ODI CWC 2023 Bowler's Performance Data Analysis

Dataset is downloaded from kaggle,using pandas libray. This dataset is based on the statistics of world education,it provides huge data regarding the students from all around the world according to all countries, it tells their literacy situation according to age groups and other different criterias.This course really helped me a lot as now I can easily manipulate datasets and I have a good grasp on basic data analysis concepts as well as using python as a data analysis tool. Data Analysis with Python: Zero to Pandas.

How to run the code

This is an executable *Jupyter notebook* hosted on Jovian.ml, a platform for sharing data science projects. You can run and experiment with the code in a couple of ways: *using free online resources* (recommended) or *on your own computer*.

Option 1: Running using free online resources (1-click, recommended)

The easiest way to start executing this notebook is to click the "Run" button at the top of this page, and select "Run on Binder". This will run the notebook on mybinder.org, a free online service for running Jupyter notebooks. You can also select "Run on Colab" or "Run on Kaggle".

Option 2: Running on your computer locally

1. Install Conda by following these instructions. Add Conda binaries to your system `PATH`, so you can use the `conda` command on your terminal.
2. Create a Conda environment and install the required libraries by running these commands on the terminal:

```
conda create -n zerotopandas -y python=3.8
conda activate zerotopandas
pip install jovian jupyter numpy pandas matplotlib seaborn opendatasets --upgrade
```

3. Press the "Clone" button above to copy the command for downloading the notebook, and run it on the terminal. This will create a new directory and download the notebook. The command will look something like this:

```
jovian clone notebook-owner/notebook-id
```

4. Enter the newly created directory using `cd directory-name` and start the Jupyter notebook.

```
jupyter notebook
```

You can now access Jupyter's web interface by clicking the link that shows up on the terminal or by visiting http://localhost:8888 on your browser. Click on the notebook file (it has a `.ipynb` extension) to open it.

Downloading the Dataset

Dataset is in .csv format and is downloaded from this url: https://www.kaggle.com/datasets/nelgiriyewithana/world-educational-data by using **opendatasets** library

```
!pip install jovian opendatasets --upgrade --quiet
```

Let's begin by downloading the data, and listing the files within the dataset.

```
dataset_url = 'https://www.kaggle.com/datasets/vikramrn/icc-mens-cricket-odi-world-cup-
```

```
import opendatasets as od
od.download(dataset_url)
```

```
Please provide your Kaggle credentials to download this dataset. Learn more:
http://bit.ly/kaggle-creds
Your Kaggle username: adeelshah41
Your Kaggle Key: ·····
Downloading icc-mens-cricket-odi-world-cup-wc-2023-bowling.zip to ./icc-mens-cricket-
odi-world-cup-wc-2023-bowling
100%|████████████████████| 9.01k/9.01k [00:00<00:00, 7.52MB/s]
```

The dataset has been downloaded and extracted.

```
data_dir = './bowlers-odi-cwc-data'
```

```
import os
os.listdir(data_dir)
```

```
['icc_wc_23_bowl.csv']
```

Let us save and upload our work to Jovian before continuing.

```
project_name = "Bowler's performance analysis in men's odi cricket world cups " # chang
```

```
!pip install jovian --upgrade -q
```

```
import jovian
```

```
jovian.commit(project=project_name)
```

```
[jovian] Creating a new project "sadeel823/Bowler's performance analysis in men's odi
cricket world cups "
[jovian] Committed successfully! https://jovian.com/sadeel823/bowlers-performance-
analysis-in-mens-odi-cricket-world-cups-51f0a
```

```
'https://jovian.com/sadeel823/bowlers-performance-analysis-in-mens-odi-cricket-world-
cups-51f0a'
```

Data Preparation and Cleaning

Analysis on Bowler's performance in odi cwc 2023 - Data provided is already in clean form, I have displayed the dataset, its weaknesses if any and its description as well.

```
import pandas as pd
```

```
#loading the dataset into a dataframe using Pandas
df=pd.read_csv('./bowlers-odi-cwc-data/icc_wc_23_bowl.csv')
```

```
df
```

	player	overs	maidens	runs	wickets	run_rate	0s	4s	6s	wd	nb	team	opponent	innings	match
0	Dilshan Madushanka	10.0	0	80	5	8.00	25	8	2	6	0	Sri Lanka	India	1	
1	Dushmantha Chameera	10.0	2	71	1	7.10	28	8	2	0	0	Sri Lanka	India	1	
2	Kasun Rajitha	9.0	0	65	0	7.22	25	7	2	1	0	Sri Lanka	India	1	
3	Angelo Mathews	3.0	0	11	0	3.66	10	1	0	0	0	Sri Lanka	India	1	
4	Maheesh Theekshana	10.0	0	67	0	6.70	18	5	1	1	0	Sri Lanka	India	1	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
569	David Willey	10.0	0	56	3	5.60	37	8	1	2	0	England	Pakistan	2	
570	Chris Woakes	5.3	0	27	1	4.90	22	5	0	0	0	England	Pakistan	2	
571	Adil Rashid	10.0	0	55	2	5.50	26	4	1	1	0	England	Pakistan	2	
572	Gus Atkinson	8.0	0	45	2	5.62	29	3	3	2	0	England	Pakistan	2	
573	Moeen Ali	10.0	0	60	2	6.00	27	7	1	0	0	England	Pakistan	2	

574 rows × 15 columns

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 574 entries, 0 to 573
Data columns (total 15 columns):
 #   Column    Non-Null Count  Dtype  
---  ------    --------------  -----  
 0   player    574 non-null    object 
 1   overs     574 non-null    float64
 2   maidens   574 non-null    int64  
 3   runs      574 non-null    int64  
 4   wickets   574 non-null    int64  
 5   run_rate  574 non-null    float64
 6   0s        574 non-null    int64  
```

