

Phase 1 : Problem Definition and Design Thinking

Problem Definition:

The project involves using IBM Cognos to predict customer churn and identify factors influencing customer retention. The goal is to help businesses reduce customer attrition by understanding the patterns and reasons behind customers leaving. This project includes defining analysis objectives, collecting customer data, designing relevant visualization as in IBM Cognos, and building a predictive model.

Design Thinking:

Analysis objective :

Conduct in-depth churn analysis to understand the patterns and trends associated with customers who have already churned. This could involve exploring when and why they left, which products or services they were using, and any commonalities among churned customers.

Data collection :

Customer churn data tells you how many customers have stopped using your product or service over a period of time.

Utilize data that the business already has in its internal records, such as customer registration information, account profiles, and purchase history. This data may include age, gender, location, and contact information.

Conduct customer surveys to gather demographic information that may not be available in internal records.

Visualization Strategy :

Data visualization allows you to quickly detect patterns, trends, and outliers in churn, enabling you to make key business decisions that can address negative patterns before they get out of control.

Visualize customer segments derived from your analysis. Use pie charts, bar charts, or tree maps to show the distribution of customers across segments.

Plan for regular updates of the visualizations as new data becomes available. Automation can help in keeping the visualizations up to date.

Predictive Model : RANDOM FOREST TECHNIQUE

Random Forest: An ensemble method that handles feature importance well and can capture complex relationships.

This technique can handle skewed data and categorical data (ordinal or non-ordinal). It is generally considered a very accurate predictive model and it can handle a large number of independent predictor variables.

