

## חלק א' - חקירת מערך נתונים

```
In [1]: #Import libraries:
import pandas as pd # pa
import numpy as np # numpy is a package for working with numerical data
```

```
In [2]: #Reading the 'HW1_data.csv' file:
df_soccer = pd.read_csv('HW1_data.csv')
```

### Question 1:


```
In [3]: #1

df_soccer.info()
```

<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 684 entries, 0 to 683  
Data columns (total 11 columns):

#	Column	Non-Null Count	Dtype
0	league	684 non-null	object
1	year	684 non-null	int64
2	position	684 non-null	int64
3	team	684 non-null	object
4	matches	684 non-null	int64
5	wins	684 non-null	int64
6	draws	684 non-null	int64
7	loses	684 non-null	int64
8	scored	684 non-null	int64
9	conceded	681 non-null	float64
10	pts	684 non-null	int64

dtypes: float64(1), int64(8), object(2)  
memory usage: 58.9+ KB



We can see according to the "info()" of the data frame that there are 684 entries.

### Question 2:

```
In [4]: #2

df_soccer.info()
display(df_soccer)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 684 entries, 0 to 683
Data columns (total 11 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   league      684 non-null    object
1   year        684 non-null    int64
2   position    684 non-null    int64
3   team        684 non-null    object
4   matches     684 non-null    int64
5   wins        684 non-null    int64
6   draws       684 non-null    int64
7   loses       684 non-null    int64
8   scored      684 non-null    int64
9   conceded    681 non-null    float64
10  pts         684 non-null    int64
dtypes: float64(1), int64(8), object(2)
memory usage: 58.9+ KB
```



	league	year	position	team	matches	wins	draws	loses	scored	conceded	pts
0	La_liga	2014	1	Barcelona	38	30	4	4	110	21.0	94
1	La_liga	2014	2	Real Madrid	38	30	2	6	118	38.0	92
2	La_liga	2014	3	Atletico Madrid	38	23	9	6	67	29.0	78
3	La_liga	2014	4	Valencia	38	22	11	5	70	32.0	77
4	La_liga	2014	5	Sevilla	38	23	7	8	71	45.0	76
...	...	...	...	...	...	...	...	...	...	...	...
679	RFPL	2019	12	PFC Sochi	30	8	9	13	40	39.0	33
680	RFPL	2019	13	FK Akhmat	30	7	10	13	27	46.0	31
681	RFPL	2019	14	Krylya Sovetov Samara	30	8	7	15	33	40.0	31
682	RFPL	2019	15	FC Tambov	30	9	4	17	37	41.0	31
683	RFPL	2019	16	FC Orenburg	30	7	6	17	28	52.0	27

684 rows × 11 columns

We can see according to the "info()" of the data frame that there are two fields of type "object", so we can be sure that they are categorical. About the other columns, we can see according to the data frame table that the other fields are continuous, so they are numerical. To sum up:

The fields that are categorical are:



1. league
2. team

The fields that are numerical are:

1. year

2. position
3. matches
4. wins
5. draws
6. loses
7. scored
8. conceded
9. pts



## Question 3:

In [5]: #3

```
print(len(df_soccer.league.unique()))
print()
print(len(df_soccer.team.unique()))
```

6

168

Number of unique values at "league" field: 6

Number of unique values at "team" field: 168



## Question 4:

In [6]: #4

```
print("Let's find out if there are missing values at our data frame: " + str(pd.isna(df_soccer).sum().max() > 0))
for i in df_soccer.columns:
    if df_soccer.isna().sum()[i] != 0:
        print('The field', i, 'have', df_soccer.isna().sum()[i], 'missing values.')
```

Let's find out if there are missing values at our data frame: True.

The field conceded have 3 missing values.

The field that has a null vlaues is conceded, and it has 3 missing values.



## Question 5:

In [7]: #5

```
min_league = ''
temp = 0
flag = True

#Loop over array of the leagues, find the number of different teams for each league
for diff_league in df_soccer.league.unique():
    if flag:
        temp = len(df_soccer[df_soccer.league == diff_league].team.unique())
        flag = False
    print(diff_league + ': ', len(df_soccer[df_soccer.league == diff_league].team.unique()), end=' ')
    if len(df_soccer[df_soccer.league == diff_league].team.unique()) < temp:
        temp = len(df_soccer[df_soccer.league == diff_league].team.unique())
    min_league = diff_league
```

```
print()
print("min:", min_league)
```

La\_liga: 30  
EPL: 30  
Bundesliga: 24  
Serie\_A: 30  
Ligue\_1: 29  
RFPL: 25



min: Bundesliga

Number of different teams over the years of: La\_liga: 30 EPL: 30 Bundesliga: 24 Serie\_A: 30  
Ligue\_1: 29 RFPL: 25

The league that has the least number of teams over the years is: Bundesliga

## Question 6:

In [8]: #6

```
#Method for calculate the avarge of scored goals per year.
def calculateAvgGoals(row):
    avg = row.scored/row.matches
    return avg

#Adds new column to the data frame that store the avarge scored goals of each team
df_soccer["avg_goals"] = df_soccer.apply(calculateAvgGoals, axis = 'columns')

#Creates a groupby object of 'team'.
gb_team = df_soccer.groupby('team')

#Prints the columns that describe the avarge scored goals of each team for every year
for team, team_df_soccer in gb_team:
    print(team)
    display(team_df_soccer[['year', 'avg_goals']])
```

AC Milan

	year	avg_goals
357	2014	1.473684
374	2015	1.289474
393	2016	1.500000
413	2017	1.473684
432	2018	1.447368
453	2019	1.657895

Alaves

	year	avg_goals
48	2016	1.078947
73	2017	1.052632
89	2018	1.026316
115	2019	0.894737



Almeria

	year	avg_goals
--	------	-----------

<b>18</b>	2014	0.921053
-----------	------	----------

Amiens

	year	avg_goals
--	------	-----------

<b>540</b>	2017	0.973684
------------	------	----------

<b>563</b>	2018	0.815789
------------	------	----------

<b>586</b>	2019	1.107143
------------	------	----------

Amkar

	year	avg_goals
--	------	-----------

<b>598</b>	2014	0.833333
------------	------	----------

<b>614</b>	2015	0.733333
------------	------	----------

<b>629</b>	2016	0.833333
------------	------	----------

<b>648</b>	2017	0.666667
------------	------	----------

Angers

	year	avg_goals
--	------	-----------

<b>496</b>	2015	1.052632
------------	------	----------

<b>519</b>	2016	1.052632
------------	------	----------

<b>541</b>	2017	1.105263
------------	------	----------

<b>560</b>	2018	1.157895
------------	------	----------

<b>577</b>	2019	1.000000
------------	------	----------

Anzhi Makhachkala

	year	avg_goals
--	------	-----------

<b>616</b>	2015	0.933333
------------	------	----------

<b>630</b>	2016	0.800000
------------	------	----------

<b>649</b>	2017	1.033333
------------	------	----------

<b>666</b>	2018	0.433333
------------	------	----------

Arsenal

	year	avg_goals
--	------	-----------

<b>122</b>	2014	1.868421
------------	------	----------

<b>141</b>	2015	1.710526
------------	------	----------

<b>164</b>	2016	2.026316
------------	------	----------

<b>185</b>	2017	1.947368
------------	------	----------

<b>204</b>	2018	1.921053
------------	------	----------

<b>227</b>	2019	1.473684
------------	------	----------

Arsenal Tula

	year	avg_goals
<b>603</b>	2014	0.666667
<b>634</b>	2016	0.600000
<b>642</b>	2017	1.166667
<b>657</b>	2018	1.333333
<b>676</b>	2019	1.233333

## Aston Villa

	year	avg_goals
<b>135</b>	2014	0.815789
<b>159</b>	2015	0.710526
<b>236</b>	2019	1.078947

## Atalanta

	year	avg_goals
<b>364</b>	2014	1.000000
<b>379</b>	2015	1.078947
<b>391</b>	2016	1.631579
<b>414</b>	2017	1.500000
<b>430</b>	2018	2.026316
<b>451</b>	2019	2.578947

