Business Requirements Document (BRD)

# **Project Name: Customer Churn Prediction Project**

## 1. Project Overview

* **Purpose:** Predict which customers are likely to cancel their subscriptions using the past three months of customer data (subscription and listening history).
* **Scope:** The project will utilize supervised learning techniques to develop a predictive model that identifies customers at high risk of churn.

**2**. Business Context

* **Problem Statement:** The company has identified an increasing trend in subscription cancellations, which impacts revenue and customer base stability.
* **Business Need:** Developing a predictive model to identify potential churn will enable proactive engagement strategies, thereby enhancing customer retention and satisfaction.

## 3. BDA Domains Integration

### Identify Research Questions (IR)

* Define specific analytics questions that the project aims to answer:
  + What patterns in subscription and listening behaviour indicate a high risk of churn?
  + Which predictive variables are most indicative of churn?

### Source Data (SD)

* **Data Sourcing Requirements**:
  + Subscription details: Start and end dates, type of subscription, active status.
  + Listening history: Track details, listening durations, user interactions.
* **Data Governance**:
  + Ensure data collection and usage comply with privacy laws and company policies.

### Analyse Data (AD)

* **Analytical Techniques**:
  + Descriptive analytics to understand customer behavior trends.
  + Predictive analytics to model churn likelihood.
* **Tools and Technologies**:
  + Use Python and its libraries like pandas, scikit-learn, and visualization tools such as matplotlib.

### Interpret and Report Results (IR)

* **Reporting and Decision Making**:
  + Develop comprehensive reports detailing the findings.
  + Provide actionable insights and recommendations based on the model predictions.
* **Visualization**:
  + Create intuitive and insightful visualizations to represent data findings.

### Use Results to Influence Business Decision Making (UR)

* **Operational Implementation**:
  + Integrate predictive model outputs into customer relationship management strategies.
* **Feedback Mechanisms**:
  + Establish channels for continuous feedback on model performance and business impact.

### Guide Company Actions (GA)

* **Policy and Strategy Development**:
  + Assist in formulating or updating policies related to customer data usage and retention strategies.
* **Ethical Considerations**:
  + Regularly assess the ethical implications of predictive modelling and its impact on customers.

## 4. Project Requirements

* **Functional Requirements**:
  + Detailed above in the BDA domains integration.
* **Non-Functional Requirements**:
  + Performance, reliability, usability, and security specifics are outlined.

## 5. Success Criteria

* **Performance Metrics**:
  + Model's accuracy, recall, precision, and AUC.
  + Reduction in churn rates.
* **Business Impact**:
  + Enhanced customer retention and satisfaction.

## 6. Project Timeline

* **Phases** and **Milestones**:
  + Include key phases like Preparation, Development, Deployment, and Evaluation.
  + Specify milestones for Project Kick-off, Model Completion, Go-Live, and First Review.

## 7. Assumptions and Constraints

* Assume data representativeness and stakeholder support.
* Note constraints related to data history, budget, and timelines.

## 8. Approval

* **Document Approval**:
  + Preparation, review, and approval details including responsible personnel.

This enhanced BRD ensures that every aspect of the project is well-documented, aligning with the BDA domains to provide a structured and comprehensive approach to achieving the project's goals through data-driven decision-making. This structure facilitates clear communication among stakeholders and guides the project toward strategic alignment and operational success.

## Code Review: Customer Data Analysis for Maven Music

**Reviewer**: Ntiyiso Ndhlovu  
**Date**:   
**File**: customer\_churn\_prediction.ipynb  
**Purpose**: To analyze customer data and listening history for Maven Music to identify trends and factors influencing customer retention and engagement.

### 1. Overview

The notebook is structured to perform data loading, cleaning, exploration, feature engineering, data merging/aggregation, statistical analysis, and visualization for customer data from Maven Music. The goal is to prepare the data for modeling and derive insights into customer behavior patterns.

### 2. Code Review

#### General Observations

* The code is logically organized, following the sequence from data loading to processing and analysis.
* Appropriate use of libraries like pandas for data manipulation and seaborn for visualization.
* Comments are sparse and could be improved to enhance readability and maintainability.

#### Specific Observations and Recommendations

#### Section: Data Loading

* **Code**: Importing data from CSV and Excel files.
* **Observations**: Data is loaded from multiple sources and formats effectively.
* **Recommendations**:
  + **Enhance Comments**: Include comments on the source and structure of the data files.
  + **Error Handling**: Add error handling for file loading to manage missing files or incorrect paths.

#### Section: Data Cleaning

* **Code**: Converting data types and handling missing values.
* **Observations**: Effective use of pandas for data type conversion and missing data management.
* **Recommendations**:
  + **Validation**: After data type conversion, add assertions or checks to confirm conversions.
  + **Documentation**: Comment on why specific default values are chosen for missing data.

