EDS Theory Activity no.1

Name : Parth Isekar

Roll No: CS7-09

Prn : 202401080062

MY dataset : IPL

IPL Dataset Assignment

Using NumPy and Pandas

Code with Output

import numpy as np

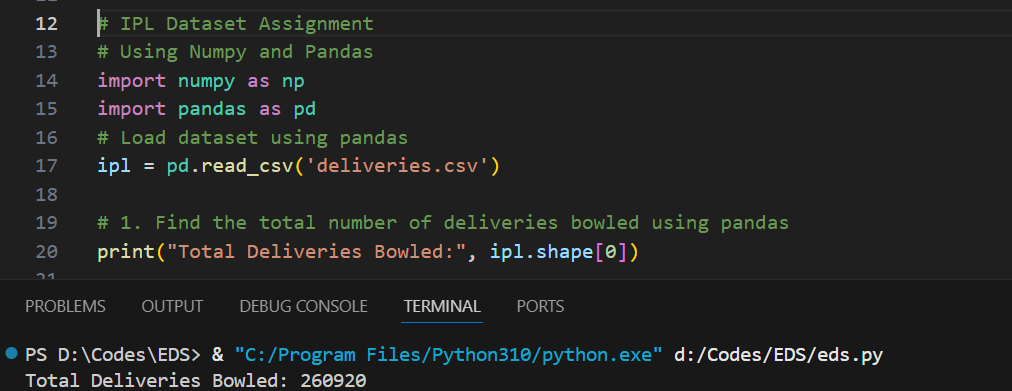
import pandas as pd

# Load dataset using pandas

ipl = pd.read\_csv('deliveries.csv')

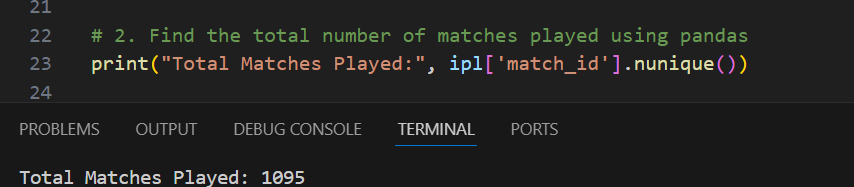
# 1. Find the total number of deliveries bowled using pandas

print("Total Deliveries Bowled:", ipl.shape[0])



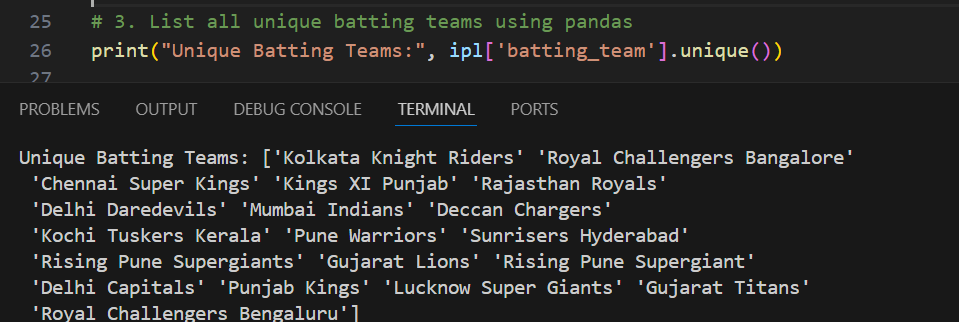
# 2. Find the total number of matches played using pandas

print("Total Matches Played:", ipl['match\_id'].nunique())



