LOVELY PROFESSIONAL UNIVERSITY

INT216

CAPSTONE- PROJECT

ON

Project Title: T20 World Cup Data Analysis

Section: K21CH

Submitted by: -

Name: Utkrist Ark

Reg. No: 12105798

Roll No: RK2CHB64

Under the Guidance of

(WASEEM UD DIN WANI)

School of Computer Science and Engineering

DECLARATION

We here by declare that the project work entitled("**T20 World Cup Data Analysis & Premier League Matches Prediction**") is an authentic record of our own work carried out as requirements of Capstone Project for the award of B. Tech degree in Python Project from Lovely Professional University, Phagwara, under the guidance of Waseem Ud Din Wani , during January to May 2023. All the information furnished in this capstone project report is based on our own intensive work and is genuine.

Name of Student: Utkrist Ark

Registration Number: 12105798

CERTIFICATE

This is to certify that the declaration statement made by this group of students is correct to the best of my knowledge

and belief. They have completed this Capstone Project under my guidance and supervision. The present work is the result of their original investigation, effort, and study. No part of the work has ever been submitted for any other degree at any University. The Capstone Project is fit for the submission and partial fulfilment of the conditions for the award of B.Tech degree in from Lovely Professional University, Phagwara.

Signature and Name of the Mentor
Designation
School of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab.

TABLE OF CONTENTS

ABSTRACT

This cricket data analysis project aims to analyze recently held t20 Cricket worldcup (held in Oman and the United Arab Emirates) data from the internet by leveraging web scraping, Python, pandas, and

Power BI. The project will involve scraping relevant data from ESPNcricinfo storing the scraped data in a structured format using pandas, and analyzing the data.

INTRODUCTION

Cricket is a widely popular sport across the world, with millions of fans, players, and stakeholders. With the advent of technology, it is now possible to access a vast amount of cricket-related data from the internet. This has led to the emergence of data analysis and visualization techniques that can be applied to gain insights into the game.

Some of the key metrics that will be analyzed include team and player statistics, match results, player performance, and rankings. The project will also explore trends and patterns in the data to gain insights into the game and make predictions for future matches.

OBJECTIVE

The primary objective of this cricket data analysis project is to leverage web scraping, Python, pandas, and Power BI to analyze historical cricket data from various online sources. The project aims to achieve the following objectives:

1.To collect and store relevant cricket data from various online sources using web scraping techniques.

2.To structure the collected data and store it in a usable format using Python and pandas.

3.To perform various data manipulation techniques such as filtering, grouping, and pivoting to analyze the collected data.

4.To gain insights into the game of cricket by analyzing key metrics such as team and player statistics, match results, player performance, and rankings.

5.To explore trends and patterns in the data and make predictions for future matches.

6.To visualize the analyzed data using Power BI and create interactive visualizations such as charts, graphs, and dashboards.

7.To provide a comprehensive and user-friendly interface to explore the data, uncover hidden insights, and make informed decisions.

DESCRIPTION

The Cricket data analysis project using web scraping, python, pandas, and power bi is a project that involves the analysis of historical cricket data from online sources to gain insights into the game. The project leverages web scraping techniques to collect relevant data sources like ESPNcricinfo. The collected data is then

stored in a structured format using the Python programming language and the pandas library.

The project involves various data manipulation techniques, such as filtering, grouping, and pivoting, to analyze the collected data. Key metrics such as team and player statistics, match results, player performance, and rankings are analyzed to gain insights into the game of cricket. The project also explores trends and patterns in the data to make predictions for future matches.

The final output of the project is visualized using Power BI, which provides a user-friendly interface to explore the data, uncover hidden insights, and make informed decisions. The visualization includes interactive visualizations such as charts, graphs, and dashboards.

The insights generated from this project can be useful for cricket enthusiasts, analysts, and stakeholders in the cricketing world, providing valuable insights and understanding of the game. Overall, the Cricket data analysis project using web scraping, python, pandas, and power bi is an essential project that leverages the power of data analysis and visualization techniques to gain insights into the game of cricket.

CODE

```
import pandas as pd
import json
with open('t20_json_files/t20_json_files/t20_wc_match_re
sults.json') as f:
    data = json.load(f)


df_match = pd.DataFrame(data[0]['matchSummary'])
df_match.head()
```

```
df_match.shape
```

```
(45, 7)
```

Use scorecard as a match id to link with other tables

```
df_match.rename({'scorecard': 'match_id'}, axis =
 1, inplace = True)
df_match.head()
```

Create a match ids dictionary that maps team names to a unique match id. This will be useful later on to link with other tables

```
match_ids_dict = {}

for index, row in df_match.iterrows():
    key1 = row['team1'] + ' Vs ' + row['team2']
```

```
    key2 = row['team2'] + ' Vs ' + row['team1']
    match_ids_dict[key1] = row['match_id']
    match_ids_dict[key2] = row['match_id']

df_match.to_csv('t20_csv_files/dim_match_summary.
csv', index = False)

with open('t20_json_files/t20_json_files/t20_wc_b
atting_summary.json') as f:
    data = json.load(f)
    all_records = []
    for rec in data:
        all_records.extend(rec['battingSummary'])

df_batting = pd.DataFrame(all_records)
df_batting.head(11)


df_batting['out/not_out'] = df_batting.dismissal.
apply(lambda x: "out" if len(x)>0 else "not_out")
df_batting.head(11)

df_batting['match_id'] = df_batting['match'].map(
match_ids_dict)
df_batting.head()

df_batting.drop(columns=["dismissal"], inplace=Tr
ue)
df_batting.head(10)
```

Cleanup weird characters

```
df_batting['batsmanName'] = df_batting['batsmanName'].apply(lambda x: x.replace('â€', ''))
df_batting['batsmanName'] = df_batting['batsmanName'].apply(lambda x: x.replace('\xa0', ''))
df_batting.head()
```

```
df_batting.shape
```

```
(699, 11)
```

```
df_batting.to_csv('t20_csv_files/t20_csv_files/fact_bating_summary.csv', index = False)
```

```
with open('t20_json_files/t20_json_files/t20_wc_bowling_summary.json') as f:
    data = json.load(f)
    all_records = []
    for rec in data:
        all_records.extend(rec['bowlingSummary'])
all_records[:2]
```

```
df_bowling = pd.DataFrame(all_records)
print(df_bowling.shape)
df_bowling.head()
```

```
df_bowling.to_csv('t20_csv_files/fact_bowling_summary.csv', index = False)

with open('t20_json_files/t20_json_files/t20_wc_player_info.json') as f:
    data = json.load(f)

df_players = pd.DataFrame(data)

print(df_players.shape)
df_players.head(10)
```

Cleanup weird characters

```
df_players['name'] = df_players['name'].apply(lambda x: x.replace('â€', ''))
df_players['name'] = df_players['name'].apply(lambda x: x.replace('†', ''))
df_players['name'] = df_players['name'].apply(lambda x: x.replace('\xa0', ''))
df_players.head(10)
```

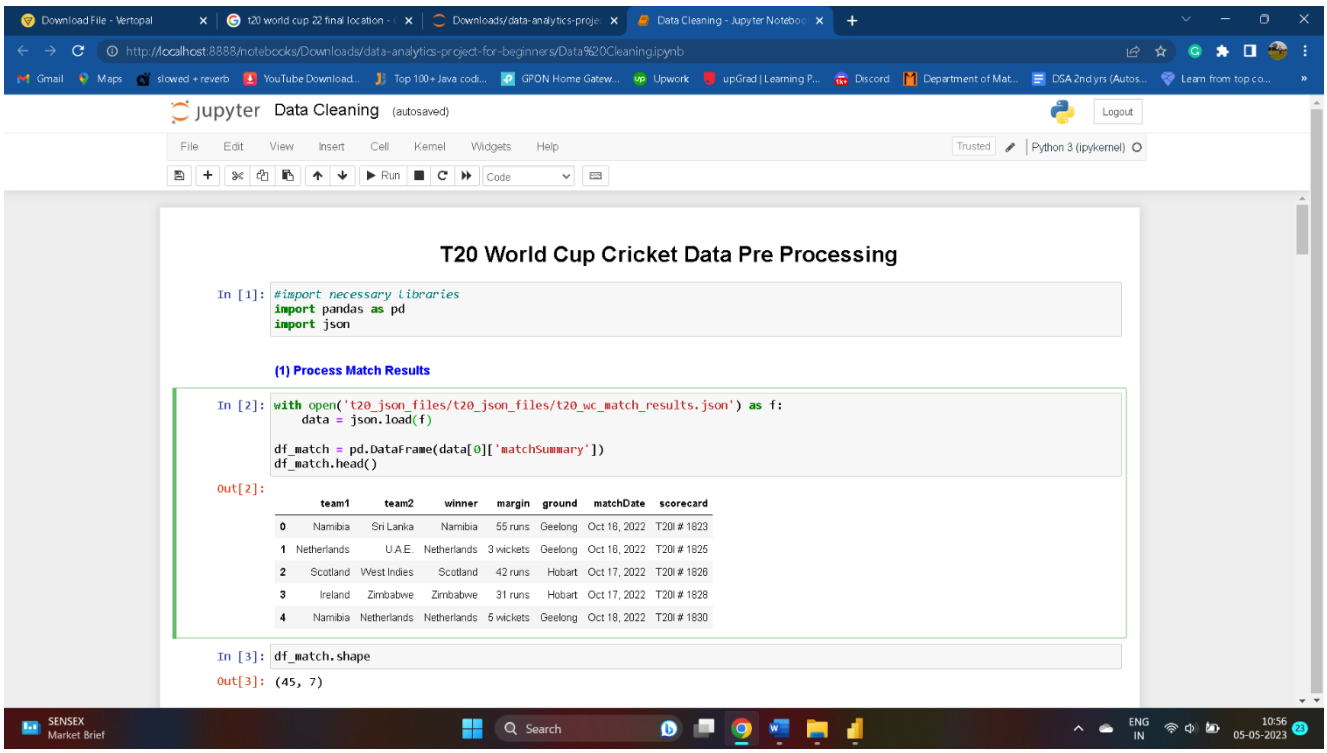
```
df_players[df_players['team'] == 'India']
```

