

```
In [1]: #Loading the required libraries
import pandas as pd
from matplotlib import pyplot as plt
import seaborn as sns

In [2]: #Loading the IPL data set
ipl = pd.read_csv('matches.csv')
ipl.head()
```

	Match_SK	match_id	team1	team2	mdate	season	venue	city	country	toss_winner	match_winner	toss_decision
0	0	335987	Royal Challengers Bangalore	Kolkata Knight Riders	4/18/2008	2008	Chinnaswamy Stadium	M Bangalore	India	Royal Challengers Bangalore	Kolkata Knight Riders	field
1	1	335988	Kings XI Punjab	Chennai Super Kings	4/19/2008	2008	Punjab Cricket Association Stadium, Mohali	Chandigarh	India	Chennai Super Kings	Chennai Super Kings	bat
2	2	335989	Delhi Daredevils	Rajasthan Royals	4/19/2008	2008	Feroz Shah Kotla	Delhi	India	Rajasthan Royals	Delhi Daredevils	bat
3	3	335990	Mumbai Indians	Royal Challengers Bangalore	4/20/2008	2008	Wankhede Stadium	Mumbai	India	Mumbai Indians	Royal Challengers Bangalore	bat
4	4	335991	Kolkata Knight Riders	Deccan Chargers	4/20/2008	2008	Eden Gardens	Kolkata	India	Deccan Chargers	Kolkata Knight Riders	bat

```
In [5]: #Shape of the data
ipl.shape
```

```
Out[5]: (637, 17)
```

```
In [6]: #Get the count/frequency of Man of the match award given to different players over the years.
ipl['manofthematch'].value_counts()
```

```
Out[6]: CH Gayle      18
        VK Patkar    16
        AB de Villiers 15
        DA Warner    15
        SR Raina     14
        RG Sharma    14
        MS Dhoni     13
        C Ganghli    13
        MEK Russey   12
        AM Rahane    12
        Name: manofthematch, dtype: int64
```

```
In [8]: #Get the frequency top 10 players with most man of the match award.
ipl['manofthematch'].value_counts()[0:10].keys()
```

```
Out[8]: CH Gayle      18
        VK Patkar    16
        AB de Villiers 15
        DA Warner    15
        SR Raina     14
        RG Sharma    14
        MS Dhoni     13
        C Ganghli    13
        MEK Russey   12
        AM Rahane    12
        Name: manofthematch, dtype: int64
```

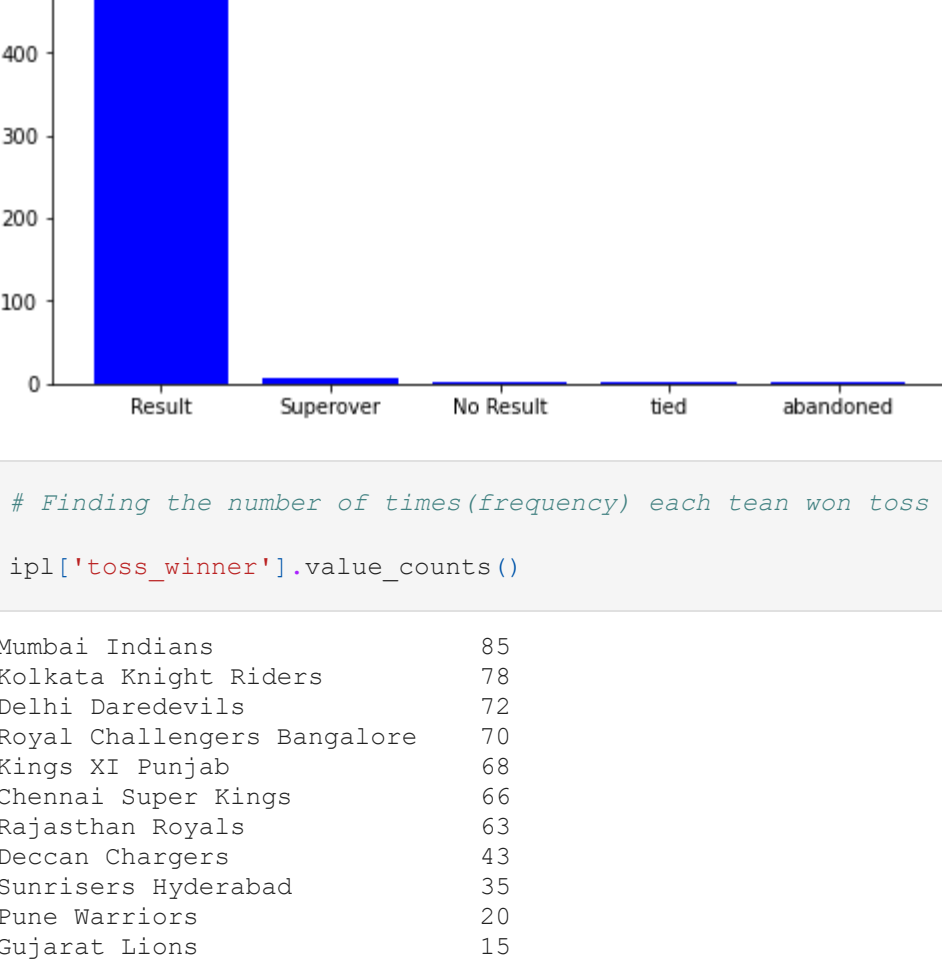
```
In [9]: #Get the frequency top 5 players with most man of the match award.
ipl['manofthematch'].value_counts()[0:5].keys()
```

```
Out[9]: CH Gayle      18
        VK Patkar    16
        AB de Villiers 15
        DA Warner    15
        SR Raina     14
        Name: manofthematch, dtype: int64
```

```
In [13]: #Getting only the names (keys) for above outcomes.
list(ipl['manofthematch'].value_counts()[0:5].keys())
```

```
Out[13]: ['CH Gayle', 'VK Patkar', 'AB de Villiers', 'DA Warner', 'SR Raina']
```