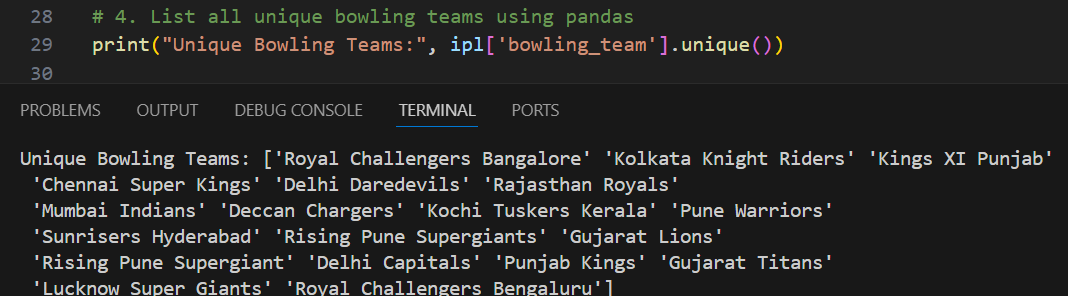
# 3. List all unique batting teams using pandas

print("Unique Batting Teams:", ipl['batting\_team'].unique())



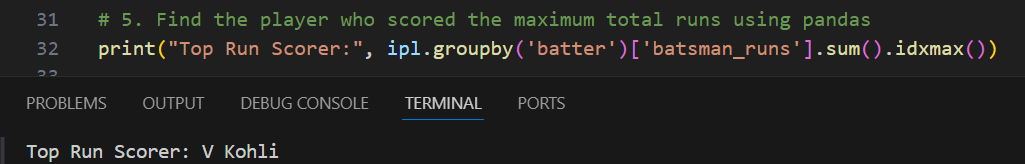
# 4. List all unique bowling teams using pandas

print("Unique Bowling Teams:", ipl['bowling\_team'].unique())



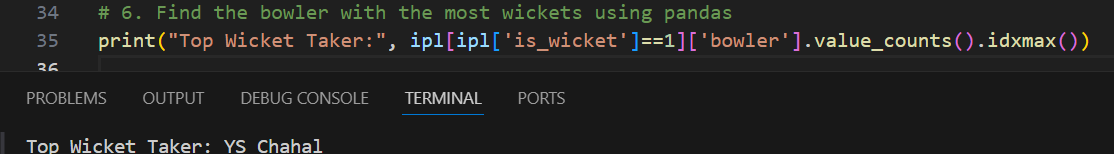
# 5. Find the player who scored the maximum total runs using pandas

print("Top Run Scorer:", ipl.groupby('batter')['batsman\_runs'].sum().idxmax())



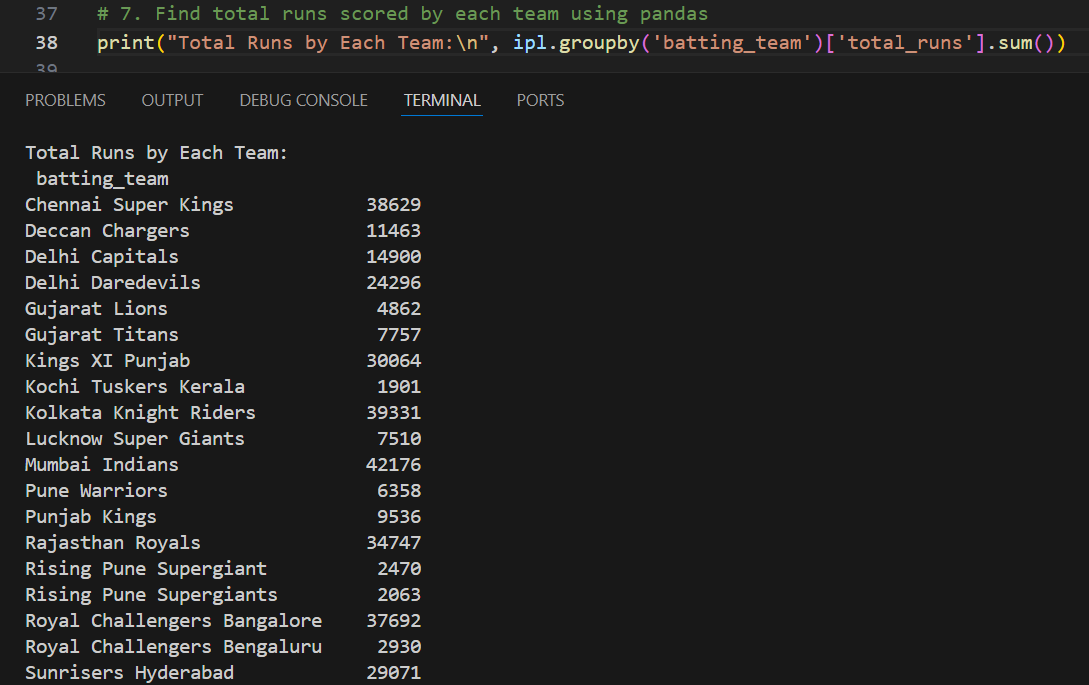
# 6. Find the bowler with the most wickets using pandas

