In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

Reading Data

In [2]:
```
icc=pd.read_csv('kaggle_data.csv')
```

In [3]:
```
icc
```

Out[3]:

	Unnamed: 0	team_1	team_2	stage	Winner_toss	Toss_descision	time
0	Match_1	Australia	SouthAfrica	Group_stage	Australia	Fielding	afternoon
1	Match_2	England	Windies	Group_stage	England	Fielding	evening
2	Match_3	Srilanka	Bangladesh	Group_stage	Srilanka	Fielding	afternoon
3	Match_4	Pakistan	India	Group_stage	Pakistan	Fielding	evening
4	Match_5	Afghanistan	Scotland	Group_stage	Afghanistan	Batting	evening
5	Match_6	SouthAfrica	Windies	Group_stage	SouthAfrica	Fielding	afternoon
6	Match_7	Pakistan	New_Zealand	Group_stage	Pakistan	Fielding	evening
7	Match_8	England	Bangladesh	Group_stage	Bangladesh	Batting	afternoon
8	Match_9	Namibia	ScotLand	Group_stage	Namibia	Fielding	evening
9	Match_10	Australia	Srilanka	Group_stage	Australia	Fielding	evening

Drop Column to remove Unnecessary data.

In [4]:
```
icc.drop(['stage','Toss_descision','avg_temperature','batting_hand','best_batter_
```

Rename Column

In [5]:
```
icc.rename(columns={"Unnamed: 0":"Matches"},inplace=True)
```

- Rename Column for Easy reading of data.

In [6]:
```
icc
```

Out[6]:

ɔest_batter	high_indvidual_scores	strike_rate	target	target_achieved	Player_of_the_match	W
n_Markram	40	111.10	119	1	Josh_Hazlewood	Au
Jos_Buttler	24	109.10	56	1	Moeen_Ali	E
ɳ_Asalanka	80	163.20	172	1	Charith_Asalanka	S
ıad_Rizwan	79	143.60	152	1	Shaheen_shah	Pa
lah_Zadran	59	173.53	191	0	Mujeeb_ur_Rehman	Afgha
Evin_Lewis	56	160.00	144	1	Anrich_Nortje	Soutl
ıad_Rizwan	33	97.06	135	1	Haris_Rauf	Pa
Jason_Roy	61	160.50	125	1	Jason_Roy	E
hael_Leask	44	162.90	110	1	Ruben_Trumpelmann	N
vid_Warner	65	154.76	155	1	Adam_Zampa	Au
Liton_Das	40	102.30	140	1	Nicholas_Pooran	W
abar_Azam	51	108.50	148	1	Asif_Ali	Pa
ɳ_Nissanka	72	124.14	143	1	Tabraiz_Shamsi	Soutl
Jos_Buttler	71	221.80	126	1	Chris_Jordan	E
d_Shahzad	45	136.36	161	0	Naveen_ul_Haq	Afgha
ryl_Mitchell	49	140.00	111	1	Trent_Boult	New_Z
Jos_Buttler	101	150.75	164	0	Jos_Buttler	E
ɔa_Bavuma	31	110.70	85	1	Kagiso_Rabada	Soutl
ıad_Rizwan	79	158.00	190	0	Muhammad_Rizwan	Pa
ırtin_Guptill	93	166.07	173	0	Martin_Guptill	New_Z
hit_Sharma	74	157.45	211	0	Rohit_Sharma	
ıaron_Finch	40	200.00	74	1	Adam_Zampa	Au
ɔn_Hetmyer	81	150.00	190	0	Chrith_Adalanka	S
ɛnn_Phillips	39	185.71	164	0	James_Neesham	New_Z
KL_Rahul	50	263.16	86	1	Ravindra_Jadeja	
vid_Warner	89	158.93	158	1	David_Warner	Au
ɜsie van der Dussen	94	156.67	190	0	Rassie_van_der_Dussen	Soutl
lah_Zadran	73	152.08	125	1	Trent_Boult	New_Z
abar_Azam	66	140.43	190	0	Shoaib_Malik	Pa
hit_Sharma	56	151.35	133	1	Ravindra_Jadeja	
ryl_Mitchell	72	153.19	167	1	Daryl_Mitchell	New_Z
ıad_Rizwan	67	128.85	177	1	Matthew_Wade	Au
_Williamson	85	177.08	173	1	Mitchell_Marsh	Au

Top 5 Match Of Teams.

In [7]:
```
icc.head()
```

Out[7]:

	Matches	team_1	team_2	Winner_toss	time	venue	best_bowler	bowling
0	Match_1	Australia	SouthAfrica	Australia	afternoon	Abu_Dhabi	Josh_Hazlewood	
1	Match_2	England	Windies	England	evening	Dubai	Adil_Rashid	
2	Match_3	Srilanka	Bangladesh	Srilanka	afternoon	Sharjah	Shakib_al_Hassan	
3	Match_4	Pakistan	India	Pakistan	evening	Dubai	Shaheen_shah	
4	Match_5	Afghanistan	Scotland	Afghanistan	evening	Sharjah	Mujeeb_ur_Rehman	

Bottom 5 Match Of Teams.