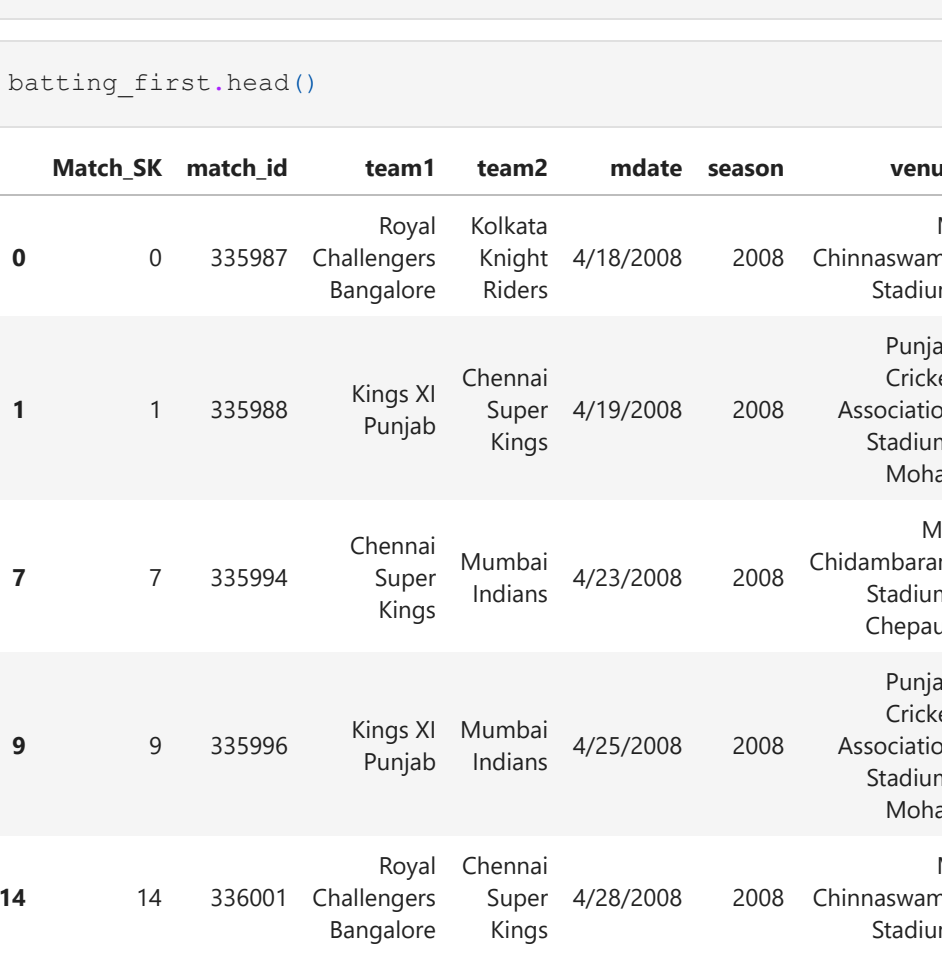
```
In [46]: #Now Bar plot for top 5 players with most man of the match award.
plt.figure(figsize=(8,5))
plt.bar(list(ipl['manofthematch'].value_counts()[0:5].keys()), list(ipl['manofthematch'].value_counts()[0:5]), color='b')
plt.show()
```



```
In [38]: #Getting the frequency of match outcome column
ipl['Outcome_Type'].value_counts()
```

```
Out[38]: Result      626
        Superover    3
        No Result    3
        tied         1
        abandoned    1
        Name: Outcome_Type, dtype: int64
```

```
In [55]: #Bar Plot for above outcome
plt.figure(figsize=(8,5))
plt.bar(list(ipl['Outcome_Type'].value_counts().keys()), list(ipl['Outcome_Type'].value_counts()), color='b')
plt.show()
```



```
In [52]: # Finding the number of times(frequency) each team won toss
ipl['toss_winner'].value_counts()
```

```
Out[52]: Mumbai Indians      85
        Kolkata Knight Riders 78
        Delhi Daredevils    70
        Royal Challengers Bangalore 70
        Kings XI Punjab     66
        Chennai Super Kings 66
        Rajasthan Royals    63
        Deccan Chargers     63
        Sunrisers Hyderabad 35
        Pune Warriors       20
        Gujarat Lions       15
        Rising Pune Supergiants 12
        Kochi Tuskers Kerala 8
        Rising Pune Supergiant 1
        Name: toss_winner, dtype: int64
```

```
In [74]: #extracting the records where a team won by batting first
batting_first = ipl[ipl['win_type']=='wickets']
```


```
In [76]: batting_first.head()
```

