WiseAnalytics

Senior Data Scientist Hands-On Assessment Assignment

Objective:

The goal of this assignment is to evaluate your ability to perform Customer Segmentation using data science techniques, including data exploration, feature engineering, model development, and deployment. You will demonstrate expertise in machine learning, statistics, cloud platforms, and effective communication of insights.

Data Source: Online Retail Dataset (https://archive.ics.uci.edu/ml/datasets/Online+Retail)

Part 1: Data Exploration & Preprocessing (Day 1)

- **Task 1: Exploratory Data Analysis (EDA)** – Perform initial data analysis to identify patterns, correlations, and potential data quality issues.

- **Task 2: Data Cleaning & Feature Engineering** – Preprocess the dataset, handle missing values, and create meaningful features for model training.

Part 2: Machine Learning Model Development (Day 2)

- **Task 3: Model Selection & Training** – Develop and compare multiple machine learning models to solve a predictive problem.

- **Task 4: Model Evaluation & Interpretation** – Assess model performance, optimize hyperparameters, and explain key insights from the model.

Part 3: Deployment & Business Impact Analysis (Day 3)

- **Task 5: Model Deployment** – Deploy the trained model as an API or a simple application for real-world usability.

- **Task 6: Business Insights & Recommendations** – Present actionable insights derived from the data and model for business decision-making.

Scalability & Optimization

- Discuss methods to scale the solution for larger datasets and improve model efficiency.

- Propose ways to monitor and retrain the model over time.

Additional Questions on RFM Customer Segmentation

- **Task 7: RFM Analysis Implementation** – Perform RFM (Recency, Frequency, Monetary) segmentation using the Online Retail Dataset (https://archive.ics.uci.edu/ml/datasets/Online+Retail). Calculate RFM scores for each customer and segment them into different categories (e.g., VIP, Frequent, Lost Customers).

- **Task 8: Customer Segmentation Insights** – Analyze the RFM segments and interpret business implications. Identify key customer groups that can be targeted for marketing campaigns.

- **Task 9: Visualization & Reporting** – Create visual representations of the RFM segments (e.g., scatter plots, bar charts) to support insights. Provide a structured report explaining the segmentation results and recommendations for customer engagement strategies.

Submission Requirements:

- **Codebase:** Submit all scripts, notebooks, and configuration files in a GitHub repository.

- **Documentation:** Provide a structured report detailing the problem, methodology, results, and business recommendations.

- **Deployment Instructions:** Explain how to deploy and test the API.

- **Dashboard Link:** If applicable, share a link or screenshots of the visualization dashboard.

Evaluation Criteria:

- **Technical Proficiency** – Code quality, efficiency, and correctness of data processing.

- **Machine Learning Expertise** – Appropriate model selection, feature engineering, and evaluation.

- **Problem-Solving Ability** – Creativity in handling data challenges and designing solutions.

- **Business Understanding** – Ability to translate data insights into strategic recommendations.

- **Communication & Documentation** – Clarity and completeness of documentation and presentations.