**Project Proposal**: Unveiling White-Labeled IoT Devices on E-commerce Platforms(Amazon)

**Objective**:

The objective of this project is to investigate and identify white-labeled IoT devices, specifically smart plugs, available on popular e-commerce platforms such as Amazon and Walmart. By employing web scraping, data analysis, and clustering techniques, we aim to reveal products that may share a common original equipment manufacturer (OEM) or supplier. Once we cover smart plugs, the plan is to then try to do the same for Smart Cameras and Smart blood pressure monitors.

**Research Question**:

Which smart plugs (or other IoT devices) on Amazon or Walmart might be white-labeled, potentially originating from the same OEM?

**Methodology**:

* Web Scraping:
  + Conduct an extensive web scraping process to collect data on smart plugs from Amazon and Walmart listings.
* Data Cleaning and Formatting:
  + Clean and format the scraped data to ensure consistency and prepare it for subsequent analysis.
* Data Preparation / Pre-processing:
  + Normalize and preprocess the data, splitting it into relevant categories for further analysis.
* Basic NLP to Cluster Similar Text:
  + Utilize Natural Language Processing (NLP) to cluster similar product descriptions and texts.
* Image Analysis:
  + Apply image analysis techniques to cluster products with similar visuals or user interfaces.
* Repeat for Smart Cameras and Smart Blood Pressure Monitors:
  + Modify and apply the established methods to investigate white-labeled products in other IoT categories.
* Prepare Presentation for Class:
  + Create a comprehensive presentation summarizing the research, methods, and findings.
* Present in Class:
  + Deliver a clear and engaging presentation to the class, sharing insights and potential implications.

Skills Required:

* Web scraping
* Basic NLP (utilizing APIs if necessary)
* Image analysis (utilizing APIs if necessary)
* Data analysis and interpretation
* Presentation skills

**Timeline**:

* Step 1: Web Scrape – November 16 - November 20
* Step 2: Data Cleaning and Formatting – November 21 - November 24
* Step 3: Data Preparation / Pre-processing – November 25 - November 28
* Step 4: Basic NLP to Cluster Similar Text – November 29 - December 2
* Step 5: Ability to Do Simple Image Analysis – December 3 - December 6
* Step 6: Repeat for Smart Cameras and Smart Blood Pressure Monitors – December 7 - December 10
* Step 7: Prepare Presentation for Class – December 11 - December 12
* Step 8: Present in Class – December 14

**Expected Output**:

Clusters of likely white-labeled smart plugs and other IoT devices, indicating potential shared OEMs, along with a presentation summarizing the findings.

**Significance**:

This project aims to shed light on the prevalence of white-labeled products in the IoT market, raising awareness among consumers and encouraging transparency in product labeling.

**Team Members**: Priyanka Bose (pb2703), Moin Khan (mk8793), Chandra Shekhar Pandey (cp3793), Priyanshi Singh (ps4609)