Certified Data Analysts

Capstone Project

Customer Retention and Sales Optimization in Retail

Part 2 - Data Science, R Programming & BI Dashboard

1. **Python (Jupyter Notebooks or Script)**
    - Data cleaning and transformation using pandas.
    - Perform RFM (Recency, Frequency, Monetary) analysis for customer segmentation.
    - Train a machine learning model (e.g., logistic regression or random forest) to predict churn.
    - Visualize feature importance using matplotlib or seaborn.

2. **R Language**
    - Perform statistical analysis:
        i. Chi-squared test to analyze relationship between gender and product category preference.
        ii. ANOVA to compare average spend across different regions.
    - Use ggplot2 for advanced visualizations.
    - Apply clustering (K-means) for customer segments based on demographic and transaction data.

3. **Tableau / Power BI**
    - Connect to your cleaned dataset (CSV or SQL).
    - Create dashboards showing:
        i. Sales trends over time
        ii. Customer retention funnel
        iii. Heatmap of product sales by location
        iv. KPI indicators (Average order value, Customer LTV, Monthly Active Users)

Project Objectives:

1. Perform data cleaning and transformation using Python (pandas) to prepare retail datasets for analysis.
2. Execute customer segmentation using RFM analysis and K-means clustering techniques.
3. Develop and evaluate predictive models (e.g., logistic regression or random forest) to forecast customer churn.
4. Create compelling data visualizations and dashboards using ggplot2, Tableau, or Power BI to communicate key metrics and insights to stakeholders.

The submission requirements:

1. **Report:**
   - **Platform:** Google Drive
   - **Format:** word and PDF
   - **Page Limit:** Maximum 10 pages
   - **Final Deliverables:**
     - Excel analysis workbook (.xlsx)
     - SQL schema and query scripts (.sql)
     - Python analysis notebook (.ipynb)
     - R analysis script (.R)
     - Tableau or Power BI dashboard file
     - Capstone Report (PDF or DOCX) summarizing insights and recommendations