eCommerce Dataset

This repository contains code and data for analyzing sales data. Below is a detailed summary of the analysis conducted.

Overview

This project aims to analyze sales data to gain insights into top-selling products, trends in sales over time, and comparisons between different periods.

Data Cleaning

The dataset (`data.csv`) underwent thorough cleaning to ensure data quality and accuracy. The cleaning process involved:

- Removal of null values: Rows containing null values in the `price`, `quantity`, or `description` columns were removed to maintain data integrity.
- Data type conversion: Appropriate data types were applied to each column for consistency and ease of analysis.

Top Selling Products

To identify the top-selling products, the `sales_analysis.py` script was utilized. The script employs Python's pandas library to:

- Calculate total sales for each product.
- Group the products based on sales volume.
- Sort the products in descending order to identify the top 10 sellers.

The results are displayed in the console or saved to an output file for further analysis.

Trend Analysis

The trend analysis focuses on the sales performance of the "WORLD WAR 2 GLIDERS ASSTD Designs" product over time. The following steps were performed:

- Filtering the dataset to include only entries related to "WORLD WAR 2 GLIDERS ASSTD Designs."
- Aggregating sales data by year to observe trends over time.
- Visualizing the trends using line plots or bar charts to highlight any patterns or fluctuations in sales volume.

Comparison of Sales Between November and March

A comparative analysis was conducted to examine the difference in sales between November and March. This analysis involved:

- Filtering the dataset to include only entries from November and March.
- Calculating the total sales for each month.
- Performing statistical tests (e.g., t-test) to determine if there is a significant difference in sales volume between the two months.

Files

- `eCommerceData.csv`: The cleaned dataset used for analysis.
- `eCommerce Analysis.ipynb`: Python script for conducting the sales analysis.
- `README.md`: This file providing a detailed overview of the analysis and repository contents.

Requirements

- Python 3
- pandas
- numpy

Author

Emmanuel Kpandeyenge