## Athletic Club

	year	avg_goals
<b>6</b>	2014	1.105263
<b>24</b>	2015	1.526316
<b>46</b>	2016	1.394737
<b>75</b>	2017	1.078947
<b>87</b>	2018	1.078947
<b>110</b>	2019	1.078947

## Atletico Madrid

	year	avg_goals
<b>2</b>	2014	1.763158
<b>22</b>	2015	1.657895
<b>42</b>	2016	1.842105
<b>61</b>	2017	1.526316
<b>81</b>	2018	1.447368
<b>103</b>	2019	1.342105

## Augsburg

	year	avg_goals
<b>244</b>	2014	1.264706
<b>271</b>	2015	1.235294
<b>288</b>	2016	1.029412
<b>305</b>	2017	1.264706
<b>326</b>	2018	1.500000
<b>343</b>	2019	1.323529

Barcelona

	year	avg_goals
<b>0</b>	2014	2.894737
<b>20</b>	2015	2.947368
<b>41</b>	2016	3.052632
<b>60</b>	2017	2.605263
<b>80</b>	2018	2.368421
<b>101</b>	2019	2.263158

Bayer Leverkusen

	year	avg_goals
<b>243</b>	2014	1.823529
<b>260</b>	2015	1.647059
<b>287</b>	2016	1.558824
<b>296</b>	2017	1.705882
<b>315</b>	2018	2.029412
<b>334</b>	2019	1.794118

Bayern Munich

	year	avg_goals
<b>240</b>	2014	2.352941
<b>258</b>	2015	2.352941
<b>276</b>	2016	2.617647
<b>294</b>	2017	2.705882
<b>312</b>	2018	2.588235
<b>330</b>	2019	2.941176

Benevento

	year	avg_goals
<b>427</b>	2017	0.868421

Bologna

	year	avg_goals
<b>381</b>	2015	0.868421
<b>402</b>	2016	1.052632
<b>422</b>	2017	1.052632
<b>437</b>	2018	1.263158
<b>458</b>	2019	1.378378

## Bordeaux

	year	avg_goals
<b>473</b>	2014	1.236842
<b>497</b>	2015	1.315789
<b>513</b>	2016	1.394737
<b>533</b>	2017	1.394737
<b>561</b>	2018	0.894737
<b>580</b>	2019	1.428571

## Borussia Dortmund

	year	avg_goals
<b>246</b>	2014	1.382353
<b>259</b>	2015	2.411765
<b>278</b>	2016	2.117647
<b>298</b>	2017	1.882353
<b>313</b>	2018	2.382353
<b>331</b>	2019	2.470588

## Borussia M.Gladbach

	year	avg_goals
<b>242</b>	2014	1.558824
<b>261</b>	2015	1.970588
<b>284</b>	2016	1.323529
<b>302</b>	2017	1.382353
<b>316</b>	2018	1.617647
<b>333</b>	2019	1.941176

## Bournemouth



	year	avg_goals
155	2015	1.184211
167	2016	1.447368
190	2017	1.184211
213	2018	1.473684
238	2019	1.052632

Brescia

	year	avg_goals
466	2019	0.921053

Brest

	year	avg_goals
582	2019	1.214286

Brighton

	year	avg_goals
194	2017	0.894737
216	2018	0.921053
234	2019	1.026316

Burnley

	year	avg_goals
138	2014	0.736842
176	2016	1.026316
186	2017	0.947368
214	2018	1.184211
228	2019	1.131579

CSKA Moscow

	year	avg_goals
589	2014	2.233333
604	2015	1.700000
621	2016	1.566667
637	2017	1.633333
655	2018	1.533333
671	2019	1.433333

Caen

	year	avg_goals
<b>480</b>	2014	1.421053
<b>494</b>	2015	1.026316
<b>523</b>	2016	0.947368
<b>543</b>	2017	0.710526
<b>566</b>	2018	0.763158

## Cagliari

	year	avg_goals
<b>365</b>	2014	1.263158
<b>399</b>	2016	1.447368
<b>423</b>	2017	0.868421
<b>442</b>	2018	0.947368
<b>460</b>	2019	1.368421

## Cardiff

	year	avg_goals
<b>217</b>	2018	0.894737

## Carpi

	year	avg_goals
<b>385</b>	2015	0.973684

## Celta Vigo

	year	avg_goals
<b>7</b>	2014	1.236842
<b>25</b>	2015	1.342105
<b>52</b>	2016	1.394737
<b>70</b>	2017	1.552632
<b>95</b>	2018	1.394737
<b>116</b>	2019	0.973684

## Cesena

	year	avg_goals
<b>367</b>	2014	0.947368

## Chelsea

	year	avg_goals
<b>120</b>	2014	1.921053
<b>149</b>	2015	1.552632
<b>160</b>	2016	2.236842
<b>184</b>	2017	1.631579
<b>202</b>	2018	1.657895
<b>223</b>	2019	1.815789

Chievo

	year	avg_goals
<b>361</b>	2014	0.736842
<b>376</b>	2015	1.131579
<b>401</b>	2016	1.131579
<b>421</b>	2017	0.947368
<b>447</b>	2018	0.657895

Cordoba

	year	avg_goals
<b>19</b>	2014	0.578947

Crotone

	year	avg_goals
<b>404</b>	2016	0.894737
<b>425</b>	2017	1.052632

Crystal Palace

	year	avg_goals
<b>129</b>	2014	1.236842
<b>154</b>	2015	1.026316
<b>173</b>	2016	1.315789
<b>189</b>	2017	1.184211
<b>211</b>	2018	1.342105
<b>233</b>	2019	0.815789

Darmstadt

	year	avg_goals
<b>269</b>	2015	1.117647
<b>293</b>	2016	0.823529

Deportivo La Coruna

	year	avg_goals
15	2014	0.921053
34	2015	1.184211
55	2016	1.131579
77	2017	1.000000

Dijon

	year	avg_goals
524	2016	1.210526
538	2017	1.447368
565	2018	0.815789
583	2019	0.964286

Dinamo Moscow

	year	avg_goals
591	2014	1.766667
618	2015	0.833333
643	2017	0.966667
663	2018	0.933333
673	2019	0.900000

Eibar

	year	avg_goals
17	2014	0.894737
33	2015	1.289474
49	2016	1.473684
68	2017	1.157895
91	2018	1.210526
113	2019	1.026316

Eintracht Frankfurt

	year	avg_goals
249	2014	1.647059
273	2015	1.000000
286	2016	1.058824
301	2017	1.323529
318	2018	1.764706
338	2019	1.735294

Elche

	year	avg_goals
--	------	-----------

12	2014	0.921053
----	------	----------

Empoli

	year	avg_goals
--	------	-----------

362	2014	1.210526
-----	------	----------

378	2015	1.052632
-----	------	----------

405	2016	0.763158
-----	------	----------

445	2018	1.342105
-----	------	----------

Espanyol

	year	avg_goals
--	------	-----------

10	2014	1.236842
----	------	----------

32	2015	1.052632
----	------	----------

47	2016	1.289474
----	------	----------

71	2017	0.947368
----	------	----------

86	2018	1.263158
----	------	----------

119	2019	0.710526
-----	------	----------

Everton

	year	avg_goals
--	------	-----------

131	2014	1.263158
-----	------	----------

151	2015	1.552632
-----	------	----------

166	2016	1.631579
-----	------	----------

187	2017	1.157895
-----	------	----------

207	2018	1.421053
-----	------	----------

231	2019	1.157895
-----	------	----------

Evian Thonon Gaillard

	year	avg_goals
--	------	-----------

485	2014	1.078947
-----	------	----------

FC Cologne

	year	avg_goals
--	------	-----------

250	2014	1.000000
-----	------	----------

266	2015	1.117647
-----	------	----------

281	2016	1.500000
-----	------	----------

311	2017	1.029412
-----	------	----------

344	2019	1.500000
-----	------	----------

FC Krasnodar

	year	avg_goals
--	------	-----------

590	2014	1.733333
-----	------	----------

607	2015	1.800000
-----	------	----------

623	2016	1.333333
-----	------	----------

639	2017	1.533333
-----	------	----------

654	2018	1.833333
-----	------	----------

670	2019	1.633333
-----	------	----------

FC Orenburg

	year	avg_goals
--	------	-----------

632	2016	0.833333
-----	------	----------

658	2018	1.300000
-----	------	----------

683	2019	0.933333
-----	------	----------

FC Rostov

	year	avg_goals
--	------	-----------

601	2014	0.900000
-----	------	----------

605	2015	1.366667
-----	------	----------

624	2016	1.200000
-----	------	----------

646	2017	0.900000
-----	------	----------

660	2018	0.833333
-----	------	----------

672	2019	1.500000
-----	------	----------

FC Tambov

	year	avg_goals
--	------	-----------

682	2019	1.233333
-----	------	----------

FC Ufa

	year	avg_goals
--	------	-----------

599	2014	0.866667
-----	------	----------

615	2015	0.833333
-----	------	----------

626	2016	0.733333
-----	------	----------

641	2017	1.133333
-----	------	----------

665	2018	0.800000
-----	------	----------

675	2019	0.733333
-----	------	----------

FC Yenisey Krasnoyarsk

	year	avg_goals
--	------	-----------

667	2018	0.8
-----	------	-----

FK Akhmat

	year	avg_goals
<b>596</b>	2014	1.000000
<b>610</b>	2015	1.166667
<b>625</b>	2016	1.266667
<b>644</b>	2017	1.000000
<b>659</b>	2018	0.933333
<b>680</b>	2019	0.900000

