

Task: Netflix User Analytics and Churn Prediction

Download dataset from [Here](#)

Problem Statement:

The goal is to analyze Netflix user data to gain insights into user behavior, revenue patterns, and predict user churn based on the available features. You need to provide actionable insights and build a machine learning model that predicts the likelihood of users churning (cancelling their subscription).

Instructions:

Part 1: Data Analysis

1. Descriptive Analysis:

- a. Perform basic exploratory data analysis (EDA) to understand the structure and trends in the dataset.
- b. Provide insights on the following:
 - i. **Subscription Type Distribution:** Breakdown of different subscription types by country and gender.
 - ii. **Monthly Revenue Analysis:** Show the distribution of monthly revenue based on subscription type, age groups, and device types.
 - iii. **Join Date Insights:** Determine trends in user sign-ups (e.g., monthly/seasonal sign-up rates).
 - iv. **Device Insights:** What are the most popular devices used by users in different countries?

Conceptual Test:

- c. Explain the importance of EDA before building any machine learning model.
- d. Identify any missing or inconsistent data and explain how to handle them.
(You have to do it on jupyter notebook)

2. Revenue Growth Trends:

- a. Calculate monthly revenue growth from the dataset.
- b. Analyze revenue trends over time and by country. Are there any countries that show a decline or significant increase in revenue?

3. User Retention Analysis:

- a. Calculate the **average plan duration** by subscription type.
- b. Determine the **churn rate** (users who did not make any payments after a certain period). Provide a definition of how you calculate churn in this dataset and any assumptions made.

Conceptual Test:

- c. Explain how you could assess user retention using this dataset.
- d. Suggest strategies Netflix could adopt to improve user retention based on the data analysis. (You have to do it on jupyter notebook)

Part 2: Machine Learning

4. Feature Engineering:

- a. Create new features that might be helpful for predicting user churn. For example:
 - i. Days between Join Date and Last Payment Date (Subscription Tenure).
 - ii. Age groups (e.g., 18-25, 26-35, etc.).
 - iii. Revenue/plan duration (average revenue per month).
- b. Use these engineered features in your machine learning model.

Conceptual Test:

- c. Explain why feature engineering is important in improving machine learning models.
(You have to do it on jupyter notebook)

5. Churn Prediction Model:

- a. Build a machine learning model to predict whether a user will churn based on the available features. Use appropriate evaluation metrics.
- b. Split the data into training and test sets.
- c. Compare at least two machine learning models (e.g., Logistic Regression, Random Forest, Gradient Boosting).
- d. Tune the hyperparameters of the best-performing model and explain why it performs better.

Conceptual Test:

- e. Explain the reasoning behind the selection of models and metrics used for this classification problem.
- f. Why did you choose the specific evaluation metric(s)? (Accuracy, Precision, Recall, F1-score, etc.)
- g. Discuss the importance of balancing the dataset (if applicable) and techniques to handle imbalanced data. (You have to do it on jupyter notebook)

6. Model Evaluation:

- a. Evaluate the performance of the models using confusion matrix, ROC curve, and precision-recall curve.
- b. Interpret the results and suggest how Netflix could use the churn prediction model to target at-risk customers and reduce churn.

7. Data Storytelling:

- a. Visualize your insights and the predictions of your model.
- b. Create a brief report summarizing your findings, methodology, and recommendations for Netflix to improve revenue and reduce churn.

Deliverables:

- A Jupyter Notebook containing:
 - Data cleaning and preprocessing steps.
 - Exploratory data analysis with detailed visualizations.
 - Feature engineering process and explanation.
 - Model training, evaluation, and interpretation.
 - A conclusion section with actionable insights for Netflix.
 - A short report summarizing the key findings and recommendations based on both data analysis and model results.

Evaluation Criteria:

1. **Understanding and Conceptual Knowledge:** Ability to explain each step and justify decisions made in the analysis and modeling process.
2. **Technical Implementation:** Quality of code, preprocessing, and model tuning. Focus on correct implementation and understanding of machine learning algorithms.
3. **Creativity in Feature Engineering:** Ability to create and justify new features that improve model performance.
4. **Interpretation and Insight:** Clear, actionable insights drawn from the data analysis and model predictions.
5. **Visualization:** Quality and clarity of visualizations used to support conclusions.

