PROBLEM STATEMENT

I am conducting a comprehensive data analytics case study on Netflix using a large-scale, real-world dataset containing over 300 million subscriber records (as of Q3 2025), including the following key variables:

- Customer ID, subscription start date, last active date, cancellation date (if applicable)
- Current subscription status (Active / Inactive / Churned)
- Subscription tier (Basic, Standard, Premium, Standard with Ads)
- Monthly viewing hours, device usage, content genre preferences, binge-watching behavior
- Country/Region, city-level geolocation, language preference

Objectives of My Analysis:

1. Determine the exact number and percentage of Active vs. Inactive/Churned customers globally and by region.
2. Calculate monthly and annual churn rates at global, regional, and country levels.
3. Identify high-risk customer segments most likely to churn in the next 30–90 days using predictive modeling (Logistic Regression, Random Forest, XGBoost).
4. Quantify the size of the "at-risk" customer pool (customers with >60% predicted churn probability) and estimate potential revenue impact if no intervention is made.