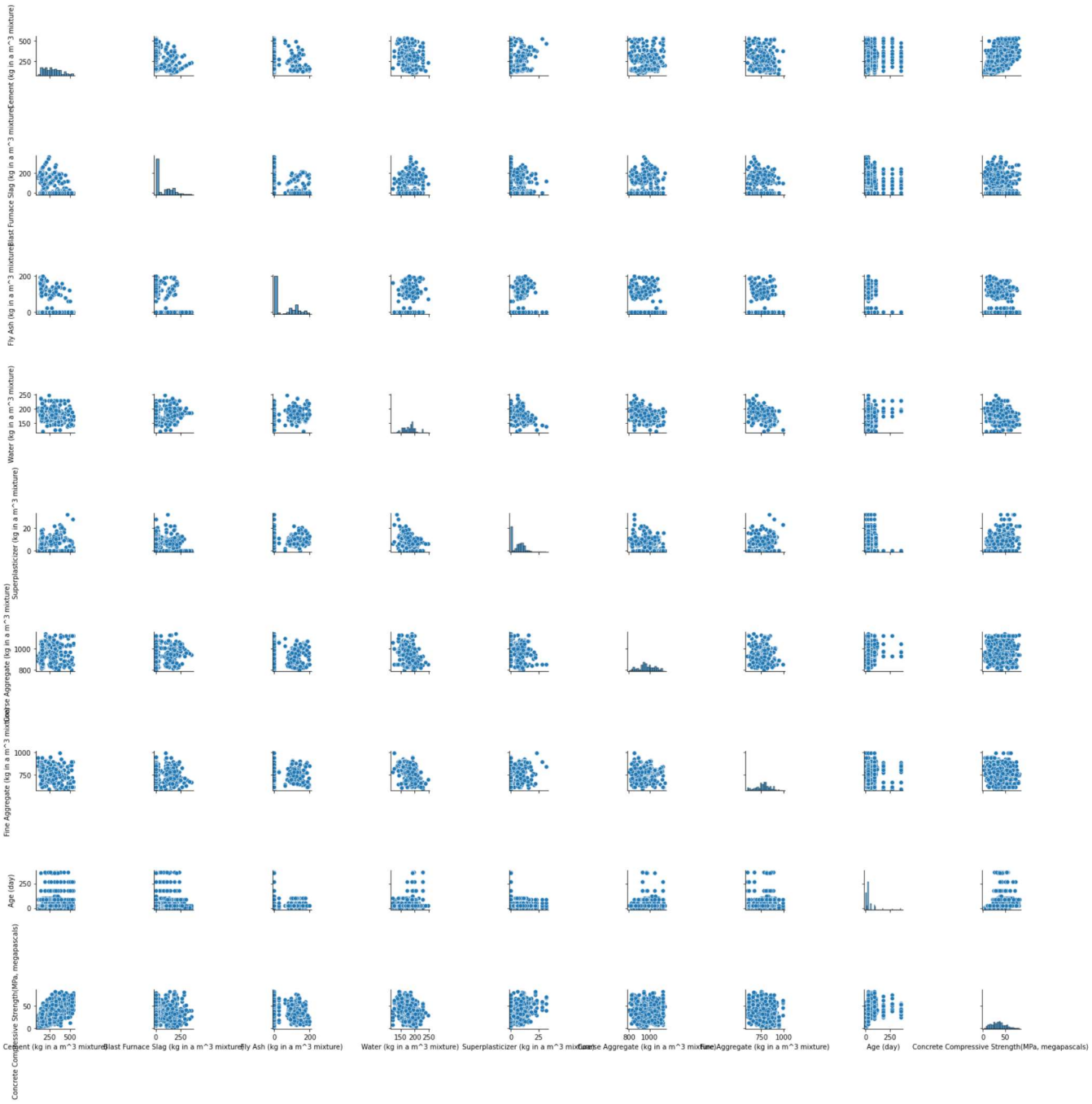
```
Cement (kg in a m^3 mixture)                        0
Blast Furnace Slag (kg in a m^3 mixture)            0
Fly Ash (kg in a m^3 mixture)                       0
Water (kg in a m^3 mixture)                         0
Superplasticizer (kg in a m^3 mixture)              0
Coarse Aggregate (kg in a m^3 mixture)              0
Fine Aggregate (kg in a m^3 mixture)                0
Age (day)                                           0
Concrete Compressive Strength(MPa, megapascals)     0
dtype: int64
```

```
# check for categories
cement.nunique()
```

```
Cement (kg in a m^3 mixture)                        280
Blast Furnace Slag (kg in a m^3 mixture)            187
Fly Ash (kg in a m^3 mixture)                       163
Water (kg in a m^3 mixture)                         205
Superplasticizer (kg in a m^3 mixture)              155
Coarse Aggregate (kg in a m^3 mixture)              284
Fine Aggregate (kg in a m^3 mixture)                304
Age (day)                                            14
Concrete Compressive Strength(MPa, megapascals)     938
dtype: int64
```

```
# visualize pairplot
sns.pairplot(cement)
```

