Customer Churn Analytics Flow (SQL → Python → ML → Insights)
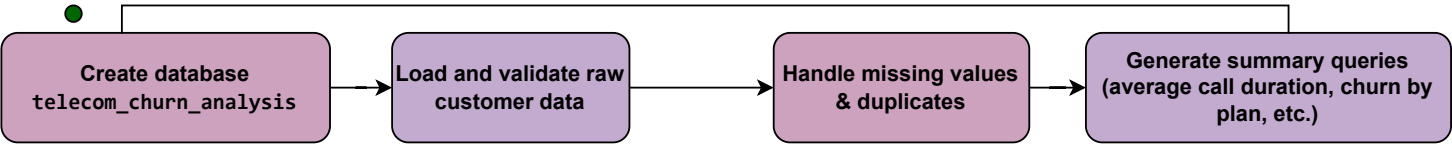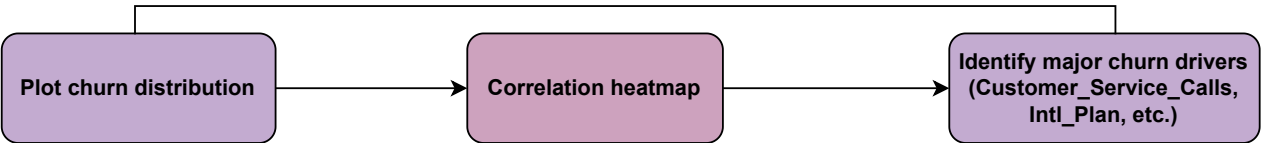
Data Acquisition (SQL)

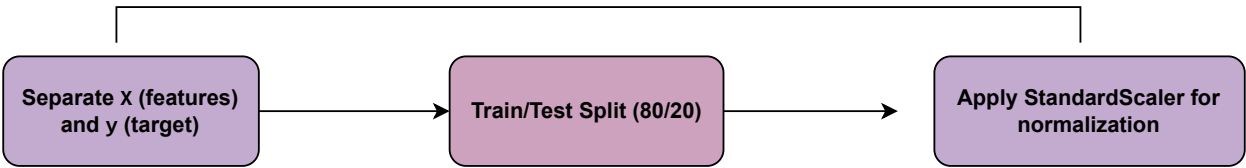| Create database telecom_churn_analysis | → | Load and validate raw customer data | → | Handle missing values & duplicates | → | Generate summary queries (average call duration, churn by plan, etc.) |

Data Cleaning & Preparation (Python)

| Import CSV into Pandas | → | Encode categorical columns (International_Plan, Voice_Mail_Plan, Churn) | → | Drop irrelevant columns (State) | → | Check missing values & basic statistics |
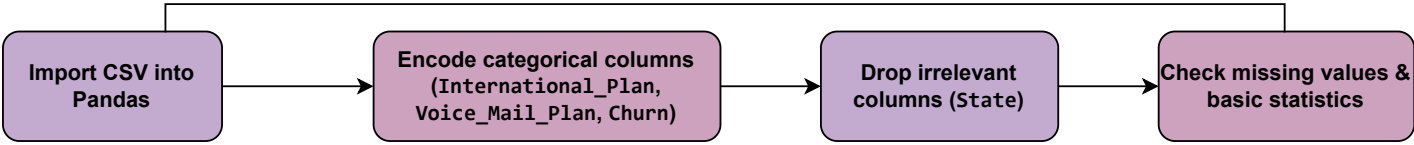
Exploratory Data Analysis (Python – Seaborn/Matplotlib)

| Plot churn distribution | → | Correlation heatmap | → | Identify major churn drivers (Customer_Service_Calls, Intl_Plan, etc.) |

Feature Engineering & Split

| Separate X (features) and y (target) | → | Train/Test Split (80/20) | → | Apply StandardScaler for normalization |

Model Building (ML Layer)

| Train Logistic Regression, Decision Tree, Random Forest | | Predict churn on test data |

Model Evaluation

| Compare models using Accuracy, Precision, Recall, F1-Score | | Random Forest = *Best Model (Accuracy ~94.76%)* |

Business Insights & Actions

| Identify high-risk churn customers | → | Recommend targeted retention strategies | → | Estimate potential revenue impact |