



Data Collection and Preprocessing Phase

Date	6 July 2024
Team ID	739916
Project Title	Predicting the Compressive Strength of Concrete
Maximum Marks	6 Marks

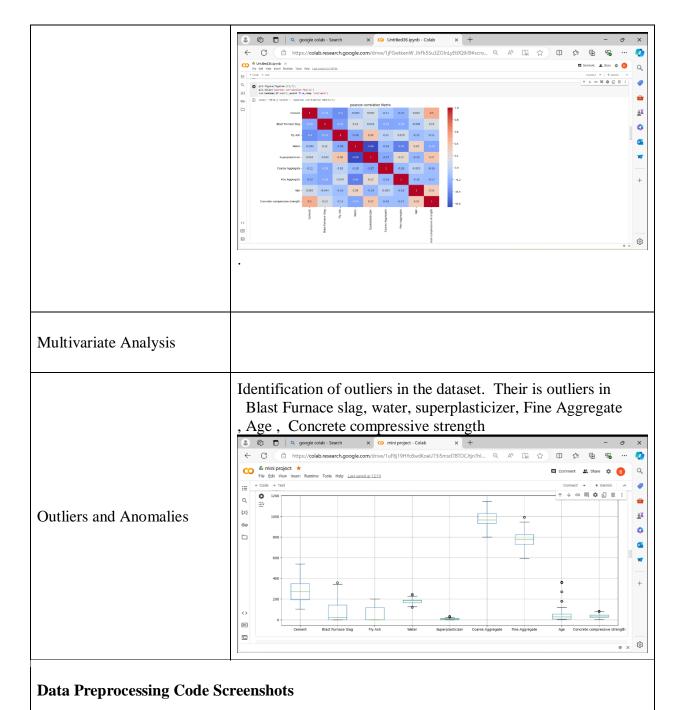
Data Exploration and Preprocessing Template

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	1030 rows x 9 columns, dtypes: float64	×
Univariate Analysis	Exploration of individual of accuracy_score,mean_squared_error,r2_score,mean_absolute_error	
Bivariate Analysis	Relationships between two variables (correlation, scatter plot	s)











Loading Data	File Edit View Insert Runtime Tools Help Lastedited on July 4 + Code + Text Connect Connec		
Handling Missing Data	df.isnull().any() Cement Blast Furnace Slag False Fly Ash Water False Superplasticizer Coarse Aggregate Fine Aggregate Age Concrete compressive strength dtype: bool False		
Data Transformation	<pre>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)</pre>		
Feature Engineering	modifying existing ones ature Engineering # 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.		



