Introduction:

The project focused on predicting customer churn in a telecommunications company. Customer churn is a crucial metric for businesses as it impacts revenue and customer retention. The goal was to develop a model that could predict which customers were likely to churn, allowing proactive retention strategies to be implemented.

Data Acquisition and Cleaning:

The data was sourced from the company's CRM system and included customer demographics, services subscribed, and churn status. Data cleaning involved handling missing values, encoding categorical variables, and removing outliers. Pre-processing steps included scaling numerical features and splitting the data into training and testing sets.

Exploratory Data Analysis:

Key findings from exploratory data analysis revealed that customers with month-to-month contracts and multiple services were more likely to churn. There was also a correlation between tenure and churn rate, with newer customers being more likely to churn.

Feature Engineering:

Features selected included contract type, internet service, tenure, and monthly charges. New features were created, such as total charges derived from monthly charges and tenure. These features were used to train the predictive model.

Model Building:

Models experimented with included logistic regression, random forest, and XGBoost. Random forest was chosen as the final model due to its superior performance in predicting customer churn. The model achieved an accuracy of 85% on the test data.

Model Evaluation:

Evaluation metrics included accuracy, precision, recall, and F1 score. Insights gained from analyzing the model's results showed that the most influential features in predicting churn were contract type and total charges. The model performed well in identifying customers at high risk of churn.

Conclusion:

The project successfully developed a predictive model for customer churn, providing valuable insights for the company to implement targeted retention strategies. Limitations included the need for more real-time data and potential bias in the dataset. Future areas of exploration could involve incorporating additional data sources and refining feature engineering techniques.