Chongqing University of technology

Md Anower Hossain(an hao ming)

Student ID: 62017010084
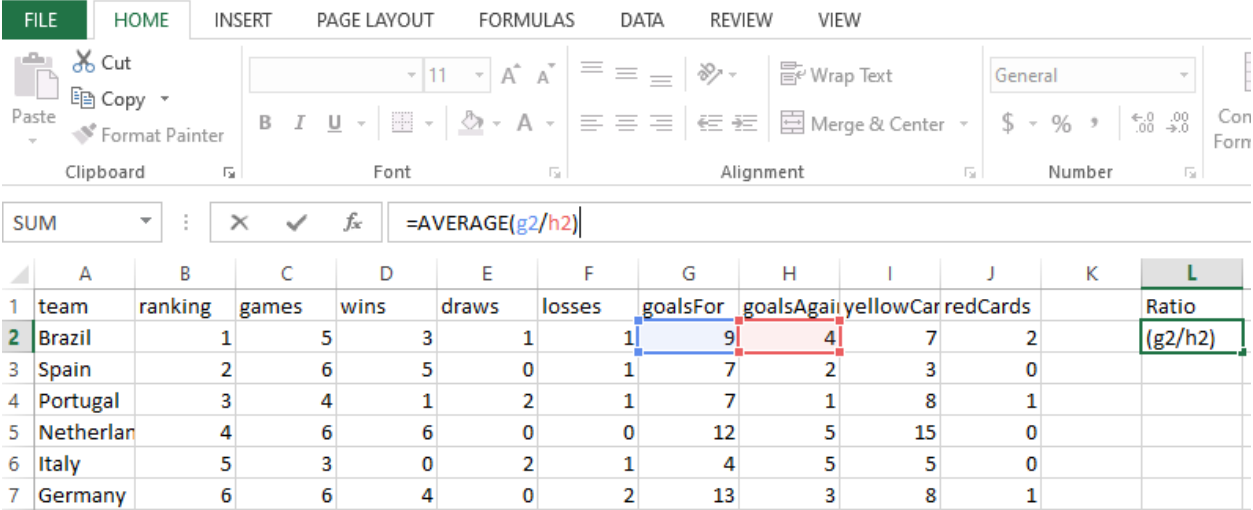
Assignments -1

7th Semester

Sub: Big data technology and practice

Problem1: Which team has the highest ratio of goals for/goals against?

Ans: in this problem we should use teams.csv data for the answer.

    I.   First of all open the teams.csv in Ms excel and use L column for answering the question.

    II.   Write Ration in L column first row.

    III.   Write the function below the Ratio =AVERAGE(G2/H2) press Enter and get the result for one country.

    IV.   Keep mouse pointer on the result and notice plus (+) sign and scroll down how much want. Sort it to Z-A for find the highest number.

	A	B	C	D	E	F	G	H	I	J	K	L
1	team	ranking	games	wins	draws	losses	goalsFor	goalsAgai	yellowCar	redCards		ratio
2	Portugal	3	4	1	2	1	7	1	8	1		7
3	Germany	6	6	4	0	2	13	3	8	1		4.333333
4	Spain	2	6	5	0	1	7	2	3	0		3.5
5	Netherlan	4	6	6	0	0	12	5	15	0		2.4
6	Brazil	1	5	3	1	1	9	4	7	2		2.25
7	Japan	45	4	2	1	1	4	2	7	0		2
8	Uruguay	16	6	3	2	1	9	5	8	2		1.8
9	Argentina	7	5	4	0	1	10	6	7	0		1.666667
10	Paraguay	31	5	1	3	1	3	2	9	0		1.5
11	Ivory Coas	27	3	1	1	1	4	3	5	0		1.333333
12	Ghana	32	5	2	2	1	5	4	11	0		1.25
13	USA	14	4	1	2	1	5	5	9	0		1
14	Switzerlan	24	3	1	1	1	1	1	8	1		1
15	Slovenia	25	3	1	1	1	3	3	9	0		1
16	New Zeala	78	3	0	3	0	2	2	6	0		1
17	Italy	5	3	0	2	1	4	5	5	0		0.8
18	Mexico	17	4	1	1	2	4	5	9	0		0.8
19	South Kor	47	4	1	1	2	6	8	6	0		0.75
20	Slovakia	34	4	1	1	2	5	7	11	0		0.714286
21	Serbia	15	3	1	0	2	2	3	8	1		0.666667

The highest ratio is 7.

