**Week 8 workshop: Pandas Retail Store Analysis**

**DAB July 2025**

**BENABDI SIDALI**

**Wednesday October 1st, 2025:**

***Discussion Questions:***

1. What are the key advantages of using Pandas for data cleaning compared to other methods?

In my opinion the key advantages are that its easier once you get the basics and its way faster because you have less steps towards your goal

1. How would you approach to handling missing values differ if the missing data was not random but had a pattern or meaning?

If the missing data has a pattern or meaning and is not random that would make it easier because you don’t have to figure out a way to deal with the missing values you just have to follow the patterns

1. What types of data quality issues might not be immediately visible through simple DataFrame inspection methods?

The most obvious one for me would be the data types because if you do not look at the data directly and read it or try to understand it and you just run an info formula you would not know for example that this is date not a text or you would not know that even though some data might look like intg but its supposed to be handled as text for example unique identifiers

1. How would you document your data cleaning process to ensure reproducibility?

Commenting you thought process would be the first step then if I notice a reusable formula that would apply to a lot of dataframes I would archive it or save it my notion notebook or just create a cleaning notebook with all the formulas ( share it on stackoverlfow)

1. In what scenarios might it be better to remove rows with missing values rather than imputing them?

There are plenty of those , a few I can name would be if the percentage of missing value is minimal or it would not have an impact on the analysis , if there no patterns and it’s all random , also if there are too much missing values in that row to the point of that row being useless and finally if imputing the data would impact analysis