World Cup 2023 Analysis Project

Made by Ahmad Ali Alsheikh Ahmad

Import libraries

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
```

```
warnings.filterwarnings('ignore')
```

Read data

```
data=pd.read_csv("CWC23_all_innings.csv")
data
```

	team	player	bat_or_bowl	bb_bf	runs	wkts	wicketball_prob	runs_per_ball	opposition	ground	start_date	overs	mdns	econ	inns	4s	6s
0	PAK	Shaheen Shah Afridi (PAK)	bowl	60	45	3.0	0.05	0.750000	v South Africa	Chennai	27-Oct-23	10.0	0.0	4.50	2	NaN	NaN
1	ENG	DJ Willey (ENG)	bowl	60	45	3.0	0.05	0.750000	v India	Lucknow	29-Oct-23	10.0	2.0	4.50	1	NaN	NaN
2	NZ	MJ Henry (NZ)	bowl	60	48	3.0	0.05	0.800000	v England	Ahmedabad	5-Oct-23	10.0	1.0	4.80	1	NaN	NaN
3	NZ	LH Ferguson (NZ)	bowl	60	49	3.0	0.05	0.816667	v Bangladesh	Chennai	13-Oct-23	10.0	0.0	4.90	1	NaN	NaN
4	AFG	Noor Ahmad (AFG)	bowl	60	49	3.0	0.05	0.816667	v Pakistan	Chennai	23-Oct-23	10.0	0.0	4.90	1	NaN	NaN

```
'''
team - team name
player - player name
bat_or_bowl - Whether the player batted (bat) or bowled (bowl) in the match
bb - Bowling figures for bowlers (overs bowled-maidens-runs conceded-wickets taken)
bf - Batting figures for batsmen (balls faced-runs scored-not out)
runs - Total runs scored by the batsman (for batting entries only)
wkts - Number of wickets taken by the bowler (for bowling entries only)
wicket - Wicket number at which the batsman was dismissed (for batting entries only)
ball - Balls bowled per over by the bowler (may not be present)
probruns_per_bal - Average runs conceded per ball by the bowler (may not be present)
opposition - Opposing team the player faced
ground - Cricket ground where the match took place
start_date - Date the match started
overs - Overs bowled by the bowling team (may not be present)
mdn - Number of maiden overs bowled by the bowler (overs bowled without conceding any runs)
seconinns - Second innings score (may not be present)
4s - Number of boundaries hit by the batsman (fours)
6s - Number of sixes hit by the batsman (sixes)
sr - Strike Rate (runs scored per 100 balls faced) for the batsman (may not be present)
not_out - Indicates if the batsman remained not out (for batting entries only)
mins - Minutes


'''
```

Exploratory Data Analysis (EDA)

Viewing the Data

```
data.shape
```

```
(1408, 20)
```

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1408 entries, 0 to 1407
Data columns (total 20 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   team             1408 non-null   object 
 1   player           1408 non-null   object 
 2   bat_or_bowl      1408 non-null   object 
 3   bb_bf            1408 non-null   int64  
 4   runs             1408 non-null   int64  
 5   wkts             562 non-null    float64
 6   wicketball_prob  1408 non-null   float64
 7   runs_per_ball    1408 non-null   float64
 8   opposition       1408 non-null   object 
 9   ground           1408 non-null   object 
 10  start_date       1408 non-null   object 
 11  overs            562 non-null    float64
 12  mdns             562 non-null    float64
 13  econ             562 non-null    float64
 14  inns             1408 non-null   int64  
 15  4s               846 non-null    float64
 16  6s               846 non-null    float64
 17  sr               846 non-null    float64
 18  not out          846 non-null    float64
```

Checking for nulls

```
data.isnull().sum()
```

```
team                0
player              0
bat_or_bowl         0
bb_bf               0
runs                0
wkts              846
wicketball_prob     0
runs_per_ball       0
opposition          0
ground              0
start_date          0
overs             846
mdns              846
econ              846
inns                0
4s                562
6s                562
sr                562
not_out           562
mins              562
dtype: int64
```

Checking for any duplicates

```
data.duplicated().sum()
```

```
2
```

```
data[data.duplicated()]
```

	team	player	bat_or_bowl	bb_bf	runs	wkts	wicketball_prob	runs_per_ball	opposition	ground	start_date	overs	mdns	econ	inns	4s	6s	sr	no
1314	ENG	AAP Atkinson (ENG)	bat	1	0	NaN	1.0	0.0	v Pakistan	Eden Gardens	11-Nov-23	NaN	NaN	NaN	1	0.0	0.0	0.0	
1315	PAK	Abdullah Shafique (PAK)	bat	2	0	NaN	0.5	0.0	v England	Eden Gardens	11-Nov-23	NaN	NaN	NaN	2	0.0	0.0	0.0	

```
data = data.drop_duplicates()
data.duplicated().sum()
```

```
0
```

overview

```
'''
According to the cricket game it is normal to have missing values in
[wkts, overs, mdns, econ, 4s, 6s, sr, not_out, mins] because:

wickets (wkts): its only for bowl not for bat and he may have missing values

Overs:its the number of overs bowled by a player.
its only for bowl not for bat and he may have missing values

Maidens (mdns): its the number of times a bowler delivers six consecutive balls.
without conceding a run. its only for bowl not for bat and he may no score.

Economy Rate (econ): is the average number of runs conceded by a bowler per over bowled.
its only for bowl not for bat and he may no score may have missing values.

Fours (4s): its the number of times a batsman hits the ball and scores four runs in a cricket match.
its only for bat not for bowl and he may have missing values.

Sixes (6s): its the number of times a batsman hits the ball and scores six runs in a cricket match.
its only for bat not for bowl and he may have missing values.

Strike Rate (sr):its the number of runs scored by a batsman per 100 balls faced.
its only for bat not for bowl and he may have missing values.

Not Out (not_out): its for bat is he not out at the end of their innings (1) or not (0).
its only for bat not for bowl and he may have missing values.

Minutes (mins):its the duration of the player's innings in minutes.
its only for bat not for bowl and he may have missing values.
'''
```

Showing the values in each column and the count of them

```
for i in data:
    print(f"******************************|{ i }|************************************")
    dfv = data[i].value_counts()
    dfv = pd.DataFrame({"Count":data[i].value_counts()})
    display(dfv)
    print()
```

```
******************************|team|************************************
```

	Count
NED	157
AUS	155
ENG	150
NZ	144
BAN	141
SA	141
SL	138
AFG	129
PAK	127
IND	124

Describing all columns

```
for i in data:
    print(f"******************************|{ i }|************************************")
    df = pd.DataFrame(data[i].describe())
    display(df)
    print()
```

******************************|team|************************************

	team
count	1406
unique	10
top	NED
freq	157

******************************|player|************************************

	player
count	1406
unique	152
top	R Ravindra (NZ)
freq	19

******************************|bat_or_bowl|************************************

	bat_or_bowl
count	1406
unique	2
top	bat
freq	844

******************************|bb_bf|************************************

	bb_bf
count	1406.000000
mean	35.353485
std	25.234407
min	0.000000
25%	14.000000
50%	32.000000
75%	54.000000
max	143.000000

**********|runs|**********

	runs
count	1406.000000
mean	33.284495
std	28.048233
min	0.000000
25%	11.000000
50%	29.000000
75%	49.000000
max	201.000000

**********|wkts|**********

	wkts
count	562.000000
mean	1.204626
std	1.198237
min	0.000000
25%	0.000000
50%	1.000000
75%	2.000000

**********|runs|**********

	runs
count	1406.000000
mean	33.284495
std	28.048233
min	0.000000
25%	11.000000
50%	29.000000
75%	49.000000
max	201.000000

**********|wkts|**********

	wkts
count	562.000000
mean	1.204626
std	1.198237
min	0.000000
25%	0.000000
50%	1.000000
75%	2.000000

**********|wicketball_prob|**********

	wicketball_prob
count	1406.000000
mean	0.069607
std	0.150176
min	0.000000
25%	0.010667
50%	0.032258
75%	0.062181
max	1.000000

**********|runs_per_ball|**********

	runs_per_ball
count	1406.000000
mean	0.900034
std	0.470163
min	0.000000
25%	0.625000
50%	0.879655
75%	1.131657
max	6.000000

**********|wicketball_prob|**********

	wicketball_prob
count	1406.000000
mean	0.069607
std	0.150176
min	0.000000
25%	0.010667
50%	0.032258
75%	0.062181
max	1.000000

**********|runs_per_ball|**********

	runs_per_ball
count	1406.000000
mean	0.900034
std	0.470163
min	0.000000
25%	0.625000
50%	0.879655
75%	1.131657
max	6.000000

**********|opposition|**********

	opposition
count	1406
unique	10
top	v India
freq	166

**********|ground|**********

	ground
count	1406
unique	10
top	Dharamsala
freq	156

**********|start_date|**********

	start_date
count	1406
unique	41
top	28-Oct-23
freq	66

**********|opposition|**********

	opposition
count	1406
unique	10
top	v India
freq	166

**********|ground|**********

	ground
count	1406
unique	10
top	Dharamsala
freq	156

**********|start_date|**********

	start_date
count	1406
unique	41
top	28-Oct-23
freq	66

**********|overs|**********

	overs
count	562.000000
mean	7.342527
std	2.679736
min	0.300000
25%	5.550000
50%	8.000000
75%	10.000000
max	10.000000

**********|mdns|**********

	mdns
count	562.000000
mean	0.256228
std	0.532547
min	0.000000
25%	0.000000
50%	0.000000
75%	0.000000
max	3.000000

**********|4s|**********

	4s
count	844.000000
mean	2.611374
std	3.148091
min	0.000000
25%	0.000000
50%	2.000000
75%	4.000000
max	21.000000

**********|6s|**********

	6s
count	844.000000
mean	0.753555
std	1.505521
min	0.000000
25%	0.000000
50%	0.000000
75%	1.000000
max	11.000000

**********|sr|**********

	sr
count	844.000000
mean	83.914976
std	52.378798
min	0.000000
25%	51.902500
50%	81.930000
75%	107.140000
max	600.000000

**********|not_out|**********

	not_out
count	844.000000
mean	0.151659
std	0.358903
min	0.000000
25%	0.000000
50%	0.000000
75%	0.000000
max	1.000000

**********|mins|**********

	mins
count	844.000000
mean	42.818720
std	41.577626
min	1.000000
25%	12.000000
50%	28.000000
75%	60.250000
max	217.000000

Team Performance Analysis:

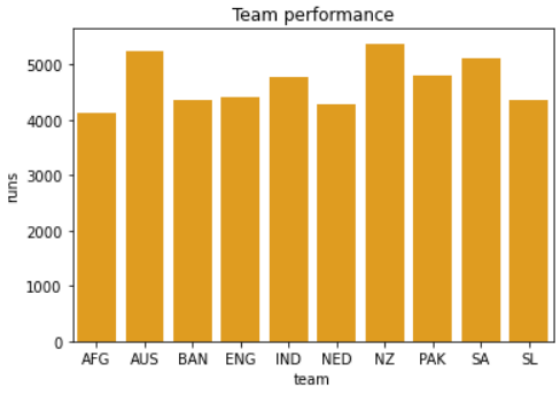
Group data

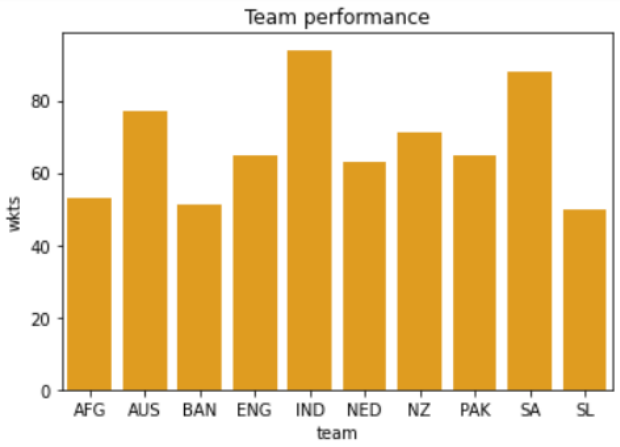
```
team_performance = data.groupby('team').agg({
    'runs': 'sum',
    'wkts': 'sum',
    'bat_or_bowl': lambda x: x.mode()
})


team_performance['avg_runs_per_match'] = team_performance['runs'] / data.groupby('team')['start_date'].nunique()
team_performance['avg_wkts_per_match'] = team_performance['wkts'] / data.groupby('team')['start_date'].nunique()
team_performance
```

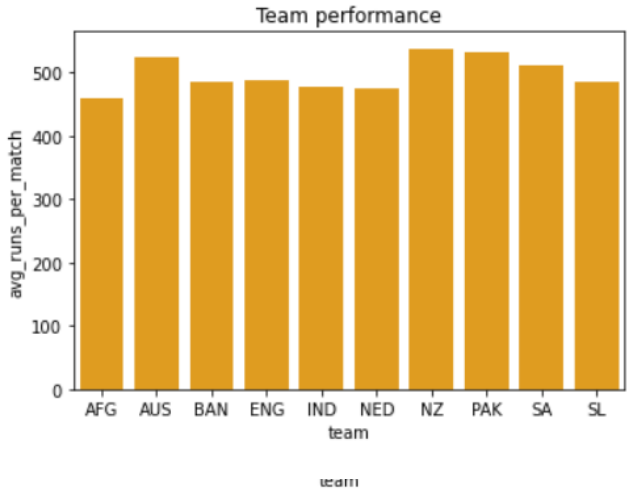
team	runs	wkts	bat_or_bowl	avg_runs_per_match	avg_wkts_per_match
AFG	4134	53.0	bat	459.333333	5.888889
AUS	5234	77.0	bat	523.400000	7.700000
BAN	4358	51.0	bat	484.222222	5.666667
ENG	4396	65.0	bat	488.444444	7.222222
IND	4783	94.0	bat	478.300000	9.400000
NED	4268	63.0	bat	474.222222	7.000000
NZ	5376	71.0	bat	537.600000	7.100000
PAK	4786	65.0	bat	531.777778	7.222222
SA	5097	88.0	bat	509.700000	8.800000
SL	4366	50.0	bat	485.111111	5.555556

Visualization

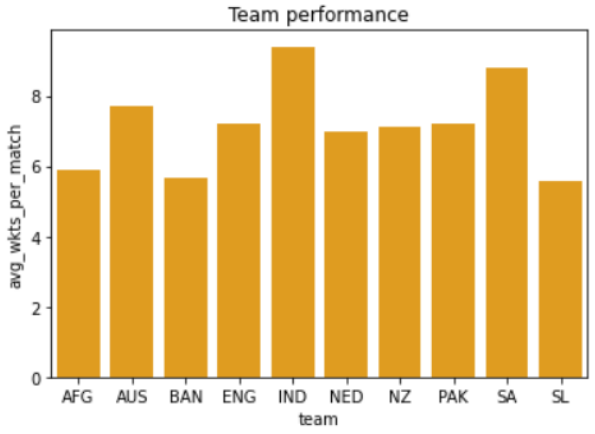
```
for i in team_performance:
    if i not in (['bat_or_bowl']):
        sns.barplot(x=team_performance.index, y=i, data=team_performance, color='orange')
        plt.title("Team performance")
        plt.show()
```



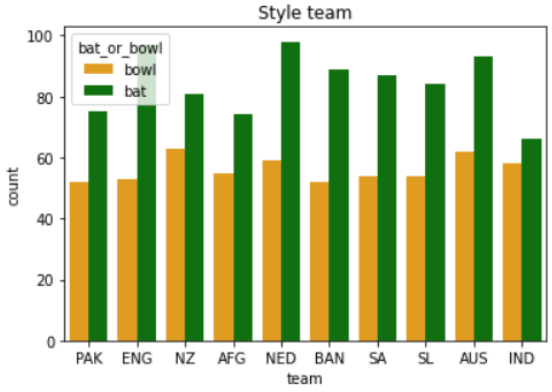




team




```
sns.countplot(x='team', data=data, hue='bat_or_bowl',palette={'bat': 'green', 'bowl': 'orange'})
plt.title("Style team")
plt.show()
```



