```
from google.colab import drive


# mount the drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```
# read file from the drive
import pandas as pd


datasetPath = '/content/drive/MyDrive/matches_1930_2022.csv'


df = pd.read_csv(datasetPath)


df
```

	home_team	away_team	home_score	home_xg	home_penalty	away_score	away_xg	away_penalty	hom
0	Argentina	France	3	3.3	4.0	3	2.2	2.0	Li
1	Croatia	Morocco	2	0.7	NaN	1	1.2	NaN	
2	France	Morocco	2	2.0	NaN	0	0.9	NaN	Didier
3	Argentina	Croatia	3	2.3	NaN	0	0.5	NaN	Li
4	Morocco	Portugal	1	1.4	NaN	0	0.9	NaN	Hoa
...	...	...	...	...	...	...	...	...	
959	Argentina	France	1	NaN	NaN	0	NaN	NaN	Fran
960	Yugoslavia	Brazil	2	NaN	NaN	1	NaN	NaN	Bosk
961	Romania	Peru	3	NaN	NaN	1	NaN	NaN	Oc
962	United States	Belgium	3	NaN	NaN	0	NaN	NaN	
963	France	Mexico	4	NaN	NaN	1	NaN	NaN	Ra

964 rows × 44 columns

```
df.columns
```

```
Index(['home_team', 'away_team', 'home_score', 'home_xg', 'home_penalty',
       'away_score', 'away_xg', 'away_penalty', 'home_manager', 'home_captain',
       'away_manager', 'away_captain', 'Attendance', 'Venue', 'Officials',
       'Round', 'Date', 'Score', 'Referee', 'Notes', 'Host', 'Year',
       'home_goal', 'away_goal', 'home_goal_long', 'away_goal_long',
       'home_own_goal', 'away_own_goal', 'home_penalty_goal',
       'away_penalty_goal', 'home_penalty_miss_long', 'away_penalty_miss_long',
       'home_penalty_shootout_goal_long', 'away_penalty_shootout_goal_long',
       'home_penalty_shootout_miss_long', 'away_penalty_shootout_miss_long',
       'home_red_card', 'away_red_card', 'home_yellow_red_card',
       'away_yellow_red_card', 'home_yellow_card_long',
       'away_yellow_card_long', 'home_substitute_in_long',
       'away_substitute_in_long'],
      dtype='object')
```

```
df['home_team'].info()
```

```
<class 'pandas.core.series.Series'>
RangeIndex: 964 entries, 0 to 963
Series name: home_team
Non-Null Count  Dtype
--------------  -----
964 non-null    object
dtypes: object(1)
memory usage: 7.7+ KB
```

```
df['home_team'].value_counts()
```

home_team	count
Brazil	85
Argentina	59
Italy	57
England	40
Germany	40
...	...
Jamaica	1
FR Yugoslavia	1
Angola	1
Trinidad and Tobago	1
Bosnia and Herzegovina	1

82 rows × 1 columns

dtype: int64

```
df['Date'].value_counts()
```

Date	count
1958-06-15	8
1958-06-08	8
1934-05-27	8
1958-06-11	7
1938-06-05	6
...	...
2006-07-08	1
2006-07-05	1
2006-07-04	1
1966-07-25	1
2022-12-18	1

378 rows × 1 columns

dtype: int64

```
df['Score'].value_counts()
```

Score	count
1–0	118
2–1	107
1–1	75
2–0	71
0–0	66
...	...
(4) 2–2 (3)	1
(3) 0–0 (2)	1
(4) 2–2 (5)	1
(4) 1–1 (3)	1
6–5	1

71 rows × 1 columns

dtype: int64

```
ScoredAboveZero = df['Score'] > '0–0'

df[ScoredAboveZero]
```

```
(929, 44)
```

```
ScoredByBrazil = df['home_team'] == 'Brazil'

df[ScoredAboveZero & ScoredByBrazil]
```

	home_team	away_team	home_score	home_xg	home_penalty	away_score	away_xg	away_penalty	
11	Brazil	Korea Republic	4	3.6	NaN	1	0.5	NaN	
34	Brazil	Switzerland	1	1.0	NaN	0	0.3	NaN	
51	Brazil	Serbia	2	2.4	NaN	0	0.2	NaN	
71	Brazil	Belgium	1	2.8	NaN	2	0.5	NaN	
74	Brazil	Mexico	2	2.7	NaN	0	0.7	NaN	
...	...	...	...	...	...	...	...	...	
911	Brazil	Sweden	4	NaN	NaN	2	NaN	NaN	
915	Brazil	Czechoslovakia	2	NaN	NaN	1	NaN	NaN	
917	Brazil	Czechoslovakia	1	NaN	NaN	1	NaN	NaN	
926	Brazil	Poland	6	NaN	NaN	5	NaN	NaN	
951	Brazil	Bolivia	4	NaN	NaN	0	NaN	NaN	Pir