	name	team	battingStyle	\
127	KL Rahul	India	Right hand Bat	
128	Rohit Sharma(c)	India	Right hand Bat	
129	Virat Kohli	India	Right hand Bat	
130	Suryakumar Yadav	India	Right hand Bat	
131	Axar Patel	India	Left hand Bat	

```
132        Hardik Pandya  India  Right hand Bat
133       Dinesh Karthik  India  Right hand Bat
134  Ravichandran Ashwin  India  Right hand Bat
135    Bhuvneshwar Kumar  India  Right hand Bat
136       Arshdeep Singh  India   Left hand Bat
137       Mohammed Shami  India  Right hand Bat
192         Deepak Hooda  India  Right hand Bat
211         Rishabh Pant  India   Left hand Bat

                              bowlingStyle          playi
ngRole  \
127                                             Opening
Batter
128                    Right arm Offbreak     Top order
Batter
129                      Right arm Medium     Top order
Batter
130  Right arm Medium, Right arm Offbreak
Batter
131                Slow Left arm Orthodox   Bowling Allr
ounder
132                 Right arm Medium fast           Allr
ounder
133                    Right arm Offbreak  Wicketkeeper
Batter
134                    Right arm Offbreak   Bowling Allr
ounder
135                      Right arm Medium
Bowler
136                   Left arm Medium fast
Bowler
137                        Right arm Fast
Bowler
```

192	Right arm Offbreak	Allrounder
211	Wicketkeeper Batter	

OUTPUT OF PROJECT



jupyter Data Cleaning (autosaved)

T20 World Cup Cricket Data Pre Processing

```
In [1]: #import necessary libraries
        import pandas as pd
        import json
```

(1) Process Match Results

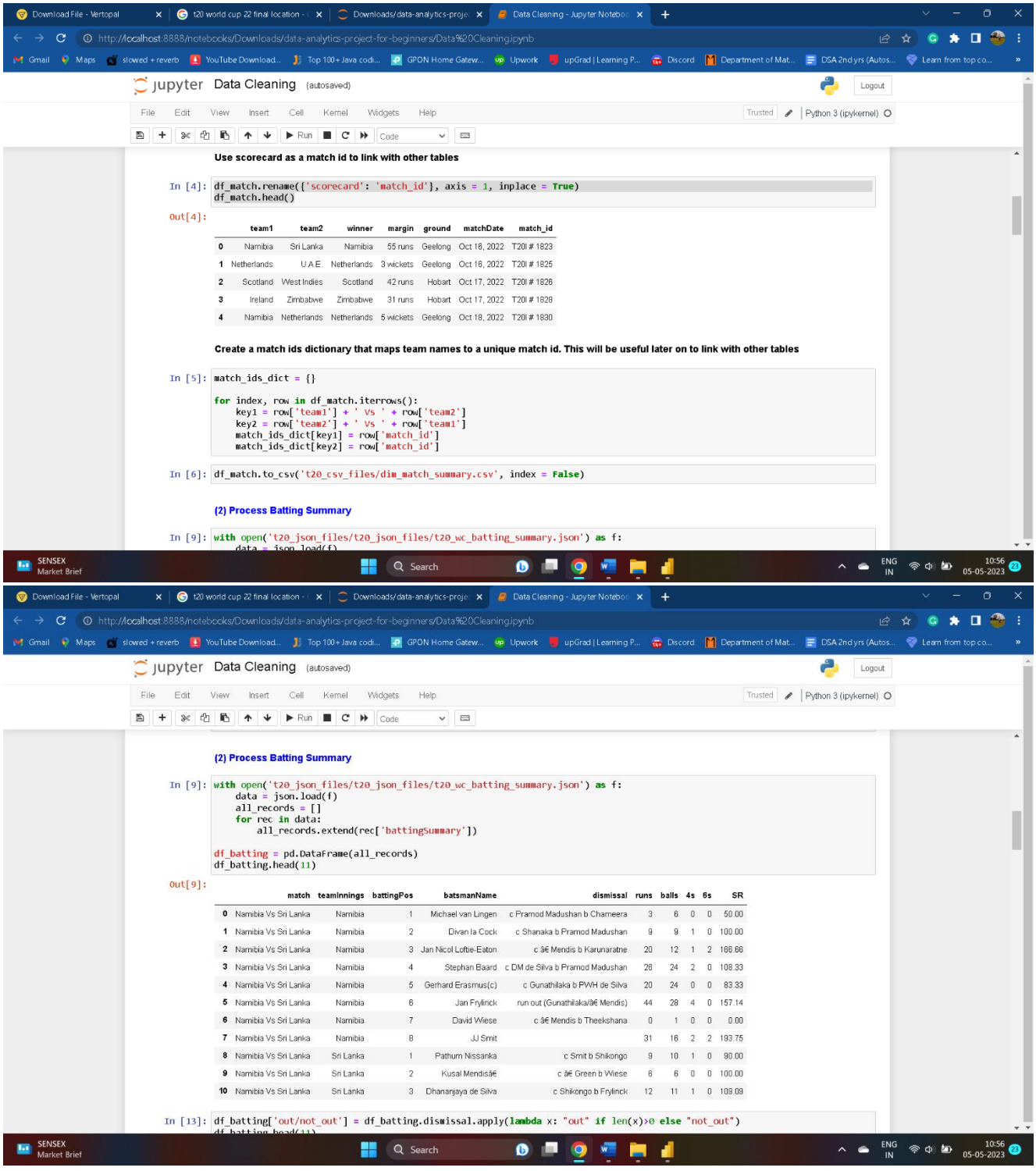
```
In [2]: with open('t20_json_files/t20_json_files/t20_wc_match_results.json') as f:
            data = json.load(f)

        df_match = pd.DataFrame(data[0]['matchSummary'])
        df_match.head()
```

Out[2]:

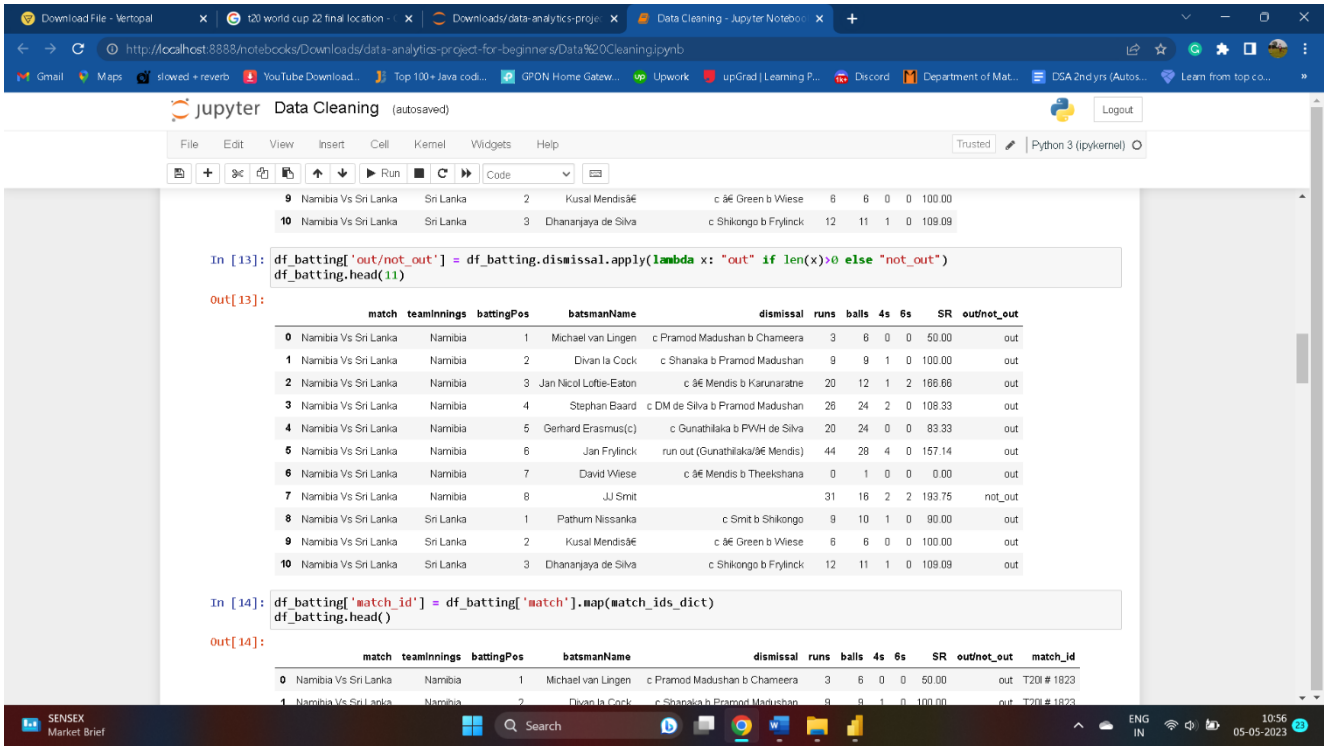
	team1	team2	winner	margin	ground	matchDate	scorecard
0	Namibia	Sri Lanka	Namibia	55 runs	Geelong	Oct 16, 2022	T20I # 1823
1	Netherlands	U.A.E.	Netherlands	3 wickets	Geelong	Oct 16, 2022	T20I # 1825
2	Scotland	West Indies	Scotland	42 runs	Hobart	Oct 17, 2022	T20I # 1826
3	Ireland	Zimbabwe	Zimbabwe	31 runs	Hobart	Oct 17, 2022	T20I # 1828
4	Namibia	Netherlands	Netherlands	5 wickets	Geelong	Oct 18, 2022	T20I # 1830

```
In [3]: df_match.shape
```