	Match_SK	match_id	team1	team2	mdate	season	venue	city	country	toss_winner	match_winner	toss_decision
0	0	335987	Royal Challengers Bangalore	Kolkata Knight Riders	4/18/2008	2008	Chinnaswamy Stadium	M Bangalore	India	Royal Challengers Bangalore	Kolkata Knight Riders	field
1	1	335988	Kings XI Punjab	Chennai Super Kings	4/19/2008	2008	Punjab Cricket Association Stadium, Mohali	Chandigarh	India	Chennai Super Kings	Chennai Super Kings	bat
7	7	335994	Chennai Super Kings	Mumbai Indians	4/23/2008	2008	MA Chidambaram Stadium, Chennai	Chennai	India	Mumbai Indians	Chennai Super Kings	field
9	9	335996	Kings XI Punjab	Mumbai Indians	4/25/2008	2008	Punjab Cricket Association Stadium, Mohali	Chandigarh	India	Mumbai Indians	Kings XI Punjab	field
14	14	336001	Royal Challengers Bangalore	Chennai Super Kings	4/28/2008	2008	M Chinnaswamy Stadium	Bangalore	India	Chennai Super Kings	Chennai Super Kings	bat

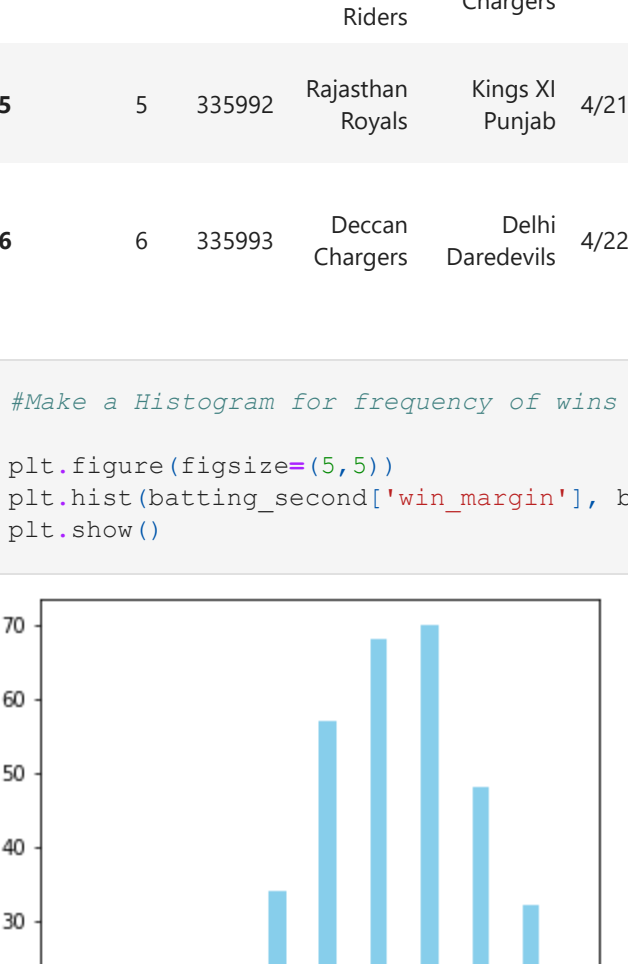
```
In [78]: #Finding out the number of wins of each team when they batted first.
batting_first['match_winner'].value_counts()
```

```
Out[78]: Mumbai Indians      47
        Chennai Super Kings 46
        Kings XI Punjab     34
        Kolkata Knight Riders 31
        Royal Challengers Bangalore 31
        Rajasthan Royals    25
        Sunrisers Hyderabad 24
        Delhi Daredevils    21
        Deccan Chargers     18
        Rising Pune Supergiants 7
        Pune Warriors       6
        Kochi Tuskers Kerala 2
        tied                1
        abandoned          1
        Gujarat Lions       1
        Name: match_winner, dtype: int64
```

