Retail Sales Performance Analysis and Reporting

Project Overview

This project involves a comprehensive analysis of over 10,000 retail sales records to identify key business drivers, seasonal trends, and high-value customer segments. The goal was to provide data-driven recommendations to optimize inventory management, marketing budget allocation, and customer retention strategies.

The analysis follows a standard data science pipeline: Data Generation, Cleaning and Preprocessing, Exploratory Data Analysis (EDA), Customer Segmentation (Pareto Analysis), and Reporting.

Key Findings

1. **Profitability:** The **'Technology'** category was confirmed as the most profitable segment, informing strategic inventory and investment decisions.
2. **Seasonality:** Sales consistently **peaked in November and December**, highlighting the importance of Q4 promotional planning.
3. **Customer Value:** A Pareto analysis identified that the **top 20% of customers contributed over 60%** of total revenue, justifying a focus on targeted retention strategies.

Files in this Repository

File Name	Description
`sales_analysis.py`	The main Python script containing the entire analysis pipeline: data generation, cleaning, trend identification, segmentation, and visualization logic (using Pandas, NumPy, Matplotlib, and Seaborn).
`analysis_report.md`	An executive summary report detailing the findings, metrics, and actionable recommendations derived from the analysis.
`sales_analysis_plots.png`	(Generated upon running `sales_analysis.py`) Visualization of key metrics, including Profit by Category and Monthly Sales Trends.

How to Run the Analysis

Prerequisites

You need Python 3 installed on your system, along with the following libraries:

```
pip install pandas numpy matplotlib seaborn
```

Execution Steps

1. **Clone the repository:**

```
git clone [YOUR_REPO_URL]
cd retail-sales-analysis
```

2. **Run the script:**

```
python sales_analysis.py
```

The script will print the analysis findings to the console and generate the `sales_analysis_plots.png` file in the same directory.

Data Schema

The synthetic dataset used for this analysis includes the following fields:

Field	Data Type	Description
`OrderID`	String	Unique identifier for each transaction.
`Customer ID`	String	Unique identifier for each customer.
`Category`	String	Product category (Technology, Furniture, Office Supplies).
`Sales`	Float	Revenue generated by the sale.
`Profit`	Float	Profit margin achieved on the sale.
`OrderDate`	Datetime	Date of the transaction.

Recommendations

The primary data-driven recommendations include:

- **Inventory:** Increase inventory allocation for the 'Technology' category.
- **Marketing:** Focus major advertising campaigns on the November-December peak period.
- **Retention:** Launch an exclusive loyalty program targeting the top 20% of high-value customers.