```
 7   4s         574 non-null    int64
 8   6s         574 non-null    int64
 9   wd         574 non-null    int64
 10  nb         574 non-null    int64
 11  team       574 non-null    object
 12  opponent   574 non-null    object
 13  innings    574 non-null    int64
 14  match_id   574 non-null    int64
dtypes: float64(2), int64(10), object(3)
memory usage: 67.4+ KB
```

```
df.dtypes
```

```
player       object
overs       float64
maidens       int64
runs          int64
wickets       int64
run_rate    float64
0s            int64
4s            int64
6s            int64
wd            int64
nb            int64
team         object
opponent     object
innings       int64
match_id      int64
dtype: object
```

```
df.shape
```

```
(574, 15)
```

```
df.describe()
```

	overs	maidens	runs	wickets	run_rate	0s	4s	6s	
count	574.000000	574.000000	574.000000	574.000000	574.000000	574.000000	574.000000	574.000000	574.
mean	7.351045	0.256098	42.322300	1.202091	5.922369	22.918118	3.900697	1.121951	1.
std	2.685601	0.533933	19.651473	1.193757	2.135935	10.251561	2.599953	1.302075	1.
min	0.300000	0.000000	1.000000	0.000000	1.350000	0.000000	0.000000	0.000000	0.
25%	6.000000	0.000000	30.000000	0.000000	4.500000	16.000000	2.000000	0.000000	0.
50%	8.000000	0.000000	41.000000	1.000000	5.660000	24.000000	4.000000	1.000000	1.
75%	10.000000	0.000000	55.000000	2.000000	7.075000	30.000000	5.000000	2.000000	2.
max	10.000000	3.000000	115.000000	7.000000	16.000000	48.000000	15.000000	6.000000	10.

```
#for missing values
missing_values_df=df.isna()
```

```
missing_values_df.sum() #if the sum returns zero for a column then that means it has nc
```

```
player      0
overs       0
maidens     0
runs        0
wickets     0
run_rate    0
0s          0
4s          0
6s          0
wd          0
nb          0
team        0
opponent    0
innings     0
match_id    0
dtype: int64
```

```
import jovian
```

```
jovian.commit()
```

```
[jovian] Updating notebook "sadeel823/bowlers-performance-analysis-in-mens-odi-cricket-world-cups-51f0a" on https://jovian.com
[jovian] Committed successfully! https://jovian.com/sadeel823/bowlers-performance-analysis-in-mens-odi-cricket-world-cups-51f0a
```

```
'https://jovian.com/sadeel823/bowlers-performance-analysis-in-mens-odi-cricket-world-cups-51f0a'
```

Exploratory Analysis and Visualization

Analysis and visualization on the bowling performance in 2023 odi cricket world cup

Let's begin by importing `matplotlib.pyplot` and `seaborn` .

```
import seaborn as sns
import matplotlib
import matplotlib.pyplot as plt
%matplotlib inline

sns.set_style('darkgrid')
matplotlib.rcParams['font.size'] = 14
```

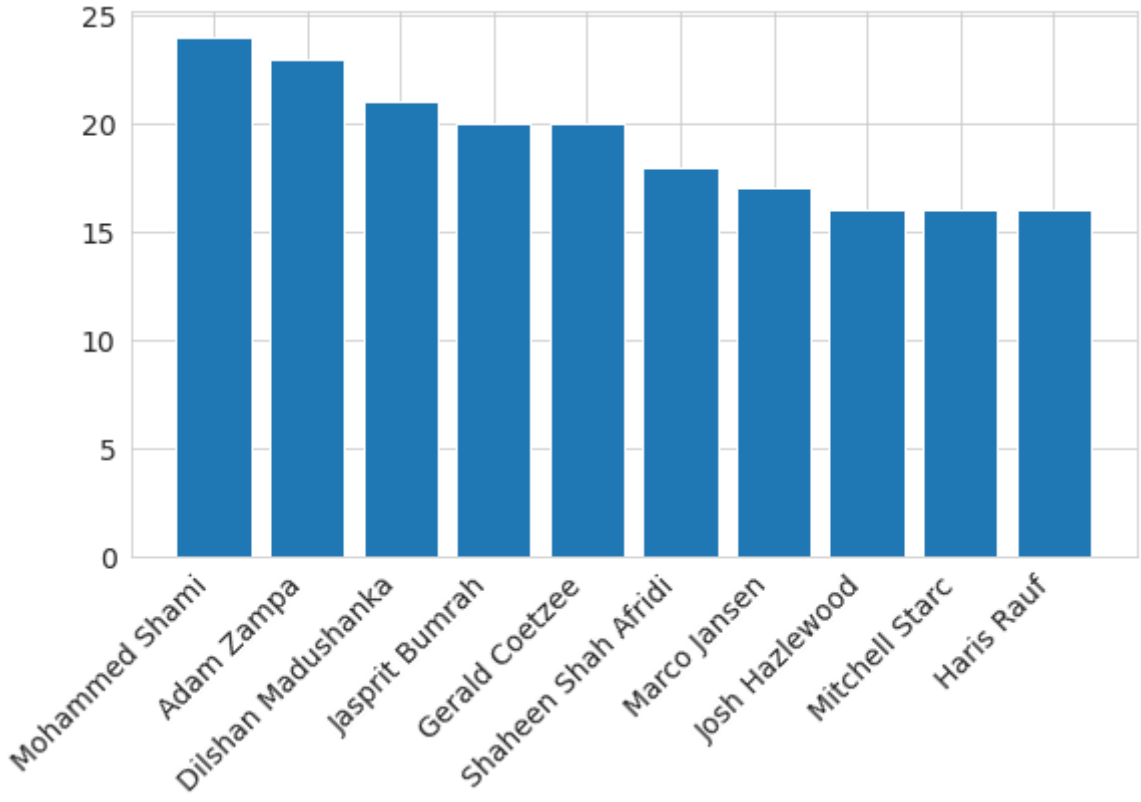
```
matplotlib.rcParams['figure.figsize'] = (9, 5)
matplotlib.rcParams['figure.facecolor'] = '#00000000'
```

Top 10 highest wicket taking bowlers by using a plot

