



Московский Государственный Технический Университет имени

Н.Э.Баумана

Факультет Информатика и системы управления

Кафедра ИУ-5

«Системы обработки информации и управления»

ОТЧЁТ

Лабораторная работа №1

Медоты машинного обучения

Выполнил: Ма Линь

студент группы: ИУ-5 21М

Москва 2022г.

Цель работы:

1, изучение различных методов визуализация данных и создание истории на основе данных.

Задание:

- Выбрать набор данных (датасет).
- Создать "историю о данных" в виде юпитер-ноутбука, с учетом следующих требований:
 1. История должна содержать не менее 5 шагов (где 5 - рекомендуемое количество шагов). Каждый шаг содержит график и его текстовую интерпретацию.
 2. На каждом шаге наряду с удачным итоговым графиком рекомендуется в юпитер-ноутбуке оставлять результаты предварительных "неудачных" графиков.
 3. Не рекомендуется повторять виды графиков, желательно создать 5 графиков различных видов.
 4. Выбор графиков должен быть обоснован использованием методологии data-to-viz. Рекомендуется учитывать типичные ошибки построения выбранного вида графика по методологии data-to-viz. Если методология Вами отвергается, то просьба обосновать Ваше решение по выбору графика.
 5. История должна содержать итоговые выводы. В реальных "историях о данных" именно эти выводы представляют собой основную ценность для предприятия.
- Сформировать отчет и разместить его в своем репозитории на github.

1,Обзор

Международная федерация футбола (фр. Fédération Internationale de Football Association, сокращённо FIFA, в русской транслитерации — ФИФА́) — главная футбольная организация, являющаяся крупнейшим международным руководящим органом в футболе, футзале и пляжном футболе. Штаб-квартира ФИФА находится в швейцарском городе Цюрихе.

Серия игр ФИФА — видеоигра - симулятор футбола , выпущенная Electronic Arts .

В этой статье будут визуализированы данные об индивидуальных способностях игроков и способностях каждой национальной сборной по футболу.

2, Конкретный процесс и реализация кода

(1) Загрузить доступные пакеты

```
In [1]: # This Python 3 environment comes with many helpful analytics libraries installed
# It is defined by the kaggle/python Docker image: https://github.com/kaggle/docker-python
# For example, here's several helpful packages to load

import numpy as np # linear algebra
import pandas as pd # data processing

# Input data files are available in the read-only "../input/" directory

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
```

(2) библиотека импорта

```
In [2]: #Importing Libraries
import numpy as np
import pandas as pd
import os
import warnings
warnings.filterwarnings('ignore')
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline

flatui = ["#9b59b6", "#3498db", "#95a5a6", "#e74c3c", "#34495e", "#2ecc71"] # defining the colour palette
flatui = sns.color_palette(flatui)
```

(3) читать данные

```
In [3]: import pandas as pd
import numpy as np
df=pd.read_csv("../data.csv")
```

(4) импорт модуля wordcloud. Wordcloud использует размеры текста для представления частоты текста.

```
In [4]: from wordcloud import WordCloud # importing the wordcloud module. Wordcloud uses the text sizes to represent the frequency of the text
```

(5) взглянув на набор данных, первые 10 строк

```
In [5]: df.head(10) # having a look at the dataset, first 10 rows
```

Out[5]:

	Unnamed: 0	ID	Name	Age	Photo	Nationality	Flag	Overall	Potential	Club	...	Cor
0	0	158023	L. Messi	31	https://cdn.sofifa.org/players/4/19/158023.png	Argentina	https://cdn.sofifa.org/flags/52.png	94	94	FC Barcelona	...	
1	1	20801	Cristiano Ronaldo	33	https://cdn.sofifa.org/players/4/19/20801.png	Portugal	https://cdn.sofifa.org/flags/38.png	94	94	Juventus	...	
2	2	190871	Neymar Jr	26	https://cdn.sofifa.org/players/4/19/190871.png	Brazil	https://cdn.sofifa.org/flags/54.png	92	93	Paris Saint-Germain	...	
3	3	193080	De Gea	27	https://cdn.sofifa.org/players/4/19/193080.png	Spain	https://cdn.sofifa.org/flags/45.png	91	93	Manchester United	...	
4	4	192985	K. De Bruyne	27	https://cdn.sofifa.org/players/4/19/192985.png	Belgium	https://cdn.sofifa.org/flags/7.png	91	92	Manchester City	...	
5	5	183277	E. Hazard	27	https://cdn.sofifa.org/players/4/19/183277.png	Belgium	https://cdn.sofifa.org/flags/7.png	91	91	Chelsea	...	
6	6	177003	L. Modrić	32	https://cdn.sofifa.org/players/4/19/177003.png	Croatia	https://cdn.sofifa.org/flags/10.png	91	91	Real Madrid	...	
7	7	176580	L. Suárez	31	https://cdn.sofifa.org/players/4/19/176580.png	Uruguay	https://cdn.sofifa.org/flags/60.png	91	91	FC Barcelona	...	
8	8	155862	Sergio Ramos	32	https://cdn.sofifa.org/players/4/19/155862.png	Spain	https://cdn.sofifa.org/flags/45.png	91	91	Real Madrid	...	
9	9	200389	J. Oblak	25	https://cdn.sofifa.org/players/4/19/200389.png	Slovenia	https://cdn.sofifa.org/flags/44.png	90	93	Atlético Madrid	...	

