LESSON HANDOUT

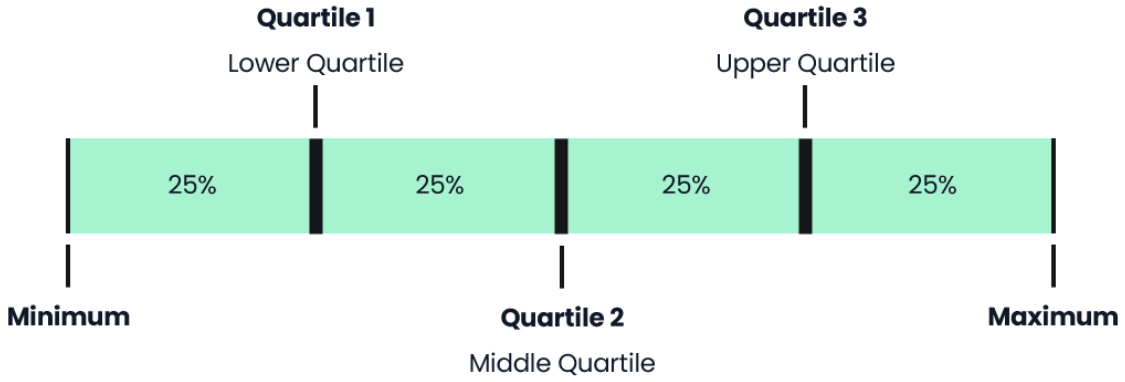
Quartiles and percentiles

Quartiles & percentiles are both ways we can divide our data to help us measure how our data is spread. Additionally, they can assist in calculating a single value's significance.

Quartiles

Quartiles divide our data into four quarters such that:

- Values below the first quartile is the lower quarter, where 25% of our data resides
- Values between the first and third quartile are the middle two quarters, where 50% of our data resides
- Values above the third quartile is the upper quarter where 25% of our data resides



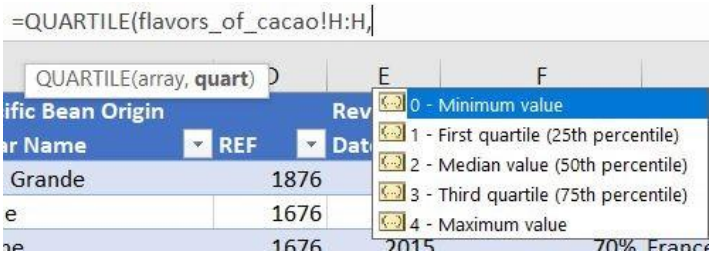
Interquartile range

The interquartile range (IQR) is the range of our middle two quarters and is calculated by subtracting the first quartile by the third quartile. The interquartile range is a measure of how spread apart our values are from the median and can be used to detect outliers. Generally, any value which lies more than 1.5 times the IQR above the third quartile or below the first quartile is considered a moderate outlier and those which lie more than 3 times the IQR are considered strong outliers.

Calculating Quartiles and the IQR

Calculating Quartiles is quite simple with the use of the Excel formula **=QUARTILE**

- **=QUARTILE**(array, quartile to return)



To calculate the **IQR**, subtract the first quartile from the third quartile.

- = (Q3 - Q1), as seen in the screenshot below:

Ratings	Quartiles
Q1	1.20
Q2	2.80
Q3	4.95
IQR	=C6-C4

The upper outlier can be calculated by Q3 plus 1.5 times the IQR: ● = Q3 + (IQR * 1.5)

Ratings	Quartiles
Q1	1.20
Q2	2.80
Q3	4.95
IQR	3.75
Upper Outlier	=C6+(C7*1.5)
Lower Outlier	

And the lower outlier can be calculated by Q1 minus 1.5 times the IQR ● = Q1 - (IQR * 1.5)

Ratings	Quartiles
Q1	1.20
Q2	2.80
Q3	4.95
IQR	3.75
Upper Outlier	10.58
Lower Outlier	=C4-(C7*1.5)

NOTE: You won't always have logical outliers - in the example screenshots above, if the 'Ratings' were from 0 - 10, you wouldn't have any upper or lower outliers as the number falls outside of the minimum and maximum ranges.

Percentiles

Quartiles are just one way to split our data apart. We could instead choose to split our data into five segment (quanties) for ten segments (deciles), but this is much less common. We can split our data into different segments by using percentiles.

- A percentile is the value below which a percentage of your data falls. For example, the 33rd percentile is above 33% of values in the dataset.