Out[3]: (45, 7)


Download File - Vertopal
t20 world cup 22 final location -
Downloads/data-analytics-proje
Data Cleaning - Jupyter Noteboo
http://localhost:8888/notebooks/Downloads/data-analytics-project-for-beginners/Data%20Cleaning.ipynb
Gmail
Maps
slowed + reverb
YouTube Download...
Top 100+ Java codi...
GPON Home Gatew...
Upwork
upGrad | Learning P...
Discord
Department of Mat...
DSA 2nd yrs (Autos...
Learn from top co...
jupyter Data Cleaning (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
Use scorecard as a match id to link with other tables
In [4]:
df_match.rename({'scorecard': 'match_id'}, axis = 1, inplace = True)
df_match.head()
Out[4]:
team1 team2 winner margin ground matchDate match_id
0 Namibia Sri Lanka Namibia 55 runs Geelong Oct 16, 2022 T20I # 1823
1 Netherlands U.A.E. Netherlands 3 wickets Geelong Oct 16, 2022 T20I # 1825
2 Scotland West Indies Scotland 42 runs Hobart Oct 17, 2022 T20I # 1826
3 Ireland Zimbabwe Zimbabwe 31 runs Hobart Oct 17, 2022 T20I # 1828
4 Namibia Netherlands Netherlands 5 wickets Geelong Oct 18, 2022 T20I # 1830
Create a match ids dictionary that maps team names to a unique match id. This will be useful later on to link with other tables
In [5]:
match_ids_dict = {}
for index, row in df_match.iterrows():
key1 = row['team1'] + ' Vs ' + row['team2']
key2 = row['team2'] + ' Vs ' + row['team1']
match_ids_dict[key1] = row['match_id']
match_ids_dict[key2] = row['match_id']
In [6]:
df_match.to_csv('t20_csv_files/dim_match_summary.csv', index = False)
(2) Process Batting Summary
In [9]:
with open('t20_json_files/t20_json_files/t20_wc_batting_summary.json') as f:
SENSEX
Market Brief
Search
ENG
IN
10:56
05-05-2023

Download File - Vertopal
t20 world cup 22 final location -
Downloads/data-analytics-proje
Data Cleaning - Jupyter Noteboo
http://localhost:8888/notebooks/Downloads/data-analytics-project-for-beginners/Data%20Cleaning.ipynb
jupyter Data Cleaning (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
(2) Process Batting Summary
In [9]:
with open('t20_json_files/t20_json_files/t20_wc_batting_summary.json') as f:
data = json.load(f)
all_records = []
for rec in data:
all_records.extend(rec['battingSummary'])
df_batting = pd.DataFrame(all_records)
df_batting.head(11)
Out[9]:
match teamInnings battingPos batsmanName dismissal runs balls 4s 6s SR
0 Namibia Vs Sri Lanka Namibia 1 Michael van Lingen c Pramod Madushan b Chameera 3 6 0 0 50.00
1 Namibia Vs Sri Lanka Namibia 2 Divan la Cock c Shanaka b Pramod Madushan 9 9 1 0 100.00
2 Namibia Vs Sri Lanka Namibia 3 Jan Nicol Loftie-Eaton c â€ Mendis b Karunaratne 20 12 1 2 166.66
3 Namibia Vs Sri Lanka Namibia 4 Stephan Baard c DM de Silva b Pramod Madushan 26 24 2 0 108.33
4 Namibia Vs Sri Lanka Namibia 5 Gerhard Erasmus(c) c Gunathilaka b PWH de Silva 20 24 0 0 83.33
5 Namibia Vs Sri Lanka Namibia 6 Jan Frylinck run out (Gunathilaka/â€ Mendis) 44 28 4 0 157.14
6 Namibia Vs Sri Lanka Namibia 7 David Wiese c â€ Mendis b Theekshana 0 1 0 0 0.00
7 Namibia Vs Sri Lanka Namibia 8 JJ Smit 31 16 2 2 193.75
8 Namibia Vs Sri Lanka Sri Lanka 1 Pathum Nissanka c Smit b Shikongo 9 10 1 0 90.00
9 Namibia Vs Sri Lanka Sri Lanka 2 Kusal Mendisâ€ c â€ Green b Wiese 6 6 0 0 100.00
10 Namibia Vs Sri Lanka Sri Lanka 3 Dhananjaya de Silva c Shikongo b Frylinck 12 11 1 0 109.09
In [13]:
df_batting['out/not_out'] = df_batting.dismissal.apply(lambda x: "out" if len(x)>0 else "not_out")
SENSEX
Market Brief
Search
ENG
IN
10:56
05-05-2023



9	Namibia Vs Sri Lanka	Sri Lanka	2	Kusal Mendisâ€	c â€ Green b Wiese	6	6	0	0	100.00	
10	Namibia Vs Sri Lanka	Sri Lanka	3	Dhananjaya de Silva	c Shikongo b Frylinck	12	11	1	0	109.09	

In [13]:
```
df_batting['out/not_out'] = df_batting.dismissal.apply(lambda x: "out" if len(x)>0 else "not_out")
df_batting.head(11)
```

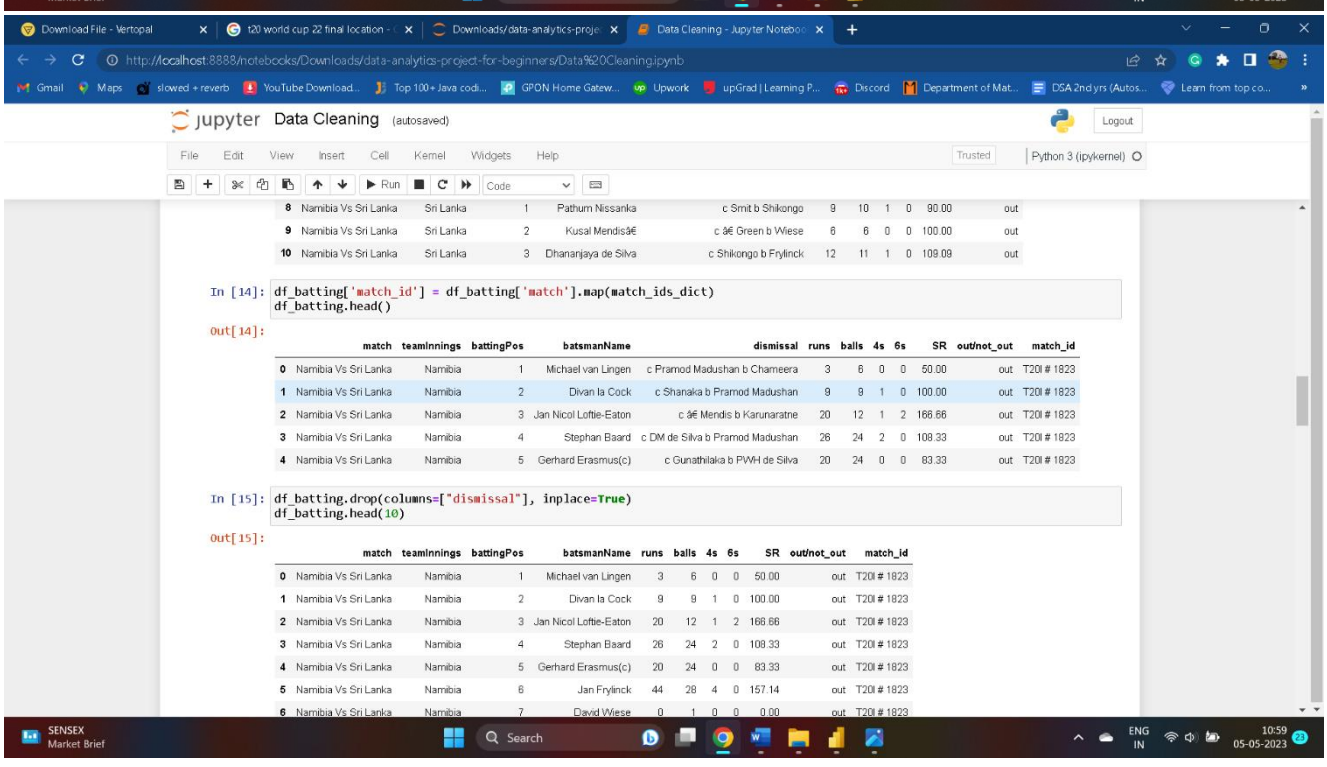
Out[13]:

	match	teaminnings	battingPos	batsmanName	dismissal	runs	balls	4s	6s	SR	out/not_out
0	Namibia Vs Sri Lanka	Namibia	1	Michael van Lingen	c Pramod Madushan b Chameera	3	6	0	0	50.00	out
1	Namibia Vs Sri Lanka	Namibia	2	Divan la Cock	c Shanaka b Pramod Madushan	9	9	1	0	100.00	out
2	Namibia Vs Sri Lanka	Namibia	3	Jan Nicol Loftie-Eaton	c â€ Mendis b Karunaratne	20	12	1	2	166.66	out
3	Namibia Vs Sri Lanka	Namibia	4	Stephan Baard	c DM de Silva b Pramod Madushan	26	24	2	0	108.33	out
4	Namibia Vs Sri Lanka	Namibia	5	Gerhard Erasmus(c)	c Gunathilaka b PWH de Silva	20	24	0	0	83.33	out
5	Namibia Vs Sri Lanka	Namibia	6	Jan Frylinck	run out (Gunathilaka/â€ Mendis)	44	28	4	0	157.14	out
6	Namibia Vs Sri Lanka	Namibia	7	David Wiese	c â€ Mendis b Theekshana	0	1	0	0	0.00	out
7	Namibia Vs Sri Lanka	Namibia	8	JJ Smit		31	16	2	2	193.75	not_out
8	Namibia Vs Sri Lanka	Sri Lanka	1	Pathum Nissanka	c Smit b Shikongo	9	10	1	0	90.00	out
9	Namibia Vs Sri Lanka	Sri Lanka	2	Kusal Mendisâ€	c â€ Green b Wiese	6	6	0	0	100.00	out
10	Namibia Vs Sri Lanka	Sri Lanka	3	Dhananjaya de Silva	c Shikongo b Frylinck	12	11	1	0	109.09	out

In [14]:
```
df_batting['match_id'] = df_batting['match'].map(match_ids_dict)
df_batting.head()
```

Out[14]:

	match	teaminnings	battingPos	batsmanName	dismissal	runs	balls	4s	6s	SR	out/not_out	match_id
0	Namibia Vs Sri Lanka	Namibia	1	Michael van Lingen	c Pramod Madushan b Chameera	3	6	0	0	50.00	out	T20I # 1823



8	Namibia Vs Sri Lanka	Sri Lanka	1	Pathum Nissanka	c Smit b Shikongo	9	10	1	0	90.00	out
9	Namibia Vs Sri Lanka	Sri Lanka	2	Kusal Mendisâ€	c â€ Green b Wiese	6	6	0	0	100.00	out
10	Namibia Vs Sri Lanka	Sri Lanka	3	Dhananjaya de Silva	c Shikongo b Frylinck	12	11	1	0	109.09	out

In [14]:
```
df_batting['match_id'] = df_batting['match'].map(match_ids_dict)
df_batting.head()
```

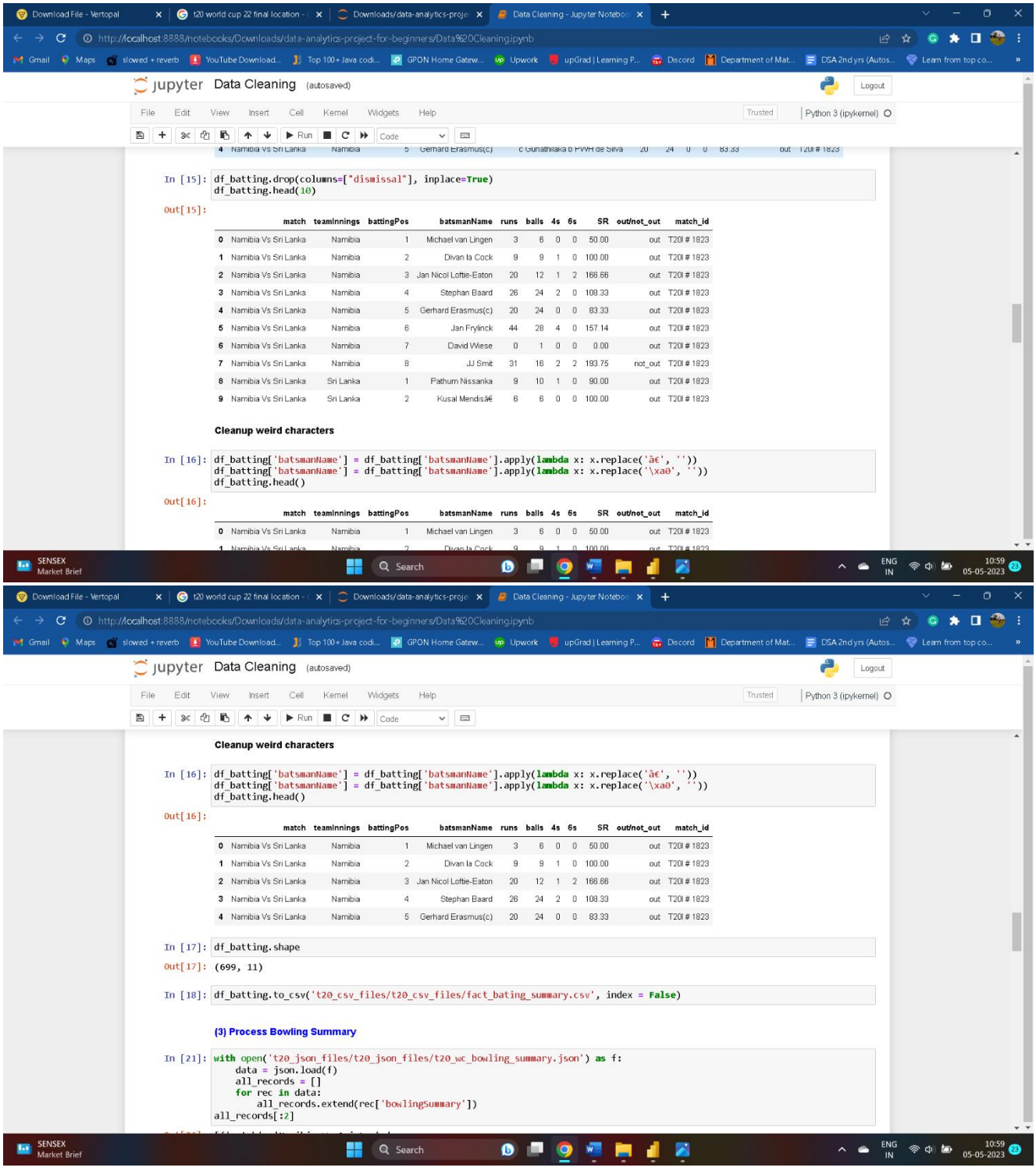
Out[14]:

	match	teaminnings	battingPos	batsmanName	dismissal	runs	balls	4s	6s	SR	out/not_out	match_id
0	Namibia Vs Sri Lanka	Namibia	1	Michael van Lingen	c Pramod Madushan b Chameera	3	6	0	0	50.00	out	T20I # 1823
1	Namibia Vs Sri Lanka	Namibia	2	Divan la Cock	c Shanaka b Pramod Madushan	9	9	1	0	100.00	out	T20I # 1823
2	Namibia Vs Sri Lanka	Namibia	3	Jan Nicol Loftie-Eaton	c â€ Mendis b Karunaratne	20	12	1	2	166.66	out	T20I # 1823
3	Namibia Vs Sri Lanka	Namibia	4	Stephan Baard	c DM de Silva b Pramod Madushan	26	24	2	0	108.33	out	T20I # 1823
4	Namibia Vs Sri Lanka	Namibia	5	Gerhard Erasmus(c)	c Gunathilaka b PWH de Silva	20	24	0	0	83.33	out	T20I # 1823

