**Requirement 1: Customer Billing Data Processing**

**Problem Statement:** The company wants to analyze customer billing data to identify patterns in payment delays and potential revenue leakage. The goal is to transform and clean the billing data, calculate the delay in payments, categorize customers based on their payment behavior, and visualize the distribution of payment delays.

**Transformations:**

1. **Data Cleaning:** Remove records with missing or null values in critical columns (e.g., invoice\_id, customer\_id, amount\_due, due\_date, payment\_date).
2. **Calculate Payment Delay:** Create a new column payment\_delay that calculates the difference between the due\_date and the payment\_date in days.
3. **Categorize Customers:** Add a new column payment\_behavior to categorize customers:
   * On Time if payment\_delay <= 0.
   * Late if payment\_delay > 0 and <= 30.
   * Very Late if payment\_delay > 30.
4. **Aggregation:** Aggregate the data to find the total amount due and total payments received by region.

**Visualization:**

* **Bar Chart:** Visualize the number of customers by payment\_behavior.
* **Histogram:** Show the distribution of payment\_delay.

**Dataset:** Download customer\_billing\_data.csv

**Requirement 2: Network Traffic Data Analysis**

**Problem Statement:** Analyze network traffic data to identify peak usage times and potential network congestion issues. The pipeline should filter, aggregate, and analyze the traffic data, categorize traffic based on usage, and visualize the peak traffic times.

**Transformations:**

1. **Data Filtering:** Filter out records where data\_usage is less than 1 MB.
2. **Aggregate Traffic:** Calculate the total data usage and the number of connections by hour and network\_type.
3. **Traffic Categorization:** Add a new column usage\_category:
   * Low for usage below 500 MB.
   * Medium for usage between 500 MB and 2 GB.
   * High for usage above 2 GB.
4. **Data Enrichment:** Add a column peak\_hour indicating whether the traffic occurred during peak hours (e.g., 6 PM to 10 PM).

**Visualization:**

* **Line Chart:** Show total data usage by hour.
* **Pie Chart:** Display the distribution of usage\_category.

**Dataset:** Download network\_traffic\_data.csv

**Requirement 3: SMS Usage Data Processing**

**Problem Statement:** The company wants to analyze SMS usage patterns to understand customer engagement and optimize messaging services. The task is to transform SMS usage data, calculate the total number of SMS sent, and identify the most active customers.

**Transformations:**

1. **Data Aggregation:** Aggregate the data to calculate the total number of SMS sent by customer\_id.
2. **Customer Categorization:** Add a new column sms\_activity:
   * Low for customers sending fewer than 10 SMS.
   * Medium for customers sending between 10 and 50 SMS.
   * High for customers sending more than 50 SMS.
3. **Data Filtering:** Filter out customers with no SMS activity.
4. **Data Enrichment:** Add a column last\_active\_date indicating the last date the customer sent an SMS.

**Visualization:**

* **Bar Chart:** Visualize the number of customers by sms\_activity.
* **Scatter Plot:** Show the relationship between customer\_id and the total number of SMS sent.

**Dataset:** Download sms\_usage\_data.csv

**Requirement 4: Customer Support Call Analysis**

**Problem Statement:** Analyze customer support call data to identify common issues, call resolution times, and customer satisfaction levels. The goal is to transform the data, calculate average resolution time, and categorize issues.

**Transformations:**

1. **Data Cleaning:** Remove records with null values in critical columns (call\_id, issue\_type, resolution\_time).
2. **Calculate Resolution Time:** Create a new column resolution\_time\_minutes that converts resolution\_time from seconds to minutes.
3. **Issue Categorization:** Add a new column issue\_severity:
   * Low for resolution times under 10 minutes.
   * Medium for resolution times between 10 and 30 minutes.
   * High for resolution times over 30 minutes.
4. **Aggregation:** Calculate the average resolution time by issue\_type.

**Visualization:**

* **Bar Chart:** Visualize the number of calls by issue\_severity.
* **Heatmap:** Show the average resolution time by issue\_type.

**Dataset:** Download support\_call\_data.csv

**Requirement 5: Data Usage Plan Analysis**

**Problem Statement:** Analyze data usage patterns to optimize mobile data plans offered by the telecom company. The task involves transforming the data to identify high and low data usage customers and understanding the distribution of usage across different plans.

**Transformations:**

1. **Data Filtering:** Filter out customers with data usage below 100 MB.
2. **Aggregate Usage:** Calculate total data usage by customer\_id and data\_plan.
3. **Plan Categorization:** Add a new column plan\_usage\_category:
   * Underutilized for customers using less than 50% of their data plan.
   * Optimal for customers using between 50% and 90% of their data plan.
   * Overutilized for customers using more than 90% of their data plan.
4. **Data Enrichment:** Add a column overage\_fee for customers who exceed their data plan limit.

**Visualization:**

* **Pie Chart:** Show the distribution of plan\_usage\_category.
* **Bar Chart:** Visualize the total overage fees by data\_plan.

1. **Visualize Data:** Use Google Colab to visualize data usage patterns and overage fees.

**Dataset:** Download data\_usage\_plan\_data.csv