Institute of Computer Technology
B. Tech Computer Science and Engineering
Sub: Data Mining and Warehousing (2CSE60E27)

PRACTICAL 2: DATA EXPLORATION

The National Basketball Association (NBA) is a professional basketball league in North America. The league is composed of 30 teams (29 in the United States and 1 in Canada) and is one of the four major professional sports leagues in the United States and Canada. It is the premier men's professional basketball league in the world. Dataset is provided to you contains the details of the NBA (National Basketball Association) like:

- Name of the player
- Team name
- Age of the player
- In how many games a player have played
- How many 3-Point Field Goal Attempts/2-Point Field Goal Attempts done by a player
- What are the 2-Point Field Goal Percentage/3-Point Field Goal Percentage
- How many free throws/ free throws attempts were made by the players
- How many Offensive Rebounds and Defensive Rebounds were done by the players
- How many Assists, Steals, Blocks, Turnovers, Personal Fouls and Points made by a player
- When the player have started the game (time) and how many minutes played the game.

Using the concept of Data exploration, try to read the given dataset using python and find out the below mentioned information from the given dataset.

1) Find out the average age of the player who have played for NBA?

```
1. Average age of the players whi have played for NBA?

26.04823151125402
```

2) Find out that each player have played how many games?

```
-----
2. How many games has each player played?

        Player  Games
0    Jayson Tatum    79
1    Kyrie Irving    67
2    Marcus Smart    80
3   Marcus Morris    75
4      Al Horford    68
5    Jaylen Brown    74
6  Gordon Hayward    72
7    Terry Rozier    79
8    Daniel Theis    66
9     Aron Baynes    51
-----
```

3) Total how many teams are there in NBA?

```
-----
3. How many teams in total are there in NBA?

30
-----
```

4) What is the minimum age of the player that have played for NBA?

```
-----
4. What is the minimum age of the player that has played for NBA?

19
-----
```

5) What is the Maximum age of the player that have played for NBA and display its details?

```
-----
5. What is the maximum age of the player that has played for NBA? His details?

    Conference   Division           Team  ... Turnovers  Personal Fouls  Points
302    Eastern  Southeast  Atlanta Hawks  ...        48             141     562

[1 rows x 30 columns]
-----
```

6) Find out how many games have been organized in Eastern region.

```
-----
6. How many games have been organized in Eastern region?

316
-----
```

7) Find out the how many regions (conference) are there where the games have been organized?

```
-----
7. How many regions (conference) are there where the games have been organized?

Eastern
Western
-----
```

8) Display the list of the player who have played in the team called "Boston Celtics"

```
-----
8. Display the list of the player who have played in the team called Boston Celtics.

Jayson Tatum
Kyrie Irving
Marcus Smart
Marcus Morris
Al Horford
Jaylen Brown
Gordon Hayward
Terry Rozier
Daniel Theis
Aron Baynes
Semi Ojeleye
Brad Wanamaker
Robert Williams
Guerschon Yabusele
PJ Dozier
R.J. Hunter
Greg Monroe
-----
```

9) Find out the total number of games organized in each division?

```
-----
9. What is the total number of games organized in each division?

Atlantic   107
Central    110
Southeast   99
Northwest   93
Pacific    105
Southwest  108
-----
```

10) Which player has scored the maximum number of goals?

```
-----
10. Which player has scored the maximum number of goals?

James Harden
-----
```

11) Which player has the lowest number of personal fouls?

```
-----
11. Which player has the lowest number of personal fouls?

            Greg Monroe
            Trevon Duval
            Ike Anigbogu
Sviatoslav Mykhailiuk
            Zach Lofton
             Tyler Ulis
           Kobi Simmons
            John Holland
           Troy Caupain
            J.P. Macura
            Yante Maten
          Emanuel Terry
           John Jenkins
            Okaro White
          Jordan Sibert
         Donte Grantham
            Tyler Davis
            Jawun Evans
            Mitch Creek
           Andre Ingram
            George King
          Vince Edwards
                Zhou Qi
          D.J. Stephens
   Kostas Antetokounmpo
           Ray Spalding
-----
```

