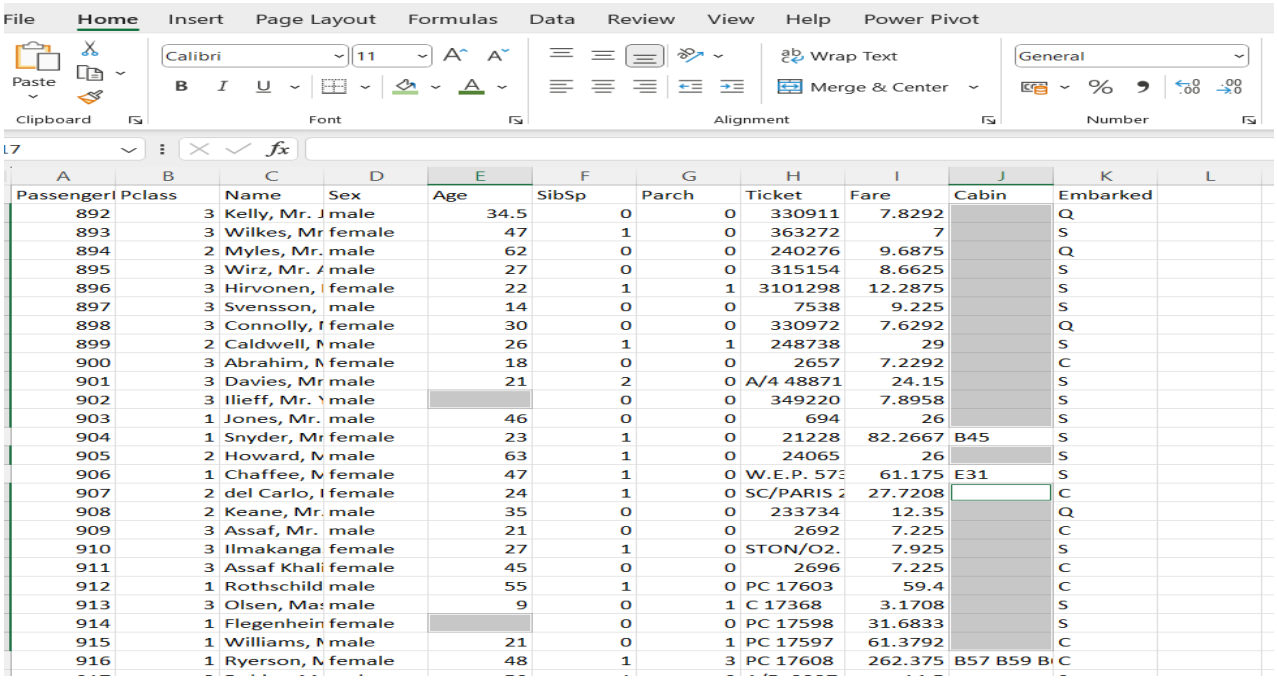
Task 1 Clean a dataset by removing missing values and outliers.

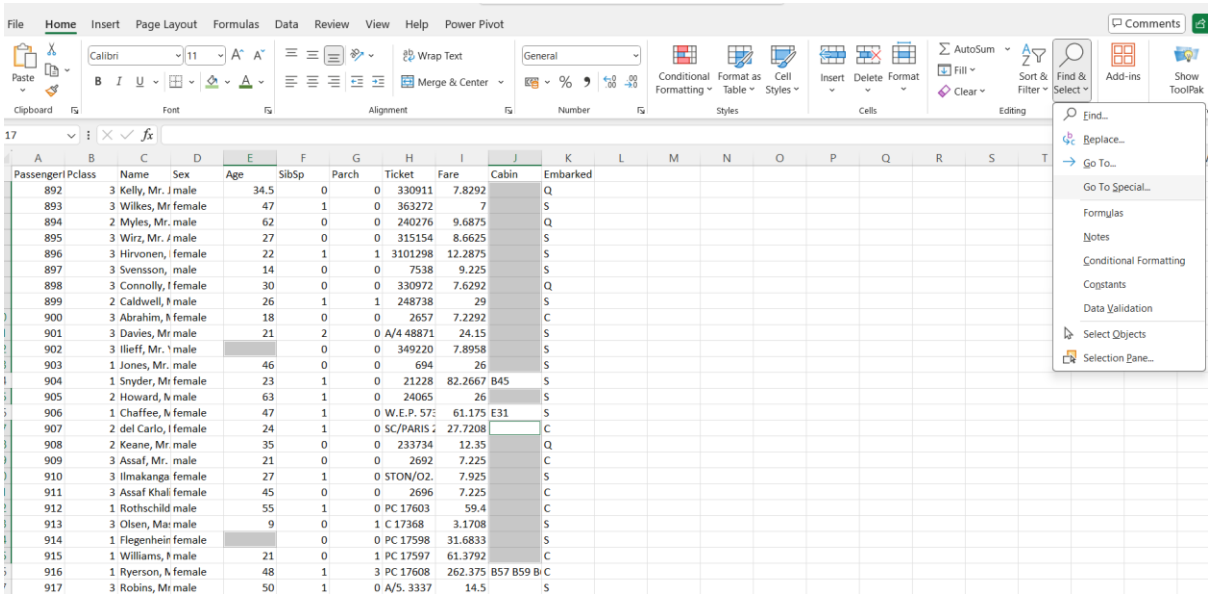
1. *For removing values first go to **Find and Select** in Home tab, Editing group.*
2. *In Find and Select, select **Go to Special.***
3. *In Go to Special**, select Blanks, click OK.***