```
df.describe()
```

	overs	maidens	runs	wickets	run_rate	0s	4s	6s	
count	574.000000	574.000000	574.000000	574.000000	574.000000	574.000000	574.000000	574.000000	574.
mean	7.351045	0.256098	42.322300	1.202091	5.922369	22.918118	3.900697	1.121951	1.
std	2.685601	0.533933	19.651473	1.193757	2.135935	10.251561	2.599953	1.302075	1.
min	0.300000	0.000000	1.000000	0.000000	1.350000	0.000000	0.000000	0.000000	0.
25%	6.000000	0.000000	30.000000	0.000000	4.500000	16.000000	2.000000	0.000000	0.
50%	8.000000	0.000000	41.000000	1.000000	5.660000	24.000000	4.000000	1.000000	1.
75%	10.000000	0.000000	55.000000	2.000000	7.075000	30.000000	5.000000	2.000000	2.
max	10.000000	3.000000	115.000000	7.000000	16.000000	48.000000	15.000000	6.000000	10.

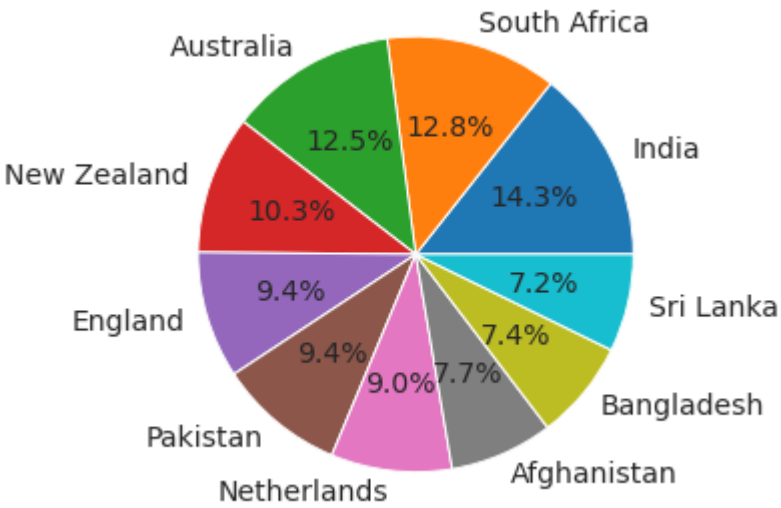
```
df2=df.groupby('player').sum()
df3=df2.sort_values('wickets',ascending=False)
df3.reset_index(inplace=True)

# Specify the new position for the 'Name' column
new_position = 0  # Change this to the desired position

# Insert the 'Name' column at the new position
df3.insert(new_position, 'player', df3.pop('player'))
df4=df3.head(10)
sns.set_style("whitegrid")
plt.bar(df4.player,df4.wickets)
plt.xticks(rotation=45,ha='right')
plt.show()
```



The team with most wickets in world cup

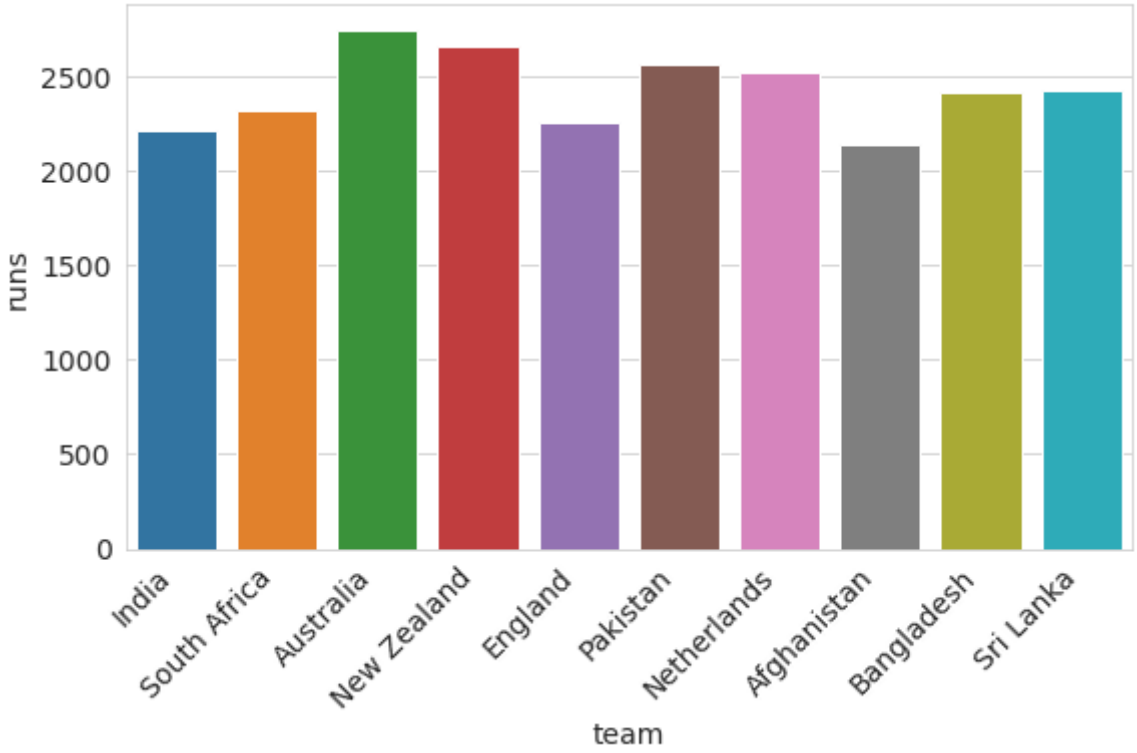
```
df5=df.groupby('team').sum()
df6=df5.sort_values('wickets',ascending=False)
df6.reset_index(inplace=True)
df6.insert(0, 'team', df6.pop('team'))
df7=df6.head(10)
```

