

# datap

November 18, 2019

```
[1]: #importing libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
from sklearn.preprocessing import LabelEncoder
```

```
↳ -----
ModuleNotFoundError                                Traceback (most recent call↳
↳ last)

<ipython-input-1-d18bd1a5d262> in <module>
      2 import numpy as np
      3 import pandas as pd
----> 4 import matplotlib.pyplot as plt
      5 import seaborn as sns
      6 import warnings

ModuleNotFoundError: No module named 'matplotlib'
```

```
[31]: #importing libraries
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
from sklearn.preprocessing import LabelEncoder
```

```
[45]: #importing dataset
dataset = pd.read_csv('World.csv')
```

```
[46]: #displaying first 5 rows of dataset
dataset.head(5)
```

```
[46]:
```

	Year	Datetime	Stage	Stadium	City \
0	1930.0	13 Jul 1930 - 15:00	Group 1	Pocitos	Montevideo
1	1930.0	13 Jul 1930 - 15:00	Group 4	Parque Central	Montevideo
2	1930.0	14 Jul 1930 - 12:45	Group 2	Parque Central	Montevideo
3	1930.0	14 Jul 1930 - 14:50	Group 3	Pocitos	Montevideo
4	1930.0	15 Jul 1930 - 16:00	Group 1	Parque Central	Montevideo

  

	Home Team Name	Home Team Goals	Away Team Goals	Away Team Name \
0	France	4.0	1.0	Mexico
1	USA	3.0	0.0	Belgium
2	Yugoslavia	2.0	1.0	Brazil
3	Romania	3.0	1.0	Peru
4	Argentina	1.0	0.0	France

  

	Win conditions	Attendance	Half-time Home Goals	Half-time Away Goals \
0		4444.0	3.0	0.0
1		18346.0	2.0	0.0
2		24059.0	2.0	0.0
3		2549.0	1.0	0.0
4		23409.0	0.0	0.0

  

	Referee	Assistant 1 \
0	LOMBARDI Domingo (URU)	CRISTOPHE Henry (BEL)
1	MACIAS Jose (ARG)	MATEUCCI Francisco (URU)
2	TEJADA Anibal (URU)	VALLARINO Ricardo (URU)
3	WARNKEN Alberto (CHI)	LANGENUS Jean (BEL)
4	REGO Gilberto (BRA)	SAUCEDO Ulises (BOL)

  

	Assistant 2	RoundID	MatchID	Home Team Initials \
0	REGO Gilberto (BRA)	201.0	1096.0	FRA
1	WARNKEN Alberto (CHI)	201.0	1090.0	USA
2	BALWAY Thomas (FRA)	201.0	1093.0	YUG
3	MATEUCCI Francisco (URU)	201.0	1098.0	ROU
4	RADULESCU Constantin (ROU)	201.0	1085.0	ARG

  

	Away Team Initials
0	MEX
1	BEL
2	BRA
3	PER
4	FRA

```
[47]: dataset.tail(5)
```

```
[47]:      Year Datetime Stage Stadium City Home Team Name Home Team Goals \
4567   NaN      NaN   NaN   NaN   NaN   NaN      NaN      NaN
4568   NaN      NaN   NaN   NaN   NaN   NaN      NaN      NaN
4569   NaN      NaN   NaN   NaN   NaN   NaN      NaN      NaN
4570   NaN      NaN   NaN   NaN   NaN   NaN      NaN      NaN
4571   NaN      NaN   NaN   NaN   NaN   NaN      NaN      NaN

      Away Team Goals Away Team Name Win conditions Attendance \
4567      NaN      NaN      NaN      NaN      NaN
4568      NaN      NaN      NaN      NaN      NaN
4569      NaN      NaN      NaN      NaN      NaN
4570      NaN      NaN      NaN      NaN      NaN
4571      NaN      NaN      NaN      NaN      NaN

      Half-time Home Goals Half-time Away Goals Referee Assistant 1 \
4567      NaN      NaN      NaN      NaN      NaN
4568      NaN      NaN      NaN      NaN      NaN
4569      NaN      NaN      NaN      NaN      NaN
4570      NaN      NaN      NaN      NaN      NaN
4571      NaN      NaN      NaN      NaN      NaN

      Assistant 2 RoundID MatchID Home Team Initials Away Team Initials
4567      NaN      NaN      NaN      NaN      NaN      NaN
4568      NaN      NaN      NaN      NaN      NaN      NaN
4569      NaN      NaN      NaN      NaN      NaN      NaN
4570      NaN      NaN      NaN      NaN      NaN      NaN
4571      NaN      NaN      NaN      NaN      NaN      NaN
```

```
[48]: dataset.shape
```

```
[48]: (4572, 20)
```

```
[49]: dataset.index
```

```
[49]: RangeIndex(start=0, stop=4572, step=1)
```

```
[50]: dataset.columns
```

```
[50]: Index(['Year', 'Datetime', 'Stage', 'Stadium', 'City', 'Home Team Name',
        'Home Team Goals', 'Away Team Goals', 'Away Team Name',
        'Win conditions', 'Attendance', 'Half-time Home Goals',
        'Half-time Away Goals', 'Referee', 'Assistant 1', 'Assistant 2',
        'RoundID', 'MatchID', 'Home Team Initials', 'Away Team Initials'],
        dtype='object')
```

```
[51]: dataset.isnull()
```

```
[51]:      Year Datetime Stage Stadium City Home Team Name Home Team Goals \
0      False      False False      False False      False      False
1      False      False False      False False      False      False
```