In [15]:
```
df_batting.drop(columns=["dismissal"], inplace=True)
df_batting.head(10)
```

Out[15]:

	match	teaminnings	battingPos	batsmanName	runs	balls	4s	6s	SR	out/not_out	match_id
0	Namibia Vs Sri Lanka	Namibia	1	Michael van Lingen	3	6	0	0	50.00	out	T20I # 1823
1	Namibia Vs Sri Lanka	Namibia	2	Divan la Cock	9	9	1	0	100.00	out	T20I # 1823
2	Namibia Vs Sri Lanka	Namibia	3	Jan Nicol Loftie-Eaton	20	12	1	2	166.66	out	T20I # 1823
3	Namibia Vs Sri Lanka	Namibia	4	Stephan Baard	26	24	2	0	108.33	out	T20I # 1823
4	Namibia Vs Sri Lanka	Namibia	5	Gerhard Erasmus(c)	20	24	0	0	83.33	out	T20I # 1823
5	Namibia Vs Sri Lanka	Namibia	6	Jan Frylinck	44	28	4	0	157.14	out	T20I # 1823
6	Namibia Vs Sri Lanka	Namibia	7	David Wiese	0	1	0	0	0.00	out	T20I # 1823



jupyter Data Cleaning (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3 (ipykernel)

Run Code

In [15]:
```
df_batting.drop(columns=["dismissal"], inplace=True)
df_batting.head(10)
```

Out[15]:

	match	teaminnings	battingPos	batsmanName	runs	balls	4s	6s	SR	out/not_out	match_id
0	Namibia Vs Sri Lanka	Namibia	1	Michael van Lingen	3	6	0	0	50.00	out	T20I # 1823
1	Namibia Vs Sri Lanka	Namibia	2	Divan la Cock	9	9	1	0	100.00	out	T20I # 1823
2	Namibia Vs Sri Lanka	Namibia	3	Jan Nicol Loftie-Eaton	20	12	1	2	166.66	out	T20I # 1823
3	Namibia Vs Sri Lanka	Namibia	4	Stephan Baard	26	24	2	0	108.33	out	T20I # 1823
4	Namibia Vs Sri Lanka	Namibia	5	Gerhard Erasmus(c)	20	24	0	0	83.33	out	T20I # 1823
5	Namibia Vs Sri Lanka	Namibia	6	Jan Frylinck	44	28	4	0	157.14	out	T20I # 1823
6	Namibia Vs Sri Lanka	Namibia	7	David Wiese	0	1	0	0	0.00	out	T20I # 1823
7	Namibia Vs Sri Lanka	Namibia	8	JJ Smit	31	16	2	2	193.75	not_out	T20I # 1823
8	Namibia Vs Sri Lanka	Sri Lanka	1	Pathum Nissanka	9	10	1	0	90.00	out	T20I # 1823
9	Namibia Vs Sri Lanka	Sri Lanka	2	Kusal Mendisâ€	6	6	0	0	100.00	out	T20I # 1823

Cleanup weird characters

In [16]:
```
df_batting['batsmanName'] = df_batting['batsmanName'].apply(lambda x: x.replace('â€', ''))
df_batting['batsmanName'] = df_batting['batsmanName'].apply(lambda x: x.replace('\xa0', ''))
df_batting.head()
```

Out[16]:

	match	teaminnings	battingPos	batsmanName	runs	balls	4s	6s	SR	out/not_out	match_id
0	Namibia Vs Sri Lanka	Namibia	1	Michael van Lingen	3	6	0	0	50.00	out	T20I # 1823
1	Namibia Vs Sri Lanka	Namibia	2	Divan la Cock	9	9	1	0	100.00	out	T20I # 1823
2	Namibia Vs Sri Lanka	Namibia	3	Jan Nicol Loftie-Eaton	20	12	1	2	166.66	out	T20I # 1823
3	Namibia Vs Sri Lanka	Namibia	4	Stephan Baard	26	24	2	0	108.33	out	T20I # 1823
4	Namibia Vs Sri Lanka	Namibia	5	Gerhard Erasmus(c)	20	24	0	0	83.33	out	T20I # 1823

