**AMOD-5430: Data Visualization**

PROJECT PROPOSAL ON

Python Driven EDA & Data Visualization in Retail

JUNE 17, 2023

# SABRINA FARZANA

STUDENT ID: 0792500

**ABDUL KAMAL AZAD**

STUDENT ID: 0774556

## Abstract

The healthcare sector generates vast amounts of data that are often underutilized. This project seeks to delve into healthcare 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 patient information, medical conditions, and healthcare services, utilizing a synthetic 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 healthcare decision-makers with actionable intelligence to refine healthcare strategies.

## Problem

The healthcare industry is a complex and dynamic environment where patient needs and medical 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 healthcare data, focusing on uncovering actionable insights that can lead to improved healthcare strategies.

.

## Intended Contribution

Our contribution will be a detailed exploratory data analysis of healthcare data, which will provide visual insights into patient demographics, medical conditions, and healthcare services. The EDA phase will be involved with a series of analyses and data visualization techniques to gain a deeper understanding of healthcare performance. We are expecting to conduct the following exploratory data analyses:

- Patient Demographics Analysis

- Medical Conditions Analysis

- Healthcare Services Analysis

## - Test Results Analysis

differences in sales across different product categories, segment, region.

**Patient Demographics Analysis:**

**Goals**: Understand patient demographic distribution and identify trends among different groups.

**Steps**: Collect, clean, and analyze data on age, gender, ethnicity, and location. Perform descriptive statistics and visualize distributions with histograms, pie charts, and bar graphs. Use correlation coefficients to explore relationships between demographics.

**Medical Conditions Analysis**

**Goals**: Identify the prevalence of medical conditions and their correlation with demographics.

**Steps**: Collect data on medical conditions and demographics. Calculate prevalence rates and identify trends over time. Use cross-tabulation to analyze the relationship between conditions and demographics. Visualize findings with heatmaps and bar charts.

**Healthcare Services Analysis**

**Goals**: Evaluate healthcare service utilization and identify usage patterns.

**Steps**: Gather data on healthcare services, calculate utilization rates, and analyze distribution across demographics and conditions. Assess service costs and visualize usage patterns with bar charts, line graphs, and pie charts.

**Test Results Analysis**

**Goals**: Analyze test results for patterns and outliers, and correlate them with demographics and conditions.

**Steps**: Collect data on test results, define normal ranges, and calculate descriptive statistics. Identify outliers and analyze correlations. Visualize results using box plots, scatter plots, and histograms.

**Differences in Sales Across Different Product Categories, Segments, and Regions**

**Goals**: Understand sales performance across categories, segments, and regions, and identify driving factors.

**Steps**: Collect sales data by category, segment, and region. Calculate sales metrics and identify trends. Analyze sales performance by segment and region. Visualize differences with bar charts, line graphs, and heatmaps.

## 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 datasets

**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.

## 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.

## Team Responsibilities

|  |  |  |
| --- | --- | --- |
| **#** | **Team Member** | **Responsibilities** |
| 1 | Sabrina Farzana  ID: 0792500 | Background research, defining objectives, writing literature review, data preprocessing & cleaning, coding, exploratory data analysis, data visualization and peer review. |
| 2 | Abdul Kamal azad  ID: 0774556 | Writing methodology, coding, exploratory data analysis, visualization, analyzing graphs, interpretating results, writing  conclusion and peer review. |

1. **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-](https://hbr.org/2023/02/why-retailers-fail-to-adopt-advanced-data-analytics) [retailers-fail-to-adopt-advanced-data-analytics](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-](https://www.analyticsvidhya.com/blog/2021/06/exploratory-data-analysis-using-data-visualization-techniques) [data-analysis-using-data-visualization-techniques](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/Ra](https://www.rit.edu/ischoolprojects/sites/rit.edu.ischoolprojects/files/document_library/Rashmi_Jeswani_Capstone.pdf) [shmi\_Jeswani\_Capstone.pdf](https://www.rit.edu/ischoolprojects/sites/rit.edu.ischoolprojects/files/document_library/Rashmi_Jeswani_Capstone.pdf)