12) Which player have attempted the 3-ponts field goals the highest number of times and what are the percentage of it?

```
-----
12. Which players have attempted the 3-ponts field goals the highest number of times and what are the percentage of it?

        Player   3-Point Field Goal Percentage
James Harden                      0.368
-----
```

13) What is the average point scored by all the players?

```
-----
13. What is the average point scored by all the players?

439.82797427652736
>>>
```

Code:

```
import pandas as pd
df = pd.read_csv(r"C:\\Users\\admin\\Desktop\\dishwa\\dmw\\Practical
2\\NBA_2018-19_Season.csv", delimiter = ',', on_bad_lines='skip') #error_bad_lines = False

print("1. Average age of the players whi have played for NBA?\n")
avg = df['Age'].mean()
print(avg)

print("-----\n2. How many games has each player played?\n")
total_games_played = df[['Player', 'Games']]
print(total_games_played.head(10))

print("-----\n3. How many teams in total are there in NBA?\n")
print(len(df['Team'].unique()))

print("-----\n4. What is the minimum age of the player that has played for NBA?\n")
print(df['Age'].min())

print("-----\n5. What is the maximum age of the player that has played for NBA? His
details?\n")
print(df[df.Age == df.Age.max()])

print("-----\n6. How many games have been organized in Eastern region?\n")
df2 = df[df['Conference']=='Eastern']
count = len(df2)
print(count)

print("-----\n7. How many regions (conference) are there where the games have been
organized?\n")
df2 = df['Conference'].unique()
for x in df2:
    print(x)

print("-----\n8. Display the list of the player who have played in the team called Boston
Celtics.\n")
df2 = df['Team']=='Boston Celtics'
player_list = df['Player'][df2].tolist()
for x in player_list:
    print(x)

print("-----\n9. What is the total number of games organized in each division?\n")
division = df['Division'].unique()
for x in division:
    print(x," ",end="")
    print(df['Division'][df['Division']==x].count())
```

print("-----\n10. Which player has scored the maximum number of goals?\n")
max = df['Field Goals'].max()
print(df.loc[df['Field Goals'] ==max]['Player'].to_string(index=False))

print("-----\n11. Which player has the lowest number of personal fouls?\n")
min = df['Personal Fouls'].min()
print(df.loc[df['Personal Fouls'] == min]['Player'].to_string(index=False))

print("-----\n12. Which players have attempted the 3-ponts field goals the highest number of times and what are the percentage of it?\n")
max = df['3-Point Field Goal Attempts'].max()
print(df.loc[df['3-Point Field Goal Attempts'] ==max][['Player','3-Point Field Goal Percentage']].to_string(index=False))

print("-----\n13. What is the average point scored by all the players?\n")
avg1 = df['Points'].mean()
print(avg1)

Output:

dmw 2.py - C:\Users\admin\Desktop\dishwa\dmw\Practical 2\dmw 2.py (3.9.1)

File Edit Format Run Options Window Help

```python
import pandas as pd
df = pd.read_csv(r"C:\\Users\\admin\\Desktop\\dishwa\\dmw\\Practical 2\\NBA_

print("1. Average age of the players whi have played for NBA?\n")
avg = df['Age'].mean()
print(avg)

print("-----\n2. How many games has each player played?\n")
total_games_played = df[['Player', 'Games']]
print(total_games_played.head(10))

print("-----\n3. How many teams in total are there in NBA?\n")
print(len(df['Team'].unique()))

print("-----\n4. What is the minimum age of the player that has played for NB
print(df['Age'].min())

print("-----\n5. What is the maximum age of the player that has played for NB
print(df[df.Age == df.Age.max()])

print("-----\n6. How many games have been organized in Eastern region?\n")
df2 = df[df['Conference']=='Eastern']
count = len(df2)
print(count)

print("-----\n7. How many regions (conference) are there where the games have
df2 = df['Conference'].unique()
for x in df2:
    print(x)

print("-----\n8. Display the list of the player who have played in the team c
df2 = df[df['Team']=='Boston Celtics']
player_list = df['Player'][df2].tolist()
for x in player_list:
    print(x)

print("-----\n9. What is the total number of games organized in each division
division = df['Division'].unique()
for x in division:
    print(x," ",end="")
    print(df['Division'][df['Division']==x].count())
```

