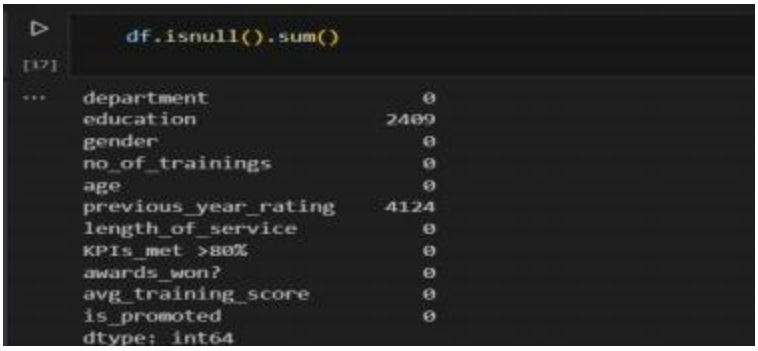
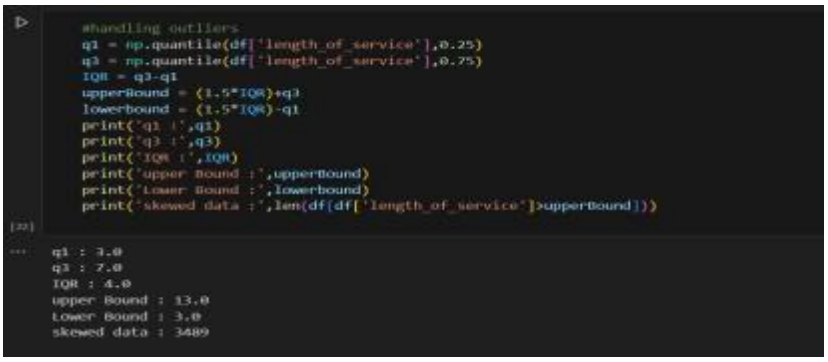
Data Collection and Preprocessing Phase

Date	10 JUNE 2024
Team ID	739689
Project Title	Human resource management:predicting employee promotion using Ml
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	Collect and describe the data available for predicting promotions. This might include employee demographics, performance scores, tenure, education, department, previous promotions, and other relevant features.
Univariate Analysis	Explore each variable individually to understand its distribution and key statistics. - Calculate and plot mean, median, mode, standard deviation, histograms, and box plots for variables like age, performance score, tenure, etc..
Bivariate Analysis	- Investigate relationships between pairs of variables. - Use correlation coefficients and scatter plots to explore the relationship between variables like tenure and promotion status, performance score and promotion status, etc. - Example: A scatter plot of performance scores vs. the number of promotions received.
Multivariate Analysis	Explore patterns and relationships involving multiple variables. - Use techniques like multiple regression analysis, logistic regression, or machine learning models (e.g., decision trees, random forests) to understand how combinations of variables

	predict promotion.
Outliers and Anomalies	Identify and address outliers that could skew the analysis. - Use techniques such as Z-scores, IQR, or robust statistical methods to detect and handle outliers.
Data Preprocessing Code Screenshots	
Loading Data	
Handling Missing Data	
Data Transformation	
Save Processed Data