**Portfolio Description**

**Title:Cardio Good Fitness Case Study: Customer Analysis and Product Optimization**

**Objective**: Analyse customer demographics and behaviour to understand product preferences, optimize marketing strategies, and enhance business outcomes for CardioGood Fitness.

**README**

**Project Name: Cardio Good Fitness Case Study**

**Overview:** This project analyses the CardioGood Fitness dataset to uncover customer trends, segment audiences, and predict product preferences using machine learning techniques. The model was fine-tuned using hyperparameter optimization and validated for improved predictive accuracy. Notably, two clusters exhibited a clear separation, yielding better segmentation for targeted marketing strategies.

**Features:**

**Data Analysis:** Summary statistics and exploratory data analysis (EDA) to understand customer demographics. Visualization of product preferences across different customer segments.

**Clustering:** Customer segmentation using K-Means clustering. Enhanced model separation, with two dominant clusters exhibiting a distinct divide for better targeting.

**Predictive Modelling:** Classification models, specifically a Random Forest model, to predict treadmill product preferences. Fine-tuned the Random Forest model using GridSearchCV for hyperparameter optimization. Cross-validated using k-fold cross-validation to ensure model robustness.

**Actionable Insights:** Recommendations for marketing strategies based on customer trends.

**Skills Demonstrated:**

**Data Cleaning and Preprocessing**: Handling missing values, encoding categorical variables, and standardizing data.

**EDA and Visualization**: Using Seaborn and Matplotlib for in-depth analysis and visual representation of customer data.

**Clustering and Segmentation**: Identifying two primary customer groups with a clear separation through K-Means clustering. Improved segmentation for focused marketing and product development strategies.

**Machine Learning and Model Tuning**: Building and evaluating classification models. Fine-tuning the Random Forest model using GridSearchCV to optimize parameters like n\_estimators, max\_depth, and min\_samples\_split for improved performance. Model validation with k-fold cross-validation to avoid overfitting.

**Python Libraries:** Pandas, NumPy, Scikit-learn, Seaborn, Matplotlib.

**Findings and Recommendations**

**Product Preferences:**

* TM195 attracts younger, budget-conscious customers.
* TM498 is preferred by middle-income customers with moderate fitness goals.
* TM798 appeals to high-income individuals aiming for advanced fitness levels.

**Customer Segments:**

* Cluster 1: Younger, budget-conscious users with higher product engagement.
* Cluster 2: Middle-aged, moderate-income users seeking steady fitness progress. The separation between these two clusters was more pronounced, offering clearer targeting opportunities.

**Predictive Insights:** The Random Forest model achieved an accuracy of 85%, after fine-tuning, indicating strong predictive capabilities for product preferences.

**Recommendations:**

**Targeted Marketing:**

* TM195: Focus on affordability and beginner-friendly features for younger customers.
* TM798: Highlight premium quality and advanced fitness capabilities for high-income users.

**Product Development: C**ustomize treadmill features for each segment, e.g., tech-integrated treadmills for younger customers seeking fitness data tracking.

Retention Strategies: Use loyalty programs and personalized offers to retain high-value customers in Cluster 2 (middle-aged, moderate-income users).

This portfolio showcases your ability to apply data science techniques, fine-tune machine learning models, and deliver actionable insights to solve real-world business challenges effectively, with a particular focus on segmenting customers more effectively for targeted actions.