**Final Capstone Proposal**

**The Data:**

I found [this](https://www.kaggle.com/datasets/awesomeasingh/vending-machine-sales) dataset on Kaggle titled “V*ending Machine Sales*”, naturally it contains data related to vending machines. Specifically, this dataset features 5 vending machines in 4 locations in Central New Jersey. This date range for the data covers is from 01/01/2022 to 08/09/2022, with 6445 rows and 18 columns, it is clear to me that it is rich with insights.

Why this vending machine data?

I wanted to find some type of sales data to analyze for insights, being a frequent user of vending machines this was a perfect opportunity to do something unique and paint a bigger picture of the “backend” of these objects we pass and use in our daily lives.

**Here is the data at a glance:**

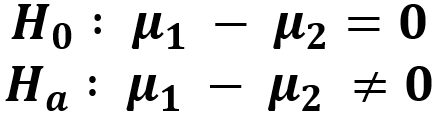
| **Variable** | **Description** | **Type** |
| --- | --- | --- |
| Status | Represents if the machine data is successfully processed | String (boolean) |
| Device ID | Unique ID for each vending machine | String (ID) |
| Location | Location of the machine | String (location reference) |
| Machine | Machine name | String (ID) |
| Product | Product vended from the machine | String (product) |
| Category | Carbonated / Food / Non-carbonated / Water | String (category) |
| Transaction | Unique identifier for every transaction | String (ID num) |
| TransDate | The Date & time of transaction | Date Time |
| Type | Type of transaction | String (boolean, cash or credit) |
| RCoil | Coil # used to vend the product | String (number) |
| RPrice | Price of the Product | Numeric (currency) |
| RQty | Quantity sold. This is usually one but machines can be configured to sell more items in a single transaction | Numeric |
| MCoil | Mapped coil # used to vend the product | String (number) |
| MPrice | Mapped price of the Product | Numeric (currency) |
| MQty | Mapped quantity sold | Numeric |
| LineTotal | Total sale per transaction | Numeric (currency) **continuous** |
| TransTotal | Represents total of all transactions that will show up on the Credit Card | Numeric (currency) **continuous** |
| Prcd Date | Date when the transaction was processed by SeedLive | Date |

**Research / Business Questions:**

There are two main questions I want to answer:

1. **What product category** **generates the most revenue, and which generate the least revenue?** 
   1. **What products do people want, and what do they not want?**
   2. I will use statistical tests such as t-tests and correlation to evaluate the different categories and other variables in this manner
   3. Carbonated / Food / Non-carbonated / Water
2. **What machine locations generate the most items purchased?** 
   1. **What is the most optimal location for a machine?**
   2. **Does the type of machine matter?**
   3. I will use statistical tests such as t-tests, correlation, and other methods to evaluate these metrics.

*These questions are not limited, and I may dive deeper into either question to get the most accurate and beneficial insights, and more research questions WILL come about as I explore insights and dive deeper.*



**Hypotheses:**

Research Question 1

* 1. Do carbonated drinks earn **more** revenue than non-carbonated drinks?
  2. Do carbonated drinks earn **more** revenue than water?
  3. Do carbonated drinks earn **more** revenue than food?

Research Question 2

* 1. Is there a **difference** between vending machine types and revenue?
  2. Is location **significant** when it comes to the frequency of purchases?
  3. Does the type of machine make a **difference** in the number of purchases?

**Using the Data:**

I will create bar charts and conduct t-tests for statistical significance. For all my tests I will be using a 95% confidence interval for consistency, and any p-value > .05 will be rejected. I will also create a time series line chart to show various trends over time between different categories and items. I will also group the data in several different ways to conduct tests among the groups for significance or correlation.

**Who will find my findings valuable:**

Specifically, anyone interested in or in the vending machine business who wants to make data-driven decisions about how to stock their machines. Also findings such as what to look for in a machine’s location, how different machines are fair in terms of revenue, and other related content.

More generally people are interested in consumer preferences on select beverages, snacks, and payment types.

This project may also offer findings that other researchers can use in fields such as social sciences, health, business, and psychology, among others.