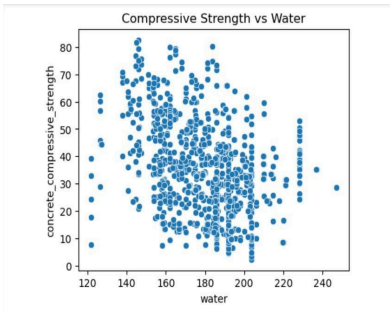
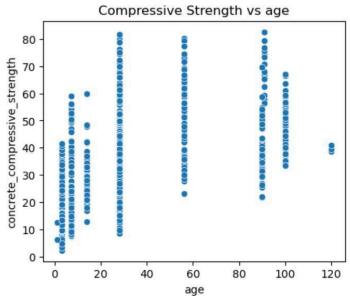
# Data Collection and Preprocessing Phase

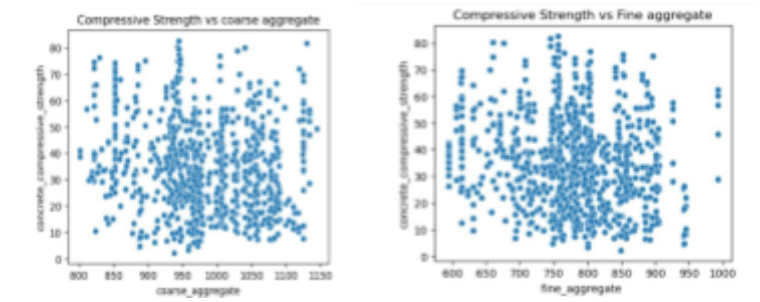
| | |
|---|---|
| Date | 08 July 2024 |
| Team ID | SWTID1720160264 |
| Project Title | Predicting Compressive Strength Of Concrete Using Machine Learning |
| Maximum Marks | 6 Marks |

## Data Exploration and Preprocessing Template

Identifies data sources, assesses quality issues like missing values and duplicates, and implements resolution plans to ensure accurate and reliable analysis.

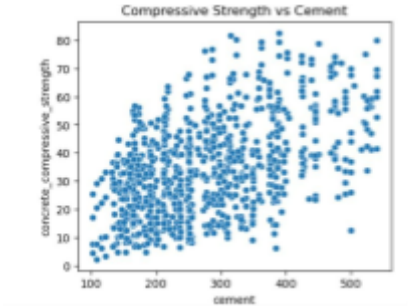
| Section | Description |
|---|---|
| Data Overview | Dimensions: `(1030, 9)`<br>Variables: `Index(['cement', 'blast_furnace_slag', 'fly_ash', 'water', 'superplasticizer', 'coarse_aggregate', 'fine_aggregate ', 'age', 'concrete_compressive_strength'], dtype='object')` |
| Univariate Analysis |  |
| Bivariate Analysis |  <br>  |

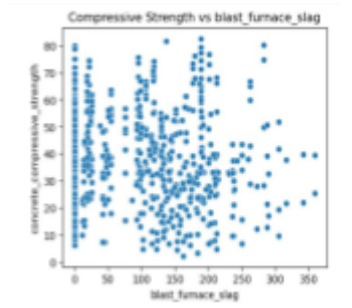
| | cement | blast_furnace_slag | fly_ash | water | superplasticizer | coarse_aggregate | fine_aggregate | age | concrete_compressive_strength |
|---|---|---|---|---|---|---|---|---|---|
| count | 1030.000000 | 1030.000000 | 1030.000000 | 1030.000000 | 1030.000000 | 1030.000000 | 1030.000000 | 1030.000000 | 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 | 77.753954 | 80.175980 | 63.169912 | 16.705742 |
| min | 102.000000 | 0.000000 | 0.000000 | 121.800000 | 0.000000 | 801.000000 | 594.000000 | 1.000000 | 2.330000 |
| 25% | 192.375000 | 0.000000 | 0.000000 | 164.900000 | 0.000000 | 932.000000 | 730.950000 | 7.000000 | 23.710000 |
| 50% | 272.900000 | 22.000000 | 0.000000 | 185.000000 | 6.400000 | 968.000000 | 779.500000 | 28.000000 | 34.445000 |
| 75% | 350.000000 | 142.950000 | 118.300000 | 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 |

