```
In [1]: import numpy as np
   import pandas as pd
   import matplotlib.pyplot as plt
   import seaborn as sns
   import sklearn
   from sklearn.linear_model import LogisticRegression
   from sklearn.model_selection import train_test_split
   from sklearn import metrics
   import warnings
   warnings.filterwarnings("ignore")
```

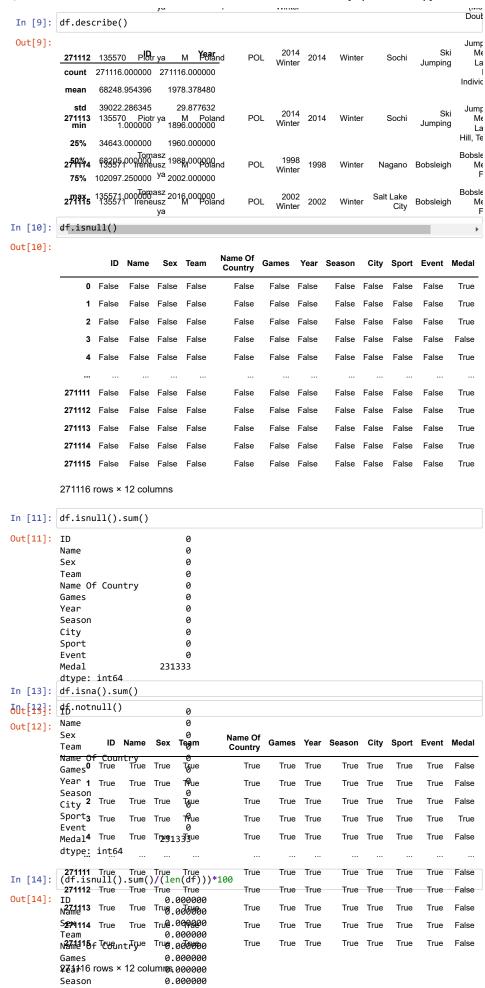
In [2]: df=pd.read_csv(r"E:\Linear regression projects\athlete_events.csv")

In [3]: df

Out[3]:

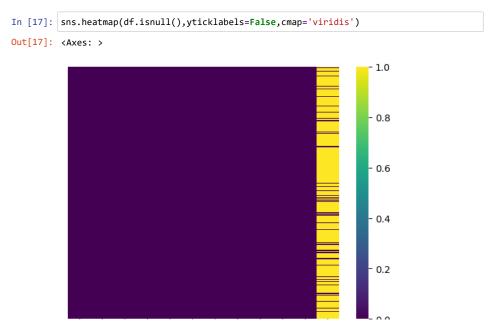
	City	Season	Year	nes	Gan	Name Of Country	Team		Sex	Name	ID	
E	Barcelona	Summer	1992	992 mer	19 Sumr	CHN	China		М	A Dijiang	1	0
	London	Summer	2012	012 mer	20 Sumr	CHN	China		М	A Lamusi	2	1
	Antwerpen	Summer A	1920	920 mer	19 Sumr	DEN	Denmark		М	Gunnar Nielsen Aaby	3	2
	Paris	Summer	1900	900 mer	19 Sumr	DEN	rk/Sweden	Denma	М	Edgar Lindenau Aabye	4	3
	Calgary	Winter	1988	988 nter		NED	etherlands	N	F	Christine Jacoba Aaftink	5	4
	Innsbruck	Winter	1976	976 nter		POL	Poland-1		М	Andrzej ya	135569	271111
	Sochi	Winter	2014	014 nter		POL	Poland		М	Piotr ya	135570	271112
	Sochi	Winter	2014	014 nter		POL	Poland		М	Piotr ya	135570	271113
	Nagano	Winter	1998	998 nter		POL	Poland		М	Tomasz Ireneusz ya	135571	
	Salt Lake City	Winter	2002	002 nter		POL	Poland Name		М	Tomasz Ireneusz ya	135571	271115
t	Sport	City	ason	Se	Year	Games	Of Country	Team		Sex 2 columns	Name	ID 271116
ı	Basketball	Barcelona	mmer	Su	1992	1992 Summer	CHN	China		M		1 0 1
						2012						df.shap
)	Judo	London	mmer	Su	2012	Summer	CHN	China		М	A,La1ln2U)si	(271216
I	Football	Antwerpen	mmer	Su	1920	1920 Summer	DEN	enmark	D	M		df.size
	Tug-Of- War	Paris	mmer	Su	1900	1900 Summer	DEN	Sweden	nmark/s	M Den	Edgar Lindenau Aabye	3 4
	Speed Skating	Calgary	Vinter	١	1988	1988 Winter	NED	erlands	Neth	F	Christine Jacoba Aaftink	4 5

```
Winter 1990
                                                                                AAIIIICI
                                                                                        ічауано вов
 In [6]: df.head()
                            Tomasz
                                                                    2002
                                                                                         Salt Lake
 Out[6]: 271115 135571
                                     М
                                                 Poland
                                                            POL
                                                                         2002
                                                                                                  Bobs
                           Ireneusz
                                                                   Winter
                                                                                             City
                                                  Name
Of
              ID
                    Name Sex
                                         Team
                                                         Games
                                                                 Year Season
                                                                                    City
                                                                                             Sport
                                                Country
          271116 rows x 12 column
                                                                                                    Ŗа
          0 1 A Dijiang M
                                         China
                                                   CHN
                                                                 1992 Summer
                                                                               Barcelona
                                                        Summer
                                                                                                    Ba
 In [4]: df.shape
 Out[4]: (12711216A, Lahau)si
                                                           2012
                                         China
                                                   CHN
                                                                2012 Summer
                                                                                             Judo
                                                        Summer
                                                                                                   Ligh
 In [5]: df.size Gunnar
                                                           1920
 Out[5]: 3253392 Nielsen
                                       Denmark
                                                   DEN
                                                                 1920 Summer Antwerpen
                                                                                           Football
                                                        Summer
                    Edgar
                                                                                           Tug-Of-
                                                           1900
                                                                                                    Wa
              4 Lindenau
                            M Denmark/Sweden
                                                   DEN
                                                                 1900 Summer
                                                                                   Paris
                                                        Summer
                                                                                              .
War
                    Aabye
                  Christine
                                                           1988
                                                                                            Speed
                   Jacoba
                                    Netherlands
                                                   NED
                                                                 1988
                                                                        Winter
                                                                                 Calgary
                                                          Winter
                                                                                            Skating
                   Aaftink
                                                                                                   500
 In [7]: df.info()
           <class 'pandas.core.frame.DataFrame'>
          RangeIndex: 271116 entries, 0 to 271115
          Data columns (total 12 columns):
           #
                Column
                                   Non-Null Count
                                                      Dtype
           0
                ID
                                   271116 non-null
                                                      int64
           1
                Name
                                   271116 non-null
                                                      object
           2
                                    271116 non-null
                Sex
           3
                Team
                                   271116 non-null
                                                      object
           4
                Name Of Country 271116 non-null
                                                      object
           5
                Games
                                   271116 non-null
                                                       object
                                   271116 non-null
           6
                Year
                                                       int64
                                   271116 non-null
           7
                Season
                                                      object
           8
                City
                                   271116 non-null
                                                      object
           9
                Sport
                                    271116 non-null
                                   271116 non-null object
           10
                Event
                                   39783 non-null
           11 Medal
                                                      object
          dtypes: int64(2), object(10)
          memory usage: 24.8+ MB
 In [8]: df.tail()
 Out[8]:
                                                  Name
                                         Team
                                                    Of
                                                        Games Year
                                                                                                    Εv
                                                Country
                                                                                                     L
                           Andrzej
                                        Poland-
                                                           1976
                                                                                                    Mi
           271111 135569
                                                   POL
                                                                1976
                                                         Winter
                                                                                                   (Me
                                                                                                   Doub
 In [9]: df.describe()
 Out[9]:
                                                                                                  Jump
                                                          2014
                                                                                             Ski
           271112 135570 Plotr ya
                                     M Poland
                                                                                                    Me
                                                   POL
                                                                2014
                                                                       Winter
                                                         Winter
                                                                                         Jumping
                                                                                                    La
                  271116.000000 271116.000000
                                                                                                 Individ
           mean
                   68248.954396
                                  1978.378480
              std
                   39022.286345
                                    29.877632
                                                                                                  Jump
                                                          2014
                                                                                             Ski
                  135570 Piotr ya
1.000000
                                                                2014
                                  M Poland
1896.000000
                                                         Winter
                                                                                         Jumping
                                                                                                 Hill, Te
                   34643.000000
                                  1960.000000
           Bobsle
                                                           1998
                                                                1998
                                                                       Winter
                                                                                Nagano Bobsleigh
                                                         Winter
             75% 102097.250000 <sup>ya</sup> 2002.000000
                                                                                                 Bobsle
           max 135571,000000 271115 135571 Ireneusz M Poland
                                                          2002
                                                                              Salt Lake
                                                   POL
                                                                2002
                                                                       Winter
                                                                                       Bobsleigh
                                                                                                    Me
                                                         Winter
                                                                                   City
In [10]: df.isnull()
Out[10]:
                                              Name Of
                        Name
                                Sex Team
                                                      Games
                                                              Year Season
                                                                             City Sport Event
                                                                                               Medal
                                              Country
                                                                                                 True
                1 False False False
                                                                      False
                                                                                         False
                                                                                                 True
```