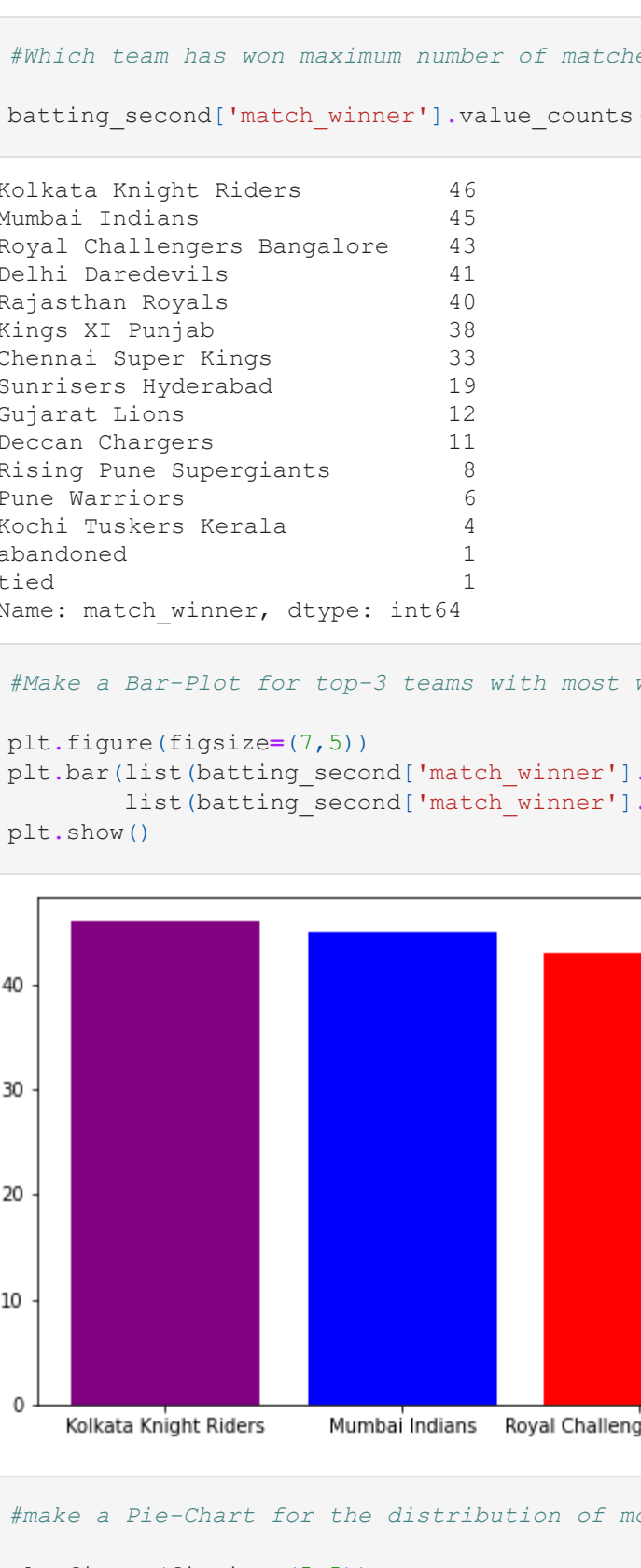
```
In [82]: #Making a histogram
plt.figure(figsize=(5,3))
plt.hist(batting_first['win_margin'])
plt.title("Distribution of runs")
plt.xlabel("Runs of Matches")
plt.ylabel("No. of Matches")
plt.show()
```



```
In [88]: #Making a bar-plot for top 3 teams with most wins after batting first.
plt.figure(figsize=(6,5))
plt.bar(list(batting_first['match_winner'].value_counts()[0:3].keys()), list(batting_first['match_winner'].value_counts()[0:3]), color = ['orange', 'yellow', 'blue'])
plt.show()
```



```
In [89]: #Make a piechart - Which team has the most win percentage after batting first.
plt.pie(list(batting_first['match_winner'].value_counts()), labels=list(batting_first['match_winner'].value_counts().keys()), autopct='%0.1f%%')
plt.show()
```



```
In [90]: #Records for winnings when batted second(win by wickets)
ipl.head()
```

