**DATA CLEANING IN EXCEL**

# Removing data duplication

Duplicated data usually was created by error and in most cases should not be there. They are unnecessary waste of space and time, moreover they are increasing chances for potential problems and costly mistakes.

Graphical user interface, text, application

Description automatically generatedGraphical user interface, application

Description automatically generated

Graphical user interface, table, Excel

Description automatically generated

Formatting and standardization – all cells withing column will be displayed the same way.

Depending on requested format, words can be formatted to upper or lower cases. Otherwise standard format will be applied using formula PROPER()

# Table Description automatically generated

# Removing Spacing

I will remove unnecessary spaces, before and/or after the word. While spacing before word its clearly visible, spotting spacing at the end of the word is not that simple. You may think that as long as data looks ‘right’, additional spaces after the word should not matter.

In general that may be okay if you are going to use, and keep data in excel. But if you want to use data in SQL or Python (especially numeric), additional spaces may cause issues.

Therefore TRIMING is necessary.

FILTERING - Checking ‘Department’ column, are all fields correct.

I was using Filter option to check values entered into column. Founded misspelled ‘Marketing’ and corrected it. This correction is needed in case we would like to present specific calculations based on department divisions. If I would not correct it, program would understood it as another department.

Graphical user interface, application, table

Description automatically generated

Next, I noticed department ‘Whig’ and ‘Whig’ with some unusual signs after. Filtered them and removed words and signs after department’s name. It allows to bring two fields together as its originally meant to be.

# CURRENCY - Adjusting currency in salary table.

If data will be used in any programming language it is advisory to remove currency sign as program may read cells as a strings not integers/ floats. This will require later on further conversion if we would like to perform any calculations using those fields.

Table

Description automatically generatedTable

Description automatically generated

For easier usage in programs I removed .00 at the end of the salary.

# DATE FORMATING

Firstly I changed date format to ‘Short date’. This removed name of the day and month and turn it into numbers, moreover changed year into 4 digits instead of 2.

Some date was written with ‘/ ’ as separator, others with ‘.’ I replaced all dot-separators with slash.

Table

Description automatically generatedA picture containing table

Description automatically generated