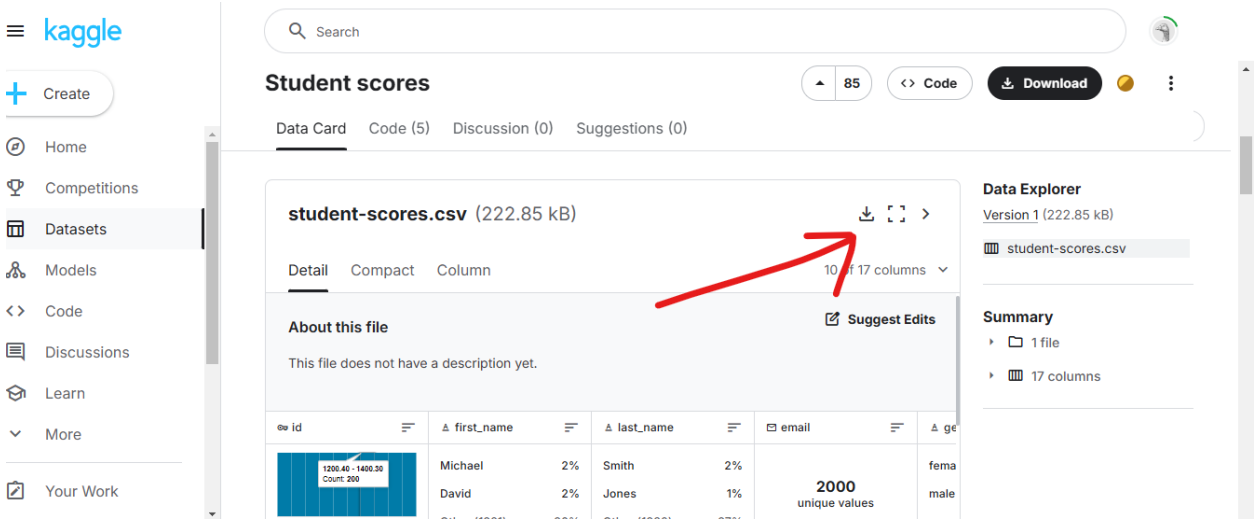
NAME : Pranav Madhukar Sirsufale

ID : 2119822

DEPARTMENT : M.Sc. Computer Science

1] Gather data

- Download dataset from Kaggle. (given link) >> click on download button
  https://www.kaggle.com/datasets/markmedhat/student-scores/data



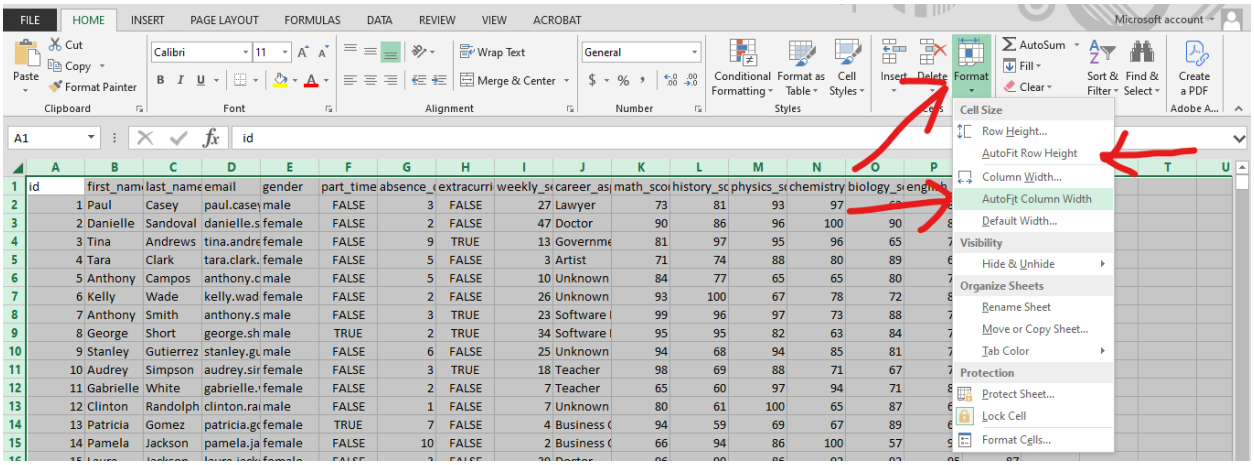
Open it in Excel.

