```
from google.colab import files
uploaded = files.upload()
```

Choose Files 2 files

- **deliveries.csv**(text/csv) - 484206 bytes, last modified: 4/6/2025 - 100% done
- **matches.csv**(text/csv) - 2753 bytes, last modified: 4/6/2025 - 100% done

```
Saving deliveries.csv to deliveries.csv
Saving matches.csv to matches.csv
```

```
import pandas as pd

deliveries_df = pd.read_csv('deliveries.csv')
matches_df = pd.read_csv('matches.csv')
```

```
# Load datasets
matches = pd.read_csv('matches.csv')
deliveries = pd.read_csv('deliveries.csv')

# Clean column names
matches.columns = matches.columns.str.strip()
deliveries.columns = deliveries.columns.str.strip()

# Check structure
matches.head()
deliveries.head()
```

ı_no	date	stage	venue	batting_team	bowling_team	innings	over	striker	
1	Mar 22, 2025	League stage	Eden Gardens, Kolkata	KKR	RCB	1	0.1	de Kock	I
1	Mar 22, 2025	League stage	Eden Gardens, Kolkata	KKR	RCB	1	0.2	de Kock	I
1	Mar 22, 2025	League stage	Eden Gardens, Kolkata	KKR	RCB	1	0.3	de Kock	I
1	Mar 22, 2025	League stage	Eden Gardens, Kolkata	KKR	RCB	1	0.4	de Kock	I
1	Mar 22, 2025	League stage	Eden Gardens, Kolkata	KKR	RCB	1	0.5	de Kock	I

Next steps: Generate code with deliveries | View recommended plots | New interactive sheet

```
# Split and combine using concat
part1 = deliveries.iloc[:5000]
part2 = deliveries.iloc[5000:]

combined = pd.concat([part1, part2])
print("Shape after concat:", combined.shape)
```

```
Shape after concat: (3778, 19)
```

```
appended = pd.concat([part1, part2], ignore_index=True)
print("Shape after appending using concat:", appended.shape)
```

```
Shape after appending using concat: (3778, 19)
```

```
import pandas as pd

# Load the CSV files (update the paths if needed)
deliveries = pd.read_csv('/content/deliveries.csv')
matches = pd.read_csv('/content/matches.csv')

# Clean column names to avoid hidden whitespace issues
deliveries.columns = deliveries.columns.str.strip()
```

```
matches.columns = matches.columns.str.strip()

# Perform inner join using appropriate keys
merged_inner = pd.merge(deliveries, matches, left_on='match_no', right_on='match_id', how='i

# Drop redundant columns from matches (those that duplicate deliveries)
merged_inner = merged_inner.drop(columns=['date_y', 'venue_y', 'stage_y'], errors='ignore')
merged_inner = merged_inner.rename(columns={
    'date_x': 'date',
    'venue_x': 'venue',
    'stage_x': 'stage'
})

# Display the result
print("Inner Join Shape:", merged_inner.shape)
merged_inner.head()
```

```
ape: (3778, 36)
```

date	stage	venue	batting_team	bowling_team	innings	over	striker	
Mar 22, 2025	League stage	Eden Gardens, Kolkata	KKR	RCB	1	0.1	de Kock	I
Mar 22, 2025	League	Eden Gardens, Kolkata	KKR	RCB	1	0.2	de Kock	I
Mar 22, 2025	League stage	Eden Gardens, Kolkata	KKR	RCB	1	0.3	de Kock	I
Mar 22, 2025	League stage	Eden Gardens, Kolkata	KKR	RCB	1	0.4	de Kock	I
Mar 22, 2025	League stage	Eden Gardens, Kolkata	KKR	RCB	1	0.5	de Kock	I

