Data Collection and Preprocessing Phase

Date	15 June 2024
Team ID	739820
Project Title	Predicting the Unpredictable: A Look inti the World of Powerlifting
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	Dimension: 614 rows × 13 columns Descriptive statistics:
Univariate Analysis	-

	playerId	Name	Sex	Equipment	Age	BodyweightKg	BestSquatKg	BestDeadliftKg
0	19391.0	Carlos Ceron	M	Raw	23.0	87.30	205.0	235.0
1	15978.0	Tito Herrera	M	Wraps	23.0	73.48	220.0	260.0
2	27209.0	Levi Lehman	M	Raw	26.0	112.40	142.5	220.0
3	27496.0	Stacy Hayford	F	Raw	35.0	59.42	95.0	102.5
4	20293.0	Brittany Hirt	F	Raw	26.5	61.40	105.0	127.5

Outliers and Anomalies	-

Data Preprocessing Code Screenshots

Loading Data

#importing the data which is in csv file

```
data1=pd.read_csv("/content/X_train.csv",header='infer')
```

```
data2=pd.read_csv("/content/y_train.csv",header='infer')
```

```
data1.head()
```

	playerId	Name	Sex	Equipment	Age	BodyweightKg	BestSquatKg	BestDeadliftKg
0	19391.0	Carlos Ceron	M	Raw	23.0	87.30	205.0	235.0
1	15978.0	Tito Herrera	M	Wraps	23.0	73.48	220.0	260.0
2	27209.0	Levi Lehman	M	Raw	26.0	112.40	142.5	220.0
3	27496.0	Stacy Hayford	F	Raw	35.0	59.42	95.0	102.5
4	20293.0	Brittany Hirt	F	Raw	26.5	61.40	105.0	127.5

Handling Missing Data

```
[ ]  data['Age'].fillna(data['Age'].mean(),inplace=True)
```

Data Transformation	`data1['Name']= LabelEncoder().fit_transform(data1['Name'])`
	
	`data['Sex']= data['Sex'].map({'M':1,'F':0})`
	`from sklearn.preprocessing import LabelEncoder`
	`data['Equipment']= LabelEncoder().fit_transform(data['Equipment'])`
	
	`data['BestSquatKg']=LabelEncoder().fit_transform(data['BestSquatKg'`

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