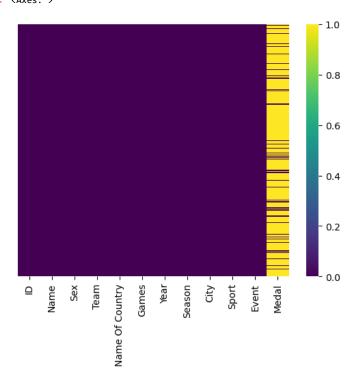


```
atype: int64
In [13]: df.isna().sum()
JUt[13]: df.notnull()
Out[12]: Name
                                      0
           Sex
                                               Name Of
                                     Team
                     ID Name
                               Sex
                                                        Games Year Season City Sport Event Medal
           Team
                                                Country
                  f Cour
True
           Name
          Games 0
                          True
                               True
                                                   True
                                                           True
                                                                True
                                                                              True
                                                                                           True
                                                                                                 False
                                      Tgue
                                                                        True
                                                                                    True
           Year 1 True
                                      T∳ue
                          True True
                                                   True
                                                                True
                                                                                           True
                                                                                                 False
                                                           True
                                                                        True
                                                                             True
                                                                                    True
           Season
City 2 True
                          True
                               True
                                      Tgue
                                                   True
                                                           True
                                                                True
                                                                        True
                                                                              True
                                                                                    True
                                                                                           True
                                                                                                 False
           Sport<sub>3</sub> True
                                      True
                          True
                                                           True
                                                                True
                                                                                           True
                                                                                                  True
                               True
                                                   True
                                                                        True
                                                                              True
                                                                                    True
           Event
          Medal<sup>4</sup> True
                          True
                                                   True
                                                           True
                                                                True
                                                                                           True
                                                                                                 False
                               Tr21931335ue
                                                                        True
                                                                              True
                                                                                    True
           dtype: int64
In [14]: (df.isnull().sum()/(len(df)))*100
                                                                                                 False
                                                   True
                                                           True
                                                                True
                                                                         True
                                                                              True
                                                                                    True
                                                                                           True
            271112 True
                                     True
                                                                                                 False
                          True True
                                                   True
                                                           True
                                                                True
                                                                        True
                                                                             True
                                                                                    True
                                                                                           True
Out[14]: ID
                                  0.000000
           271113 True
                         True True.00000
                                                   True
                                                           True
                                                                True
                                                                        True
                                                                             True
                                                                                    True
                                                                                           True
                                                                                                 False
           S271114 True
                               Tru@ . 00 00 00 00 00
                          True
                                                   True
                                                           True
                                                                True
                                                                         True
                                                                              True
                                                                                    True
                                                                                           True
                                                                                                 False
                                  0.000000
           Name of Teventrye
                               Tru8.000000
                                                   True
                                                           True
                                                                True
                                                                        True
                                                                              True
                                                                                    True
                                                                                           True
                                                                                                 False
           Games
                                  0.000000
           2₹1116 rows × 12 columnes 000000
                                  0.000000
           Season
           City
                                  0.000000
                                  0.000000
           Sport
                                 0.000000
           Event
          Medal
                                 85.326207
           dtype: float64
In [15]: df.drop
Out[15]: <bound method DataFrame.drop of
                                                             ID
                                                                                         Name Sex
           Team Name Of Country \
           0
                         1
                                              A Dijiang
                                                            М
                                                                          China
                                                                                              CHN
          1
                         2
                                               A Lamusi
                                                            Μ
                                                                          China
                                                                                              CHN
           2
                          3
                                   Gunnar Nielsen Aaby
                                                                        Denmark
                                                                                              DEN
           3
                          4
                                  Edgar Lindenau Aabye
                                                            М
                                                               Denmark/Sweden
                                                                                              DEN
           4
                          5
                             Christine Jacoba Aaftink
                                                            F
                                                                   Netherlands
                                                                                              NED
                                             Andrzej ya
           271111 135569
                                                                      Poland-1
                                                            М
                                                                                              POL
           271112
                   135570
                                               Piotr ya
                                                            М
                                                                         Poland
                                                                                              P<sub>0</sub>L
                   135570
           271113
                                               Piotr ya
                                                                         Poland
                                                                                              POL
           271114
                   135571
                                    Tomasz Ireneusz ya
                                                            Μ
                                                                         Poland
                                                                                              POL
           271115 135571
                                                            М
                                                                                              P<sub>0</sub>L
                                    Tomasz Ireneusz ya
                                                                         Poland
                           Games
                                   Year
                                          Season
                                                              City
                                                                              Sport
          0
                    1992 Summer
                                   1992
                                                         Barcelona
                                                                         Basketball
                                          Summer
          1
                    2012 Summer
                                   2012
                                          Summer
                                                            London
                                                                                Judo
           2
                    1920 Summer
                                   1920
                                                                           Football
                                          Summer
                                                         Antwerpen
           3
                    1900 Summer
                                   1900
                                          Summer
                                                             Paris
                                                                         Tug-Of-War
                    1988 Winter
           4
                                   1988
                                          Winter
                                                                     Speed Skating
                                                           Calgary
           271111
                    1976 Winter
                                   1976
                                          Winter
                                                         Innsbruck
                                                                               Luge
                    2014 Winter
                                                                        Ski Jumping
           271112
                                   2014
                                          Winter
                                                             Sochi
In [16]: 271113
271114
                    2014 Winter
                                   2014
                                          Winter
                                                             Sochi
                                                                        Ski Jumping
                   1998 Winter
                                   1998
                                          Winter
                                                            Nagano
                                                                          Bobsleigh
Out[16]: 271115
                   2002 Winter
                                   2002
                                                                          Bobsleigh
                                         Winter Salt Lake City
                                                           Name
                             Name Sex
                                                   Team
                                                                                              City
                                                                                                      5
                                                            Event Medias
                                                                          Year
                                                                                Season
                                   Basketball Men's Basketball
                                                                     NaN
                                  Judo Men's Extra-Lightweight
          1
                                                                     NaN
                                                                     1892
          2
                                    MFootball Menina Football
                                                                           1992 Summer
                                                                                         Barcelona Bask
                                                                  Summer
Gold
                                   Tug-Of-War Men's Tug-Of-War
           3
                             Speed Skating Women's 500 metres
          4
                                                                     NaN
                                                                    2012
                        2 A Lamusi Luge Mixed (Mehinas Doch) essum Mah
                                                                          2012 Summer
                                                                                           London
           271111
           271112
                   Ski Jumping Men's Large Hill, Individual
                                                                     NaN
                           Ski Jumping Men's Large Hill, Team
           271113
                                                                     NaN
           27111<del>4</del>
                                           Bobs leangharMen's DEQUE
                                                                     1192A
                                     М
                            Nielsen
                                                                          1920 Summer Antwerpen
                                                                                                     Fo
                                           Bobsleigh Men's Four Summer
           271115
                              Aabv
           [271116 rows x 12Edgailumns]>
                                                                     1900
                                                                                                     Tu
                                      M Denmark/Sweden
                                                            DEN
                                                                          1900 Summer
                                                                                             Paris
                        4 Lindenau
                                                                  Summer
                             Aabye
                            Christine
                                                                     1988
                                                                                                      S
                            Jacoba
                                      F
                                              Netherlands
                                                            NED
                                                                           1988
                                                                                  Winter
                                                                                           Calgary
                                                                    Winter
                             Aaftink
```

