**Executive Summary:**

Customer Churn Prediction & Personalized Recommendation Pipeline  
Developed an end-to-end data science workflow to predict high-risk customers, deliver personalized product recommendations, and evaluate retention strategies using A/B testing. Built a Random Forest model to identify churn based on tenure, usage, and support interactions, and implemented item-item collaborative filtering for tailored recommendations. Conducted A/B testing to measure the impact of interventions on reducing churn, producing ETL-ready outputs for dashboard integration.

**Customer Churn Prediction, Personalized Recommendations, and A/B Testing Pipeline**

**Project Overview**

This project demonstrates an end-to-end data science pipeline designed to help businesses **predict customer churn**, **deliver personalized product recommendations**, and **measure the impact of retention strategies through A/B testing**. The workflow integrates feature engineering, machine learning, recommendation systems, and experimental design into a single, actionable pipeline.

**Key Objectives**

1. Identify high-risk customers likely to churn and prioritise them for intervention.
2. Generate personalized product recommendations (movies/music) to increase engagement.
3. Evaluate the effectiveness of targeted interventions using A/B testing.
4. Produce ETL-ready outputs for seamless integration with dashboards or BI tools.

**Data Simulation and Features**

* **Customer Dataset:** 500 simulated customers with features: tenure, monthly usage, and support tickets.
* **Churn Modeling:** Churn probability is realistically based on feature patterns—short tenure, low usage, or high support ticket counts increase the likelihood of churn.
* **Product Ratings Dataset:** 10 users with ratings for 10 movies/music products, simulating user preferences for personalized recommendations.

**Methodology**

1. **Churn Prediction:**
   * Trained a RandomForestClassifier to predict customer churn probabilities and identify high-risk customers.
   * Predictions occur at the start, enabling proactive retention strategies.
2. **Recommendation System:**
   * Implemented an **item-item collaborative filtering model** using cosine similarity.
   * Weighted product recommendations for each user based on their ratings, providing tailored engagement opportunities.
3. **A/B Testing:**
   * High-risk customers are randomly assigned to a control group (B) or intervention group (A).
   * Simulated intervention reduces churn probability for group A, allowing evaluation of retention strategy effectiveness.
   * Churn outcomes post-intervention are compared between groups to assess impact.

**Expected Results**

1. **High-Risk Customers:**
   * Approximately 26% identified as high-risk with predicted churn probability > 0.7.
2. **A/B Test Results:**
   * **Group A (intervention):** 65% churn rate
   * **Group B (control):** 87% churn rate
   * Demonstrates a 22%-point reduction in churn due to targeted interventions.
3. **Recommendations:**
   * Each user receives **3 top recommended products** based on item similarity and their existing ratings.
   * Example: User\_1 → Inception (3.8), Interstellar (3.6), Parasite (3.5)
4. **CSV Outputs:**
   * customer\_churn\_predictions.csv → all 500 customers with predicted churn probability and label.
   * customer\_recommendations.csv → 30 total recommendations (10 users × 3 top products).
   * high\_risk\_customers\_ab\_test.csv → all high-risk customers with A/B group assignment, adjusted churn probability, and simulated outcomes.
5. Example Visuals:

A green and orange pie chart

AI-generated content may be incorrect.

| **User ID** | **Recommended Product 1** | **Recommended Product 2** | **Recommended Product 3** |
| --- | --- | --- | --- |
| User\_1 | Inception (3.8) | Interstellar (3.6) | Parasite (3.5) |
| User\_2 | The Godfather (3.7) | La La Land (3.6) | Avengers: Endgame (3.5) |
| User\_3 | Bohemian Rhapsody (4.0) | The Dark Knight (3.8) | Inception (3.7) |

A blue and grey squares

AI-generated content may be incorrect.

**Business Impact**

* **Proactive Retention:** High-risk customers can be targeted with tailored campaigns, reducing churn by ~15%.
* **Personalized Engagement:** Recommendations improve customer satisfaction and engagement with products.
* **Data-Driven Insights:** A/B testing allows measurement of intervention effectiveness.
* **Dashboard Integration:** ETL-ready outputs enable real-time monitoring in Tableau or Power BI.

**Skills Demonstrated**

* Python (Pandas, NumPy) for data manipulation and simulation
* Machine Learning (Random Forest) for predictive modeling
* Evaluation metrics: ROC AUC, confusion matrix, classification report
* Collaborative filtering for recommendations
* A/B testing design and analysis
* End-to-end pipeline design for actionable business insights

**Tools & Technologies**

Python, Pandas, NumPy, scikit-learn, PyCharm, CSV outputs for dashboard integration