[illegible]

4560	True	True	True	True	True	True	True
4561	True	True	True	True	True	True	True
4562	True	True	True	True	True	True	True
4563	True	True	True	True	True	True	True
4564	True	True	True	True	True	True	True
4565	True	True	True	True	True	True	True
4566	True	True	True	True	True	True	True
4567	True	True	True	True	True	True	True
4568	True	True	True	True	True	True	True
4569	True	True	True	True	True	True	True
4570	True	True	True	True	True	True	True
4571	True	True	True	True	True	True	True

	Away Team	Goals	Away Team Name	Win conditions	Attendance	\
0		False	False	False	False	
1		False	False	False	False	
2		False	False	False	False	
3		False	False	False	False	
4		False	False	False	False	
5		False	False	False	False	
6		False	False	False	False	
7		False	False	False	False	
8		False	False	False	False	
9		False	False	False	False	
10		False	False	False	False	
11		False	False	False	False	
12		False	False	False	False	
13		False	False	False	False	
14		False	False	False	False	
15		False	False	False	False	
16		False	False	False	False	
17		False	False	False	False	
18		False	False	False	False	
19		False	False	False	False	
20		False	False	False	False	
21		False	False	False	False	
22		False	False	False	False	
23		False	False	False	False	
24		False	False	False	False	
25		False	False	False	False	
26		False	False	False	False	
27		False	False	False	False	
28		False	False	False	False	
29		False	False	False	False	
...		...	...	...	...	
4542		True	True	True	True	
4543		True	True	True	True	

4544	True	True	True	True
4545	True	True	True	True
4546	True	True	True	True
4547	True	True	True	True
4548	True	True	True	True
4549	True	True	True	True
4550	True	True	True	True
4551	True	True	True	True
4552	True	True	True	True
4553	True	True	True	True
4554	True	True	True	True
4555	True	True	True	True
4556	True	True	True	True
4557	True	True	True	True
4558	True	True	True	True
4559	True	True	True	True
4560	True	True	True	True
4561	True	True	True	True
4562	True	True	True	True
4563	True	True	True	True
4564	True	True	True	True
4565	True	True	True	True
4566	True	True	True	True
4567	True	True	True	True
4568	True	True	True	True
4569	True	True	True	True
4570	True	True	True	True
4571	True	True	True	True

	Half-time Home Goals	Half-time Away Goals	Referee	Assistant 1	\
0	False	False	False	False	
1	False	False	False	False	
2	False	False	False	False	
3	False	False	False	False	
4	False	False	False	False	
5	False	False	False	False	
6	False	False	False	False	
7	False	False	False	False	
8	False	False	False	False	
9	False	False	False	False	
10	False	False	False	False	
11	False	False	False	False	
12	False	False	False	False	
13	False	False	False	False	
14	False	False	False	False	
15	False	False	False	False	
16	False	False	False	False	

