法希姆-62019010100

DATASETS

Titanic Datasets: Titanic.csv Description: Data on passengers of the RMS Titanic. Entries include the name, age, class, fare, gender, and whether or not the passenger survived.

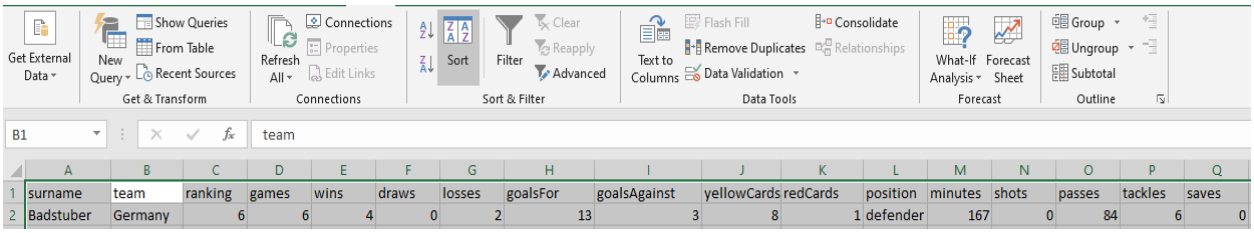
World Cup Datasets: Players.csv, Teams.csv Description: 2010 World Cup data including last name, team, position, minutes played, and game statistics for each player (Players.csv) as well as world ranking, games played in tournaments, and game statistics for each team (Teams.csv).

- **World Cup Data Problem**

Problem 1. Which team has the highest ratio of goalsFor/ goalsAgainst?

I use the sorting method to find highest ratio of “goalsFor”.

Answer: The highest ratio for “goalsFor” is 13.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	surname	team	ranking	games	wins	draws	losses	goalsFor	goalsAgainst	yellowCards	redCards	position	minutes	shots	passes	tackles	saves
2	Badstuber	Germany	6	6	4	0	2	13	3	8	1	defender	167	0	84	6	0

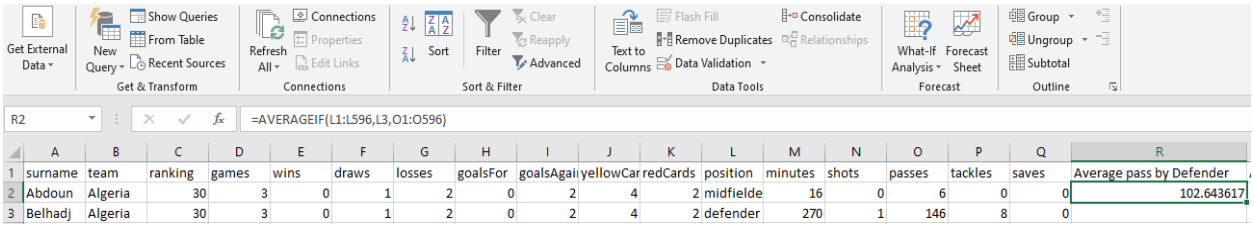
Figure-1: Highest ration “goalsFor”

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

Used formula **=AVERAGEIF(L1:L596,L3,O1:O596)** to solve made by defenders

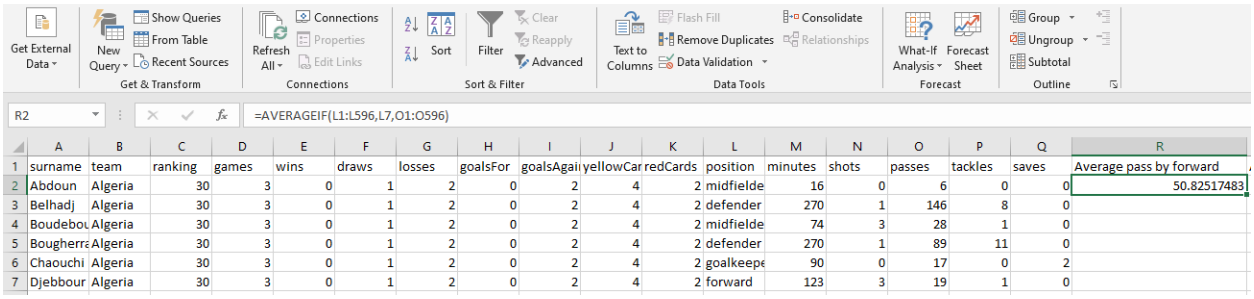
Used formula **=AVERAGEIF(L1:L596,L7,O1:O596)** to solve made by forwards.

Answer2: Average Pass made by defender is **102.643617** and Average Pass made by forward is **50.82517483**.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	surname	team	ranking	games	wins	draws	losses	goalsFor	goalsAgai	yellowCar	redCards	position	minutes	shots	passes	tackles	saves	Average pass by Defender
2	Abdoun	Algeria	30	3	0	1	2	0	2	4	2	midfielde	16	0	6	0	0	102.643617
3	Belhadj	Algeria	30	3	0	1	2	0	2	4	2	defender	270	1	146	8	0	

Figure-2.1: Made by Defender



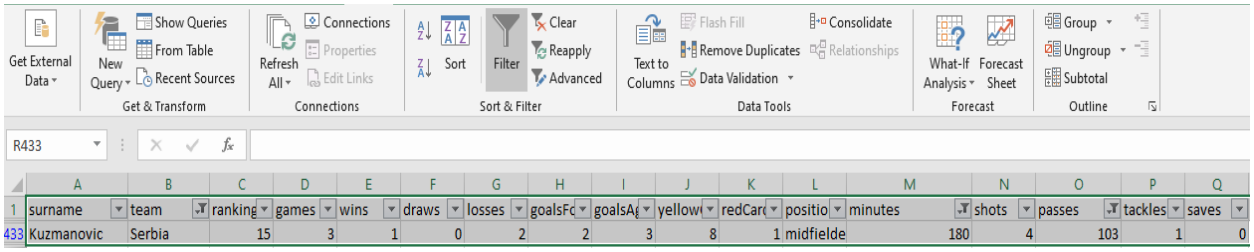
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	surname	team	ranking	games	wins	draws	losses	goalsFor	goalsAgai	yellowCar	redCards	position	minutes	shots	passes	tackles	saves	Average pass by forward
2	Abdoun	Algeria	30	3	0	1	2	0	2	4	2	midfielde	16	0	6	0	0	50.82517483
3	Belhadj	Algeria	30	3	0	1	2	0	2	4	2	defender	270	1	146	8	0	
4	Boudebou	Algeria	30	3	0	1	2	0	2	4	2	midfielde	74	3	28	1	0	
5	Bougherra	Algeria	30	3	0	1	2	0	2	4	2	defender	270	1	89	11	0	
6	Chaouchi	Algeria	30	3	0	1	2	0	2	4	2	goalkeepe	90	0	17	0	2	
7	Djebbour	Algeria	30	3	0	1	2	0	2	4	2	forward	123	3	19	1	0	

Figure-2.2: Made by Forwards

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

I used Filter to solve this according to team name, played time and passes quantity.