Top-performing teams

```
print("the strength of team:")
print("-------------------------------------")
print("Total Runs:", team_performance['runs'].max(),"Team:", team_performance['runs'].idxmax())
print("Total Wickets:", team_performance['wkts'].max(),"Team:", team_performance['wkts'].idxmax())
print("Average Runs per Match:", team_performance['avg_runs_per_match'].max(),"Team:" ,team_performance['avg_runs_per_match'].idx
print("Average Wickets per Match:", team_performance['avg_wkts_per_match'].max(),"Team:", team_performance['avg_wkts_per_match'].
print("Team Style:")
print("-------------------------------------")
for team, style in data.groupby(['team', 'bat_or_bowl']).size().unstack().idxmax(axis=1).items():
    print(f"{team}: {style}")
```

```
the strength of team:
-------------------------------------
Total Runs: 5376 Team: NZ
Total Wickets: 94.0 Team: IND
Average Runs per Match: 537.6 Team: NZ
Average Wickets per Match: 9.4 Team: IND
Team Style:
-------------------------------------
AFG: bat
AUS: bat
BAN: bat
ENG: bat
IND: bat
NED: bat
NZ: bat
PAK: bat
SA: bat
SL: bat
```

Player Performance Analysis:

```
data['player'].nunique()
```

```
152
```

Group data

```
batting_data = data[data['bat_or_bowl'] == 'bat']
bowling_data = data[data['bat_or_bowl'] == 'bowl']
batting_players = batting_data.groupby('player').agg({
    'runs': 'sum',
    'bb_bf': 'sum',
    '4s': 'sum',
    '6s': 'sum',
    'sr': 'mean',
    'not_out': 'sum',
    'mins': 'mean'
})
bowling_players = bowling_data.groupby('player').agg({
    'runs': 'sum',
    'wkts': 'sum',
    'bb_bf': 'sum',
    'overs': 'sum',
    'mdns': 'sum',
    'econ': 'mean',
})
display(batting_players)
display(bowling_players)
```

player	runs	bb_bf	4s	6s	sr	not_out	mins
A Dutt (NED)	70	87	1.0	5.0	88.586667	3.0	13.777778
A Zampa (AUS)	48	60	5.0	0.0	75.230000	3.0	14.000000
AAP Atkinson (ENG)	37	23	7.0	0.0	122.220000	1.0	13.000000
AD Mathews (SL)	51	78	4.0	1.0	48.927500	0.0	38.250000
AK Markram (SA)	406	366	44.0	9.0	151.525000	1.0	51.000000
...	...	...	...	...	...	...	...
Usama Mir (PAK)	0	3	0.0	0.0	0.000000	0.0	4.000000
V Kohli (IND)	711	784	64.0	9.0	82.956000	3.0	109.700000
Vikramjit Singh (NED)	98	150	12.0	1.0	52.366667	0.0	33.166667
W Barresi (NED)	83	112	11.0	1.0	66.167500	0.0	34.250000
WA Young (NZ)	206	246	23.0	6.0	68.415000	0.0	57.333333

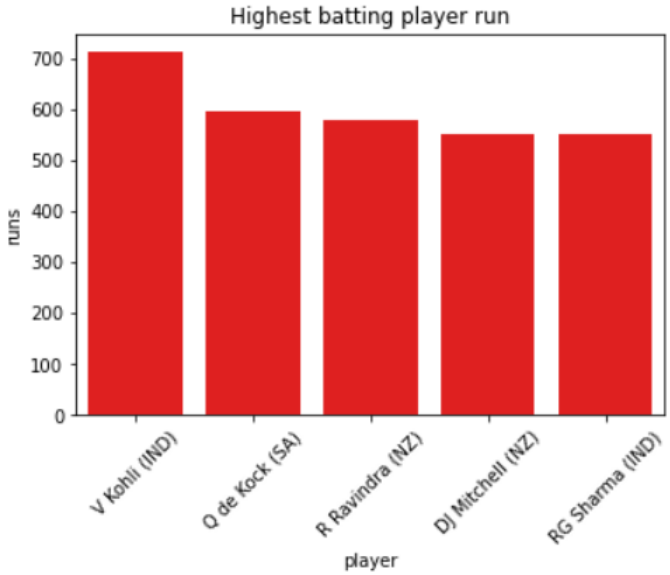
146 rows × 7 columns

player	runs	wkts	bb_bf	overs	mdns	econ
A Dutt (NED)	426	10.0	465	77.3	6.0	5.566667
A Zampa (AUS)	471	22.0	516	86.0	1.0	5.380000
AAP Atkinson (ENG)	146	4.0	144	24.0	0.0	6.043333
AD Mathews (SL)	107	6.0	133	22.1	2.0	4.918000
AK Markram (SA)	85	1.0	111	18.3	1.0	5.480000
...	...	...	...	...	...	...
Tanzim Hasan Sakib (BAN)	80	3.0	60	10.0	0.0	8.000000
Taskin Ahmed (BAN)	305	5.0	330	55.0	3.0	5.632857
Usama Mir (PAK)	248	4.0	210	35.0	0.0	7.050000
V Kohli (IND)	15	1.0	21	3.3	0.0	4.165000
Vikramjit Singh (NED)	52	0.0	42	7.0	0.0	7.916667

102 rows × 6 columns

Batting

```
bat=batting_players.sort_values(by='runs', ascending=False).head(5)
sns.barplot(x=bat.index,y='runs', data=bat, color='red')
plt.title("Highest batting player run")
plt.xticks(rotation=45)
plt.show()
```



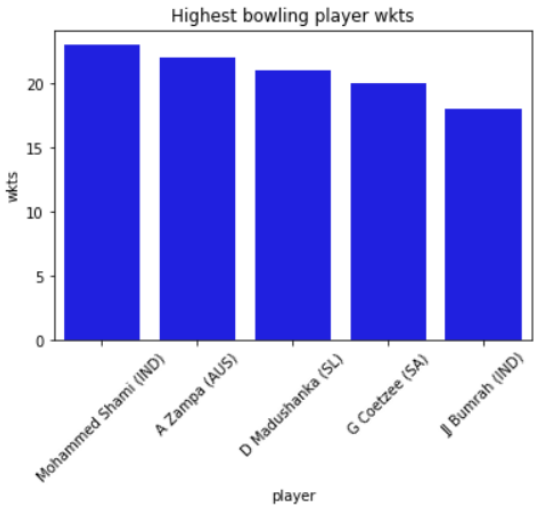
Impact Batting players

```
print("Best Performing Batting Players")
print("-------------------------------------")
print("Highest player runs:", batting_players['runs'].idxmax(),batting_players['runs'].max())
print("Highest player bb_bf:", batting_players['bb_bf'].idxmax(),batting_players['bb_bf'].max())
print("Highest player 4s:", batting_players['4s'].idxmax(),batting_players['4s'].max())
print("Highest player 6s:", batting_players['6s'].idxmax(),batting_players['6s'].max())
print("Highest player sr:", batting_players['sr'].idxmax(),batting_players['sr'].max())
print("Highest player not_out:", batting_players['not_out'].idxmax(),batting_players['not_out'].max())
print("Highest player mins:", batting_players['mins'].idxmax(),batting_players['mins'].max())
```

```
Best Performing Batting Players
-------------------------------------
Highest player runs: V Kohli (IND) 711
Highest player bb_bf: V Kohli (IND) 784
Highest player 4s: V Kohli (IND) 64.0
Highest player 6s: RG Sharma (IND) 28.0
Highest player sr: DJ Willey (ENG) 163.93333333333334
Highest player not_out: KL Rahul (IND) 4.0
Highest player mins: V Kohli (IND) 109.7
```

Bowling

```
bowl=bowling_players.sort_values(by='wkts', ascending=False).head(5)
sns.barplot(x=bowl.index,y='wkts', data=bowl, color='blue')
plt.title("Highest bowling player wkts")
plt.xticks(rotation=45)
plt.show()
```



Impact Bowling players

```
print("Best Performing Bowling Players")
print("-------------------------------------")
print("Highest player wkts:",bowling_players['wkts'].idxmax(),bowling_players['wkts'].max())
print("Highest player runs:",bowling_players['runs'].idxmax(),bowling_players['runs'].max())
print("Highest player bb_bf:",bowling_players['bb_bf'].idxmax(),bowling_players['bb_bf'].max())
print("Highest player overs:",bowling_players['overs'].idxmax(),bowling_players['overs'].max())
print("Highest player mdns:",bowling_players['mdns'].idxmax(),bowling_players['mdns'].max())
print("Highest player econ:",bowling_players['econ'].idxmax(),bowling_players['econ'].max())
```

```
Best Performing Bowling Players
-------------------------------------
Highest player wkts: Mohammed Shami (IND) 23.0
Highest player runs: Haris Rauf (PAK) 533
Highest player bb_bf: MJ Santner (NZ) 556
Highest player overs: MJ Santner (NZ) 92.4
Highest player mdns: JR Hazlewood (AUS) 8.0
Highest player econ: JDS Neesham (NZ) 12.18
```

Opposition and Ground Analysis:

Teams

Group data

```
team_opposition= data.groupby(['team','opposition']).agg({
    'runs': 'sum',
    'wkts': 'sum'
}).reset_index()
display(team_opposition)
```

	team	opposition	runs	wkts
0	AFG	v Australia	563	6.0
1	AFG	v Bangladesh	303	3.0
2	AFG	v England	480	10.0
3	AFG	v India	525	2.0
4	AFG	v Netherlands	346	6.0
...	...	...	...	...
86	SL	v India	391	6.0
87	SL	v Netherlands	495	9.0
88	SL	v New Zealand	339	4.0
89	SL	v Pakistan	675	4.0

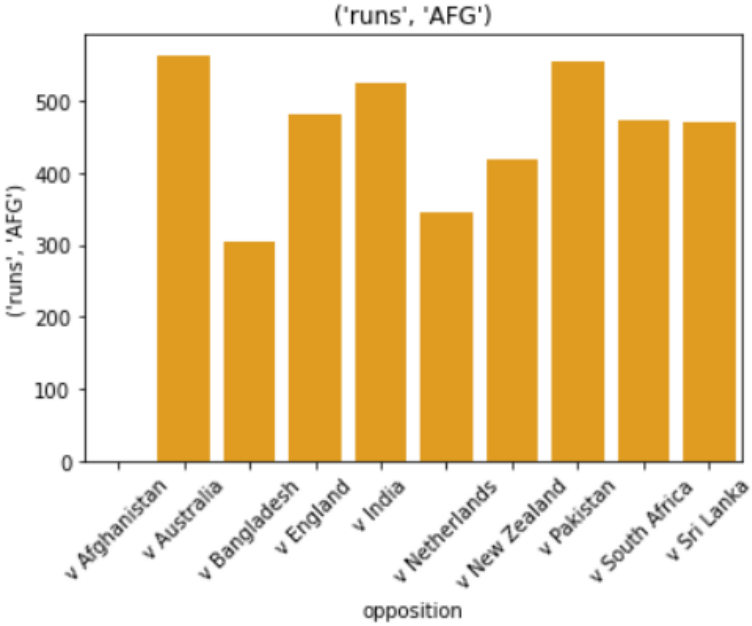
Pivot table

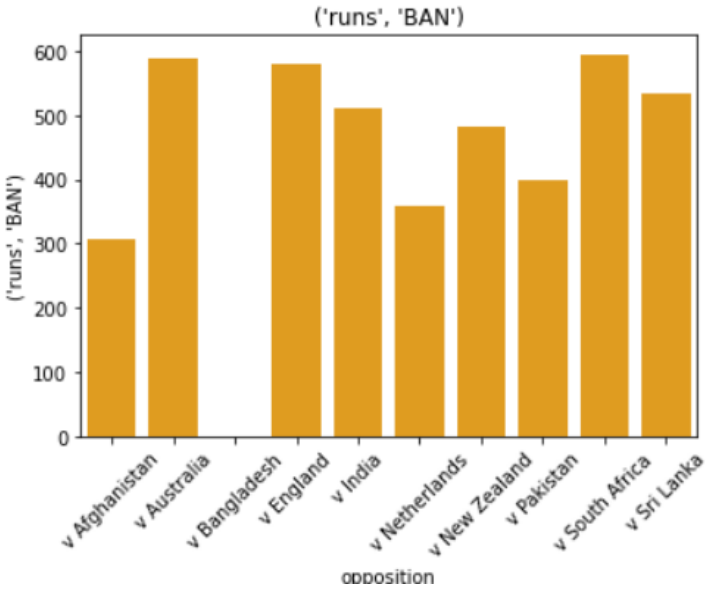
Investigate how teams against different oppositions in wkts

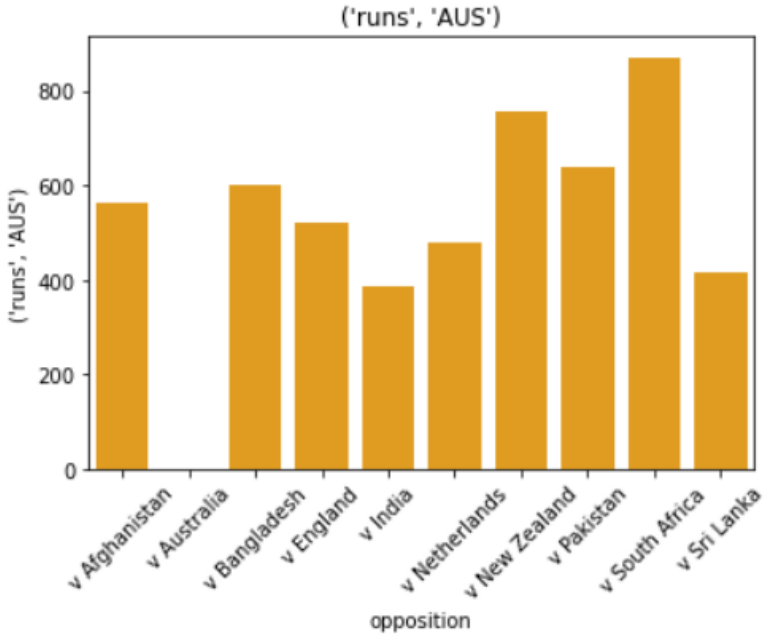
```
data1=pd.pivot_table(team_opposition,values=["runs"],index=["opposition"],columns=["team"])
data1
```

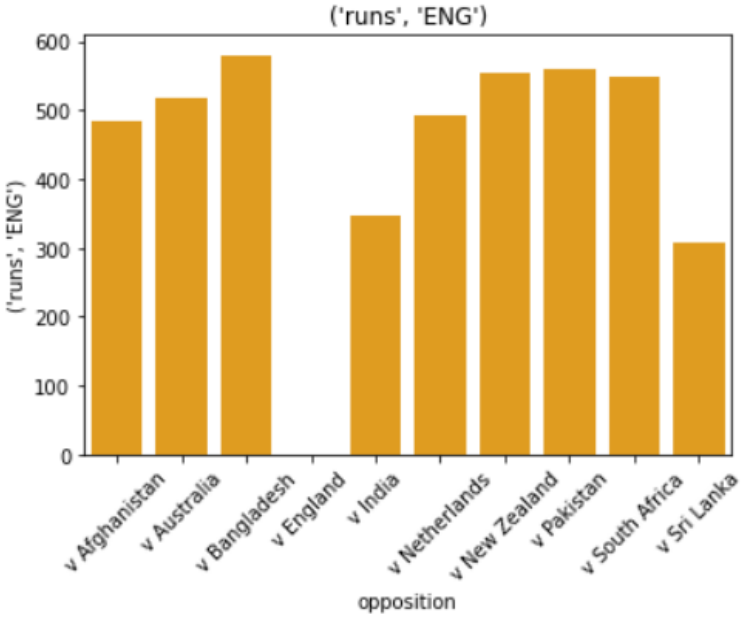
	runs									
team	AFG	AUS	BAN	ENG	IND	NED	NZ	PAK	SA	SL
opposition										
v Afghanistan	NaN	564.0	306.0	485.0	525.0	341.0	417.0	551.0	477.0	471.0
v Australia	563.0	NaN	589.0	519.0	386.0	482.0	744.0	642.0	884.0	411.0
v Bangladesh	303.0	599.0	NaN	580.0	512.0	359.0	481.0	401.0	603.0	541.0
v England	480.0	523.0	581.0	NaN	349.0	510.0	557.0	568.0	545.0	310.0
v India	525.0	386.0	511.0	346.0	NaN	646.0	1224.0	378.0	405.0	391.0
v Netherlands	346.0	477.0	360.0	493.0	623.0	13.0	531.0	482.0	434.0	495.0
v New Zealand	418.0	758.0	483.0	554.0	1241.0	538.0	NaN	589.0	507.0	339.0
v Pakistan	555.0	639.0	398.0	561.0	378.0	482.0	574.0	NaN	516.0	675.0
v South Africa	473.0	871.0	595.0	549.0	383.0	418.0	508.0	516.0	NaN	733.0
v Sri Lanka	471.0	417.0	535.0	309.0	386.0	479.0	340.0	659.0	726.0	NaN

