1. Team Member's Details:

- **Group Name:** Data Innovators

- **Name 1:** Stephanie Dawsonn-Andoh
  - o **Email:** stephandoh@gmail.com
  - o **Country:** Ghana
  - o **Specialization:** Data Analytics

- **Name 2:** Sreedhar Rongala
  - o **Email:** rongalasreedhar@gmail.com
  - o **Country:** India
  - o **Specialization:** Data Analytics

- **Name 3:** Brittney Smith
  - o **Email:** brittneysintership@gmail.com
  - o **Country:** USA
  - o **Specialization:** Data Analytics

---

2. Problem Description:

XYZ Bank is aiming to enhance its marketing campaign by delivering personalized Christmas offers to its customers. Instead of rolling out a generic offer for everyone, they want to target specific customer segments with relevant offers. To solve this problem efficiently, XYZ Bank approached ABC Analytics Company to help with customer segmentation. The bank's requirement is to group customers into no more than 5 segments, as more segments would be inefficient for their marketing efforts.

---

3. Business Understanding:

The key business goal is to implement personalized marketing to increase customer engagement and improve conversion rates during the Christmas campaign. By understanding customer behavior, demographics, and preferences, the bank can create relevant offers for each group. The segmentation should also uncover hidden patterns in customer data that are not easily noticeable through manual analysis. The results will help XYZ Bank to tailor their products and services to better suit the needs of each group, resulting in higher customer satisfaction and loyalty.

Key Points:

- Increase customer engagement with tailored offers.
- Reduce inefficiencies by limiting segments to 5 groups.
- Enhance customer satisfaction and retention.

---

4. Project Lifecycle Along With Deadlines:

Phase	Task	Deadline
Week 7 (Sep 19)	Submit project plan and business understanding. Prepare data intake report.	19th September 2024
Week 8 (Sep 26)	Data Understanding: Analyze data structure, missing values, and outliers. Report data issues and strategies for handling them.	26th September 2024
Week 9 (Oct 3)	Data Cleansing and Transformation: Handle missing values and outliers using multiple techniques. Ensure data is ready for modeling.	3rd October 2024
Week 10 (Oct 10)	EDA: Perform detailed analysis of the dataset. Identify key insights and trends. Provide a final recommendation for business users.	10th October 2024
Week 11 (Oct 17)	EDA Presentation: Prepare slides for business users and provide a technical overview for model recommendations.	17th October 2024
Week 12 (Oct 24)	Model Selection and Building: Select the best base model for segmentation and experiment with various model types (e.g., K-means, Ensemble models). Develop a dashboard for data visualization.	24th October 2024
Final Submission (Oct 31)	Submit final report, code, and presentation, including results of all team member contributions and the chosen model solution.	31st October 2024

5. Data Intake Report:

- Name: Customer Segmentation Group Project

- Report date: 9/19/2024

- Internship Batch: LISUM36

- Version: 1.0

- Data intake by:

  o Stephanie Dawsonn-Andoh

- Data intake reviewer:

  o Sreedhar Rongala

- o Brittney Smith

- Presentation and Reporting
    - ■ Brittney Smith

- Data storage location:

- https://github.com/BrittneyMSmith/Data-Glacier-Internship/upload/main/Week_7

Tabular data details:

Total number of observations	1000000
Total number of files	1
Total number of features	48
Base format of the file	csv
Size of the data	1146,663 KB

Proposed Approach:

1. Handling Outliers

- Age Column:
    - Some outliers in the age column were identified, particularly customers with ages above 100. The maximum age in the dataset was capped based on the highest plausible age value found.
    - This adjustment ensures that extreme outliers do not distort our analysis of customer demographics.

2. Handling Null Values

- Current Assumptions:
    - There are null values in some columns. We have yet to finalize the handling strategy, but potential approaches include:
        - Dropping rows where critical fields are null, particularly for fields necessary for analysis like customer type and relationship status.
        - For non-critical columns (e.g., account-related indicators), null values may be assumed to represent inactive services.
    - Final decisions on how to handle null values will be agreed upon during further discussions.