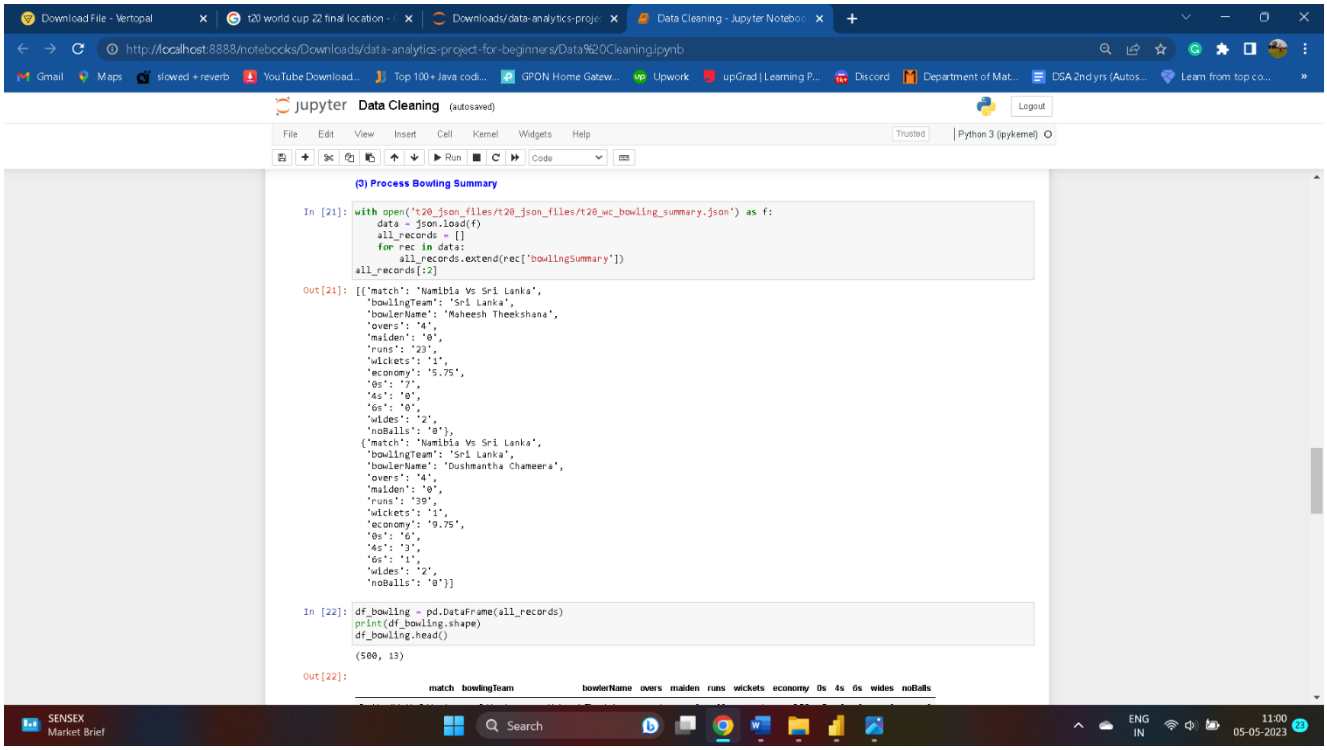
In [17]:
```
df_batting.shape
```

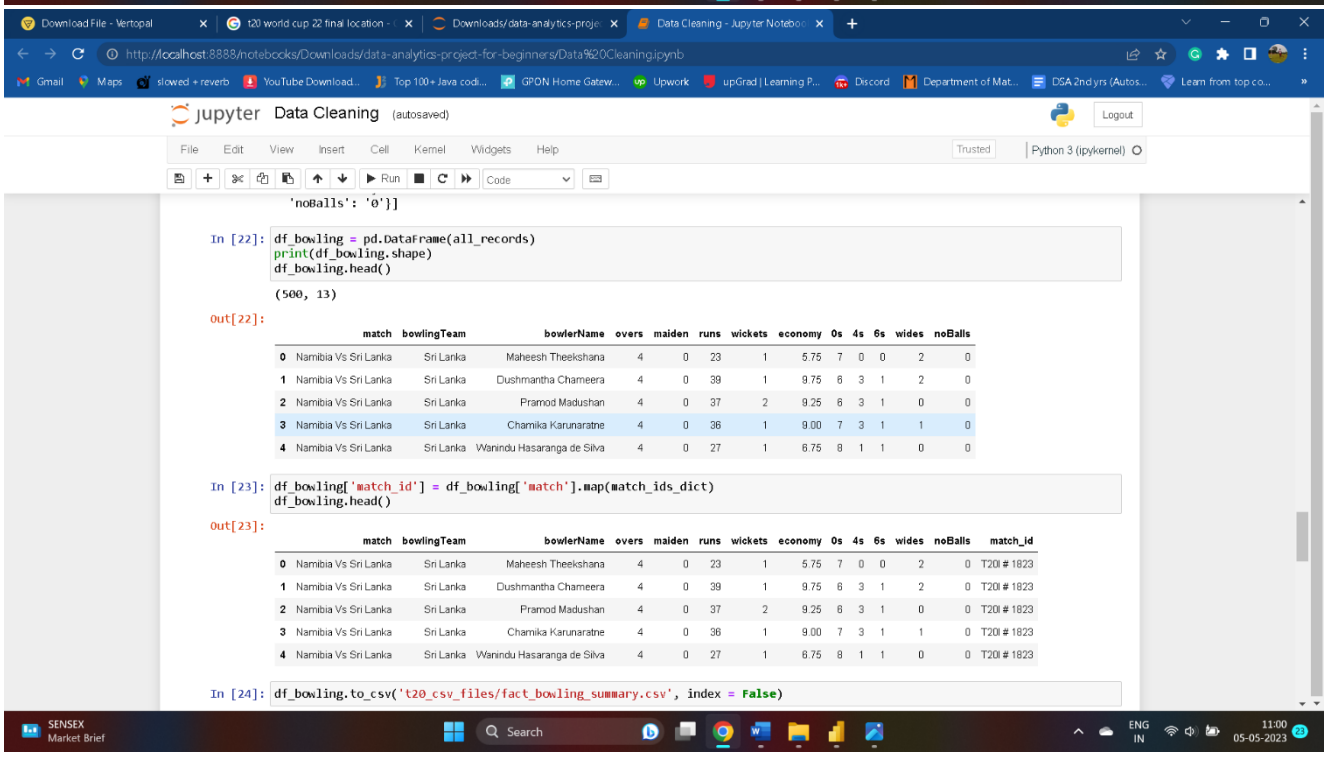
Out[17]: (699, 11)

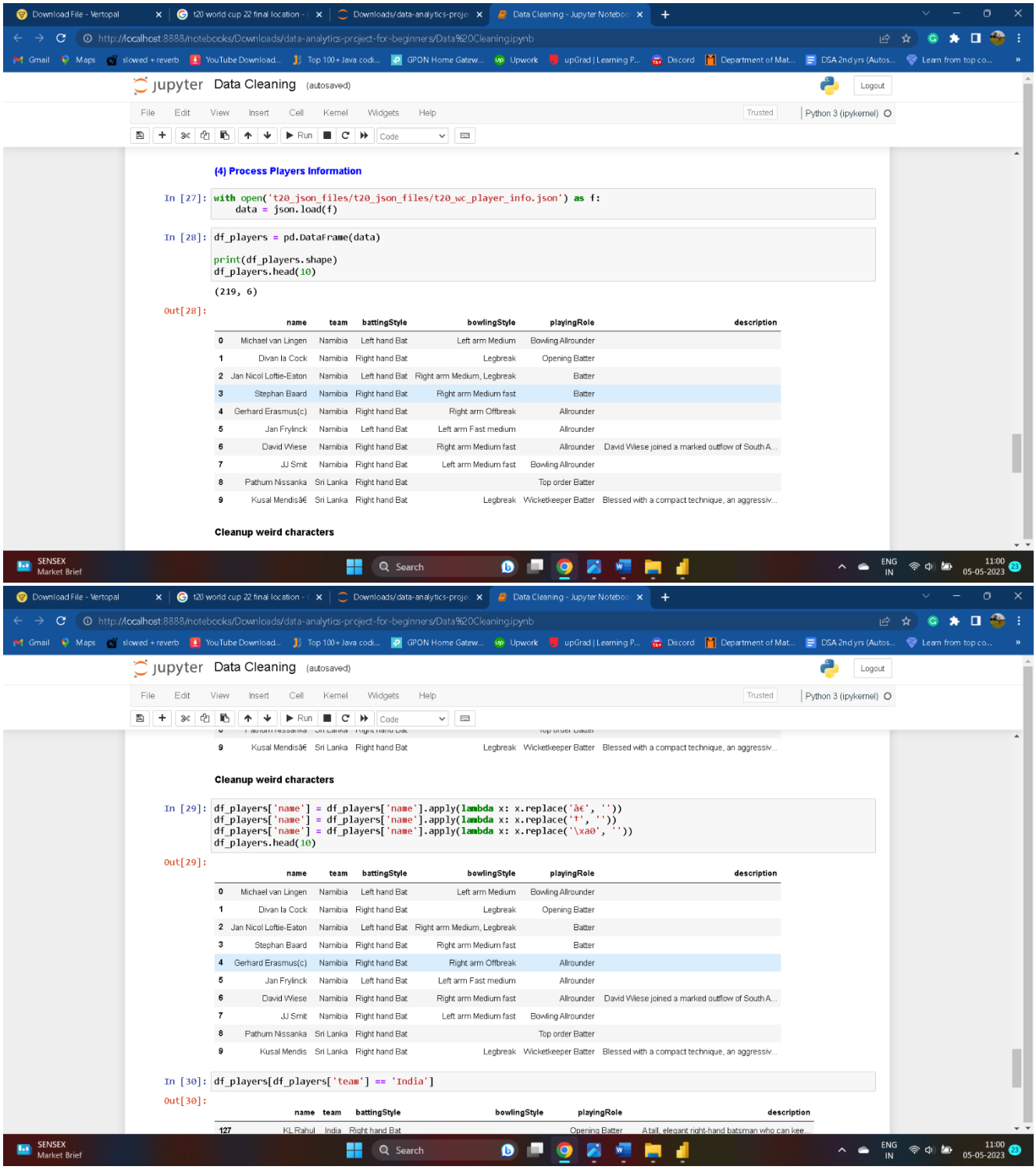
In [18]:
```
df_batting.to_csv('t20_csv_files/t20_csv_files/fact_bating_summary.csv', index = False)
```

(3) Process Bowling Summary

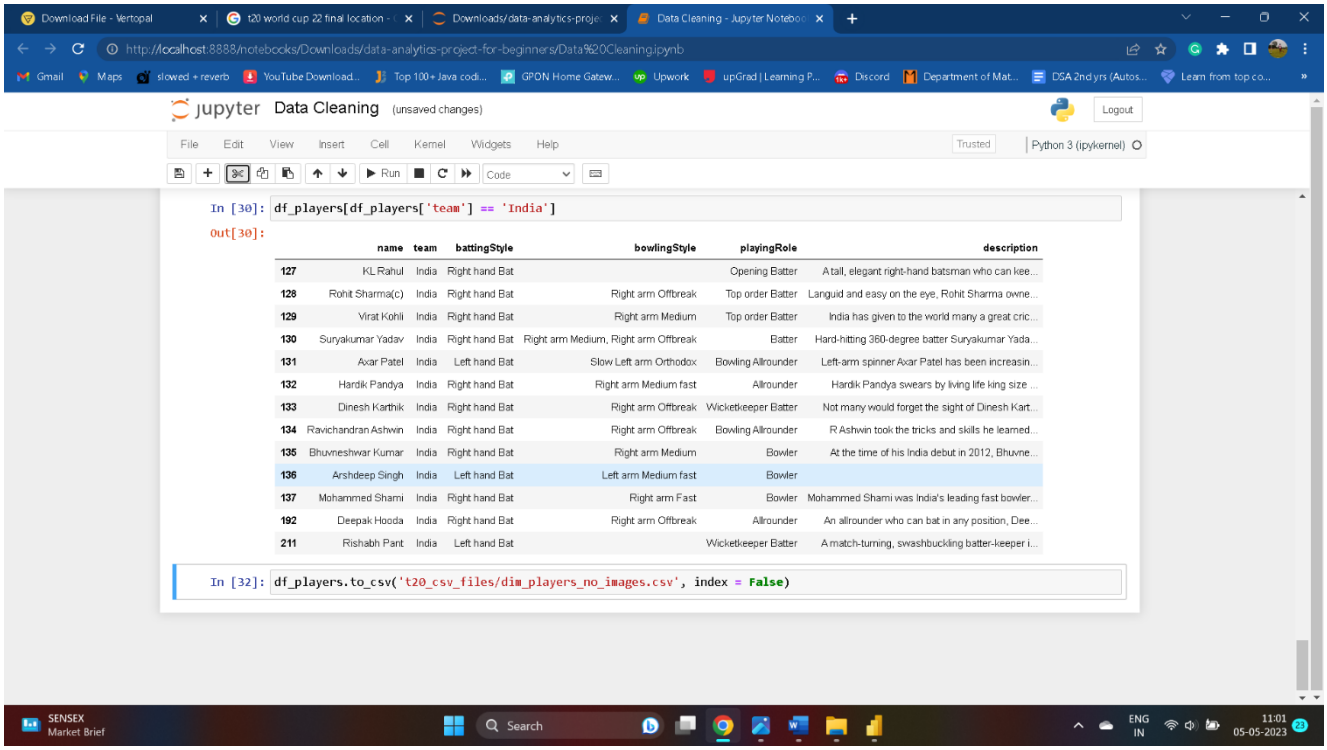
In [21]:
```
with open('t20_json_files/t20_json_files/t20_wc_bowling_summary.json') as f:
    data = json.load(f)
    all_records = []
    for rec in data:
        all_records.extend(rec['bowlingSummary'])
all_records[:2]
```