IDLE Shell 3.9.1

File Edit Shell Debug Options Window Help

```
-----
7. How many regions (conference) are there where the games have been organized?

Eastern
Western
-----
8. Display the list of the player who have played in the team called Boston Celtics.

Jayson Tatum
Kyrie Irving
Marcus Smart
Marcus Morris
Al Horford
Jaylen Brown
Gordon Hayward
Terry Rozier
Daniel Theis
Aron Baynes
Semi Ojeleye
Brad Wanamaker
Robert Williams
Guerschon Yabusele
PJ Dozier
R.J. Hunter
Greg Monroe
-----
9. What is the total number of games organized in each division?

Atlantic  107
Central  110
Southeast  99
Northwest  93
Pacific  105
Southwest  108
-----
10. Which player has scored the maximum number of goals?

James Harden
-----
11. Which player has the lowest number of personal fouls?
```
Activate Windows
Go to Settings to activate Windows.

dmw 2.py - C:\Users\admin\Desktop\dishwa\dmw\Practical 2\dmw 2.py (3.9.1)

File Edit Format Run Options Window Help

```python
import pandas as pd
df = pd.read_csv(r"C:\\Users\\admin\\Desktop\\dishwa\\dmw\\Practical 2\\

print("1. Average age of the players whi have played for NBA?\n")
avg = df['Age'].mean()
print(avg)

print("-----\n2. How many games has each player played?\n")
total_games_played = df[['Player', 'Games']]
print(total_games_played.head(10))

print("-----\n3. How many teams in total are there in NBA?\n")
print(len(df['Team'].unique()))

print("-----\n4. What is the minimum age of the player that has played f
print(df['Age'].min())

print("-----\n5. What is the maximum age of the player that has played f
print(df[df.Age == df.Age.max()])

print("-----\n6. How many games have been organized in Eastern region?\n
df2 = df[df['Conference']=='Eastern']
count = len(df2)
print(count)

print("-----\n7. How many regions (conference) are there where the games
df2 = df['Conference'].unique()
for x in df2:
    print(x)

print("-----\n8. Display the list of the player who have played in the t
df2 = df[df['Team']=='Boston Celtics']
player_list = df['Player'][df2].tolist()
for x in player_list:
    print(x)

print("-----\n9. What is the total number of games organized in each div
division = df['Division'].unique()
for x in division:
    print(x," ",end="")
    print(df['Division'][df['Division']==x].count())
```

IDLE Shell 3.9.1

File Edit Shell Debug Options Window Help

```
James Harden
-----
11. Which player has the lowest number of personal fouls?

          Greg Monroe
          Trevon Duval
          Ike Anigbogu
Sviatoslav Mykhailiuk
          Zach Lofton
          Tyler Ulis
          Kobi Simmons
          John Holland
          Troy Caupain
          J.P. Macura
          Yante Maten
          Emanuel Terry
          John Jenkins
          Okaro White
          Jordan Sibert
          Donte Grantham
          Tyler Davis
          Jawun Evans
          Mitch Creek
          Andre Ingram
          George King
          Vince Edwards
          Zhou Qi
          D.J. Stephens
Kostas Antetokounmpo
          Ray Spalding
-----
12. Which players have attempted the 3-ponts field goals the highest number of times and what are
    Player  3-Point Field Goal Percentage
James Harden                    0.368
-----
13. What is the average point scored by all the players?

439.82797427652736
>>>
```
Activate Windows
Go to Settings to activate Windows.