```
fig, ax = plt.subplots()
ax.pie(df7.wickets, labels=df7.team,autopct='%1.1f%%')
plt.show()
```



Total runs conceded by each team

```
sns.barplot(x='team', y='runs', data=df7);
plt.xticks(rotation=45,ha='right')
plt.show()
```



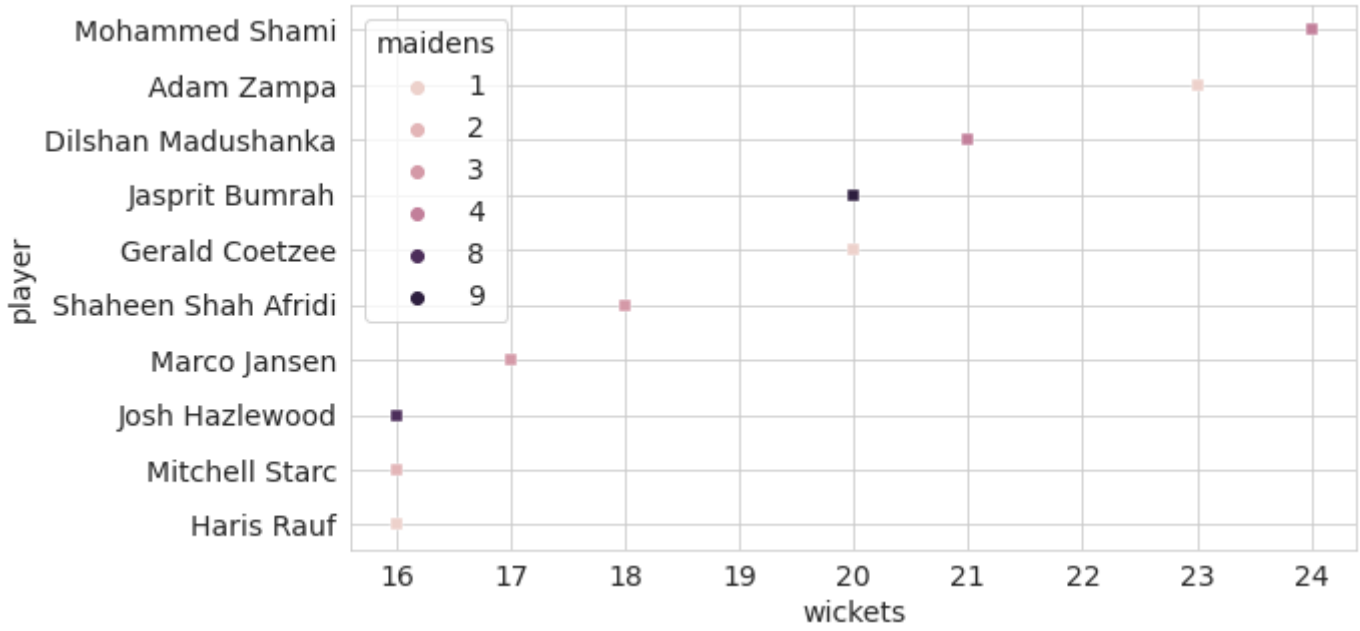
Number of Dot balls bowled by each team

```
sns.barplot(x='team', y='0s', data=df7);
plt.xticks(rotation=45,ha='right')
plt.show()
```

```

```

Top 10 bowlers as per their wickets and their maiden overs

```
sns.scatterplot(x='wickets',y='player',marker='s',hue='maidens',data=df3.head(10))
plt.show()
```



Let us save and upload our work to Jovian before continuing

```
import jovian
```

```
jovian.commit()
```

```
[jovian] Updating notebook "sadeel823/bowlers-performance-analysis-in-mens-odi-cricket-world-cups-51f0a" on https://jovian.com
[jovian] Committed successfully! https://jovian.com/sadeel823/bowlers-performance-analysis-in-mens-odi-cricket-world-cups-51f0a
```

```
'https://jovian.com/sadeel823/bowlers-performance-analysis-in-mens-odi-cricket-world-cups-51f0a'
```

Asking and Answering Questions

Now I will make some question marks and answer through data analysis on the provided dataset.

Q1: Which 5 bowlers bowled the most maiden overs in the odi cricket world cup 2023?

