**Case Study: Data Cleaning and Exploratory Data Analysis (EDA) on Telecom Customer Data**

**1. Introduction**

**Objective:** Perform data cleaning and exploratory data analysis (EDA) on the given dataset to identify trends, inconsistencies, and key insights related to customer churn.

**Dataset Overview:** The dataset consists of telecom customer information, including demographic details, service subscriptions, billing information, and churn status.

**2. Data Cleaning Process**

**2.1 Handling Missing Values**

* Checked for missing values across all columns.
* Identified missing values in TotalCharges and imputed them appropriately (e.g., replacing empty strings with the mean or median of the column).

**2.2 Data Type Corrections**

* Converted TotalCharges from string to numeric.
* Ensured categorical variables (e.g., Gender, Contract, PaymentMethod) were properly encoded for analysis.

**2.3 Handling Duplicates**

* Checked for duplicate CustomerID values and confirmed dataset integrity.

**2.4 Addressing Outliers**

* Used boxplots and IQR method to detect outliers in MonthlyCharges and TotalCharges.
* Analyzed impact and decided on appropriate handling methods (e.g., capping extreme values).

**3. Exploratory Data Analysis (EDA)**

**3.1 Demographic Analysis**

* **Gender Distribution:** Checked male vs. female proportions.
* **Senior Citizens:** Assessed whether senior citizens have a higher churn rate.
* **Partner and Dependents:** Evaluated their impact on churn.

**3.2 Service Subscription Analysis**

* **Phone and Internet Services:** Analyzed how different service combinations impact churn.
* **Additional Features (Security, Backup, Device Protection, Tech Support, Streaming Services):**
  + Identified which services have the highest association with churn.

**3.3 Contract and Billing Analysis**

* **Contract Types:** Compared churn rates for month-to-month, one-year, and two-year contracts.
* **Payment Methods:** Assessed the relationship between payment methods and churn likelihood.
* **Monthly Charges vs. Churn:** Identified whether high charges correlate with higher churn rates.
* **Paperless Billing:** Evaluated whether customers who opt for paperless billing churn more frequently.

**3.4 Churn Trend Analysis**

* **Churn Rate Overview:**
  + Percentage of customers who churned.
  + Key characteristics of churned customers.
* **Tenure vs. Churn:**
  + Identified churn patterns based on customer tenure.
  + Determined if newer customers are more likely to leave.

**4. Key Findings & Insights**

* Customers with **month-to-month contracts** had the highest churn rate.
* **Senior citizens and single customers (without a partner or dependents)** exhibited higher churn rates.
* Customers without **tech support or security services** were more likely to churn.
* **Higher monthly charges** correlated with higher churn probability, but long-term customers tended to stay despite high charges.
* **Electronic check payment method** had the highest churn rate among payment options.

**5. Recommendations**

* **Customer Retention Strategy:** Encourage long-term contracts by offering discounts or incentives.
* **Enhance Customer Support:** Promote tech support and security services to increase customer satisfaction.
* **Billing Optimization:** Address high churn in customers with electronic check payments by promoting alternative methods.
* **Loyalty Programs:** Implement strategies to retain new customers, such as first-year discounts or engagement programs.

**6. Conclusion**

The EDA provided valuable insights into customer churn trends and factors influencing churn behavior. These findings can guide strategic decisions to improve customer retention and reduce churn rates in the telecom industry.

**Customer Churn Analysis Presentation**

**Slide 1: Title Slide**

* **Title:** Customer Churn Analysis
* **Subtitle:** Understanding Factors Influencing Churn
* **Presented by:** [Your Name]
* **Date:** [Date]

**Slide 2: Project Overview**

* **Objective:** Identify factors contributing to customer churn and propose solutions.
* **Dataset:** 7043 customer records with 21 attributes.
* **Key Challenge:** Imbalanced dataset with only 27% churned customers.

**Slide 3: Data Summary**

* **Total Customers:** 7043
* **Churned Customers:** 1869 (27%)
* **Features Include:** Demographics, service usage, contract details, billing information.
* **Data Cleaning:** Converted 'TotalCharges' to numeric, handled missing values.

**Slide 4: Key Insights from Exploratory Data Analysis**

* **Demographics:** Most customers are younger individuals without dependents.
* **Monthly Charges:** Higher charges correlate with higher churn.
* **Contract Type:** Month-to-month contracts see the highest churn.
* **Internet Service:** Fiber optic users churn more frequently than DSL users.
* **Tech Support & Online Backup:** Lack of support services increases churn risk.

**Slide 5: Visualizations & Findings**

* **Bar Charts:** Distribution of churn by contract type, internet service, and tech support.
* **Boxplot:** Monthly charges vs. churn status.
* **Count Plots:** Churn rates across different categorical variables.

**Slide 6: Data Preparation & Feature Engineering**

* **Categorical Encoding:** Transformed categorical variables into numerical values.
* **Feature Selection:** Identified key predictors of churn.
* **Oversampling:** Used SMOTE to balance dataset.

**Slide 7: Predictive Modeling Strategy**

* **Models Considered:** Logistic Regression, Random Forest, XGBoost.
* **Evaluation Metrics:** Accuracy, Precision, Recall, F1-score.
* **Preliminary Findings:** Contract type and monthly charges are strong churn predictors.

**Slide 8: Business Recommendations**

* **Reduce Monthly Churn:** Offer discounts or loyalty incentives for high-paying customers.
* **Improve Retention:** Encourage annual contracts instead of month-to-month.
* **Enhance Support:** Offer free tech support for new customers.
* **Customer Engagement:** Target at-risk customers with personalized offers.

**Slide 9: Next Steps**

* **Model Optimization:** Fine-tune hyperparameters for better prediction.
* **Deployment:** Implement real-time churn prediction model.
* **Continuous Monitoring:** Track customer behavior and refine strategies.

**Slide 10: Q&A**

* **Thank you!**
* Questions & Discussion
* Contact: [Your Email]