

In [2]: `import pandas as pd
import numpy as np`

In [93]: `df = pd.read_csv("deliveries.csv")`

Which bowler has hit for maximum number of 1s

In [148]: `df2 = df.query(" total_runs == '1' ")
df2.groupby('bowler')[['total_runs']].count().sort_values('total_runs',ascending = False).head(5)`

Out[148]:

	total_runs
bowler	
Harbhajan Singh	1549
R Ashwin	1415
A Mishra	1414
PP Chawla	1311
DJ Bravo	1189

Which batsman got stumped out maximum number of times?

In [92]: `df.query("dismissal_kind=='stumped'") [['player_dismissed','dismissal_kind']].groupby('player_dismissed').count().sort_values('dismissal_kind',ascending = False).head(3)`

Out[92]:

	dismissal_kind
player_dismissed	
SK Raina	8
RV Uthappa	7
AT Rayudu	7

In [8]: `df2 = pd.read_csv("matches.csv")`

Which team has won maximum number of matches while chasing?

In [14]: `df3= df2.query(("toss_decision=='field' and toss_winner == winner or toss_decision=='bat' and toss_winner != winner")) [['id','winner']].groupby('winner').count().reset_index().sort_values('id',ascending=False).head(5)`

Out[14]:

	winner	id
7	Kolkata Knight Riders	58
8	Mumbai Indians	51
0	Chennai Super Kings	49
13	Royal Challengers Bangalore	49
10	Rajasthan Royals	46

Which team has lost maximum number of matches while chasing?

In [16]: `df4 = df2.query(("toss_decision=='field' & (toss_winner != winner) or toss_decision=='bat' & (toss_winner == winner)")) [['id','winner']].groupby('winner').count().reset_index().sort_values('id',ascending=False).head(5)`

Out[16]:

	winner	id
8	Mumbai Indians	58
0	Chennai Super Kings	51
5	Kings XI Punjab	38
12	Royal Challengers Bangalore	35
7	Kolkata Knight Riders	34

Which toss decision is best for each ground?

In [9]: `def decision(row):
 if row.winner_field > row.winner_bat:
 return 'field'
 else:
 return 'bat'`

In [10]: `df_field_win = df2[df2['toss_winner']==df2['winner']].query("toss_decision=='field'") [['winner','venue']].groupby('venue').count()
df_bat_win = df2[df2['toss_winner']==df2['winner']].query("toss_decision=='bat'") [['winner','venue']].groupby('venue').count()
df_bat_win
df_new = df_field_win.merge(df_bat_win,on='venue',suffixes=('_field','_bat'))
df_new['decision'] = df_new.apply(decision,axis=1)
df_new[['decision']]`

Out[10]:

	decision
venue	
Barabati Stadium	field
Brabourne Stadium	bat
De Beers Diamond Oval	bat
Dr DY Patil Sports Academy	field
Dr. Y.S. Rajasekhara Reddy ACA-VDCA Cricket Stadium	bat
Dubai International Cricket Stadium	field
Eden Gardens	field
Feroz Shah Kotla	field
Feroz Shah Kotla Ground	field
Himachal Pradesh Cricket Association Stadium	field
JSCA International Stadium Complex	field
Kingsmead	bat
M Chinnaswamy Stadium	field
M. A. Chidambaram Stadium	field
MA Chidambaram Stadium, Chepauk	bat
Maharashtra Cricket Association Stadium	field
Nehru Stadium	bat
Newlands	bat
OUTsurance Oval	bat
Punjab Cricket Association IS Bindra Stadium, Mohali	field
Punjab Cricket Association Stadium, Mohali	field
Rajiv Gandhi International Stadium, Uppal	field
Rajiv Gandhi Intl. Cricket Stadium	bat
Sardar Patel Stadium, Motera	bat
Sawai Mansingh Stadium	field
Shaheed Veer Narayan Singh International Stadium	field
Sheikh Zayed Stadium	bat
SuperSport Park	field
Wankhede Stadium	field

Which player has won maximum number of MoM awards while chasing?

In [77]: `df2 = df2[df2['win_by_wickets']>0][['player_of_match','win_by_wickets']]
playerdm= df2.groupby('player_of_match').count()
playerdm[playerdm.win_by_wickets==playerdm.win_by_wickets.max()]`

Out[77]:

	win_by_wickets
player_of_match	
YK Pathan	11

Which batsman has played maximum number of dot balls?

In [87]: `dot_balls_df = df.query('total_runs==0')[['batsman','total_runs']]
new_df = dot_balls_df.groupby('batsman').count()
new_df[new_df.total_runs == new_df.total_runs.max()]`

Out[87]:

	total_runs
batsman	
V Kohli	1369

Which non-striker has been part of maximum number of runouts?

In [191]: `df_new = df.query('dismissal_kind == "run out"')[['dismissal_kind','non_striker']]
df4 = df_new.groupby('non_striker').count()
df4[df4.dismissal_kind == df4.dismissal_kind.max()]`

Out[191]:

	dismissal_kind
non_striker	
KD Karthik	16
SK Raina	16

The number of venues in each season

In [4]: `matches = pd.read_csv("matches.csv")
matches[['season','venue']].groupby('season').count()`

Out[4]:

	venue
season	
2008	58
2009	57
2010	60
2011	73
2012	74
2013	76
2014	60
2015	59
2016	60
2017	59
2018	60
2019	60

How many matches played by each team and how many won

In [158]: `new_df = pd.concat([matches['team1'],matches['team2']]).value_counts().reset_index()
new_df.columns=['index','matches_played']
match_won = matches['winner'].value_counts().reset_index()
new_df.merge(match_won,on='index')`

Out[158]:

	index	matches_played	winner
0	Mumbai Indians	187	109
1	Royal Challengers Bangalore	180	84
2	Kolkata Knight Riders	178	92
3	Kings XI Punjab	176	82
4	Chennai Super Kings	164	100
5	Delhi Daredevils	161	67
6	Rajasthan Royals	147	75
7	Sunrisers Hyderabad	108	58
8	Deccan Chargers	75	29
9	Pune Warriors	46	12
10	Gujarat Lions	30	13
11	Rising Pune Supergiant	16	10
12	Delhi Capitals	16	10
13	Kochi Tuskers Kerala	14	6
14	Rising Pune Supergiants	14	5

Which venue was the most used throughout the history of IPL

In [166]: `df2['venue'].value_counts().reset_index().head(5)`

Out[166]:

	index	venue
0	Eden Gardens	77
1	Wankhede Stadium	73
2	M Chinnaswamy Stadium	73
3	Feroz Shah Kotla	67
4	Rajiv Gandhi International Stadium, Uppal	56

Most favorite umpire

In [178]: `df_umpires = pd.concat([df2['umpire1'],df2['umpire2']])
df_umpires = df_umpires.value_counts().reset_index()
df_umpires.columns = ['umpire','matches']
df_umpires.head(3)`

Out[178]:

	umpire	matches
0	S Ravi	106
1	HDPK Dharmasena	87
2	C Shamshuddin	73

Total number of ties in each season

In [189]: `df2.query('result=="tie"')[['result','season']].groupby('season').count()`

Out[189]:

	result
season	
2009	1
2010	1
2013	2
2014	1