Fiorentina

	year	avg_goals
<b>351</b>	2014	1.605263
<b>372</b>	2015	1.578947
<b>395</b>	2016	1.657895
<b>415</b>	2017	1.421053
<b>441</b>	2018	1.236842
<b>457</b>	2019	1.297297

Fortuna Duesseldorf

	year	avg_goals
<b>321</b>	2018	1.441176
<b>346</b>	2019	1.058824

Freiburg

	year	avg_goals
<b>256</b>	2014	1.058824
<b>282</b>	2016	1.235294
<b>308</b>	2017	0.941176
<b>324</b>	2018	1.352941
<b>337</b>	2019	1.411765

Frosinone

	year	avg_goals
<b>386</b>	2015	0.921053
<b>446</b>	2018	0.763158

Fulham

	year	avg_goals
<b>218</b>	2018	0.894737

GFC Ajaccio

	year	avg_goals
<b>506</b>	2015	0.973684

Genoa

	year	avg_goals
353	2014	1.631579
377	2015	1.184211
403	2016	1.000000
419	2017	0.868421
444	2018	1.026316
464	2019	1.189189

Getafe

	year	avg_goals
13	2014	0.868421
38	2015	0.973684
67	2017	1.105263
85	2018	1.263158
107	2019	1.131579

Girona

	year	avg_goals
69	2017	1.315789
97	2018	0.973684

Granada

	year	avg_goals
16	2014	0.763158
36	2015	1.210526
59	2016	0.789474
106	2019	1.368421

Guingamp

	year	avg_goals
477	2014	1.078947
503	2015	1.236842
517	2016	1.210526
539	2017	1.263158
567	2018	0.736842

Hamburger SV



	year	avg_goals
<b>254</b>	2014	0.735294
<b>267</b>	2015	1.176471
<b>289</b>	2016	0.970588
<b>310</b>	2017	0.852941

Hannover 96

	year	avg_goals
<b>252</b>	2014	1.176471
<b>275</b>	2015	0.911765
<b>306</b>	2017	1.294118
<b>328</b>	2018	0.911765

Hertha Berlin

	year	avg_goals
<b>255</b>	2014	1.058824
<b>264</b>	2015	1.235294
<b>280</b>	2016	1.264706
<b>303</b>	2017	1.264706
<b>322</b>	2018	1.441176
<b>339</b>	2019	1.411765

Hoffenheim

	year	avg_goals
<b>247</b>	2014	1.441176
<b>272</b>	2015	1.147059
<b>279</b>	2016	1.882353
<b>297</b>	2017	1.941176
<b>320</b>	2018	2.058824
<b>335</b>	2019	1.558824

Huddersfield

	year	avg_goals
<b>195</b>	2017	0.736842
<b>219</b>	2018	0.578947

Hull

	year	avg_goals
<b>137</b>	2014	0.868421
<b>177</b>	2016	0.973684

Ingolstadt

	year	avg_goals
--	------	-----------

268	2015	0.970588
-----	------	----------

292	2016	1.058824
-----	------	----------

Inter

	year	avg_goals
--	------	-----------

355	2014	1.552632
-----	------	----------

371	2015	1.315789
-----	------	----------

394	2016	1.894737
-----	------	----------

411	2017	1.736842
-----	------	----------

431	2018	1.500000
-----	------	----------

449	2019	2.131579
-----	------	----------

Juventus

	year	avg_goals
--	------	-----------

348	2014	1.894737
-----	------	----------

368	2015	1.973684
-----	------	----------

388	2016	2.026316
-----	------	----------

408	2017	2.263158
-----	------	----------

428	2018	1.842105
-----	------	----------

448	2019	2.000000
-----	------	----------

Krylya Sovetov Samara

	year	avg_goals
--	------	-----------

612	2015	0.633333
-----	------	----------

633	2016	1.033333
-----	------	----------

664	2018	0.833333
-----	------	----------

681	2019	1.100000
-----	------	----------

Kuban Krasnodar

	year	avg_goals
--	------	-----------

597	2014	1.066667
-----	------	----------

617	2015	1.133333
-----	------	----------

Las Palmas

	year	avg_goals
--	------	-----------

31	2015	1.184211
----	------	----------

53	2016	1.394737
----	------	----------

78	2017	0.631579
----	------	----------

Lazio

	year	avg_goals
350	2014	1.868421
375	2015	1.368421
392	2016	1.947368
412	2017	2.342105
435	2018	1.473684
450	2019	2.078947

Lecce

	year	avg_goals
465	2019	1.324324

Leganes

	year	avg_goals
56	2016	0.947368
76	2017	0.894737
92	2018	0.973684
117	2019	0.789474

Leicester

	year	avg_goals
133	2014	1.210526
140	2015	1.789474
172	2016	1.263158
188	2017	1.473684
209	2018	1.342105
224	2019	1.763158

Lens

	year	avg_goals
487	2014	0.842105

Levante

	year	avg_goals
14	2014	0.894737
39	2015	0.973684
74	2017	1.157895
93	2018	1.552632
111	2019	1.236842

Lille

	year	avg_goals
<b>474</b>	2014	1.131579
<b>492</b>	2015	1.026316
<b>518</b>	2016	1.052632
<b>542</b>	2017	1.078947
<b>549</b>	2018	1.789474
<b>571</b>	2019	1.250000

Liverpool

	year	avg_goals
<b>125</b>	2014	1.368421
<b>147</b>	2015	1.657895
<b>163</b>	2016	2.052632
<b>183</b>	2017	2.210526
<b>201</b>	2018	2.342105
<b>220</b>	2019	2.236842

Lokomotiv Moscow

	year	avg_goals
<b>594</b>	2014	1.033333
<b>609</b>	2015	1.433333
<b>627</b>	2016	1.300000
<b>636</b>	2017	1.366667
<b>653</b>	2018	1.500000
<b>669</b>	2019	1.366667

Lorient

	year	avg_goals
<b>483</b>	2014	1.157895
<b>502</b>	2015	1.236842
<b>525</b>	2016	1.157895

Lyon

	year	avg_goals
<b>469</b>	2014	1.894737
<b>489</b>	2015	1.763158
<b>511</b>	2016	2.105263
<b>530</b>	2017	2.289474
<b>550</b>	2018	1.842105
<b>576</b>	2019	1.500000

Mainz 05

	year	avg_goals
251	2014	1.323529
263	2015	1.352941
290	2016	1.294118
307	2017	1.117647
323	2018	1.352941
342	2019	1.294118

## Malaga

	year	avg_goals
8	2014	1.105263
27	2015	1.000000
50	2016	1.289474
79	2017	0.631579

## Mallorca

	year	avg_goals
118	2019	1.052632

## Manchester City

	year	avg_goals
121	2014	2.184211
144	2015	1.868421
162	2016	2.105263
180	2017	2.789474
200	2018	2.500000
221	2019	2.684211

## Manchester United

	year	avg_goals
123	2014	1.631579
143	2015	1.289474
165	2016	1.421053
181	2017	1.789474
205	2018	1.710526
222	2019	1.736842