Answer: The player name is “Kuzmanovic”. Which is shown below:



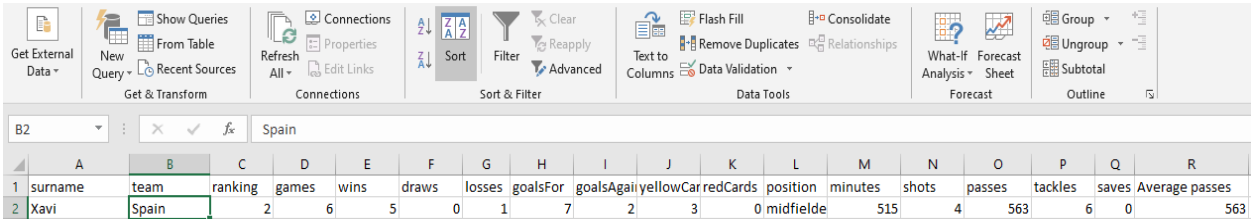
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	surname	team	ranking	games	wins	draws	losses	goalsFo	goalsAg	yellow	redCar	positio	minutes	shots	passes	tackles	saves
433	Kuzmanovic	Serbia	15	3	1	0	2	2	3	8	1	midfielde	180	4	103	1	0

Figure-3.1: Player Name

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

First, I found the average pass of each team and then I used “sorting” method to get highest average passes team name.

Answer4: The team “Spain’ has passed highest average. Shown in Figure-4:



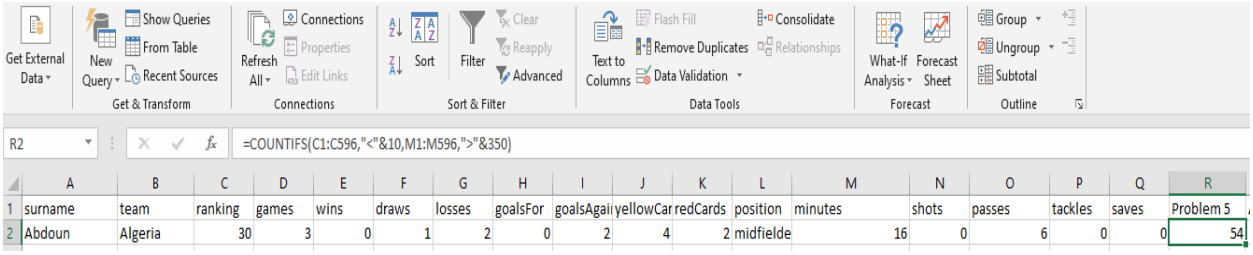
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	surname	team	ranking	games	wins	draws	losses	goalsFor	goalsAgai	yellowCar	redCards	position	minutes	shots	passes	tackles	saves	Average passes
2	Xavi	Spain	2	6	5	0	1	7	2	3	0	midfielde	515	4	563	6	0	563

Figure-4: Team Name for highest pass

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

Used formula **=COUNTIFS(C1:C596,"<"&10,M1:M596,">"&350)** to solve this problem.

Answer5: 54 players. Shown in figure-5 below:



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	surname	team	ranking	games	wins	draws	losses	goalsFor	goalsAgai	yellowCar	redCards	position	minutes	shots	passes	tackles	saves	Problem 5
2	Abdoun	Algeria	30	3	0	1	2	0	2	4	2	midfielde	16	0	6	0	0	54

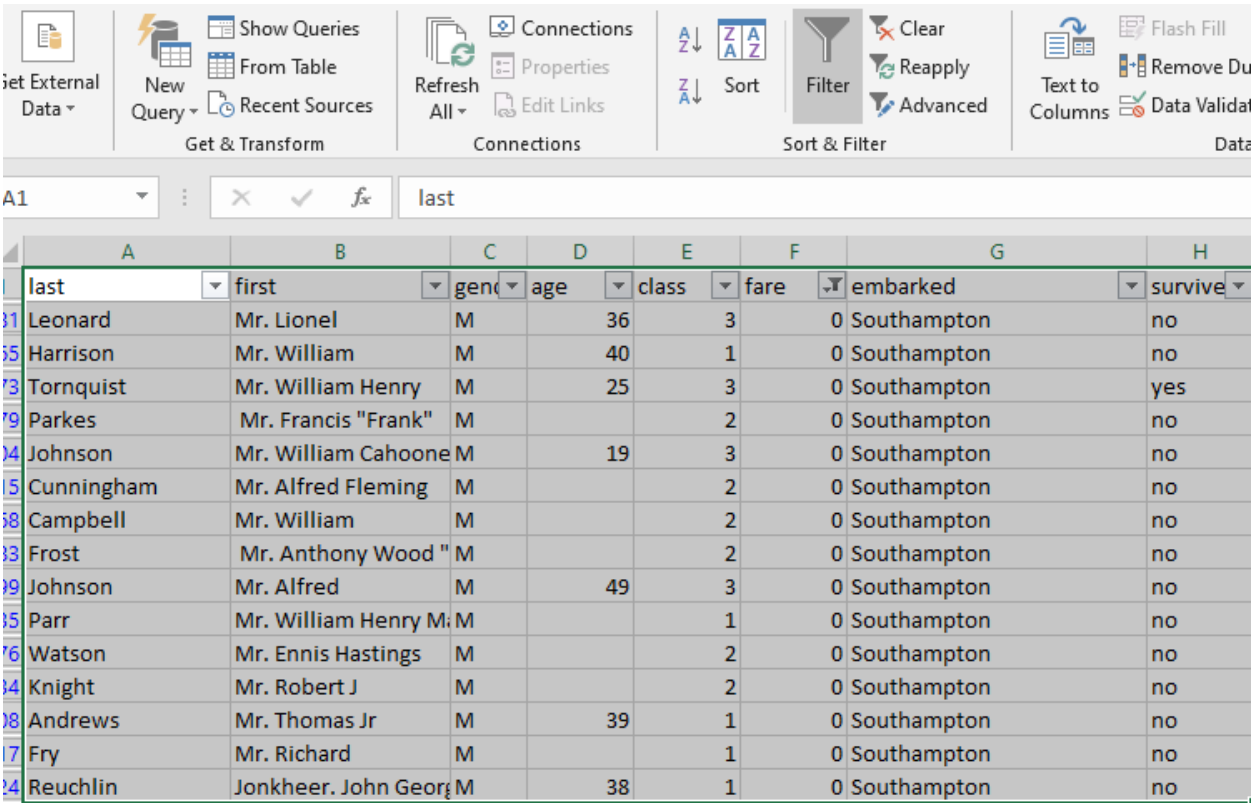
Figure-5: Players Number

- **Titanic Data**

Problem 6. What characteristics are shared by all passengers whose fare is 0?

