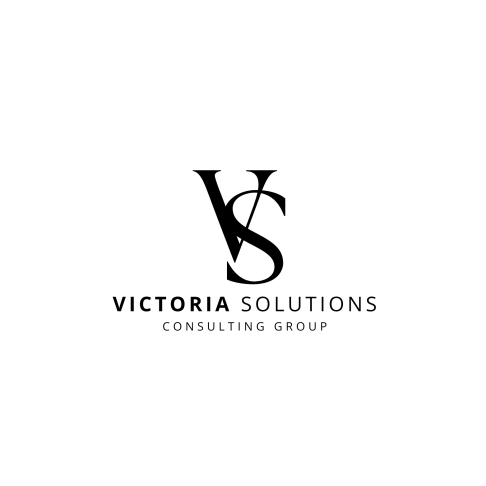
|  |  |
| --- | --- |
| Name | Asad |
| Contact Number | 07443892214 |
| Project Title (Example – Week1, Week2, Week3) | Week 3 project |



**Project Guidelines and Rules**

1. **Formatting and Submission**
   * **Format:** Use a readable font (e.g., Arial/Times New Roman), size 12, 1.5 line spacing.
   * **Title:** Include Week and Title (Example - Week 1: TravelEase Case Study.)
   * **File Format:** Submit as PDF or Word file to contact@victoriasolutions.co.uk
   * **Page Limit:** 4–5 pages, including the title and references.
2. **Answer Requirements**
   * **Word Count:** Each answer should be 100–150 words; total 800–1,200 words.
   * **Clarity:** Write concise, structured answers with key points.
   * **Tone:** Use formal, professional language.
3. **Content Rules**
   * Answer all questions thoroughly, referencing case study concepts.
   * Use examples where possible (e.g., risk assessment techniques).
   * Break complex answers into bullet points or lists.
4. **Plagiarism Policy**
   * Submit original work; no copy-pasting.
   * Cite external material in a consistent format (e.g., APA, MLA).
5. **Evaluation Criteria**
   * **Understanding:** Clear grasp of business analysis principles.
   * **Application:** Effective use of concepts like cost-benefit analysis and Agile/Waterfall.
   * **Clarity:** Logical, well-structured responses.
   * **Creativity:** Innovative problem-solving and examples.
   * **Completeness:** Answer all questions within the word limit.
6. **Deadlines and Late Submissions**
   * **Deadline:** Submit on time; trainees who submit fail to submit the project will miss the “Certificate of Excellence”

1. **Additional Resources**
   * Refer to lecture notes and recommended readings.
   * Contact the instructor or peers for clarifications before the deadline.

**START YOUR PROJECT FROM HERE:**

**Week 3 project**

**1. Introduction**

This project involves analysing a retail sales dataset containing customer information, spending behaviour, and churn status. The dataset includes variables such as *Region, Total Spend, Purchase Frequency, Marketing Spend, Seasonality Index,* and *Churned*. The primary objective is to apply advanced data analysis techniques in Python to uncover customer behaviour patterns, build predictive models for churn and spending, and perform statistical analysis. By combining classification, regression, and clustering methods, the analysis aims to support data-driven decision-making, improve marketing strategies, and enhance customer retention efforts. The outcome will provide actionable insights to guide business strategy.

**2. Data Preparation & Cleaning**

The dataset was first inspected in Excel to verify structure and formatting. A duplicates check was performed, confirming no repeated records. Column headers were reviewed for consistency, and data types were validated to ensure numerical fields contained only numbers and categorical fields were consistent.

In Python, preprocessing involved encoding categorical variables using Label Encoder (e.g., *Region* and *Churned*) to prepare them for modelling. Numerical features were scaled using StandardScaler to standardise ranges and improve model performance. These steps ensured the dataset was clean, well-structured, and ready for predictive modelling and statistical analysis

**3. Predictive Modelling**  
Logistic Regression and Random Forest were applied for classification to predict customer churn. Logistic Regression provides a statistical approach, modelling the probability of churn based on features like spending, frequency, and seasonality. Random Forest, an ensemble of decision trees, enhances classification accuracy by capturing complex, non-linear relationships and ranking feature importance. For regression tasks, Linear Regression was used to predict Total Spend, offering a straightforward interpretation of how independent variables affect spending. This model assumes linear relationships and was evaluated using R² and error metrics. Combining these models ensures robust predictive capability for both classification and continuous outcome prediction.

**4. Statistical Analysis**  
The classification models were evaluated using accuracy, precision, recall, and F1-score. Random Forest achieved high accuracy, with Seasonality Index and Total Spend emerging as top predictors of churn. Regression performance was assessed using R², Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). The Linear Regression model showed a strong R² value, indicating a good fit between predicted and actual spending. K-Means clustering revealed three distinct customer segments with clear differences in spending, purchase frequency, and marketing responsiveness. Summary statistics from clusters provided actionable insights into segment behaviour, enabling targeted marketing strategies to improve retention and increase overall revenue.

**5. Clustering Insights**  
The K-Means clustering identified three distinct customer segments based on Total Spend and Purchase Frequency.  
**Segment 0** represents low-value customers with low spend and purchase frequency, requiring targeted promotions to encourage engagement.  
**Segment 1** consists of high-value customers with the highest spend and frequency, making them ideal for loyalty programs and retention strategies.  
**Segment 2** includes mid-tier customers with moderate spending and engagement, offering potential for upselling and cross-selling. The clear separation between clusters indicates well-defined behavioural differences, enabling the business to tailor marketing campaigns and resource allocation for each segment, ultimately improving customer retention and revenue growth.

**6. Visualisations**

A red and blue grid with white text

AI-generated content may be incorrect.

A graph of blue rectangular bars with white text

AI-generated content may be incorrect.

A chart with numbers and dots

AI-generated content may be incorrect.

A graph with a red line and blue dots

AI-generated content may be incorrect.

1. **Business Insights & Recommendations**  
   Analysis shows that high-spend, high-frequency customers (Segment 1) are the most valuable and least likely to churn, while low-spend customers (Segment 0) are at higher risk.
2. **VIP Retention:** Implement loyalty rewards and exclusive offers for Segment 1 to maintain engagement.
3. **Reactivation Campaigns:** Use targeted promotions and personalized marketing to encourage purchases from Segment 0.
4. **Upselling Core Customers:** For Segment 2, introduce bundle offers or subscription plans to increase purchase frequency.
5. **Seasonal Targeting:** Leverage high Seasonality Index periods for campaigns, focusing on products or regions with strong seasonal demand.