## Marseille

	year	avg_goals
<b>471</b>	2014	2.000000
<b>501</b>	2015	1.263158
<b>512</b>	2016	1.500000
<b>531</b>	2017	2.105263
<b>552</b>	2018	1.578947
<b>569</b>	2019	1.464286

Metz

	year	avg_goals
<b>486</b>	2014	0.815789
<b>521</b>	2016	1.026316
<b>547</b>	2017	0.894737
<b>581</b>	2019	0.964286

Middlesbrough

	year	avg_goals
<b>178</b>	2016	0.710526

Monaco

	year	avg_goals
<b>470</b>	2014	1.342105
<b>490</b>	2015	1.500000
<b>508</b>	2016	2.815789
<b>529</b>	2017	2.236842
<b>564</b>	2018	1.000000
<b>575</b>	2019	1.571429

Montpellier

	year	avg_goals
<b>475</b>	2014	1.210526
<b>499</b>	2015	1.289474
<b>522</b>	2016	1.263158
<b>537</b>	2017	0.947368
<b>553</b>	2018	1.394737
<b>574</b>	2019	1.250000

Mordovya

	year	avg_goals
<b>595</b>	2014	0.733333
<b>619</b>	2015	1.000000

Nancy

	year	avg_goals
526	2016	0.763158
Nantes		
	year	avg_goals
481	2014	0.763158
500	2015	0.868421
514	2016	1.052632
536	2017	0.947368
559	2018	1.263158
579	2019	1.000000
Napoli		
	year	avg_goals
352	2014	1.842105
369	2015	2.105263
390	2016	2.473684
409	2017	2.026316
429	2018	1.947368
454	2019	1.605263
Newcastle United		
	year	avg_goals
134	2014	1.052632
157	2015	1.157895
191	2017	1.026316
212	2018	1.105263
232	2019	1.000000
Nice		
	year	avg_goals
478	2014	1.157895
491	2015	1.526316
510	2016	1.657895
535	2017	1.394737
554	2018	0.789474
572	2019	1.464286
Nimes		

	year	avg_goals
556	2018	1.500000
585	2019	1.035714
Norwich		
	year	avg_goals
158	2015	1.026316
239	2019	0.684211
Nuernberg		
	year	avg_goals
329	2018	0.764706
Osasuna		
	year	avg_goals
58	2016	1.052632
109	2019	1.210526
PFC Sochi		
	year	avg_goals
679	2019	1.333333
Paderborn		
	year	avg_goals
257	2014	0.911765
347	2019	1.088235
Palermo		
	year	avg_goals
358	2014	1.394737
383	2015	1.000000
406	2016	0.868421
Paris Saint Germain		
	year	avg_goals
468	2014	2.184211
488	2015	2.684211
509	2016	2.184211
528	2017	2.842105
548	2018	2.763158
568	2019	2.777778
Parma		



	year	avg_goals
--	------	-----------

366	2014	0.868421
-----	------	----------

Parma Calcio 1913

	year	avg_goals
--	------	-----------

443	2018	1.078947
-----	------	----------

459	2019	1.405405
-----	------	----------

Pescara

	year	avg_goals
--	------	-----------

407	2016	0.921053
-----	------	----------

Queens Park Rangers

	year	avg_goals
--	------	-----------

139	2014	1.105263
-----	------	----------

RasenBallsport Leipzig

	year	avg_goals
--	------	-----------

277	2016	1.941176
-----	------	----------

299	2017	1.676471
-----	------	----------

314	2018	1.852941
-----	------	----------

332	2019	2.382353
-----	------	----------

Rayo Vallecano

	year	avg_goals
--	------	-----------

9	2014	1.210526
---	------	----------

37	2015	1.368421
----	------	----------

99	2018	1.078947
----	------	----------

Real Betis

	year	avg_goals
--	------	-----------

29	2015	0.894737
----	------	----------

54	2016	1.078947
----	------	----------

65	2017	1.578947
----	------	----------

90	2018	1.157895
----	------	----------

114	2019	1.263158
-----	------	----------

Real Madrid

	<b>year</b>	<b>avg_goals</b>
<b>1</b>	2014	3.105263
<b>21</b>	2015	2.894737
<b>40</b>	2016	2.789474
<b>62</b>	2017	2.473684
<b>82</b>	2018	1.657895
<b>100</b>	2019	1.842105

Real Sociedad

	<b>year</b>	<b>avg_goals</b>
<b>11</b>	2014	1.157895
<b>28</b>	2015	1.184211
<b>45</b>	2016	1.552632
<b>72</b>	2017	1.736842
<b>88</b>	2018	1.184211
<b>105</b>	2019	1.473684

Real Valladolid

	<b>year</b>	<b>avg_goals</b>
<b>96</b>	2018	0.842105
<b>112</b>	2019	0.842105

Reims

	<b>year</b>	<b>avg_goals</b>
<b>482</b>	2014	1.236842
<b>505</b>	2015	1.157895
<b>555</b>	2018	1.026316
<b>573</b>	2019	0.928571

Rennes

	<b>year</b>	<b>avg_goals</b>
<b>476</b>	2014	0.921053
<b>495</b>	2015	1.368421
<b>515</b>	2016	0.947368
<b>532</b>	2017	1.315789
<b>557</b>	2018	1.447368
<b>570</b>	2019	1.357143

Roma

	year	avg_goals
<b>349</b>	2014	1.421053
<b>370</b>	2015	2.184211
<b>389</b>	2016	2.368421
<b>410</b>	2017	1.605263
<b>433</b>	2018	1.736842
<b>452</b>	2019	2.026316

Rubin Kazan

	year	avg_goals
<b>592</b>	2014	1.300000
<b>613</b>	2015	1.100000
<b>628</b>	2016	1.000000
<b>645</b>	2017	1.066667
<b>662</b>	2018	0.800000
<b>677</b>	2019	0.600000

SC Bastia

	year	avg_goals
<b>479</b>	2014	0.973684
<b>498</b>	2015	0.947368
<b>527</b>	2016	0.763158

SD Huesca

	year	avg_goals
<b>98</b>	2018	1.131579

SKA-Khabarovsk

	year	avg_goals
<b>651</b>	2017	0.533333

SPAL 2013

	year	avg_goals
<b>424</b>	2017	1.026316
<b>440</b>	2018	1.157895
<b>467</b>	2019	0.702703

Saint-Etienne

	year	avg_goals
<b>472</b>	2014	1.342105
<b>493</b>	2015	1.105263
<b>516</b>	2016	1.078947
<b>534</b>	2017	1.236842
<b>551</b>	2018	1.552632
<b>584</b>	2019	1.035714

Sampdoria

	year	avg_goals
<b>354</b>	2014	1.263158
<b>382</b>	2015	1.263158
<b>398</b>	2016	1.289474
<b>416</b>	2017	1.473684
<b>436</b>	2018	1.578947
<b>461</b>	2019	1.263158

Sassuolo

	year	avg_goals
<b>359</b>	2014	1.289474
<b>373</b>	2015	1.289474
<b>397</b>	2016	1.578947
<b>418</b>	2017	0.763158
<b>438</b>	2018	1.394737
<b>455</b>	2019	1.864865

Schalke 04

	year	avg_goals
<b>245</b>	2014	1.235294
<b>262</b>	2015	1.500000
<b>285</b>	2016	1.323529
<b>295</b>	2017	1.558824
<b>325</b>	2018	1.088235
<b>341</b>	2019	1.117647

Sevilla

	year	avg_goals
4	2014	1.868421
26	2015	1.342105
43	2016	1.815789
66	2017	1.289474
84	2018	1.631579
102	2019	1.421053

Sheffield United

	year	avg_goals
229	2019	1.026316

Southampton

	year	avg_goals
126	2014	1.421053
145	2015	1.552632
168	2016	1.078947
196	2017	0.973684
215	2018	1.184211
230	2019	1.342105

Spartak Moscow

	year	avg_goals
593	2014	1.400000
608	2015	1.600000
620	2016	1.533333
638	2017	1.700000
656	2018	1.200000
674	2019	1.166667

Sporting Gijon

	year	avg_goals
35	2015	1.052632
57	2016	1.105263

Stoke

	year	avg_goals
128	2014	1.263158
148	2015	1.078947
171	2016	1.078947
197	2017	0.921053