Problem 2: What is the average number of passes made by defenders? By forwards?

Ans: here is 2 questions have to find out the average number of passes by defender and another is average number passes by forwards.

Passes by defenders: in this case open the World cup-players csv file in the excel and specify a column for store the result for passes by defender. Here is j column named Defender. In this column we will put the result.

Step-1: put the curser in the below of defender box.
Step-2: write the formula =AVERAGEIF(c2:c596,c3,f2:f596)
Step-3: hit enter

	Get External Data			Connections			Sort & Filter		

SUM	▾ ⋮	✕ ✓	f_x	=AVERAGEIF(c2:c596,c3,f2:f596)					

	A	B	C	D	E	F	G	H	I	J
1	surname	team	position	minutes	shots	passes	tackles	saves		Defender
2	Abdoun	Algeria	midfielde	16	0	6	0	0		3,f2:f596)
3	Belhadj	Algeria	defender	270	1	146	8	0		
4	Boudebou	Algeria	midfielde	74	3	28	1	0		
5	Bougherra	Algeria	defender	270	1	89	11	0		
6	Chaouchi	Algeria	goalkeepe	90	0	17	0	2		

Passes by Forward: in this case open the World cup-players csv file in the excel and specify a column for store the result for passes by forward. Here is k column named forward. In this column we will put the result.

Step-1: put the curser in the below of defender box.
Step-2: write the formula =AVERAGEIF(c2:c596,c3,f2:f596)
Step-3: hit enter

Problem-3: What player on a team with "ia" in the team name played less than 200 minutes and made more than 100 passes?

Ans: in the team csv file what is the name with "ia" and have played less then 200 minutes and made passes more than 100.
Step-1: open the following file.
Step-3: select all file by pressing ctrl+a.
Step-4: click filter icon and filter team with contains "ia" hit enter.

step-5: again filter minutes less than 200 minutes

A	B	C	D	E	F	G	H	I	J
surnam	team	positio	minute	shots	passes	tackles	saves		
Abdoun	Algeria	midfielde	16	0	6	0	0		
Belhadj	Algeria	defender							
Boudebou	Algeria	midfielde							
Bougherra	Algeria	defender							
Chaouchi	Algeria	goalkeepe							
Djebbour	Algeria	forward							
Ghezzal	Algeria	forward							
Guedioura	Algeria	midfielde							
Halliche	Algeria	defender							
Kadir	Algeria	midfielde							
Lacen	Algeria	midfielde							
M'Bolhi	Algeria	goalkeepe							
Matmour	Algeria	midfielde	255	3	68	3	0		

Custom AutoFilter   ?   ✕

Show rows where:
    minutes

    is less than   ∨   200

        ● And   ○ Or

                          ∨   ∨

Use ? to represent any single character
Use * to represent any series of characters

                              OK   Cancel

step-6: again filter made passes more than 100.

	surnam	team	positio	minute	shots	passes	tackles	saves		
1										
2	Abdoun	Algeria	Sort Smallest to Largest				0	0		
4	Boudebou	Algeria	Sort Largest to Smallest				1	0		
6	Chaouchi	Algeria	Sort by Color			▸	0	2		
7	Djebbour	Algeria					1	0		
8	Ghezzal	Algeria	Clear Filter From "passes"				0	0		
9	Guedioura	Algeria	Filter by Color			▸	1	0		
13	M'Bolhi	Algeria	Number Filters			▸				
15	Mesbah	Algeria								
16	Saifi	Algeria	Search			🔎				
40	Beaucham	Australi	☑ (Select All)			⌃				
41	Bresciano	Australi	☑ 0							
42	Cahill	Australi	☑ 1							
43	Carney	Australi	☑ 3							
44	Chipperfi	Australi	☑ 4							
47	Garcia	Australi	☑ 5							
48	Grella	Australi	☑ 6							
49	Holman	Australi	☑ 7							
50	Jedinak	Australi	☑ 8							
51	Kennedy	Australi	☑ 9			⌄				
52	Kewell	Australi	OK Cancel							