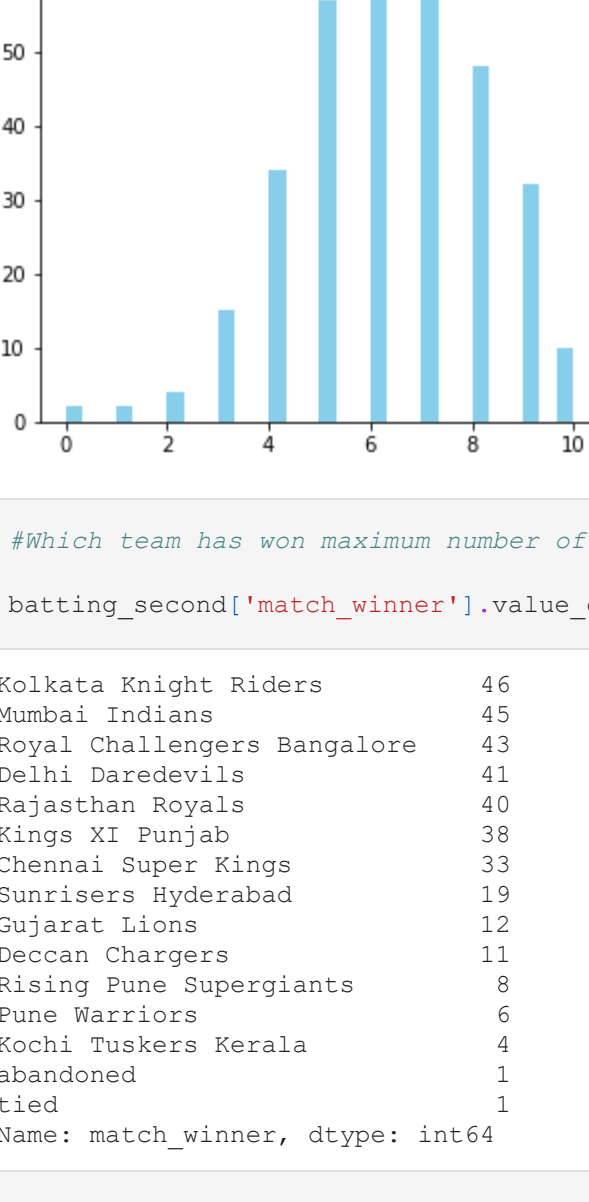
	Match_SK	match_id	team1	team2	mdate	season	venue	city	country	toss_winner	match_winner	toss_decision
0	0	335987	Royal Challengers Bangalore	Kolkata Knight Riders	4/18/2008	2008	Chinnaswamy Stadium	M Bangalore	India	Royal Challengers Bangalore	Kolkata Knight Riders	field
1	1	335988	Kings XI Punjab	Chennai Super Kings	4/19/2008	2008	Punjab Cricket Association Stadium, Mohali	Chandigarh	India	Chennai Super Kings	Chennai Super Kings	bat
2	2	335989	Delhi Daredevils	Rajasthan Royals	4/19/2008	2008	Feroz Shah Kotla	Delhi	India	Rajasthan Royals	Delhi Daredevils	bat
3	3	335990	Mumbai Indians	Royal Challengers Bangalore	4/20/2008	2008	Wankhede Stadium	Mumbai	India	Mumbai Indians	Royal Challengers Bangalore	bat
4	4	335991	Kolkata Knight Riders	Deccan Chargers	4/20/2008	2008	Eden Gardens	Kolkata	India	Deccan Chargers	Kolkata Knight Riders	bat

```
In [92]: batting_second = ipl[ipl['win_type'] != 'runs']
```

```
In [93]: batting_second.head()
```