print("Top Wicket Taker:", ipl[ipl['is\_wicket']==1]['bowler'].value\_counts().idxmax())



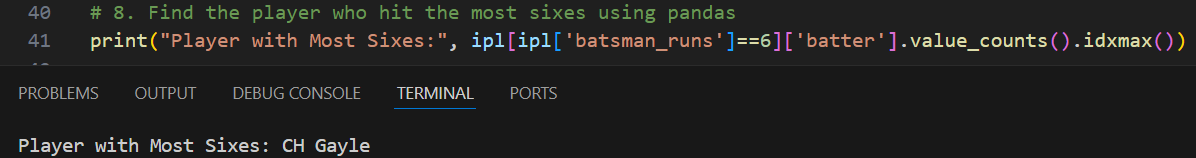
# 7. Find total runs scored by each team using pandas

print("Total Runs by Each Team:\n", ipl.groupby('batting\_team')['total\_runs'].sum())



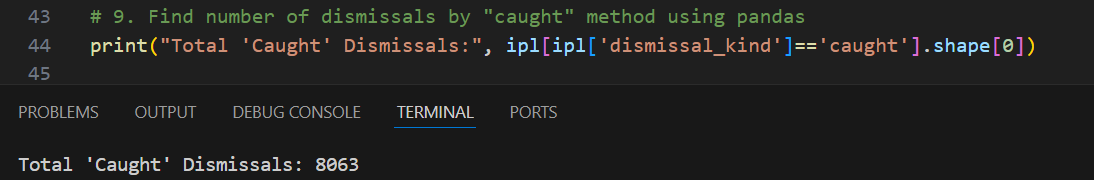
# 8. Find the player who hit the most sixes using pandas

print("Player with Most Sixes:", ipl[ipl['batsman\_runs']==6]['batter'].value\_counts().idxmax())

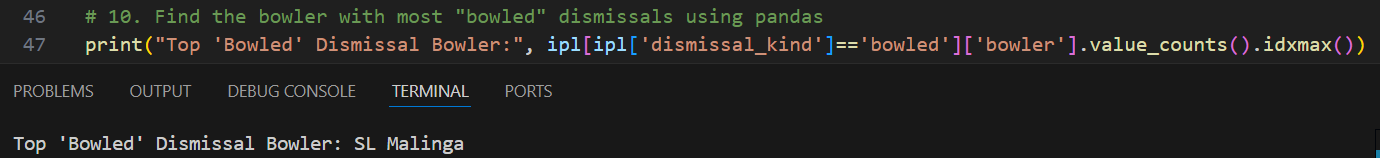


# 9. Find number of dismissals by "caught" method using pandas

print("Total 'Caught' Dismissals:", ipl[ipl['dismissal\_kind']=='caught'].shape[0])