Number Filters submenu:
- Equals...
- Does Not Equal...
- Greater Than...
- Greater Than Or Equal To...
- Less Than...
- Less Than Or Equal To...
- Between...
- Top 10...
- Above Average
- Below Average
- Custom Filter...

Finally the result is:

Kuzmanovic	Serbia	midfielder	180	4	103	1	0

A433   ⋮   ✕ ✓ fx   Kuzmanovic

	A	B	C	D	E	F	G	H
1	surnam	team	positio	minute	shots	passes	tackles	saves
433	Kuzmanov	Serbia	midfielde	180	4	103	1	0
597								
598								

Problem-4: Which team has the highest average number of passes per minute played (and what is that average)?

Ans: indicate a cell for put the result here is j2.
Step-1: keep curser in j2 and write the formula =AVERAGEIF(d:d,d2,f:f)

	A	B	C	D	E	F	G	H	I	J
1	surname	team	position	minutes	shots	passes	tackles	saves		Problem_4
2	Abdoun	Algeria	midfielde	16	0	6	0	0		(d:d,d2,f:f)
3	Belhadj	Algeria	defender	270	1	146	8	0		
4	Boudebou	Algeria	midfielde	74	3	28	1	0		
5	Bougherra	Algeria	defender	270	1	89	11	0		
6	Chaouchi	Algeria	goalkeepe	90	0	17	0	2		
7	Djebbour	Algeria	forward	123	3	19	1	0		
8	Ghezzal	Algeria	forward	40	3	8	0	0		
9	Guedioura	Algeria	midfielde	38	0	18	1	0		
10	Halliche	Algeria	defender	270	2	94	4	0		
11	Kadir	Algeria	midfielde	262	0	104	3	0		
12	Lacen	Algeria	midfielde	270	0	158	8	0		

Step-2: hit enter and get result 10.5 for first cell.
Step-3: scroll down on + sign of that result and get all of them.
Step-4: for the highest average select the j column and sort them into Z-A. in the first row is the highest average. That is 169.

	A	B	C	D	E	F	G	H	I	J
1	Torres	USA	midfielde	45	1	32	0	0		169
2	Villa	Spain	forward	529	22	169	2	0		8.5
3	Arbeloa	Spain	defender	13	0	12	0	0		84
4	Yeom Ki-H	South Kor	midfielde	249	2	80	6	0		320
5	Podolski	Germany	forward	531	17	217	9	0		80
6	Pedro Roc	Spain	midfielde	116	5	80	0	0		250
7	Neuer	Germany	goalkeepe	540	0	99	0	20		82
8	Fabregas	Spain	midfielde	94	1	116	2	0		254
9	Puyol	Spain	defender	534	3	254	8	0		44
10	Jansen	Germany	defender	73	1	44	3	0		115.6667
11	Robben	Netherlan	forward	267	6	81	5	0		29.5
12	de Zeeuw	Netherlan	midfielde	47	0	37	2	0		48.5

Problem 5. How many players on a team ranked under 10 (i.e. ranks 1 to 9) played more than 350 minutes?

Ans: based on this questions there are 54 players.

