**Phase 4 - Submission**

**TEAM NAME: Data Digits**

Sravani Ganipisetti - [sgani2@unh.newhaven.edu](mailto:sgani2@unh.newhaven.edu)

Divya Astalapuram – [dasta1@unh.newhaven.edu](mailto:dasta1@unh.newhaven.edu)

Bindu Priya Basa- [bbasa4@unh.newhaven.edu](mailto:bbasa4@unh.newhaven.edu)

**Please introduce your selected data set and research question.**

The dataset we are working with is “H&M sales 2018 data”. The dataset is taken from Kaggle website (<https://www.kaggle.com/datasets/tulasiram574/hm-sales-data> ). This dataset contains 15 attributes and 100 rows. This dataset is about different products purchased from H&M company in United States and about the profits.

* This dataset has information on products sold by H&M in the year of 2018.
* Each product is represented by a unique ID(Product ID) and Order ID number and includes information such as the Customer Id, Order Date, Sales, Quantity, Discount and Profit.
* And this dataset includes information about the Category and Sub-Category of each product and ordered from which City, State and Region.

**Please put a list of the exploration techniques, which you used in this work:**

Below are the data exploration techniques we used for this project.

* Data Pre-Processing

1. Filtering

* Data Cleaning
* Data Visualization

1. Heat Map
2. Bar Chart
3. Line Graph
4. Scatter Plot
5. Correlation Matrix

**Please describe your data explorations from different perspectives using varied visualization techniques such as tables and charts:**

Below are the tables and graphs that are generated by using visualization techniques:

A screenshot of a graph

Description automatically generated

* The above graph represents the correlation between Sales, Profit , Quantity and Discount.
* By using the Heat Map, we can see the correlation among Sales, Quantity, Discount with Profit.
* And we can see from graph Profits are highly correlated with sales.
* This graph is generated using Matplotlib functions.

A graph with blue dots

Description automatically generated

* The above graph represents the relation between Sales and Quantity.
* X-axis represents Quantity and Y-axis represents Sales.
* To visualize this I have plotted a scatter plot between the two. And we can see that when the quantity increase sales increasing.

A white background with black dots

Description automatically generated

* Above graph represents Distribution of sales from which range customer purchasing the products.
* This graph is generated using Box plot functions.

A graph of different colored columns

Description automatically generated

* Above graph represents relation between Sales and Sales by State.
* X-axis represents Sales by State and Y-axis represents Overall sales.
* This graph is generated using Bar plot functions.

A graph of different colored rectangles

Description automatically generated

* Above graph represents the relation between total sales and product wise categories.
* The graph is generated using Bar plot functions.
* X-axis represents Categories and Y-axis represents total count of that.
* From this graph we can get total sales of all products based on Categories.

A graph showing a bar chart

Description automatically generated with medium confidence

* This graph represents relation between Sales and Sales by Product Category.
* This graph is generated using Matplotlib functions.
* X-axis represents Product category and Y-axis represents Sales by product category.
* From this graph we get sales variations for particular category.

A graph of a bar chart

Description automatically generated with medium confidence

* This graph represents relation between Sales and Sales by Ship Mode.
* This graph is generated using Matplotlib functions.
* X-axis represents Shipping mode of the order and Y-axis represents Sales.
* From this graph we get shipping mode variations for sales.

A screenshot of a computer screen

Description automatically generated

* This graph represents relation of Sales, discount and Profit.
* This graph is generated using Pairplot functions.
* This is Correlation matrix for Sales, Discount and Profit fields. The matrix alone can show us different relations among the fields present.

**GITHUB\_REPO:** https://github.com/Sravani1698/Data-Digits.git