Assignment 2 - Pandas

## Import the attached Netflix csv file in Jupyter notebook and perform following operations using Pandas:

1. Print the first 5 rows and last 5 rows of the dataframe
2. Check how many rows and columns are there using Pandas function.
3. Print all the column names.
4. Calculate the descriptive statistics of all the variables(integer/float/object etc).
5. Check the number of unique values for each column.
6. Check the percentage of missing values for each column.
7. Delete all the rows where Director column has missing values.
8. Print all the records where country has Germany value (including West Germany). If any other country is there along with Germany, then that row should also come in output.
9. Expand Duration column into 2 separate columns – First column having the numeric value and other having String. Eg: 3 seasons should be split in 2 columns having 3 in 1st column and seasons in 2nd column.
10. Split Date added into 3 separate columns having date value in 1st column, month value in 2nd column and year value in 3rd.
11. Print the number of TV shows/Movies released in each year.
12. Rename the column title with movie\_title.
13. Split Listed\_in column into 3 different columns with col name (Genre1, Genre2, Genre3). Split the column based on comma.



## Import both the attached files (student.csv and marks.csv) in Jupyter notebook and perform following operations:

1. Combine both the dataframes into single dataframe which will have all the records from both the tables.
2. Print the maximum and minimum marks Gender wise.
3. Print all the students IDs and their marks who have scored more than the average marks of the class.
4. Print the dataframe who are Males and are Employed.
5. Create a new Column ‘IQ\_level’ which will have 3 values (Intelligent, Mediocre, weak). If student scored than 80 then Tag him as Intelligent, if student scored between 50-80, then Mediocre, else weak.
6. Count the number of males and females from each city.
7. Print the top 5 Male scorers.
8. Replace the Male value with M and Female value with F and export this dataframe to excel file in D: (D drive) and name the file as test.csv.
9. Check if any student\_ID is duplicated.
10. Create a separate dataframe which will have all the Integer/Float variables.
11. Get those Student\_IDs which are present in Students table but not in Marks table.

 

1. Explain the concept of missing values. How can you identify the missing values in a Pandas DataFrame ?

What are the different ways of treating/Imputing/Deleting the missing values.

Explain with example.