Problem 6. What characteristics are shared by all passengers whose fare is 0?
Ans: Select whole cell and filter it. Filter on fare where unselect all without 0.

	last	first	gender	age	class	fare	embarked	survive
181	Leonard	Mr. Lionel	M	36	3	0	Southampton	no
265	Harrison	Mr. William	M	40	1	0	Southampton	no
273	Tornquist	Mr. William Henry	M	25	3	0	Southampton	yes
279	Parkes	Mr. Francis "Frank"	M		2	0	Southampton	no
304	Johnson	Mr. William Cahoone Jr	M	19	3	0	Southampton	no
415	Cunningh	Mr. Alfred Fleming	M		2	0	Southampton	no
468	Campbell	Mr. William	M		2	0	Southampton	no
483	Frost	Mr. Anthony Wood "Archie"	M		2	0	Southampton	no
599	Johnson	Mr. Alfred	M	49	3	0	Southampton	no
635	Parr	Mr. William Henry Marsh	M		1	0	Southampton	no
676	Watson	Mr. Ennis Hastings	M		2	0	Southampton	no
734	Knight	Mr. Robert J	M		2	0	Southampton	no
808	Andrews	Mr. Thomas Jr	M	39	1	0	Southampton	no
817	Fry	Mr. Richard	M		1	0	Southampton	no
824	Reuchlin	Jonkheer. John George	M	38	1	0	Southampton	no

Problem 7. How many married women over age 50 embarked in Cherbourg? (Married women are denoted by "Mrs.")

Ans: open the sheet and select a cell for store the result and write the formula =COUNTIFS(B:B,"Mrs*",D:D,">"&50,G:G,G3) hit enter the answer will be 4.

	AVERAGE		:	×	✓	fx	=COUNTIFS(b:b,"Mrs*",d:d,">"&50,g:g,g3)			

	A	B	C	D	E	F	G	H	I	J
1	last	first	gender	age	class	fare	embarked	survived		problem-7
2	Braund	Mr. Owen Harris	M	22	3	7.25	Southampton	no		">"&50,g:g,g3)
3	Ward	Miss Anna	F	35	1	512.3292	Cherbourg	yes		
4	Heikkiner	Miss Laina	F	26	3	7.925	Southampton	yes		
5	Futrelle	Mrs. Jacques Heath (Lily May Peel)	F	35	1	53.1	Southampton	yes		
6	Allen	Mr. William Henry	M	35	3	8.05	Southampton	no		
7	Moran	Mr. James	M		3	8.4583	Queenstown	no		
8	McCarthy	Mr. Timothy J	M	54	1	51.8625	Southampton	no		
9	Palsson	Master Gosta Leonard	M	2	3	21.075	Southampton	no		
10	Johnson	Mrs. Oscar W (Elisabeth Vilhelmina Berg)	F	27	3	11.1333	Southampton	yes		
11	Cardeza	Mr. Thomas Drake Martinez	M	36	1	512.3292	Cherbourg	yes		
12	Sandstron	Miss Marguerite Rut	F	4	3	16.7	Southampton	yes		
13	Bonnell	Miss Elizabeth	F	58	1	26.55	Southampton	yes		
14	Saunderco	Mr. William Henry	M	20	3	8.05	Southampton	no		
15	Andersso	Mr. Anders Johan	M	39	3	31.275	Southampton	no		
16	Vestrom	Miss Hulda Amanda Adolfina	F	14	3	7.8542	Southampton	no		
17	Hewlett	Mrs. (Mary D Kingcome)	F	55	2	16	Southampton	yes		
18	Rice	Master Eugene	M	2	3	29.125	Queenstown	no		
19	Williams	Mr. Charles Eugene	M		2	13	Southampton	yes		
20	Vander Pl	Mrs. Julius (Emelia Maria Vandemoortele)	F	31	3	18	Southampton	no		
21	Lesurer	Mr. Gustave J	M	35	1	512.3292	Cherbourg	yes		