Strasbourg

	year	avg_goals
544	2017	1.157895
558	2018	1.526316
578	2019	1.185185

Sunderland

	year	avg_goals
136	2014	0.815789
156	2015	1.263158
179	2016	0.763158

Swansea

	year	avg_goals
127	2014	1.210526
150	2015	1.105263
174	2016	1.184211
198	2017	0.736842

Tom Tomsk

	year	avg_goals
635	2016	0.566667

Torino

	year	avg_goals
356	2014	1.263158
380	2015	1.368421
396	2016	1.868421
417	2017	1.421053
434	2018	1.368421
463	2019	1.216216

Torpedo Moscow

	year	avg_goals
602	2014	0.933333

Tosno

	year	avg_goals
650	2017	0.766667

Tottenham

	year	avg_goals
124	2014	1.526316
142	2015	1.815789
161	2016	2.263158
182	2017	1.947368
203	2018	1.763158
225	2019	1.605263

Toulouse

	year	avg_goals
484	2014	1.131579
504	2015	1.184211
520	2016	0.973684
545	2017	1.000000
562	2018	0.921053
587	2019	0.785714

Troyes

	year	avg_goals
507	2015	0.736842
546	2017	0.842105

Udinese

	year	avg_goals
363	2014	1.131579
384	2015	0.921053
400	2016	1.236842
420	2017	1.263158
439	2018	1.026316
462	2019	0.972973

Union Berlin

	year	avg_goals
340	2019	1.205882

Ural

	year	avg_goals
<b>600</b>	2014	1.033333
<b>611</b>	2015	1.300000
<b>631</b>	2016	0.800000
<b>647</b>	2017	1.033333
<b>661</b>	2018	1.100000
<b>678</b>	2019	1.200000

## Valencia

	year	avg_goals
<b>3</b>	2014	1.842105
<b>30</b>	2015	1.210526
<b>51</b>	2016	1.473684
<b>63</b>	2017	1.710526
<b>83</b>	2018	1.342105
<b>108</b>	2019	1.210526

## Verona

	year	avg_goals
<b>360</b>	2014	1.289474
<b>387</b>	2015	0.894737
<b>426</b>	2017	0.789474
<b>456</b>	2019	1.270270

## VfB Stuttgart

	year	avg_goals
<b>253</b>	2014	1.235294
<b>274</b>	2015	1.470588
<b>300</b>	2017	1.058824
<b>327</b>	2018	0.941176

## Villarreal

	year	avg_goals
<b>5</b>	2014	1.263158
<b>23</b>	2015	1.157895
<b>44</b>	2016	1.473684
<b>64</b>	2017	1.500000
<b>94</b>	2018	1.289474
<b>104</b>	2019	1.657895

## Watford



	year	avg_goals
<b>152</b>	2015	1.052632
<b>175</b>	2016	1.052632
<b>193</b>	2017	1.157895
<b>210</b>	2018	1.368421
<b>237</b>	2019	0.947368

Werder Bremen

	year	avg_goals
<b>248</b>	2014	1.470588
<b>270</b>	2015	1.470588
<b>283</b>	2016	1.794118
<b>304</b>	2017	1.088235
<b>319</b>	2018	1.705882
<b>345</b>	2019	1.235294

West Bromwich Albion

	year	avg_goals
<b>132</b>	2014	1.000000
<b>153</b>	2015	0.894737
<b>170</b>	2016	1.131579
<b>199</b>	2017	0.815789

West Ham

	year	avg_goals
<b>130</b>	2014	1.157895
<b>146</b>	2015	1.710526
<b>169</b>	2016	1.236842
<b>192</b>	2017	1.263158
<b>208</b>	2018	1.368421
<b>235</b>	2019	1.289474

Wolfsburg

	year	avg_goals
<b>241</b>	2014	2.117647
<b>265</b>	2015	1.382353
<b>291</b>	2016	1.000000
<b>309</b>	2017	1.058824
<b>317</b>	2018	1.823529
<b>336</b>	2019	1.411765

Wolverhampton Wanderers


	year	avg_goals
206	2018	1.236842
226	2019	1.342105

Zenit St. Petersburg

	year	avg_goals
588	2014	1.933333
606	2015	2.033333
622	2016	1.666667
640	2017	1.533333
652	2018	1.900000
668	2019	2.166667

```
In [9]: #checking the correlation
print(df_soccer.position.corr(df_soccer.avg_goals))
```

```
-0.7587818572237601
```

The correlation we have got (-0.7587) tells us that there is a linear correlation ( $-1 < -0.7587 < 0$ ), and even a strong correlation according to the high value of the correlation. 

## Question 7:

```
In [10]: #7

#Creates a groupby object of 'League' and 'year', and calculate the median value of
gb_median = df_soccer.groupby(['league', 'year']).median()['scored']

#Prints the groupby object we have created.
print(gb_median)
```

league	year	
Bundesliga	2014	44.0
	2015	44.0
	2016	44.5
	2017	43.5
	2018	53.0
	2019	48.0
EPL	2014	46.5
	2015	48.5
	2016	47.5
	2017	44.5
	2018	51.5
	2019	46.5
La_liga	2014	43.0
	2015	45.5
	2016	53.0
	2017	46.5
	2018	47.0
	2019	46.0
Ligue_1	2014	44.0
	2015	46.0
	2016	42.5
	2017	45.5
	2018	46.0
	2019	33.0
RFPL	2014	31.0
	2015	34.5
	2016	30.5
	2017	31.5
	2018	30.5
	2019	36.5
Serie_A	2014	49.0
	2015	46.5
	2016	56.0
	2017	51.0
	2018	51.5
	2019	51.5

Name: scored, dtype: float64

C:\Users\liorb\AppData\Local\Temp\ipykernel\_24608\2117126896.py:4: FutureWarning: The default value of numeric\_only in DataFrameGroupBy.median is deprecated. In a future version, numeric\_only will default to False. Either specify numeric\_only or select only columns which should be valid for the function.

```
gb_median = df_soccer.groupby(['league', 'year']).median()['scored']
```

We have found that the league that has the highest difference between the medians over the years is "ligue\_1"(difference of 13 goals scored(46 is the highest at year 2018/2015 and 33 is the lowest at year 2019).



## Question 8:

In [11]: #8

*#Method for calculate the sum of pts each team got at each year.*

```
def calculatePts(row):
    sumOfPts = row.wins*3 + row.draws*1
    return sumOfPts
```

*#Adds new column to the data frame that store the sum of pts each team got at each year*

```
df_soccer['our_pts'] = df_soccer.apply(calculatePts, axis='columns')
```

*#Method for compare our values of pts to the values that given to us.*

```
def comparePts(row):
    if row.pts == row.our_pts:
        return True
    else:
        return None

#Adds new column to the data frame that store boolean value - "True" if our values
df_soccer['compare_pts'] = df_soccer.apply(comparePts, axis='columns')
display(df_soccer[['year', 'team', 'matches', 'wins', 'draws', 'loses', 'pts', 'our_pts', 'compare_pts']])

df_soccer.info()
print()
#we can use the `count()` method to count the number of non-NA values of 'compare_pts'
print("the number of non-NA values of 'compare_pts' column is: " + str(df_soccer['compare_pts'].count()))
print()
print("We will Find information about the mistake:")
display(df_soccer[df_soccer['compare_pts'].isna() == True])

print('the mistake happens to be in league is "Serie_A", year 2017, the team "Crotone")
```

	year	team	matches	wins	draws	loses	pts	our_pts	compare_pts
0	2014	Barcelona	38	30	4	4	94	94	True
1	2014	Real Madrid	38	30	2	6	92	92	True
2	2014	Atletico Madrid	38	23	9	6	78	78	True
3	2014	Valencia	38	22	11	5	77	77	True
4	2014	Sevilla	38	23	7	8	76	76	True
...	...	...	...	...	...	...	...	...	...
679	2019	PFC Sochi	30	8	9	13	33	33	True
680	2019	FK Akhmat	30	7	10	13	31	31	True
681	2019	Krylya Sovetov Samara	30	8	7	15	31	31	True
682	2019	FC Tambov	30	9	4	17	31	31	True
683	2019	FC Orenburg	30	7	6	17	27	27	True

684 rows × 9 columns



```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 684 entries, 0 to 683
Data columns (total 14 columns):
#   Column      Non-Null Count  Dtype
---  -
0   league      684 non-null    object
1   year        684 non-null    int64
2   position    684 non-null    int64
3   team        684 non-null    object
4   matches     684 non-null    int64
5   wins        684 non-null    int64
6   draws       684 non-null    int64
7   loses       684 non-null    int64
8   scored      684 non-null    int64
9   conceded    681 non-null    float64
10  pts         684 non-null    int64
11  avg_goals   684 non-null    float64
12  our_pts     684 non-null    int64
13  compare_pts 683 non-null    object
dtypes: float64(2), int64(9), object(3)
memory usage: 74.9+ KB
```

the number of non-NA values of 'compare\_pts' column is: 683 of total 684 entries.

