AMOD-5430: Data Visualization

PROJECT PROPOSAL ON

Python Driven EDA & Data Visualization in Retail

NOVEMBER 8, 2023

MUHAMMAD RIAD HOSSAIN
STUDENT ID: 0793364

MUHAMMAD SAJID SALMAN
STUDENT ID: 0774556

1. Abstract

The retail sector generates vast amounts of data that are often underutilized. This project seeks to delve into retail sales data to extract meaningful insights through meticulous Exploratory Data Analysis (EDA) and the employment of advanced visualization techniques with Python. The exploration will revolve around dissecting product performance, customer buying behaviors, and inventory efficiencies, utilizing a robust dataset sourced from Kaggle. We will leverage Python libraries such as Pandas for data wrangling, Matplotlib and Seaborn for visualization. Our anticipated contribution is a set of detailed visual analyses that illuminate hidden patterns and trends, equipping retail decision-makers with actionable intelligence to refine sales tactics.

2. Problem

The retail industry is a complex and dynamic environment where consumer preferences and market trends change rapidly. One of the significant challenges within this sector is the effective use of data for informed decision-making. Our project addresses the problem of untapped potential in retail data, focusing on uncovering actionable insights that can lead to improved sales strategies.

3. Intended Contribution

Our contribution will be a detailed exploratory data analysis of retail data, which will provide visual insights into sales trends and product performance. The EDA phase will be involved with a series of analyses and data visualization techniques to gain a deeper understanding of retail performance. We are expecting to conduct the following exploratory data analyses:

- **Country-Wise Sales & Profit Analysis**
  Identify sales, profits trends across different countries. Visualizations such as bar charts and plots will be used to showcase the distribution of sales among countries.

- **Segment-Wise Contribution to Sales & Profit**
  This analysis will be focused on understanding the contributions of various customer segments to overall sales and profit.

- **Year Wise Analysis of Sales & Profit**
  This analysis will investigate sales and profit trends over different years.

- **Regression Analysis between Discount & Sales**
  A regression analysis will be performed to examine the relationship between discounts and sales. This will help understand how changes in discounts affect sales.

- **ANOVA Test**
  An analysis of variance (ANOVA) test will be conducted to determine if there were significant differences in sales across different product categories, segment, region.

- **Chi-Square Independence Test**
  A Chi-square test of independence will be performed to assess whether the choice of ship mode was associated with the region.

4. Methodology

The methodology detailed below will provide a systematic approach to exploring retail performance using EDA and statistical testing. By implementing these methods, valuable insights and evidence-based conclusions will be drawn from the dataset.

- **Data Set**: To explore retail performance through Exploratory Data Analysis (EDA) and statistical testing, we will use a dataset of retail industry from Kaggle. Before conducting any analyses, certain preprocessing tasks will be performed to ensure data quality and consistency.
- **Programming Language**: Python
- **Environments/Tools**: Jupyter Notebook for interactive development.
- **Libraries**: Pandas, Numpy, Matplotlib, Seaborn, Scikit-learn, SciPy, Statsmodels
- **Computational Resources**: Personal computers with adequate processing power and memory.
- **EDA Structure**: We will perform univariate analysis to understand single-variable distributions, followed by bivariate and multivariate analysis to uncover relationships and patterns.

5. Expected Results

We anticipate revealing insightful trends and correlations within the retail data that could influence decision-making. While our results may deviate from initial expectations, the analysis will provide a solid foundation for understanding retail data characteristics.

6. Team Responsibilities

#	Team Member	Responsibilities
1	Muhammad Riad Hossain ID: 0793364	Background research, defining objectives, writing literature review, data preprocessing & cleaning, coding, exploratory data analysis, data visualization and peer review.
2	Muhammad Sajid Salman ID: 0774556	Writing methodology, coding, exploratory data analysis, visualization, analyzing graphs, interpretating results, writing conclusion and peer review.

7. Key References

- DeHoratius, N., Musalem, A., & Rooderkerk, R. (2023, February 27). Why retailers fail to adopt advanced data analytics. Harvard Business Review. https://hbr.org/2023/02/why-retailers-fail-to-adopt-advanced-data-analytics.
- Agrawal, R. (2022, August 31). Exploratory data analysis using data visualization techniques! Analytics Vidhya. https://www.analyticsvidhya.com/blog/2021/06/exploratory-data-analysis-using-data-visualization-techniques.
- Jeswani, R. (2021, December). Predicting Walmart sales, exploratory data analysis, and Walmart sales dashboard. B. Thomas Golisano College of Computing and Information Sciences, Rochester Institute of Technology. https://www.rit.edu/ischoolprojects/sites/rit.edu.ischoolprojects/files/document_library/Rashmi_Jeswani_Capstone.pdf