Assignment 1:- Shirish Patel

|  |
| --- |
| **1. Data Loading and Inspection** |
| **- Load the sales, customer, and product datasets into Pandas DataFrames** |
| df = pd.read\_csv(r'C:\Users\shiripatel\Desktop\PBA\_Assignment\sales\_data\_sample.csv', encoding='latin1')  *Note: - Used latin1 as it was giving encoding error while reading the file* |
|  |
| **- Inspect the data for missing values, inconsistencies, and outliers.** |
| df.isnull().sum() |
| Null values |
| Address line 2, State and Territory have more than 50% of values as NaN |
| **2. Object-Oriented Programming (OOP) Implementation** |
| - Define classes for Customer, Product, and SalesTransaction  - Implement constructors and destructors for these classes.  - Demonstrate inheritance by creating a subclass for a specific type of customer (e.g., VIPCustomer) that inherits from the Customer class. |
| With Delete/ Destructor function    - Demonstrate inheritance by creating a subclass for a specific type of customer (e.g., VIPCustomer) that inherits from the Customer class. |
| **3. Data Cleaning and Transformation** |
| - Handle missing values and correct data inconsistencies.  - Transform data types where necessary (e.g., date formatting).  - Merge the datasets to create a unified DataFrame for analysis. |
| There is no missing value; however to clean data and create class, we have converted “NaN” data into default value |
|  |
| Merge the datasets to create a unified DataFrame for analysis |
| **4. Python Modules and Packages**  - Create a custom Python module for data cleaning functions.  - Utilize built-in modules like os, time, and sys to manage files and system operations.  - Fetch additional data from a REST API using the requests module (e.g., exchange rates for currency conversion).  Can Skip for now as discussed with professor |
| **File Handling**  - Write functions to read, append, and handle files in Python.  - Save cleaned data to new CSV files.  Read file.    Clean file – drop column wit missing details    Save updated file |
| **6. Web Scraping**  - Use BeautifulSoup or Scrapy to scrape additional data (e.g., product reviews) from a relevant e-commerce website.  Can Skip for now as discussed with professor |
| **7. Data Analysis with NumPy and Pandas**  - Perform numerical operations and linear algebra using NumPy.  - Manipulate and analyze data using Pandas (e.g., filtering, grouping, and aggregating data).  - Load data from various formats (CSV, Excel, JSON) and export analysis results to different formats. |
| Country wise sale – grouping    Aggregate function – sales summary product wise |
| **8. Data Visualization**  - Create basic plots (line plot, scatter plot, pie chart, bar plot) using Matplotlib.  - Enhance visualizations with annotations and customizations.  - Create interactive plots using Plotly (e.g., histograms, bar plots).  - Generate advanced visualizations with Seaborn (e.g., countplot, barplot, scatterplot). |
| Line Plot |
| **9. Geospatial Analysis**  - Utilize latitude and longitude data for creating maps and charts.  - Plot customer distribution and sales hotspots on a map. |
|  |
| **10. Programming Best Practices**  - Follow PEP standards for code style and documentation.  - Write efficient, readable, and reusable code.  - Implement error handling and debugging strategies.  Can Skip for now as discussed with professor |
| **11. Problem-solving and Debugging with ChatGPT**  - Use ChatGPT to assist in solving coding problems and debugging your scripts.  Using ChatGPT to identify error.    Used GPT to find solution to add trend line in scatterplot. |
| **12. Final Report and Presentation**  - Summarize your findings and insights in a comprehensive report.  - Create a presentation with key visualizations and insights for stakeholders.  Attached separate PPT |