Problem Statement: Predicting Customer Churn in the Telecom Industry

Objective: The objective of this project is to analyze customer data in the telecom industry and develop a predictive model to identify customers who are likely to churn. By understanding the factors that contribute to customer churn, telecom companies can proactively implement strategies to retain customers, improve customer satisfaction, and reduce revenue loss.

Dataset Description: The dataset provided consists of 100 variables and approximately 100,000 records. It contains various attributes related to the telecom industry and factors considered important for customer management. The target variable, "churn," indicates whether a customer is likely to churn or not.

Approach and Roadmap:

1. Data Understanding: Explore and familiarize yourself with the dataset, including the meaning and significance of each variable. Identify any missing values, outliers, or data quality issues that need to be addressed.

2. Data Preprocessing: Cleanse and preprocess the data to ensure its quality and suitability for analysis. Handle missing values, outliers, and perform data transformations as needed. This step may involve techniques such as data imputation, outlier detection, and feature engineering.

3. Exploratory Data Analysis (EDA): Conduct a comprehensive EDA to gain insights into the dataset. Perform statistical analysis, data visualization, and identify correlations between variables. Explore the distribution of churn and other important variables related to customer behavior and telecom industry attributes.

4. Feature Selection: Select the most relevant features for predicting churn. Use techniques such as correlation analysis, feature importance, and domain knowledge to identify the key drivers of churn in the telecom industry.

5. Model Development: Build a predictive model to identify customers at risk of churn. Apply appropriate machine learning algorithms such as logistic regression, decision trees, random forests, or gradient boosting. Split the dataset into training and testing sets, and evaluate the model's performance using appropriate metrics such as accuracy, precision, recall, and F1-score.

6. Model Interpretation: Interpret the model results to gain insights into the factors influencing customer churn. Identify the most significant variables and their impact on churn prediction. Explain the model's predictions to stakeholders and decision-makers using clear and concise explanations.

7. Solution and Recommendations: Based on the insights gained from the analysis, provide actionable recommendations to reduce customer churn. Propose strategies and initiatives that telecom companies can implement to improve customer retention, enhance service offerings, and address regulatory requirements effectively.

8. Visualizations and Reporting: Create informative visualizations, charts, and dashboards to present the analysis findings and model performance. Use tools such as Tableau, Power BI, or Python libraries (e.g., Matplotlib, Seaborn) to create compelling visual representations of the data and insights.

9. Integration with Jira and Project Management: Utilize Jira to track the progress of the project, manage tasks, and collaborate with stakeholders. Document the project steps, findings, and deliverables in Jira to ensure transparency and effective project management.

10. Documentation and Presentation: Prepare a comprehensive report summarizing the entire project, including the problem statement, data preprocessing steps, EDA insights, model development, interpretation, recommendations, and visualizations. Present the findings and recommendations to stakeholders, highlighting the value of the analysis in addressing regulatory requirements and reducing customer churn.

Throughout the project, leverage your skills in data analysis, SQL querying, Jira dashboard handling, and other relevant tools and techniques to perform various analyses, develop models, and derive meaningful insights. Ensure compliance with regulatory requirements and maintain data privacy and security throughout the analysis process.