mns

```
merged_left = pd.merge(deliveries, matches, left_on='match_no', right_on='match_id', how='le
print("Left join result:", merged_left.shape)
```

```
Left join result: (3778, 39)
```

```
merged_outer = pd.merge(deliveries, matches, left_on='match_no', right_on='match_id', how='c
print("Outer join result:", merged_outer.shape)
```

```
Outer join result: (3778, 39)
```

```
team_runs = deliveries.groupby('batting_team')['runs_of_bat'].sum().sort_values(ascending=Fa
print(team_runs)
```

```
batting_team
LSG     740
SRH     715
KKR     608
MI      597
GT      567
RR      548
RCB     521
CSK     468
PBKS    401
DC      361
Name: runs_of_bat, dtype: int64
```

```
wickets = deliveries[deliveries['player_dismissed'].notnull()]
top_bowlers = wickets.groupby('bowler')['player_dismissed'].count().sort_values(ascending=Fa
print(top_bowlers.head(10))
```

```
bowler
Mitchell Starc    9
Noor Ahmad        9
Thakur            8
Hardik Pandya     8
Khaleel Ahmed     7
Arshdeep Singh    6
Vaibhav Arora     6
Sai Kishore       6
Kuldeep Yadav     6
Hazlewood         6
Name: player_dismissed, dtype: int64
```

```
matches_per_stage = matches.groupby('stage')['match_id'].count()
print(matches_per_stage)
```

```
stage
League    16
Name: match_id, dtype: int64
```

```
pivot_team_runs = pd.pivot_table(deliveries, index='match_no', columns='batting_team',
                                  values='runs_of_bat', aggfunc='sum', fill_value=0)
```

```
print(pivot_team_runs.head())
```

```
batting_team  CSK   DC   GT  KKR  LSG   MI  PBKS  RCB   RR  SRH
match_no
1               0    0    0  168    0    0     0  174    0    0
2               0    0    0    0    0    0     0    0  229  268
3             153    0    0    0    0  145     0    0    0    0
4               0  203    0    0  202    0     0    0    0    0
5               0    0  220    0    0    0   229    0    0    0
```

```
wickets['wicket'] = 1  # Add a column to count
pivot_wickets = pd.pivot_table(wickets, index='match_no', columns='bowler',
                                values='wicket', aggfunc='sum', fill_value=0)

print(pivot_wickets.head())
```

```
bowler    Akash Deep  Arshad Khan  Arshdeep Singh  Ashwani Kumar  Ashwin  \
match_no
1                  0            0               0              0       0
2                  0            0               0              0       0
3                  0            0               0              0       1
4                  0            0               0              0       0
5                  0            0               3              0       0

bowler    Avesh Khan  Bhuvneshwar  Boult  Chahal  Chahar  ...  Theekshana  \
match_no                                                  ...
1                  0            0      0       0       0  ...           0
2                  0            0      0       0       0  ...           2
3                  0            0      0       0       1  ...           0
4                  0            0      0       0       0  ...           0
5                  0            0      0       0       0  ...           0

bowler    Tushar Deshpande  Vaibhav Arora  Varun Chakaravarthy  \
match_no
1                        0              1                    1
2                        3              0                    0
3                        0              0                    0
4                        0              0                    0
5                        0              0                    0

bowler    Vignesh Puthur  Vipraj Nigam  Will Jacks  Yash Dayal  Zampa  \
match_no
1                      0             0           0           1      0
2                      0             0           0           0      1
3                      3             0           1           0      0
4                      0             1           0           0      0
5                      0             0           0           0      0

bowler    Zeeshan Ansari
match_no
1                      0
2                      0
```

```
3                       0
4                       0
5                       0

[5 rows x 67 columns]
<ipython-input-36-1118b74cbfa5>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_
  wickets['wicket'] = 1  # Add a column to count
```

```
# Step 1: Clean column names
deliveries.columns = deliveries.columns.str.strip()
matches.columns = matches.columns.str.strip()

# Step 2: Merge with correct columns
merged_df = pd.merge(
    deliveries,
    matches[['match_id', 'venue']],
    left_on='match_no',
    right_on='match_id',
    how='left'
)

# Step 3: Verify if 'venue' exists
print("Columns in merged_df:", merged_df.columns.tolist())

# Step 4: Create pivot table if 'venue' is present
if 'venue' in merged_df.columns:
    pivot_batsman_venue = pd.pivot_table(
        merged_df,
        index='striker',
        columns='venue',
        values='runs_of_bat',
        aggfunc='sum',
        fill_value=0
    )
    display(pivot_batsman_venue.head())
else:
    print("❌ 'venue' column not found in merged DataFrame.")
```

```
Columns in merged_df: ['match_no', 'date', 'stage', 'venue_x', 'batting_team', 'bowling_
❌ 'venue' column not found in merged DataFrame.
```