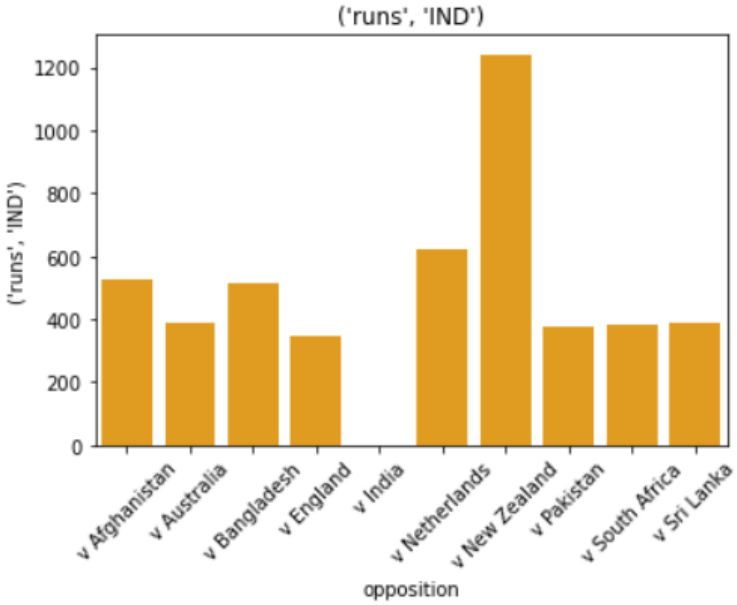
```
for i in data1:
    sns.barplot(x=data1.index, y=i, data=data1,color='orange')
    plt.title(f"{i}")
    plt.xticks(rotation=45)
    plt.show()
```

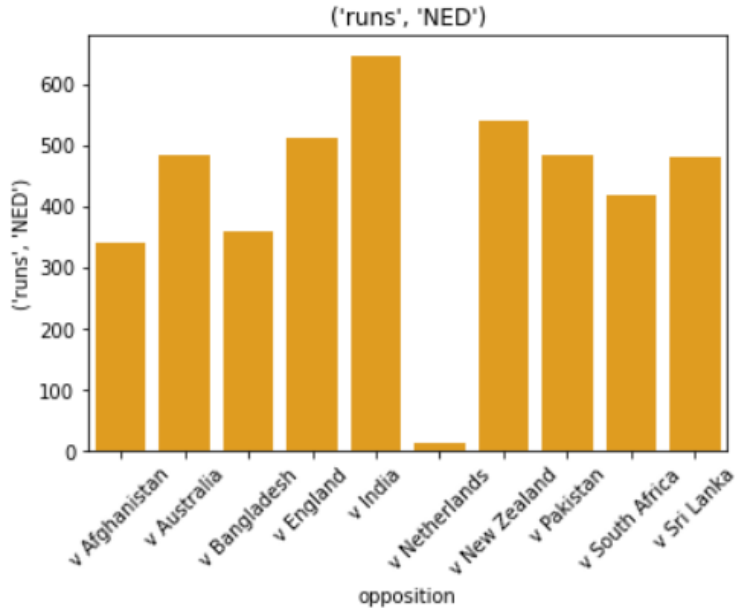

('runs', 'AFG')
('runs', 'AFG')
0
100
200
300
400
500
v Afghanistan
v Australia
v Bangladesh
v England
v India
v Netherlands
v New Zealand
v Pakistan
v South Africa
v Sri Lanka
opposition

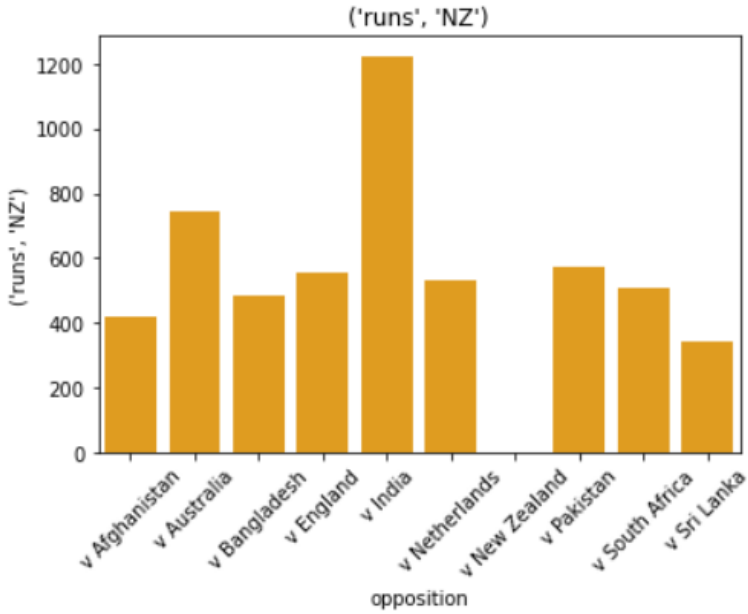

('runs', 'BAN')
('runs', 'BAN')
0
100
200
300
400
500
600
v Afghanistan
v Australia
v Bangladesh
v England
v India
v Netherlands
v New Zealand
v Pakistan
v South Africa
v Sri Lanka
opposition

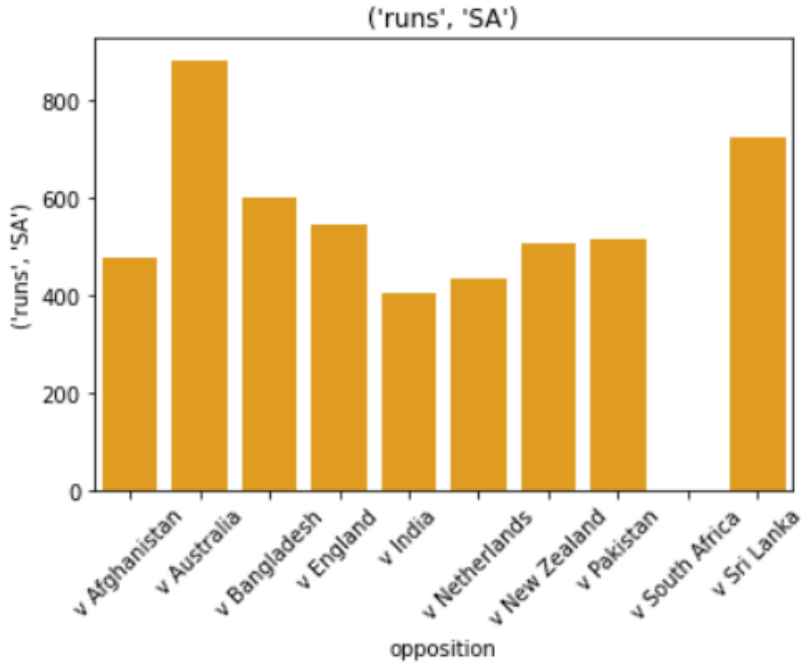

('runs', 'AUS')
('runs', 'AUS')
0
200
400
600
800
v Afghanistan
v Australia
v Bangladesh
v England
v India
v Netherlands
v New Zealand
v Pakistan
v South Africa
v Sri Lanka
opposition

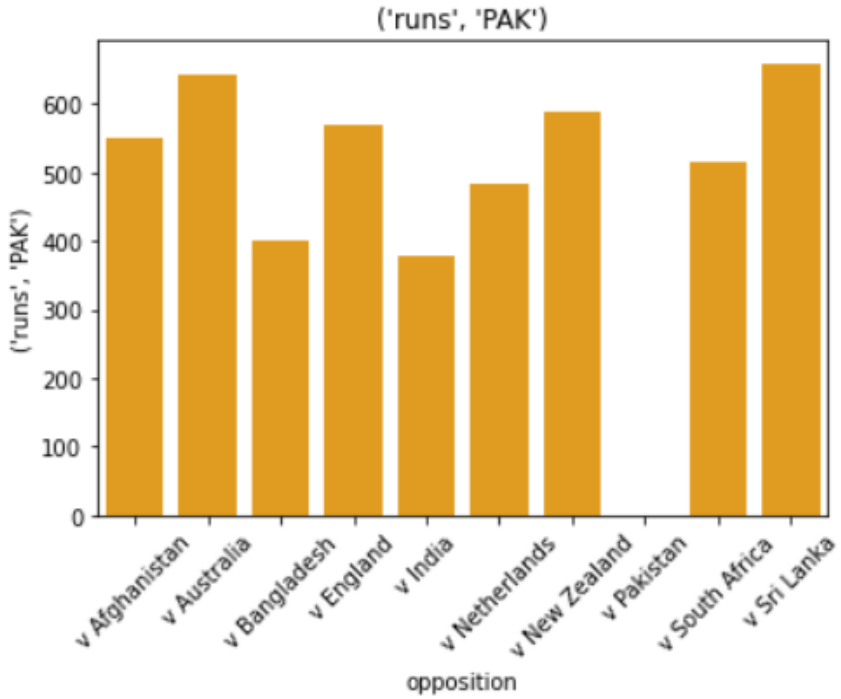

('runs', 'ENG')
('runs', 'ENG')
600
500
400
300
200
100
0
v Afghanistan
v Australia
v Bangladesh
v England
v India
v Netherlands
v New Zealand
v Pakistan
v South Africa
v Sri Lanka
opposition

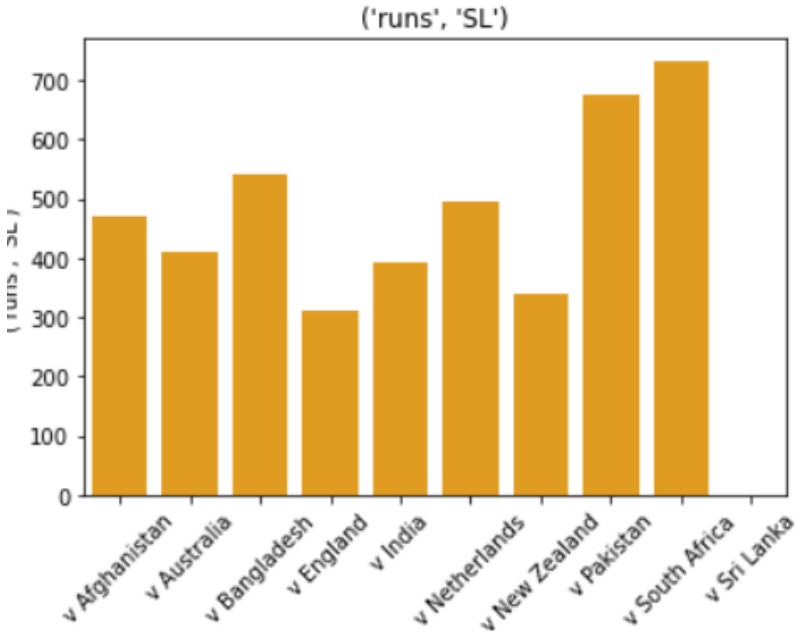

('runs', 'IND')
('runs', 'IND')
1200
1000
800
600
400
200
0
v Afghanistan
v Australia
v Bangladesh
v England
v India
v Netherlands
v New Zealand
v Pakistan
v South Africa
v Sri Lanka
opposition


('runs', 'NED')
('runs', 'NED')
600
500
400
300
200
100
0
v Afghanistan
v Australia
v Bangladesh
v England
v India
v Netherlands
v New Zealand
v Pakistan
v South Africa
v Sri Lanka
opposition


('runs', 'NZ')
('runs', 'NZ')
1200
1000
800
600
400
200
0
v Afghanistan
v Australia
v Bangladesh
v England
v India
v Netherlands
v New Zealand
v Pakistan
v South Africa
v Sri Lanka
opposition


('runs', 'SA')
('runs', 'SA')
800
600
400
200
0
v Afghanistan
v Australia
v Bangladesh
v England
v India
v Netherlands
v New Zealand
v Pakistan
v South Africa
v Sri Lanka
opposition


('runs', 'PAK')
('runs', 'PAK')
0
100
200
300
400
500
600
v Afghanistan
v Australia
v Bangladesh
v England
v India
v Netherlands
v New Zealand
v Pakistan
v South Africa
v Sri Lanka
opposition


('runs', 'SL')
0
100
200
300
400
500
600
700
v Afghanistan
v Australia
v Bangladesh
v England
v India
v Netherlands
v New Zealand
v Pakistan
v South Africa
v Sri Lanka

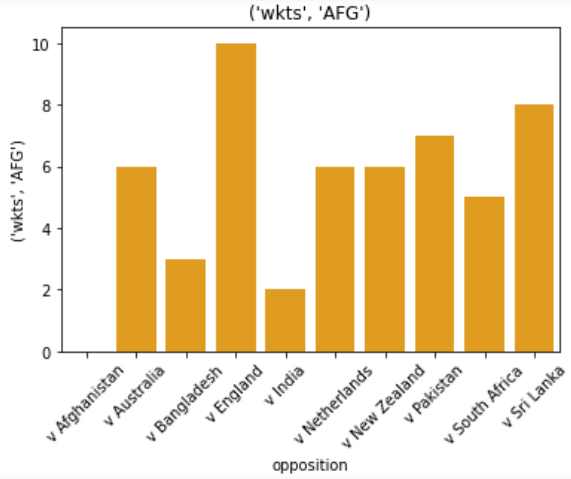
Pivout table

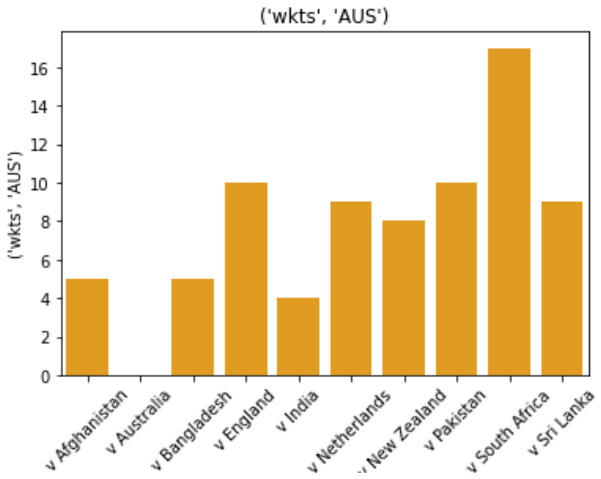
Investigate how teams against different oppositions in runs

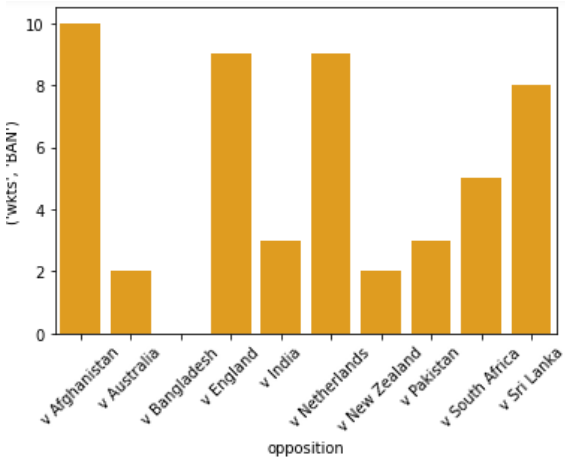
```
data2=pd.pivot_table(team_opposition,values=["wkts"],index=["opposition"],columns=["team"])
data2
```

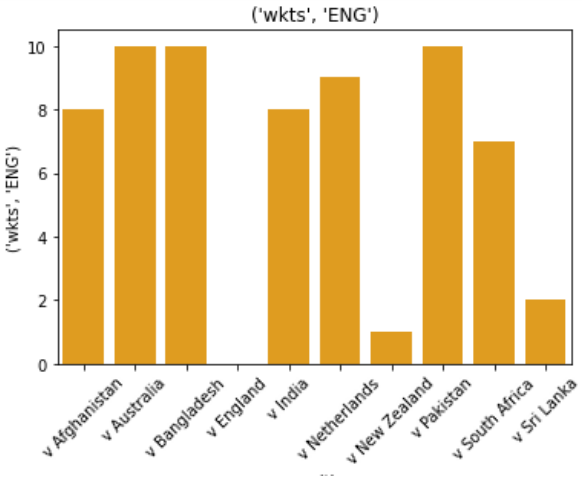
	wkts									
team	AFG	AUS	BAN	ENG	IND	NED	NZ	PAK	SA	SL
opposition										
v Afghanistan	NaN	5.0	10.0	8.0	8.0	3.0	10.0	2.0	9.0	3.0
v Australia	6.0	NaN	2.0	10.0	10.0	7.0	10.0	9.0	17.0	4.0
v Bangladesh	3.0	5.0	NaN	10.0	8.0	9.0	9.0	10.0	10.0	7.0
v England	10.0	10.0	9.0	NaN	10.0	8.0	9.0	8.0	9.0	8.0
v India	2.0	4.0	3.0	8.0	NaN	4.0	9.0	3.0	5.0	6.0
v Netherlands	6.0	9.0	9.0	9.0	9.0	1.0	9.0	9.0	8.0	9.0
v New Zealand	6.0	8.0	2.0	1.0	19.0	7.0	NaN	6.0	10.0	4.0
v Pakistan	7.0	10.0	3.0	10.0	10.0	9.0	1.0	NaN	10.0	4.0
v South Africa	5.0	17.0	5.0	7.0	10.0	10.0	4.0	9.0	NaN	5.0
v Sri Lanka	8.0	9.0	8.0	2.0	10.0	5.0	10.0	9.0	10.0	NaN