In [8]:
```
icc.tail()
```

Out[8]:

	Matches	team_1	team_2	Winner_toss	time	venue	best_bowler	bowl
28	Match_29	Pakistan	ScotLand	Pakistan	evening	Sharjah	Shadab_Khan	
29	Match_30	India	Namibia	India	evening	Dubai	Ravindra_Jadeja	
30	Match_31	New_Zealand	England	New_Zealand	evening	Abu_Dhabi	Liam_Livingstone	
31	Match_32	Australia	Pakistan	Australia	evening	Dubai	Shadab_Khan	
32	Match_33	Australia	New_Zealand	Australia	evening	Dubai	Josh_Hazlewood	

Number of Rows and Columns

In [9]:
```
icc.shape
```

Out[9]:
```
(33, 19)
```

Columns in Dataset

In [10]:
```
icc.columns
```

Out[10]:
```
Index(['Matches', 'team_1', 'team_2', 'Winner_toss', 'time', 'venue',
       'best_bowler', 'bowling_arm', 'bowling_style',
       'most_individual_wickets', 'economy', 'best_bowler_country',
       'best_batter', 'high_indvidual_scores', 'strike_rate', 'target',
       'target_achieved', 'Player_of_the_match', 'Winner'],
      dtype='object')
```

Null Values in Dataset

In [11]:
```
icc.isnull().sum()
```

Out[11]:
```
Matches                    0
team_1                     0
team_2                     0
Winner_toss                0
time                       0
venue                      0
best_bowler                0
bowling_arm                0
bowling_style              0
most_individual_wickets    0
economy                    0
best_bowler_country        0
best_batter                0
high_indvidual_scores      0
strike_rate                0
target                     0
target_achieved            0
Player_of_the_match        0
Winner                     0
dtype: int64
```

Dataset Information.

In [12]:
```
icc.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 33 entries, 0 to 32
Data columns (total 19 columns):
 #   Column                   Non-Null Count  Dtype  
---  ------                   --------------  -----  
 0   Matches                  33 non-null     object 
 1   team_1                   33 non-null     object 
 2   team_2                   33 non-null     object 
 3   Winner_toss              33 non-null     object 
 4   time                     33 non-null     object 
 5   venue                    33 non-null     object 
 6   best_bowler              33 non-null     object 
 7   bowling_arm              33 non-null     object 
 8   bowling_style            33 non-null     object 
 9   most_individual_wickets  33 non-null     int64  
 10  economy                  33 non-null     float64
 11  best_bowler_country      33 non-null     object 
 12  best_batter              33 non-null     object 
 13  high_indvidual_scores    33 non-null     int64  
 14  strike_rate              33 non-null     float64
 15  target                   33 non-null     int64  
 16  target_achieved          33 non-null     int64  
 17  Player_of_the_match      33 non-null     object 
 18  Winner                   33 non-null     object 
dtypes: float64(2), int64(4), object(13)
memory usage: 5.0+ KB
```

Statistical Information

In [13]:
```
icc.describe()
```

Out[13]:

	most_individual_wickets	economy	high_indvidual_scores	strike_rate	target	target_ac
count	33.000000	33.000000	33.000000	33.000000	33.000000	33.
mean	2.878788	5.215758	62.393939	152.402121	146.606061	0.
std	0.892944	2.187966	20.135385	33.999070	37.136521	0.
min	1.000000	0.900000	24.000000	97.060000	56.000000	0.
25%	2.000000	4.000000	45.000000	136.360000	125.000000	0.
50%	3.000000	4.800000	65.000000	153.190000	152.000000	1.
75%	3.000000	6.250000	79.000000	162.900000	173.000000	1.
max	5.000000	12.000000	101.000000	263.160000	211.000000	1.