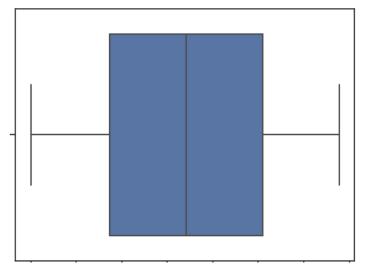
	~/ + + + 4	2017	MTILCEL	2017	MTILCEL	200	114 2	N J UI	שיי±איי		
In [16]:	271113	2014	Winter	2014	Winter	Soc	hi S	ki Jur	nping		
111 [10].	271114	1998	Winter	1998	Winter	Naga	no	Bobs	leigh		
Out[16]:	271115			2002		t Lake Ci		Bobs	•		
out[10].	2/1115	2002	WITTEET	2002	WINCEL SUI	Name	cy	0003.	LCISII		
		ID	Name	Sav	Tear		. Gamas	Voor	Season	City	٤
		טו				LVCII	t Mewses	rear	Season	City	•
	0				etball Men's						
	1		J	udo N	Men's Extra-L	ightweigh	t NaN				
	2 0	1	A Dijiang	MF	ootball M en in Of-War Men's	s Footchal	1, 1997	1992	Summer	Barcelona	Bask
	3			Tug-(Of-War Men's	Tug-Of-Wa	Summer Gold				
	4				ing Women's						
	-		эрееи	Skat	Tilg Wollien 3	Joo metre					
	1	2	A Lamusi	. М	Chin	a - CHN	2012	2012	Summer	London	
	27111 1				e Mixed (Mehin			2012	Odminion	London	
	271112	Ski J	lumping M	en's	Large Hill,	Individua	l NaN				
	271113		Ski Jum	ping	Men's Large	Hill, Tea	m NaN				
	27111 4	3	Curina	М	Bobs leeighar			1020	Summer	Antwerpen	For
	271115	3	Aaby	IVI	Bobsleigh			1920	Summer	Antwerpen	FUI
	2,1113		Аару		DODDICIBIL	nen 5 rou	·				
	F07444	_	40-1-3	-							
			x 12Edgail				1900		_		Tu
	3	4	Lindenau	IVI	Denmark/Swede	n DEN	Summer	1900	Summer	Paris	
			Aabye								
			Christine								
	4	5		F	Netherland	s NED	1988	1988	Winter	Calgary	S
		·	Aaftink	•	110110110110		Winter	.000	********	ou.gu. y	Sk
			,								
										•••	
			Andrzej				1976				
	271111	135569	yá	M	Poland-	1 POL	Winter	1976	Winter	Innsbruck	
			,								
							2014				
	271112	135570	Piotr ya	M	Polan	d POL	Winter	2014	Winter	Sochi	Jun
							viiitoi				oun
	274442	135570	Diotr vo		Polan	d POL	2014	2014	Winter	Sochi	
	2/1113	133370	Piotr ya	М	Polan	a POL	Winter	2014	vviriter	Sociii	Jun
			Tomasz				4000				
	271114	135571	Ireneusz	M	Polan	d POL	1998	1998	Winter	Nagano	Bobs
			ya				Winter			Ü	
			-								
			Tomasz		_		2002			Salt Lake	
	271115	135571	Ireneusz	M	Polan	d POL	Winter	2002	Winter	City	Bobs
			ya							City	
	271116	rows × 1	12 columns	3							
	27 1110		ooiaiiiii	•							

