

## Quiz 5 - Handling Missing Values in KNIME and Spark

### Quiz 5 - Handling Missing Values in KNIME and Spark

---

1. If we remove all missing values from the data, how many air pressure at 9am measurements have values between 911.736 and 914.67?

- **77**
- 287
- 80

2. If we impute the missing values with the minimum value, how many air temperature at 9am measurements are less than 42.292?

- **28**
- 23
- 1
- 5

3. How many samples have missing values for air\_pressure\_9am?

- **3**
- 5
- 1092
- 0

4. Which column in the weather dataset has the most number of missing values?

- **rain\_accumulation\_9am**
- number
- They are all the same
- air\_temp\_9am

5. When we remove all the missing values from the dataset, the number of rows is 1064, yet the variable with most missing values has 1089 rows. Why did the number of rows decrease so much?

- **Because the missing values in each column are not necessarily in the same row**
- Because rows with missing values as well as rows with 0s are removed
- Because rows with missing values as well as rows with duplicate values are removed