	Match_SK	match_id	team1	team2	mdate	season	venue	city	country	toss_winner	match_winner	toss_decision
2	2	335989	Delhi Daredevils	Rajasthan Royals	4/19/2008	2008	Feroz Shah Kotla	Delhi	India	Rajasthan Royals	Delhi Daredevils	bat
3	3	335990	Mumbai Indians	Royal Challengers Bangalore	4/20/2008	2008	Wankhede Stadium	Mumbai	India	Mumbai Indians	Royal Challengers Bangalore	bat
4	4	335991	Kolkata Knight Riders	Deccan Chargers	4/20/2008	2008	Eden Gardens	Kolkata	India	Deccan Chargers	Kolkata Knight Riders	bat
5	5	335992	Rajasthan Royals	Kings XI Punjab	4/21/2008	2008	Sawai Mansingh Stadium	Jaipur	India	Kings XI Punjab	Rajasthan Royals	bat
6	6	335993	Deccan Chargers	Delhi Daredevils	4/22/2008	2008	Rajiv Gandhi International Stadium, Uppal	Hyderabad	India	Deccan Chargers	Delhi Daredevils	bat

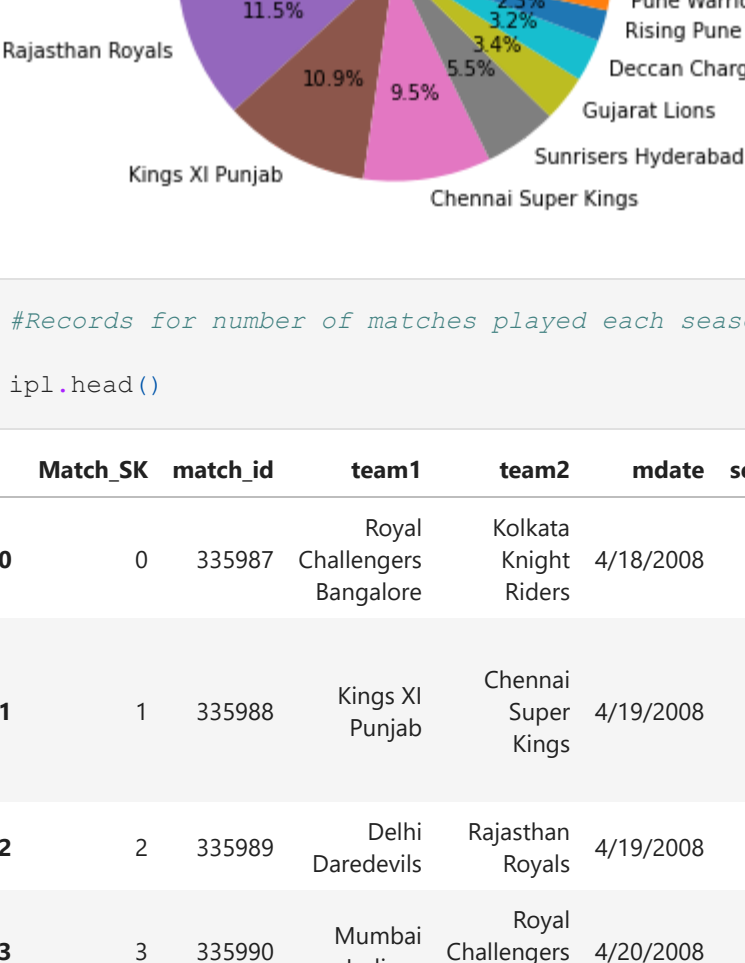
```
In [95]: #Making a Histogram for frequency of wins w.r.t number of wickets
plt.figure(figsize=(5,3))
plt.hist(batting_second['win_margin'], bins=30, color='skyblue')
plt.show()
```