In this image you can see blank or missing values

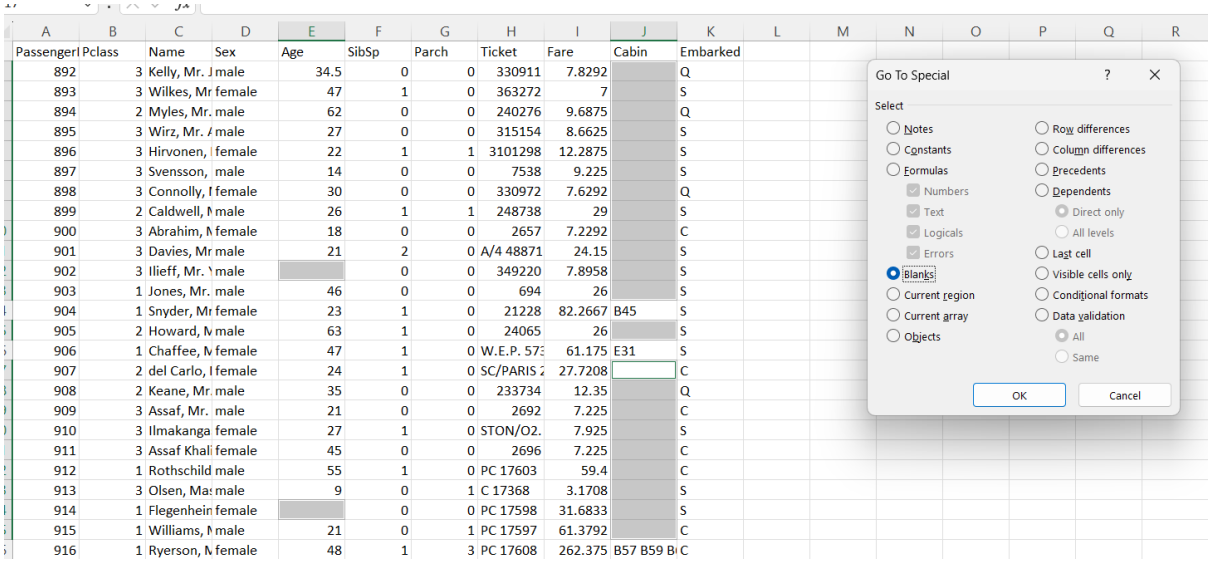


PassengerI	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
892	3	Kelly, Mr. J	male	34.5	0	0	330911	7.8292		Q
893	3	Wilkes, Mr	female	47	1	0	363272	7		S
894	2	Myles, Mr.	male	62	0	0	240276	9.6875		Q
895	3	Wirz, Mr. A	male	27	0	0	315154	8.6625		S
896	3	Hirvonen, I	female	22	1	1	3101298	12.2875		S
897	3	Svensson,	male	14	0	0	7538	9.225		S
898	3	Connolly, I	female	30	0	0	330972	7.6292		Q
899	2	Caldwell, M	male	26	1	1	248738	29		S
900	3	Abrahim, M	female	18	0	0	2657	7.2292		C
901	3	Davies, Mr	male	21	2	0	A/4 48871	24.15		S
902	3	Ilieff, Mr. Y	male		0	0	349220	7.8958		S
903	1	Jones, Mr.	male	46	0	0	694	26		S