Download File - Vertopal
t20 world cup 22 final location -
Downloads/data-analytics-proje
Data Cleaning - Jupyter Notebook
http://localhost:8888/notebooks/Downloads/data-analytics-project-for-beginners/Data%20Cleaning.ipynb
jupyter Data Cleaning (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
(3) Process Bowling Summary
In [21]:
with open('t20_json_files/t20_json_files/t20_wc_bowling_summary.json') as f:
data = json.load(f)
all_records = []
for rec in data:
all_records.extend(rec['bowlingSummary'])
all_records[:2]
Out[21]:
[{'match': 'Namibia Vs Sri Lanka',
'bowlingTeam': 'Sri Lanka',
'bowlerName': 'Maheesh Theekshana',
'overs': '4',
'maiden': '0',
'runs': '23',
'wickets': '1',
'economy': '5.75',
'0s': '7',
'4s': '0',
'6s': '0',
'wides': '2',
'noBalls': '0'},
{'match': 'Namibia Vs Sri Lanka',
'bowlingTeam': 'Sri Lanka',
'bowlerName': 'Dushmantha Chameera',
'overs': '4',
'maiden': '0',
'runs': '39',
'wickets': '1',
'economy': '9.75',
'0s': '6',
'4s': '3',
'6s': '1',
'wides': '2',
'noBalls': '0'}]
In [22]:
df_bowling = pd.DataFrame(all_records)
print(df_bowling.shape)
df_bowling.head()
(500, 13)
Out[22]:
match bowlingTeam bowlerName overs maiden runs wickets economy 0s 4s 6s wides noBalls
SENSEX
Market Brief
Search
ENG
IN
11:00
05-05-2023

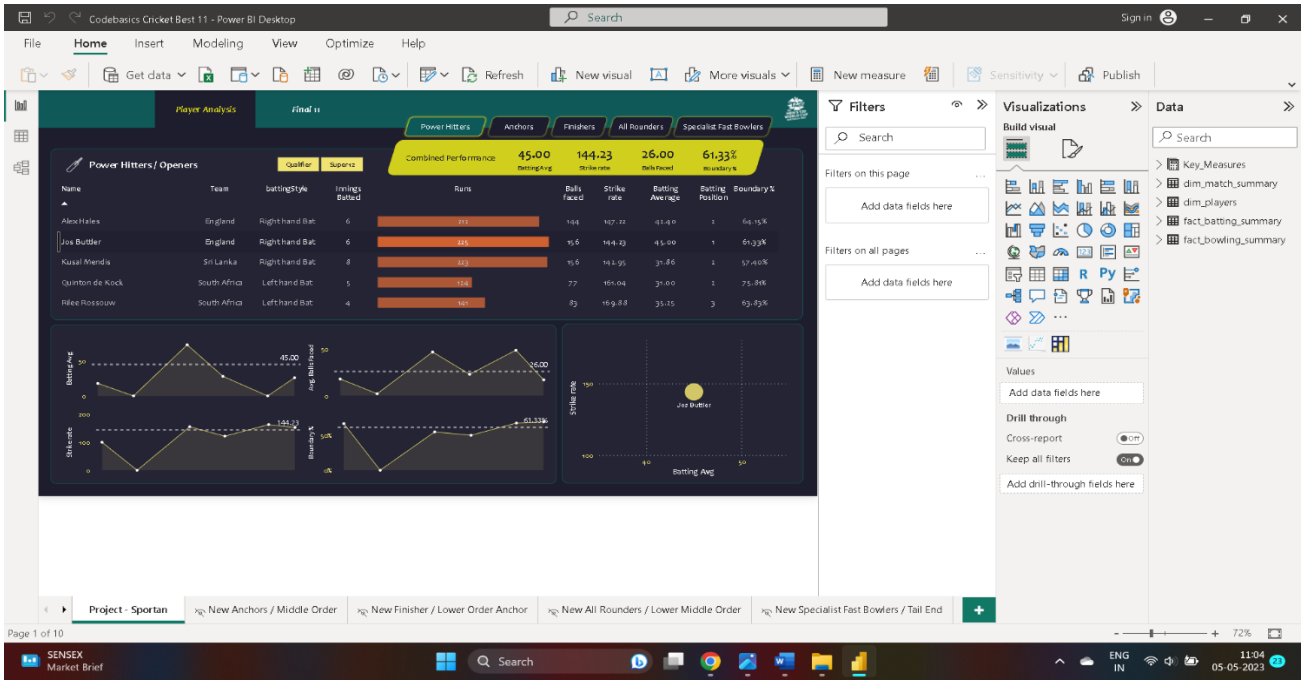

Download File - Vertopal
t20 world cup 22 final location -
Downloads/data-analytics-proje
Data Cleaning - Jupyter Notebook
http://localhost:8888/notebooks/Downloads/data-analytics-project-for-beginners/Data%20Cleaning.ipynb
jupyter Data Cleaning (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
'noBalls': '0'}]
In [22]:
df_bowling = pd.DataFrame(all_records)
print(df_bowling.shape)
df_bowling.head()
(500, 13)
Out[22]:
match bowlingTeam bowlerName overs maiden runs wickets economy 0s 4s 6s wides noBalls
0 Namibia Vs Sri Lanka Sri Lanka Maheesh Theekshana 4 0 23 1 5.75 7 0 0 2 0
1 Namibia Vs Sri Lanka Sri Lanka Dushmantha Chameera 4 0 39 1 9.75 6 3 1 2 0
2 Namibia Vs Sri Lanka Sri Lanka Pramod Madushan 4 0 37 2 9.25 6 3 1 0 0
3 Namibia Vs Sri Lanka Sri Lanka Chamika Karunaratne 4 0 36 1 9.00 7 3 1 1 0
4 Namibia Vs Sri Lanka Sri Lanka Wanindu Hasaranga de Silva 4 0 27 1 6.75 8 1 1 0 0
In [23]:
df_bowling['match_id'] = df_bowling['match'].map(match_ids_dict)
df_bowling.head()
Out[23]:
match bowlingTeam bowlerName overs maiden runs wickets economy 0s 4s 6s wides noBalls match_id
0 Namibia Vs Sri Lanka Sri Lanka Maheesh Theekshana 4 0 23 1 5.75 7 0 0 2 0 T20I # 1823
1 Namibia Vs Sri Lanka Sri Lanka Dushmantha Chameera 4 0 39 1 9.75 6 3 1 2 0 T20I # 1823
2 Namibia Vs Sri Lanka Sri Lanka Pramod Madushan 4 0 37 2 9.25 6 3 1 0 0 T20I # 1823
3 Namibia Vs Sri Lanka Sri Lanka Chamika Karunaratne 4 0 36 1 9.00 7 3 1 1 0 T20I # 1823
4 Namibia Vs Sri Lanka Sri Lanka Wanindu Hasaranga de Silva 4 0 27 1 6.75 8 1 1 0 0 T20I # 1823
In [24]:
df_bowling.to_csv('t20_csv_files/fact_bowling_summary.csv', index = False)
SENSEX
Market Brief
Search
ENG
IN
11:00
05-05-2023