```
df_1=df.groupby('player').sum()
df_2=df_1.sort_values('maidens',ascending=False)
df_2.reset_index(inplace=True)

df_2.insert(0, 'player', df_2.pop('player'))
df_2.head()
```

```
plt.bar(df_2.head().player,df_2.head().maidens)
plt.xticks(rotation=45,ha='right')
plt.show()
```

Q2: What was the overall performance by Pakistani bowlers in world cup 2023 ?

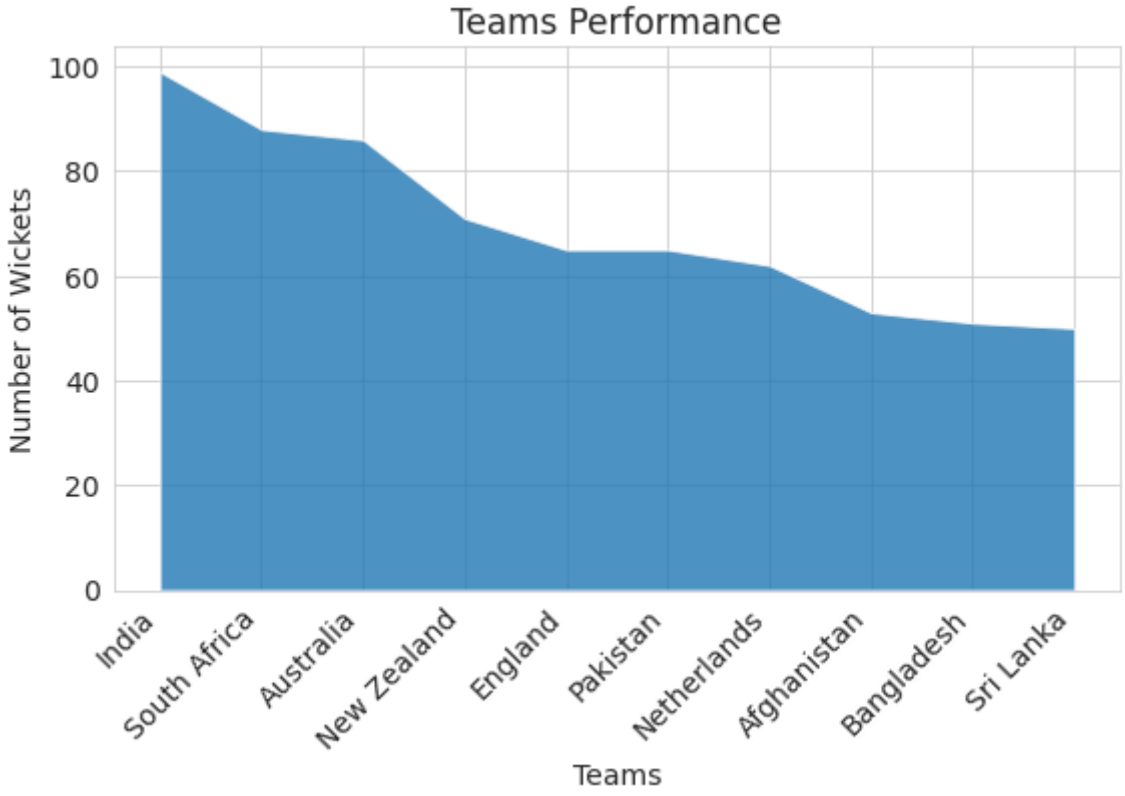
```
df7.iloc[5]
```

```
df_Pakistan=df.groupby(df.team=='Pakistan').sum()
df_Pakistan.iloc[1]
```

Q3: Which team's bowlers dominated the world cup 2023 ?

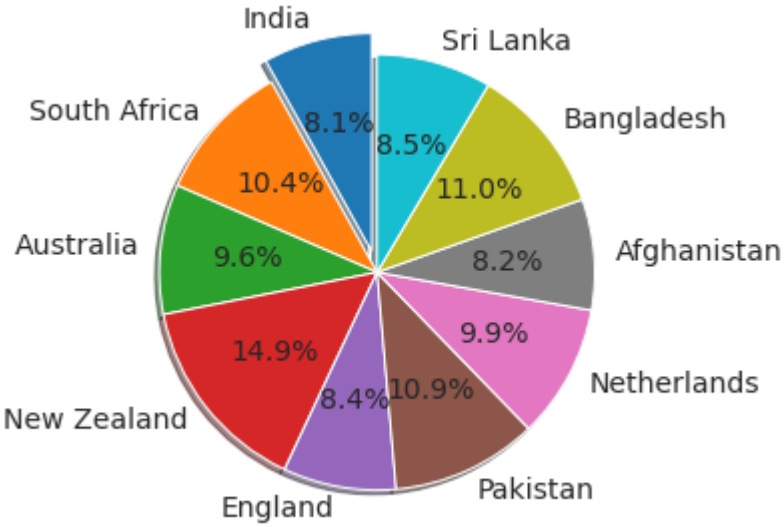
```
sns.barplot(x='team', y='0s', data=df7);
plt.xticks(rotation=45,ha='right')
plt.show()
#Teams ranked by the no. of dot balls they have bowled
```

```
fig, ax = plt.subplots()

ax.stackplot(df7['team'], df7['wickets'], alpha=0.8)

ax.set_title('Teams Performance')
ax.set_xlabel('Teams')
ax.set_ylabel('Number of Wickets')
plt.xticks(rotation=45,ha='right')

plt.show()
```




```
explode = (0.1, 0, 0, 0,0,0,0,0,0,0)
labels=df7.team
fig, ax = plt.subplots()
ax.pie(df7['6s'], labels=labels,explode=explode, autopct='%1.1f%%',
       shadow=True, startangle=90)
```