904	1	Snyder, Mr	female	23	1	0	21228	82.2667	B45	S
905	2	Howard, M	male	63	1	0	24065	26		S
906	1	Chaffee, M	female	47	1	0	W.E.P. 573	61.175	E31	S
907	2	del Carlo, I	female	24	1	0	SC/PARIS 2	27.7208		C
908	2	Keane, Mr.	male	35	0	0	233734	12.35		Q
909	3	Assaf, Mr.	male	21	0	0	2692	7.225		C
910	3	Ilmakanga	female	27	1	0	STON/O2.	7.925		S
911	3	Assaf Khali	female	45	0	0	2696	7.225		C
912	1	Rothschild	male	55	1	0	PC 17603	59.4		C
913	3	Olsen, Mas	male	9	0	1	C 17368	3.1708		S
914	1	Flegenhein	female		0	0	PC 17598	31.6833		S
915	1	Williams, M	male	21	0	1	PC 17597	61.3792		C
916	1	Ryerson, M	female	48	1	3	PC 17608	262.375	B57 B59 B	C

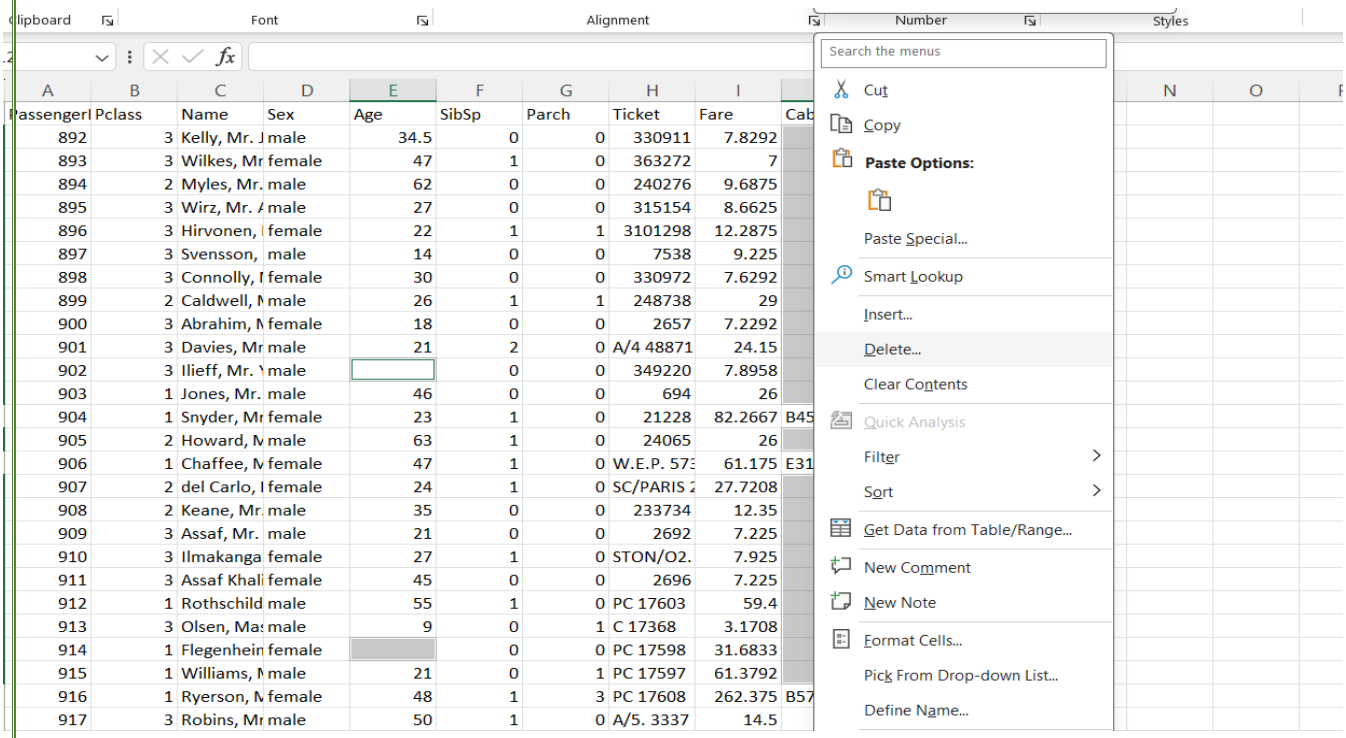
*For removing values first go to **Find and Select** in Home tab, Editing group.*



