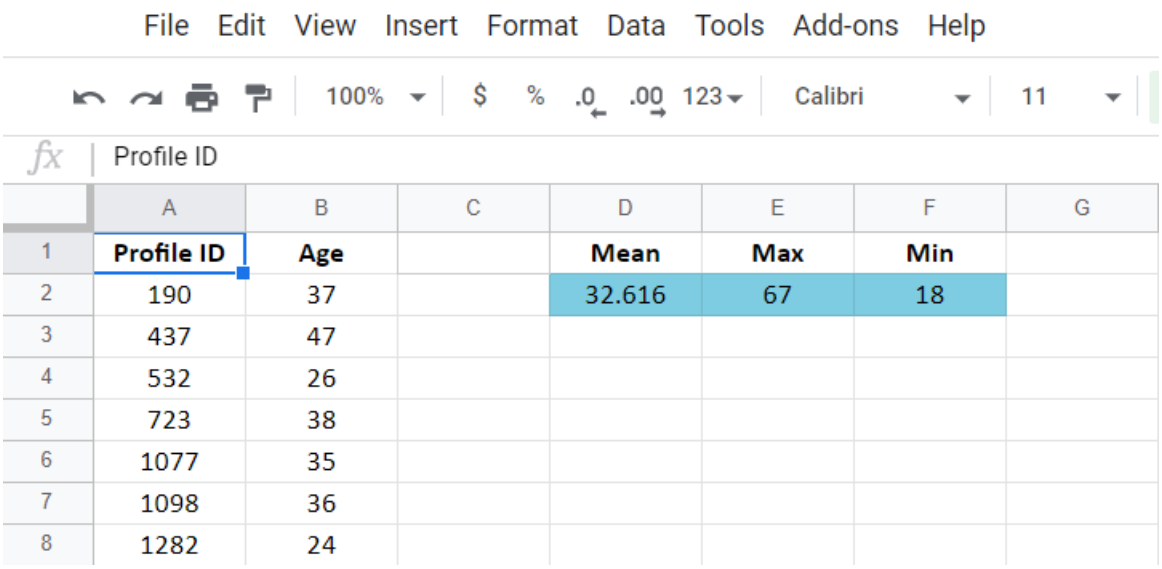
CASE STUDY

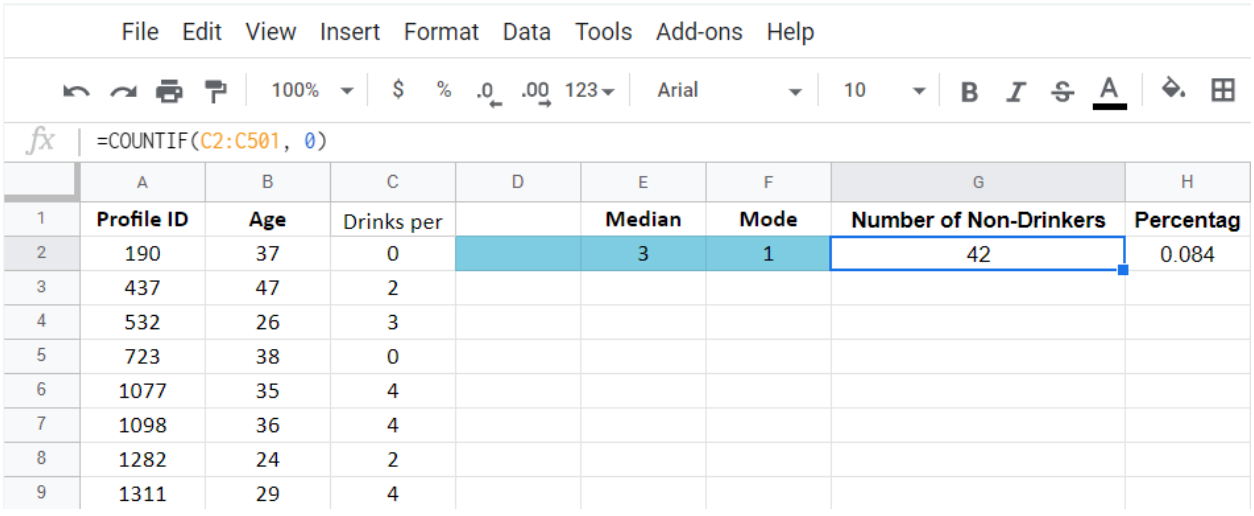
Apply statistical knowledge and solidify everything by working through a case study consisting of online dating profile data.

1.1 Calculate mean, max and min of age



	A	B	C	D	E	F	G
1	**Profile ID**	**Age**		**Mean**	**Max**	**Min**	
2	190	37		32.616	67	18	
3	437	47					
4	532	26					
5	723	38					
6	1077	35					
7	1098	36					
8	1282	24					

1.2 Calculate median, mode and number of non-drinkers



	A	B	C	D	E	F	G	H
1	**Profile ID**	**Age**	Drinks per		**Median**	**Mode**	**Number of Non-Drinkers**	**Percentag**
2	190	37	0		3	1	42	0.084
3	437	47	2					
4	532	26	3					
5	723	38	0					
6	1077	35	4					
7	1098	36	4					
8	1282	24	2					
9	1311	29	4					