We will Find information about the mistake:

	league	year	position	team	matches	wins	draws	loses	scored	conceded	pts	avg_g
425	Serie_A	2017	18	Crotone	38	9	8	21	40	66.0	180	1.05

According to that the "Serie\_A" league have 38 matches per year according to the information we got, and the max number that a team at that league and yaer can got is  $38 \times 3 = 116$  (if the team won all the games at that year), the given value of "pts" we have got is not possible, so the origin of the mistake is the given "pts" column

## Question 9:

In [12]: #9

```
#Method for compare the values of conceded goals to the values of scored goals.
def isFailure(row):
    if row.conceded != None:
        if row.conceded > row.scored:
            return True
        else:
            return None

#Adds new column to the data frame that store boolean value - "True" if a team is
df_soccer['is_failure'] = df_soccer.apply(isFailure, axis='columns')

#Gets an array that store the different leagues in the data frame.
arrOfLeagues = df_soccer.league.unique()

listCountFailure = []
#find the number of "failed teams" for over the years of each league and append to
for i in arrOfLeagues:
    filt = (df_soccer['league'] == i) & (df_soccer['is_failure'] == True)
    val = df_soccer.loc[filt].count()['is_failure']
    leagueAndVal = i + " " + str(val)
```

```
listCountFailure.append(leagueAndVal)  
  
display(listCountFailure)
```

```
['La_liga 75',  
 'EPL 71',  
 'Bundesliga 63',  
 'Serie_A 63',  
 'Ligue_1 69',  
 'RFPL 52']
```

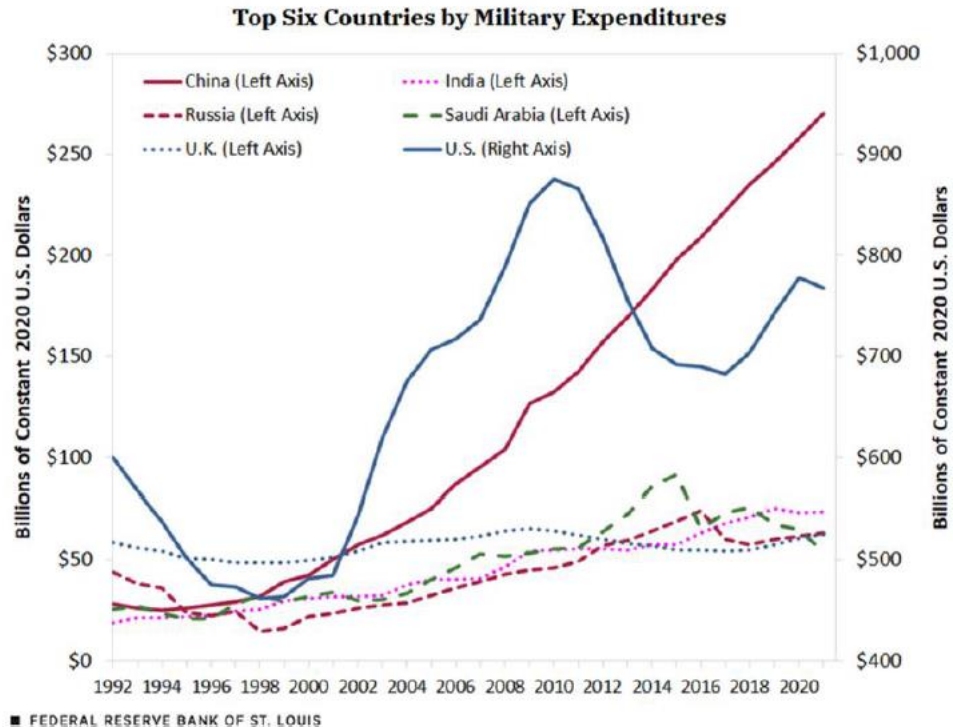
According to the results we got, the league with the most "failed teams" for over the years is "La\_liga" league.



## חלק ב' - הצגת מידע

### שאלה 1

בחנו את הגרף הבא וענו על השאלות שאחריו



A. מה תוכלו להסיק מן הגרף המוצג לגבי ההוצאה הצבאית של מדינות שונות בעולם לאורך השנים 1992 - 2020?

#### פתרון:

ניתן להסיק לפי הגרף המוצג על ההוצאה הצבאית של מדינות שונות בעולם בשנים 1992-2020 כי באופן כללי קיימת מגמת עלייה עם השנים, למעט אולי בריטניה שנראה שההוצאות הכספיות שלהם על הצבא די נשארו באותו סכום. כמו כן, ניתן לראות בבירור כי לסין ישנה עליה מאוד מגמתית במהלך השנים שבבדקו, ובנוסף, ניתן גם כן לראות כי לארצות הברית היה שיא בהוצאה הצבאית בסביבות 2010.

B. הצביעו על בעיה בגרף שעלולה לגרום לצופה בו להסיק מסקנות שגויות לגבי ההוצאה הצבאית של מדינות בעולם לאורך השנים 1992 - 2020.

#### פתרון:

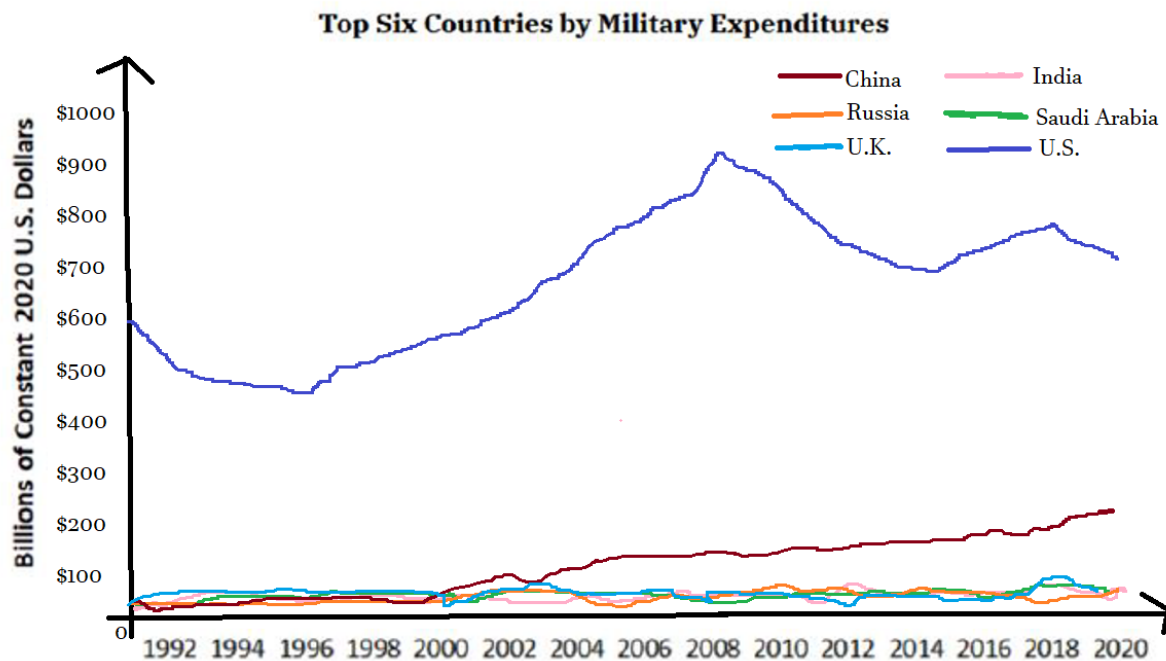
יוצר הגרף פיצל את ציר ה-y לשנים ובחר בסדרי גודל שונים לאותו משתנה, כאשר במקרא הוא מודיע לצופה שעבור כל המדינות צריך להתייחס לגרף שלהן ביחס לציר y השמאלי למעט אחת עבורה צריך להתייחס לגרף שלה ביחס לציר y השמאלי.

