Declaration on Plagiarism

This form must be filled in and completed by the student(s) submitting an assignment

Name:	Akshai Ramesh
Student Number:	19210382
Programme:	Data Analytics
Module Code:	CA682
Assignment Title:	Data Visualisation
Submission Date:	13 Dec 2019
Module Coordinator:	Dr Suzanne Little

I declare that this material, which I now submit for assessment, is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work. I understand that plagiarism, collusion, and copying are grave and serious offences in the university and accept the penalties that would be imposed should I engage in plagiarism, collusion or copying. I have read and understood the Assignment Regulations. I have identified and included the source of all facts, ideas, opinions, and viewpoints of others in the assignment references. Direct quotations from books, journal articles, internet sources, module text, or any other source whatsoever are acknowledged and the source cited are identified in the assignment references. This assignment, or any part of it, has not been previously submitted by me or any other person for assessment on this or any other course of study.

I have read and understood the referencing guidelines found at http://www.dcu.ie/info/regulations/plagiarism.shtml, https://www4.dcu.ie/students/az/plagiarism and/or recommended in the assignment guidelines

Name: Akshai Ramesh Date: 16.12.2019

Project Title or Headline

Abstract
This story relates to my favorite sport and one of the most common ball games, Basketball. Since the 3-point line was introduced in 1979 season, there has been a noticable increase in number of 3 points shots being played. 3 points are given to the player if he/she could shoot from behind the three-point line. My story draws insights on the evolution of 3 pointer shot, consequential changes in game strategy and its very high efficiency.

What is the conclusion that you reached?

- A clear insight of how players transformed their game style after the '3 pointer rule'.
- Based on the results obtained, the 3 point shot is highly efficient.
- The accuracy of 3 point shoot depends on how far is the distance from the 3 point line.

Dataset Description

Data Retrieval : The data was obtained through website scraping from Basketball Reference (URL : **https://www.basketball-reference.com/**). For each player the seasons total stats from 1979 has been scrapped.
The code used for scraping is present in ~/source/scrap_data_soup.py.
Datasize : The dataset is 106.5 MB in size. The data consists of :
seasons_total.csv and -
1000_plus_shot_charts_2011_2016.csv

seasons_total.csv :
The dataset consists of 18295 rows × 31 columns. There are large number of columns, but for our case study, the table contains the following :
yr – season year
g – number of goals scored by a player
fg3 – number of 3 pointer goals scored by a player
fg3a – number of 3 pointer goals attempted by a player
pts – number of points scored by a player in a season

1000_plus_shot_charts_2011_2016.csv :
The dataset consists of 243719 rows x 7 columns. Most of the columns are required for the analysis.
yr – season year
game date – date on which the game was played
shot_distance – the distance from which the shoot was taken.
X – The x coordinate of shooting position.
Y – The y coordinate of shooting position.
shot_made_flag – Indicates that the shot taken wont be counted.

The Idea is to use both the datasets together and analyze the possible combinations to find something interesting.

Data Acquisition and Processing:

Dataset Preparation : The data was obtained by scraping. The code is given below :

Welcome Guide | scrap_data_soup.py

```
from bs4 import BeautifulSoup
import requests
def parse_html_table(html_table):
    '''For parsing basketball reference stats html table.
    We will apply this function to every seasons html page.'''
    data = []
    cur_row = []
    row_names = []
    for ele in html_table:
        stat_name  = ele['data-stat']
        stat_value = ele.string
        new_row = (stat_name == 'player')
        if new_row:
            if cur_row:
                data.append(cur_row)
            cur_row = []
            col_names = []
            cur_row.append(ele['csk']) # fix asterisk error
            col_names.append(stat_name)
            continue
        cur_row.append(stat_value)
        col_names.append(stat_name)
    return data, col_names


```

The code was run on every html page containing seasons data. The data was then loaded into the pandas dataframe and cleaned before EDA.