	J3		:	×	✓	fx				

	A	B	C	D	E	F	G	H	I	J
1	last	first	gender	age	class	fare	embarked	survived		problem-7
2	Braund	Mr. Owen Harris	M	22	3	7.25	Southampton	no		4
3	Ward	Miss Anna	F	35	1	512.3292	Cherbourg	yes		
4	Heikkiner	Miss Laina	F	26	3	7.925	Southampton	yes		
5	Futrelle	Mrs. Jacques Heath (Lily May Peel)	F	35	1	53.1	Southampton	yes		
6	Allen	Mr. William Henry	M	35	3	8.05	Southampton	no		
7	Moran	Mr. James	M		3	8.4583	Queenstown	no		
8	McCarthy	Mr. Timothy J	M	54	1	51.8625	Southampton	no		
9	Palsson	Master Gosta Leonard	M	2	3	21.075	Southampton	no		
10	Johnson	Mrs. Oscar W (Elisabeth Vilhelmina Berg)	F	27	3	11.1333	Southampton	yes		
11	Cardeza	Mr. Thomas Drake Martinez	M	36	1	512.3292	Cherbourg	yes		
12	Sandstron	Miss Marguerite Rut	F	4	3	16.7	Southampton	yes		
13	Bonnell	Miss Elizabeth	F	58	1	26.55	Southampton	yes		
14	Saunderco	Mr. William Henry	M	20	3	8.05	Southampton	no		
15	Andersso	Mr. Anders Johan	M	39	3	31.275	Southampton	no		

Problem 8. Which embarkation city had the highest-paying passengers on average?
Ans: Cherbourg city had the highest-paying passengers on average.
Step-1: select whole cell and sort it to Z-A the first one is the result.

	A	B	C	D	E	F	G	H
1	last	first	gender	age	class	fare	embarked	survive
2	Cumings	Mrs. John	F	38	1	71.2833	Cherbourg	yes

Problem 9. What is the most common last name among passengers? What is the average number of passengers per last name?

Ans: the first answer write the formula =INDEX(A1:A892,MODE(MATCH(A1:A892,A1:A892,0))) in any cell where will have put the result. The result Andersson is the most common last name among passengers

J2			×	✓	fx	=INDEX(A1:A892,MODE(MATCH(A1:A892,A1:A892,0)))			

	A	B	C	D	E	F	G	H	I	J
1	last	first	gender	age	class	fare	embarked	survived		prblm-9-1
2	Braund	Mr. Owen	M	22	3	7.25	Southampton	no		Andersson
3	Cumings	Mrs. John	F	38	1	71.2833	Cherbourg	yes		
4	Heikkinen	Miss Laina	F	26	3	7.925	Southampton	yes		
5	Futrelle	Mrs. Jacqu	F	35	1	53.1	Southampton	yes		
6	Allen	Mr. Willia	M	35	3	8.05	Southampton	no		
7	Moran	Mr. James	M		3	8.4583	Queenstown	no		

2nd Ans: write the formula =AVERAGE(COUNTIF(A1:A892,A2)) hit enter for one cell's result by scrolling down on that result it will get all of them .

AVERAGE			×	✓	fx	=AVERAGE(COUNTIF(A1:A892,A2))					

