Pandas

Pandas is a Python library used for working with data sets.

It has functions for analyzing, cleaning, exploring, and manipulating data.

Pandas series

A Pandas Series is like a column in a table.

It is a one-dimensional array holding data of any type.

2 values describe the pandas= data and indicies

Pandas Data Frame

A Pandas Data Frame is a 2 dimensional data structure, like a 2 dimensional array, or a table with rows and columns.

Read CSV Files

A simple way to store big data sets is to use CSV files (comma separated files).

CSV files contains plain text and is a well know format that can be read by everyone including Pandas.

Pandas functions

Viewing the first few rows of the dataset: *head()* method of pandas Data Frame

Viewing the last few rows of the dataset: tail*()* method of pandas Data Frame

Viewing the dimensionality of the dataset: ***shape*** attribute of the Data Frame class

Getting a concise summary of the dataset: ***info()*** method of pandas Data Frame

Getting descriptive statistics of the data: ***describe()*** method of pandas Data Frame

Components of the dataset: Values — ***values*** attribute of the Data Frame class, Columns — ***columns*** attribute of the Data Frame class, Index — ***index*** attribute of the Data Frame class

isnull Detect missing values for an array-like object.
df.columns:-When you have a big dataset like that it can be hard to see all the columns. using .columns function, you can print out all the columns of the dataset

df.drop():-You can drop some unnecessary columns using df.drop(). In this dataset we have so many columns we are not going to use all of them for this tutorial.

df.iloc():-This function takes as a parameter the rows and column indices and gives you the subset of the DataFrame accordingly.

df.loc():-This function does almost the similar operation as .iloc() function. But here we can specify exactly which row index we want and also the name of the columns we want in our subset.

Dtypes:-Another very basic and widely used functions. Because it is necessary to know the data types of the variables before we dive into the analysis, visualization, or predictive modeling.

insert():-As the name of the function suggests, it inserts a column in the specified position. To demonstrate that I will first create an array of random numbers that have the length of our DataFrame
df.groupby():-This is the most popular function for data summarizing. You can group the data as per a certain variable and find out useful information about those groups.

value_counts():-We can get the value counts of each category using this function. Here I am getting how many values are there in each league_rank.