```
plt.show()
#Teams ranked by the no. of sixes they have conceded
```



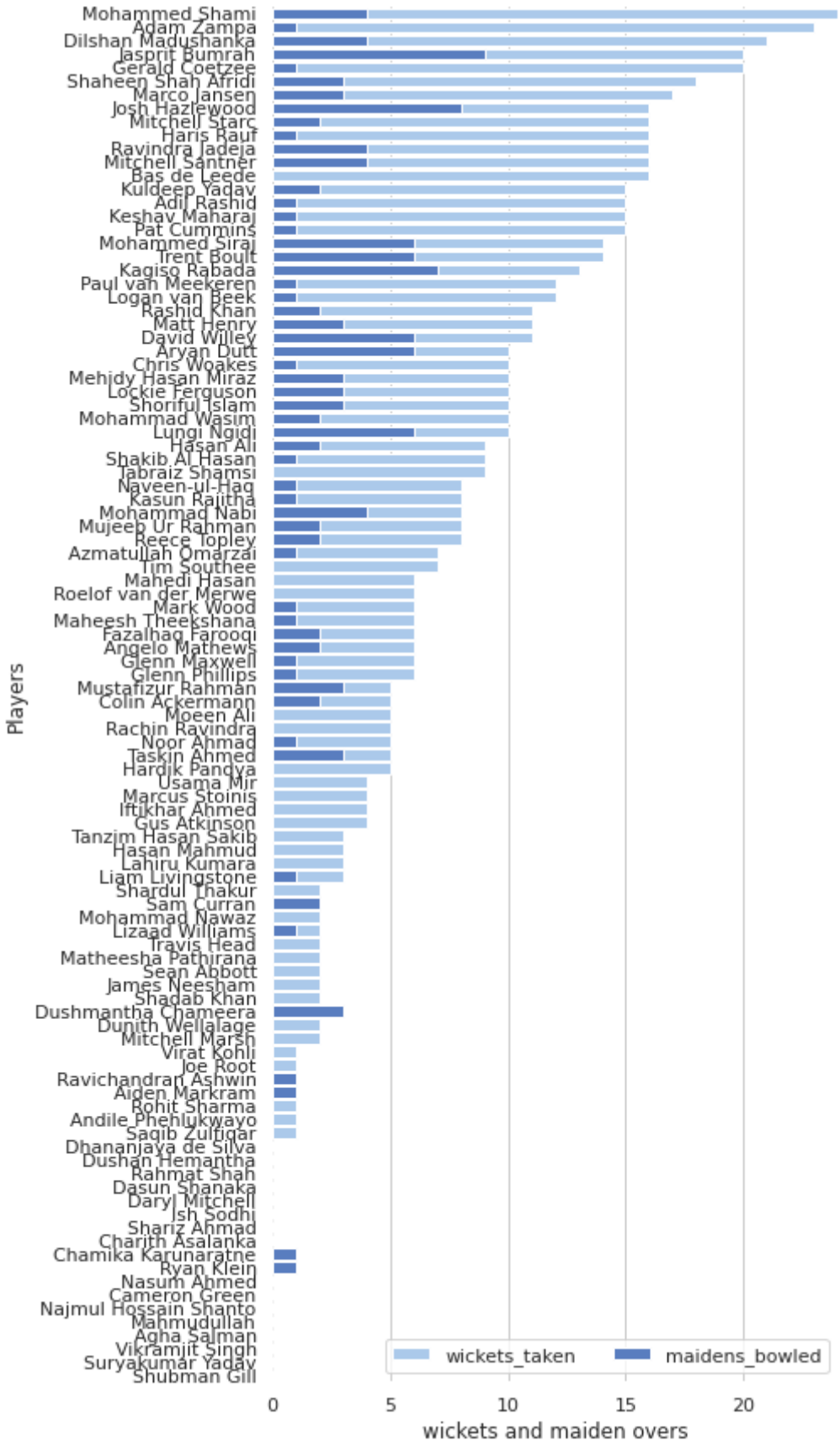
Q4: Which bowlers came out on top in world cup 2023?

```
df_1=df.groupby('player').sum()
df_2=df_1.sort_values('wickets',ascending=False)
df_2.reset_index(inplace=True)

df_2.insert(0, 'player', df_2.pop('player'))
df_2.head(10)
```

	player	overs	maidens	runs	wickets	run_rate	0s	4s	6s	wd	nb	innings	match_id
0	Mohammed Shami	48.5	4	257	24	35.97	183	27	7	9	0	13	103
1	Adam Zampa	96.0	1	515	23	58.20	261	33	14	9	0	16	240
2	Dilshan Madushanka	77.8	4	525	21	61.08	241	63	8	23	3	14	171
3	Jasprit Bumrah	91.5	9	373	20	42.62	372	38	6	22	0	17	205
4	Gerald Coetzee	63.3	1	396	20	51.54	214	47	11	10	2	13	182
5	Shaheen Shah Afridi	81.0	3	481	18	53.06	273	55	10	22	0	13	285
6	Marco Jansen	68.6	3	450	17	58.58	259	59	10	36	5	15	206
7	Josh Hazlewood	93.1	8	449	16	52.16	334	49	8	13	0	16	240
8	Mitchell Starc	87.0	2	528	16	60.82	287	49	13	41	0	15	218
9	Haris Rauf	79.0	1	533	16	60.92	227	50	16	18	0	13	285

```
sns.set_theme(style="whitegrid")