In [96]:
```
icc.duplicated().sum()
```

Out[96]:
```
0
```

In [14]:
```
icc.nunique()
```

Out[14]:
```
Matches                    33
team_1                     10
team_2                     13
Winner_toss                11
time                        2
venue                       3
best_bowler                26
bowling_arm                 2
bowling_style               5
most_individual_wickets     5
economy                    24
best_bowler_country        11
best_batter                23
high_indvidual_scores      28
strike_rate                33
target                     27
target_achieved             2
Player_of_the_match        30
Winner                     10
dtype: int64
```

Statistical Figures

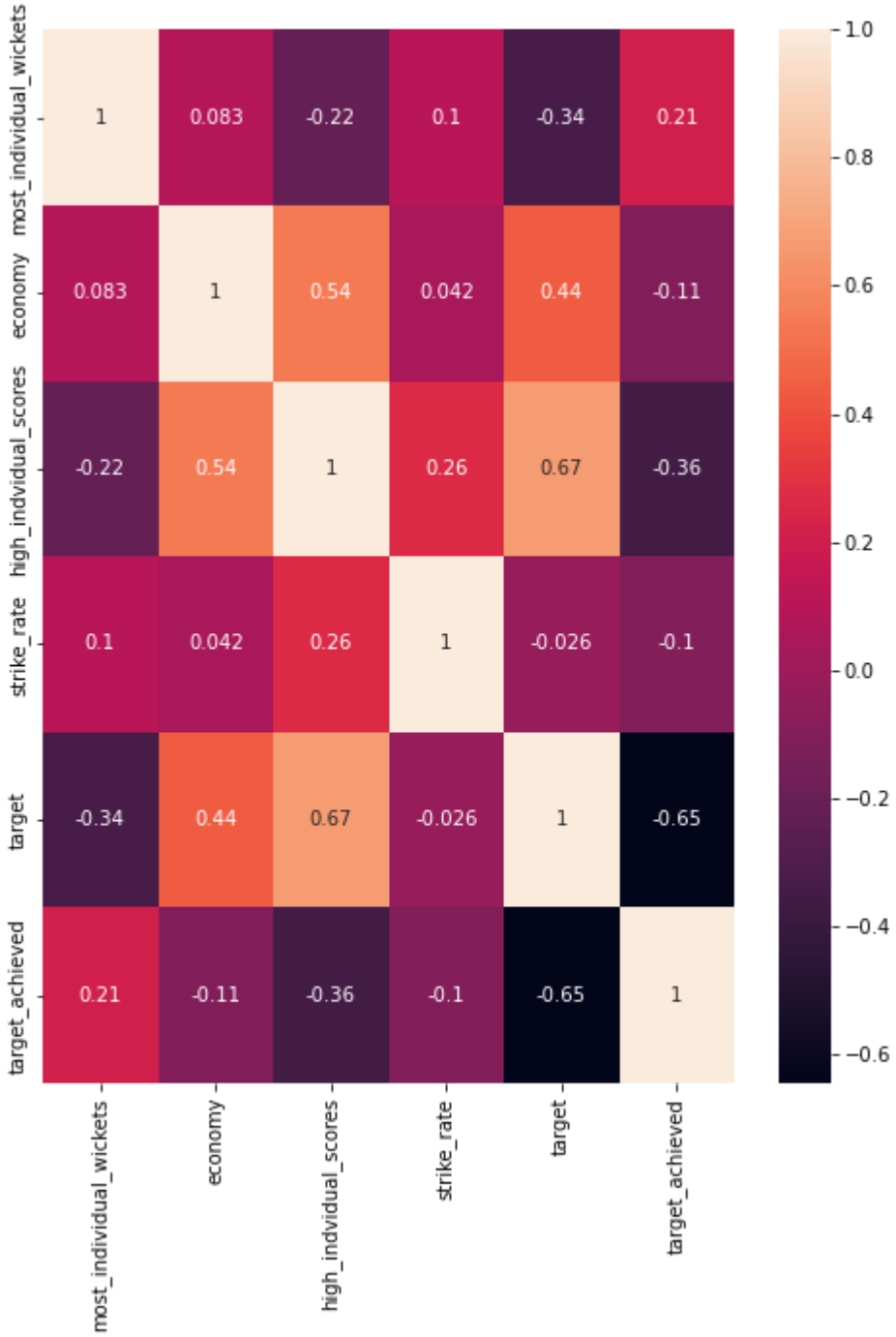
In [15]:
```
icc.corr()
```

Out[15]:

	most_individual_wickets	economy	high_indvidual_scores	strike_rate	t
most_individual_wickets	1.000000	0.083063	-0.216257	0.102263	-0.33
economy	0.083063	1.000000	0.540763	0.042015	0.43
high_indvidual_scores	-0.216257	0.540763	1.000000	0.256137	0.66
strike_rate	0.102263	0.042015	0.256137	1.000000	-0.02
target	-0.335086	0.439099	0.668378	-0.025931	1.00
target_achieved	0.209059	-0.111473	-0.356031	-0.100342	-0.64

Heatmap:

In [16]:
```
icc_corr=icc.corr()
plt.figure(figsize=(8,10))
sns.heatmap(icc_corr,annot=True)
plt.show()
```



Heatmap Plot Info:

- In the plot it shows the correlation between the all attributes in the dataset.
- By the observation we see the relation between the order quantity and cost of the product is very strong.