Answer6: I used “Filter” to solve this problem where all passengers “fare” is 0 and

all characteristics about those passengers has been appeared. Shown in figure-6:



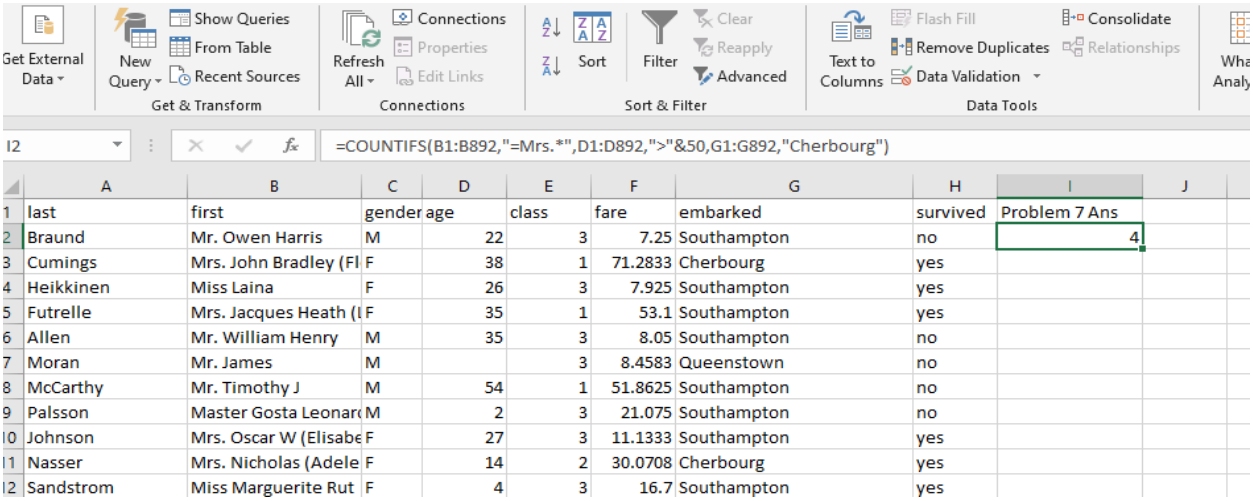
last	first	gen	age	class	fare	embarked	survive
Leonard	Mr. Lionel	M	36	3	0	Southampton	no
Harrison	Mr. William	M	40	1	0	Southampton	no
Tornquist	Mr. William Henry	M	25	3	0	Southampton	yes
Parkes	Mr. Francis "Frank"	M		2	0	Southampton	no
Johnson	Mr. William Cahoone	M	19	3	0	Southampton	no
Cunningham	Mr. Alfred Fleming	M		2	0	Southampton	no
Campbell	Mr. William	M		2	0	Southampton	no
Frost	Mr. Anthony Wood "	M		2	0	Southampton	no
Johnson	Mr. Alfred	M	49	3	0	Southampton	no
Parr	Mr. William Henry M	M		1	0	Southampton	no
Watson	Mr. Ennis Hastings	M		2	0	Southampton	no
Knight	Mr. Robert J	M		2	0	Southampton	no
Andrews	Mr. Thomas Jr	M	39	1	0	Southampton	no
Fry	Mr. Richard	M		1	0	Southampton	no
Reuchlin	Jonkheer. John Georg	M	38	1	0	Southampton	no

Figure-6: Fare 0 Passenger's Characteristics

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

I used this formula **=COUNTIFS(B1:B892,"=Mrs.*",D1:D892,">"&50,G1:G892,"Cherbourg")** to solve this problem.

Answer7: 4 women. Shown figure-7 below:

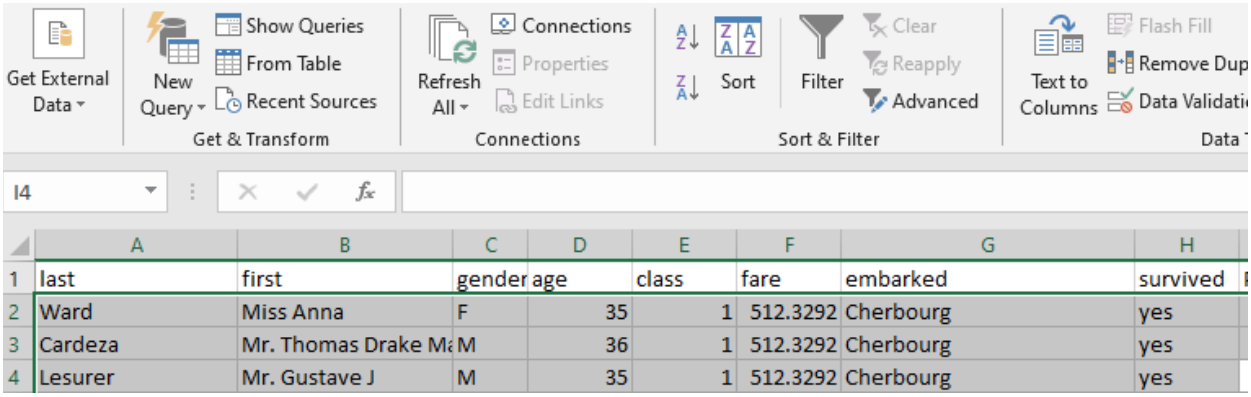


	A	B	C	D	E	F	G	H	I	J
1	last	first	gender	age	class	fare	embarked	survived	Problem 7 Ans	
2	Braund	Mr. Owen Harris	M	22	3	7.25	Southampton	no	4	
3	Cumings	Mrs. John Bradley (Fl	F	38	1	71.2833	Cherbourg	yes		
4	Heikkinen	Miss Laina	F	26	3	7.925	Southampton	yes		
5	Futrelle	Mrs. Jacques Heath (I	F	35	1	53.1	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 Leonarc	M	2	3	21.075	Southampton	no		
10	Johnson	Mrs. Oscar W (Elisabe	F	27	3	11.1333	Southampton	yes		
11	Nasser	Mrs. Nicholas (Adele	F	14	2	30.0708	Cherbourg	yes		
12	Sandstrom	Miss Marguerite Rut	F	4	3	16.7	Southampton	yes		

Figure-7: Women Satisfy these criteria

Problem 8. Which embarkation city had the highest-paying passengers on average?

Answer8: The embarkation city name is “Cherbourg”. Shown in figure-8 below:



	A	B	C	D	E	F	G	H
1	last	first	gender	age	class	fare	embarked	survived
2	Ward	Miss Anna	F	35	1	512.3292	Cherbourg	yes
3	Cardeza	Mr. Thomas Drake Ma	M	36	1	512.3292	Cherbourg	yes
4	Lesurer	Mr. Gustave J	M	35	1	512.3292	Cherbourg	yes

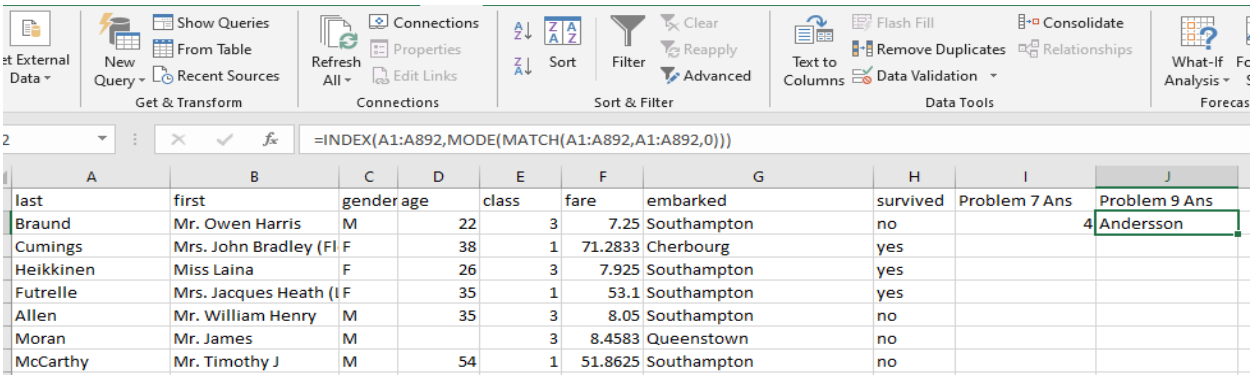
Figure-8: Embarkation city name

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

Used formula **=INDEX(A1:A892,MODE(MATCH(A1:A892,A1:A892,0)))** to find common name. Figure-9.1.