id	first_nam	last_name	email	gender	part_time	absence_(	extracurri	weekly_s	career_as	math_sco	history_sc	physics_s	chemistry	biology_s	english_s	geography_score
1	Paul	Casey	paul.casey	male	FALSE	3	FALSE	27	Lawyer	73	81	93	97	63	80	87
2	Danielle	Sandoval	danielle.s	female	FALSE	2	FALSE	47	Doctor	90	86	96	100	90	88	90
3	Tina	Andrews	tina.andre	female	FALSE	9	TRUE	13	Governme	81	97	95	96	65	77	94
4	Tara	Clark	tara.clark.	female	FALSE	5	FALSE	3	Artist	71	74	88	80	89	63	86
5	Anthony	Campos	anthony.c	male	FALSE	5	FALSE	10	Unknown	84	77	65	65	80	74	76
6	Kelly	Wade	kelly.wad	female	FALSE	2	FALSE	26	Unknown	93	100	67	78	72	80	84
7	Anthony	Smith	anthony.s	male	FALSE	3	TRUE	23	Software I	99	96	97	73	88	76	64
8	George	Short	george.sh	male	TRUE	2	TRUE	34	Software I	95	95	82	63	84	70	85
9	Stanley	Gutierrez	stanley.gu	male	FALSE	6	FALSE	25	Unknown	94	68	94	85	81	74	72
10	Audrey	Simpson	audrey.sir	female	FALSE	3	TRUE	18	Teacher	98	69	88	71	67	71	73
11	Gabrielle	White	gabrielle.(	female	FALSE	2	FALSE	7	Teacher	65	60	97	94	71	81	66
12	Clinton	Randolph	clinton.ra	male	FALSE	1	FALSE	7	Unknown	80	61	100	65	87	64	61

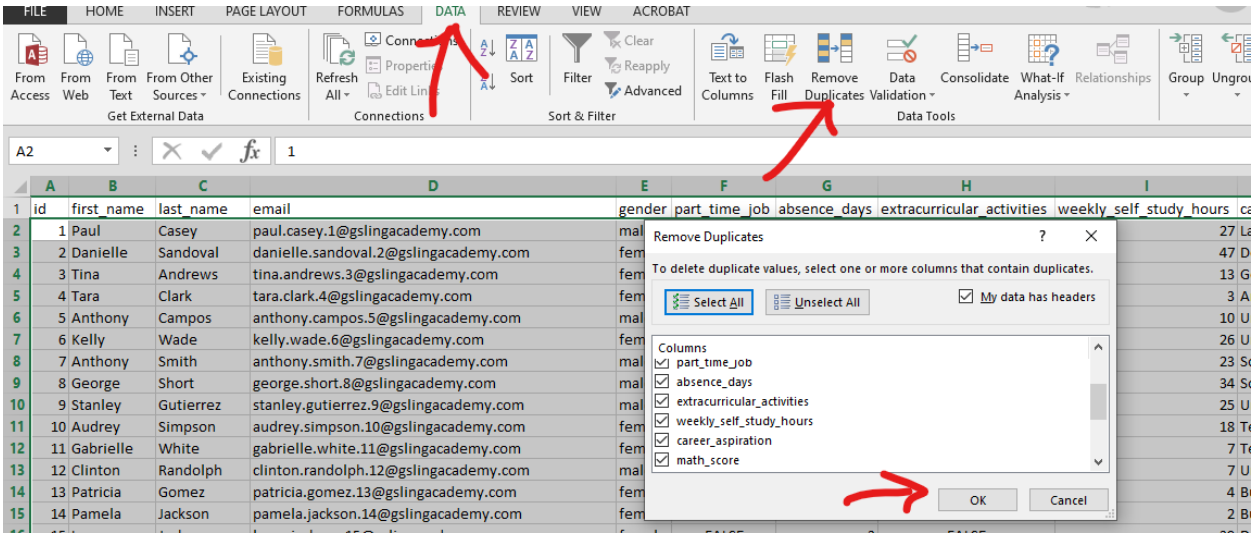
Shape = (17 x 2000)