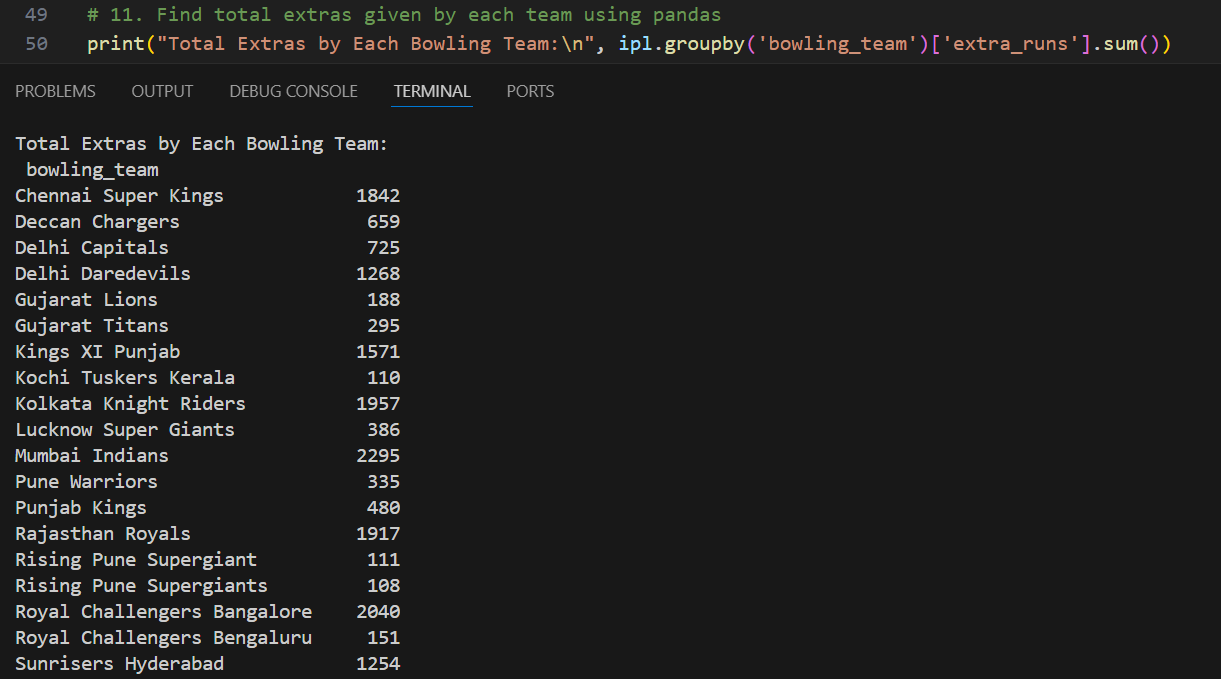


# 10. Find the bowler with most "bowled" dismissals using pandas

print("Top 'Bowled' Dismissal Bowler:", ipl[ipl['dismissal\_kind']=='bowled']['bowler'].value\_counts().idxmax())

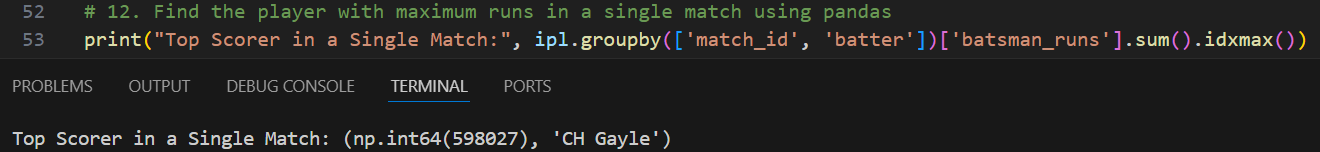
# 11. Find total extras given by each team using pandas

print("Total Extras by Each Bowling Team:\n", ipl.groupby('bowling\_team')['extra\_runs'].sum())



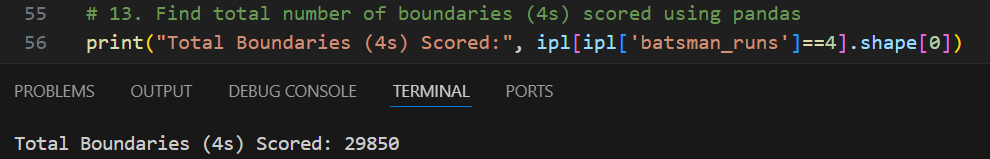
# 12. Find the player with maximum runs in a single match using pandas

print("Top Scorer in a Single Match:", ipl.groupby(['match\_id', 'batter'])['batsman\_runs'].sum().idxmax())