- We can also find out where a value sits in comparison to the dataset with its percentile rank - the % of values this value is above. For example, a person who is in the 50th percentile rank height is taller than half of those included in the dataset.

Classifying your dataset:

You can classify your dataset by using Quartiles and VLOOKUP.

First, build your table so you have classes from Very Low through to Very high (you can use more appropriate wording for the data you are using) and use your Quarterlies and outliers to fill in the classes. For example:

Fig 1:

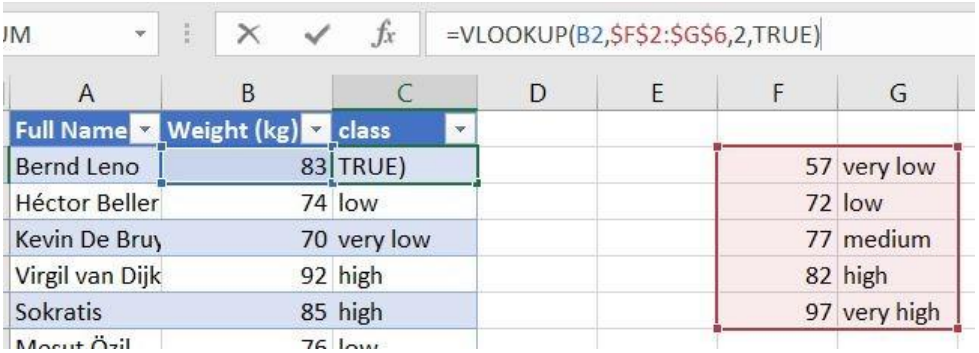
Lower Outlier	**Very Low**
Q1	**Low**
Q2	**Medium**
Q3	**High**
Upper Outlier	**Very High**

Fig 2:

57	very low
72	low
77	medium
82	high
97	very high

Then, use VLOOKUP and an APPROXIMATE MATCH to calculate the class. What the approximate match does is take the number and find the closest value *lower* than itself.

- =VLOOKUP(cell to calculate, range of your classification table, column index number, TRUE)



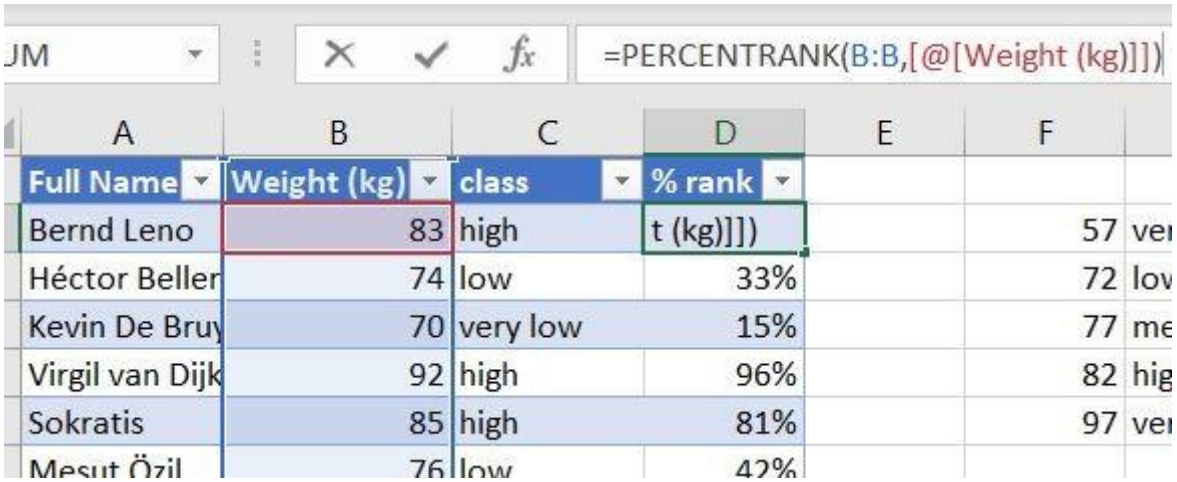
When classifying in this way, it's a good idea to provide your classification within the data dictionary.

Calculating Percentiles:

You can quickly calculate percentiles by using the formula **=PERCENT RANK.** It returns the rank of a value in a dataset as a percentage of the dataset.

- **=PERCENTRANK**(array, x) - Where "x" is the value
- Remember to change the format to Percentage using the % button

Screenshot example below:



Further reading:

Quartile Function

Percent Rank Function