Certainly! Here's a comprehensive summary of the project:

---

### Project Summary

#### Objective:

The primary goal of this project was to build a Demographic-Based Offer Recommendation System for Starbucks. Using data provided by Starbucks, the recommendation system aimed to predict which offers would likely be completed by different demographic groups.

#### Data:

Three data files were provided:

1. \*\*portfolio.json\*\*: Contains metadata about each offer (e.g., type, difficulty, reward).

2. \*\*profile.json\*\*: Contains demographic data for each customer (e.g., age, gender, income).

3. \*\*transcript.json\*\*: Contains records for transactions, offers received, viewed, and completed.

#### Methodology:

1. \*\*Data Loading and Exploration\*\*:

- Loaded the datasets from their respective paths.

- Explored the datasets to understand their structure, features, and potential challenges (e.g., missing values).

2. \*\*Data Preprocessing\*\*:

- Cleaned data to handle missing values, especially in the profile dataset.

- Extracted relevant information from the transcript dataset, such as offer\_id and transaction amounts.

- Merged the three datasets to form a comprehensive dataframe that could be used for analysis and modeling.

3. \*\*Data Analysis\*\*:

- Segmented the merged dataset based on age groups (Young, Middle-aged, Senior) to identify patterns in offer completions across these groups.

- Analyzed how different demographic groups responded to different offer types, resulting in a summary table showcasing these preferences.

4. \*\*Machine Learning Modeling\*\*:

- Prepared the merged data for modeling by converting categorical variables (offer type and gender) into dummy variables.

- Created a binary target variable to indicate offer completion.

- Split the data into training and testing sets.

- Trained a Logistic Regression model on the training data and evaluated its performance on the test data.

#### Results:

The Logistic Regression model achieved a training accuracy of approximately 86.34% and a test accuracy of approximately 86.18%. However, the model showed a bias towards predicting "offer not completed", indicating potential imbalances in the dataset.

#### Challenges and Future Improvements:

- \*\*Data Imbalance\*\*: The model's bias towards a particular class suggested that there might be an imbalance in the dataset. This can be addressed through techniques like resampling.

- \*\*Model Choice\*\*: In future iterations, experimenting with other algorithms, such as Random Forest or Gradient Boosting, might yield better results.

- \*\*Feature Engineering\*\*: Creating new features or modifying existing ones might improve the model's predictive power.

- \*\*Hyperparameter Tuning\*\*: Further tuning of the model's hyperparameters might lead to better performance.