פעולה זו של פיצול ציר ה-y לשני צירי הגרף עלול לגרום לצופה להסיק באופן שגוי שהעלייה בהוצאות כספיות של מדינות לצורכי צבא והיחסים בהוצאות אלה בין המדינות המוצגות בשנים 1992-2021 עלו בצורה הרבה יותר חדה ממה שקרה במציאות.

C. הציעו ויזואליזציה חלופית שתציג את המידע שאמור להיות מועבר לצופה ואת הנתונים המוצגים בצורה אפקטיבית. אין צורך בכתיבת קוד, אך הוסיפו סקיצה פשוטה שתבהיר כיצד הוויזואליזציה החלופית שהצעתם תיראה.

### פתרון:

הוויזואליזציה שאנחנו חשבנו שתהיה יעילה יותר בהעברת המידע והנתונים לצופה היא גרף דומה לגרף הנתון רק נאחד את ציר הY לציר אחד רציף עם שינויים מתאימים ביחידות המידה וביחס שלהן.



בוויזואליזציה שיצרנו כל המדינות מתייחסות לאותו ציר Y בהתאם לוויזואליזציה החלופית שהצענו. היתרון של גרף זה הוא שלכל המדינות יש את אותו ציר שהן מתייחסות אליו, ועל כן נמנע הבלבול של הגרף שנתון לנו בשאלה. אולם, בגרף זה ישנם 4 מדינות עם הוצאות לצרכים צבאיים שנעות סביב ה-\$50 ועל כן קשה לשים לב להבדלים ביניהן.

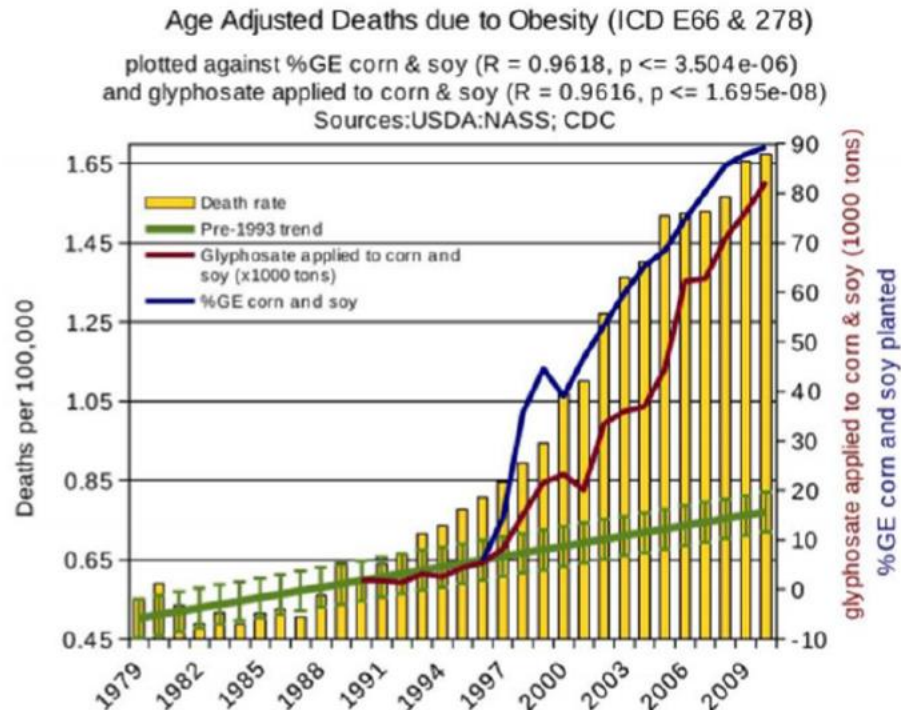


## שאלה 2

בחנו את הוויזואליזציה הבאה וענו על השאלות שאחריה.

**שימו לב:** Glyphosate הוא סוג של קוטל עשבים הנפוץ בגידולים חקלאיים רבים;

GE = Genetically engineered = מהונדס גנטית; obesity = עודף משקל.



A. מה ניתן להסיק לגבי הקשר שבין מוות מעודף משקל, השימוש ב-Glyphosate בגידולי תירס וסויה, ואחוז גידולי תירס וסויה המהונדסים גנטית?

### פתרון:

ניתן לראות בגרף כי ישנה מגמה של עלייה במס' המתים בכל שנה (per 100,000) ובזמנית, עליה בשימוש ב-Glyphosate לאורך השנים וכן, של אחוז גידולי תירס וסויה המהונדסים גנטית מידי שנה. לכן, ניתן להסיק, בהתאם לגרף המצורף, כי קיים קשר כלשהו בין מוות מעודף משקל ובין השימוש ב-Glyphosate לאורך השנים וכן, של אחוז גידולי תירס וסויה המהונדסים גנטית מידי שנה.

B. בעיצוב הגרף, יוצרו קיבלו מספר החלטות לא סטנדרטיות על מנת להעביר מסר מסוים. מהו לדעתכם המסר שיוצרי הגרף ביקשו לייצר באמצעותו? הסבירו. האם אתם מסכימים עם המסר?

### פתרון:

לדעתנו, המסר אותו ניסו להעביר יוצרי הגרף הינו כי ישנו קשר סיבתי לפיו עליה בשימוש ב-Glyphosate לאורך השנים וכן, של אחוז גידולי תירס וסויה המהונדסים גנטית מידי שנה מובילה באופן ישיר לעודף משקל וכתוצאה מעודף משקל ישנה עליה במס' המתים מידי שנה. ההחלטות שבחרו יוצרי הגרף על מנת להעביר את אותו מסר הן:

- עליה בשימוש ב-Glyphosate לאורך השנים וכן, של אחוז גידולי תירס וסויה המהונדסים גנטית מידי שנה על אותו ציר ה-Y, ומכך נובע כי ישנו עומס של נתונים באותו הגרף, וכן נוצר הרושם לפיו שני המשתנים על ציר ה-Y הם כמו משתנה אחד, וזאת במקום להפרידם לשני גרפים שונים וכך להתייחס להשפעה של כל אחד מהם על התמותה בנפרד, ללא הצגת מצב בו לכאורה הם "זהים".
- הוספת קו המגמה של התמותה עד לשנת 1993, וזאת על מנת לגרום לצופה להסיק כי התמותה הצפויה

נמוכה מזאת שבפועל מאז שהחל השימוש ב שימוש ב- Glyphosate לאורך השנים וכן, של אחוז גידולי תירס וסויה המהונדסים גנטית.

3. הגרף מראה, לכאורה קשר בין עליה בשימוש ב- Glyphosate לאורך השנים וכן, של אחוז גידולי תירס וסויה המהונדסים גנטית מידי שנה לבין עליה במס' המתים מידי שנה, אולם, לא ניתן להסיק מכך שום דבר על הקשר בין עודף משקל לעליה במס' המתים מכיוון שהגרף אינו מציג כלל שעליה בשימוש ב- Glyphosate לאורך השנים וכן, של אחוז גידולי תירס וסויה המהונדסים גנטית מידי שנה מובילה באופן ישיר לעודף משקל, ועל כן, לא ניתן להסיק שום דבר על קשר כלשהו בין עודף משקל למס' מתים מידי שנה, ועל כן, איננו מסכימים עם המסר אותו הם ניסו, לדעתנו, להעביר.