10 rows × 89 columns

(6) Проверьте общее количество строк и столбцов в данных.

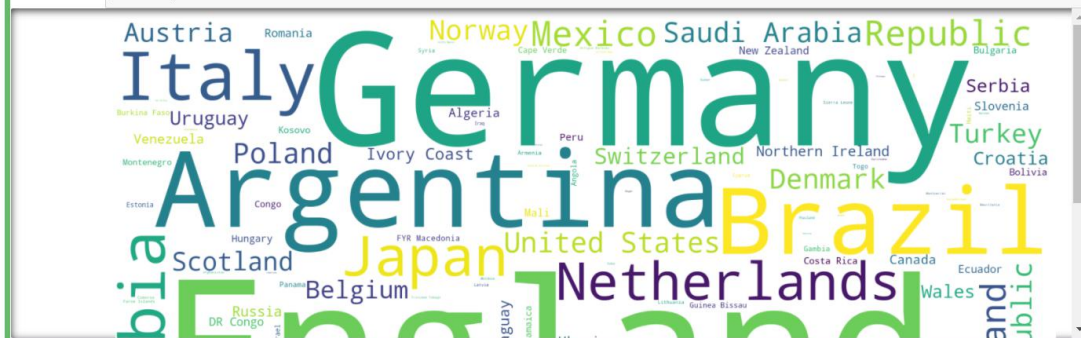
```
In [6]: df.shape # checking the number of rows and columns in the dataset
Out[6]: (18207, 89)
```

(7) Используйте облако слов, чтобы показать количество спортсменов в разных странах.

```
In [7]: #Nationality Text Size = Nationality Player Count
```

```
In [8]: # Plotting the wordcloud for the Nationalit column
plt.subplots(figsize=(25,15))
wordcloud = WordCloud(
    background_color='white',
    width=1920,
    height=1080
).generate(" ".join(df.Nationality))

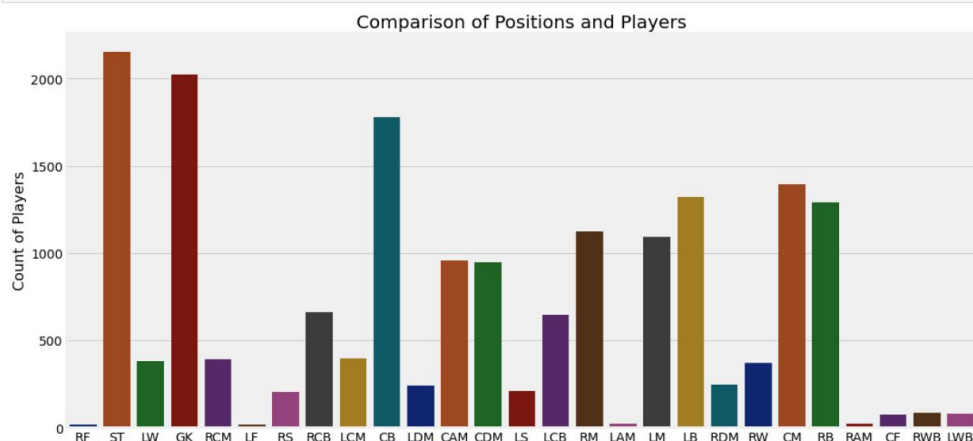
plt.imshow(wordcloud)
plt.axis('off')
plt.savefig('graph.png')
plt.show()
```



(8) Распределение позиций игроков на поле.