	A	B	C	D	E	F	G	H	I	J	K	L
1	last	first	gender	age	class	fare	embarked	survived		prblm-9-1	problem 9-2	
2	Braund	Mr. Owen	M	22	3	7.25	Southampton	no		Andersson	A892,A2))	
3	Cumings	Mrs. John	F	38	1	71.2833	Cherbourg	yes			1	
4	Heikkinen	Miss Laina	F	26	3	7.925	Southampton	yes			1	
5	Futrelle	Mrs. Jacqu	F	35	1	53.1	Southampton	yes			2	
6	Allen	Mr. Willia	M	35	3	8.05	Southampton	no			2	
7	Moran	Mr. James	M		3	8.4583	Queenstown	no			3	
8	McCarthy	Mr. Timot	M	54	1	51.8625	Southampton	no			1	
9	Palsson	Master Gc	M	2	3	21.075	Southampton	no			4	
10	Johnson	Mrs. Oscar	F	27	3	11.1333	Southampton	yes			6	
11	Nasser	Mrs. Nich	F	14	2	30.0708	Cherbourg	yes			2	
12	Sandstron	Miss Marg	F	4	3	16.7	Southampton	yes			2	
13	Bonnell	Miss Elizal	F	58	1	26.55	Southampton	yes			1	
14	Saunderco	Mr. Willia	M	20	3	8.05	Southampton	no			1	
15	Andersson	Mr. Ander	M	39	3	31.275	Southampton	no			9	
16	Vestrom	Miss Huld	F	14	3	7.8542	Southampton	no			1	
17	Hewlett	Mrs. (Mar	F	55	2	16	Southampton	yes			1	
18	Rice	Master Eu	M	2	3	29.125	Queenstown	no			5	
19	Williams	Mr. Charle	M		2	13	Southampton	yes			4	
20	Vander Pl	Mrs. Julius	F	31	3	18	Southampton	no			3	
21	Masselma	Mrs. Fatin	F		3	7.225	Cherbourg	yes			1	
22	Fynney	Mr. Josepl	M	35	2	26	Southampton	no			1	
23	Beesley	Mr. Lawre	M	34	2	13	Southampton	yes			1	

The marked part is the following result.

Problem 10. What's the survival rate for passengers in the three different classes, i.e., what fraction of passengers in each class survived? Find the answer using spreadsheet functions only - don't perform any arithmetic by hand!
Ans: step-1: first filter class and servived to 1 and yes and copy it to another sheet.
Step-2: use a cell for count the cell that is 136.

	last	first	gender	age	class	fare	embarked	survive
1	last	first	gender	age	class	fare	embarked	survive
3	Cumings	Mrs. John Bradley (Florenc	F	38	1	71.2833	Cherbourg	yes
5	Futrelle	Mrs. Jacques Heath (Lily M:	F	35	1	53.1	Southampton	yes
13	Bonnell	Miss Elizabeth	F	58	1	26.55	Southampton	yes
25	Sloper	Mr. William Thompson	M	28	1	35.5	Southampton	yes
33	Spencer	Mrs. William Augustus (Ma	F		1	146.5208	Cherbourg	yes
54	Harper	Mrs. Henry Sleeper (Myna	F	49	1	76.7292	Cherbourg	yes
57	Woolner	Mr. Hugh	M		1	35.5	Southampton	yes
63	Icard	Miss Amelie	F	38	1	80	Southampton	yes
90	Fortune	Miss Mabel Helen	F	23	1	263	Southampton	yes
99	Greenfiel	Mr. William Bertram	M	23	1	63.3583	Cherbourg	yes
138	Newsom	Miss Helen Monypeny	F	19	1	26.2833	Southampton	yes
153	Pears	Mrs. Thomas (Edith Wearn	F	22	1	66.6	Southampton	yes
168	Chibnall	Mrs. (Edith Martha Bowern	F		1	55	Southampton	yes
189	Romaine	Mr. Charles Hallace ("Mr C	M	45	1	26.55	Southampton	yes
196	Brown	Mrs. James Joseph (Margar	F	44	1	27.7208	Cherbourg	yes
197	Lurette	Miss Elise	F	58	1	146.5208	Cherbourg	yes
211	Blank	Mr. Henry	M	40	1	31	Cherbourg	yes
217	Newell	Miss Madeleine	F	31	1	113.275	Cherbourg	yes
220	Bazzani	Miss Albina	F	32	1	76.2917	Cherbourg	yes
226	Hoyt	Mr. Frederick Maxfield	M	38	1	90	Southampton	yes
232	Harris	Mrs. Henry Birkhardt (Irene	F	35	1	83.475	Southampton	yes
250	Beckwith	Mr. Richard Leonard	M	37	1	52.5542	Southampton	yes

Make the same things as class 2, 3.
Take all the result in one sheet and use the formula =(J2/SUM(J2:J4)*100)
For final result

	I	J	K	L
			fraction	
		136	39.76608	
		87	25.4386	
		119	34.79532	