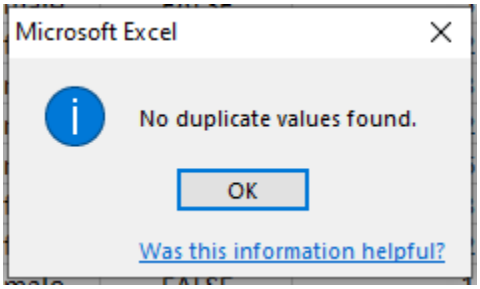
2] Organize and clean data (remove duplicates , handling missing values)

- Go To Home tab > Format > Select Autofit column width and autofit row height (so we can see it properly)

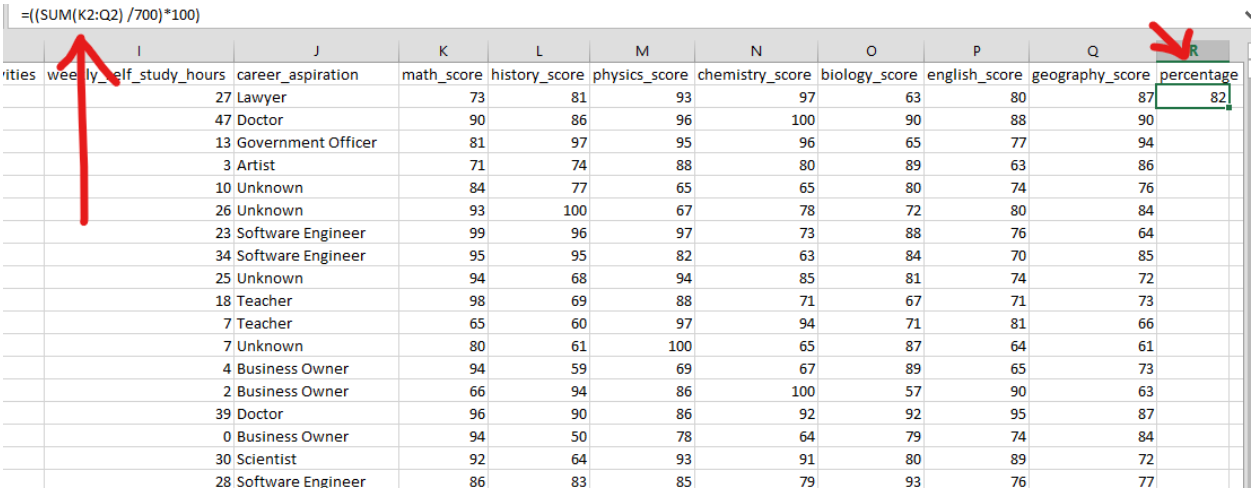


- Go to Data Rebun >> Remove Duplicates >> a window will appear then click ok it will remove all duplicates if there are.