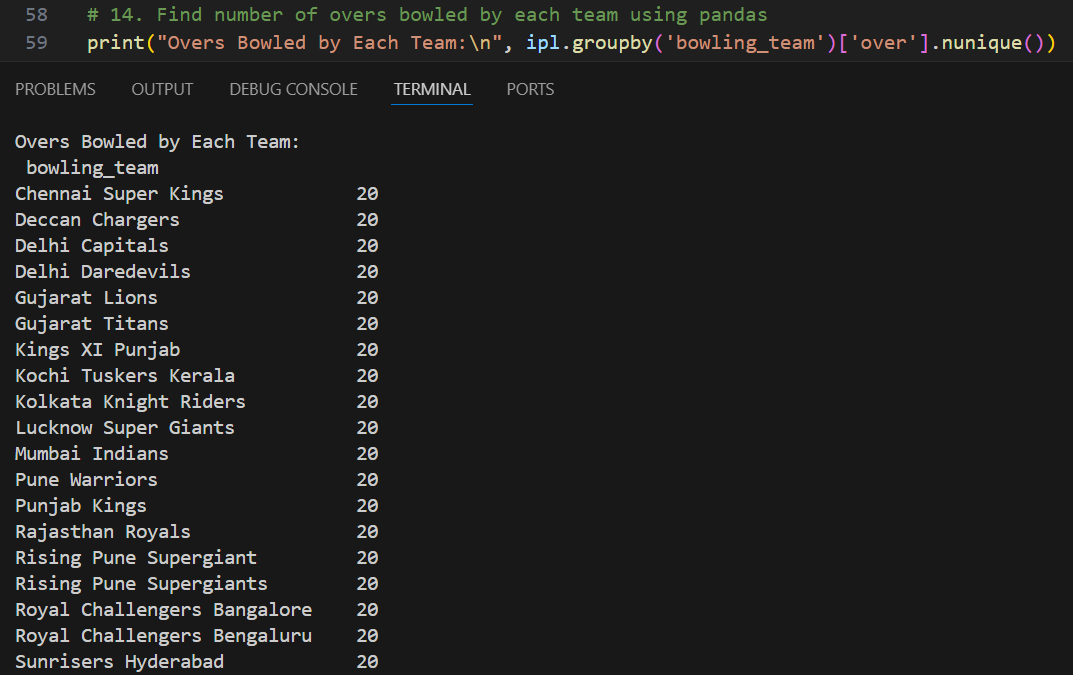
# 13. Find total number of boundaries (4s) scored using pandas

print("Total Boundaries (4s) Scored:", ipl[ipl['batsman\_runs']==4].shape[0])



# 14. Find number of overs bowled by each team using pandas

print("Overs Bowled by Each Team:\n", ipl.groupby('bowling\_team')['over'].nunique())



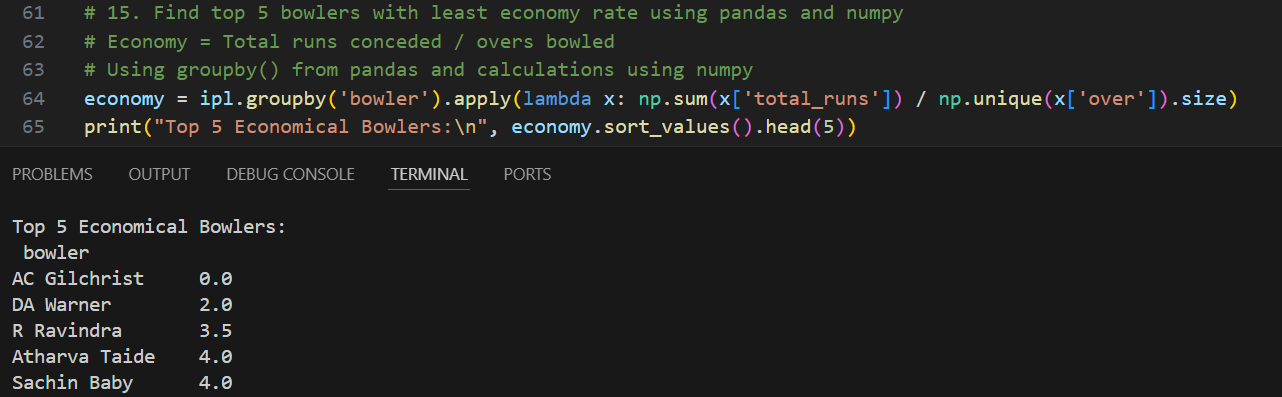
# 15. Find top 5 bowlers with least economy rate using pandas and numpy