81 rows × 44 columns

```
brazildf = df[ScoredAboveZero & ScoredByBrazil].sort_values(by = 'home_score',ascending= True)
```

```
brazildf
```

	home_team	away_team	home_score	home_xg	home_penalty	away_score	away_xg	away_penalty	
512	Brazil	Argentina	0	NaN	NaN	1	NaN	NaN	
261	Brazil	France	0	NaN	NaN	1	NaN	NaN	Ca
695	Brazil	Poland	0	NaN	NaN	1	NaN	NaN	
817	Brazil	Czechoslovakia	0	NaN	NaN	0	NaN	NaN	
679	Brazil	Spain	0	NaN	NaN	0	NaN	NaN	
...	...	...	...	...	...	...	...	...	
887	Brazil	Mexico	5	NaN	NaN	0	NaN	NaN	
831	Brazil	France	5	NaN	NaN	2	NaN	NaN	
891	Brazil	Spain	6	NaN	NaN	1	NaN	NaN	
926	Brazil	Poland	6	NaN	NaN	5	NaN	NaN	
893	Brazil	Sweden	7	NaN	NaN	1	NaN	NaN	

81 rows × 44 columns

```
brazildf.info()
```

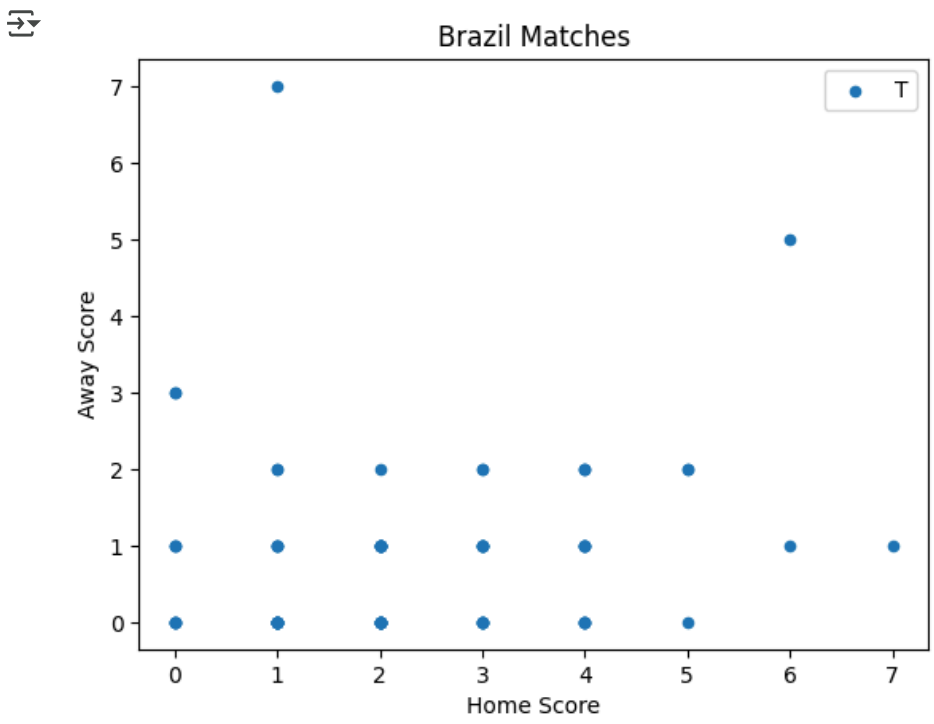
```
<class 'pandas.core.frame.DataFrame'>
Index: 81 entries, 512 to 893
Data columns (total 44 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   home_team                       81 non-null     object
 1   away_team                       81 non-null     object
 2   home_score                      81 non-null     int64
 3   home_xg                         7 non-null      float64
 4   home_penalty                    0 non-null      float64
 5   away_score                      81 non-null     int64
 6   away_xg                         7 non-null      float64
 7   away_penalty                    0 non-null      float64
 8   home_manager                    81 non-null     object
 9   home_captain                    59 non-null     object
 10  away_manager                    81 non-null     object
 11  away_captain                    59 non-null     object
 12  Attendance                      81 non-null     int64
 13  Venue                           81 non-null     object
 14  Officials                       64 non-null     object
 15  Round                           81 non-null     object
 16  Date                            81 non-null     object
 17  Score                           81 non-null     object
 18  Referee                         64 non-null     object
 19  Notes                           3 non-null      object
 20  Host                            81 non-null     object
 21  Year                            81 non-null     int64
 22  home_goal                       71 non-null     object
 23  away_goal                       38 non-null     object
 24  home_goal_long                  71 non-null     object
 25  away_goal_long                  38 non-null     object
 26  home_own_goal                   2 non-null      object
 27  away_own_goal                   2 non-null      object
 28  home_penalty_goal               8 non-null      object
 29  away_penalty_goal               8 non-null      object
 30  home_penalty_miss_long          0 non-null      object
 31  away_penalty_miss_long          0 non-null      object
 32  home_penalty_shootout_goal_long 0 non-null      object
```

```
 33  away_penalty_shootout_goal_long  0 non-null      object
 34  home_penalty_shootout_miss_long  0 non-null      object
 35  away_penalty_shootout_miss_long  0 non-null      object
 36  home_red_card                    4 non-null      object
 37  away_red_card                    3 non-null      object
 38  home_yellow_red_card             1 non-null      object
 39  away_yellow_red_card             3 non-null      object
 40  home_yellow_card_long            43 non-null     object
 41  away_yellow_card_long            47 non-null     object
 42  home_substitute_in_long          51 non-null     object
 43  away_substitute_in_long          55 non-null     object
dtypes: float64(4), int64(4), object(36)
memory usage: 28.5+ KB
```