Data Cleaning :

- *Missing Data :*

There were some missing values in ‘fga’ and ‘year’ field, such rows were dropped for consistency.

- *Handling Irrelevant Data :*

Some of the fields like ‘player’, ‘age’ , ‘team_id’ , have been dropped since they are not involved in the analysis process.

3. ***Visualisation:*** In visualisation process, three primary cases were analysed and visualised, in the order as given below :

- *What is the average number of 3 pointer attempts ?*

The 2 tables were joined together using the year criteria and the graph for average number of 3 point attempts was constructed. The left side graph represents the steady rise in the number of 3 point attempts during the 1995 , where NBA brought in the 3 point line. This data was used for training purpose and a logistic regression model was trained to predict the avg. 3 point attempts in future.

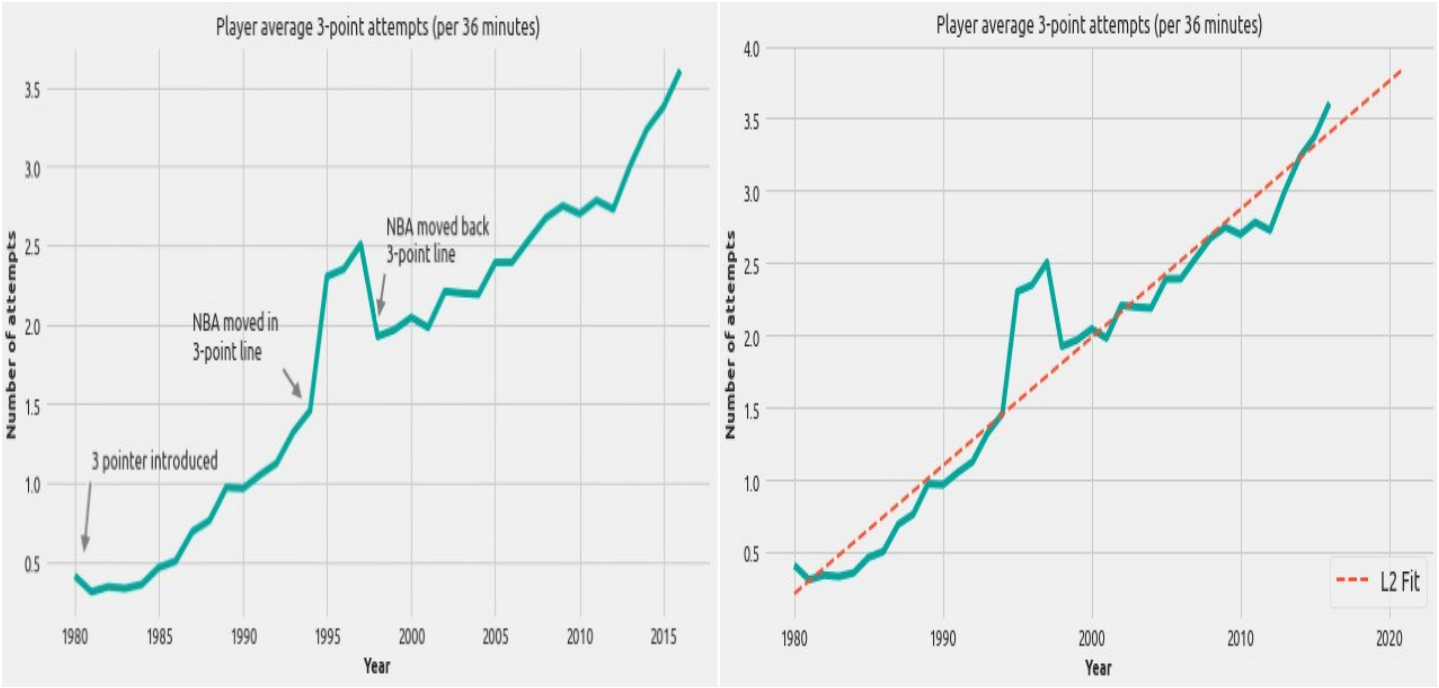

Fig 1: *Player Average 3 – point attempts (per 36 minutes)*

The stable trend graph had year on the x-axis and avg. number of attempts on the y-axis. In order to show the predicted plot, along with its reference in a neat manner, the 2 plots were put together horizontally next to each other.