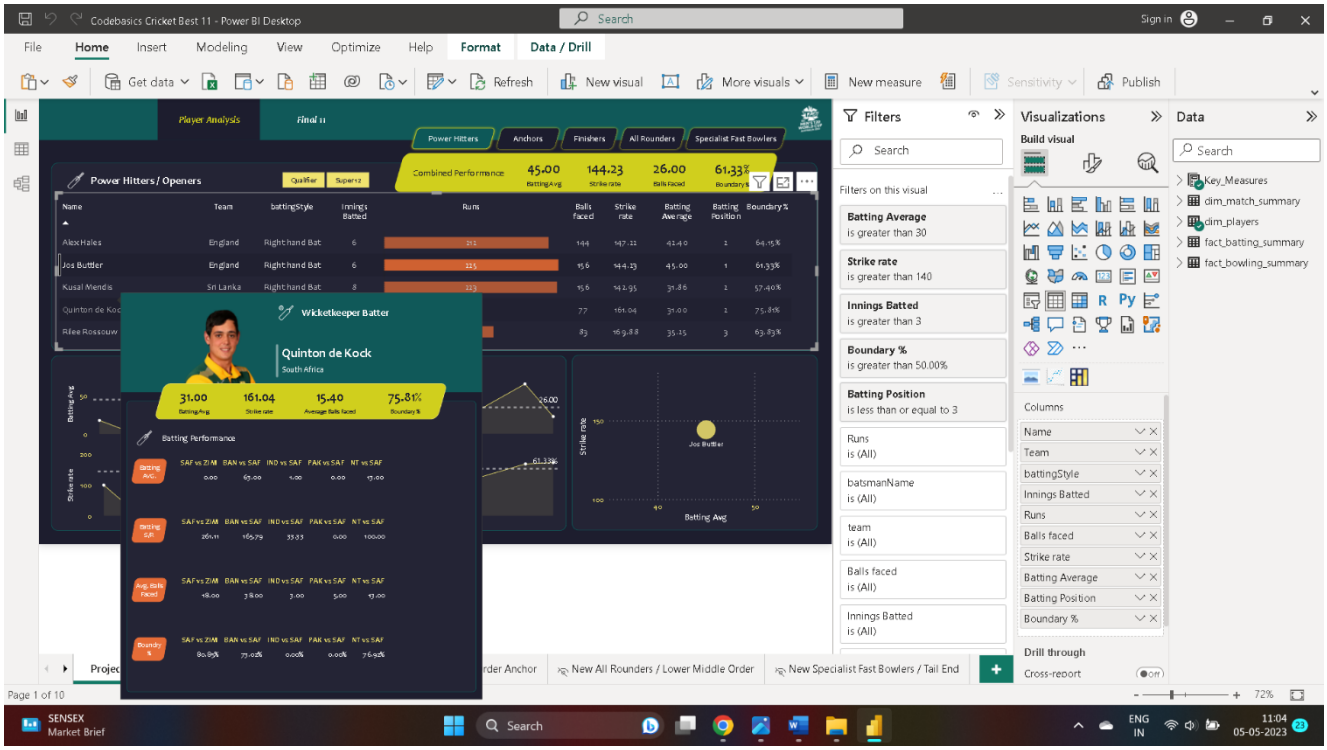

Download File - Vertopal
t20 world cup 22 final location -
Downloads/data-analytics-proje
Data Cleaning - Jupyter Notebook
http://localhost:8888/notebooks/Downloads/data-analytics-project-for-beginners/Data%20Cleaning.ipynb
jupyter Data Cleaning (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
(4) Process Players Information
In [27]:
with open('t20_json_files/t20_json_files/t20_wc_player_info.json') as f:
data = json.load(f)
In [28]:
df_players = pd.DataFrame(data)
print(df_players.shape)
df_players.head(10)
(219, 6)
Out[28]:
name team battingStyle bowlingStyle playingRole description
0 Michael van Lingen Namibia Left hand Bat Left arm Medium Bowling Allrounder
1 Divan la Cock Namibia Right hand Bat Legbreak Opening Batter
2 Jan Nicol Loftie-Eaton Namibia Left hand Bat Right arm Medium, Legbreak Batter
3 Stephan Baard Namibia Right hand Bat Right arm Medium fast Batter
4 Gerhard Erasmus(c) Namibia Right hand Bat Right arm Offbreak Allrounder
5 Jan Frylinck Namibia Left hand Bat Left arm Fast medium Allrounder
6 David Wiese Namibia Right hand Bat Right arm Medium fast Allrounder David Wiese joined a marked outflow of South A...
7 JJ Smit Namibia Right hand Bat Left arm Medium fast Bowling Allrounder
8 Pathum Nissanka Sri Lanka Right hand Bat Top order Batter
9 Kusal Mendisâ€ Sri Lanka Right hand Bat Legbreak Wicketkeeper Batter Blessed with a compact technique, an aggressiv...
Cleanup weird characters
SENSEX
Market Brief
Search
ENG
IN
11:00
05-05-2023

Download File - Vertopal
t20 world cup 22 final location -
Downloads/data-analytics-proje
Data Cleaning - Jupyter Notebook
http://localhost:8888/notebooks/Downloads/data-analytics-project-for-beginners/Data%20Cleaning.ipynb
jupyter Data Cleaning (autosaved)
Logout
File Edit View Insert Cell Kernel Widgets Help
Trusted
Python 3 (ipykernel)
Run
Code
9 Kusal Mendisâ€ Sri Lanka Right hand Bat Legbreak Wicketkeeper Batter Blessed with a compact technique, an aggressiv...
Cleanup weird characters
In [29]:
df_players['name'] = df_players['name'].apply(lambda x: x.replace('â€', ''))
df_players['name'] = df_players['name'].apply(lambda x: x.replace('†', ''))
df_players['name'] = df_players['name'].apply(lambda x: x.replace('\xa0', ''))
df_players.head(10)
Out[29]:
name team battingStyle bowlingStyle playingRole description
0 Michael van Lingen Namibia Left hand Bat Left arm Medium Bowling Allrounder
1 Divan la Cock Namibia Right hand Bat Legbreak Opening Batter
2 Jan Nicol Loftie-Eaton Namibia Left hand Bat Right arm Medium, Legbreak Batter
3 Stephan Baard Namibia Right hand Bat Right arm Medium fast Batter
4 Gerhard Erasmus(c) Namibia Right hand Bat Right arm Offbreak Allrounder
5 Jan Frylinck Namibia Left hand Bat Left arm Fast medium Allrounder
6 David Wiese Namibia Right hand Bat Right arm Medium fast Allrounder David Wiese joined a marked outflow of South A...
7 JJ Smit Namibia Right hand Bat Left arm Medium fast Bowling Allrounder
8 Pathum Nissanka Sri Lanka Right hand Bat Top order Batter
9 Kusal Mendis Sri Lanka Right hand Bat Legbreak Wicketkeeper Batter Blessed with a compact technique, an aggressiv...
In [30]:
df_players[df_players['team'] == 'India']
Out[30]:
name team battingStyle bowlingStyle playingRole description
127 KL Rahul India Right hand Bat Opening Batter A tall, elegant right-hand batsman who can kee...
SENSEX
Market Brief
Search
ENG
IN
11:00
05-05-2023



DashBoard Visualization Using PowerBI





SCOPE OF PROJECT

The scope of the Cricket data analysis project using web scraping, python, pandas, and power bi is vast, and it can cover multiple areas of cricket analysis. The primary focus of the project is to collect historical cricket data from various online sources using web scraping techniques and analyze it to gain insights into the game.

The scope of the project includes the analysis of various metrics such as team and player statistics, match results, player performance, and rankings. The project can also explore trends and patterns in the data to make predictions for future matches.

The project can be used to analyze data for various formats of cricket, including test matches, one-day internationals, and T20s. It can cover data for both international and domestic cricket matches and can be customized to suit the requirements of different stakeholders in the cricketing world.The project can also be extended to include analysis of social media data to gain insights into public sentiment around cricket matches and players. The scope can further be expanded to track player movements and analyze their performance on the field using data from sensors and cameras placed on the field.

Overall, the scope of the Cricket data analysis project using web scraping, python, pandas, and power bi is broad, and it has the potential to provide valuable insights into the game of cricket. The project can be customized to suit different requirements and can be extended to cover new areas of cricket analysis in the future

FUTURE DEVELOPMENT OF THE PROJECT

The Cricket data analysis project using web scraping, python, pandas, and power bi has a lot of potential for future development. Some of the possible future development areas for this project include:

Real-time Data Analysis: One of the future development areas for the project is to incorporate real-time data analysis to provide up-to-date insights into ongoing matches. This will require the project to collect live data and analyze it in real-time.

Machine Learning: Another future development area for the project is to incorporate machine learning techniques to enhance the accuracy of predictions and insights generated from the data. This will require the use of advanced algorithms such as neural networks and decision trees.

Social Media Analysis: The project can also be extended to include analysis of social media data to gain insights into public sentiment around cricket matches and players.

Player Tracking: The project can also be extended to track player movements and analyze their performance on the field. This can be done using data from sensors and cameras placed on the field.

Mobile Application: The project can also be developed into a mobile application that provides live updates, insights, and predictions on ongoing matches.

Overall, there are numerous possibilities for future development of the Cricket data analysis project using web scraping, python, pandas, and power bi. These developments will enhance the accuracy and usefulness of insights generated from the project and provide more value to cricket enthusiasts, analysts, and stakeholders in the cricketing world.

CONCLUSION

The Cricket data analysis project using web scraping, python, pandas, and power bi is a comprehensive project that leverages various techniques to analyze historical cricket data from multiple online sources. The project aims to gain insights into the game of cricket and explore trends and patterns in the data.

By leveraging web scraping techniques, relevant data is collected and stored in a structured format using Python and pandas. The project involves various data manipulation techniques, such as filtering, grouping, and pivoting, to analyze the collected data. Key metrics such as team and player statistics, match results, player performance, and rankings are analyzed to gain insights into the game.The final output of the project is visualized using Power BI, which provides a user-friendly interface to explore the data and create interactive visualizations such as charts, graphs, and dashboards.The insights generated from this project can be useful for cricket enthusiasts, analysts, and stakeholders in the cricketing world, providing valuable insights and understanding of the game. The project demonstrates the power of data analysis and

visualization techniques in gaining insights into complex datasets and making informed decisions.

Overall, the Cricket data analysis project using web scraping, python, pandas, and power bi is an essential project that showcases the potential of technology in the field of sports analysis and provides valuable insights into the game of cricket..