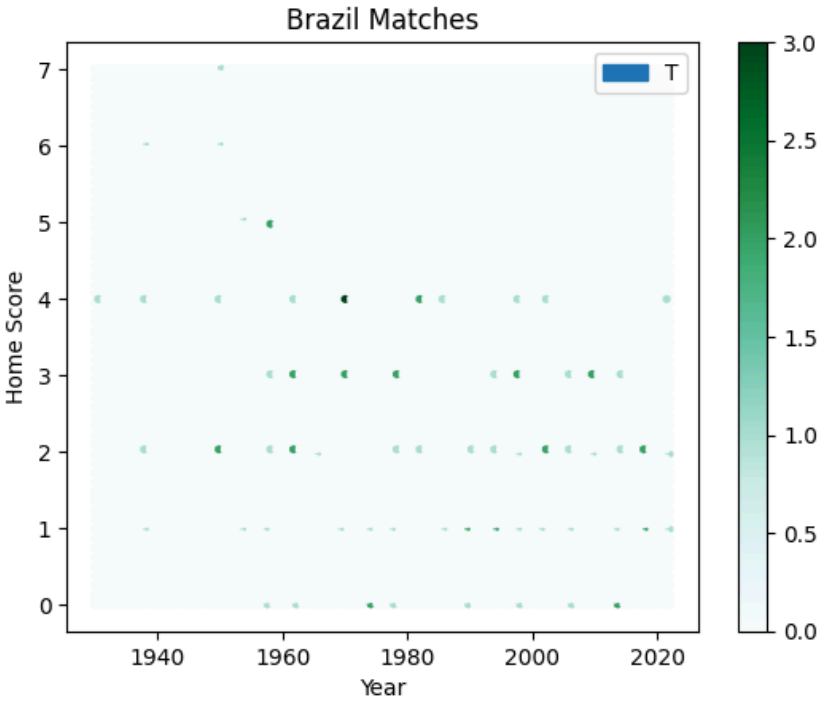
```
%matplotlib inline
```

```
from matplotlib import pyplot as plt
```

```
brazildf.plot(x='home_score',y='away_score',kind='scatter')
plt.xlabel('Home Score')
plt.ylabel('Away Score')
plt.legend('Takeaway')
plt.title('Brazil Matches')
plt.show()
```




```
brazildf.plot(x='Year',y='home_score',kind='hexbin')
plt.xlabel('Year')
plt.ylabel('Home Score')
plt.legend('Takeaway')
plt.title('Brazil Matches')
plt.show()
```


Brazil Matches
T
Home Score
Year
0
1
2
3
4
5
6
7
1940
1960
1980
2000
2020
0.0
0.5
1.0
1.5
2.0
2.5
3.0