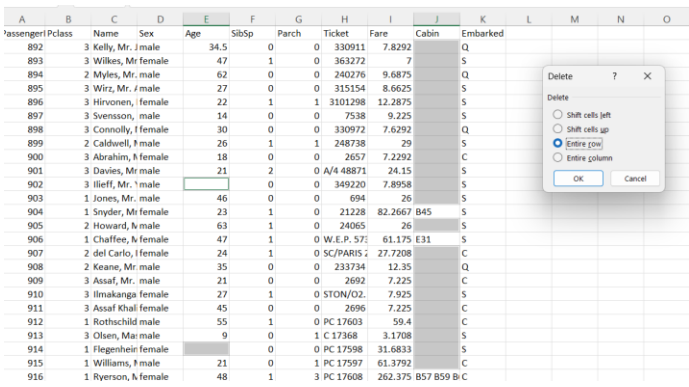
In Go to Special, ***select Blanks, click OK.***

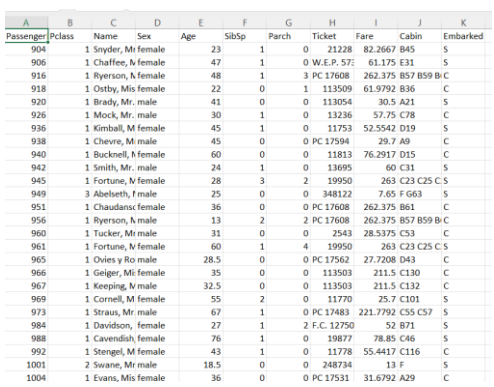


After Click Ok, Right Click on Blank Cells and Click on Delete

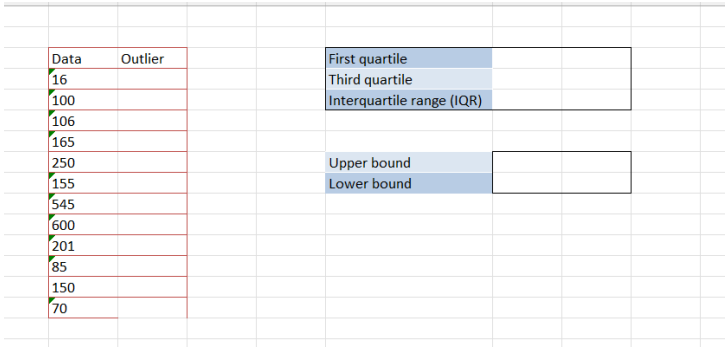


After Click on Delete, Select Entire Row and Click OK, After OK you can see all blank is deleted