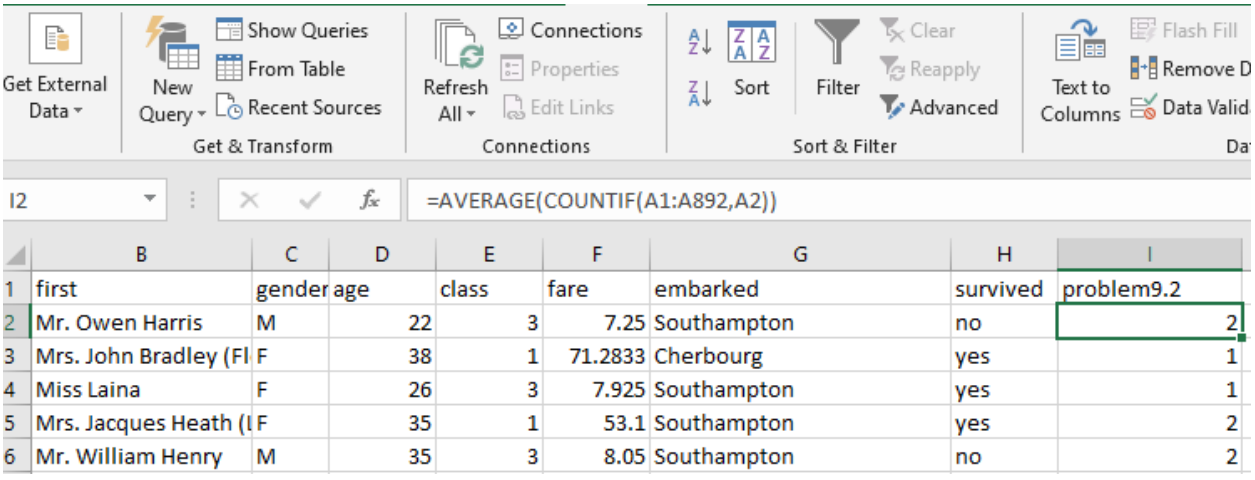
Used formula **=AVERAGE(COUNTIF(A1:A892,A2))** to solve average number of passengers per last name. Figure-9.2.

Answer9: Most common name is “Andersson” and average number of passengers per last name is shown figure9.2below:



	A	B	C	D	E	F	G	H	I	J
	last	first	gender	age	class	fare	embarked	survived	Problem 7 Ans	Problem 9 Ans
	Braund	Mr. Owen Harris	M	22	3	7.25	Southampton	no	4	Andersson
	Cumings	Mrs. John Bradley (Fl	F	38	1	71.2833	Cherbourg	yes		
	Heikkinen	Miss Laina	F	26	3	7.925	Southampton	yes		
	Futrelle	Mrs. Jacques Heath (I	F	35	1	53.1	Southampton	yes		
	Allen	Mr. William Henry	M	35	3	8.05	Southampton	no		
	Moran	Mr. James	M		3	8.4583	Queenstown	no		
	McCarthy	Mr. Timothy J	M	54	1	51.8625	Southampton	no		

Figure-9.1: Most common name



	B	C	D	E	F	G	H	I
1	first	gender	age	class	fare	embarked	survived	problem9.2
2	Mr. Owen Harris	M	22	3	7.25	Southampton	no	2
3	Mrs. John Bradley (Fl	F	38	1	71.2833	Cherbourg	yes	1
4	Miss Laina	F	26	3	7.925	Southampton	yes	1
5	Mrs. Jacques Heath (I	F	35	1	53.1	Southampton	yes	2
6	Mr. William Henry	M	35	3	8.05	Southampton	no	2

Figure-9.2: Passengers per last name

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 function s only - don't perform any arithmetic by hand!

Used formula **=COUNTIFS(E1:E892,"="&3,H1:H892,"=yes")** to get survival rate for passengers in three different classes. Figure-10.1 below.

Answer10: The survival rate for passengers in the three different classes is 119.

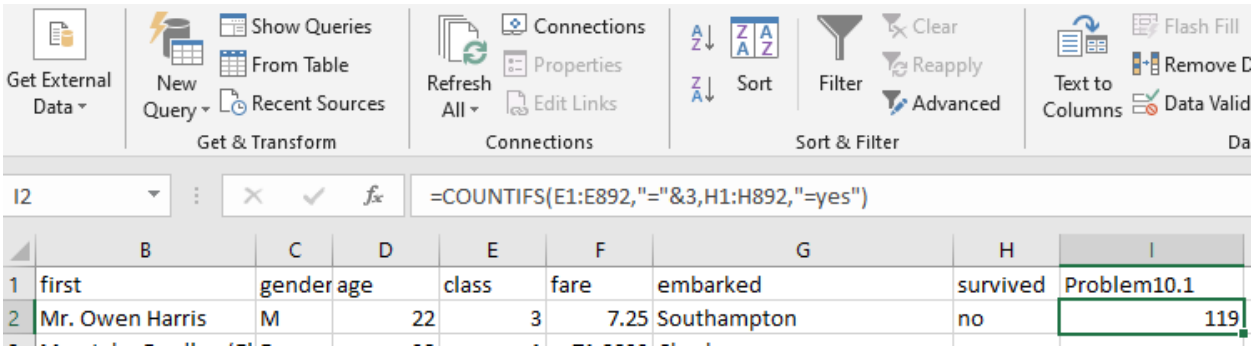


Figure-10.1: Survival rate

I used this **=COUNTIFS(E1:E892,E2,H1:H892,"yes")/3** formula to get the fraction of passengers in each class survived. Figure-10.2 below.

