# UCD PA: Data Analytics for Business – Final Project

Charlotte McCarthy

10/02/2022

## GitHub URL

## https://github.com/CharlottePPB/UDCPA\_CharlotteMcCarthy

## Abstract

(Short overview of the entire project and features)

## Introduction

(Explain why you chose this project use case)

## Dataset

The dataset chosen for this project came from Kaggle. I chose this dataset because it was opensource and came from a real supermarket company. The one disadvantage of this dataset is the limited number of rows, just 1000. A full dataset capturing the same information would likely have millions of rows.

The dataset captures purchases across three branches of a large supermarket chain. It contains a mix of customer data, like which branch the customer visited and whether they’re a member of the supermarket loyalty scheme, and transaction data, like what the customer purchased and how much profit the supermarket made.

|  |  |
| --- | --- |
| Column | Description |
| Invoice ID | The auto generated invoice ID |
| Branch | The supermarket branch |
| City | The city where the branch is located |
| Customer Type | Is the customer a member of the loyalty scheme? |
| Gender | Gender of the customer |
| Product Line | Product line of item purchased |
| Unit Price | Unit price of item purchased |
| Quantity | Quantity of items purchased |
| Tax 5% | Total tax paid by customer |
| Total | Total amount paid by customer |
| Date | Date the purchase was made |
| Time | Time the purchase was made |
| Payment | Payment method used |
| COGS | Cost of the goods sold |
| Gross Margin % | Margin percentage for the supermarket |
| Gross Income | Gross profit for the supermarket |
| Rating | Customer satisfaction rating for their visit |

Figure 1: Description of all the columns in the dataset

The data has nine object columns, seven float columns and one integer column. The pandas dtypes function was used to easily view this information.

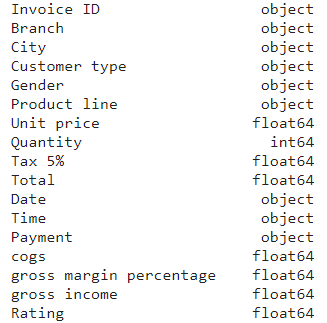
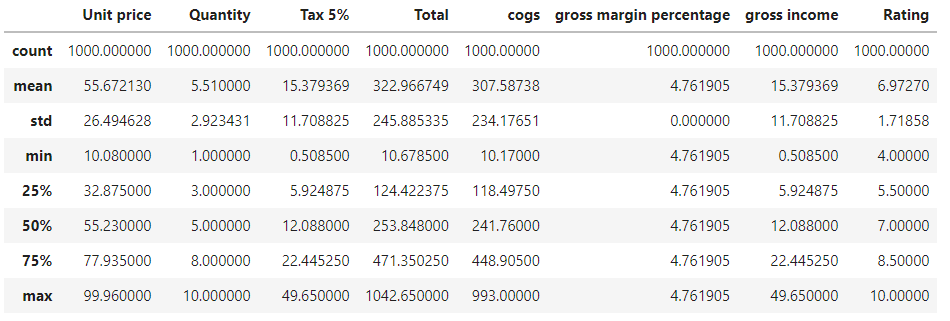


Figure 2: The data types for all columns in the dataset

The pandas describe function was used to view the summary statistics for all numeric columns within the dataset.



## Implementation Process

(Describe your entire process in detail)

## Results

(Include the charts and describe them)

## Insights

(Point out at least 5 insights in bullet points)