Data Collection and Preprocessing Phase
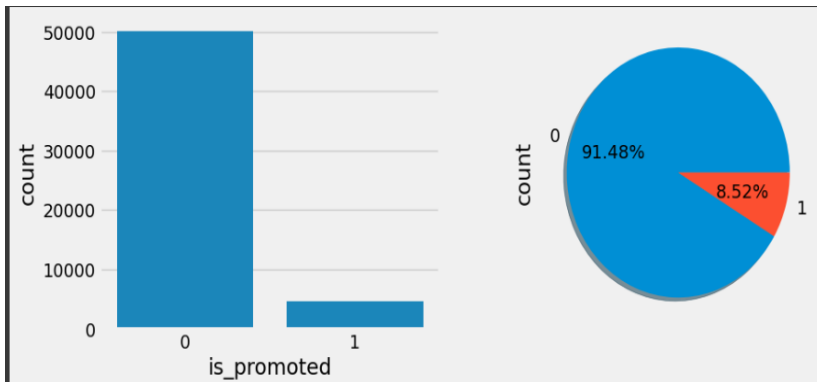
Date	15 March 2024
Team ID	xxxxxx
Project Title	Human Resource Management: Predicting Employee Promotions Using Machine Learning
Maximum Marks	6 Marks

Data Exploration and Preprocessing Template

Optimizing Human Resource Management by employing machine learning techniques to predict employee promotions, enhancing workforce planning and development.

Section	Description
Data Overview	<u>54808 rows x 14 columns</u> <u>Descriptive statistics:</u> -
Univariate Analysis	

Bivariate Analysis	-
Multivariate Analysis	
Outliers and Anomalies	

Data Preprocessing code Screenshots

Loading Data	

| Handling Missing Data | ```# Replacing nan with mode

print(df['education'].value_counts())
df['education']=df['education'].fillna(df['education'].mode()[0])

education
Bachelor's   39078
Master's & above   14925
Below Secondary   805
Name: count, dtype: int64

Replacing nan with mode

print(df['previous_year_rating'].value_counts())
df['previous_year_rating']=df['previous_year_rating'].fillna(df['previous_year_rating'].mode()[0])

previous_year_rating
3.0   18618
5.0   11741
4.0   9877
1.0   6223
2.0   4225
Name: count, dtype: int64``` |
|---|---|
| Data Transformation | ```# Feature mapping is done on education column

from sklearn.preprocessing import LabelEncoder

df['education'] = df['education'].replace(('Below Secondary', 'Bachelor', 'Master', ' & above'), ('0', '1', '2', '3'))

lb = LabelEncoder()

df['department'] = lb.fit_transform(df['department'])``` |
| Feature Engineering | Attached the codes in the final submission |
| Save Processed Data | - |