



```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
df=pd.read_csv("/content/hockey_teams.csv") # load the data
df.shape
```



 (582, 9)

```
df.info()
```


 <class 'pandas.core.frame.DataFrame'>
 RangeIndex: 582 entries, 0 to 581
 Data columns (total 9 columns):
 # Column Non-Null Count Dtype

 0 team_name 582 non-null object
 1 year 582 non-null int64
 2 wins 582 non-null int64
 3 losses 582 non-null int64
 4 ot_losses 358 non-null float64
 5 win_pct 582 non-null float64
 6 goals_for 582 non-null int64
 7 goals_against 582 non-null int64
 8 goal_difference 582 non-null int64
 dtypes: float64(2), int64(6), object(1)
 memory usage: 41.1+ KB

```
df.head() #head of the data
```

	team_name	year	wins	losses	ot_losses	win_pct	goals_for	goals_against	goal_difference
0	Boston Bruins	1990	44	24	NaN	0.550	299	264	35
1	Buffalo Sabres	1990	31	30	NaN	0.388	292	278	14
2	Calgary Flames	1990	46	26	NaN	0.575	344	263	81
3	Chicago Blackhawks	1990	49	23	NaN	0.613	284	211	73
4	Detroit Red Wings	1990	34	38	NaN	0.425	273	298	-25



Next steps: [View recommended plots](#) [New interactive sheet](#)

```
df.duplicated() # checking the duplication in the dataset
```

	0
0	False
1	False
2	False
3	False
4	False
...	...
577	False
578	False
579	False
580	False
581	False

582 rows × 1 columns

dtype: bool

```
df.isnull().sum()
```



	0
team_name	0
year	0
wins	0
losses	0
ot_losses	224
win_pct	0
goals_for	0
goals_against	0
goal_difference	0

dtype: int64

```
df.drop("ot_losses",axis=1) # removeing the column
```



	team_name	year	wins	losses	win_pct	goals_for	goals_against	goal_difference
0	Boston Bruins	1990	44	24	0.550	299	264	35
1	Buffalo Sabres	1990	31	30	0.388	292	278	14
2	Calgary Flames	1990	46	26	0.575	344	263	81
3	Chicago Blackhawks	1990	49	23	0.613	284	211	73
4	Detroit Red Wings	1990	34	38	0.425	273	298	-25
...
577	Tampa Bay Lightning	2011	38	36	0.463	235	281	-46
578	Toronto Maple Leafs	2011	35	37	0.427	231	264	-33
579	Vancouver Canucks	2011	51	22	0.622	249	198	51
580	Washington Capitals	2011	42	32	0.512	222	230	-8
581	Winnipeg Jets	2011	37	35	0.451	225	246	-21

582 rows × 8 columns

```
fd=df.groupby(['team_name']).nunique()
fd
```



	year	wins	losses	ot_losses	win_pct	goals_for	goals_against	goal_difference	
team_name									
Anaheim Ducks	6	5	6	6	5	6	6	6	
Atlanta Thrashers	11	9	11	8	9	11	11	11	
Boston Bruins	21	16	13	10	16	21	21	20	
Buffalo Sabres	21	16	14	9	17	19	20	19	
Calgary Flames	21	14	11	8	15	20	19	21	
Carolina Hurricanes	14	11	11	9	11	13	14	14	
Chicago Blackhawks	21	18	19	9	17	21	21	21	
Colorado Avalanche	16	12	12	7	12	15	13	15	
Columbus Blue Jackets	11	9	9	9	9	11	11	9	
Dallas Stars	18	14	13	7	14	15	18	17	
Detroit Red Wings	21	14	16	8	15	20	19	18	
Edmonton Oilers	21	13	13	10	14	21	19	20	
Florida Panthers	18	12	14	8	12	17	17	16	
Hartford Whalers	7	6	6	0	7	7	7	7	
Los Angeles Kings	21	13	14	9	16	19	20	20	
Mighty Ducks of Anaheim	12	10	9	5	10	12	12	12	
Minnesota North Stars	3	3	3	0	3	3	3	3	
Minnesota Wild	11	10	7	7	10	10	11	9	
Montreal Canadiens	21	15	13	8	16	19	17	20	
Nashville Predators	13	11	12	8	11	13	13	13	
New Jersey Devils	21	13	13	7	15	20	18	17	
New York Islanders	21	14	16	9	16	18	19	21	
New York Rangers	21	16	14	11	17	20	16	20	
Ottawa Senators	19	14	15	9	14	19	19	19	
Philadelphia Flyers	21	14	14	6	14	19	21	21	
Phoenix Coyotes	15	10	9	6	10	13	15	14	
Pittsburgh Penguins	21	17	16	9	16	21	21	18	
Quebec Nordiques	5	5	5	0	5	5	5	4	
San Jose Sharks	20	17	17	8	17	17	19	19	
St. Louis Blues	21	16	14	9	18	21	20	21	
Tampa Bay Lightning	19	14	12	9	15	17	18	18	
Toronto Maple Leafs	21	12	17	8	14	20	20	16	
Vancouver Canucks	21	17	16	7	18	21	19	20	
Washington Capitals	21	18	17	9	19	20	19	20	
Winnipeg Jets	7	7	7	1	7	7	7	7	