- *Does the shot accuracy and shot distance parameters follow a same trend ?*

Shot accuracy was computed by identifying the type of score obtained. The accuracy metric score was then plotted with the year in range of 2012 – 2016. The Shot distance from 3 point line while shooting 3 pointer was plotted against the year parameter. The linear stable increase and sudden drop was plotted by using 2 scatter charts horizontally.

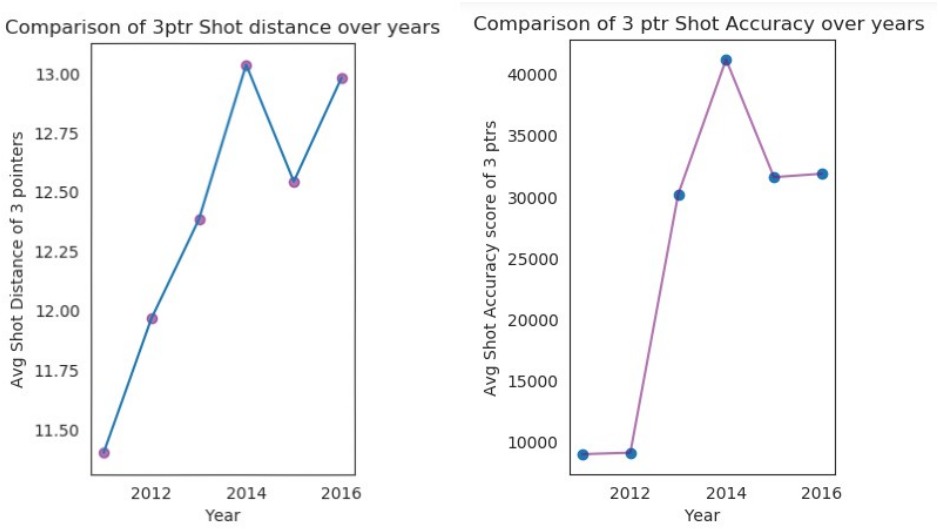

Fig 2 : *Similar trends in shot distance and shot accuracy over years*

- *Analyzing the Geospatial Shot Chart Data*

The Graph was put together as a combination of heat map and pie chart. The heatmap was used to show the distribution of shot positions and pie chart was used to represent the corresponding year percentage of shots inside. The change was significant over a period of 6 years and hence was finally plotted into a gif image.