# Initialize the matplotlib figure
f, ax = plt.subplots(figsize=(6, 15))

# Load the example car crash dataset
crashes = df_2.sort_values("wickets", ascending=False)

# Plot the total crashes
```

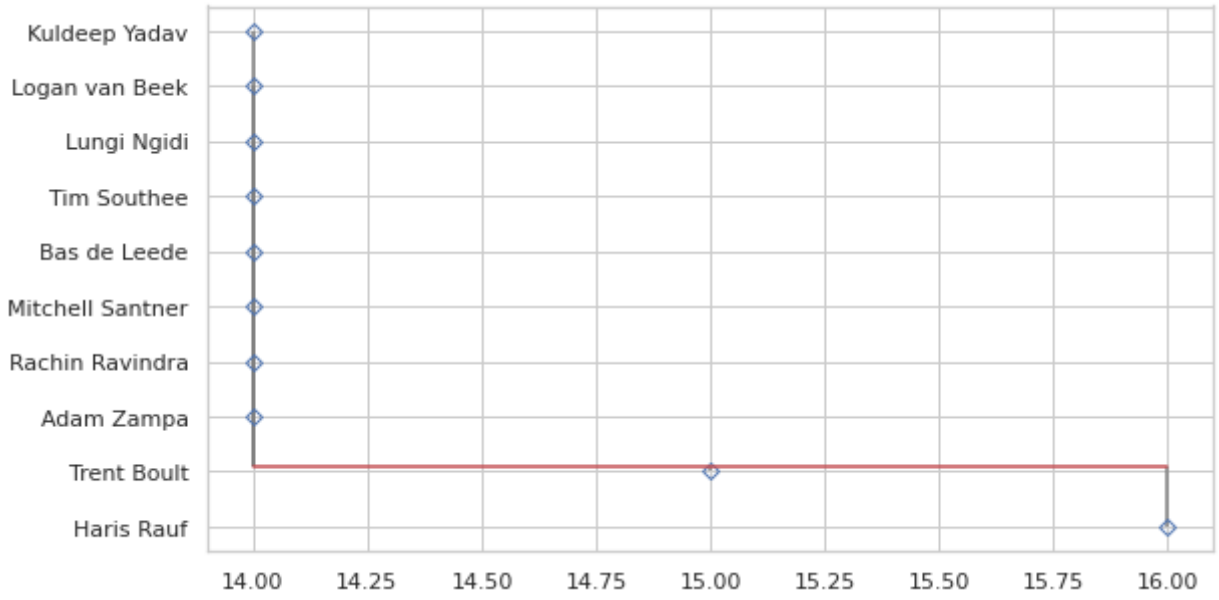
```
sns.set_color_codes("pastel")
sns.barplot(x="wickets", y="player", data=df_2,
            label="wickets_taken", color="b")

# Plot the crashes where alcohol was involved
sns.set_color_codes("muted")
sns.barplot(x="maidens", y="player", data=crashes,
            label="maidens_bowled", color="b")

# Add a legend and informative axis label
ax.legend(ncol=2, loc="lower right", frameon=True)
ax.set(xlim=(0, 24), ylabel="Players",
    xlabel="wickets and maiden overs")
sns.despine(left=True, bottom=True)
```



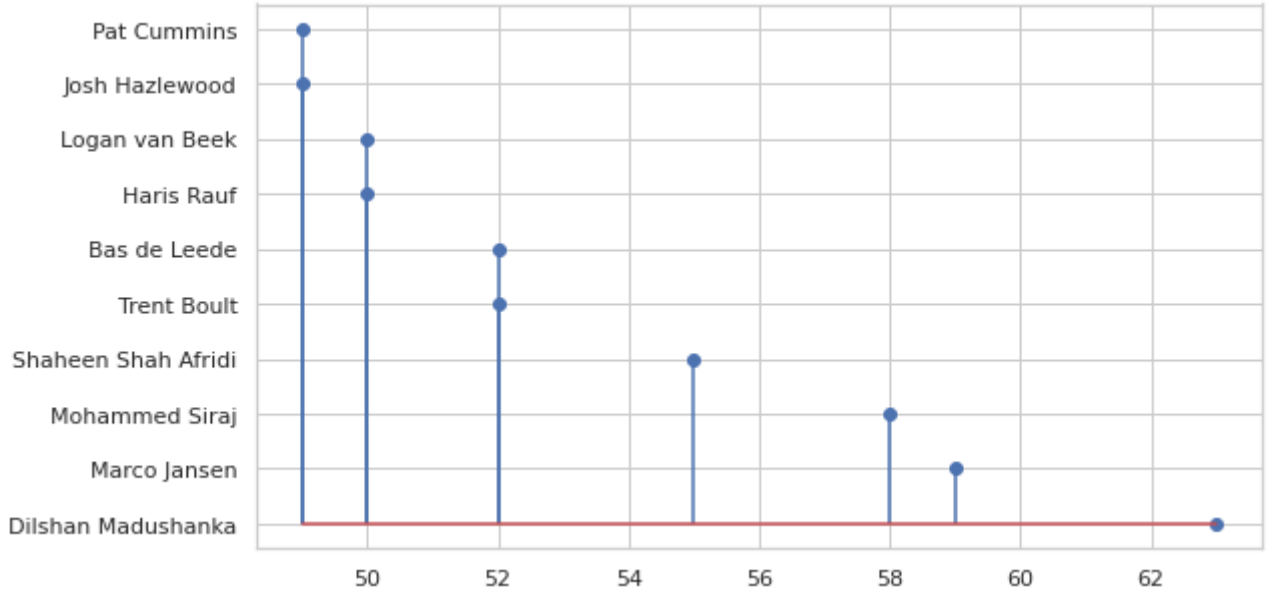
Q5: Most boundaries conceded by a bowler in world cup 2023?