17	False	False	False	False
18	False	False	False	False
19	False	False	False	False
20	False	False	False	False
21	False	False	False	False
22	False	False	False	False
23	False	False	False	False
24	False	False	False	False
25	False	False	False	False
26	False	False	False	False
27	False	False	False	False
28	False	False	False	False
29	False	False	False	False
...	...	...	...	...
4542	True	True	True	True
4543	True	True	True	True
4544	True	True	True	True
4545	True	True	True	True
4546	True	True	True	True
4547	True	True	True	True
4548	True	True	True	True
4549	True	True	True	True
4550	True	True	True	True
4551	True	True	True	True
4552	True	True	True	True
4553	True	True	True	True
4554	True	True	True	True
4555	True	True	True	True
4556	True	True	True	True
4557	True	True	True	True
4558	True	True	True	True
4559	True	True	True	True
4560	True	True	True	True
4561	True	True	True	True
4562	True	True	True	True
4563	True	True	True	True
4564	True	True	True	True
4565	True	True	True	True
4566	True	True	True	True
4567	True	True	True	True
4568	True	True	True	True
4569	True	True	True	True
4570	True	True	True	True
4571	True	True	True	True

	Assistant 2	RoundID	MatchID	Home Team Initials	Away Team Initials
0	False	False	False	False	False

1	False	False	False	False	False
2	False	False	False	False	False
3	False	False	False	False	False
4	False	False	False	False	False
5	False	False	False	False	False
6	False	False	False	False	False
7	False	False	False	False	False
8	False	False	False	False	False
9	False	False	False	False	False
10	False	False	False	False	False
11	False	False	False	False	False
12	False	False	False	False	False
13	False	False	False	False	False
14	False	False	False	False	False
15	False	False	False	False	False
16	False	False	False	False	False
17	False	False	False	False	False
18	False	False	False	False	False
19	False	False	False	False	False
20	False	False	False	False	False
21	False	False	False	False	False
22	False	False	False	False	False
23	False	False	False	False	False
24	False	False	False	False	False
25	False	False	False	False	False
26	False	False	False	False	False
27	False	False	False	False	False
28	False	False	False	False	False
29	False	False	False	False	False
...	...	...	...	...	...
4542	True	True	True	True	True
4543	True	True	True	True	True
4544	True	True	True	True	True
4545	True	True	True	True	True
4546	True	True	True	True	True
4547	True	True	True	True	True
4548	True	True	True	True	True
4549	True	True	True	True	True
4550	True	True	True	True	True
4551	True	True	True	True	True
4552	True	True	True	True	True
4553	True	True	True	True	True
4554	True	True	True	True	True
4555	True	True	True	True	True
4556	True	True	True	True	True
4557	True	True	True	True	True
4558	True	True	True	True	True



4559	True	True	True	True	True
4560	True	True	True	True	True
4561	True	True	True	True	True
4562	True	True	True	True	True
4563	True	True	True	True	True
4564	True	True	True	True	True
4565	True	True	True	True	True
4566	True	True	True	True	True
4567	True	True	True	True	True
4568	True	True	True	True	True
4569	True	True	True	True	True
4570	True	True	True	True	True
4571	True	True	True	True	True

[4572 rows x 20 columns]

```
[52]: dataset.isnull().sum()
```

```
[52]: Year                3720
      Datetime            3720
      Stage              3720
      Stadium            3720
      City               3720
      Home Team Name     3720
      Home Team Goals    3720
      Away Team Goals    3720
      Away Team Name     3720
      Win conditions     3720
      Attendance         3722
      Half-time Home Goals 3720
      Half-time Away Goals 3720
      Referee            3720
      Assistant 1        3720
      Assistant 2        3720
      RoundID            3720
      MatchID            3720
      Home Team Initials  3720
      Away Team Initials  3720
      dtype: int64
```

```
[53]: dataset.dropna(inplace= True)
      dataset.isnull().sum()
```

```
[53]: Year                0
      Datetime            0
      Stage              0
      Stadium            0
      City               0
      Home Team Name     0
```

```

Home Team Goals      0
Away Team Goals      0
Away Team Name       0
Win conditions       0
Attendance           0
Half-time Home Goals 0
Half-time Away Goals 0
Referee              0
Assistant 1          0
Assistant 2          0
RoundID              0
MatchID              0
Home Team Initials   0
Away Team Initials   0
dtype: int64