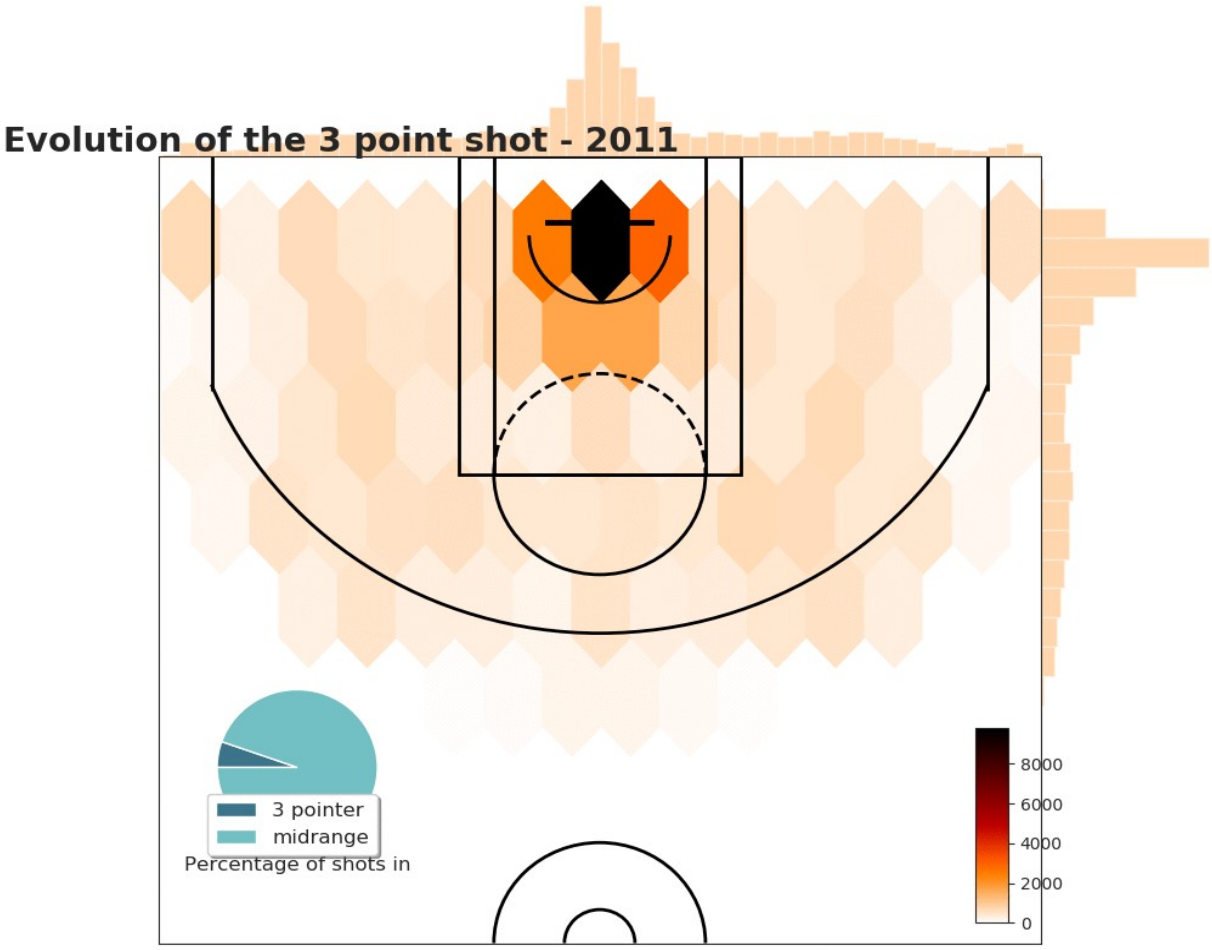

Fig 3 : *Evolution of the 3 point shot (2012-2016)*

Conclusion

- The player average number of 3 point attempts graph shows the steady increase in number of attempts since the 1995, when the 3 pointer line was introduced.
- It is interesting to note the gradual increase in number of attempts of 3 point shot after NBA moved in 3 point line, followed by a gradual decrease after the time that NBA moved back. It shows the huge impact NBA has on the style of basketball being played.
- In naive terms, more closer to the ring , greater should be the accuracy. But it was surprising to see that although players started to shoot from long distances, there accuracy was gradually increasing with time.
- The evolution of 3 point shot clearly shows a gradual increase in distribution across the 3 point line, it can be noticed that in initial stages, the 3 point shots were made

dominantly from infront of the ring, but over a period of time there are significant heat map coordinates around the corner.

- The increasing number of 3 point attempts over years shows that the players have adapted to using it in their game strategy.

References

https://eazybi.com/blog/data_visualization_and_chart_types/ [Choosing the right chart]
https://www.basketball-reference.com/ [Dataset scraped from this website]
http://savvastjortjoglou.com/nba-shot-sharts.html [Creating shot chart tutorial]
https://matplotlib.org/3.1.1/tutorials/introductory/pyplot.html [matplotlib reference]
https://www.dataquest.io/blog/web-scraping-beautifulsoup/ [Beautiful Soup Tutorial]