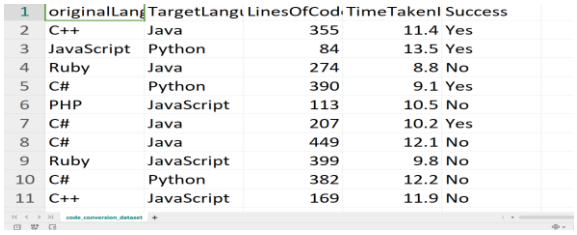
Data Collection and Preprocessing Phase

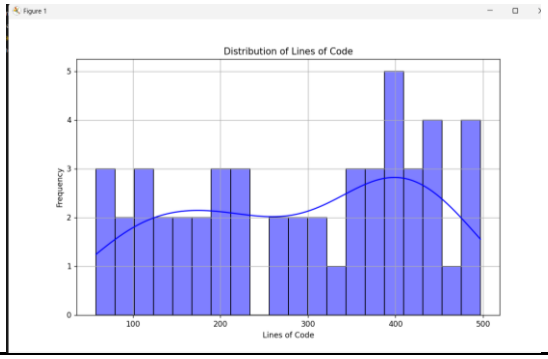
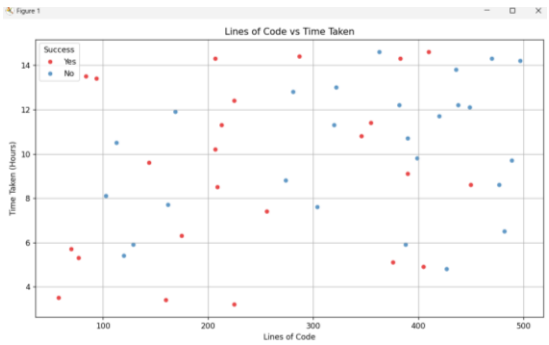
Date	16 July 2024
Team ID	
Project Title	CodeXchange: An AI-Powered Code Translator Tool using Palm's chat-bison-001
Maximum Marks	6 Marks

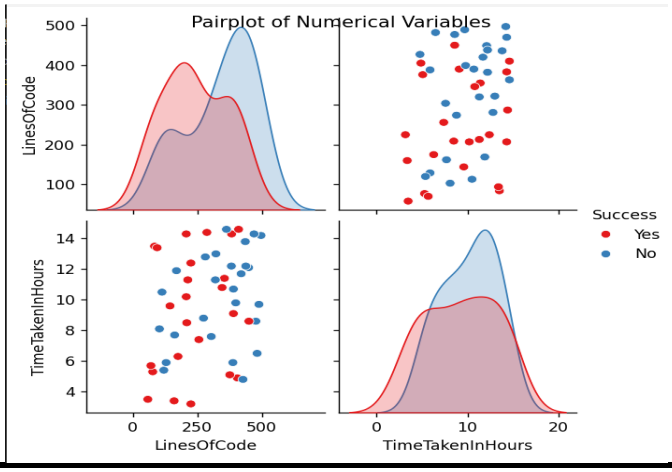
Data Exploration and Preprocessing Report

Dataset variables will be statistically analyzed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature engineering. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong foundation for insights and predictions.

Section	Description
Data Overview	<u>Dimension:</u> 51 rows × 5 columns <u>Descriptive statistics:</u> 
Univariate Analysis	

	originalLang	TargetLang	LinesOfCod	TimeTaken	Success
2	C++	Java	355	11.4	Yes
3	JavaScript	Python	84	13.5	Yes
4	Ruby	Java	274	8.8	No
5	C#	Python	390	9.1	Yes
6	PHP	JavaScript	113	10.5	No
7	C#	Java	207	10.2	Yes
8	C#	Java	449	12.1	No
9	Ruby	JavaScript	399	9.8	No
10	C#	Python	382	12.2	No
11	C++	JavaScript	169	11.9	No

	
Bivariate Analysis	

Multivariate Analysis	
Outliers and Anomalies	-
Data Preprocessing Code Screenshots	

Handling Missing Data	`data['Gender'] = data['Gender'].fillna(data['Gender'].mode()[0])`
	`data['Married'] = data['Married'].fillna(data['Married'].mode()[0])`
	`#replacing + with space for filling the nan values`
	`data['Dependents']=data['Dependents'].str.replace('+','')`
	`<ipython-input-71-6ac39c248773>:2: FutureWarning: The default value of regex will change from`
	`data['Dependents']=data['Dependents'].str.replace('+','')`
	`data['Dependents'] = data['Dependents'].fillna(data['Dependents'].mode()[0])`
	`data['Self_Employed'] = data['Self_Employed'].fillna(data['Self_Employed'].mode()[0])`
	`data['LoanAmount'] = data['LoanAmount'].fillna(data['LoanAmount'].mode()[0])`
	`data['Loan_Amount_Term'] = data['Loan_Amount_Term'].fillna(data['Loan_Amount_Term'].mode()[0])`
	`data['Credit_History'] = data['Credit_History'].fillna(data['Credit_History'].mode()[0])`

Feature Engineering	Attached the codes in final submission.
Save Processed Data	-