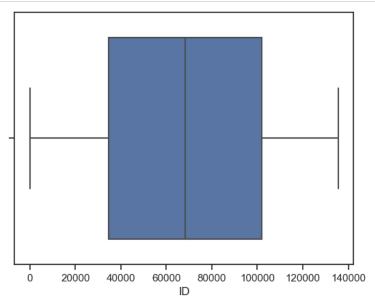


```
In [17]: sns.heatmap(df.isnull(),yticklabels=False,cmap='viridis')
Out[17]: <Axes: >
```



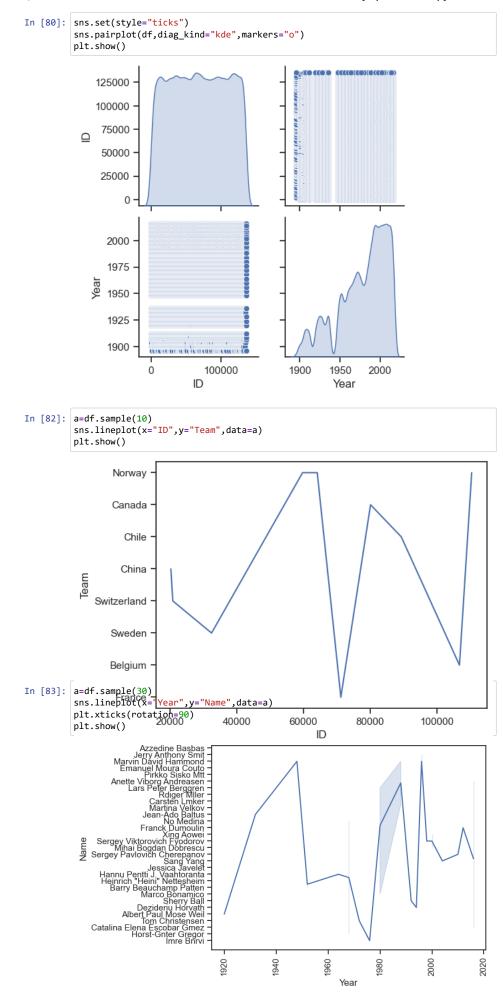


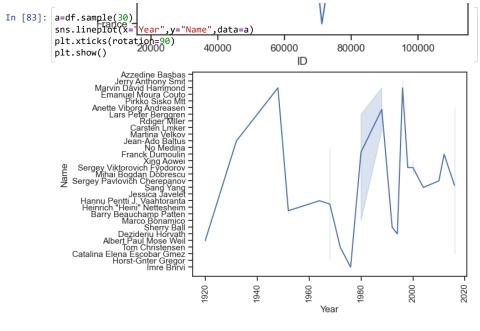




```
TypeError
                                        Traceback (most recent call last)
Cell In[87], line 2
     1 for i in df.columns:
----> 2
           sns.boxplot(x=df[i])
            plt.show()
File ~\anaconda3\Lib\site-packages\seaborn\categorical.py:2231, in boxplot(dat
a, x, y, hue, order, hue_order, orient, color, palette, saturation, width, dodg
e, fliersize, linewidth, whis, ax, **kwargs)
   2224 def boxplot(
   2225
            data=None, *, x=None, y=None, hue=None, order=None, hue_order=None,
            orient=None, color=None, palette=None, saturation=.75, width=.8,
   2226
            dodge=True, fliersize=5, linewidth=None, whis=1.5, ax=None,
   2227
   2228
            **kwargs
   2229 ):
-> 2231
            plotter = _BoxPlotter(x, y, hue, data, order, hue_order,
   2232
                                  orient, color, palette, saturation,
   2233
                                  width, dodge, fliersize, linewidth)
   2235
            if ax is None:
   2236
                ax = plt.gca()
File ~\anaconda3\Lib\site-packages\seaborn\categorical.py:785, in BoxPlotter.
_init__(self, x, y, hue, data, order, hue_order, orient, color, palette, satura
tion, width, dodge, fliersize, linewidth)
    781 def __init__(self, x, y, hue, data, order, hue_order,
```