<seaborn.axisgrid.PairGrid at 0x7fb8158a6610>



```
# columns name
cement.columns
```

```
Index(['Cement (kg in a m^3 mixture)',
       'Blast Furnace Slag (kg in a m^3 mixture)',
       'Fly Ash (kg in a m^3 mixture)', 'Water (kg in a m^3 mixture)',
       'Superplasticizer (kg in a m^3 mixture)',
       'Coarse Aggregate (kg in a m^3 mixture)',
       'Fine Aggregate (kg in a m^3 mixture)', 'Age (day)',
       'Concrete Compressive Strength(MPa, megapascals) '],
      dtype='object')
```

```
# define y
y = cement['Concrete Compressive Strength(MPa, megapascals) ']
```

```
# define X
X = cement[['Cement (kg in a m^3 mixture)',
'Blast Furnace Slag (kg in a m^3 mixture)',
'Fly Ash (kg in a m^3 mixture)', 'Water (kg in a m^3 mixture)',
'Superplasticizer (kg in a m^3 mixture)',
'Coarse Aggregate (kg in a m^3 mixture)',
'Fine Aggregate (kg in a m^3 mixture)', 'Age (day)']]
```

```
# split data
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(
X, y, test_size=.7, random_state=2529)
```

```
# verify shape
X_train.shape, X_test.shape, y_train.shape, y_test.shape
```

```
((309, 8), (721, 8), (309,), (721,))
```

```
# select model
from sklearn.linear_model import LinearRegression
model = LinearRegression()
```

```
# train model
```

```
model.fit(X_train, y_train)
```

```
LinearRegression()
```

```
# predict with model
y_pred = model.predict(X_test)
```

```
# model evaluation
from sklearn.metrics import mean_absolute_error, mean_absolute_percentage_error, mean_squa
```

```
# model MAE
mean_absolute_error(y_test,y_pred)
```

```
8.396967949621466
```

```
# model MAPE
mean_absolute_percentage_error(y_test, y_pred)
```

```
0.31619362297440723
```

```
# model MSE
mean_squared_error(y_test, y_pred)
```

```
114.78747173290049
```

```
# future prediction
sample = cement.sample()
sample
```

	Cement (kg in a m^3 mixture)	Blast Furnace Slag (kg in a m^3 mixture)	Fly Ash (kg in a m^3 mixture)	Water (kg in a m^3 mixture)	Superplasticizer (kg in a m^3 mixture)	Coarse Aggregate (kg in a m^3 mixture)	Fine Aggregate (kg in a m^3 mixture)
468	213.5	0.0	174.24	159.21	11.66	1043.6	771.9

```
# define X_new
```

```
X_new = sample.loc[:,X.columns]
X_new
```

	Cement (kg in a m^3 mixture)	Blast Furnace Slag (kg in a m^3 mixture)	Fly Ash (kg in a m^3 mixture)	Water (kg in a m^3 mixture)	Superplasticizer (kg in a m^3 mixture)	Coarse Aggregate (kg in a m^3 mixture)	Fine Aggregate (kg in a m^3 mixture)
468	213.5	0.0	174.24	159.21	11.66	1043.6	771.9

```
# predict for X_new
model.predict(X_new)
```

```
array([29.90205555])
```