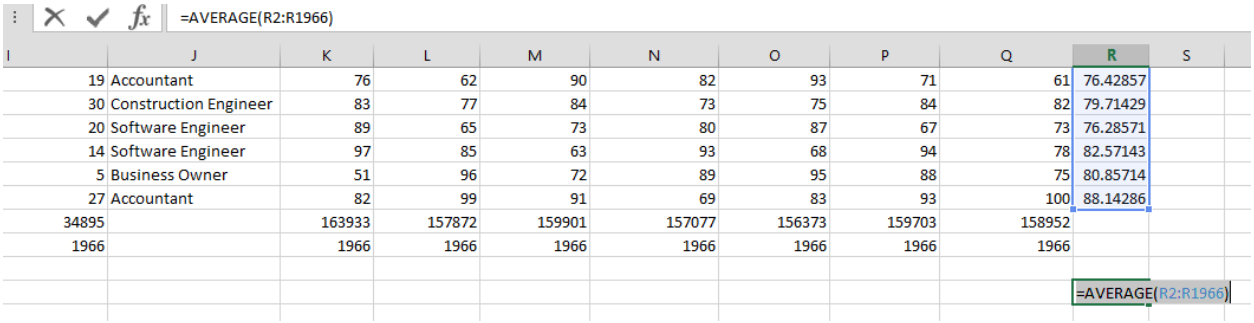
- calculating percentage column using =((SUM(K2:Q2) /700)*100) this formula



=((SUM(K2:Q2) /700)*100)

I	J	K	L	M	N	O	P	Q	R
weekly_self_study_hours	career_aspiration	math_score	history_score	physics_score	chemistry_score	biology_score	english_score	geography_score	percentage
27	Lawyer	73	81	93	97	63	80	87	82
47	Doctor	90	86	96	100	90	88	90	
13	Government Officer	81	97	95	96	65	77	94	
3	Artist	71	74	88	80	89	63	86	
10	Unknown	84	77	65	65	80	74	76	
26	Unknown	93	100	67	78	72	80	84	
23	Software Engineer	99	96	97	73	88	76	64	
34	Software Engineer	95	95	82	63	84	70	85	
25	Unknown	94	68	94	85	81	74	72	
18	Teacher	98	69	88	71	67	71	73	
7	Teacher	65	60	97	94	71	81	66	
7	Unknown	80	61	100	65	87	64	61	
4	Business Owner	94	59	69	67	89	65	73	
2	Business Owner	66	94	86	100	57	90	63	
39	Doctor	96	90	86	92	92	95	87	
0	Business Owner	94	50	78	64	79	74	84	
30	Scientist	92	64	93	91	80	89	72	
28	Software Engineer	86	83	85	79	93	76	77	

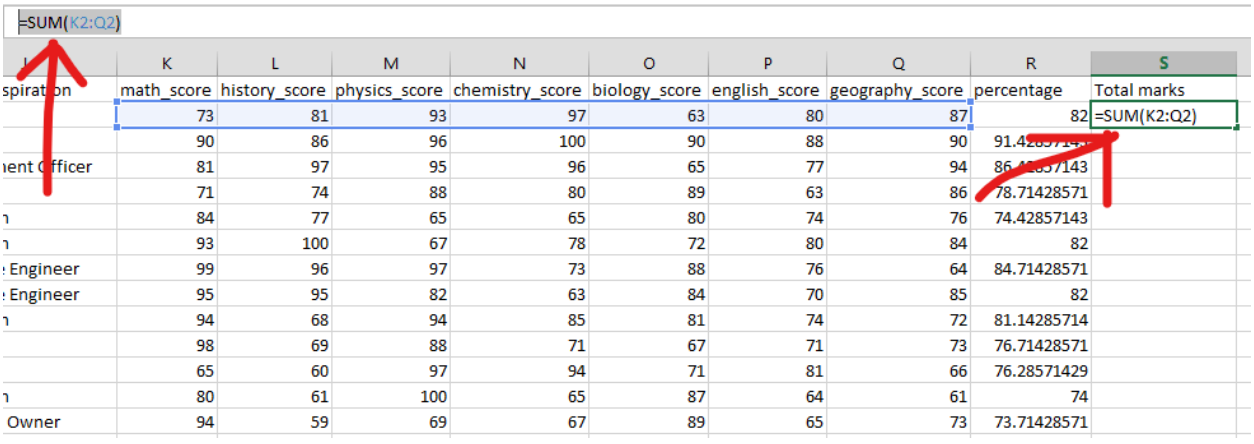
- calculating the average percentages using this formula =AVERAGE(R2:R1966)



=AVERAGE(R2:R1966)