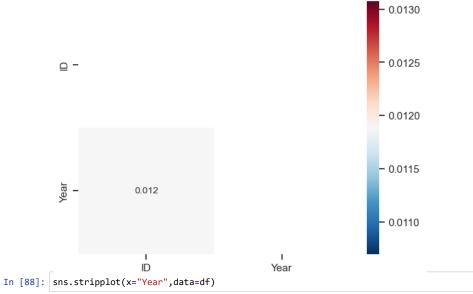
```
TypeError
                                                    Traceback (most recent call last)
         Cell In[87], line 2
               1 for i in df.columns:
                     sns.boxplot(x=df[i])
                     plt.show()
               3
         File ~\anaconda3\Lib\site-packages\seaborn\categorical.py:2231, in boxplot(dat
         a, x, y, hue, order, hue_order, orient, color, palette, saturation, width, dodg
         e, fliersize, linewidth, whis, ax, **kwargs)
            2224 def boxplot(
                      data=None, *, x=None, y=None, hue=None, order=None, hue_order=None,
            2225
            2226
                     orient=None, color=None, palette=None, saturation=.75, width=.8,
            2227
                      dodge=True, fliersize=5, linewidth=None, whis=1.5, ax=None,
            2228
            2229 ):
                      plotter = _BoxPlotter(x, y, hue, data, order, hue_order,
          -> 2231
            2232
                                            orient, color, palette, saturation,
                                            width, dodge, fliersize, linewidth)
            2233
            2235
                     if ax is None:
            2236
                          ax = plt.gca()
         File ~\anaconda3\Lib\site-packages\seaborn\categorical.py:785, in _BoxPlotter.
          _init__(self, x, y, hue, data, order, hue_order, orient, color, palette, satura
         tion, width, dodge, fliersize, linewidth)
             781 def __init__(self, x, y, hue, data, order, hue_order,
             782
                               orient, color, palette, saturation,
             783
                               width, dodge, fliersize, linewidth):
          --> 785
                      {\tt self.establish\_variables}({\tt x},\ {\tt y},\ {\tt hue},\ {\tt data},\ {\tt orient},\ {\tt order},\ {\tt hue\_order})
             786
                     self.establish_colors(color, palette, saturation)
             788
                     self.dodge = dodge
         File ~\anaconda3\Lib\site-packages\seaborn\categorical.py:544, in _CategoricalP
         lotter.establish_variables(self, x, y, hue, data, orient, order, hue_order, uni
                         raise ValueError(err)
             543 # Figure out the plotting orientation
          --> 544 orient = infer_orient(
                     x, y, orient, require_numeric=self.require_numeric
             546 )
             548 # Option 2a:
             549 # We are plotting a single set of data
             550 # ---
             551 if x is None or y is None:
             552
                     # Determine where the data are
         File ~\anaconda3\Lib\site-packages\seaborn\_oldcore.py:1601, in infer_orient(x,
         y, orient, require_numeric)
                         warnings.warn(single_var_warning.format("Vertical", "x"))
            1599
            1600
                      if require_numeric and x_type != "numeric":
          -> 1601
                          raise TypeError(nonnumeric_dv_error.format("Horizontal", "x"))
            1602
                      return "h"
            1604 elif str(orient).startswith("v"):
         TypeError: Horizontal orientation requires numeric `x` variable.
In [80]: sns.set(style="ticks")
         sns.pairplot(df,diag_kind="kde",markers="o")
         plt.show()
                                                     125000
              100000
               75000
               50000
               25000
                   0
                2000
                1975
                1950
                400E
```