In [17]:
```
icc.cov()
```

Out[17]:

	most_individual_wickets	economy	high_indvidual_scores	strike_rate	
most_individual_wickets	0.797348	0.162282	-3.888258	3.104640	
economy	0.162282	4.787194	23.823598	3.125419	
high_indvidual_scores	-3.888258	23.823598	405.433712	175.347263	4
strike_rate	3.104640	3.125419	175.347263	1155.936792	·
target	-11.111742	35.678277	499.785038	-32.740076	13
target_achieved	0.087121	-0.113826	-3.345644	-1.592150	

Statistical Figures

In [18]:
```
icc['team_2'].count()
```

Out[18]: 33

In [19]:
```
icc['team_1'].count()
```

Out[19]: 33

In [20]:
```
icc['target'].mean()
```

Out[20]: 146.6060606060606

In [21]:
```
icc['target'].max()
```

Out[21]: 211

In [22]:
```
icc['target'].min()
```

Out[22]: 56

- Count: The count of team 1 & team 2 is 33
- Mean: Average of the target score by all over the tournament.
- Max: Maximum target by individual team.
- Min: Minimum target by individual team.

Count Number Of Winner

In [40]:
```
icc['Winner'].value_counts()
```

Out[40]:
```
Australia      6
Pakistan       5
New_Zealand    5
England        4
SouthAfrica    4
India          3
Srilanka       2
Afghanistan    2
Namibia        1
Windies        1
Name: Winner, dtype: int64
```

Pivot Table

In [76]:
```
cb_team1=pd.pivot_table(data=icc[['Matches','team_1','team_2']],
                        index=['team_1','team_2'],
                        values='Matches',
                        aggfunc='max')
cb_team1
```

Out[76]:

		Matches
team_1	**team_2**	
Afghanistan	**Namibia**	Match_15
	Scotland	Match_5
Australia	**Bangladesh**	Match_22
	New_Zealand	Match_33
	Pakistan	Match_32
	SouthAfrica	Match_1
	Srilanka	Match_10
	Windies	Match_26
England	**Australia**	Match_14
	Bangladesh	Match_8
	Srilanka	Match_17
	Windies	Match_2
India	**Afghanistan**	Match_21
	Namibia	Match_30
	ScotLand	Match_25
Namibia	**ScotLand**	Match_9
New_Zealand	**Afghanistan**	Match_28
	England	Match_31
	India	Match_16
	Namibia	Match_24
	ScotLand	Match_20
Pakistan	**Afghanistan**	Match_12
	India	Match_4
	Namibia	Match_19
	New_Zealand	Match_7
	ScotLand	Match_29
SouthAfrica	**Bangladesh**	Match_18
	England	Match_27
	Srilanka	Match_13
	Windies	Match_6

		Matches
team_1	**team_2**	
Srilanka	**Bangladesh**	Match_3
	Windies	Match_23
Windies	**Bangladesh**	Match_11

- This pivot table shows the matches played between team_1 and team_2.

In [23]:
```
cb_team=pd.pivot_table(data=icc[['team_1','team_2','target']],
                       index=['team_1','team_2'],
                       values='target',
                       aggfunc='max')
cb_team
```

Out[23]:

		target
team_1	**team_2**	
Afghanistan	**Namibia**	161
	Scotland	191
Australia	**Bangladesh**	74
	New_Zealand	173
	Pakistan	177
	SouthAfrica	119
	Srilanka	155
	Windies	158
England	**Australia**	126
	Bangladesh	125
	Srilanka	164
	Windies	56
India	**Afghanistan**	211
	Namibia	133
	ScotLand	86
Namibia	**ScotLand**	110
New_Zealand	**Afghanistan**	125
	England	167
	India	111
	Namibia	164
	ScotLand	173
Pakistan	**Afghanistan**	148
	India	152
	Namibia	190
	New_Zealand	135
	ScotLand	190
SouthAfrica	**Bangladesh**	85
	England	190
	Srilanka	143
	Windies	144