I	J	K	L	M	N	O	P	Q	R	S
19	Accountant	76	62	90	82	93	71	61	76.42857	
30	Construction Engineer	83	77	84	73	75	84	82	79.71429	
20	Software Engineer	89	65	73	80	87	67	73	76.28571	
14	Software Engineer	97	85	63	93	68	94	78	82.57143	
5	Business Owner	51	96	72	89	95	88	75	80.85714	
27	Accountant	82	99	91	69	83	93	100	88.14286	
34895		163933	157872	159901	157077	156373	159703	158952		
1966		1966	1966	1966	1966	1966	1966	1966		
									=AVERAGE(R2:R1966)	

- adding column **Total marks** using **summation** =sum()

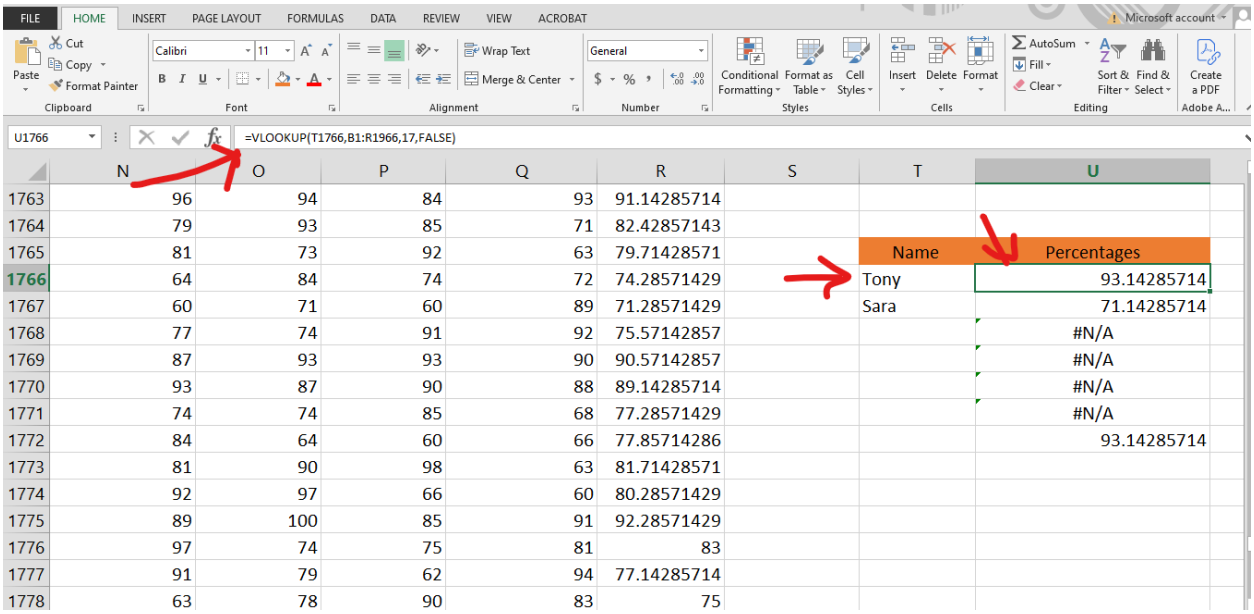


=SUM(K2:Q2)

I	K	L	M	N	O	P	Q	R	S
spiration	math_score	history_score	physics_score	chemistry_score	biology_score	english_score	geography_score	percentage	Total marks
	73	81	93	97	63	80	87	82	=SUM(K2:Q2)
	90	86	96	100	90	88	90	91.42857143	
ient Officer	81	97	95	96	65	77	94	86.42857143	
	71	74	88	80	89	63	86	78.71428571	
n	84	77	65	65	80	74	76	74.42857143	
n	93	100	67	78	72	80	84	82	
Engineer	99	96	97	73	88	76	64	84.71428571	
Engineer	95	95	82	63	84	70	85	82	
n	94	68	94	85	81	74	72	81.14285714	
	98	69	88	71	67	71	73	76.71428571	
	65	60	97	94	71	81	66	76.28571429	
n	80	61	100	65	87	64	61	74	
Owner	94	59	69	67	89	65	73	73.71428571	