```

```
[54]: dataset.shape
```

```
[54]: (850, 20)
```

```
[55]: dataset.tail(10)
```

```
[55]:
      Year      Datetime      Stage \
842  2014.0  01 Jul 2014 - 13:00      Round of 16
843  2014.0  01 Jul 2014 - 17:00      Round of 16
844  2014.0  04 Jul 2014 - 13:00  Quarter-finals
845  2014.0  04 Jul 2014 - 17:00  Quarter-finals
846  2014.0  05 Jul 2014 - 13:00  Quarter-finals
847  2014.0  05 Jul 2014 - 17:00  Quarter-finals
848  2014.0  08 Jul 2014 - 17:00      Semi-finals
849  2014.0  09 Jul 2014 - 17:00      Semi-finals
850  2014.0  12 Jul 2014 - 17:00  Play-off for third place
851  2014.0  13 Jul 2014 - 16:00      Final

      Stadium      City Home Team Name  Home Team Goals \
842  Arena de Sao Paulo  Sao Paulo  Argentina          1.0
843  Arena Fonte Nova  Salvador  Belgium            2.0
844  Estadio do Maracana  Rio De Janeiro  France            0.0
845  Estadio Castelao  Fortaleza  Brazil            2.0
846  Estadio Nacional  Brasilia  Argentina          1.0
847  Arena Fonte Nova  Salvador  Netherlands        0.0
848  Estadio Mineirao  Belo Horizonte  Brazil            1.0
849  Arena de Sao Paulo  Sao Paulo  Netherlands        0.0
850  Estadio Nacional  Brasilia  Brazil            0.0
851  Estadio do Maracana  Rio De Janeiro  Germany            1.0

      Away Team Goals  Away Team Name      Win conditions \
842              0.0  Switzerland  Argentina win after extra time
843              1.0      USA      Belgium win after extra time

```

844	1.0	Germany	
845	1.0	Colombia	
846	0.0	Belgium	
847	0.0	Costa Rica	Netherlands win on penalties (4 - 3)
848	7.0	Germany	
849	0.0	Argentina	Argentina win on penalties (2 - 4)
850	3.0	Netherlands	
851	0.0	Argentina	Germany win after extra time

	Attendance	Half-time Home Goals	Half-time Away Goals \
842	63255.0	0.0	0.0
843	51227.0	0.0	0.0
844	74240.0	0.0	1.0
845	60342.0	1.0	0.0
846	68551.0	1.0	0.0
847	51179.0	0.0	0.0
848	58141.0	0.0	5.0
849	63267.0	0.0	0.0
850	68034.0	0.0	2.0
851	74738.0	0.0	0.0

	Referee	Assistant 1 \
842	ERIKSSON Jonas (SWE)	KLASENIUS Mathias (SWE)
843	HAIMOUDI Djamel (ALG)	ACHIK Redouane (MAR)
844	PITANA Nestor (ARG)	MAIDANA Hernan (ARG)
845	Carlos VELASCO CARBALLO (ESP)	ALONSO FERNANDEZ Roberto (ESP)
846	Nicola RIZZOLI (ITA)	Renato FAVERANI (ITA)
847	Ravshan IRMATOV (UZB)	RASULOV Abduxamidullo (UZB)
848	RODRIGUEZ Marco (MEX)	TORRETERA Marvin (MEX)
849	Cneyt AKIR (TUR)	DURAN Bahattin (TUR)
850	HAIMOUDI Djamel (ALG)	ACHIK Redouane (MAR)
851	Nicola RIZZOLI (ITA)	Renato FAVERANI (ITA)

	Assistant 2	RoundID	MatchID	Home Team Initials \
842	WARNSMARK Daniel (SWE)	255951.0	300186503.0	ARG
843	ETCHIALI Abdelhak (ALG)	255951.0	300186497.0	BEL
844	BELATTI Juan Pablo (ARG)	255953.0	300186485.0	FRA
845	YUSTE Juan (ESP)	255953.0	300186461.0	BRA
846	Andrea STEFANI (ITA)	255953.0	300186504.0	ARG
847	KOCHKAROV Bakhadyr (KGZ)	255953.0	300186488.0	NED
848	QUINTERO Marcos (MEX)	255955.0	300186474.0	BRA
849	ONGUN Tarik (TUR)	255955.0	300186490.0	NED
850	ETCHIALI Abdelhak (ALG)	255957.0	300186502.0	BRA
851	Andrea STEFANI (ITA)	255959.0	300186501.0	GER