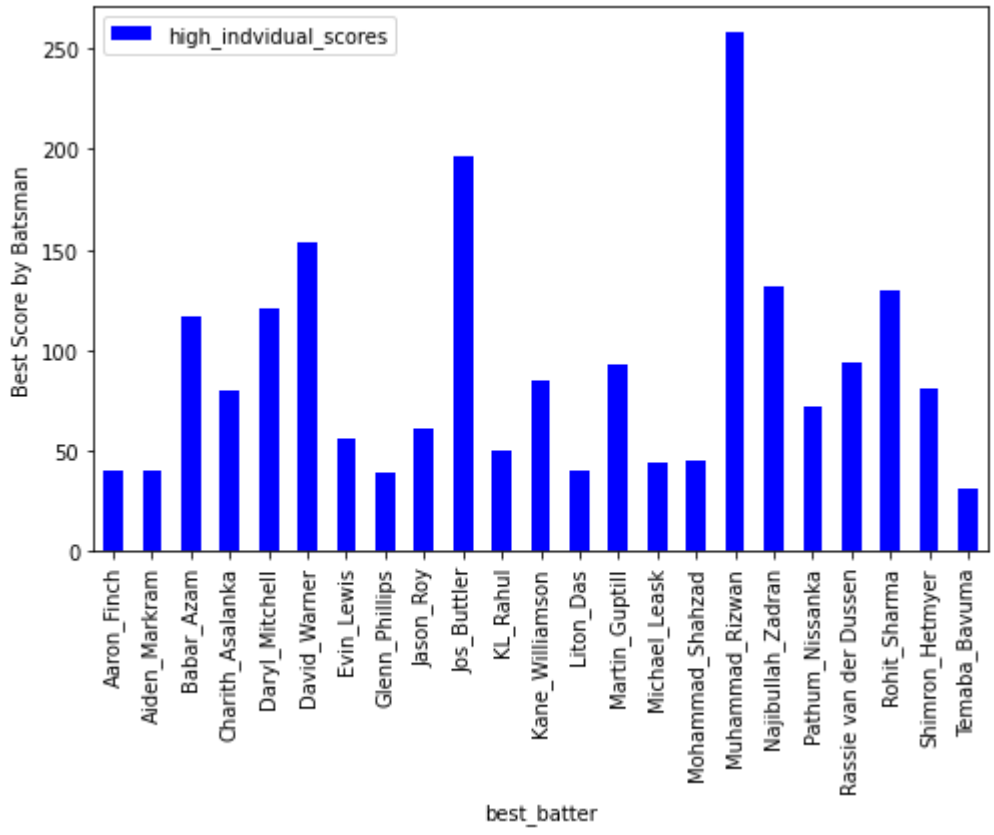
		target
team_1	team_2	
Srilanka	Bangladesh	172
	Windies	190
Windies	Bangladesh	140

- In this pivot it shows the target by the teams in the tournament.

Bargraph:

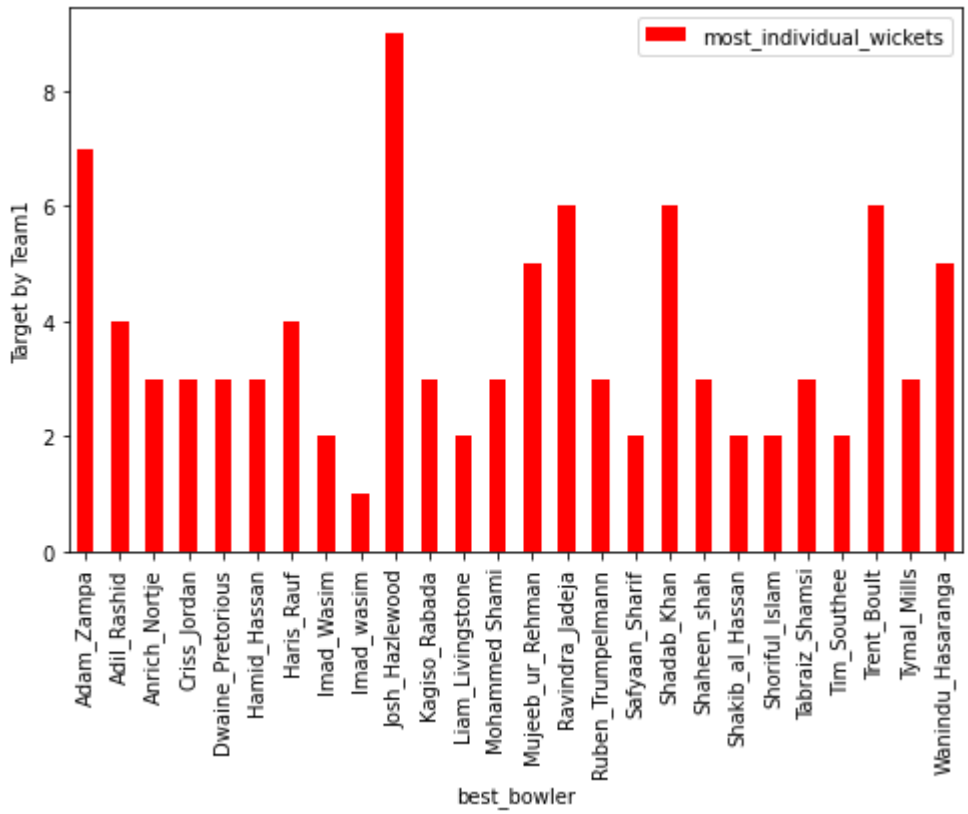
- To show Target score by Team 1 & Team 2.