#### Section: Exploratory Data Analysis (EDA)

* **Code**: Checking unique values and distribution of data.
* **Observations**: Use of basic pandas functions to explore the data.
* **Recommendations**:
  + **Visualization**: Include more plots (histograms, boxplots) to visualize distributions and outliers.
  + **Detailed Analysis**: Deeper analysis of anomalies or unusual patterns observed in the data.

#### Section: Feature Engineering

* **Code**: Creating dummy variables and aggregating data.
* **Observations**: Effective transformation of categorical variables into a usable format for modeling.
* **Recommendations**:
  + **Efficiency**: Investigate more efficient ways to handle large datasets during dummy variable creation.
  + **Robustness**: Add checks to ensure all expected categories are included in the dummy variables.

#### Section: Data Merging and Aggregation

* **Code**: Merging various dataframes to create a comprehensive dataset.
* **Observations**: Correct use of merge operations to consolidate data.
* **Recommendations**:
  + **Key Integrity**: Ensure the keys used for merging have no duplicates and are not missing.

#### Section: Calculating Derived Metrics

* **Code**: Calculating percentages and derived statistics.
* **Observations**: Accurate calculations that are essential for insights.
* **Recommendations**:
  + **Accuracy**: Add checks to confirm that calculations do not result in unexpected values (e.g., percentages over 100%).

#### Section: Visualization

* **Code**: Using seaborn for pair plots.
* **Observations**: Effective visualization to understand pairwise relationships.
* **Recommendations**:
  + **Interactivity**: Consider using interactive plotting libraries like Plotly for dynamic insights.

#### Section: Correlation Analysis

* **Code**: Computing correlation matrices.
* **Observations**: Essential step for understanding feature relationships.
* **Recommendations**:
  + **Heatmap**: Use a heatmap for better visualization of correlation matrices.

### 3. Summary

The notebook demonstrates a thorough approach to data analysis with robust data manipulation techniques. However, there is a need for enhanced commenting, error handling, and more detailed EDA to improve the robustness and usability of the analysis. Further, some additional visualizations and checks could help in making the notebook an even more powerful tool for insight generation.

**Simplified Customer Churn Analysis Report**

**Summary**

This report looks at why customers of a music streaming service might decide to leave (churn) and suggests ways to keep them subscribed. By studying how often customers use discounts, how much they listen, and what kind of music they like, we identify key trends and offer strategies to improve customer retention.

**Goals**

The main goal is to understand what behaviors and characteristics show us a customer might soon cancel their service.

**Questions We Explored**

1. **What behaviors and subscription details suggest a customer might leave us soon?**
2. **What are the best clues that a customer might stop using our service?**

**Detailed Findings**

**Discounts and Cancelling**

* **What We Saw**: Customers who get discounts are more likely to cancel their subscriptions.
* **What It Means**: It seems that discounts don't solve the main reasons customers are unhappy. Instead, they might only delay the customer's decision to leave.

**Listening Habits and Keeping Customers**

* **What We Saw**: Customers who use the service a lot are less likely to cancel.
* **What It Means**: The more a customer uses the service, the happier they seem to be with it, making them want to stay longer.

**Types of Content and Customer Loyalty**

* **Pop Music**: Customers who listen to a lot of pop music are more likely to cancel.
  + **What It Means**: There may be problems with the selection or variety of pop music available.
* **Podcasts**: Listening to podcasts doesn't really affect whether customers cancel.
  + **What It Means**: Customers who listen to podcasts might have different reasons for staying or leaving compared to those who listen to music.

**Best Clues for Predicting If a Customer Will Cancel**

1. **Using Discounts**: If a customer often uses discounts, they might be thinking about cancelling.
2. **Frequency of Use**: Customers who don't use the service much are more likely to stop using it.
3. **Listening to Pop Music**: Customers who listen to a lot of pop music are at higher risk of cancelling.

**Suggestions for Keeping Customers**

**Better Use of Discounts**

* **Smarter Discounts**: Give discounts based on a customer’s habits to try to truly make them happier and more likely to stay.
* **Check If Discounts Work**: Regularly check to see if our discounts are actually making customers want to stay longer.

**Make Listening More Enjoyable**

* **Personal Touches**: Use technology to recommend music and features that fit each customer’s taste.
* **Rewards for Using More**: Create fun programs that reward customers for using the service more, like earning badges for listening to a variety of songs.

**Improving Music Choices**

* **Better Pop Music Options**: Look closely at our pop music choices. Ask customers what they like or don’t like and try to fix any issues.
* **Special Features for Podcast Fans**: Since podcast listeners have different habits, think about offering special content or features just for them.

**Using Data to Prevent Cancelling**

* **Predictive Models**: Use data science to predict which customers might cancel and try to make them happier before they decide to leave.
* **Act Early**: Use the information from our data to reach out to customers who might be thinking of leaving. Offer them things that could change their minds.

**Closing Thoughts**

By understanding why customers might cancel and acting on these insights, we can make their experience better and keep them for longer. This approach needs to be flexible and adapt to new feedback and changes in what customers expect from us. This way, we can keep improving and keep our customers happy