Apply **=VLOOPUP** function .(getting percentages corresponding names) using the expression >

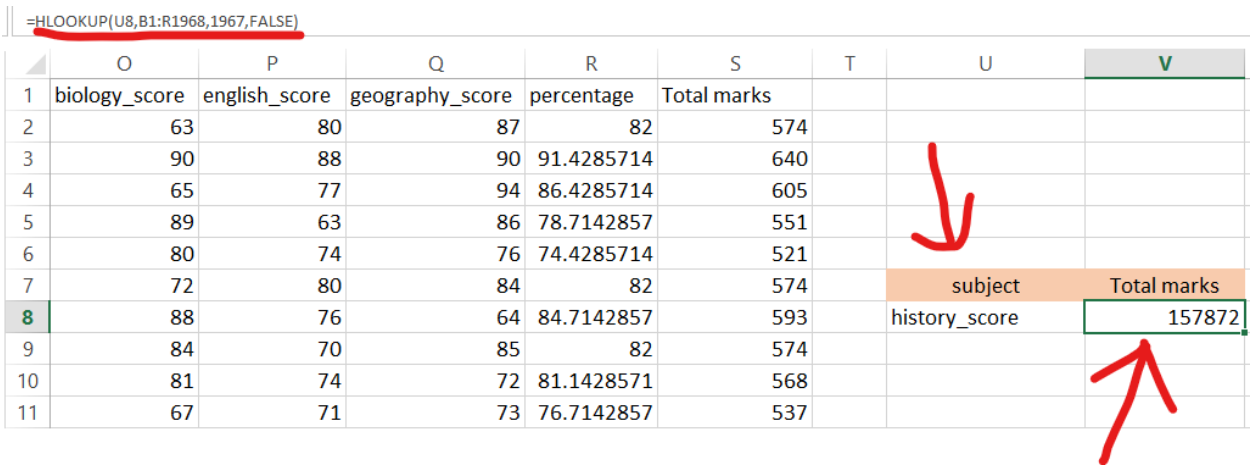
=VLOOKUP(T1766,B1:R1966,17,FALSE)



	N	O	P	Q	R	S	T	U
1763	96	94	84	93	91.14285714			
1764	79	93	85	71	82.42857143			
1765	81	73	92	63	79.71428571		Name	Percentages
1766	64	84	74	72	74.28571429		Tony	93.14285714
1767	60	71	60	89	71.28571429		Sara	71.14285714
1768	77	74	91	92	75.57142857			#N/A
1769	87	93	93	90	90.57142857			#N/A
1770	93	87	90	88	89.14285714			#N/A
1771	74	74	85	68	77.28571429			#N/A
1772	84	64	60	66	77.85714286			93.14285714
1773	81	90	98	63	81.71428571			
1774	92	97	66	60	80.28571429			
1775	89	100	85	91	92.28571429			
1776	97	74	75	81	83			
1777	91	79	62	94	77.14285714			
1778	63	78	90	83	75			