In [92]:
```
icc.groupby('best_batter')[['high_indvidual_scores']].sum().plot.bar(color=['blue
plt.ylabel('Best Score by Batsman')
plt.show()
```



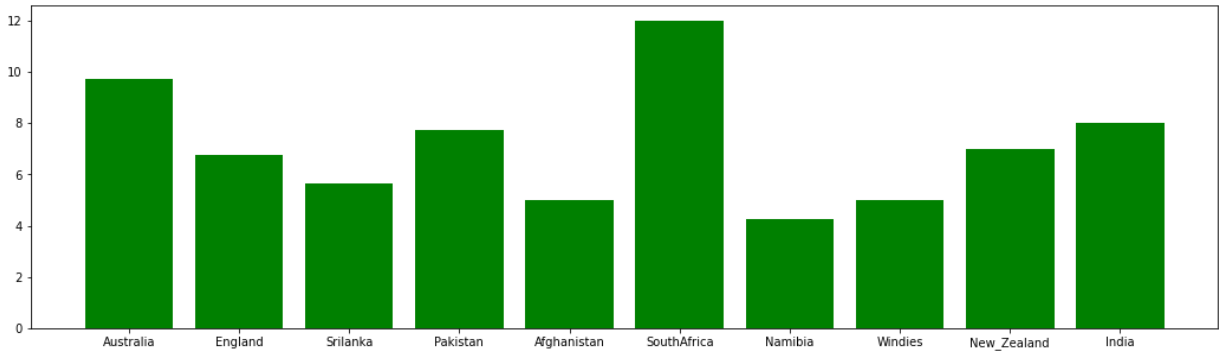
- In the given plot we observe that the Highest Score from Najibullah_Zadran from Team Afganistan.

In [93]:
```
icc.groupby('best_bowler')[['most_individual_wickets']].sum().plot.bar(color=['r'
plt.ylabel('Target by Team1')
plt.show()
```



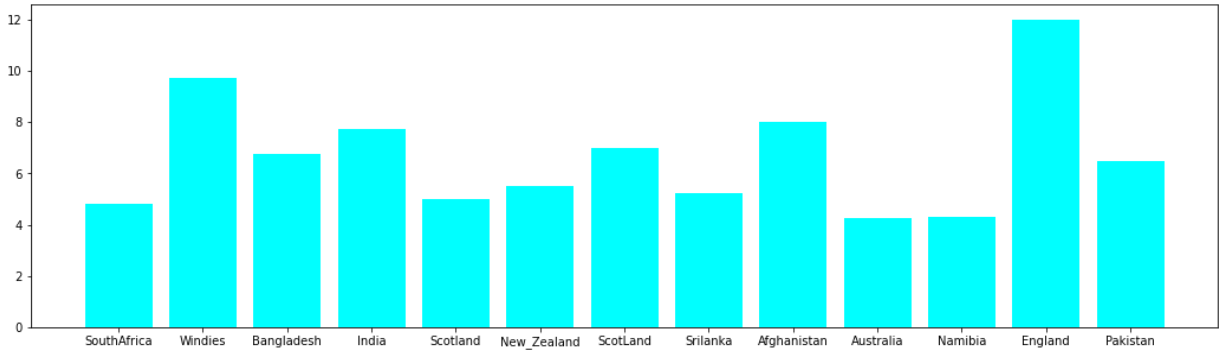
- In the given plot we observe that the Highest Wicket taken from Josh Hazlewood from Team Australlia.

In [52]:
```
plt.figure(figsize=(18,5))
plt.bar('team_1','economy',data=icc,color='g')
plt.show()
```



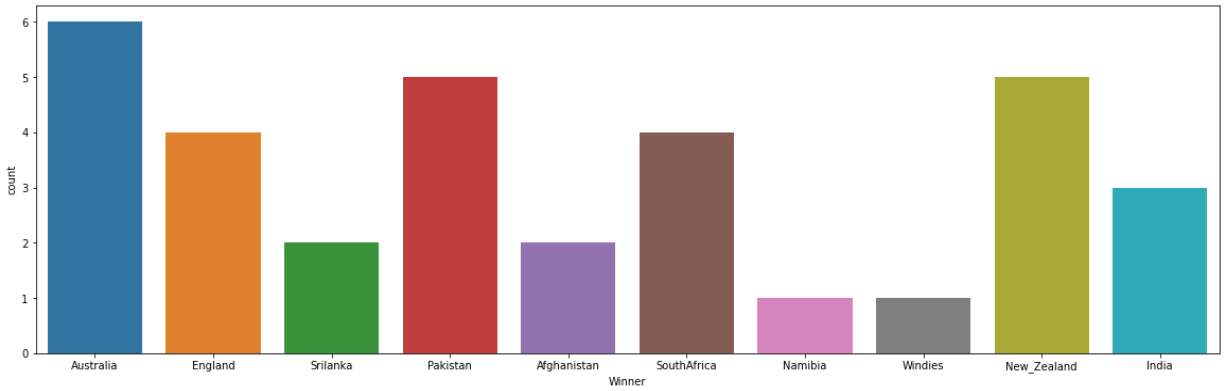
- SouthAfrica has the highest economy in the tournament with reference of the Score they made allover the tournament.

