CUSTOMER CHURN ANALYSIS PROJECT REPORT

1. Objective

The objective of this project is to analyze telecom customer data and predict customer churn using a logistic regression model. This analysis provides insights into key factors influencing churn and helps in business decision-making to reduce churn rate.

2. Dataset Overview

The dataset used is 'WA_Fn-UseC_-Telco-Customer-Churn.csv', which includes information on telecom customer demographics, services subscribed, tenure, charges, and churn labels (Yes/No).

3. Tools & Technologies Used

- Python (Pandas, NumPy, Scikit-learn, Joblib)
- Power BI for data visualization
- Jupyter Notebook for code development

4. Data Preprocessing

Steps included:

- Dropping customerID column
- Converting TotalCharges to numeric
- Handling missing values
- Encoding categorical variables using one-hot encoding
- Standardizing features using StandardScaler

5. Model Development

Machine Learning Model:

- Algorithm: Logistic Regression
- Train-Test Split: 80:20

Metrics:

- Accuracy
- Confusion Matrix
- Classification Report

A Logistic Regression model was trained using an 80-20 train-test split. The model's performance was evaluated using accuracy, confusion matrix, and classification report.

6. Model Saving

The trained model and scaler were saved using joblib for later use in deployment or prediction.

7. Dashboard Visualization

A Power BI dashboard was created with multiple visuals to explore churn distribution, service usage, and financial metrics:

- **Bar Chart**: Internet Service vs Churn
- **Clustered Column Chart**: Contract Type vs Churn
- **Pie Chart**: Gender vs Churn
- **Line Chart**: Tenure vs Churn
- **Decomposition Tree**: Drivers of Churn
- **Key Influencers Visual**: Identify top features contributing to churn

8. Insights Derived

- Customers with month-to-month contracts churn more.
- Fiber optic users tend to churn more than DSL users.
- Lower tenure and higher monthly charges are common among churned customers.
- Support services, contract type, and tech support significantly affect churn.

9. Conclusion

This project helped:

- Identify the key reasons for customer churn
- Predict churn using a trained logistic regression model
- Create business-friendly visuals to assist decision-makers

10. Future Improvements

- Test more advanced ML algorithms (Random Forest, XGBoost)
- Deploy model using Flask or Streamlit
- Automate monthly churn reports