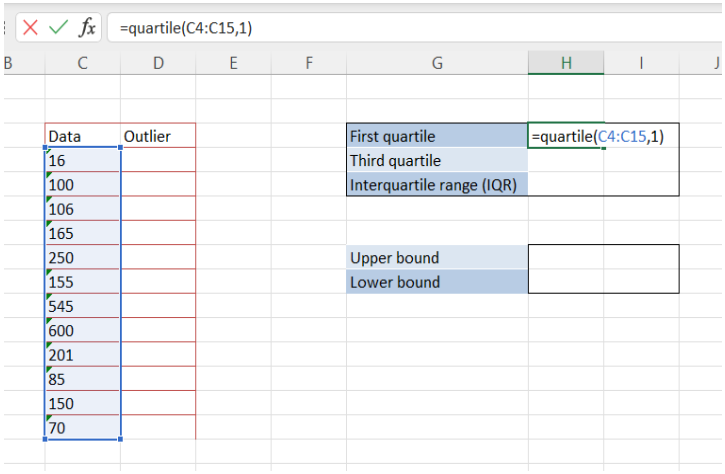
Cleaning dataset of removing Outliers



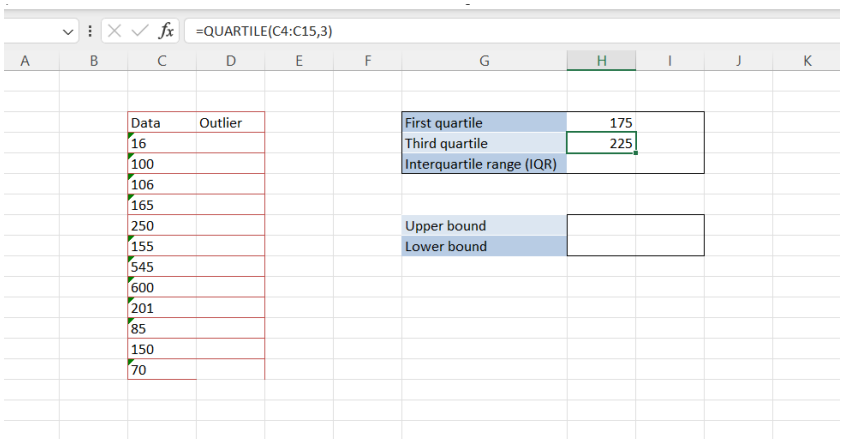
Data	Outlier
16	
100	
106	
165	
250	
155	
545	
600	
201	
85	
150	
70	

First quartile	
Third quartile	
Interquartile range (IQR)	
Upper bound	
Lower bound	

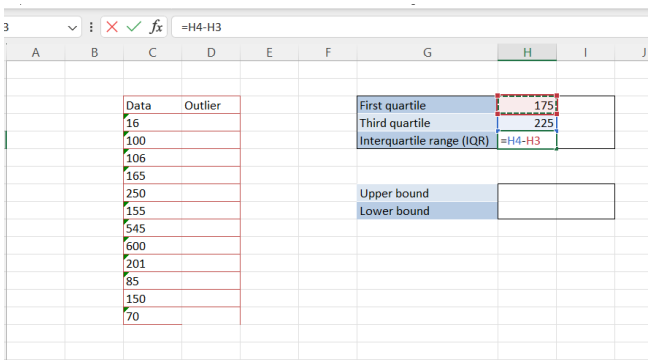
First, Go to First quartile, Go for the formula =quartile(select the data column range, minimum value)

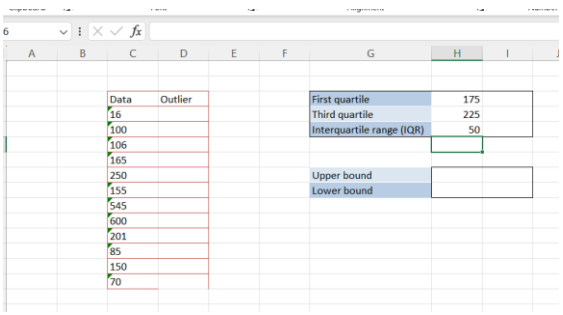


First, Go to Third quartile, Go for the formula =quartile(select the data column range, Third quartile)

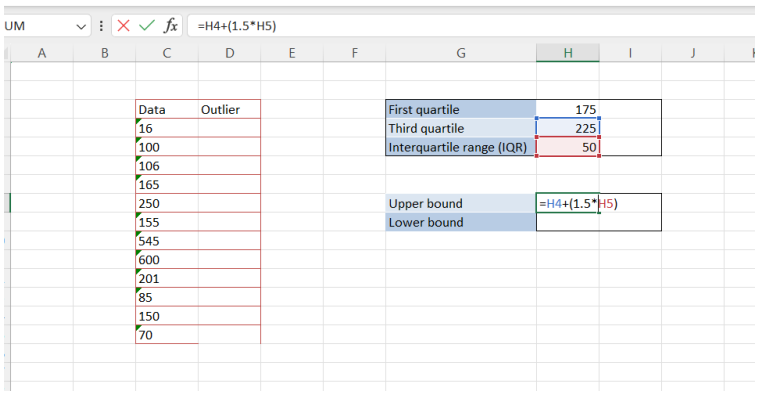


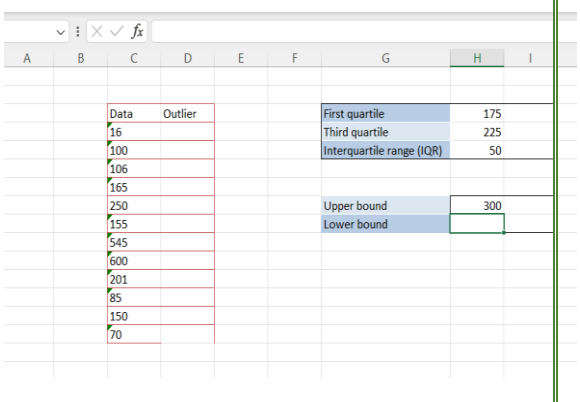
For Interquartile, Go for the formula = quartile(Third quartile – First quartile)