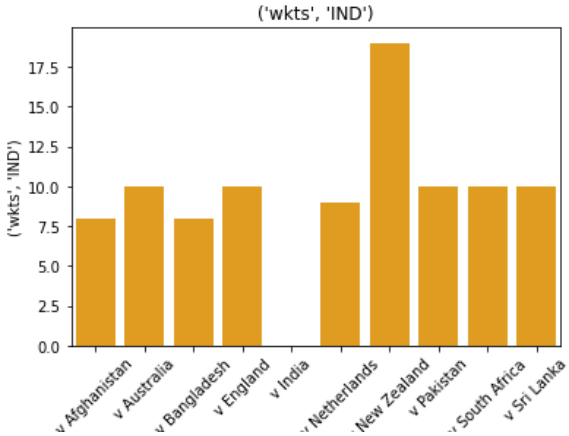
```
for i in data2:
    sns.barplot(x=data2.index, y=i, data=data2,color='orange')
    plt.title(f"{i}")
    plt.xticks(rotation=45)
    plt.show()
```

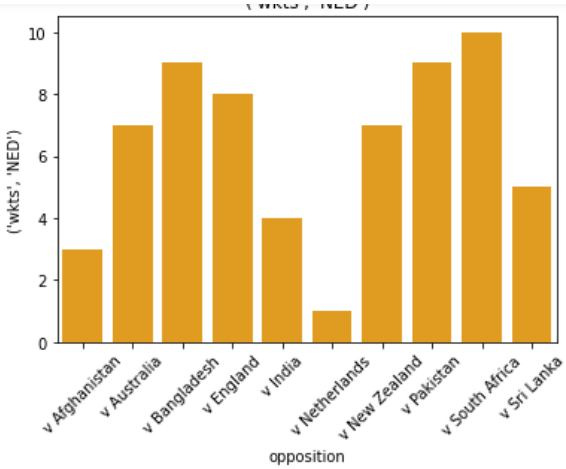


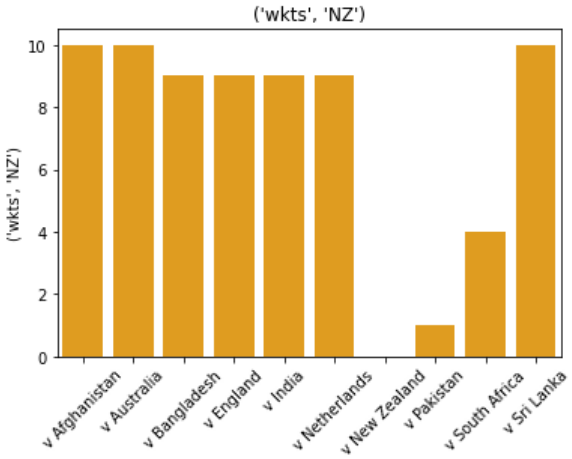

('wkts', 'AUS')
('wkts', 'AUS')
0
2
4
6
8
10
12
14
16
v Afghanistan
v Australia
v Bangladesh
v England
v India
v Netherlands
v New Zealand
v Pakistan
v South Africa
v Sri Lanka

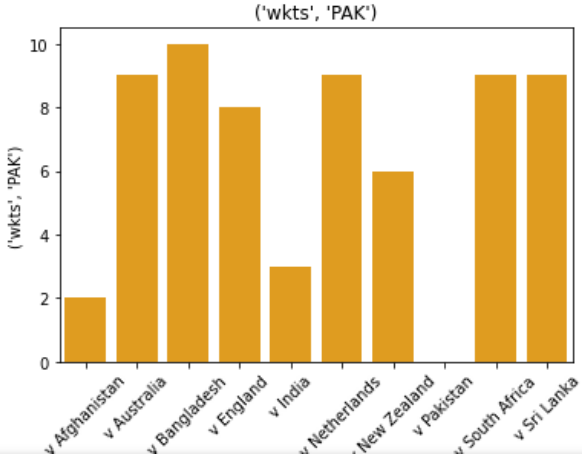

('wkts', 'BAN')
0
2
4
6
8
10
v Afghanistan
v Australia
v Bangladesh
v England
v India
v Netherlands
v New Zealand
v Pakistan
v South Africa
v Sri Lanka
opposition

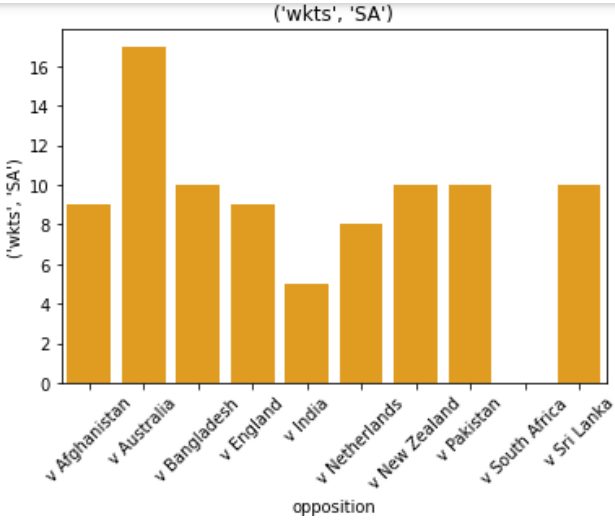

('wkts', 'ENG')
('wkts', 'ENG')
0
2
4
6
8
10
v Afghanistan
v Australia
v Bangladesh
v England
v India
v Netherlands
v New Zealand
v Pakistan
v South Africa
v Sri Lanka

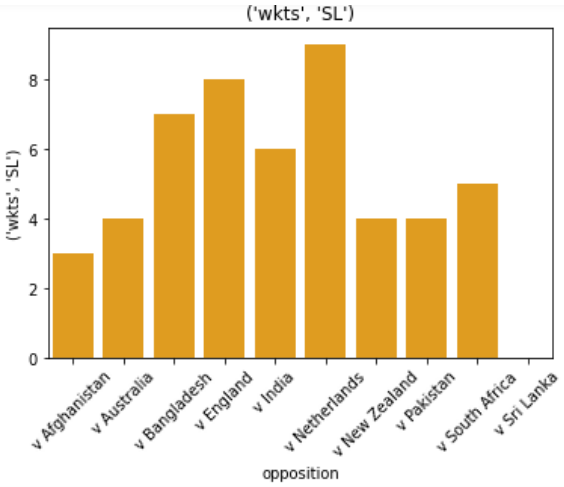

('wkts', 'IND')
('wkts', 'IND')
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
v Afghanistan
v Australia
v Bangladesh
v England
v India
v Netherlands
v New Zealand
v Pakistan
v South Africa
v Sri Lanka


('wkts', 'NED')
0
2
4
6
8
10
v Afghanistan
v Australia
v Bangladesh
v England
v India
v Netherlands
v New Zealand
v Pakistan
v South Africa
v Sri Lanka
opposition


('wkts', 'NZ')
0
2
4
6
8
10
v Afghanistan
v Australia
v Bangladesh
v England
v India
v Netherlands
v New Zealand
v Pakistan
v South Africa
v Sri Lanka


('wkts', 'PAK')
0
2
4
6
8
10
v Afghanistan
v Australia
v Bangladesh
v England
v India
v Netherlands
v New Zealand
v Pakistan
v South Africa
v Sri Lanka


('wkts', 'SA')
0
2
4
6
8
10
12
14
16
v Afghanistan
v Australia
v Bangladesh
v England
v India
v Netherlands
v New Zealand
v Pakistan
v South Africa
v Sri Lanka
opposition



Players

```
player_opposition= data.groupby(['player','opposition']).agg({
    'runs': 'sum',
    'wkts': 'sum'
}).reset_index()
display(player_opposition)
```

	player	opposition	runs	wkts
0	A Dutt (NED)	v Afghanistan	59	0.0
1	A Dutt (NED)	v Australia	60	1.0
2	A Dutt (NED)	v Bangladesh	35	1.0
3	A Dutt (NED)	v England	68	2.0
4	A Dutt (NED)	v India	57	0.0
...	...	...	...	...
973	WA Young (NZ)	v Australia	32	0.0
974	WA Young (NZ)	v England	0	0.0
975	WA Young (NZ)	v India	17	0.0
976	WA Young (NZ)	v Netherlands	70	0.0
977	WA Young (NZ)	v South Africa	33	0.0

Investigate how players against different oppositions in runs

```
player_runs=pd.pivot_table(player_opposition,values=["runs"],index=["player"],columns=["opposition"])
player_runs
```

	runs									
opposition	**v Afghanistan**	**v Australia**	**v Bangladesh**	**v England**	**v India**	**v Netherlands**	**v New Zealand**	**v Pakistan**	**v South Africa**	**v Sri Lanka**
player										
A Dutt (NED)	59.0	60.0	35.0	68.0	57.0	NaN	73.0	49.0	42.0	53.0
A Zampa (AUS)	58.0	NaN	32.0	50.0	59.0	9.0	74.0	54.0	136.0	47.0
AAP Atkinson (ENG)	NaN	NaN	NaN	NaN	NaN	43.0	NaN	45.0	95.0	NaN
AD Mathews (SL)	41.0	NaN	35.0	14.0	23.0	NaN	45.0	NaN	NaN	NaN
AK Markram (SA)	50.0	89.0	60.0	42.0	26.0	1.0	6.0	111.0	NaN	106.0
...	...	...	...	...	...	...	...	...	...	...
Usama Mir (PAK)	55.0	82.0	66.0	NaN	NaN	NaN	NaN	NaN	45.0	NaN
V Kohli (IND)	55.0	85.0	105.0	0.0	NaN	64.0	212.0	16.0	101.0	88.0
Vikramjit Singh (NED)	NaN	52.0	3.0	NaN	NaN	NaN	21.0	68.0	2.0	4.0
W Barresi (NED)	1.0	NaN	41.0	37.0	4.0	NaN	NaN	NaN	NaN	NaN
WA Young (NZ)	54.0	32.0	NaN	0.0	17.0	70.0	NaN	NaN	33.0	NaN

152 rows × 10 columns

```
x1=player_runs.sum(1).sort_values(ascending=False).head(5)
print("Highest five players runs against oppositions")
print("-------------------------------------------")
x1
```

```
Highest five players runs against oppositions
-------------------------------------------

player
R Ravindra (NZ)             971.0
V Kohli (IND)               726.0
GJ Maxwell (AUS)            693.0
BFW de Leede (NED)          626.0
Azmatullah Omarzai (AFG)    623.0
dtype: float64
```

```
player_runs=pd.pivot_table(player_opposition,values=["runs"],index=["player"],columns=["opposition"]).loc[x1.index]
player_runs
```

	runs									
opposition	v Afghanistan	v Australia	v Bangladesh	v England	v India	v Netherlands	v New Zealand	v Pakistan	v South Africa	v Sri Lanka
player										
R Ravindra (NZ)	66.0	172.0	46.0	199.0	194.0	97.0	NaN	108.0	26.0	63.0
V Kohli (IND)	55.0	85.0	105.0	0.0	NaN	64.0	212.0	16.0	101.0	88.0
GJ Maxwell (AUS)	256.0	NaN	NaN	NaN	48.0	106.0	103.0	40.0	73.0	67.0
BFW de Leede (NED)	3.0	119.0	42.0	84.0	94.0	NaN	82.0	129.0	38.0	35.0
Azmatullah Omarzai (AFG)	NaN	74.0	31.0	32.0	96.0	42.0	83.0	50.0	105.0	110.0

```
player_runs.columns
```

```
MultiIndex([('runs',  'v Afghanistan'),
            ('runs',    'v Australia'),
            ('runs',   'v Bangladesh'),
            ('runs',      'v England'),
            ('runs',        'v India'),
            ('runs',  'v Netherlands'),
            ('runs',  'v New Zealand'),
            ('runs',     'v Pakistan'),
            ('runs', 'v South Africa'),
            ('runs',    'v Sri Lanka')],
           names=[None, 'opposition'])
```

```
player_runs1=pd.pivot_table(player_opposition,values=["runs"],index=["opposition"],columns=["player"])
player_runs1
```

v Australia	60.0	NaN	NaN	NaN	89.0	NaN	NaN	14.0	58.0	64.0	...	21.0	36.0	NaN	61.0	74.0	8
v Bangladesh	35.0	32.0	NaN	35.0	60.0	NaN	NaN	NaN	53.0	68.0	...	NaN	NaN	NaN	NaN	NaN	6
v England	68.0	50.0	NaN	14.0	42.0	NaN	NaN	41.0	NaN	0.0	...	NaN	1.0	NaN	53.0	39.0	N
v India	57.0	59.0	NaN	23.0	26.0	NaN	0.0	54.0	48.0	20.0	...	5.0	51.0	NaN	NaN	16.0	N
v Netherlands	NaN	9.0	43.0	NaN	1.0	NaN	NaN	NaN	55.0	NaN	...	53.0	15.0	NaN	54.0	NaN	
v New Zealand	73.0	74.0	NaN	45.0	6.0	NaN	NaN	21.0	62.0	4.0	...	NaN	16.0	NaN	73.0	13.0	
v Pakistan	49.0	54.0	45.0	NaN	111.0	NaN	NaN	5.0	55.0	NaN	...	2.0	0.0	NaN	42.0	7.0	
v South Africa	42.0	136.0	95.0	NaN	NaN	NaN	NaN	20.0	71.0	9.0	...	4.0	12.0	NaN	NaN	NaN	4
v Sri Lanka	53.0	47.0	NaN	NaN	106.0	NaN	NaN	9.0	41.0	113.0	...	2.0	9.0	85.0	39.0	15.0	

10 rows × 152 columns

```
player_runs1.columns
```

```
MultiIndex([('runs',             'A Dutt (NED)'),
            ('runs',            'A Zampa (AUS)'),
            ('runs',        'AAP Atkinson (ENG)'),
            ('runs',           'AD Mathews (SL)'),
            ('runs',           'AK Markram (SA)'),
            ('runs',       'AL Phehlukwayo (SA)'),
            ('runs',            'AT Carey (AUS)'),
            ('runs',       'AT Nidamanuru (NED)'),
            ('runs',           'AU Rashid (ENG)'),
            ('runs',   'Abdullah Shafique (PAK)'),
            ...
            ('runs',           'TWM Latham (NZ)'),
            ('runs',        'Tanzid Hasan (BAN)'),
            ('runs', 'Tanzim Hasan Sakib (BAN)'),
            ('runs',        'Taskin Ahmed (BAN)'),
            ('runs',       'Towhid Hridoy (BAN)'),
            ('runs',           'Usama Mir (PAK)'),
            ('runs',             'V Kohli (IND)'),
            ('runs',     'Vikramjit Singh (NED)'),
            ('runs',           'W Barresi (NED)'),
            ('runs',             'WA Young (NZ)')],
           names=[None, 'player'], length=152)
```

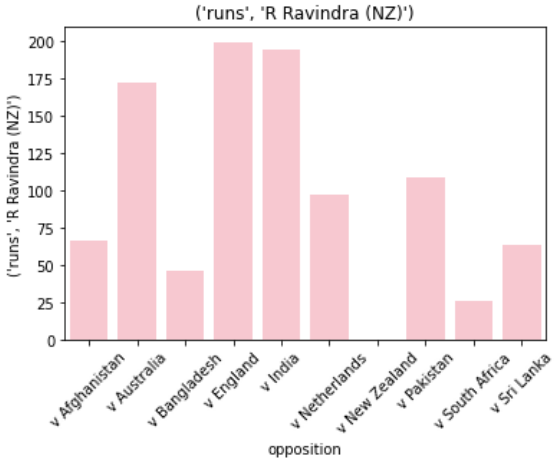
```
# take list to slice data by players name
l=[]
for i in player_runs.index:
    l.append(("runs",i))
l
```

