

NAME: MANAS WARKE

DIVISION: ET2

BATCH: E23

ROLL NO: ET2 – 51

PRN: 202401070214

20 operations on IPL dataset

import pandas as pd

# Load the dataset

df = pd.read\_csv('matches.csv')

# 1. View first 5 rows

print(df.head())

# 2. View last 5 rows

print(df.tail())

# 3. Number of rows and columns

print(df.shape)

# 4. Column names

print(df.columns)

# 5. Dataset information

print(df.info())

# 6. Statistical summary

print(df.describe())

# 7. Check for missing values

print(df.isnull().sum())

# 8. Fill missing 'city' with 'Unknown'

df['city'] = df['city'].fillna('Unknown')

# 9. Fill missing 'umpire1' with 'None'

df['umpire1'] = df['umpire1'].fillna('None')

# 10. Rename column 'player\_of\_match' to 'POM'

df = df.rename(columns={'player\_of\_match': 'POM'})

# 11. Select one column

print(df['winner'])

# 12. Select multiple columns

print(df[['team1', 'team2', 'winner']])

# 13. Filter matches where "Chennai Super Kings" played

print(df[(df['team1'] == 'Chennai Super Kings') | (df['team2'] == 'Chennai Super Kings')])

# 14. Sort matches by season

print(df.sort\_values('season'))

# 15. Group by winner and count matches

print(df['winner'].value\_counts())

# 16. Add a new column "win\_by\_total" = win\_by\_runs + win\_by\_wickets

df['win\_by\_total'] = df['win\_by\_runs'] + df['win\_by\_wickets']

# 17. Drop the "umpire3" column

df = df.drop(columns=['umpire3'])

# 18. Reset the index

df = df.reset\_index(drop=True)

# 19. Set "id" as the new index

df = df.set\_index('id')

# 20. Save the new dataframe to a new CSV

df.to\_csv('simple\_ipl.csv')

output :

@ **First 5 rows**:  
→ Matches from 2017 (Sunrisers Hyderabad vs RCB, MI vs RPS, etc.)

@ **Last 5 rows**:  
→ Matches till 2019 (MI vs KKR, CSK vs MI, MI vs CSK final, etc.)

@ **Shape**:  
→ (756 rows, 18 columns)

@ **Columns**:  
→ ['id', 'Season', 'city', 'date', 'team1', 'team2', ...]

@ **Info**:  
→ Dataset has **some missing values** (e.g., city, umpire3 heavily missing)

@ **Describe (Stats)**:  
→ win\_by\_runs maximum is **146 runs**

@ **Missing Values**:  
→ 'umpire3' missing in **637 rows!**

@ **Winner column sample**:  
→ Sunrisers Hyderabad, RPS, KKR, KXIP, RCB...

@ **Selected Columns (team1, team2, winner)**:  
→ First few matches.

@ **Matches where CSK played**:  
→ Starting from 2008 season itself

@ **Sorted by Season**:  
→ **'season'**

@ **Wins per Team** (top 5 teams):

* Mumbai Indians : 109 wins
* Chennai Super Kings : 100 wins
* Kolkata Knight Riders: 92 wins
* RCB: 84 wins
* KXIP: 82 wins

@ **New CSV saved at**:  
 simple\_ipl\_output.csv