Telco Customer Churn Analysis Project

Overview

This project analyzes **Telco customer churn** to identify patterns and factors influencing whether a customer will leave the service. It includes **data cleaning, feature engineering, exploratory data analysis (EDA), and visualizations** to provide actionable insights for retention strategies.

◆ Dataset

- **Source:** Telco Customer Churn dataset
- **Features:** Customer demographics, account information, services, payment method, tenure, charges, etc.
- **Target Variable:** Churn (Yes/No)

🛠 Pipeline Workflow

1. **Data Cleaning (datacleaning.py)**
    - Removed duplicates
    - Converted TotalCharges to numeric
    - Filled missing values with median
    - Generated log file (datacleaning.log)
2. **Feature Engineering & Encoding (feature_encoding.py)**
    - Dropped customerID
    - Encoded target Churn (Yes=1, No=0)
    - Encoded binary features (gender, Partner, Dependents, etc.)
    - One-hot encoded categorical features (InternetService, Contract, PaymentMethod)
    - Saved encoded dataset (encoded_telco_churn.csv)
    - Logged all steps (feature_encoding.log)
3. **Visualization & EDA (visualisation.py)**
    - Added **SIM_Operator** column (synthetic for analysis)
    - Plots generated and **saved automatically** in outputs/plots/
    - Key plots:
        - Churn Distribution
        - Gender Distribution
        - Tenure vs Churn
        - Monthly Charges vs Churn
        - Payment Method vs Churn
        - Gender vs Internet Service
        - Churn vs SIM Operator
        - SIM Operator vs Gender

    - All plots logged (visualisation.log)
4. **Pipeline Orchestration (main.py)**
    - Runs all modules sequentially
    - Handles exceptions
    - Generates logs for the full workflow (main.log)

📊 Key Insights

- **Churn Rate:** High among certain contract types and payment methods.
- **Tenure:** Customers with shorter tenure are more likely to churn.
- **Monthly Charges:** Customers in higher quartiles tend to churn more.
- **SIM Operator:** Differences in churn across network providers.
- **Demographics:** Gender and senior citizen status influence churn slightly.

All insights are derived from **automated plots saved in outputs/plots/.**

Technologies Used

- Python 3.x
- pandas, numpy
- matplotlib
- Logging module for audit trails
- Modular Python scripts for pipeline workflow

📝 Notes

- SIM_Operator column is **synthetically generated** for visualization purposes.
- All modules include **try–except blocks** and logging for safe execution.
- Pipeline is designed to be **reproducible and non-interactive**.

📌 Future Enhancements

- Add **machine learning model** to predict churn
- Create a **Streamlit dashboard** to visualize churn interactively
- Generate **automated report PDFs** combining all plots