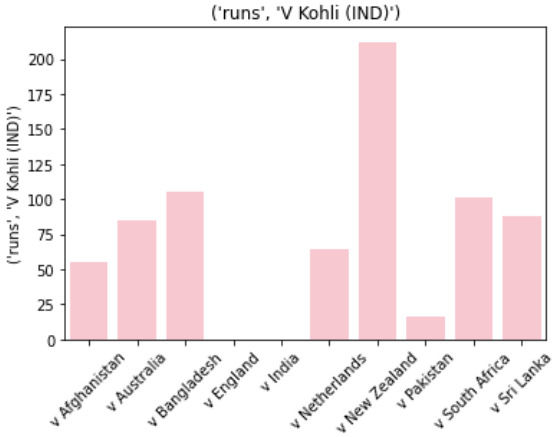
```
[('runs', 'R Ravindra (NZ)'),
 ('runs', 'V Kohli (IND)'),
 ('runs', 'GJ Maxwell (AUS)'),
 ('runs', 'BFW de Leede (NED)'),
 ('runs', 'Azmatullah Omarzai (AFG)')]
```

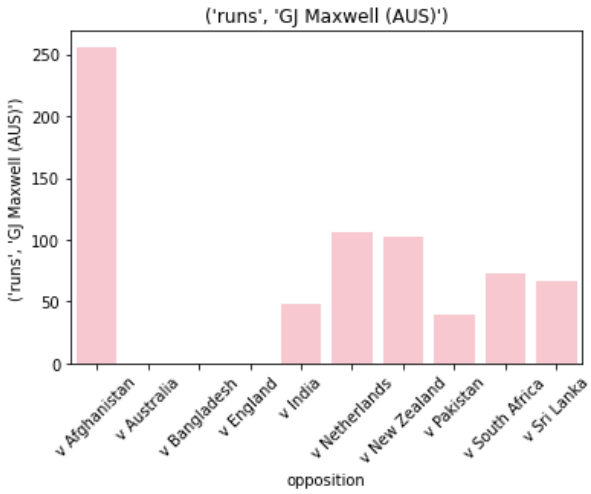
```
player_runs2=player_runs1.loc[:,l]
player_runs2
```

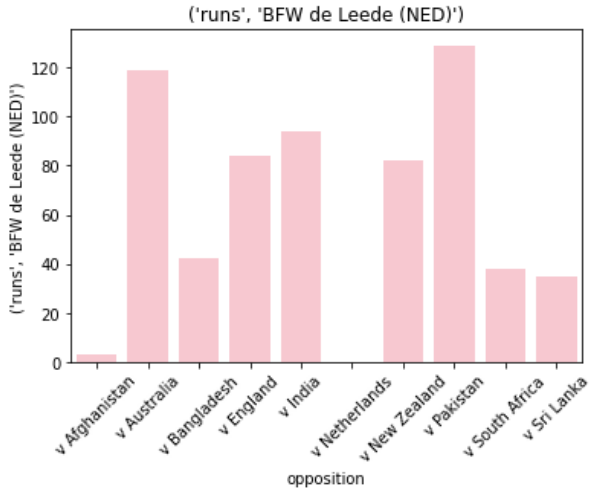
	runs				
player	R Ravindra (NZ)	V Kohli (IND)	GJ Maxwell (AUS)	BFW de Leede (NED)	Azmatullah Omarzai (AFG)
opposition					
v Afghanistan	66.0	55.0	256.0	3.0	NaN
v Australia	172.0	85.0	NaN	119.0	74.0
v Bangladesh	46.0	105.0	NaN	42.0	31.0
v England	199.0	0.0	NaN	84.0	32.0
v India	194.0	NaN	48.0	94.0	96.0
v Netherlands	97.0	64.0	106.0	NaN	42.0
v New Zealand	NaN	212.0	103.0	82.0	83.0
v Pakistan	108.0	16.0	40.0	129.0	50.0
v South Africa	26.0	101.0	73.0	38.0	105.0
v Sri Lanka	63.0	88.0	67.0	35.0	110.0

```
for i in player_runs2:
    sns.barplot(x=player_runs2.index, y=i, data=player_runs2,color='pink')
    plt.title(f"{i}")
    plt.xticks(rotation=45)
    plt.show()
```









Investigate how players against different oppositions in wkts

```
player_wkts=pd.pivot_table(player_opposition,values=["wkts"],index=["player"],columns=["opposition"])
player_wkts
```

	wkts									
opposition	**v Afghanistan**	**v Australia**	**v Bangladesh**	**v England**	**v India**	**v Netherlands**	**v New Zealand**	**v Pakistan**	**v South Africa**	**v Sri Lanka**
player										
A Dutt (NED)	0.0	1.0	1.0	2.0	0.0	NaN	2.0	1.0	0.0	3.0
A Zampa (AUS)	1.0	NaN	2.0	3.0	0.0	4.0	3.0	4.0	1.0	4.0
AAP Atkinson (ENG)	NaN	NaN	NaN	NaN	NaN	0.0	NaN	2.0	2.0	NaN
AD Mathews (SL)	0.0	NaN	2.0	2.0	0.0	NaN	2.0	NaN	NaN	NaN
AK Markram (SA)	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	NaN	0.0
...	...	...	...	...	...	...	...	...	...	...
Usama Mir (PAK)	0.0	1.0	1.0	NaN	NaN	NaN	NaN	NaN	2.0	NaN
V Kohli (IND)	0.0	0.0	0.0	0.0	NaN	1.0	0.0	0.0	0.0	0.0
Vikramjit Singh (NED)	NaN	0.0	0.0	NaN	NaN	NaN	0.0	0.0	0.0	0.0
W Barresi (NED)	0.0	NaN	0.0	0.0	0.0	NaN	NaN	NaN	NaN	NaN
WA Young (NZ)	0.0	0.0	NaN	0.0	0.0	0.0	NaN	NaN	0.0	NaN

152 rows × 10 columns

```
x2=player_wkts.sum(1).sort_values(ascending=False).head(5)
print("Highest five players wkts against oppositions")
print("-------------------------------------------")
x2
```

```
Highest five players wkts against oppositions
-------------------------------------------

player
Mohammed Shami (IND)         23.0
A Zampa (AUS)                22.0
D Madushanka (SL)            21.0
G Coetzee (SA)               20.0
Shaheen Shah Afridi (PAK)    18.0
dtype: float64
```

```
player_wkts=pd.pivot_table(player_opposition,values=["wkts"],index=["player"],columns=["opposition"]).loc[x2.index]
player_wkts
```

	wkts									
opposition	v Afghanistan	v Australia	v Bangladesh	v England	v India	v Netherlands	v New Zealand	v Pakistan	v South Africa	v Sri Lanka
player										
Mohammed Shami (IND)	NaN	NaN	NaN	4.0	NaN	0.0	12.0	NaN	2.0	5.0
A Zampa (AUS)	1.0	NaN	2.0	3.0	0.0	4.0	3.0	4.0	1.0	4.0
D Madushanka (SL)	2.0	3.0	3.0	0.0	5.0	4.0	0.0	2.0	2.0	NaN
G Coetzee (SA)	4.0	2.0	3.0	3.0	NaN	1.0	2.0	2.0	NaN	3.0
Shaheen Shah Afridi (PAK)	1.0	5.0	3.0	2.0	2.0	1.0	0.0	NaN	3.0	1.0

```
player_wkts1=pd.pivot_table(player_opposition,values=["wkts"],index=["opposition"],columns=["player"])
player_wkts1
```

player	A Dutt (NED)	A Zampa (AUS)	AAP Atkinson (ENG)	AD Mathews (SL)	AK Markram (SA)	AL Phehlukwayo (SA)	AT Carey (AUS)	AT Nidamanuru (NED)	AU Rashid (ENG)	Abdullah Shafique (PAK)	...	TWM Latham (NZ)	Tanzid Hasan (BAN)	Tanzim Hasan Sakib (BAN)	Taskin Ahmed (BAN)	Towhid Hridoy (BAN)	Us
opposition																	
v Afghanistan	0.0	1.0	NaN	0.0	0.0	1.0	NaN	NaN	3.0	0.0	...	0.0	0.0	NaN	1.0	NaN	
v Australia	1.0	NaN	NaN	NaN	1.0	NaN	NaN	0.0	2.0	0.0	...	0.0	0.0	NaN	1.0	0.0	
v Bangladesh	1.0	2.0	NaN	2.0	0.0	NaN	NaN	NaN	1.0	0.0	...	NaN	NaN	NaN	NaN	NaN	
v England	2.0	3.0	NaN	2.0	0.0	NaN	NaN	0.0	NaN	0.0	...	NaN	0.0	NaN	1.0	0.0	
v India	0.0	0.0	NaN	0.0	0.0	NaN	0.0	0.0	2.0	0.0	...	0.0	0.0	NaN	NaN	0.0	
v Netherlands	NaN	4.0	0.0	NaN	0.0	NaN	NaN	NaN	3.0	NaN	...	0.0	0.0	NaN	2.0	NaN	
v New Zealand	2.0	3.0	NaN	2.0	0.0	NaN	NaN	0.0	0.0	0.0	...	NaN	0.0	NaN	0.0	0.0	
v Pakistan	1.0	4.0	2.0	NaN	0.0	NaN	NaN	0.0	2.0	NaN	...	0.0	0.0	NaN	0.0	0.0	
v South Africa	0.0	1.0	2.0	NaN	NaN	NaN	NaN	0.0	2.0	0.0	...	0.0	0.0	NaN	NaN	NaN	
v Sri Lanka	3.0	4.0	NaN	NaN	0.0	NaN	NaN	0.0	0.0	0.0	...	0.0	0.0	3.0	0.0	0.0	

10 rows × 152 columns

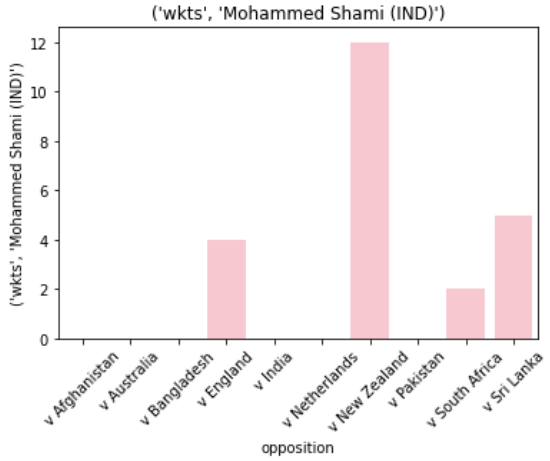
```
# take list to slice data by players name
l1=[]
for i in player_wkts.index:
    l1.append(("wkts",i))
l1
```

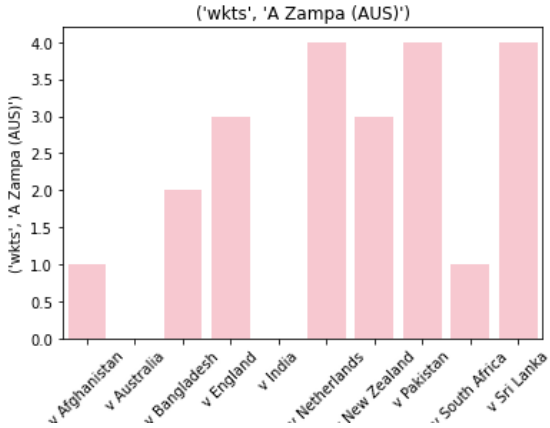
```
[('wkts', 'Mohammed Shami (IND)'),
 ('wkts', 'A Zampa (AUS)'),
 ('wkts', 'D Madushanka (SL)'),
 ('wkts', 'G Coetzee (SA)'),
 ('wkts', 'Shaheen Shah Afridi (PAK)')]
```

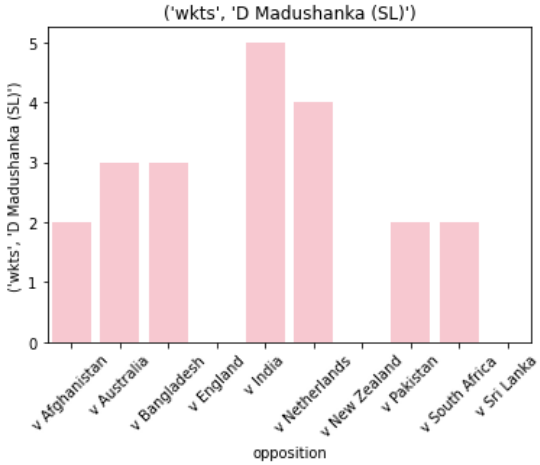
```
player_wkts2=player_wkts1.loc[:,l1]
player_wkts2
```

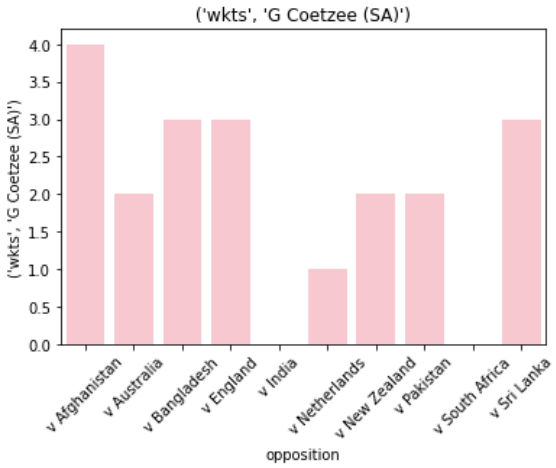
	wkts				
player	Mohammed Shami (IND)	A Zampa (AUS)	D Madushanka (SL)	G Coetzee (SA)	Shaheen Shah Afridi (PAK)
opposition					
v Afghanistan	NaN	1.0	2.0	4.0	1.0
v Australia	NaN	NaN	3.0	2.0	5.0
v Bangladesh	NaN	2.0	3.0	3.0	3.0
v England	4.0	3.0	0.0	3.0	2.0
v India	NaN	0.0	5.0	NaN	2.0
v Netherlands	0.0	4.0	4.0	1.0	1.0
v New Zealand	12.0	3.0	0.0	2.0	0.0
v Pakistan	NaN	4.0	2.0	2.0	NaN
v South Africa	2.0	1.0	2.0	NaN	3.0
v Sri Lanka	5.0	4.0	NaN	3.0	1.0

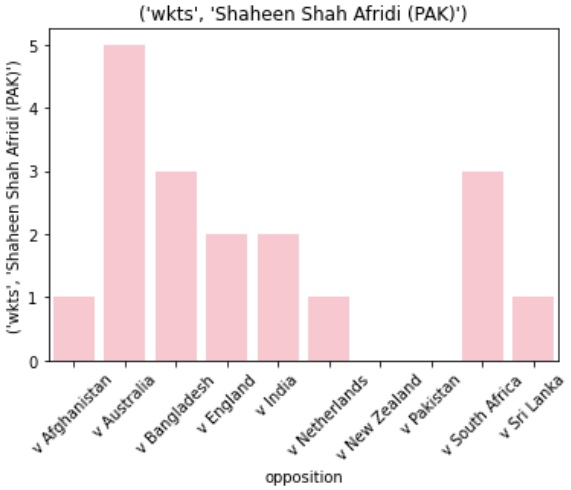
```
for i in player_wkts2:
    sns.barplot(x=player_wkts2.index, y=i, data=player_wkts2,color='pink')
    plt.title(f"{i}")
    plt.xticks(rotation=45)
    plt.show()
```






('wkts', 'D Madushanka (SL)')
('wkts', 'D Madushanka (SL)')
0
1
2
3
4
5
v Afghanistan
v Australia
v Bangladesh
v England
v India
v Netherlands
v New Zealand
v Pakistan
v South Africa
v Sri Lanka
opposition


('wkts', 'G Coetzee (SA)')
('wkts', 'G Coetzee (SA)')
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
v Afghanistan
v Australia
v Bangladesh
v England
v India
v Netherlands
v New Zealand
v Pakistan
v South Africa
v Sri Lanka
opposition


('wkts', 'Shaheen Shah Afridi (PAK)')
('wkts', 'Shaheen Shah Afridi (PAK)')
0
1
2
3
4
5
v Afghanistan
v Australia
v Bangladesh
v England
v India
v Netherlands
v New Zealand
v Pakistan
v South Africa
v Sri Lanka
opposition

