**Analysis**

**Total Observations in 2018: 17981**

**Total Observations in 2019: 18207**

**Duplicate records in 2018 dataset :52  (Refer DuplicateData\_2018.xlsx)**

**Join the data based on ID column  (Refer Final\_JoinedData\_UsingIDColumn.xlsx)**

**Player match 2018 and 2019 :12539**

**Player available in 2018 and not available in 2019 :5442**

**Join the data based on Name Column  (Refer JoinedData\_UsingNameColumn.xlsx)**

**Player match 2018 and 2019 : 15041**

**Player available in 2018 and not available in 2019 : 5187**

**Removed Column From 2018 Dataset**

**Column 0, Photo, Flag, Club Logo, Value, Wage**

**Predicting “Value” Attribute of the player with respect to 2018 data set of the players**

**Challenges**

**There are multiple players with exact same name who are playing for different clubs and have different age, so when we join data we get duplicate results.**

**Found 52 Duplicate records in 2018 dataset**

**So Joined the 2018 data set using the ID Attribute of the player as Each player has unique ID which distinguishes the player and will help to predict the Player value.**

**Group Members:**

**Dhaval Pandya**

**You fu**

**Drishti Shishodiya**

**Supriya Murty**