Away Team Initials	
842	SUI

843	USA
844	GER
845	COL
846	BEL
847	CRC
848	GER
849	ARG
850	NED
851	ARG

```
[56]: dataset['Year'].tail()
```

```
[56]: 847    2014.0
      848    2014.0
      849    2014.0
      850    2014.0
      851    2014.0
      Name: Year, dtype: float64
```

```
[57]: dataset['Year'].replace(np.NaN, dataset['Year'].mean())
```

```
[57]: 0      1930.0
      1      1930.0
      2      1930.0
      3      1930.0
      4      1930.0
      5      1930.0
      6      1930.0
      7      1930.0
      8      1930.0
      9      1930.0
     10      1930.0
     11      1930.0
     12      1930.0
     13      1930.0
     14      1930.0
     15      1930.0
     16      1930.0
     17      1930.0
     18      1934.0
     19      1934.0
     20      1934.0
     21      1934.0
     22      1934.0
     23      1934.0
     24      1934.0
     25      1934.0
     26      1934.0
     27      1934.0
```

```

28      1934.0
29      1934.0
...
820     2014.0
821     2014.0
822     2014.0
824     2014.0
825     2014.0
826     2014.0
827     2014.0
828     2014.0
829     2014.0
830     2014.0
831     2014.0
832     2014.0
833     2014.0
834     2014.0
835     2014.0
836     2014.0
837     2014.0
838     2014.0
839     2014.0
840     2014.0
842     2014.0
843     2014.0
844     2014.0
845     2014.0
846     2014.0
847     2014.0
848     2014.0
849     2014.0
850     2014.0
851     2014.0

```

Name: Year, Length: 850, dtype: float64

```
[58]: dataset['Year'].mean()
```

```
[58]: 1985.0211764705882
```

```
[59]: ds = pd.read_csv('Data.csv')
      ds1 = ds
      ds
```

```
[59]:
```

	Country	Age	Salary	Purchased
0	France	44.0	72000.0	No
1	Spain	27.0	48000.0	Yes
2	Germany	30.0	54000.0	No
3	Spain	38.0	61000.0	No
4	Germany	40.0	NaN	Yes

5	France	35.0	58000.0	Yes
6	Spain	NaN	52000.0	No
7	France	48.0	79000.0	Yes
8	Germany	50.0	83000.0	No
9	France	37.0	67000.0	Yes

```
[60]: ds["Age"].fillna("30",inplace=True)
```

```
[61]: ds
```

```
[61]:
```

	Country	Age	Salary	Purchased
0	France	44	72000.0	No
1	Spain	27	48000.0	Yes
2	Germany	30	54000.0	No
3	Spain	38	61000.0	No
4	Germany	40	NaN	Yes
5	France	35	58000.0	Yes
6	Spain	30	52000.0	No
7	France	48	79000.0	Yes
8	Germany	50	83000.0	No
9	France	37	67000.0	Yes

```
[62]: ds["Salary"].fillna("00",inplace=True)
```

```
[63]: ds
```

```
[63]:
```

	Country	Age	Salary	Purchased
0	France	44	72000	No
1	Spain	27	48000	Yes
2	Germany	30	54000	No
3	Spain	38	61000	No
4	Germany	40	00	Yes
5	France	35	58000	Yes
6	Spain	30	52000	No
7	France	48	79000	Yes
8	Germany	50	83000	No
9	France	37	67000	Yes

```
[71]: x = ds.iloc[-1: , :-1].values
x
```

```
[71]: array([[ 'France', 37.0, 67000.0]], dtype=object)
```

```
[76]: label_encoder = LabelEncoder()
```

```
[79]: x[:, 0] = label_encoder.fit_transform(x[:, 0])
```

```
[80]: x
```

```
[80]: array([[0, 37.0, 67000.0]], dtype=object)
```

```
[ ]:
```