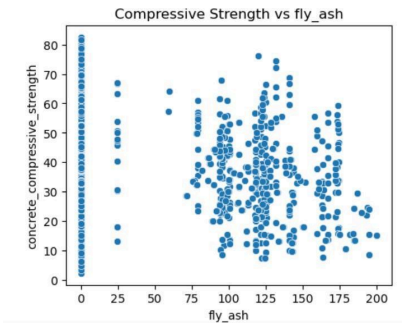


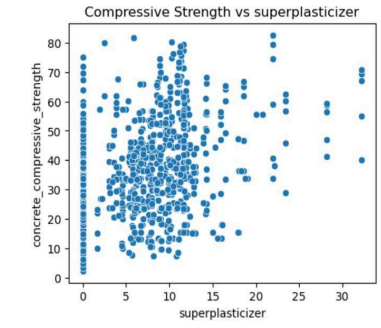






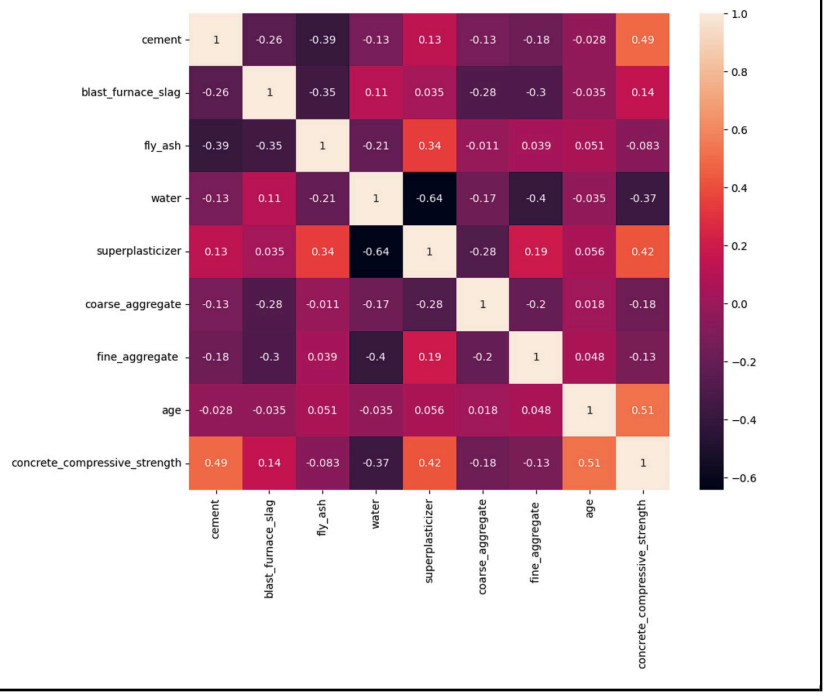



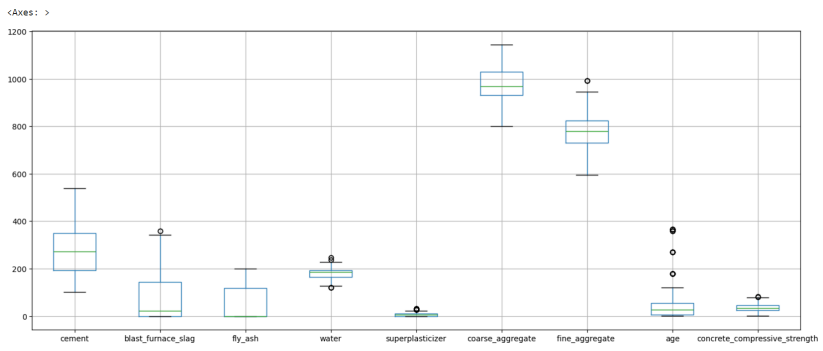




Multivariate Analysis

```
plt.figure(figsize=(10,8))
sns.heatmap(data.corr(),annot=True)
plt.show()
```



| | |
|---|---|
| Outliers and Anomalies |  |
| **Data Preprocessing Code Screenshots** | |
| Loading Data | `data=pd.read_csv('concrete_data.csv')` |
| Handling Missing Data | `data.isnull().any()` |
| Data Transformation | `from sklearn.preprocessing import StandardScaler`<br>`scaler=StandardScaler()`<br>`scaler.fit_transform(x)` |
| Feature Engineering | — |
| Save Processed Data | — |