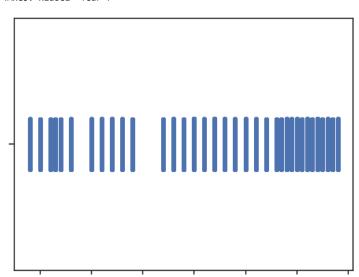




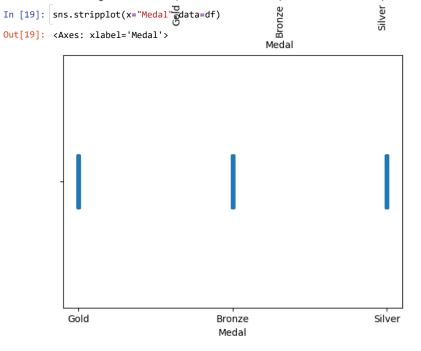


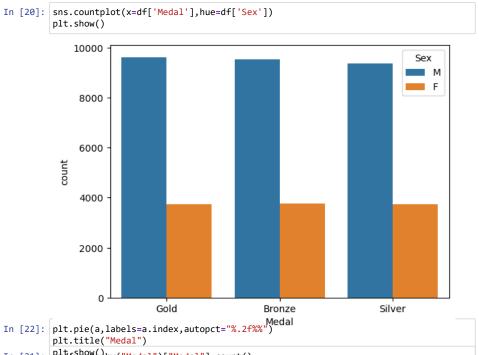


Out[88]: <Axes: xlabel='Year'>

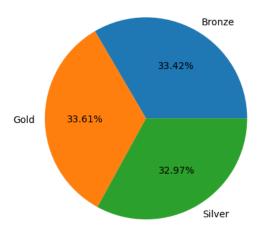


```
ıcaı
In [88]: sns.stripplot(x="Year",data=df)
Out[88]: <Axes: xlabel='Year'>
            1900
                      1920
                                                1980
                                                         2000
                                                                  2020
                               1940
                                       1960
                                      Year
In [18]: sns.histplot(df["Medal"],bins=10,kde=True)
        plt.xticks(rotation=90)
        plt.show()
            50000
            40000
         Count 30000
            20000
            10000
                                              Bronze
                                                                Silver
In [19]: sns.stripplot(x="Medal" data=df)
Out[19]: <Axes: xlabel='Medal'>
                                             Medal
```





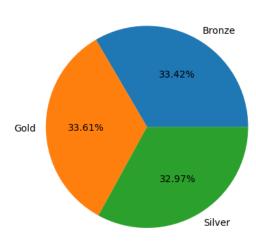




```
In [22]: plt.pie(a,labels=a.index,autopct="%.2f%%")

In [21]: Bldf:BP%upby("Medal")["Medal"].count()
```