In [55]:
```
plt.figure(figsize=(18,5))
plt.bar('team_2','economy',data=icc,color='cyan')
plt.show()
```



- England has the highest economy in the tournament with reference of the Score they made allover the tournament.

In [67]:
```
plt.figure(figsize=(20,6))
sns.countplot(x=icc['Winner'])
plt.xticks(rotation=0)
plt.show()
```



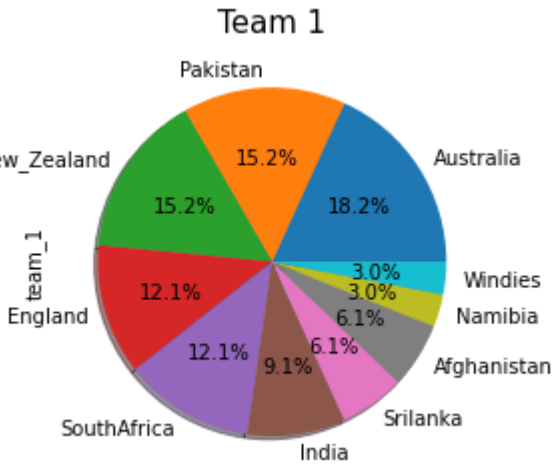
- Match Winning of the Teams in the tournament.
- Australia Wins the highest number of matches.

In [ ]:
```
```

Pie Charts

- In the given pie chart it shows the score distribution of the teams.

In [61]:
```
plt.title('Team 1',fontsize=15)
icc['team_1'].value_counts().plot.pie(autopct='%1.1f%%',shadow=True)
plt.show()
```



- Australia has the highest score portion achieved.

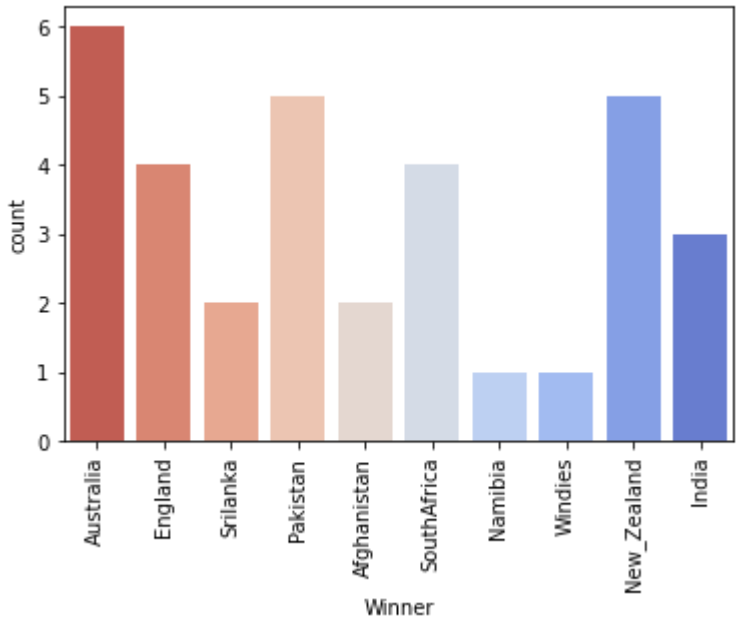
In [62]:
```
plt.title('Team 2',fontsize=15)
icc['team_2'].value_counts().plot.pie(autopct='%1.1f%%',shadow=True)
plt.show()
```

- Bangladesh has the highest score portion achieved.

Count Plot

- Count the number of match winning by the individual team in the tournament.

In [73]:
```
sns.countplot(x='Winner',data=icc,palette='coolwarm_r')
plt.xticks(rotation=90)
plt.show()
```