Next steps:

View recommended plots

New interactive sheet

```
fg=df.groupby(['team_name']).agg({'wins':'sum'})
fg
```



wins



team_name



Anaheim Ducks	257
Atlanta Thrashers	342
Boston Bruins	816
Buffalo Sabres	803
Calgary Flames	764
Carolina Hurricanes	515
Chicago Blackhawks	765
Colorado Avalanche	677
Columbus Blue Jackets	342
Dallas Stars	752
Detroit Red Wings	986
Edmonton Oilers	692
Florida Panthers	573
Hartford Whalers	195
Los Angeles Kings	730
Mighty Ducks of Anaheim	381
Minnesota North Stars	95
Minnesota Wild	405
Montreal Canadiens	783
Nashville Predators	503
New Jersey Devils	905
New York Islanders	650
New York Rangers	792
Ottawa Senators	679
Philadelphia Flyers	836
Phoenix Coyotes	557
Pittsburgh Penguins	839
Quebec Nordiques	147
San Jose Sharks	721
St. Louis Blues	811
Tampa Bay Lightning	588
Toronto Maple Leafs	767
Vancouver Canucks	815
Washington Capitals	806
Winnipeg Jets	212

Next steps:

[View recommended plots](#)[New interactive sheet](#)

```
# Filter DataFrame for a single team
team_df = df[df['team_name'] == 'Anaheim Ducks']
```

```
# Create pivot table for this team
ty = pd.pivot_table(
    team_df,
    index='team_name',
    columns=['year', 'wins'],
    values='losses')
```

```
print(ty)
```

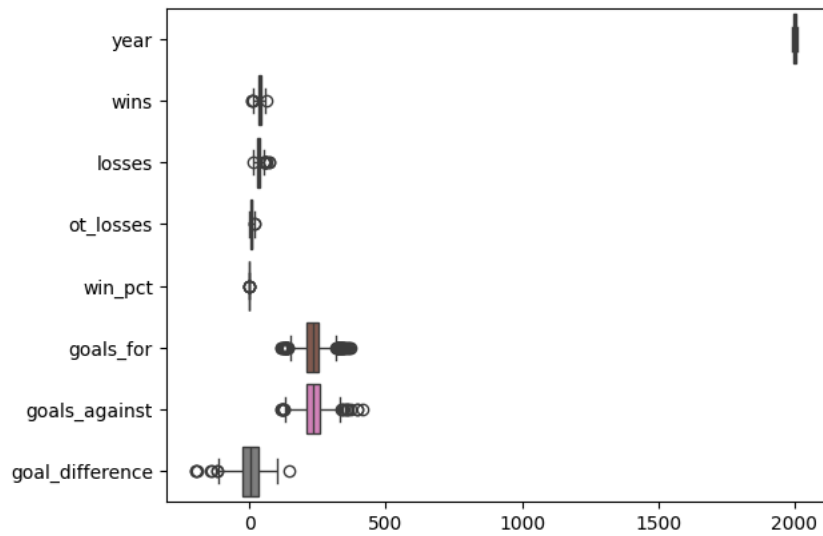


year	2006	2007	2008	2009	2010	2011
wins	48	47	42	39	47	34
team_name						
Anaheim Ducks	20.0	27.0	33.0	32.0	30.0	36.0