Medal



```
In [23]: from sklearn.preprocessing import OneHotEncoder,StandardScaler
                                       categorical_cols=["Year","City"]
                                      encoder=OneHotEncoder(drop='first', sparse=False)
                                       encoder_cols=pd.DataFrame(encoder.fit_transform(df[categorical_cols]),columns=encoder_cols=pd.DataFrame(encoder.fit_transform(df[categorical_cols]),columns=encoder_cols=pd.DataFrame(encoder.fit_transform(df[categorical_cols])),columns=encoder_cols=pd.DataFrame(encoder.fit_transform(df[categorical_cols])),columns=encoder_cols=pd.DataFrame(encoder.fit_transform(df[categorical_cols])),columns=encoder_cols=pd.DataFrame(encoder.fit_transform(df[categorical_cols])),columns=encoder_cols=pd.DataFrame(encoder.fit_transform(df[categorical_cols])),columns=encoder_cols=pd.DataFrame(encoder.fit_transform(df[categorical_cols])),columns=encoder_cols=pd.DataFrame(encoder.fit_transform(df[categorical_cols])),columns=encoder_cols=pd.DataFrame(encoder.fit_transform(df[categorical_cols])),columns=encoder_cols=pd.DataFrame(encoder.fit_transform(df[categorical_cols])),columns=encoder_cols=pd.DataFrame(encoder.fit_transform(df[categorical_cols])),columns=encoder_cols=pd.DataFrame(encoder.fit_transform(df[categorical_cols])),columns=encoder_cols=pd.DataFrame(encoder.fit_transform(df[categorical_cols=pd.DataFrame(encoder.fit_transform(df[categorical_cols=pd.DataFrame(encoder.fit_transform(df[categorical_cols=pd.DataFrame(encoder.fit_transform(df[categorical_cols=pd.DataFrame(encoder.fit_transform(df[categorical_cols=pd.DataFrame(encoder.fit_transform(df[categorical_cols=pd.DataFrame(encoder.fit_transform(df[categorical_cols=pd.DataFrame(encoder.fit_transform(df[categorical_cols=pd.DataFrame(encoder.fit_transform(df[categorical_cols=pd.DataFrame(encoder.fit_transform(df[categorical_cols=pd.DataFrame(encoder.fit_transform(df[categorical_cols=pd.DataFrame(encoder.fit_transform(df[categorical_cols=pd.DataFrame(encoder.fit_transform(df[categorical_cols=pd.DataFrame(encoder.fit_transform(df[categorical_cols=pd.DataFrame(encoder.fit_transform(df[categorical_cols=pd.DataFrame(encoder.fit_transform(df[categorical_cols=pd.DataFrame(encoder.fit_transform(df[categorical_cols=pd.DataFrame(encoder.fit_transform(df[categorical_cols=pd.DataFrame(encod
    In [24]: numerical_cols=["ID","Year"]
                                       scaler=StandardScaler()
                                       scaled\_cols=pd.DataFrame(scaler.fit\_transform(df[numerical\_cols]), columns=scaler.fit\_transform(df[numerical\_cols]), columns=scaler.fit\_transfor
In [113]: x=pd.concat([encoder_cols,scaled_cols],axis=1)
                                      y=df['Sex']
In [114]: x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=42)
In [115]: log=LogisticRegression()
In [116]: log.fit(x_train,y_train)
Out[116]:
                                         ▼ LogisticRegression
                                         LogisticRegression()
In [117]: |print('Train Score:',log.score(x_train,y_train))
                                       Train Score: 0.7248031278239861
In [121]: print(metrics.classification_report(y_test,pred_test))
Test Score: 0.7264311006196518
                                                                                                                                                0.00
                                                                                                                                                                                     0.00
                                                                                                                                                                                                                      14834
                                                                                                            0.00
In [119]: pred_{train} = \log.pred_{ct}^{0.73} (x_{train})^{00}
                                                                                                                                                                                     0.84
                                                                                                                                                                                                                      39390
                                      pred_test=log.predict(x_test)
                                                                                                                                                                                     0.73
                                                                                                                                                                                                                      54224
                                                 macro avg
                                                                                                           0.36
                                                                                                                                               0.50
                                                                                                                                                                                   0.42
                                                                                                                                                                                                                      54224
In [120]: weightmedtaigs.class@f56ation_@epdrt(y_t0a6m,pred54224n))
                                                                                                                                         recall f1-score
                                                                                         precision
                                                                                                                                                                                                              support
In [122]: from sklearp.metrica_import matabews_cogragef, assugacy_score, precision_score,
                                                                                                            0.72
                                                                                                                                                1.00
                                                                                                                                                                                     0.84
                                                                                                                                                                                                                  157204
accuracy
                                                                                                                                                0.50
                                                 macro avg
                                                                                                            0.36
                                                                                                                                                                                     0.42
                                                                                                                                                                                                                  216892
                                       ₩βåghŧβø,a¥Breshold0<sub>x</sub>5metrics9r32_curve(y6train2b6882y[:, 0],roc)
                                     plt.plot([0, 1], [0, 1], "k--")
plt.plot(fpr, tpr, label='logistic')
                                      plt.show()
                                            1.0
```

```
In [121]: print(metrics.classification_report(y_test,pred_test))
Test Score: 0.7264311006196518
                                       0.00
                                                 0.00
                                                          14834
                             0.00
In [119]: pred_train=Nog.predict(x_train)^{00}
                                                          39390
                                                 0.84
          pred_test=log.predict(x_test)
                                                 0.73
             macro avg
                             0.36
                                       0.50
                                                 0.42
                                                          54224
In [120]: weightmdtaigs.class@f56ation_@epdrt(y_t0a6m,pred54224n))
                                     recall f1-score
                        precision
                                                        support
In [122]: from sklearp.metrica.jmport matthews_cogregef, assyggcy_score, precision_score,
                             0.72
                                       1.00
                                                 0.84
                                                         157204
0.36
                                       0.50
                                                 0.42
                                                         216892
             macro avg
          ₩βåghŧβø,a¥Breshold0<sub>x</sub>5@etrics0r32_curve0y64rain286882y[:, 0],roc)
          plt.plot([0, 1], [0, 1], "k--")
plt.plot(fpr, tpr, label='logistic')
          plt.show()
           1.0
           0.8
           0.6
           0.4
           0.2
           0.0
                  0.0
                             0.2
                                         0.4
                                                     0.6
                                                                0.8
                                                                            1.0
In [124]: metrics.roc_auc_score(y_train,roc)
Out[124]: 0.6901432210989167
In [125]: from sklearn.metrics import matthews_corrcoef
          mcc=matthews_corrcoef(y_test,pred_test)
          print('MCC:',mcc)
 In [57]: Mest_param = grid.best_params_
          best_model = grid.best_estimator_
In [126]: param_grid={
               'penalty':['l1','l2'],
          y_pred'=[Bestomodel:predict(x_test)
 In [58]:
          from sklearn.metrics import accuracy_score , precision_score,recall_score,f1_score
from sklearn.model_selection import GridSearchCV
In [ ]:
In [128]:
          grid=GridSearchCV(estimator=log, param grid=param grid, cv= 5)
In [129]: grid.fit(x_train,y_train)
Out[129]:
                      GridSearchCV
            ▶ estimator: LogisticRegression
                 ▶ LogisticRegression
```