```
df_1=df.groupby('player').sum()
df_2=df_1.sort_values('6s',ascending=False)
df_2.reset_index(inplace=True)

df_2.insert(0, 'player', df_2.pop('player'))
```

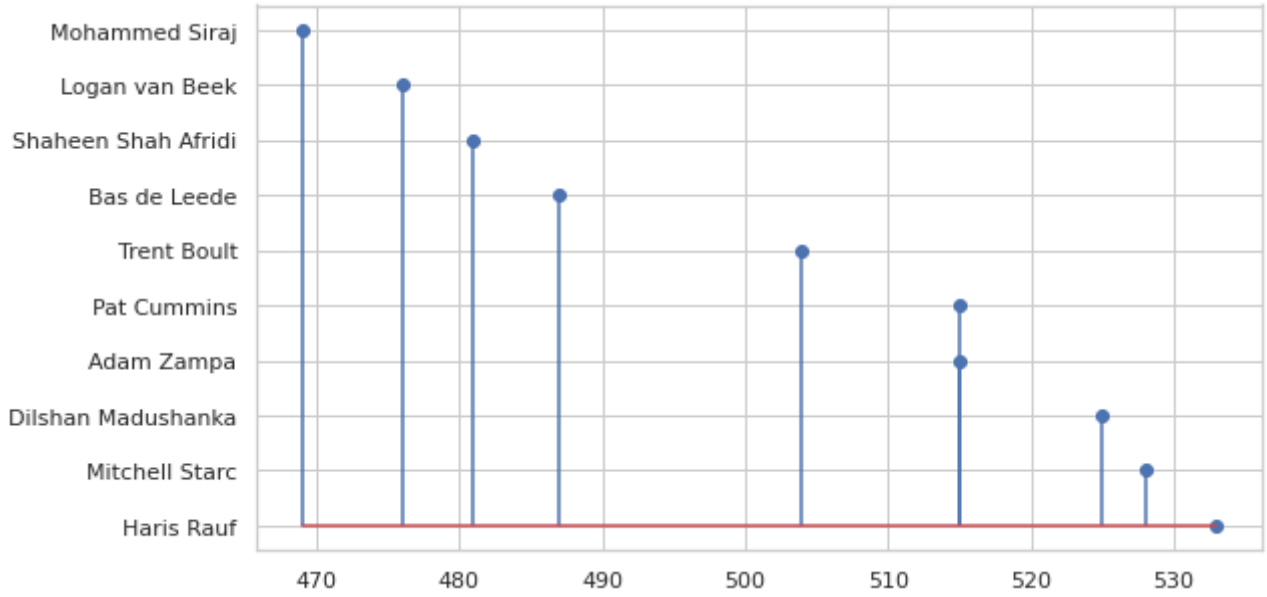
```
df_2.rename(columns={'6s':'sixes','4s':'fours'},inplace=True)
#boundaries in terms of sixes
markerline, stemlines, baseline = plt.stem(
    df_2.head(10).sixes, df_2.head(10).player, linefmt='grey', markerfmt='D', bottom=1.
markerline.set_markerfacecolor('none')
plt.show()
```




```
df_1=df.groupby('player').sum()
df_3=df_1.sort_values('4s',ascending=False)
df_3.reset_index(inplace=True)

df_3.insert(0, 'player', df_3.pop('player'))
df_3.rename(columns={'6s':'sixes','4s':'fours'},inplace=True)
df_3.head(10)

plt.stem(df_3.head(10).fours, df_3.head(10).player)
plt.show()
#most fours hit to a bowler in world cup 2023
```



```
plt.stem(df_2.sort_values('runs',ascending=False).head(10).runs, df_2.sort_values('runs
plt.show()
#most runs to a bowler in world cup 2023
```



Let us save and upload our work to Jovian before continuing.

```
import jovian
```

```
jovian.commit()
```

```
[jovian] Updating notebook "sadeel823/bowlers-performance-analysis-in-mens-odi-cricket-world-cups-51f0a" on https://jovian.com
[jovian] Committed successfully! https://jovian.com/sadeel823/bowlers-performance-analysis-in-mens-odi-cricket-world-cups-51f0a
```

```
'https://jovian.com/sadeel823/bowlers-performance-analysis-in-mens-odi-cricket-world-cups-51f0a'
```

Inferences and Conclusion

Analysis on bowler's performance in odi cwc 2023 - In the above shown analysis, bowler's performance has been displayed in different aspects, the bowlers have been ranked in respect to their wickets, maiden overs as well as boundaries and runs conceded, plots, bar plots, line graphs, scatter plots are drawn on the basis of this analysis which clearly determines the best bowler and the best performing team in term's of bowling.

```
import jovian
```

```
jovian.commit()
```

```
[jovian] Updating notebook "sadeel823/bowlers-performance-analysis-in-mens-odi-cricket-world-cups-51f0a" on https://jovian.com
[jovian] Committed successfully! https://jovian.com/sadeel823/bowlers-performance-analysis-in-mens-odi-cricket-world-cups-51f0a
```

```
'https://jovian.com/sadeel823/bowlers-performance-analysis-in-mens-odi-cricket-world-cups-51f0a'
```

References and Future Work

TODO - In future I plan to work on large data sets, and work on improving my data analyst skills. In order to get skilled I will also be working on various other data analysis tools such as excel, Tableau, sql. Sql is going to be my current learning goal.
Dataset used in current project: https://www.kaggle.com/datasets/vikramrn/icc-mens-odi-world-cup-wc-2023

```
import jovian
```

```
jovian.commit()
```

```

```