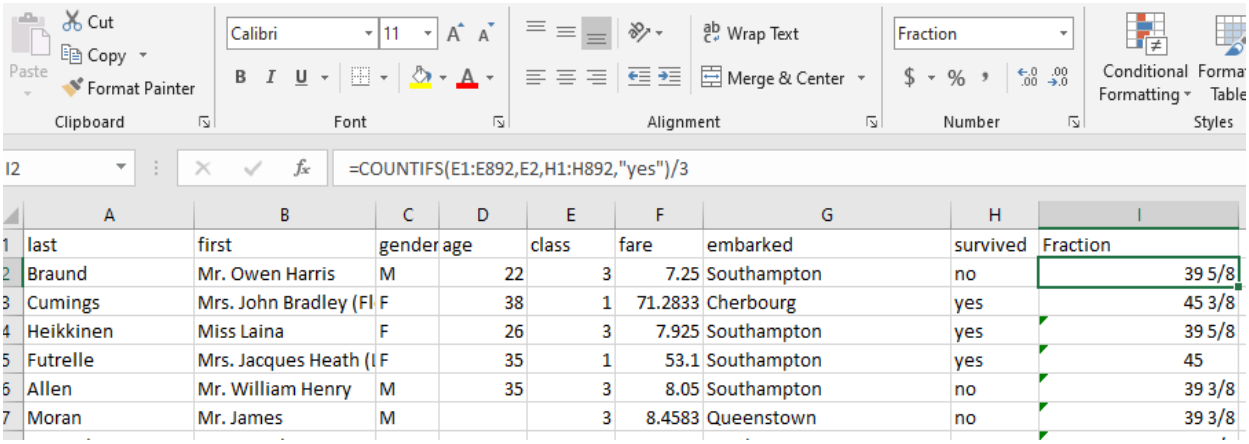
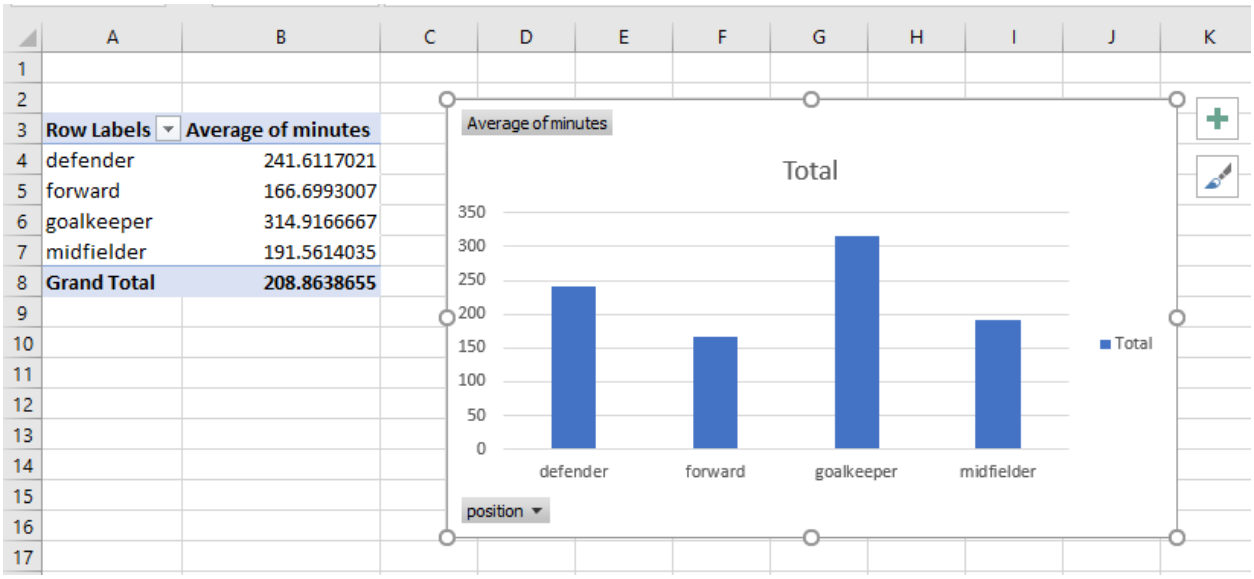


Figure-10.2: Fraction

- # World Cup Data

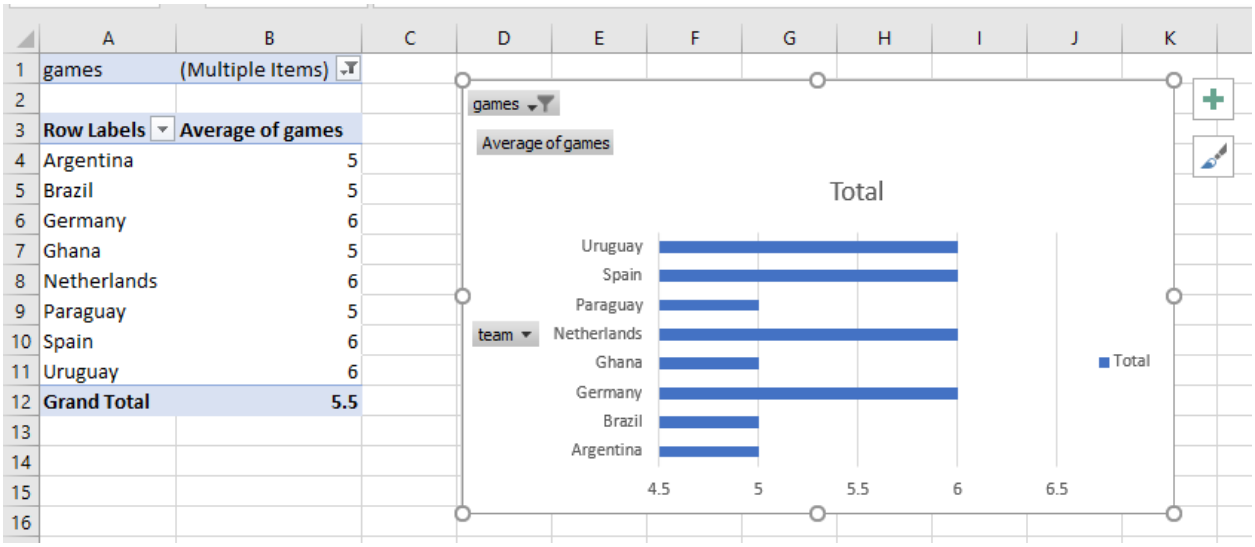
Problem 11. Create a bar chart showing the average number of minutes played by players in each position.

Answer-11: “pivot table” for Data analyzing and “Bar chart” for Data Visualization.