Investigate how players against different ground in runs

```
player_ground= data.groupby(['player','ground']).agg({
    'runs': 'sum',
    'wkts': 'sum'
}).reset_index()
display(player_ground)
```

	player	ground	runs	wkts
0	A Dutt (NED)	Bengaluru	57	0.0
1	A Dutt (NED)	Delhi	60	1.0
2	A Dutt (NED)	Dharamsala	42	0.0
3	A Dutt (NED)	Eden Gardens	35	1.0
4	A Dutt (NED)	Hyderabad	122	3.0
...	...	...	...	...
810	WA Young (NZ)	Ahmedabad	0	0.0
811	WA Young (NZ)	Chennai	54	0.0
812	WA Young (NZ)	Dharamsala	49	0.0
813	WA Young (NZ)	Hyderabad	70	0.0
814	WA Young (NZ)	Pune	33	0.0

```
player_runs=pd.pivot_table(player_ground,values=["runs"],index=["player"],columns=["ground"])
player_runs
```

	runs									
ground	**Ahmedabad**	**Bengaluru**	**Chennai**	**Delhi**	**Dharamsala**	**Eden Gardens**	**Hyderabad**	**Lucknow**	**Pune**	**Wankhede**
player										
A Dutt (NED)	NaN	57.0	NaN	60.0	42.0	35.0	122.0	112.0	68.0	NaN
A Zampa (AUS)	50.0	54.0	59.0	9.0	74.0	55.0	NaN	128.0	32.0	58.0
AAP Atkinson (ENG)	NaN	NaN	NaN	NaN	NaN	45.0	NaN	NaN	43.0	95.0
AD Mathews (SL)	NaN	59.0	NaN	35.0	NaN	NaN	NaN	NaN	41.0	23.0
AK Markram (SA)	50.0	NaN	111.0	106.0	1.0	59.0	NaN	56.0	6.0	102.0
...	...	...	...	...	...	...	...	...	...	...
Usama Mir (PAK)	NaN	82.0	100.0	NaN	NaN	66.0	NaN	NaN	NaN	NaN
V Kohli (IND)	16.0	64.0	85.0	55.0	95.0	101.0	NaN	0.0	105.0	205.0
Vikramjit Singh (NED)	NaN	NaN	NaN	52.0	2.0	3.0	89.0	4.0	NaN	NaN
W Barresi (NED)	NaN	4.0	NaN	NaN	NaN	41.0	NaN	1.0	37.0	NaN
WA Young (NZ)	0.0	NaN	54.0	NaN	49.0	NaN	70.0	NaN	33.0	NaN

152 rows × 10 columns

```
x1=player_runs.sum(1).sort_values(ascending=False).head(5)
print("Highest five players runs against ground")
print("----------------------------------------------")
x1
```

```
Highest five players runs against ground
----------------------------------------------

player
R Ravindra (NZ)             971.0
V Kohli (IND)               726.0
GJ Maxwell (AUS)            693.0
BFW de Leede (NED)          626.0
Azmatullah Omarzai (AFG)    623.0
dtype: float64
```

```
player_runs=pd.pivot_table(player_ground,values=["runs"],index=["player"],columns=["ground"]).loc[x1.index]
player_runs
```

	runs									
ground	Ahmedabad	Bengaluru	Chennai	Delhi	Dharamsala	Eden Gardens	Hyderabad	Lucknow	Pune	Wankhede
player										
R Ravindra (NZ)	199.0	171.0	112.0	NaN	293.0	NaN	97.0	NaN	26.0	73.0
V Kohli (IND)	16.0	64.0	85.0	55.0	95.0	101.0	NaN	0.0	105.0	205.0
GJ Maxwell (AUS)	NaN	40.0	48.0	106.0	103.0	36.0	NaN	104.0	NaN	256.0
BFW de Leede (NED)	NaN	94.0	NaN	119.0	38.0	42.0	211.0	38.0	84.0	NaN
Azmatullah Omarzai (AFG)	105.0	NaN	133.0	128.0	31.0	NaN	NaN	42.0	110.0	74.0

```
player_runs1=pd.pivot_table(player_ground,values=["runs"],index=["ground"],columns=["player"])
player_runs1
```

player	A Dutt (NED)	A Zampa (AUS)	AAP Atkinson (ENG)	AD Mathews (SL)	AK Markram (SA)	AL Phehlukwayo (SA)	AT Carey (AUS)	AT Nidamanuru (NED)	AU Rashid (ENG)	Abdullah Shafique (PAK)	...	TWM Latham (NZ)	Tanzid Hasan (BAN)	Tanzim Hasan Sakib (BAN)	Taskin Ahmed (BAN)	Towhid Hridoy (BAN)	Us
ground																	
Ahmedabad	NaN	50.0	NaN	NaN	50.0	75.0	NaN	NaN	120.0	20.0	...	NaN	NaN	NaN	NaN	NaN	
Bengaluru	57.0	54.0	NaN	59.0	NaN	NaN	NaN	54.0	41.0	68.0	...	4.0	NaN	NaN	NaN	NaN	
Chennai	NaN	59.0	NaN	NaN	111.0	NaN	0.0	NaN	NaN	67.0	...	68.0	16.0	NaN	73.0	13.0	
Delhi	60.0	9.0	NaN	35.0	106.0	NaN	NaN	14.0	62.0	NaN	...	NaN	9.0	85.0	39.0	15.0	
Dharamsala	42.0	74.0	NaN	NaN	1.0	NaN	NaN	20.0	53.0	NaN	...	26.0	6.0	NaN	85.0	39.0	
Eden Gardens	35.0	55.0	45.0	NaN	59.0	NaN	NaN	NaN	55.0	68.0	...	NaN	15.0	NaN	96.0	7.0	
Hyderabad	122.0	NaN	NaN	NaN	NaN	NaN	NaN	26.0	NaN	113.0	...	53.0	NaN	NaN	NaN	NaN	
Lucknow	112.0	128.0	NaN	NaN	56.0	NaN	NaN	9.0	48.0	NaN	...	NaN	NaN	NaN	NaN	NaN	
Pune	68.0	32.0	43.0	41.0	6.0	NaN	NaN	41.0	55.0	NaN	...	4.0	87.0	NaN	61.0	90.0	
Wankhede	NaN	58.0	95.0	23.0	102.0	NaN	NaN	NaN	71.0	NaN	...	0.0	12.0	NaN	NaN	NaN	

10 rows × 152 columns

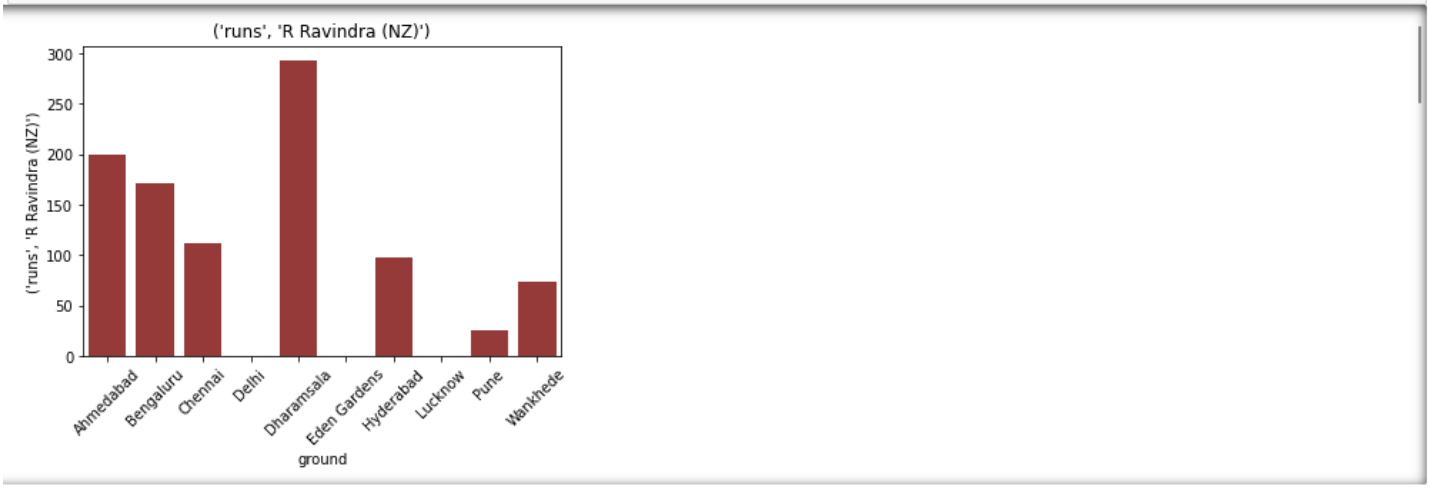
```
# take list to slice data by players name
l=[]
for i in player_runs.index:
    l.append(("runs",i))
l
```

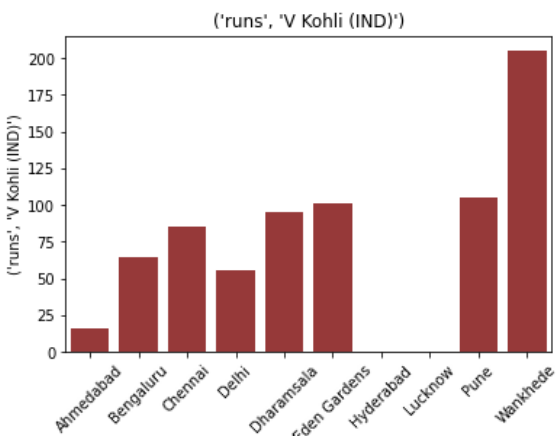
```
[('runs', 'R Ravindra (NZ)'),
 ('runs', 'V Kohli (IND)'),
 ('runs', 'GJ Maxwell (AUS)'),
 ('runs', 'BFW de Leede (NED)'),
 ('runs', 'Azmatullah Omarzai (AFG)')]
```

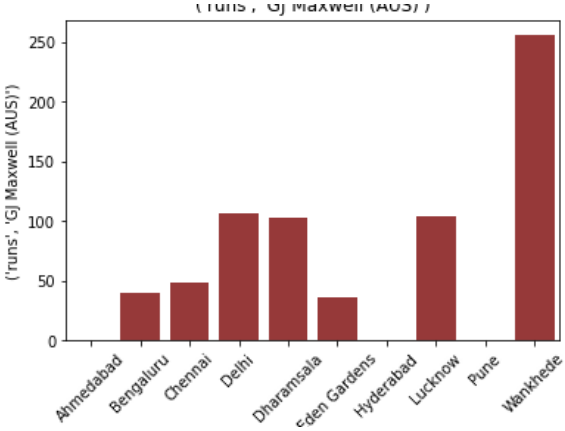
```
player_runs2=player_runs1.loc[:,l]
player_runs2
```

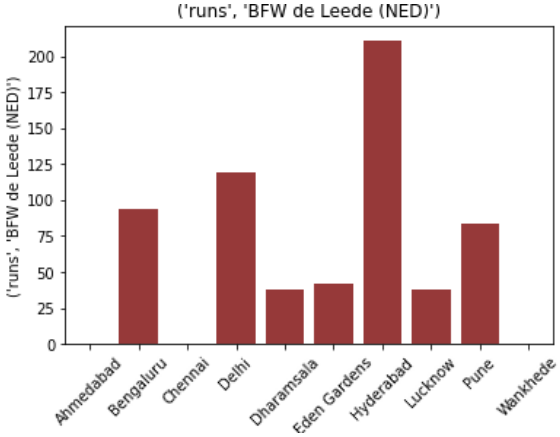
	runs				
player	R Ravindra (NZ)	V Kohli (IND)	GJ Maxwell (AUS)	BFW de Leede (NED)	Azmatullah Omarzai (AFG)
ground					
Ahmedabad	199.0	16.0	NaN	NaN	105.0
Bengaluru	171.0	64.0	40.0	94.0	NaN
Chennai	112.0	85.0	48.0	NaN	133.0
Delhi	NaN	55.0	106.0	119.0	128.0
Dharamsala	293.0	95.0	103.0	38.0	31.0
Eden Gardens	NaN	101.0	36.0	42.0	NaN
Hyderabad	97.0	NaN	NaN	211.0	NaN
Lucknow	NaN	0.0	104.0	38.0	42.0
Pune	26.0	105.0	NaN	84.0	110.0
Wankhede	73.0	205.0	256.0	NaN	74.0

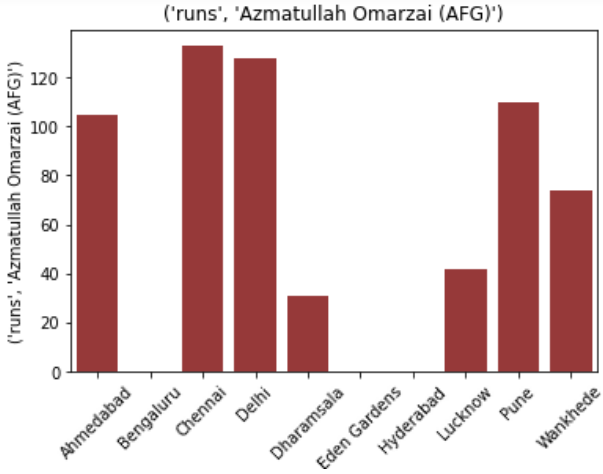
```
for i in player_runs2:
    sns.barplot(x=player_runs2.index, y=i, data=player_runs2,color='brown')
    plt.title(f"{i}")
    plt.xticks(rotation=45)
    plt.show()
```






('runs', 'GJ Maxwell (AUS)')
250
200
150
100
50
0
Ahmedabad
Bengaluru
Chennai
Delhi
Dharamsala
Hyderabad
Lucknow
Pune
Wankhede


('runs', 'BFW de Leede (NED)')
('runs', 'BFW de Leede (NED)')
200
175
150
125
100
75
50
25
0
Ahmedabad
Bengaluru
Chennai
Delhi
Dharamsala
Eden Gardens
Hyderabad
Lucknow
Pune
Wankhede


('runs', 'Azmatullah Omarzai (AFG)')
('runs', 'Azmatullah Omarzai (AFG)')
120
100
80
60
40
20
0
Ahmedabad
Bengaluru
Chennai
Delhi
Dharamsala
Eden Gardens
Hyderabad
Lucknow
Pune
Wankhede

Investigate how players against different ground in wkts

```
player_wkts=pd.pivot_table(player_ground,values=["wkts"],index=["player"],columns=["ground"])
player_wkts
```

	wkts									
ground	Ahmedabad	Bengaluru	Chennai	Delhi	Dharamsala	Eden Gardens	Hyderabad	Lucknow	Pune	Wankhede
player										
A Dutt (NED)	NaN	0.0	NaN	1.0	0.0	1.0	3.0	3.0	2.0	NaN
A Zampa (AUS)	3.0	4.0	0.0	4.0	3.0	0.0	NaN	5.0	2.0	1.0
AAP Atkinson (ENG)	NaN	NaN	NaN	NaN	NaN	2.0	NaN	NaN	0.0	2.0
AD Mathews (SL)	NaN	4.0	NaN	2.0	NaN	NaN	NaN	NaN	0.0	0.0
AK Markram (SA)	0.0	NaN	0.0	0.0	0.0	1.0	NaN	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...
Usama Mir (PAK)	NaN	1.0	2.0	NaN	NaN	1.0	NaN	NaN	NaN	NaN
V Kohli (IND)	0.0	1.0	0.0	0.0	0.0	0.0	NaN	0.0	0.0	0.0
Vikramjit Singh (NED)	NaN	NaN	NaN	0.0	0.0	0.0	0.0	0.0	NaN	NaN
W Barresi (NED)	NaN	0.0	NaN	NaN	NaN	0.0	NaN	0.0	0.0	NaN
WA Young (NZ)	0.0	NaN	0.0	NaN	0.0	NaN	0.0	NaN	0.0	NaN