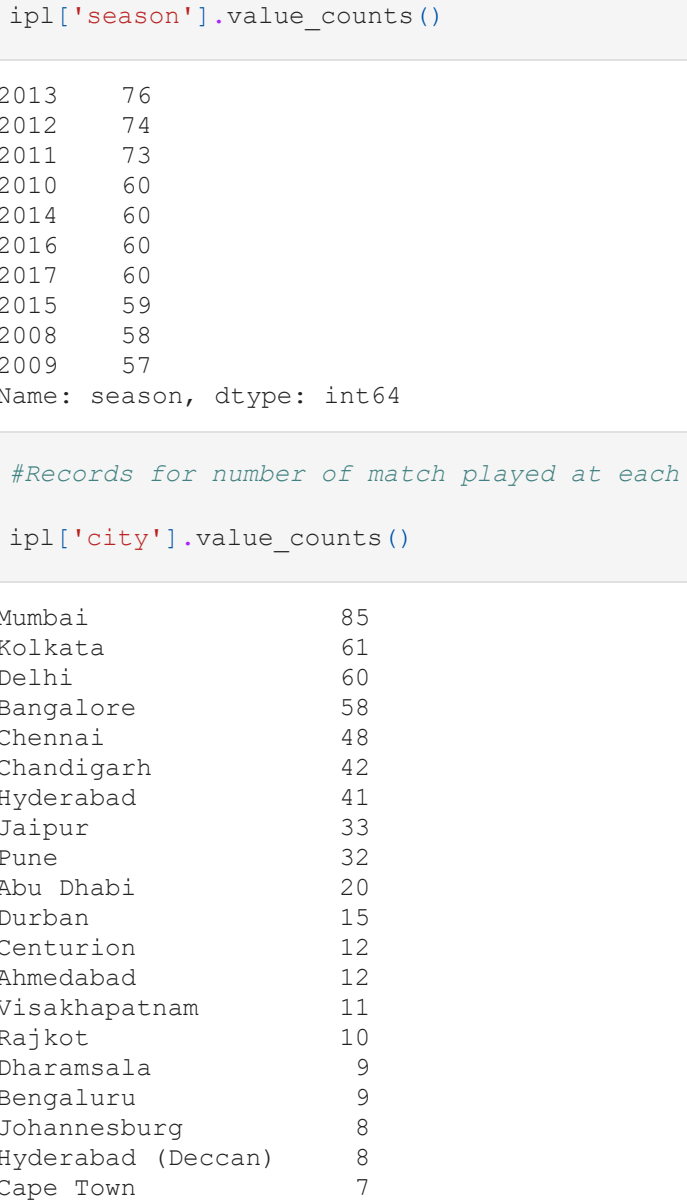
```
In [96]: #Which team has won maximum number of matches when batted second
batting_second['match_winner'].value_counts()
```

```
Out[96]: Kolkata Knight Riders      46
        Mumbai Indians          45
        Royal Challengers Bangalore 43
        Delhi Daredevils        41
        Rajasthan Royals        40
        Kings XI Punjab         38
        Chennai Super Kings     33
        Sunrisers Hyderabad     19
        Gujarat Lions           12
        Deccan Chargers         11
        Rising Pune Supergiants 6
        Kochi Tuskers Kerala    4
        tied                    1
        Name: match_winner, dtype: int64
```

```
In [98]: #Make a Bar-Plot for top-3 teams with most wins after batting second.
plt.figure(figsize=(7,5))
plt.bar(list(batting_second['match_winner'].value_counts()[0:3].keys()), list(batting_second['match_winner'].value_counts()[0:3]), color= ['purple', 'blue', 'red'])
plt.show()
```



```
In [100]: #Make a Pie-Chart for the distribution of most wins after batting second.
plt.figure(figsize=(5,3))
plt.pie(list(batting_second['match_winner'].value_counts()), labels=list(batting_second['match_winner'].value_counts().keys()), autopct='%0.1f%%')
plt.show()
```



```
In [101]: #Records for number of matches played each season.
ipl.head()
```