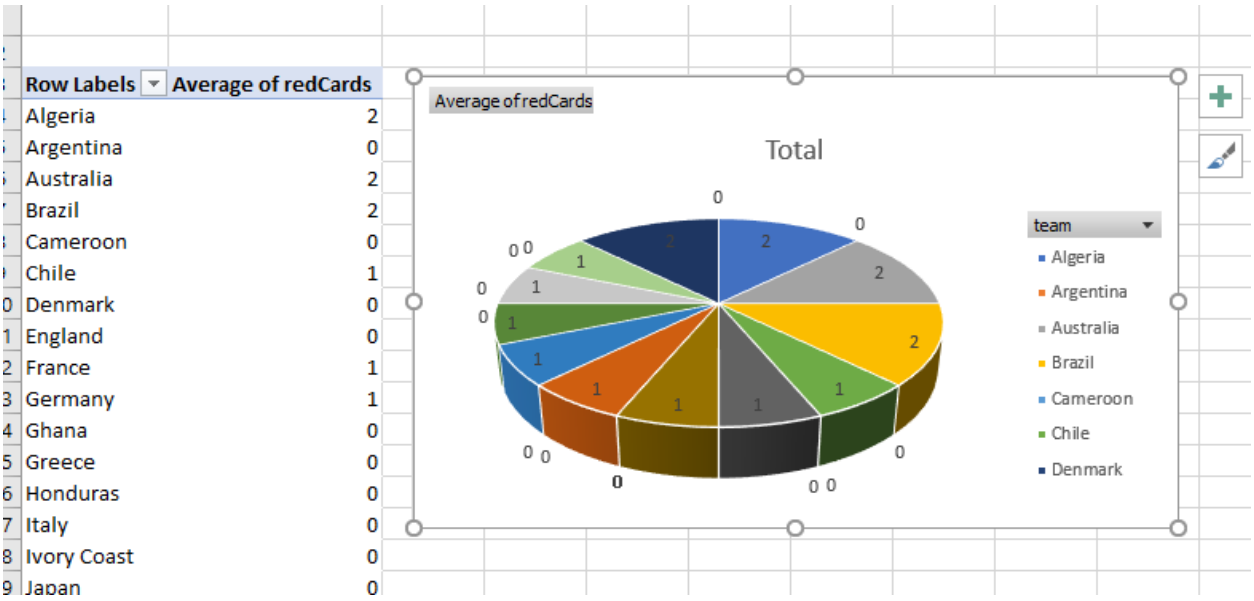
Problem 12. Create a stacked bar chart showing teams that played more than 4 games, with their total number of games divided into wins, draws, and losses.

Answer-12: “pivot table” for Data analyzing and “Stacked Bar chart” for Data Visualization.



Problem 13. Create a pie chart showing the relative percentage of teams with 0, 1, and 2 red cards.

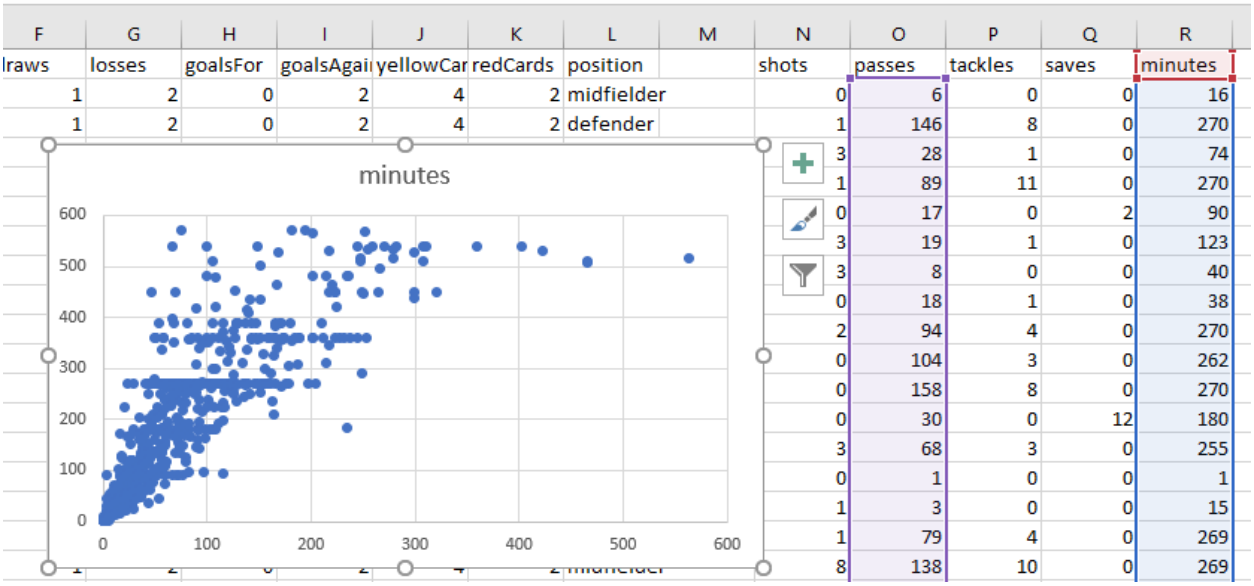
Answer-13: "pivot table" for Data analyzing and "Pie chart" for Data Visualization.



Row Labels	Average of redCards
Algeria	2
Argentina	0
Australia	2
Brazil	2
Cameroon	0
Chile	1
Denmark	0
England	0
France	1
Germany	1
Ghana	0
Greece	0
Honduras	0
Italy	0
Ivory Coast	0
Japan	0

Problem 14. Create a scatterplot of players showing passes (x-axis) versus minutes (y-axis). (If you know anything about the World Cup you might think about why there are lines of dots.)

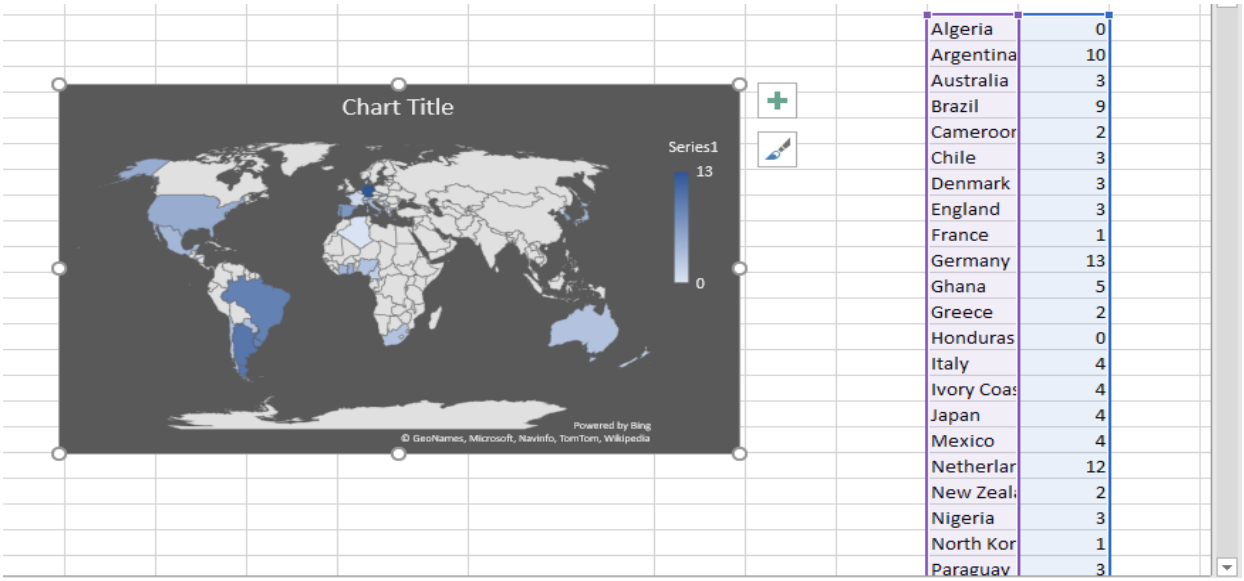
Answer-14: Used "Scatterplot" for Data Visualization.



