FINAL PROJECT PROPOSAL – SAT4650

Title

Analysis and Visualization of a Sales Dataset

Basic Idea

The basic idea of this project is to conduct exploratory data analysis (EDA) and create interactive visualizations for a sales dataset. This project will leverage various Python programming concepts, including data types, variables, functions, file operations, error handling, object-oriented programming, GUI, and data analysis and visualization.

Problem To Be Solved

The main problem to be solved with this project is to derive actionable insights from a sales dataset through Exploratory Data Analysis (EDA) and interactive visualization. This involves collecting a sales dataset, cleaning, and organizing the data, and employing statistical and visual exploration techniques to uncover patterns, trends, and correlations. The project aims to address challenges related to data consistency, missing values, and overall data quality. Ultimately, the goal is to provide a comprehensive understanding of the sales data, facilitating informed decision-making for business strategies and highlighting areas for potential improvement. These techniques can also be tailored to any other kind of dataset such as data from sensors.

Implementation Plan

My implementation will make use of the following packages/modules: *csv, pandas, numpy, matplotlib, seaborn, plotly, tkinter,* and *Pillow*.

Pandas will be employed to read the CSV file containing the dataset and create a structured DataFrame. Utilizing Python's data preparation capabilities, we will handle missing values and ensure data consistency.

For the Exploratory Data Analysis (EDA), statistical analysis will be conducted using Pandas, while key Python libraries such as Matplotlib and Seaborn will facilitate static visualizations. Trends over time will be explored through line charts and time series plots. For deeper insights, correlations between variables will be identified. The Data Visualization phase will include dynamic and interactive visualizations implemented with Plotly, offering a more engaging user experience.

The project will adhere to Object-Oriented Programming (OOP) principles, implementing a structured code with classes for data handling and visualization. This approach enhances code organization and maintainability. Finally, a Graphical User Interface (GUI) will be developed using Tkinter, allowing users to interactively select and customize visualizations based on different aspects of the sales data.