C. מצאו בוויזואליזציה לפחות 2 בעיות שגורמות לה להיות פחות אפקטיבית. הסבירו מהי הבעיה ומדוע זו נחשבת בעיה.

### פתרון:

- בעיות שמצאנו בוויזואליזציה שגורמות לה להיות פחות אפקטיבית הן:
1. להתחיל למדוד את השימוש ב- Glyphosate לאורך השנים וכן, של אחוז גידולי תירס וסויה המהונדסים גנטית מידי שנה מערך התחלתי של 10- למרות שברור כי לא ניתן להתחיל למדוד כך וזה יוצר אשליה לפיה העלייה בשימוש ב- Glyphosate לאורך השנים וכן, של אחוז גידולי תירס וסויה המהונדסים גנטית מידי שנה צמודה לעלייה במס' המתים מידי שנה.
  2. ניתן לראות כי הגרף הנתון מכיל בתוכו יותר מידי מידע ועל כן נוצר עומס שמקשה על הקורא להבין את המשמעות של כל ייצוג ויזואלי.
  3. לא מוצג לנו שום קשר בין עליה בשימוש ב- Glyphosate לאורך השנים וכן, של אחוז גידולי תירס וסויה המהונדסים גנטית מידי שנה לבין עליה במס' האנשים הסובלים מעודף משקל, אלא לבין מס' מקרי מוות בשנה(לכאורה).
  4. שני צירים אנכיים שמייצגים ערכים שונים עם שנתות שונות שגם כן יוצרת בלבול אצל הקורא שדורשת ממנו לעשות מאמץ מיותר כדי להבין את הוויזואליזציה.
  5. ישנם פסים אופקיים ברקע של הגרף שהם די עבים שגם כן גורמים לעומס אצל הקורא.



### שאלה 3

בחנו את הוויזואליזציה הבאה וענו על השאלות שאחריה. שימו לב: CAGR = קצב צמיחה שנתי ממוצע



A. מהו המידע שאמור להיות מועבר לצופה באמצעות הוויזואליזציה הזו?

#### פתרון:

המידע שאמור להיות מועבר לצופה בוויזואליזציה שמולנו הוא שהתחזיות הן שנתח השוק של תחליפי הגבינה הטבעוניים יגדל בין השנים 2018-2026 מתוצר של 2.48 מיליארד דולר ל 4.72 מיליארד דולר שהם עליה ממוצעת של 8.3 אחוז בכל שנה.

B. מצאו בוויזואליזציה לפחות 3 בעיות שגורמות לה להיות פחות אפקטיבית. הסבירו מהי הבעיה ומדוע זו בעיה.

#### פתרון:

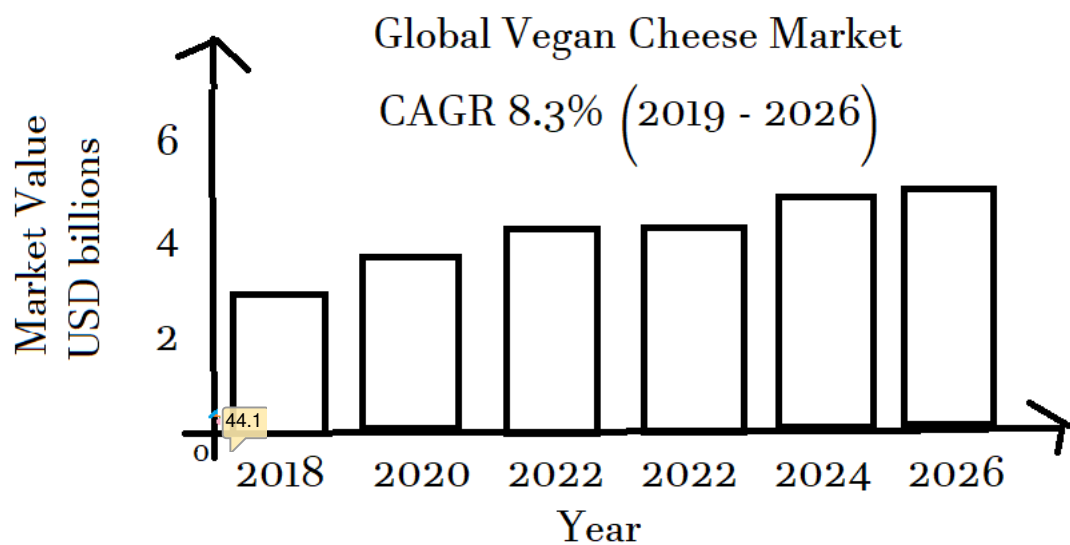
הוויזואליזציה שמולנו מכילה כמה בעיות אשר גורמות לה להיות פחות אפקטיבית:

1. בנוסף הבחירה בתרשים פאי ובפרט בתלת מימד הוא בעייתי, כמו שלמדנו בהרצאות, לאדם קשה עד בלתי אפשרי להבדיל בין החלקים שגודלם די קרוב ( כמו למשל הכחול והאדום) ולכן פוגע באפקטיביות של הגרף בכך שאינו נותן לצופה להבין את היחס האמיתי שבין הגדלים של נתחי השוק של תחליפי הגבינה.
2. לבסוף יוצר הגרף בחר לקחת שני נתחים מהפאי בלבד ולספק עליהם מידע מאוד מועט של שנה וכמות של כסף כאשר לא ניתן לדעת למה כוונת יוצר הגרף ביחסי הגדלים שבין נתחי הפאי. בעיה זאת גורמת לוויזואליזציה להיות פחות אפקטיבית מאחר והיא לא נותן לצופה את מלוא המידע לו הוא זקוק כדי להבין ביחס למה גדלי החלקים של 2018 ו2026 הם כאלה ולא אחרים.
3. נעשה שימוש ביצירת הגרף בהמון אלמנטים וויזואליים שמסיחים את הדעת מהמסר אותו הם רצו להעביר(ההסחה באה לידי ביטוי ברקע של גבינות שלא תורם שום דבר, וכן בשימוש בצבעים עם ניגודיות גבוהה שגם כן לא משקפים דבר.
4. יוצר הגרף רצה להעביר מידע תחזיתי לגבי שינוי בקצב צמיחה ביחס לזמן, כאשר ראינו בהרצאות שגרף זה אמור להיות עם לפחות משתנה אחד רציף (של שנים) ובמקרה שלפנינו אולי אפילו שניים. הבעיה היא בחירה לא מדויקת של הגרף בגרף שלפנינו אי הדיוק בבחירת הגרף פוגעת ביכולת של הצופה לקבל מידע ולהסיק מסקנות באופן פשוט ויעיל לגבי הצמיחה של שוק תחליפי הגבינה.

C. הציעו ויזואליזציה חלופית שתציג את המידע שאמור להיות מועבר לצופה ואת הנתונים המוצגים בצורה אפקטיבית. אין צורך בכתיבת קוד, אך הוסיפו סקיצה פשוטה שתבהיר כיצד הויזואליזציה החלופית שהצעתם תיראה.

### פתרון:

לדעתנו ניתן להציג את המידע והנתונים לצופה בצורה אפקטיבית יותר באמצעות גרף עמודות שבנוי מציר X על גביו נעות השנים וציר Y שבו סכומי כסף במיליוני דולרים. כל עמודה מציגת שנה וגובהה הוא הסכום ששוק תחליפי הגבינה הכניס באותה שנה, כך שקיימת עליה (נראית לעין) בין עמודה לעמודה כפי שמצוין בגרף שקיבלנו של 8.3 אחוז בממוצע בשנה בסה"כ ניתן לראות את העלייה הכללית בהבדלים בין שנת 2018 לשנת 2026.



## שאלה 4

מהי הבעיה בכותרת הבאה (NC = North Carolina):

**Average NC teacher pay is nearly \$58,000, state says. But educators argue many earn less.**

### פתרון:

הבעיה בכותרת הבאה הינה כי הממוצע הוא מושפע מערכים קיצוניים, ועל כן הוא אינו מדד אמין לייצוג משכורות המורים בצפון קרוליינה. מה שקורה זה שלמרות הנכונות של המידע, הוא איננו אינדיקטור אמיתי למשכורות המורים במדינה. לכן, המורים שחולקים על טענת המדינה יכול להיות שאכן צודקים בכך שלמרות הממוצע שהוצג, הרבה מורים מרוויחים פחות מכך. ראינו בהרצאה שעל מנת לייצג נאמנה את שכר המורים החודשי, ניתן לחשב את חציון שכר המורים אשר אינו מושפע מערכים קיצוניים כמו הממוצע.

נקודות 2- .אותו להניח ואין חסר הוא? הגיע המידע מאיפה 3C 44.1