1.3 Calculate variance, standard deviation, max and z-score of hours since last login data

File Edit View Insert Format Data Tools Add-ons Help

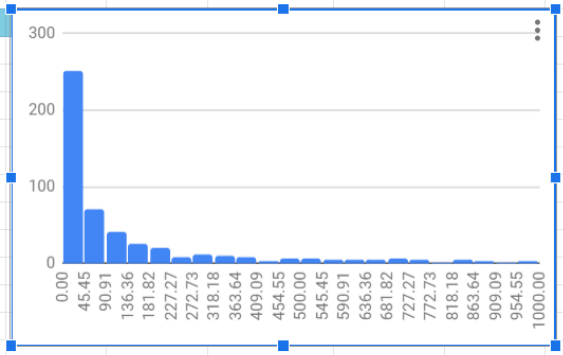
100% $ % .0 .00 123 Arial 10

fx =(J2-F2)/I2

	A	B	C	D	E	F	G	H	I	J	K
1	**Profile ID**	**Hours**		**Mode**	**2nd**	**Mean**		**Variance**	**Standard**	**Max**	**Z-Score**
2	41857	184		31	45	139		42259.76726	205.5718056	963	4.008331773
3	5580	33									
4	715	68									
5	35291	31									
6	29289	271									
7	1183	29									

2 Insert a histogram

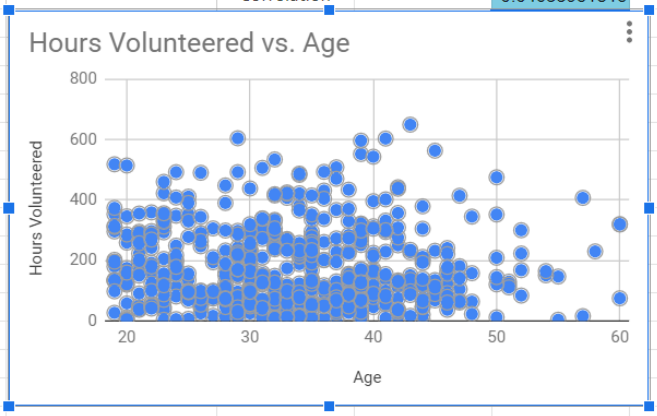
A	B	C	D	E	F	G	H	I	J	K
Profile ID	**Hours**		**Mode:**	**31**						
41857	184		2nd	45						
5580	33		Maximum:	963						
715	68		Mean:	139.174						
35291	31		Standard	206						
29289	271									
1183	29									
34143	589									
9883	127									
21387	18									
20648	62									
18759	47									
30222	9									
15062	86									



3 Calculate the correlation between age and hours of going to church and make a scatter chart.

fx

	A	B	C	D	E	F	G	H
1	**Profile ID**	**Age**	Hours		Correlation		-0.04033961816	
2	19971	29	186					
3	27485	27	17					
4	3655	35	282					
5	12052	40	313					
6	28096	55	147					
7	18008	46	124					
8	36352	32	249					
9	7970	23	46					
10	57506	20	259					
11	32573	23	293					
12	23335	38	156					
13	39702	20	143					
14	58860	37	3					
15	40443	25	411					



4.1 Make chi-squared test of:

H0: The `roommates` & `gender` variables are independent

H1: The `roommates` & `gender` variables are not independent

With P value=0.05 cutoff

File Edit View Insert Format Data Tools Add-ons Help

=CHITEST(B3:C7, G3:H7)

	Observed Frequencies					Expected Frequencies					chiSq P-Value	Result
	roommates	female	male	Grand Total		*roommates*	female	male	Grand Total		0.8167780085	FAIL TO REJECT
	1	3	2	5		1	2.6	2.4	5			
	2	24	23	47		2	24.44	22.56	47			
	3	57	47	104		3	54.08	49.92	104			
	4	20	23	43		4	22.36	20.64	43			
	5		1	1		5	0.52	0.48	1			
	Grand Total	**104**	**96**	**200**		**Grand**	**104**	**96**	**200**			

4.2 Calculate the correlation between income and age, a best fit line for income and age, calculate the "fit" of the linest function using "intercept" + ("age" * "slope") with B2

=H2+(B2*G2)

	A	B	C	D	E	F	G	H
1	**Profile ID**	**Age**	**Income**		**Fit**	**Correlation**	**Line Slope**	**Y-Intercept**
2	41307	41	186996		186,455.70	0.25	3,946.86	24,634.57
3	46003	48	212721					
4	18665	26	129236					
5	11564	36	164546					
6	14366	22	109203					

4.3 Calculate the coefficients using LINEST(), follow the multiple linear regression equation, calculate the first observation's estimate in cell K2. Compare it with the actual value in B2

=(C2*G2)+(D2*H2)+(E2*I2)+J2

	A	B	C	D	E	F	G	H	I	J	K
1	**Profile ID**	**Hours since last**	**Age**	**Income**	**Hours Volunteered**		**age_coeff**	**Income_coeff**	**Hours_coeff**	**y-intercept**	**Estimate**
2	1197	63.65	57	22	49		0.14	-0.01	1.30	-11.28	60.20
3	4127	7.1167	24	260	8						
4	2881	11.9667	24	25	9						
5	4131	8.8833	24	217	8						
6	5522	6.75	24	178	6						