Apply =HLOOKUP() function getting subjectwise total marks using formula

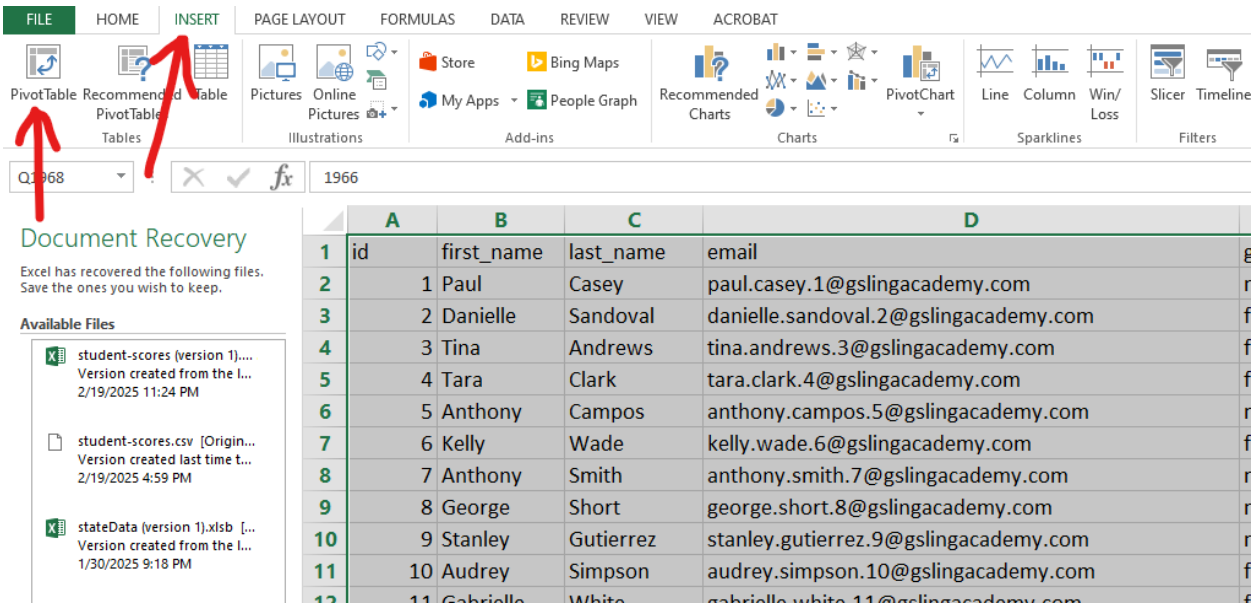
=HLOOKUP(U8,B1:R1968,1967,FALSE)



	O	P	Q	R	S	T	U	V
1	biology_score	english_score	geography_score	percentage	Total marks			
2	63	80	87	82	574			
3	90	88	90	91.4285714	640			
4	65	77	94	86.4285714	605			
5	89	63	86	78.7142857	551			
6	80	74	76	74.4285714	521			
7	72	80	84	82	574		subject	Total marks
8	88	76	64	84.7142857	593		history_score	157872
9	84	70	85	82	574			
10	81	74	72	81.1428571	568			
11	67	71	73	76.7142857	537			

