

Achievement 6 Project Information

Data Source

<https://www.kaggle.com/datasets/roopacalistus/superstore?resource=download&select=SampleSuperstore.csv>

The sales details of different stores of a supermarket chain that has multiple stores in different parts of the US.

Column Name	Description	Variable Type
Ship Mode	The Method used to Ship Products	Categorical
Segment	The market segment the customer belongs to	Categorical
Country	The country where the order was placed	Categorical
City	The city where the order was placed	Categorical
State	The state where the order was places	Categorical
Postal Code	The postal code of delivery location	Categorical
Region	The region where the order was placed	Categorical
Category	The category of the product	Categorical
Sub-Category	The sub-category of the product	Categorical
Sales	The sales revenue generated by the order	Continuous
Quantity	The number of units sold	Continuous
Discount	The discount applied to the order	Continuous
Profit	The profit generated from the order	Continuous

I chose this dataset as I am interested in going into retail analytics as a future career. With analyzing this data, I am able to understand more information on the retail industry and what insights are important in this industry.

Data Profile

Consistency checks done to ensure there are no duplicates, mixed-type data, missing values, or outliers that are unreasonable given the context of this data set. The following changes were made.

- Dropped the Ship Mode column since this column is not necessary to understand the data.
- Removed NA or Null cells to clean the data

Limitations

- The limitations in this data set are such as limited time range, since some of this data is from 2014 to 2017 making it out dated. Data duplication and inconsistencies are issues that could be in the data as well as missing values that could cause limitations.

Ethical Considerations

- Ethical use of this data set requires transparency about how the data is collected and what is its intended use. Bias based on the geography or demographic can also influence the results of the data. Ethically when using this data retail store should only use it to understand the need of the community and not use it to over price their customers on high demand products

Key Questions

- Patterns from customer purchasing behavior
- Key trend in sales and profit across different products
- Impact of discounts in overall sales
- What products are more profitable than others
- What region has the highest amount of sale per population
- Based off of this data what is the predictive analysis for future sales.