Assignment 3: Data Foundations, Histograms & Tooltip **(20 points)**

Due: 15.05.2023 8AM

Task 1: Data Foundations **(5 Points)**

Task 1a) Data Types: Imagine you work for a fruit trade company, and your job is to create a categorization scheme for apples (e.g., to sell them to customers). Name **three descriptive attributes** of **nominal, ordinal, and numerical data type**, and possible value ranges.

An example for cars:
Numerical: Horsepower, 50-500 HP

Attributes for apples:

1. Nominal: Type of apple. E.g. Granny Smith, Honey Crisps or Gala

2. Ordinal: Size of the apple. E.g. Small, Medium, Big

3. Numeric: Weight of the apples. E.g. 50g-400g

Task 1b) Missing values – When using a **global constant to fill missing values** of an attribute, why should we not use this value for further calculations? Can you think of a siution where it might be useful to consider missing values?

Answer: Because it does not have anything to do with the real values. It is useful to analyze the behavior of missing values for example if they come from some systematic error or the data collection is correlated with a cofounder that also affects the variables we are interested in.

Task 1c) Binning – In which cases should we prefer equal-depth binning compared to equal-width binning?

Answer: When the distribution is uneven and/or there are outliers.

➔ **After putting in your answers, export the docx-File to PDF and upload it alongside the source code files.**

➔ **Continue with the programming tasks on the next page, you do not need to change this document any further**