F (draws)	G (losses)	H (goalsFor)	I (goalsAgai)	J (yellowCar)	K (redCards)	L (position)	M	N (shots)	O (passes)	P (tackles)	Q (saves)	R (minutes)
1	2	0	2	4	2	midfielder		0	6	0	0	16
1	2	0	2	4	2	defender		1	146	8	0	270
								3	28	1	0	74
								1	89	11	0	270
								0	17	0	2	90
								3	19	1	0	123
								3	8	0	0	40
								0	18	1	0	38
								2	94	4	0	270
								0	104	3	0	262
								0	158	8	0	270
								0	30	0	12	180
								3	68	3	0	255
								0	1	0	0	1
								1	3	0	0	15
								1	79	4	0	269
								8	138	10	0	269

Problem 15. Create a map of countries colored light to dark blue based on how many goals their team made ("goalsFor").

Answer-15: "pivot table" for Data analyzing and "Map" for Data Visualization.

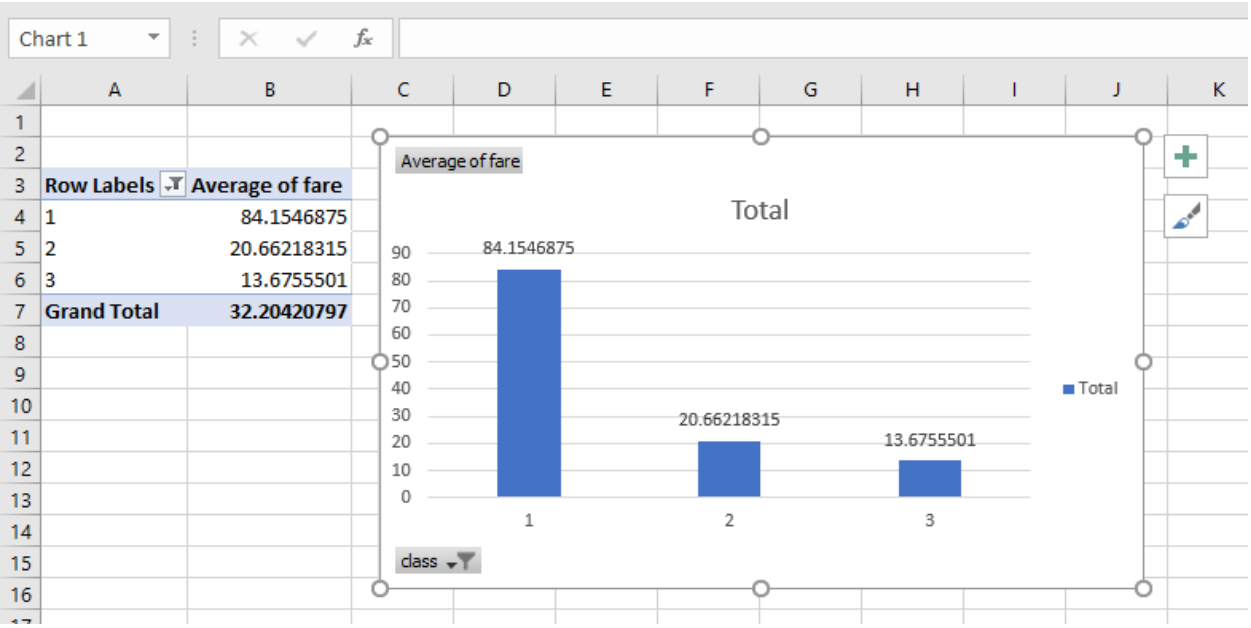


Algeria	0
Argentina	10
Australia	3
Brazil	9
Cameroor	2
Chile	3
Denmark	3
England	3
France	1
Germany	13
Ghana	5
Greece	2
Honduras	0
Italy	4
Ivory Coas	4
Japan	4
Mexico	4
Netherlar	12
New Zeal	2
Nigeria	3
North Kor	1
Paraguay	3

- Titanic Data

Problem 16. Create a bar chart showing the average fare paid by passengers in each class. The three bars should be labeled "first", "second", "third".

Answer-16: "pivot table" for Data analyzing and "Bar chart" for Data Visualization. Data has been labeled as well.



Problem 17. Create a stacked bar chart showing the number of passengers in each class, divided into male and female (three bars). Then reverse roles and show the number of passengers of each gender, divided into class (two bars).

Answer-17: "pivot table" for Data analyzing and "Stacked Bar chart" for Data Visualization. Number of passengers in each class has shown is figure-17.1 and number of passengers of each gender has shown in figure-17.2.