# Economy = Total runs conceded / overs bowled

# Using groupby() from pandas and calculations using numpy

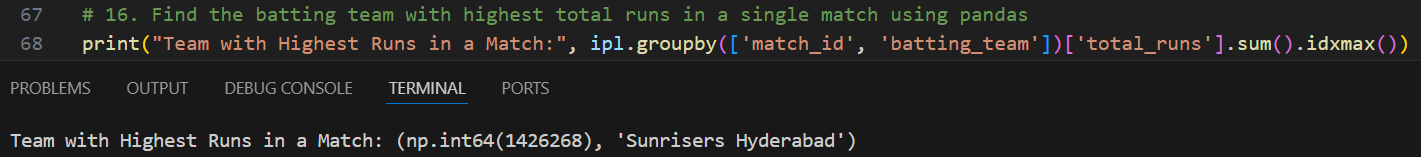
economy = ipl.groupby('bowler').apply(lambda x: np.sum(x['total\_runs']) / np.unique(x['over']).size)

print("Top 5 Economical Bowlers:\n", economy.sort\_values().head(5))



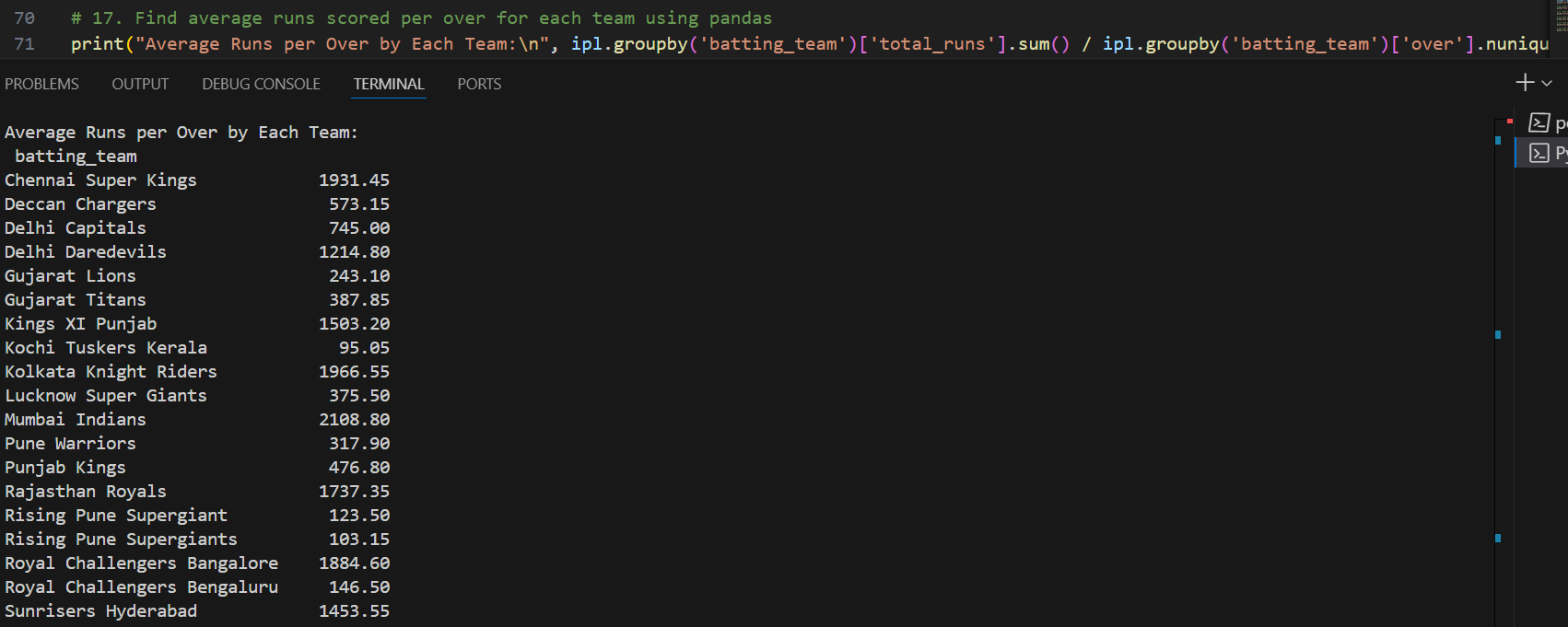
# 16. Find the batting team with highest total runs in a single match using pandas