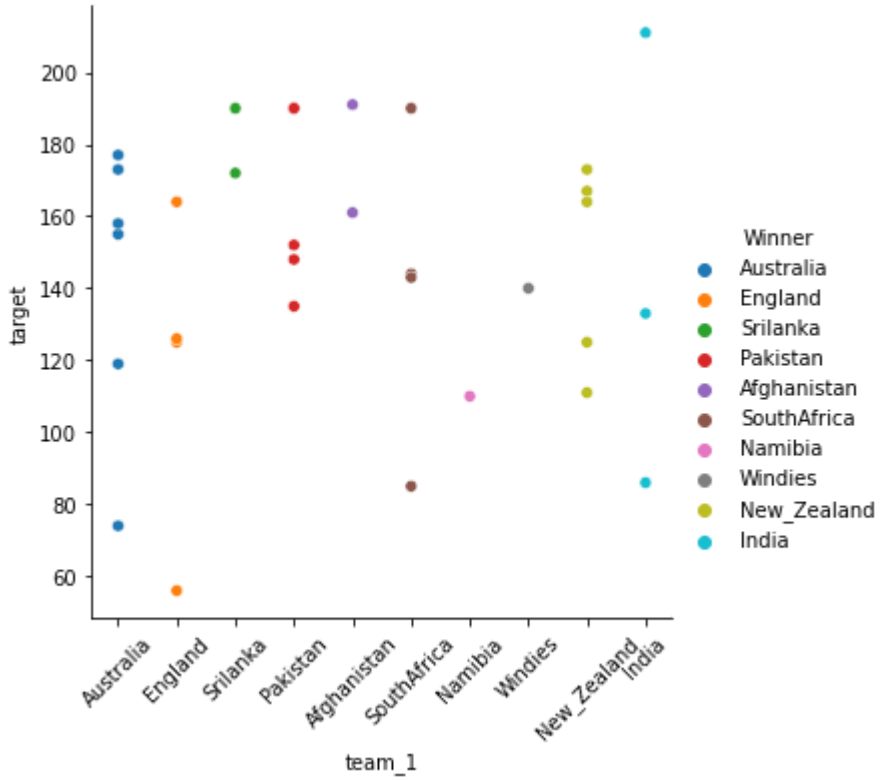


- In the given count plot we observe Australia Win the maximum matches according to the plot.

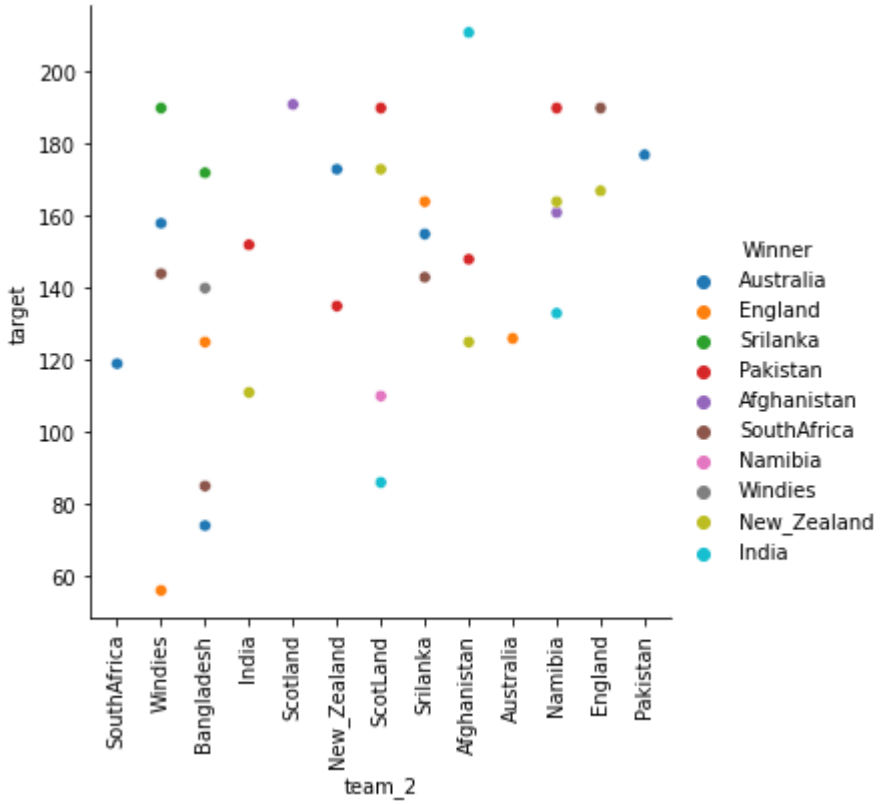
Rel plot

- It shows the team_1's Countrys target with the help of dots.

In [83]:
```
sns.relplot(x='team_1',y='target',hue='Winner',data=icc)
plt.xticks(rotation=45)
plt.show()
```



In [90]:
```
sns.relplot(x='team_2',y='target',hue='Winner',data=icc)
plt.xticks(rotation=90)
plt.show()
```



Conclusion

- In ICC World Cup 2021 there are Around 16 Teams compete for the World Cup.
- Cricket Format T20 and Tournament Format is of Group Stage and Knockout.
- Total Matches held are 33 Matches of the team to qualify and the 12 matches are of qualifiers so there are total 45 Matches conducted in the tournament.
- Champion Team is 'Australia' and Runner Up is 'NewZealand'.

In [ ]: