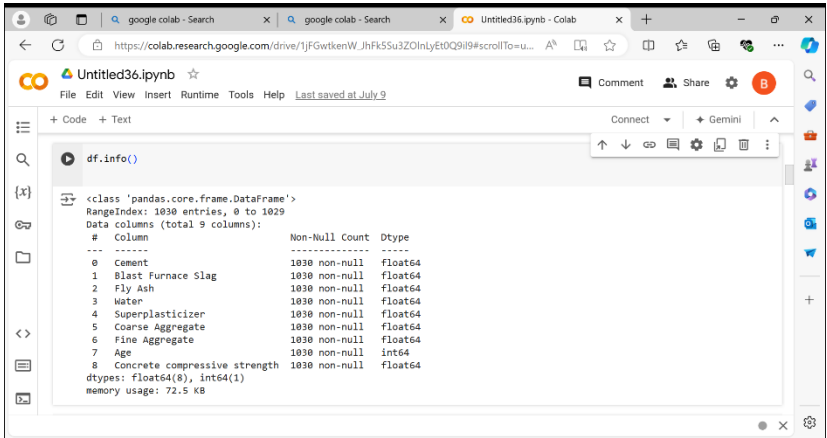


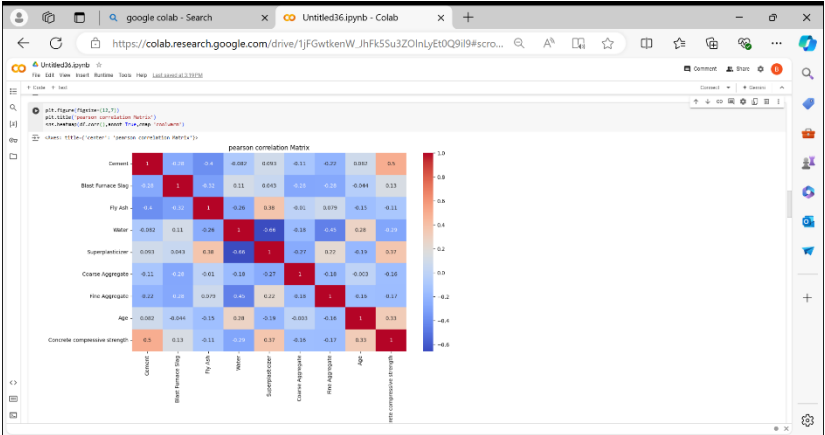
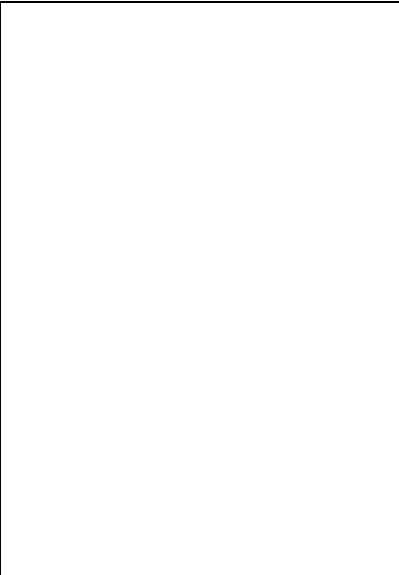
Data Collection and Preprocessing Phase

Date	6 July 2024
Team ID	739695
Project Title	Predicting the Compressive Strength of Concrete
Maximum Marks	6 Marks

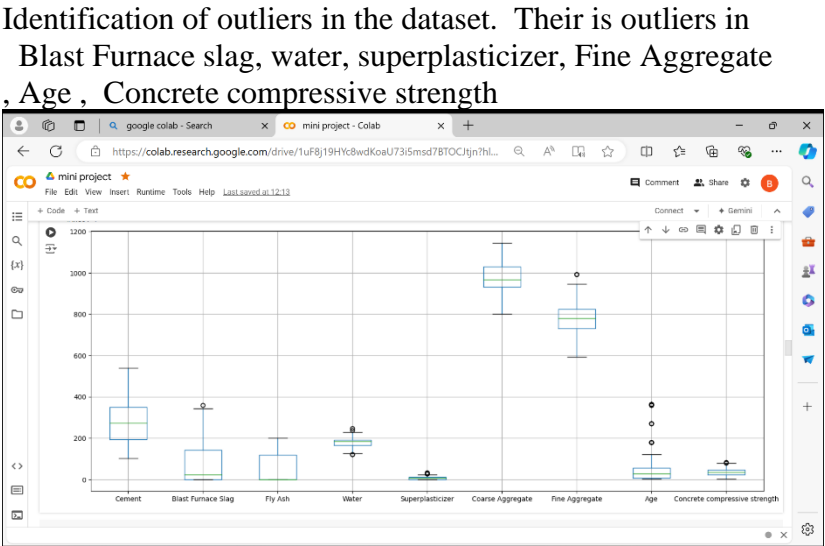
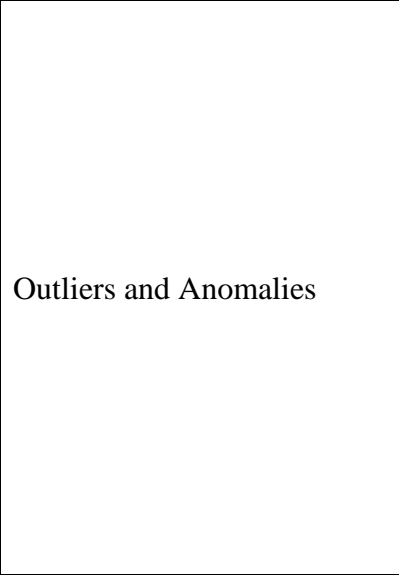
Data Exploration and Preprocessing Template

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	<div>1030 rows x 9 columns, dtypes: float64</div> <div></div>
Univariate Analysis	Exploration of individual of accuracy_score,mean_squared_error,r2_score,mean_absolute_error
Bivariate Analysis	Relationships between two variables (correlation, scatter plots)



Multivariate Analysis



Outliers and Anomalies

Data Preprocessing Code Screenshots

## Loading Data

```
File Edit View Insert Runtime Tools Help Last edited on July 4
+ Code + Text
Reading the dataset
df = pd.read_excel('/content/Concrete_Data.xls')
df
```

	Cement (component 1)(kg in a m³ mixture)	Blast Furnace Slag (component 2)(kg in a m³ mixture)	Fly Ash (component 3)(kg in a m³ mixture)	Water (component 4)(kg in a m³ mixture)	Superplasticizer (component 5)(kg in a m³ mixture)	Coarse Aggregate (component 6)(kg in a m³ mixture)	Fine Aggregate (component 7)(kg in a m³ mixture)	Age (day)	Concrete compressive strength(MPa, megapascals)
0	540.0	0.0	0.0	162.0	2.5	1040.0	676.0	28	79.986111
1	540.0	0.0	0.0	162.0	2.5	1055.0	676.0	28	61.887366
2	332.5	142.5	0.0	228.0	0.0	932.0	594.0	270	40.269535
3	332.5	142.5	0.0	228.0	0.0	932.0	594.0	365	41.052780
4	198.6	132.4	0.0	192.0	0.0	978.4	825.5	360	44.296075

## Handling Missing Data

```
df.isnull().any()
```

Cement	False
Blast Furnace Slag	False
Fly Ash	False
Water	False
Superplasticizer	False
Coarse Aggregate	False
Fine Aggregate	False
Age	False
Concrete compressive strength	False
dtype:	bool

## Data Transformation

Scaling:

```
[ ] scaler = StandardScaler()

x_train = scaler.fit_transform(x_train)
x_test = scaler.transform(x_test)

[ ] x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=42)
```

## Feature Engineering

modifying existing ones

```
[ ] # Splitting the features and target variable

cols = df.columns.drop('Concrete compressive strength')
x = df[cols]
y = df['Concrete compressive strength']
```

## Save Processed Data

Code to save the cleaned and processed data for future use.

```
[ ] pickle.dump(gbr,open('cement.pkl','wb'))
```