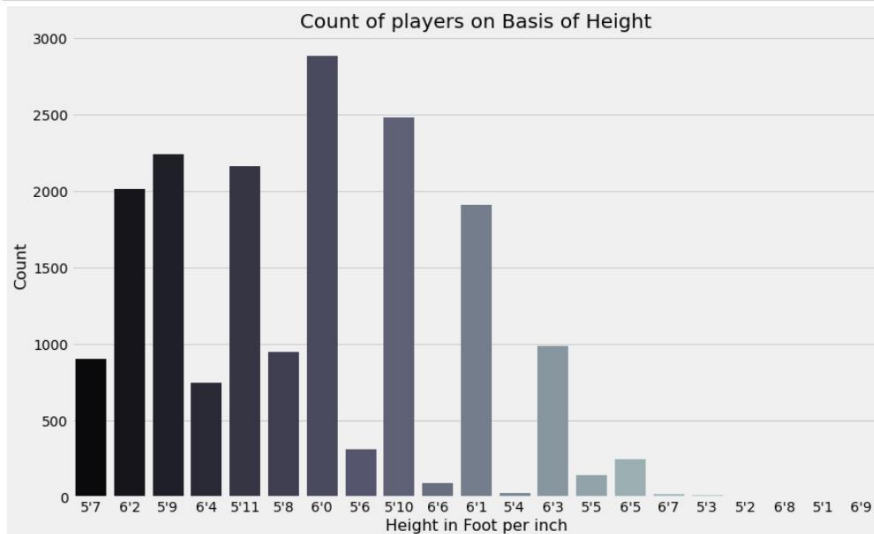
```
In [9]: # different positions acquired by the players

plt.figure(figsize = (18, 8))
plt.style.use('fivethirtyeight')
ax = sns.countplot('Position', data = df, palette = 'dark')
ax.set_xlabel(xlabel = 'Different Positions in Football', fontsize = 16)
ax.set_ylabel(ylabel = 'Count of Players', fontsize = 16)
ax.set_title(label = 'Comparison of Positions and Players', fontsize = 20)
plt.show()
```



(9) Распределение роста игроков.

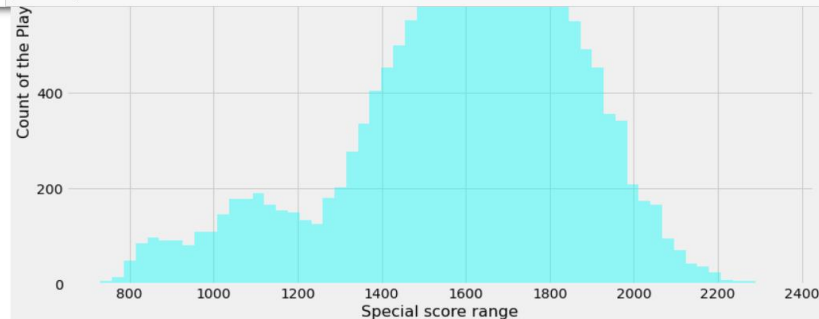
```
In [10]: # plotting coun of players based on their heights
plt.figure(figsize = (13, 8))
ax = sns.countplot(x = 'Height', data = df, palette = 'bone')
ax.set_title(label = 'Count of players on Basis of Height', fontsize = 20)
ax.set_xlabel(xlabel = 'Height in Foot per inch', fontsize = 16)
ax.set_ylabel(ylabel = 'Count', fontsize = 16)
plt.show()
```



(10) Распределение количества игроков, забивших.

```
In [11]: x = df.Special
plt.figure(figsize = (12, 8))
plt.style.use('tableau-colorblind10')

ax = sns.distplot(x, bins = 58, kde = False, color = 'cyan')
ax.set_xlabel(xlabel = 'Special score range', fontsize = 16)
ax.set_ylabel(ylabel = 'Count of the Players', fontsize = 16)
ax.set_title(label = 'Histogram for the Speciality Scores of the Players', fontsize = 20)
plt.show()
```



(11) Информация о пяти лучших игроках с наибольшим количеством голов.

```
In [12]: # Every Nations' Player and their overall scores

some_countries = ('England', 'Germany', 'Spain', 'Argentina', 'France', 'Brazil', 'Italy', 'Columbia') # defining a tuple consisting of co
data_countries = df.loc[df['Nationality'].isin(some_countries) & df['Overall']] # extracting the overall data of the countries selected in
data_countries.head()
```

Out[12]:

Unnamed: 0	ID	Name	Age	Photo	Nationality	Flag	Overall	Potential	Club	...	Con
3	3	De Gea	27	https://cdn.sofifa.org/players/4/19/193080.png	Spain	https://cdn.sofifa.org/flags/45.png	91	93	Manchester United	...	
8	8	Sergio Ramos	32	https://cdn.sofifa.org/players/4/19/155862.png	Spain	https://cdn.sofifa.org/flags/45.png	91	91	Real Madrid	...	
14	14	N. Kanté	27	https://cdn.sofifa.org/players/4/19/215914.png	France	https://cdn.sofifa.org/flags/18.png	89	90	Chelsea	...	
15	15	P. Dybala	24	https://cdn.sofifa.org/players/4/19/211110.png	Argentina	https://cdn.sofifa.org/flags/52.png	89	94	Juventus	...	
16	16	H. Kane	24	https://cdn.sofifa.org/players/4/19/202126.png	England	https://cdn.sofifa.org/flags/14.png	89	91	Tottenham Hotspur	...	

5 rows x 89 columns

(12) Количество голов, забитых национальной сборной по футболу.



(13) Отображение количества вхождений имен среди 20 лучших игроков в виде облака слов.