print("Team with Highest Runs in a Match:", ipl.groupby(['match\_id', 'batting\_team'])['total\_runs'].sum().idxmax())



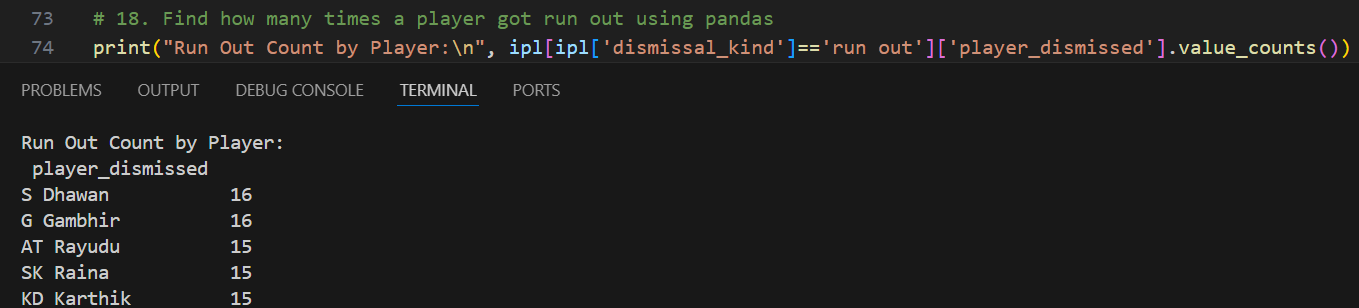
# 17. Find average runs scored per over for each team using pandas

print("Average Runs per Over by Each Team:\n", ipl.groupby('batting\_team')['total\_runs'].sum() / ipl.groupby('batting\_team')['over'].nunique())



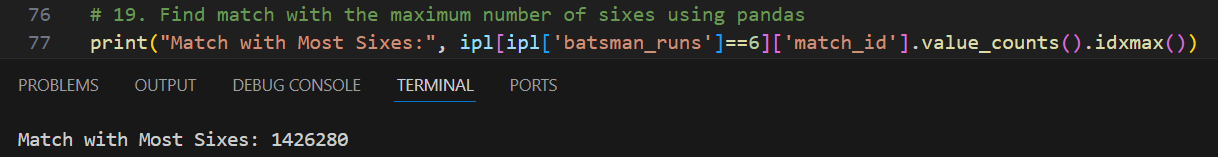
# 18. Find how many times a player got run out using pandas

print("Run Out Count by Player:\n", ipl[ipl['dismissal\_kind']=='run out']['player\_dismissed'].value\_counts())



# 19. Find match with the maximum number of sixes using pandas

print("Match with Most Sixes:", ipl[ipl['batsman\_runs']==6]['match\_id'].value\_counts().idxmax())



# 20. Find the fielder with most dismissals using pandas

print("Top Fielder in Dismissals:", ipl['fielder'].value\_counts().idxmax())