	Match_SK	match_id	team1	team2	mdate	season	venue	city	country	toss_winner	match_winner	toss_decision
0	0	335987	Royal Challengers Bangalore	Kolkata Knight Riders	4/18/2008	2008	Chinnaswamy Stadium	M Bangalore	India	Royal Challengers Bangalore	Kolkata Knight Riders	field
1	1	335988	Kings XI Punjab	Chennai Super Kings	4/19/2008	2008	Punjab Cricket Association Stadium, Mohali	Chandigarh	India	Chennai Super Kings	Chennai Super Kings	bat
2	2	335989	Delhi Daredevils	Rajasthan Royals	4/19/2008	2008	Feroz Shah Kotla	Delhi	India	Rajasthan Royals	Delhi Daredevils	bat
3	3	335990	Mumbai Indians	Royal Challengers Bangalore	4/20/2008	2008	Wankhede Stadium	Mumbai	India	Mumbai Indians	Royal Challengers Bangalore	bat
4	4	335991	Kolkata Knight Riders	Deccan Chargers	4/20/2008	2008	Eden Gardens	Kolkata	India	Deccan Chargers	Kolkata Knight Riders	bat

```
In [102]: ipl['season'].value_counts()
```

```
Out[102]: 2013      76
        2012     73
        2011     73
        2010     60
        2014     60
        2016     60
        2017     60
        2015     59
        2008     58
        2009     57
        Name: season, dtype: int64
```

```
In [104]: #Records for number of match played at each city.
ipl['city'].value_counts()
```

```
Out[104]: Mumbai      85
        Kolkata    81
        Delhi      60
        Bangalore  58
        Chennai    48
        Chandigarh 42
        Hyderabad  41
        Jaipur     32
        Pune       32
        Abu Dhabi  20
        Durban     15
        Centurion  12
        Ahmedabad  12
        Visakhapatnam 11
        Rajkot     10
        Dharamsala 9
        Bengaluru  9
        Johannesburg 8
        Hyderabad (Deccan) 8
        Cape Town  7
        Cuttack    7
        Port Elizabeth 7
        Ranchi     7
        Raipur     6
        Kochi      5
        Indore     5
        Kanpur     4
        Mohali     4
        Nagpur     3
        Kimberley  3
        East London 3
        Bloemfontein 2
        Name: city, dtype: int64
```

```
In [105]: #Finding out how many times a team has won the match after winning the toss.
import numpy as np
np.sum(ipl['toss_winner']==ipl['match_winner'])
```

```
Out[105]: 324
```

```
In [106]: ipl['toss_winner']==ipl['match_winner']
```

```
Out[106]: 0    False
        1     True
        2    False
        3    False
        4    False
        ...
        632    True
        633    False
        634    True
        635    True
        636    True
        Length: 637, dtype: bool
```

```
In [107]: 324/637
```

```
Out[107]: 0.508634229199372
```

```
In [108]: deliveries=pd.read_csv('deliveries.csv')
deliveries.head()
```

	match_id	inning	batting_team	bowling_team	over	ball	batsman	non_striker	bowler	is_super_over	...	bye_runs	legbye_runs	noball_runs
0	1	1	1	1	1	1	DA Warner	S Dhawan	TS Mills	0	...	0	0	0
1	1	1	1	1	2	2	DA Warner	S Dhawan	TS Mills	0	...	0	0	0
2	1	1	1	1	3	3	DA Warner	S Dhawan	TS Mills	0	...	0	0	0
3	1	1	1	1	4	4	DA Warner	S Dhawan	TS Mills	0	...	0	0	0
4	1	1	1	1	5	5	DA Warner	S Dhawan	TS Mills	0	...	0	0	0

5 rows × 21 columns

```
In [110]: deliveries['match_id'].unique()
```

```
Out[110]: array([ 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637], dtype=int64)
```

```
In [ ]: #The above results says that there are 636 different matches present in this record.
```

```
In [111]: match_1 = deliveries[deliveries['match_id']==1]
```

```
In [112]: match_1.head()
```

	match_id	inning	batting_team	bowling_team	over	ball	batsman	non_striker	bowler	is_super_over	...	bye_runs	legbye_runs	noball_runs
0	1	1	1	1	1	1	DA Warner	S Dhawan	TS Mills	0	...	0	0	0
1	1	1	1	1	2	2	DA Warner	S Dhawan	TS Mills	0	...	0	0	0
2	1	1	1	1	3	3	DA Warner	S Dhawan	TS Mills	0	...	0	0	0
3	1	1	1	1	4	4	DA Warner	S Dhawan	TS Mills	0	...	0	0	0
4	1	1	1	1	5	5	DA Warner	S Dhawan	TS Mills	0	...	0	0	0

5 rows × 21 columns

```
In [113]: match_1.shape
```

```
Out[113]: (248, 21)
```

```
In [119]: SRH = match_1[match_1['inning']==1]
```

```
In [120]: SRH['batsman_runs'].value_counts()
```

```
Out[120]: 1    57
        0    32
        17    7
        2     9
        6     9
        8     8
        Name: batsman_runs, dtype: int64
```

```
In [121]: SRH['dismissal_kind'].value_counts()
```

```
Out[121]: caught      3
        bowled    3
        Name: dismissal_kind, dtype: int64
```

```
In [122]: RCB = match_1[match_1['inning']==2]
```

```
In [123]: RCB['batsman_runs'].value_counts()
```

```
Out[123]: 0    49
        1    44
        4    15
        6     8
        2     7
        Name: batsman_runs, dtype: int64
```

```
In [124]: RCB['dismissal_kind'].value_counts()
```

```
Out[124]: caught      6
        run out    2
        bowled     2
        Name: dismissal_kind, dtype: int64
```