This document contains just the first step to VDA (Visual Data Analysis) and then comes in the second part of the learning, the self learning part. There will be an assignment on all of it to help consolidate your learning. So let's continue with the IPL dataset we were using yesterday.

Let's try to find which team has a trait of snatching it from the death most times while chasing a total, i.e. which team has a trait of winning a match by a small number of wickets the most.

```
import seaborn as sns          #seaborn is a library used for VDA
```

First let's create a new dataframe with matches won by chasing.

```
dfNew=df[df['win_by_wickets']!=0]
```

```
sns.set(rc={'figure.figsize':(45,8.27)})
#this is used to set the size of the diagram you get
sns.countplot(x='winner', hue='win_by_wickets', data=dfNew);
#used to plot the graph with the various teams across the x-axis and the
other details are self explanatory
```

It can be seen from this plot that CSK and KKR are probably the teams with highest wins on the lower win by wickets spectrum.

Let's try and check if our visual analysis is actually right.

So, let's learn about the crosstab function which is like a multidimensional value_counts() function.
To use this,

```
pd.crosstab(dfNew['winner'], dfNew['win_by_wickets'])
```

Aah, perfect, seems like our assumption was right.

Let's try and find which team has the maximum wins by a certain number of wickets using this table.

```
dfFinal=pd.crosstab(dfNew['winner'], dfNew['win_by_wickets'])
dfFinal
```

```
dfFinal.max()
```

```
[25] dfFinal=pd.crosstab(dfNew['winner'], dfNew['win_by_wickets'])
     dfFinal
```

win_by_wickets	1	2	3	4	5	6	7	8	9	10
winner										
Chennai Super Kings	1	1	1	5	11	4	7	6	4	1
Deccan Chargers	0	0	0	0	2	5	2	1	0	1
Delhi Daredevils	0	1	1	3	6	10	8	6	6	1
Gujarat Lions	0	0	2	1	2	4	3	0	0	0
Kings XI Punjab	0	0	1	4	2	16	9	4	2	1
Kochi Tuskers Kerala	0	0	0	0	0	0	2	2	0	0
Kolkata Knight Riders	1	1	2	5	8	10	10	10	4	1
Mumbai Indians	0	1	0	7	11	8	7	7	4	1
Pune Warriors	0	0	0	1	1	1	3	0	0	0
Rajasthan Royals	0	0	6	4	6	7	7	8	2	1
Rising Pune Supergiant	0	0	0	1	1	1	1	0	1	0
Rising Pune Supergiants	0	0	0	1	0	0	1	0	1	0
Royal Challengers Bangalore	0	1	2	4	8	8	8	4	8	3
Sunrisers Hyderabad	1	0	1	1	5	2	6	3	3	1

```
dfFinal.max()
```

```
win_by_wickets
1       1
2       1
3       6
4       7
5      11
6      16
7      10
8      10
9       8
10      3
dtype: int64
```

From the above tables, match values and there, you have the answer.

Links for further reading and consolidation of what you learn yesterday:

- https://www.kaggle.com/kashnitsky/topic-1-exploratory-data-analysis-with-pandas : This helps you understand how you can have a basic probabilistic model to guess a feature. Teaches you how to use the methods learnt yesterday to get something substantial from the data.
- https://www.kaggle.com/kashnitsky/topic-2-visual-data-analysis-in-python
- https://www.datacamp.com/community/tutorials/exploratory-data-analysis-python
- https://medium.com/python-pandemonium/introduction-to-exploratory-data-analysis-in-python-8b6bcb55c190

There are infinite resources online, you are free to explore and ask doubts :)