✓ Data Visualization

```
# checking for outliers
sns.boxplot(df,orient='h')
```

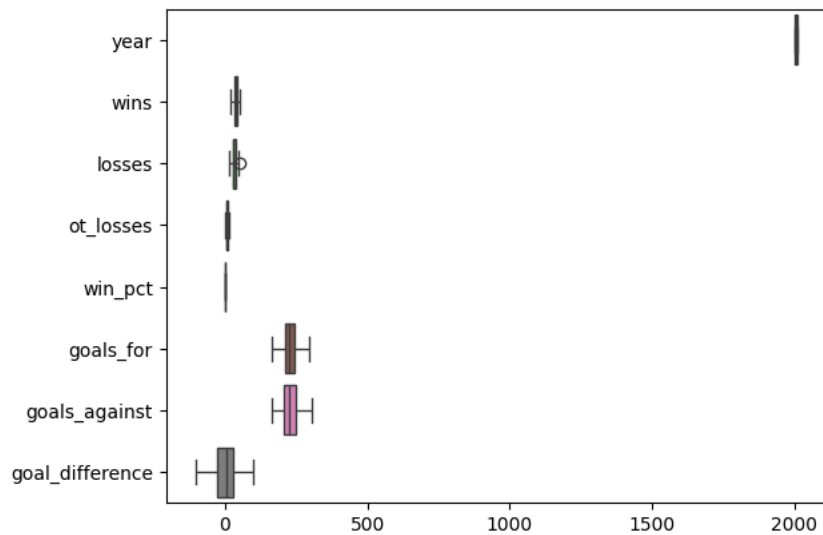
<Axes: >



```
for col in df.select_dtypes(include=['float', 'int']).columns:
    Q1 = df[col].quantile(0.25)
    Q3 = df[col].quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
    df = df[(df[col] >= lower_bound) & (df[col] <= upper_bound)]
```

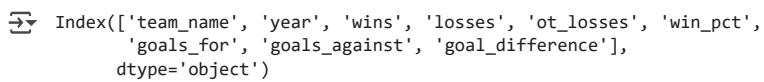
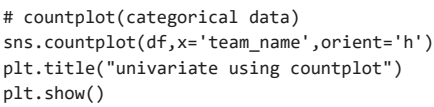
```
sns.boxplot(df,orient='h')
```

<Axes: >



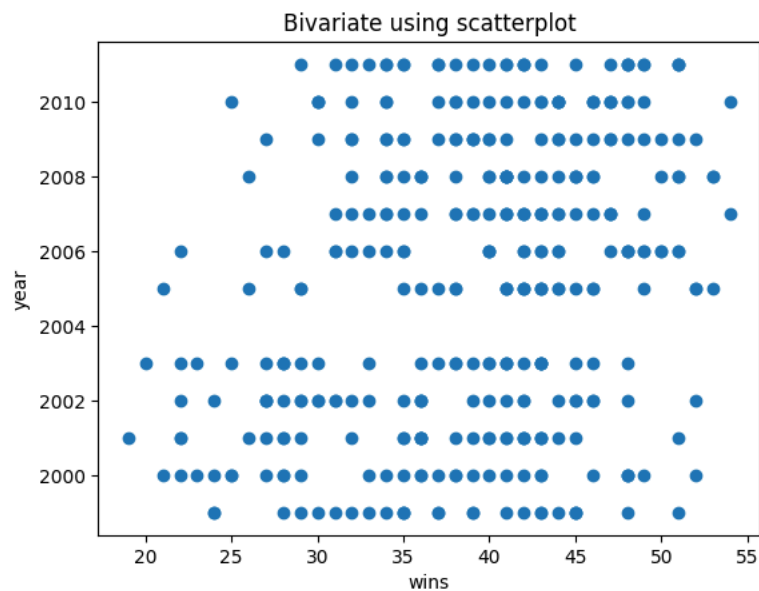
✓ Univariate Analysis

```
# Histogram(Numerical data)
plt.hist(df['wins'],bins=8,color='pink')
plt.title("univariate using histogram")
plt.show()
```



```
# num v/s num
plt.scatter(df['wins'],df['year'])
plt.title("Bivariate using scatterplot")
plt.xlabel("wins")
plt.ylabel("year")
```

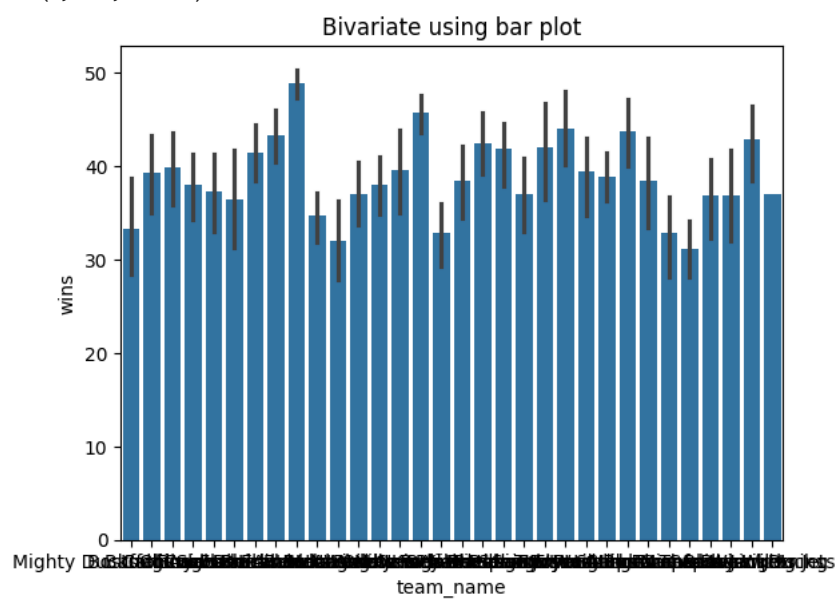
↻ Text(0, 0.5, 'year')



```
# cat v/s cat
# there is no two categorical
```