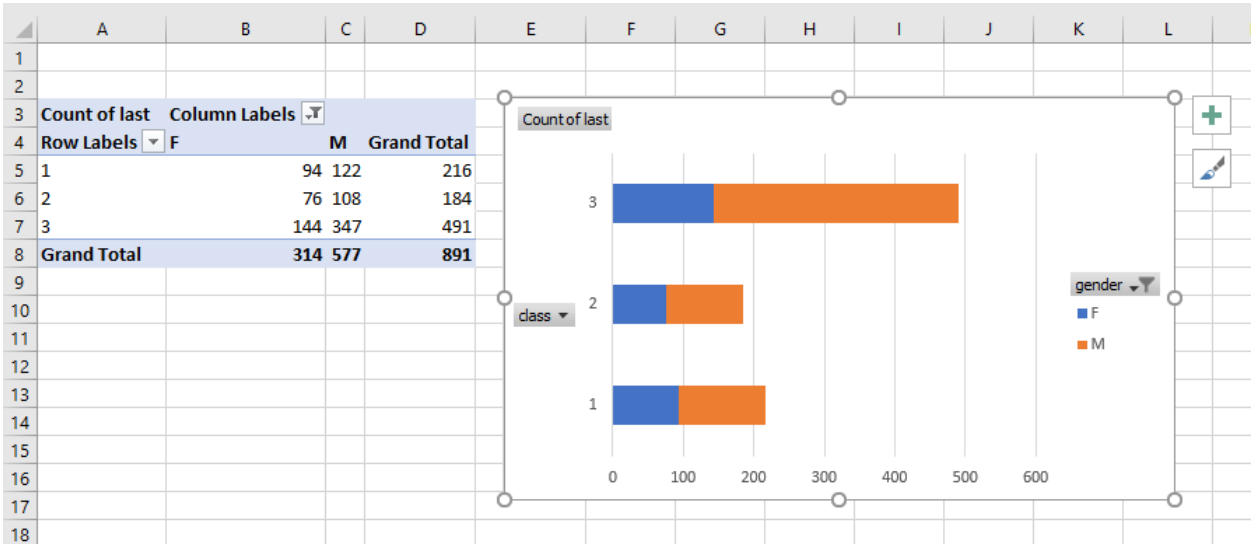


Figure-17.1: Passengers in each class

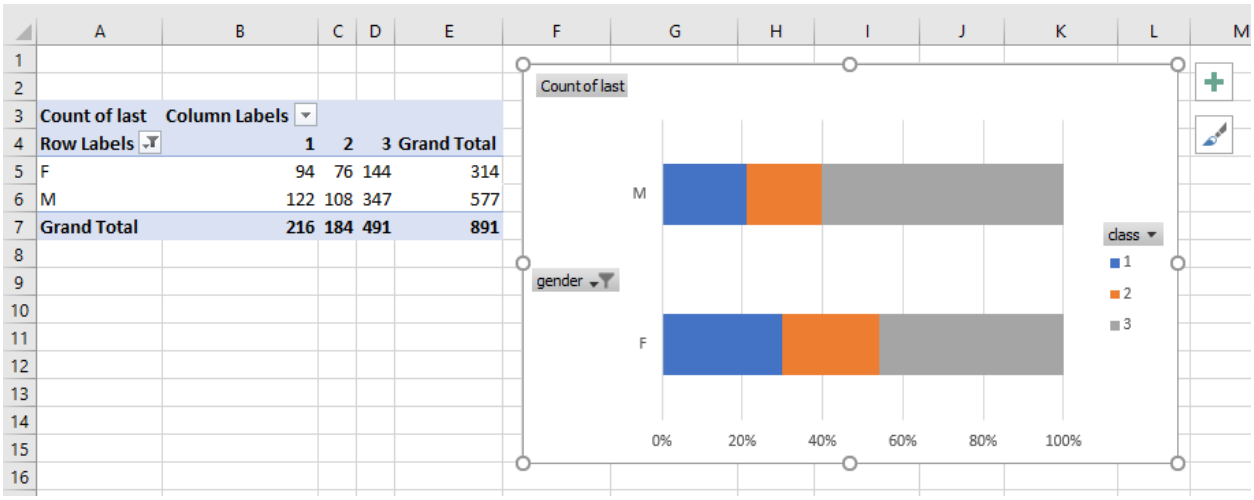
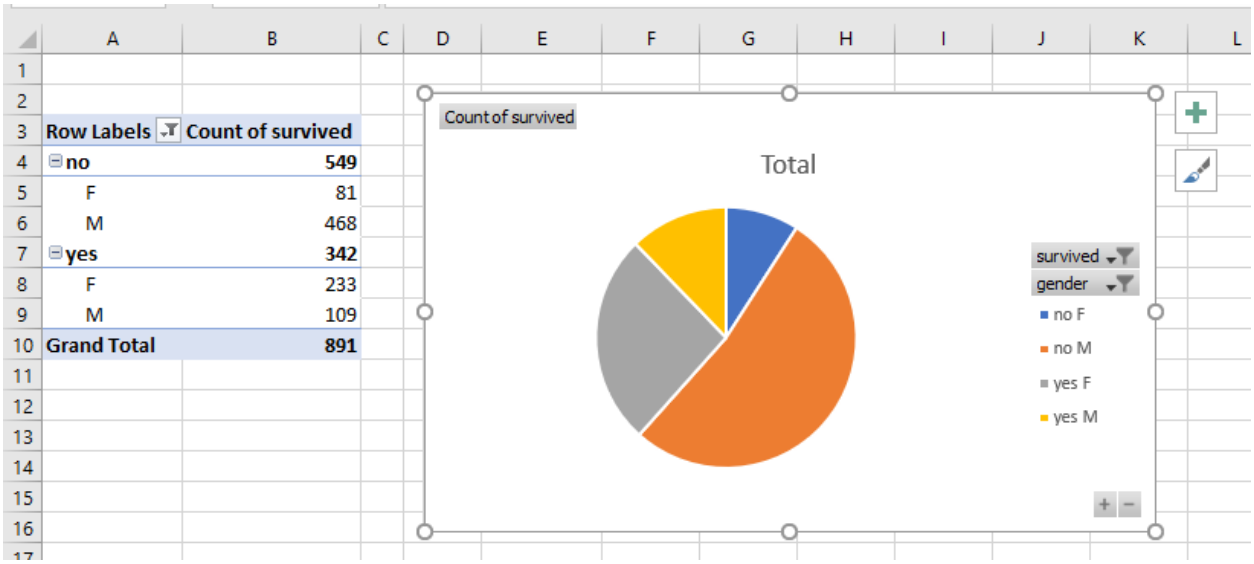


Figure-17.2: Passengers in each gender

Problem 18. Create a pie chart showing the relative number of male survivors, male non-survivors, female survivors, and female non-survivors (four slices).

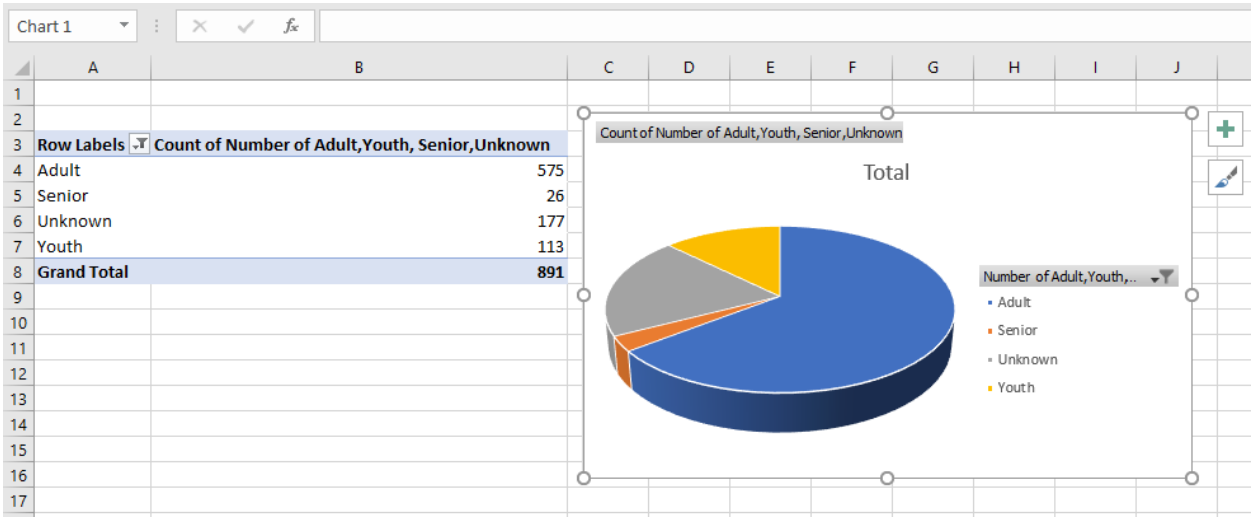
Answer-18: “pivot table” for Data analyzing and “Pie chart” for Data Visualization.



Row Labels	Count of survived
no	549
F	81
M	468
yes	342
F	233
M	109
Grand Total	891

Problem 19. Let "youth" denote passengers whose age is under 18, "adult" denote passengers age 18-59, and "senior" denote passengers whose age is 60 and above. Create a pie chart with four slices showing the relative number of youth, adult, senior, and those whose age is unknown. Hint: consider using function ==if(*,*,if(*,*,*))..

Answer-19: “pivot table” for Data analyzing and “Pie chart” for Data Visualization.



Row Labels	Count of Number of Adult,Youth, Senior,Unknown
Adult	575
Senior	26
Unknown	177
Youth	113
Grand Total	891