Purpose:

This is a project I undertook at a retail job to better understand the popularity of our items in Department 1 and Department 2 and a strategy to more efficiently manage inventory. I am creating this document as a means of demonstrating WHY and HOW I did certain things to achieve the result.

Note:

This data has been sanitized with respect to the types of products as a way to keep the company confidential. The sanitation was done by hand, by me, using mainly INDEX MATCH in Google Sheets.

Data set overview:

The data pulled is raw sales data from a point of sale system called ‘Shopkeep’. This data ranged from Jan 1, 2019 until April 30, 2021 with each row representing an individual item sold through the register.

Columns:

Time: DateTime value of when the purchase was made

Operation Type: Denotes whether the transaction was a SALE or RETURN

Store Code: Individual POS generated product number

Department: The department of which the item belongs

Cost: Inventory cost of the item representing Cost of Good

Price: Base price the item sells for

Quantity: Amount of that product sold for that transaction

Subtotal: The price x Quantity of the item before tax

Discount: Amount, represented as a positive number, discounted off the item

Net Total: Subtotal - Discount

Tax: Figure representing sales tax value for the line item

Total Due: Net Total + Sales Tax

Transaction ID: Auto Generated identifier for each transaction

Steps:

1. Sanitize data for anonymity as RAW\_DATA
2. Use Google Sheets to begin cleaning and normalizing data as CLEAN\_DATA
3. Use Jupyter Notebooks to finish any cleaning and manipulation of data and exploratory data analysis
4. Use Tableau to create a dashboard and visualizations
5. Final presentation using Google Slides

Files and File Types:

* RAW\_DATA.csv is a .csv and can be opened using any standard text editor or csv reading program
* CLEAN\_DATA\_FOR\_PYTHON comes in 2 file types, one a csv, and the other in excel spreadsheet(.xlsx) to keep comment integrity.
* SALES\_EDA.ipynb is a Google Colab jupyter notebook for final cleaning and exploratory data analysis of the sales information to massage it into a format ready for visualizations in Tableau.
* DEPT1\_FINAL & DEPT2\_FINAL are excel spreadsheets that I used to create dashboards and charts in Tableau to communicate to my stakeholder the performance of the products. In these I replaced the store\_code with a sanitized version of Brand and product identifiers like Brand1Prod1, Brand1Prod2, Brand2Prod1. The report was based on understanding the brands and what is popular in those brands.