```python
import pandas as pd
import matplotlib.pyplot as plt

# Step 2: Load the Dataset
# Replace the URL or file path with the actual dataset path
dataset_url = "ipl_matches_smalldata.csv" # Example dataset file path
df = pd.read_csv(dataset_url)

print("Dataset Overview:")
print(df.head())

# Step 3: Explore the Dataset
print("\nDataset Info:")
print(df.info())
print("\nDataset Statistics:")
print(df.describe())

# Step 4: Check for Missing Values
print("\nMissing Values:")
print(df.isnull().sum())

# Step 5: Clean the Data
# Fill missing values with 0
df.fillna(0, inplace=True)

# Rename columns for clarity
df.rename(columns={"batsman_runs": "Runs_Scored"}, inplace=True)

# Step 6: Analyze Data
# Example 1: Top Batsmen Analysis
#Total runs scored by each batsman:
print(df.head())
top_batsmen = df.groupby('batter')['Runs_Scored'].sum().sort_values(ascending=False).hea
print("\nTop 10 Batsmen by Runs:")
print(top_batsmen)

# Visualize the data
top_batsmen.plot(kind='bar', title='Top 10 Batsmen by Runs', figsize=(10, 5))
plt.xlabel('Batsman')
plt.ylabel('Total Runs')
plt.show()
```

```
Dataset Overview:
   match_id  inning           batting_team              bowling_team  over  \
0    335982       1  Kolkata Knight Riders  Royal Challengers Bangalore     0
1    335982       1  Kolkata Knight Riders  Royal Challengers Bangalore     0
2    335982       1  Kolkata Knight Riders  Royal Challengers Bangalore     0
3    335982       1  Kolkata Knight Riders  Royal Challengers Bangalore     0
4    335982       1  Kolkata Knight Riders  Royal Challengers Bangalore     0

   ball       batter    bowler    non_striker  batsman_runs  extra_runs  \
0     1  SC Ganguly   P Kumar   BB McCullum             0           1
1     2  BB McCullum  P Kumar    SC Ganguly             0           0
2     3  BB McCullum  P Kumar    SC Ganguly             0           1
3     4  BB McCullum  P Kumar    SC Ganguly             0           0
4     5  BB McCullum  P Kumar    SC Ganguly             0           0

   total_runs extras_type  is_wicket player_dismissed dismissal_kind fielder
```

```
0        1     legbyes      0           NaN          NaN   NaN
1        0         NaN      0           NaN          NaN   NaN
2        1       wides      0           NaN          NaN   NaN
3        0         NaN      0           NaN          NaN   NaN
4        0         NaN      0           NaN          NaN   NaN

Dataset Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3999 entries, 0 to 3998
Data columns (total 17 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   match_id          3999 non-null   int64
 1   inning            3999 non-null   int64
 2   batting_team      3999 non-null   object
 3   bowling_team      3999 non-null   object
 4   over              3999 non-null   int64
 5   ball              3999 non-null   int64
 6   batter            3999 non-null   object
 7   bowler            3999 non-null   object
 8   non_striker       3999 non-null   object
 9   batsman_runs      3999 non-null   int64
 10  extra_runs        3999 non-null   int64
 11  total_runs        3999 non-null   int64
 12  extras_type       263 non-null    object
 13  is_wicket         3999 non-null   int64
 14  player_dismissed  199 non-null    object
 15  dismissal_kind    199 non-null    object
 16  fielder           141 non-null    object
dtypes: int64(8), object(9)
memory usage: 390.6+ KB
None

Dataset Statistics:
             match_id        inning         over          ball  batsman_runs  \
count     3999.000000   3999.000000  3999.000000   3999.000000   3999.000000
mean    335989.992498      1.474369     8.998500      3.639660      1.282571
std          4.880866      0.499405     5.601177      1.825542      1.701359
min     335982.000000      1.000000     0.000000      1.000000      0.000000
25%     335986.000000      1.000000     4.000000      2.000000      0.000000
50%     335990.000000      1.000000     9.000000      4.000000      1.000000
75%     335994.000000      2.000000    14.000000      5.000000      1.000000
max     335998.000000      2.000000    19.000000      9.000000      6.000000

         extra_runs   total_runs    is_wicket
count   3999.000000  3999.000000  3999.000000
mean       0.086022     1.368592     0.049762
std        0.403960     1.689408     0.217481
min        0.000000     0.000000     0.000000
25%        0.000000     0.000000     0.000000
50%        0.000000     1.000000     0.000000
75%        0.000000     1.000000     0.000000
max        5.000000     7.000000     1.000000

Missing Values:
match_id             0
inning               0
batting_team         0
bowling_team         0
over                 0
ball                 0
batter               0
bowler               0
non_striker          0
batsman_runs         0
extra_runs           0
total_runs           0
extras_type       3736
is_wicket            0
player_dismissed  3800
dismissal_kind    3800
fielder           3858
dtype: int64
   match_id  inning           batting_team                 bowling_team  over  \
0    335982       1  Kolkata Knight Riders  Royal Challengers Bangalore     0
1    335982       1  Kolkata Knight Riders  Royal Challengers Bangalore     0
2    335982       1  Kolkata Knight Riders  Royal Challengers Bangalore     0
3    335982       1  Kolkata Knight Riders  Royal Challengers Bangalore     0
4    335982       1  Kolkata Knight Riders  Royal Challengers Bangalore     0

   ball       batter    bowler  non_striker  Runs_Scored  extra_runs  \
0     1   SC Ganguly  P Kumar  BB McCullum            0           1
1     2  BB McCullum  P Kumar   SC Ganguly            0           0
2     3  BB McCullum  P Kumar   SC Ganguly            0           1
```

```
3     4  BB McCullum  P Kumar    SC Ganguly           0           0
4     5  BB McCullum  P Kumar    SC Ganguly           0           0

   total_runs extras_type  is_wicket player_dismissed dismissal_kind fielder
0           1     legbyes          0                0              0       0
1           0           0          0                0              0       0
2           1       wides          0                0              0       0
3           0           0          0                0              0       0
4           0           0          0                0              0       0

Top 10 Batsmen by Runs:
batter
ML Hayden         189
BB McCullum       188
G Gambhir         174
KC Sangakkara     172
MEK Hussey        168
SR Watson         162
A Symonds         161
RV Uthappa        155
AC Gilchrist      153
LRPL Taylor       149
Name: Runs_Scored, dtype: int64
```


Top 10 Batsmen by Runs