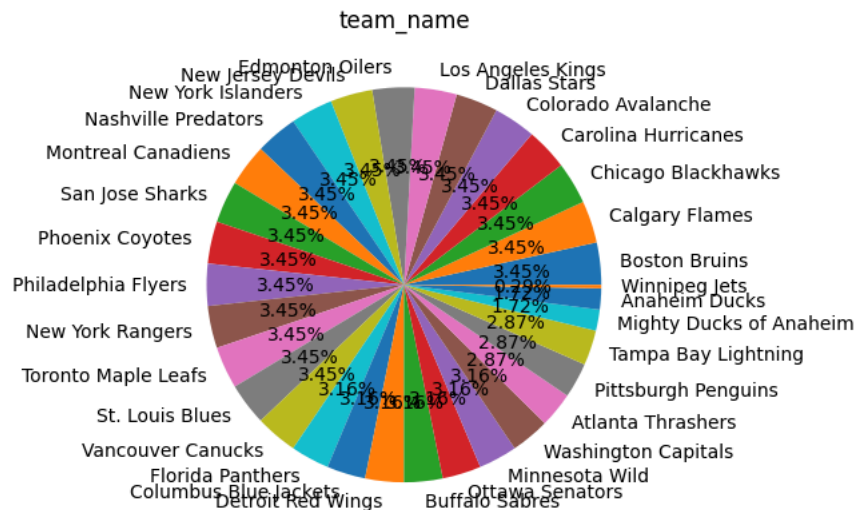
```
# num v/s cat
sns.barplot(df,x='team_name',y='wins')
plt.title("Bivariate using bar plot")
plt.xlabel("team_name")
plt.ylabel("wins")
```

↻ Text(0, 0.5, 'wins')



```
# pie chart
dependents=df['team_name'].value_counts()
plt.pie(dependents,labels=dependents.index,autopct='%0.2f%%')
plt.title("team_name")
```

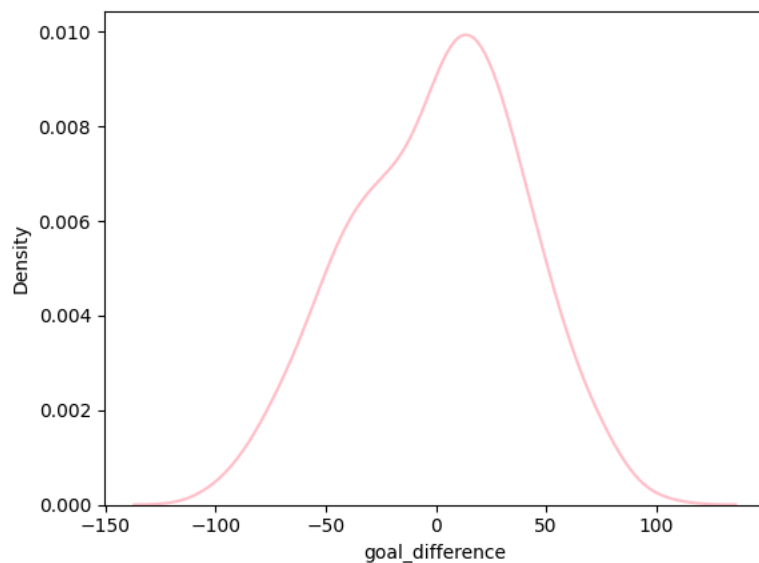
```
Text(0.5, 1.0, 'team_name')
```



✓ Data Distribution

```
sns.kdeplot(df['goal_difference'],color='pink')
print(df['goal_difference'].skew())
print(df['goal_difference'].kurt())
```

```
-0.18704240365207966
-0.47609407750784927
```



```
from scipy import stats
stats.probplot(df['goal_difference'],dist=stats.norm,plot=plt)
```