Supermarket Sales Analysis Using Python
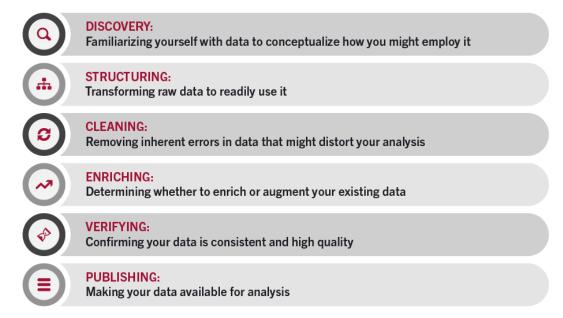

Data Wrangling Report

Prepared by: Mohammed Amr Abass

Objective: The goal of this analysis is to get business insights as shown in the business insights report.

This is done by the process of Data Wrangling.

The steps of data wrangling are explained according to Harvard Business school as:

DISCOVERY:
Familiarizing yourself with data to conceptualize how you might employ it

STRUCTURING:
Transforming raw data to readily use it

CLEANING:
Removing inherent errors in data that might distort your analysis

ENRICHING:
Determining whether to enrich or augment your existing data

VERIFYING:
Confirming your data is consistent and high quality

PUBLISHING:
Making your data available for analysis

- Discovery: After importing the proper libraries (Pandas – Matplotlib – Seaborn – NumPy), the data was uploaded as csv file and transformed to data frame (df). Using the .head() and .info() methods enables us to take a look inside the data frame and discover the problems.

- Structuring: The steps where we worked on the problems. The problems that arose were:
  – Branch – Date & Time formats – 'Total' Calculation

- Cleaning: After solving data issues, there were null values representing less than 0.6% of the data. The drop of this percentage of data is acceptable.

- Enriching: The data was enough to give informative insights

- Verifying: After dropping the null values, we left with 1000 records of non-null values with the proper data type to proceed with the following steps of visualizing the data

- Publishing: The process of visualizing the data that produces the charts in the business insights report, which helps in making data-driven decisions.