152 rows × 10 columns

```
x2=player_wkts.sum(1).sort_values(ascending=False).head(5)
print("Highest five players wkts against ground")
print("-------------------------------------------")
x2
```

```
Highest five players wkts against ground
-------------------------------------------

player
Mohammed Shami (IND)         23.0
A Zampa (AUS)                22.0
D Madushanka (SL)            21.0
G Coetzee (SA)               20.0
Shaheen Shah Afridi (PAK)    18.0
dtype: float64
```

```
player_wkts=pd.pivot_table(player_ground,values=["wkts"],index=["player"],columns=["ground"]).loc[x1.index]
player_wkts
```

	wkts									
ground	Ahmedabad	Bengaluru	Chennai	Delhi	Dharamsala	Eden Gardens	Hyderabad	Lucknow	Pune	Wankhede
player										
R Ravindra (NZ)	1.0	2.0	1.0	NaN	0.0	NaN	1.0	NaN	0.0	0.0
V Kohli (IND)	0.0	1.0	0.0	0.0	0.0	0.0	NaN	0.0	0.0	0.0
GJ Maxwell (AUS)	NaN	0.0	0.0	0.0	1.0	0.0	NaN	3.0	NaN	1.0
BFW de Leede (NED)	NaN	2.0	NaN	2.0	2.0	2.0	5.0	0.0	3.0	NaN
Azmatullah Omarzai (AFG)	0.0	NaN	3.0	0.0	1.0	NaN	NaN	0.0	1.0	2.0

```
player_wkts1=pd.pivot_table(player_ground,values=["wkts"],index=["ground"],columns=["player"])
player_wkts1
```

player	A Dutt (NED)	A Zampa (AUS)	AAP Atkinson (ENG)	AD Mathews (SL)	AK Markram (SA)	AL Phehlukwayo (SA)	AT Carey (AUS)	AT Nidamanuru (NED)	AU Rashid (ENG)	Abdullah Shafique (PAK)	...	TWM Latham (NZ)	Tanzid Hasan (BAN)	Tanzim Hasan Sakib (BAN)	Taskin Ahmed (BAN)	Towhid Hridoy (BAN)	Us
ground																	
Ahmedabad	NaN	3.0	NaN	NaN	0.0	1.0	NaN	NaN	2.0	0.0	...	NaN	NaN	NaN	NaN	NaN	
Bengaluru	0.0	4.0	NaN	4.0	NaN	NaN	NaN	0.0	0.0	0.0	...	0.0	NaN	NaN	NaN	NaN	
Chennai	NaN	0.0	NaN	NaN	0.0	NaN	0.0	NaN	NaN	0.0	...	0.0	0.0	NaN	0.0	0.0	
Delhi	1.0	4.0	NaN	2.0	0.0	NaN	NaN	0.0	3.0	NaN	...	NaN	0.0	3.0	0.0	0.0	
Dharamsala	0.0	3.0	NaN	NaN	0.0	NaN	NaN	0.0	1.0	NaN	...	0.0	0.0	NaN	2.0	0.0	
Eden Gardens	1.0	0.0	2.0	NaN	1.0	NaN	NaN	NaN	2.0	0.0	...	NaN	0.0	NaN	2.0	0.0	
Hyderabad	3.0	NaN	NaN	NaN	NaN	NaN	NaN	0.0	NaN	0.0	...	0.0	NaN	NaN	NaN	NaN	
Lucknow	3.0	5.0	NaN	NaN	0.0	NaN	NaN	0.0	2.0	NaN	...	NaN	NaN	NaN	NaN	NaN	
Pune	2.0	2.0	0.0	0.0	0.0	NaN	NaN	0.0	3.0	NaN	...	0.0	0.0	NaN	1.0	0.0	
Wankhede	NaN	1.0	2.0	0.0	0.0	NaN	NaN	NaN	2.0	NaN	...	0.0	0.0	NaN	NaN	NaN	

10 rows × 152 columns

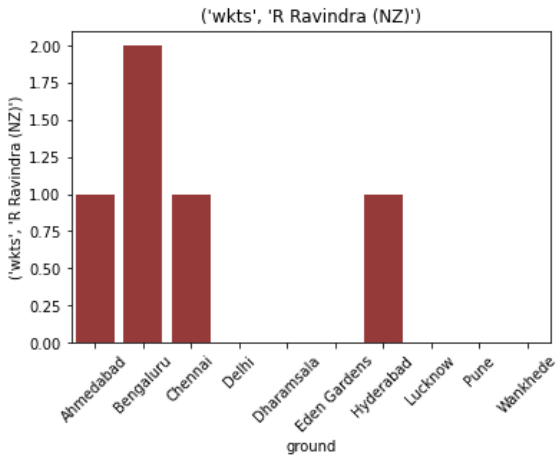
```
# take list to slice data by players name
l1=[]
for i in player_wkts.index:
    l1.append(("wkts",i))
l1
```

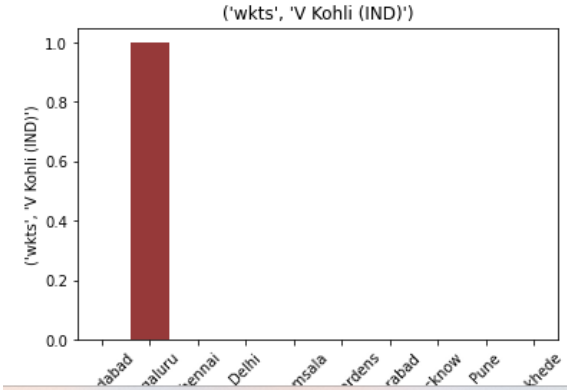
```
[('wkts', 'R Ravindra (NZ)'),
 ('wkts', 'V Kohli (IND)'),
 ('wkts', 'GJ Maxwell (AUS)'),
 ('wkts', 'BFW de Leede (NED)'),
 ('wkts', 'Azmatullah Omarzai (AFG)')]
```

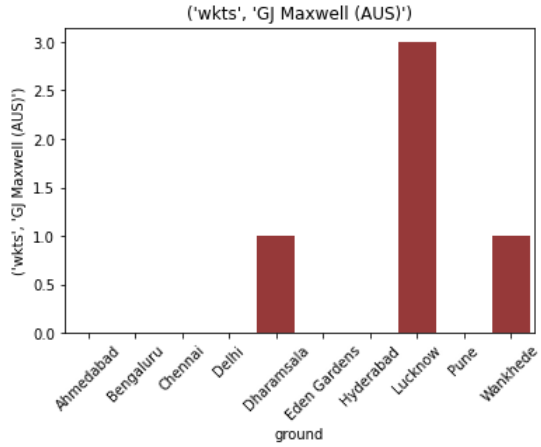
```
player_wkts2=player_wkts1.loc[:,l1]
player_wkts2
```

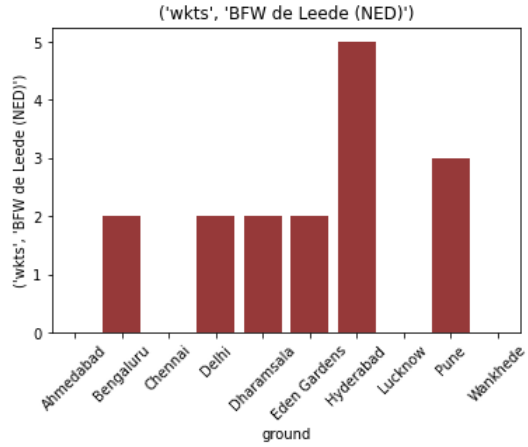
	wkts				
player	R Ravindra (NZ)	V Kohli (IND)	GJ Maxwell (AUS)	BFW de Leede (NED)	Azmatullah Omarzai (AFG)
ground					
Ahmedabad	1.0	0.0	NaN	NaN	0.0
Bengaluru	2.0	1.0	0.0	2.0	NaN
Chennai	1.0	0.0	0.0	NaN	3.0
Delhi	NaN	0.0	0.0	2.0	0.0
Dharamsala	0.0	0.0	1.0	2.0	1.0
Eden Gardens	NaN	0.0	0.0	2.0	NaN
Hyderabad	1.0	NaN	NaN	5.0	NaN
Lucknow	NaN	0.0	3.0	0.0	0.0
Pune	0.0	0.0	NaN	3.0	1.0
Wankhede	0.0	0.0	1.0	NaN	2.0

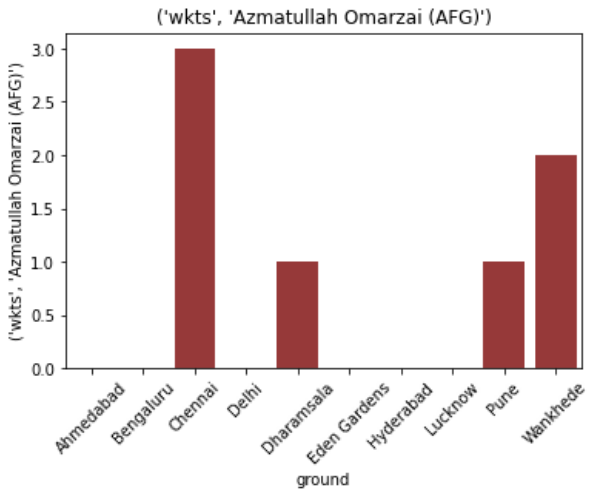
```
for i in player_wkts2:
    sns.barplot(x=player_runs2.index, y=i, data=player_wkts2,color='brown')
    plt.title(f"{i}")
    plt.xticks(rotation=45)
    plt.show()
```











Temporal Analysis:

performance trends over time in runs and wkts for teams

```
team_date= data.groupby(['team','start_date']).agg({
    'runs': 'sum',
    'wkts': 'sum'
}).reset_index().sort_values('start_date')
display(team_date)
```

	team	start_date	runs	wkts
56	NZ	1-Nov-23	508	4.0
75	SA	1-Nov-23	507	10.0
0	AFG	10-Nov-23	473	5.0
76	SA	10-Nov-23	477	9.0
28	ENG	10-Oct-23	580	10.0
...	...	...	...	...
18	AUS	8-Oct-23	386	4.0
64	NZ	9-Nov-23	340	10.0
93	SL	9-Nov-23	339	4.0
55	NED	9-Oct-23	538	7.0
65	NZ	9-Oct-23	531	9.0

```
team_date['start_date'] = pd.to_datetime(team_date['start_date'],format='%d-%b-%y')
team_date1=team_date.sort_values('start_date')
team_date1
```

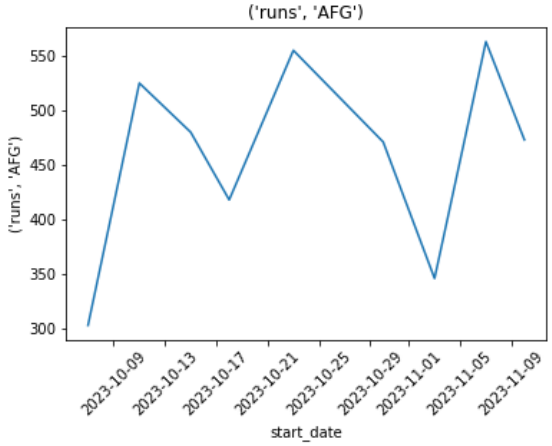
	team	start_date	runs	wkts
63	NZ	2023-10-05	557	9.0
35	ENG	2023-10-05	554	1.0
74	PAK	2023-10-06	482	9.0
53	NED	2023-10-06	482	9.0
92	SL	2023-10-07	733	5.0
...	...	...	...	...
47	NED	2023-11-12	659	5.0
40	IND	2023-11-15	707	10.0
58	NZ	2023-11-15	693	4.0
78	SA	2023-11-16	413	7.0
11	AUS	2023-11-16	406	10.0

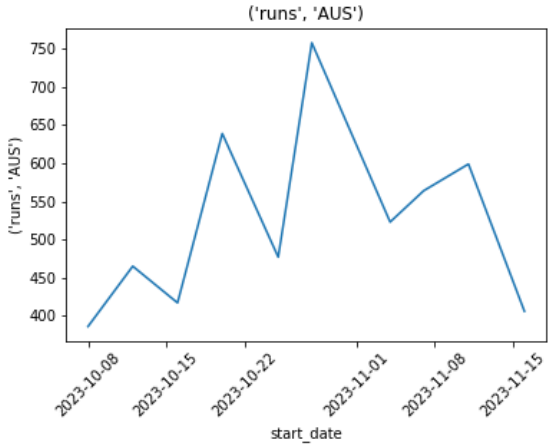
94 rows × 4 columns

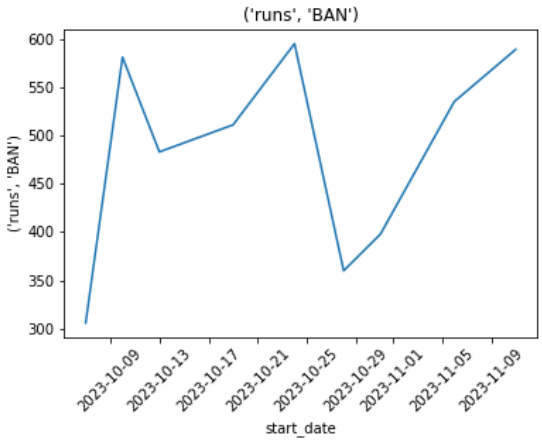
```
team_date1=pd.pivot_table(team_date1,values=["runs"],index=["start_date"],columns=["team"])
team_date1
```

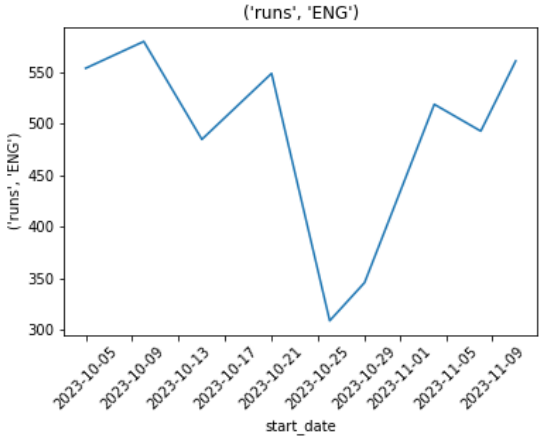
	runs									
team	AFG	AUS	BAN	ENG	IND	NED	NZ	PAK	SA	SL
start_date										
2023-10-05	NaN	NaN	NaN	554.0	NaN	NaN	557.0	NaN	NaN	NaN
2023-10-06	NaN	NaN	NaN	NaN	NaN	482.0	NaN	482.0	NaN	NaN
2023-10-07	303.0	NaN	306.0	NaN	NaN	NaN	NaN	NaN	726.0	733.0
2023-10-08	NaN	386.0	NaN	NaN	386.0	NaN	NaN	NaN	NaN	NaN
2023-10-09	NaN	NaN	NaN	NaN	NaN	538.0	531.0	NaN	NaN	NaN
2023-10-10	NaN	NaN	581.0	580.0	NaN	NaN	NaN	659.0	NaN	675.0
2023-10-11	525.0	NaN	NaN	NaN	525.0	NaN	NaN	NaN	NaN	NaN
2023-10-12	NaN	465.0	NaN	NaN	NaN	NaN	NaN	NaN	471.0	NaN
2023-10-13	NaN	NaN	483.0	NaN	NaN	NaN	481.0	NaN	NaN	NaN
2023-10-14	NaN	NaN	NaN	NaN	378.0	NaN	NaN	378.0	NaN	NaN
2023-10-15	480.0	NaN	NaN	485.0	NaN	NaN	NaN	NaN	NaN	NaN
2023-10-16	NaN	417.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	411.0
2023-10-17	NaN	NaN	NaN	NaN	NaN	418.0	NaN	NaN	434.0	NaN
2023-10-18	418.0	NaN	NaN	NaN	NaN	NaN	417.0	NaN	NaN	NaN
2023-10-19	NaN	NaN	511.0	NaN	512.0	NaN	NaN	NaN	NaN	NaN
2023-10-20	NaN	639.0	NaN	NaN	NaN	NaN	NaN	642.0	NaN	NaN
2023-10-21	NaN	NaN	NaN	549.0	NaN	479.0	NaN	NaN	545.0	495.0
2023-10-22	NaN	NaN	NaN	NaN	534.0	NaN	531.0	NaN	NaN	NaN
2023-10-23	555.0	NaN	NaN	NaN	NaN	NaN	NaN	551.0	NaN	NaN
2023-10-24	NaN	NaN	595.0	NaN	NaN	NaN	NaN	NaN	603.0	NaN
2023-10-25	NaN	477.0	NaN	NaN	NaN	482.0	NaN	NaN	NaN	NaN

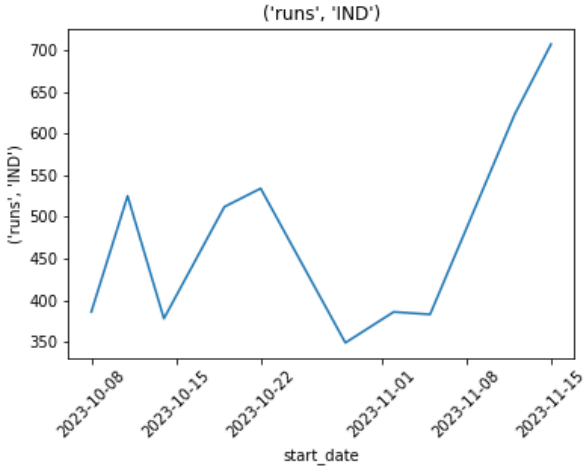
```
for i in team_date1:
    sns.lineplot(x=team_date1.index, y=i, data=team_date1)
    plt.title(f"{i}")
    plt.xticks(rotation=45)
    plt.show()
```

