**Introduction To Data Science Tools & Techniques**

Sir Farrukh Hassan – Fall 2020

Term Project Report

# **Dataset Description**

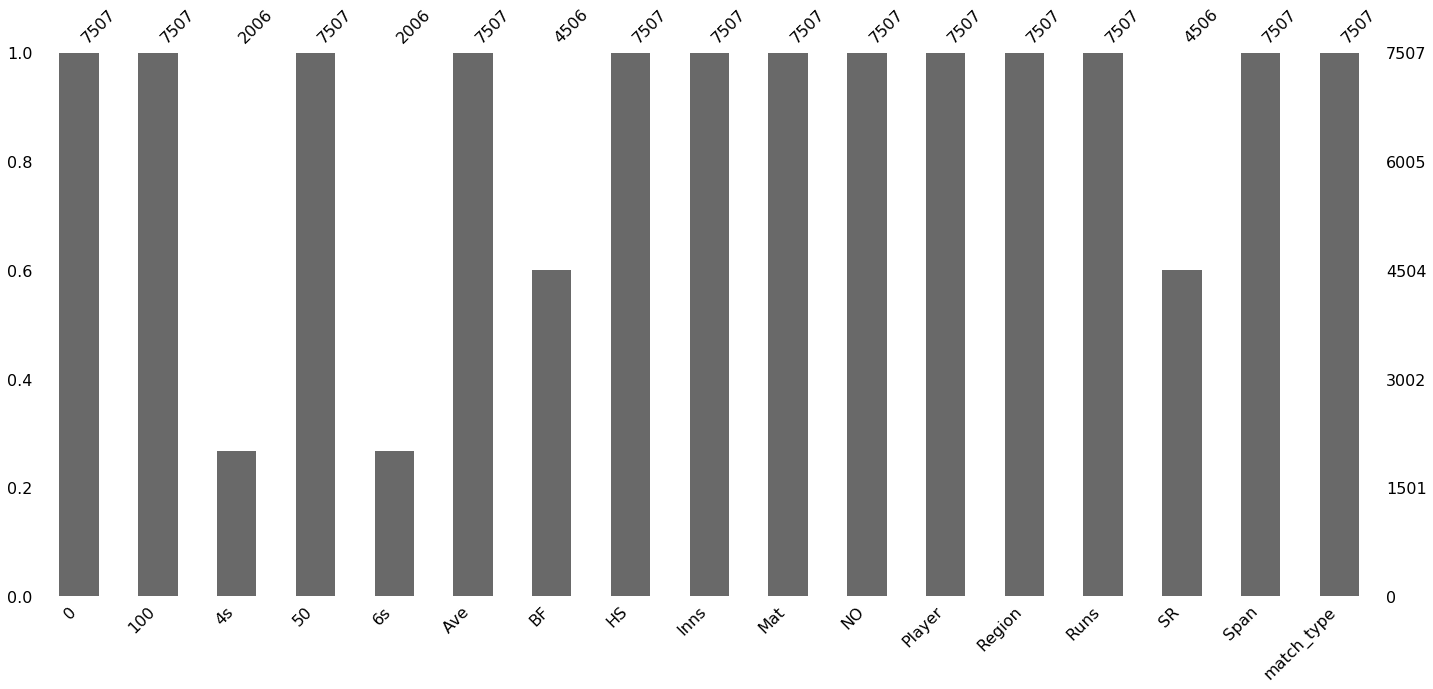
The chosen dataset has 9 CSV files in total containing statistics about various cricketers. The data recorded is from 1877 till 2019. There are 3 different files for different match formats and 3 files for each for different types of players.

## Types of Players

1. Batter
2. Baller
3. Fielder

## Types of Players

1. ODI
2. Test
3. T20

There are common attributes between all files, however, after merging all match formats for into a single file for each player type, missing values were encountered. These missing values were result of a player not participating in a different match format.

As shown visually, in the above diagram for ballers dataset, after merging ODI, T20 and Test matches datasets, missing values were encountered.