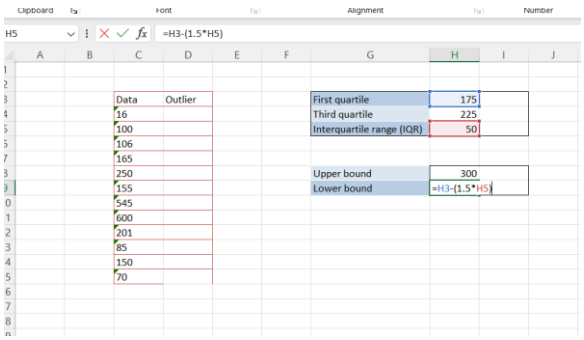


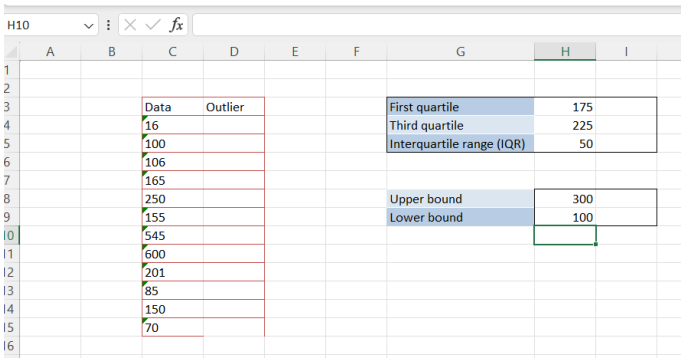
For Upper Bound Go to the formula Third quartile +(1.5 Interquartile)*





For Lower Bound Go to the Formula First quartile – (1.5 Interquartile)*

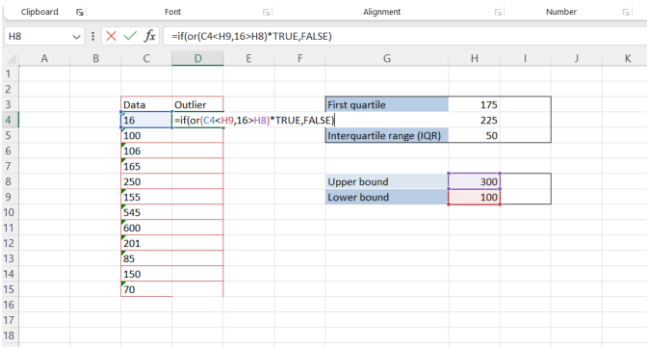


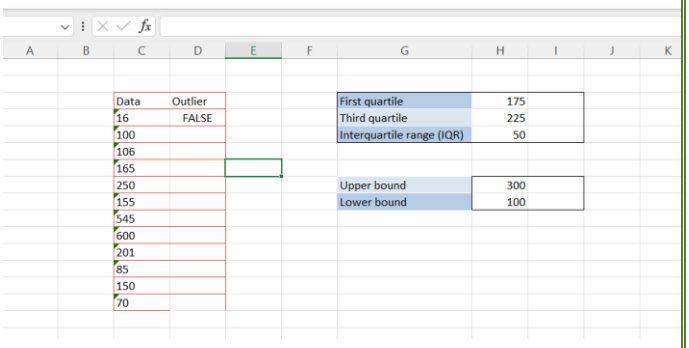


How to find Outliers

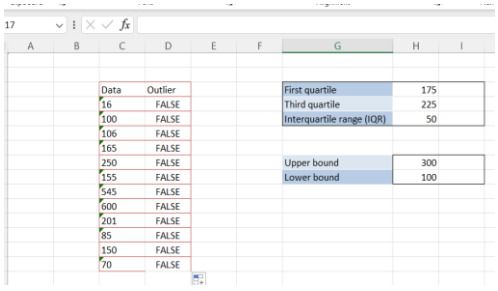
Having 1000 in place of 1 is a false outlier

For Outliers Go to the formula if(or(16<LowerBound, 16>UpperBound)*TRUE,FALSE)





Autofill all the outliers here



Shaikh Jasmin Kauser