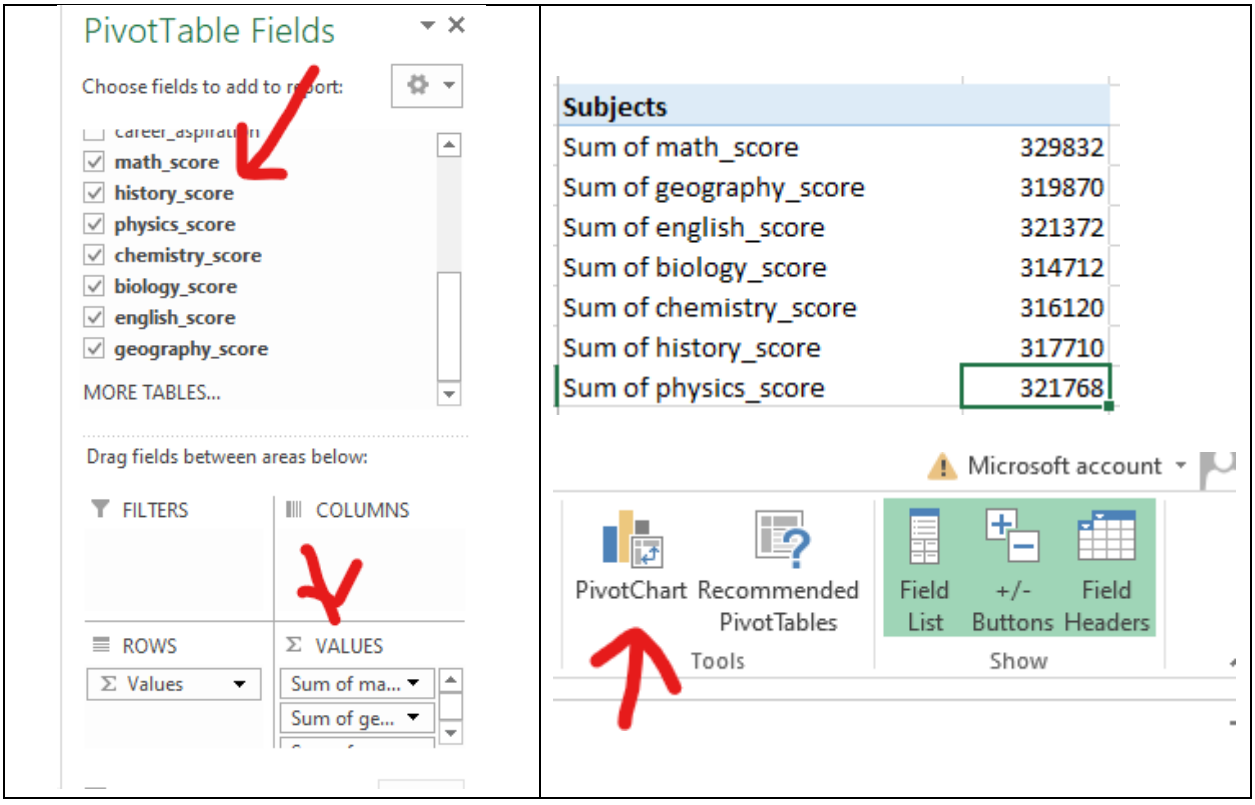
PIVOT TABLE

- go to Insert rebon



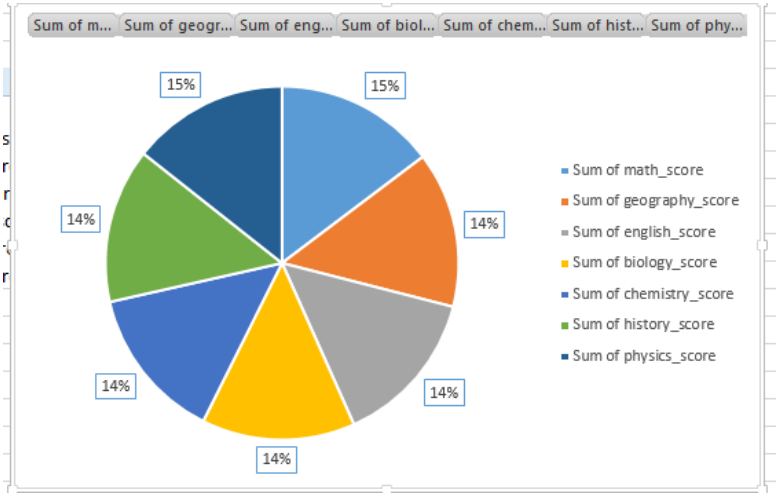
	A	B	C	D
1	id	first_name	last_name	email
2	1	Paul	Casey	paul.casey.1@gslingacademy.com
3	2	Danielle	Sandoval	danielle.sandoval.2@gslingacademy.com
4	3	Tina	Andrews	tina.andrews.3@gslingacademy.com
5	4	Tara	Clark	tara.clark.4@gslingacademy.com
6	5	Anthony	Campos	anthony.campos.5@gslingacademy.com
7	6	Kelly	Wade	kelly.wade.6@gslingacademy.com
8	7	Anthony	Smith	anthony.smith.7@gslingacademy.com
9	8	George	Short	george.short.8@gslingacademy.com
10	9	Stanley	Gutierrez	stanley.gutierrez.9@gslingacademy.com
11	10	Audrey	Simpson	audrey.simpson.10@gslingacademy.com

- drag the PivotTable Fields into VALES tab and ROWS it will create a table then go to analayze rebone and click on PivotChart



- Select an Appropriate Chart for the visualization

Pie Chart



Bar chart