

Ryan Tomas

028210102

4/23/2025

## Uncovering Product and Customer Purchase Patterns in Amazon Sales Data

### **Goal:**

This project will explore the associative relationships between products, categories in Amazon sales data. The goal is to find which products or categories are frequently purchased together and to understand purchasing behavior across payment methods.

### **Features of interest:**

- Order ID; This is important to see what products were bought in the same order and the amount
- Data: this feature is data that can be used to see the frequent customers or find a relationship between items and days.
- Category: help filter through the items.

The features that the dataset has are the data on when the order was made and the quantity of the product ordered. Each product has a category that can be used to find patterns that relate to other products. This project will explore the associative relationships between products, categories in Amazon sales data.

### **Plan to handle Continuous Variables:**

The way I plan to handle continuous variables is to make a low, medium, and high for the prices and total sales. The changes that must be made to a feature are to dataOrder because it is not set to month or quarter to find seasonal trends.

**Proposed rule mining approach:**

I'm interested in this dataset because it relates to the in-class examples and seems like a real-world use case of the material we are learning. I'm planning to use FP-Growth for efficiency on transaction level data.

**Expected outcome:**

What I expected to get from this project was to identify the frequent co-purchases. Another is to find an association, like using a type of payment method, that relates to the chance of buying a tech product

**Why this matters:**

This matters because we can use the expected outcome to improve the recommendations algorithm of Amazon with items that have a relationship. Another way that can be used is to sell the data to a third-party company to use for ads.

I'm going to use a dataset from Kaggle that is called [Amazon Sales 2025](#), which was licensed by Apache and is expected to get updates annually. This dataset revealed around retail and shipping, which is similar to the subject we talked about in class. The features that the dataset has are the data on when the order was made and the quantity of the product ordered. Each product has a category that can be used to find patterns that relate to other products.