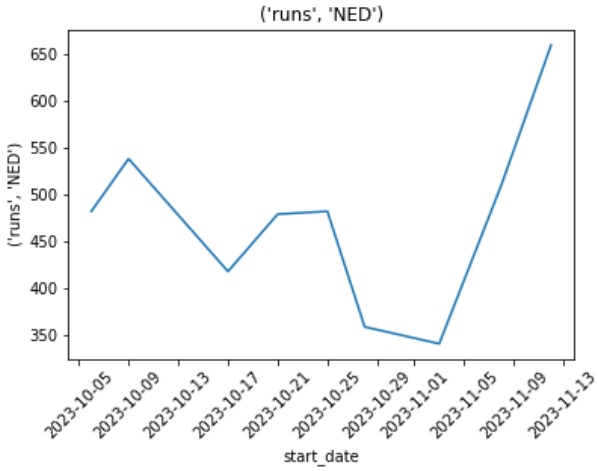


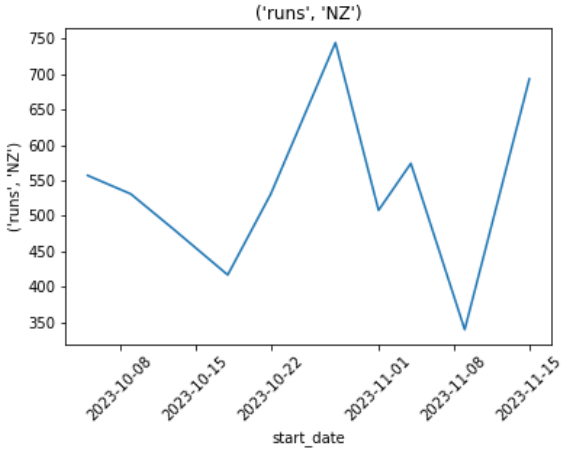


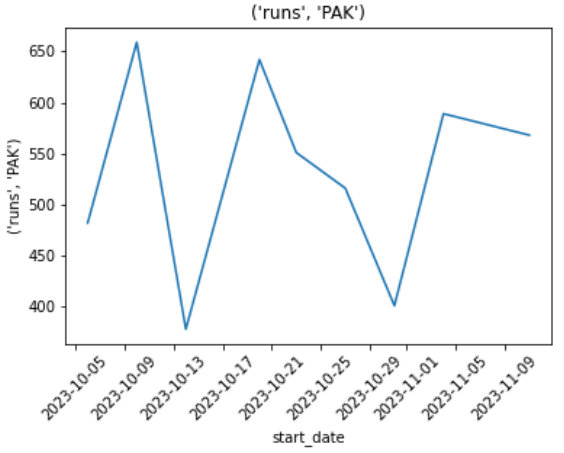


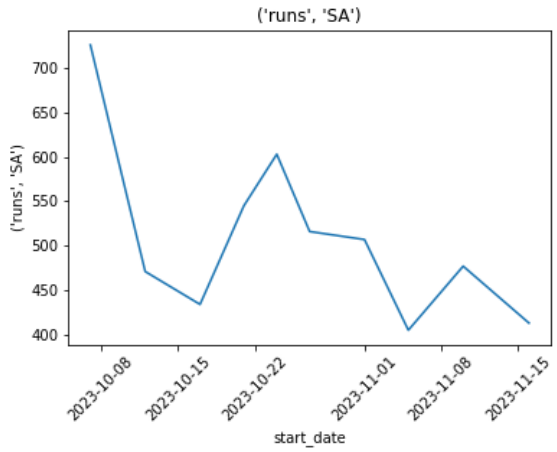


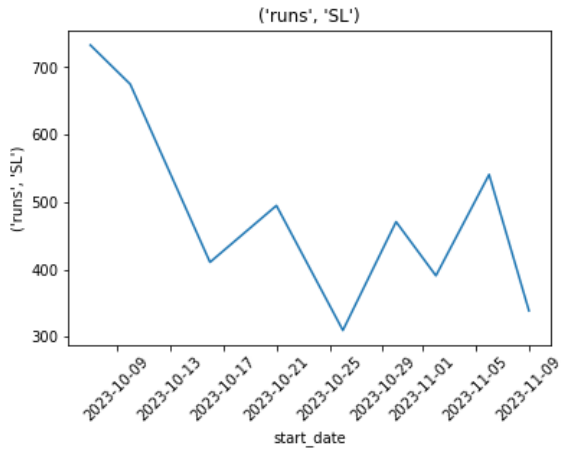

('runs', 'IND')
('runs', 'IND')
700
650
600
550
500
450
400
350
2023-10-08
2023-10-15
2023-10-22
2023-11-01
2023-11-08
2023-11-15
start_date


('runs', 'NED')
('runs', 'NED')
650
600
550
500
450
400
350
2023-10-05
2023-10-09
2023-10-13
2023-10-17
2023-10-21
2023-10-25
2023-10-29
2023-11-01
2023-11-05
2023-11-09
2023-11-13
start_date


('runs', 'NZ')
('runs', 'NZ')
750
700
650
600
550
500
450
400
350
2023-10-08
2023-10-15
2023-10-22
2023-11-01
2023-11-08
2023-11-15
start_date


('runs', 'PAK')
('runs', 'PAK')
650
600
550
500
450
400
2023-10-05
2023-10-09
2023-10-13
2023-10-17
2023-10-21
2023-10-25
2023-10-29
2023-11-01
2023-11-05
2023-11-09
start_date






```
team_date['start_date'] = pd.to_datetime(team_date['start_date'],format='%d-%b-%y')
team_date1=team_date.sort_values('start_date')
team_date1
```

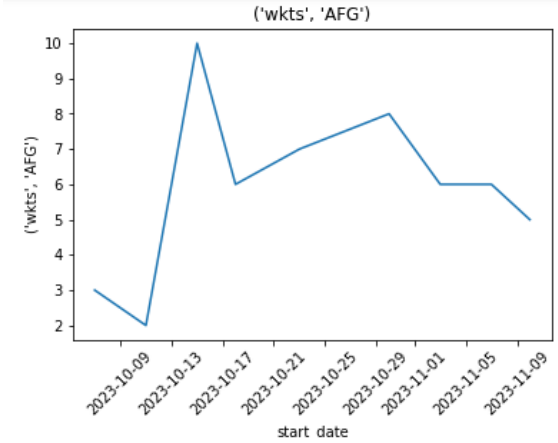
	team	start_date	runs	wkts
63	NZ	2023-10-05	557	9.0
35	ENG	2023-10-05	554	1.0
74	PAK	2023-10-06	482	9.0
53	NED	2023-10-06	482	9.0
92	SL	2023-10-07	733	5.0
...	...	...	...	...
47	NED	2023-11-12	659	5.0
40	IND	2023-11-15	707	10.0
58	NZ	2023-11-15	693	4.0
78	SA	2023-11-16	413	7.0
11	AUS	2023-11-16	406	10.0

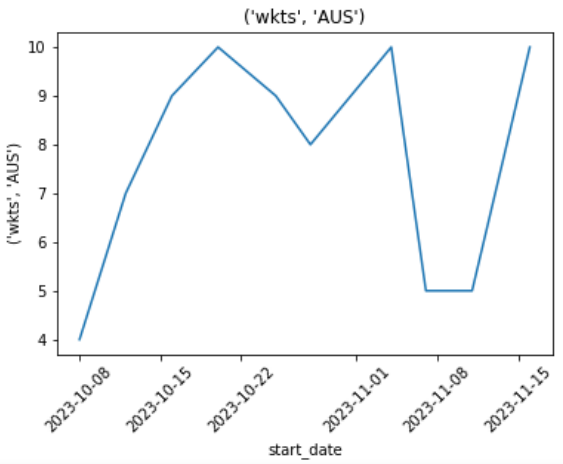
94 rows × 4 columns

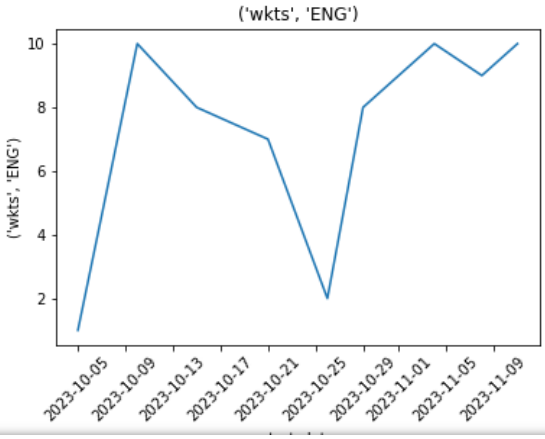
```
team_date1=pd.pivot_table(team_date1,values=["wkts"],index=["start_date"],columns=["team"])
team_date1
```

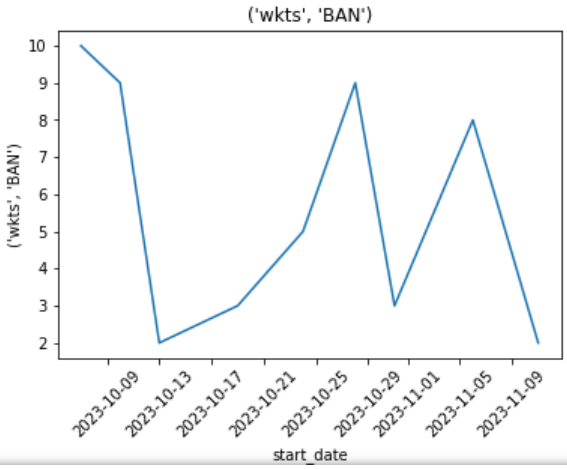
	wkts									
team	AFG	AUS	BAN	ENG	IND	NED	NZ	PAK	SA	SL
start_date										
2023-10-05	NaN	NaN	NaN	1.0	NaN	NaN	9.0	NaN	NaN	NaN
2023-10-06	NaN	NaN	NaN	NaN	NaN	9.0	NaN	9.0	NaN	NaN
2023-10-07	3.0	NaN	10.0	NaN	NaN	NaN	NaN	NaN	10.0	5.0
2023-10-08	NaN	4.0	NaN	NaN	10.0	NaN	NaN	NaN	NaN	NaN
2023-10-09	NaN	NaN	NaN	NaN	NaN	7.0	9.0	NaN	NaN	NaN
2023-10-10	NaN	NaN	9.0	10.0	NaN	NaN	NaN	9.0	NaN	4.0
2023-10-11	2.0	NaN	NaN	NaN	8.0	NaN	NaN	NaN	NaN	NaN
2023-10-12	NaN	7.0	NaN	NaN	NaN	NaN	NaN	NaN	10.0	NaN
2023-10-13	NaN	NaN	2.0	NaN	NaN	NaN	9.0	NaN	NaN	NaN
2023-10-14	NaN	NaN	NaN	NaN	10.0	NaN	NaN	3.0	NaN	NaN
2023-10-15	10.0	NaN	NaN	8.0	NaN	NaN	NaN	NaN	NaN	NaN
2023-10-16	NaN	9.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	4.0
2023-10-17	NaN	NaN	NaN	NaN	NaN	10.0	NaN	NaN	8.0	NaN
2023-10-18	6.0	NaN	NaN	NaN	NaN	NaN	10.0	NaN	NaN	NaN
2023-10-19	NaN	NaN	3.0	NaN	8.0	NaN	NaN	NaN	NaN	NaN
2023-10-20	NaN	10.0	NaN	NaN	NaN	NaN	NaN	9.0	NaN	NaN
2023-10-21	NaN	NaN	NaN	7.0	NaN	5.0	NaN	NaN	9.0	9.0
2023-10-22	NaN	NaN	NaN	NaN	9.0	NaN	5.0	NaN	NaN	NaN
2023-10-23	7.0	NaN	NaN	NaN	NaN	NaN	NaN	2.0	NaN	NaN
2023-10-24	NaN	NaN	5.0	NaN	NaN	NaN	NaN	NaN	10.0	NaN

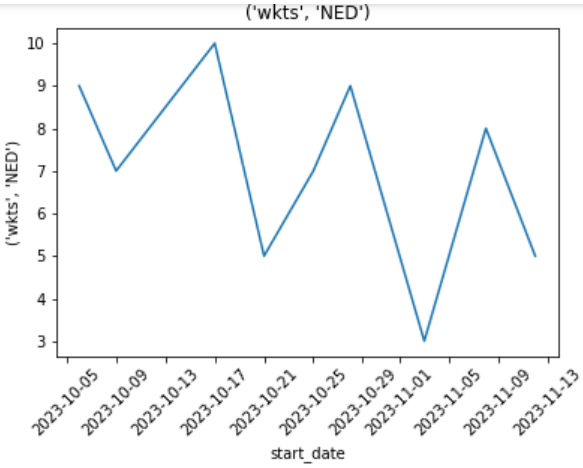
```
for i in team_date1:
    sns.lineplot(x=team_date1.index, y=i, data=team_date1)
    plt.title(f"{i}")
    plt.xticks(rotation=45)
    plt.show()
```

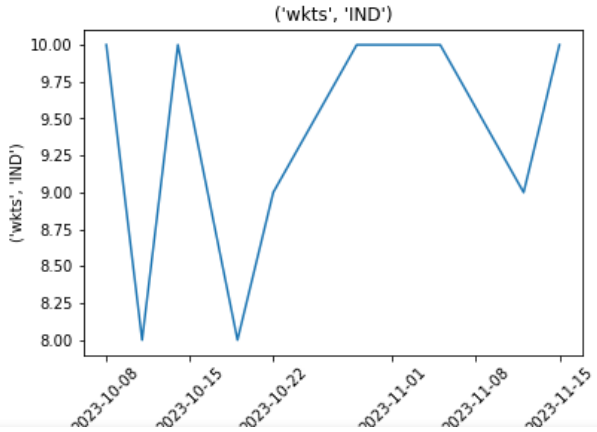


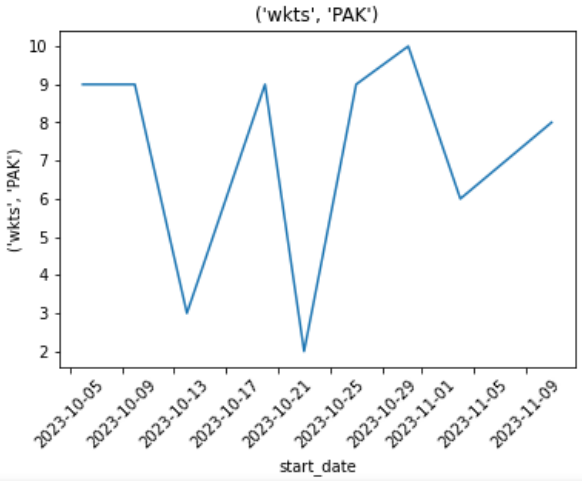

('wkts', 'AUS')
('wkts', 'AUS')
start_date
2023-10-08
2023-10-15
2023-10-22
2023-11-01
2023-11-08
2023-11-15

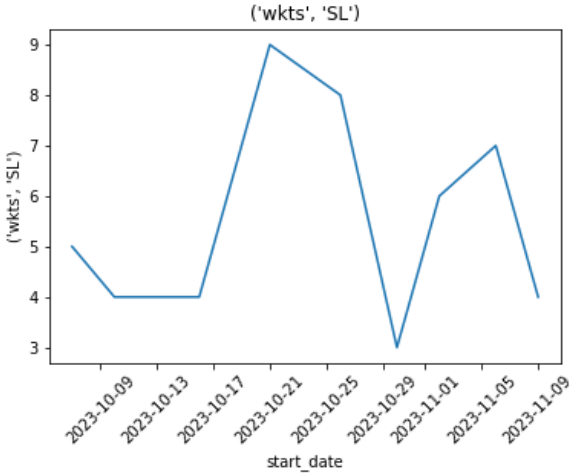

('wkts', 'ENG')
('wkts', 'ENG')
2023-10-05
2023-10-09
2023-10-13
2023-10-17
2023-10-21
2023-10-25
2023-10-29
2023-11-01
2023-11-05
2023-11-09


('wkts', 'BAN')
('wkts', 'BAN')
start_date
2023-10-09
2023-10-13
2023-10-17
2023-10-21
2023-10-25
2023-10-29
2023-11-01
2023-11-05
2023-11-09


('wkts', 'NED')
('wkts', 'NED')
start_date
2023-10-05
2023-10-09
2023-10-13
2023-10-17
2023-10-21
2023-10-25
2023-10-29
2023-11-01
2023-11-05
2023-11-09
2023-11-13


('wkts', 'IND')
('wkts', 'IND')
10.00
9.75
9.50
9.25
9.00
8.75
8.50
8.25
8.00
2023-10-08
2023-10-15
2023-10-22
2023-11-01
2023-11-08
2023-11-15


('wkts', 'PAK')
('wkts', 'PAK')
10
9
8
7
6
5
4
3
2
2023-10-05
2023-10-09
2023-10-13
2023-10-17
2023-10-21
2023-10-25
2023-10-29
2023-11-01
2023-11-05
2023-11-09
start_date


('wkts', 'SL')
('wkts', 'SL')
9
8
7
6
5
4
3
2023-10-09
2023-10-13
2023-10-17
2023-10-21
2023-10-25
2023-10-29
2023-11-01
2023-11-05
2023-11-09
start_date