

1. Error Log Analysis

- **Dataset:** [IBM Log Analysis \(Logytics\)](#)
 - **Problem:** Identify the most common error messages (lines containing "ERR" or "Error") and their counts.
 - **Example:**
 - **Input:** Log entries with error messages.
 - **Output:** {"ERR OR console - uaaabaam(node": 3, "Error: aaa: Key already exists: 32}
-

2. Customer Segmentation

- **Dataset:** [Customer purchase history data.](#)
 - **Problem:** Group customers by purchasing frequency.
 - **Example:**
 - **Input:** Customer purchase records (e.g., {"Customer A": 5 purchases, "Customer B": 10 purchases}).
 - **Output:** {"Low Frequency": [Customer A], "High Frequency": [Customer B]}
-

3. Climate Data Analysis

- **Dataset:** [Temperature and precipitation records by location.](#)
 - **Problem:** Calculate average temperature per region over a given period.
 - **Example:**
 - **Input:** Location-based temperature data (e.g., {"Region A": [70, 72, 68]}).
 - **Output:** {"Region A": 70}
-

4. Financial Transaction Analysis

- **Dataset:** [Bank transaction logs.](#)
- **Problem:** Identify the most common transaction type (e.g., deposit, withdrawal).
- **Example:**
 - **Input:** Transaction records.
 - **Output:** {"Deposit": 500, "Withdrawal": 300}

5. Patient Data Aggregation

- **Dataset:** [Medical records with patient age and diagnosis.](#)
- **Problem:** Count occurrences of specific diagnoses.
- **Example:**
 - **Input:** Patient records (e.g., {"Patient1": "Hypertension", "Patient2": "Diabetes"}).
 - **Output:** {"Hypertension": 20, "Diabetes": 15}

6. Product Review Summary

- **Dataset:** [Customer reviews for different product categories.](#)
- **Problem:** Calculate the average review rating by category.
- **Example:**
 - **Input:** Review ratings by category.
 - **Output:** {"Electronics": 4.2, "Clothing": 3.8}

7. Traffic Analysis

- **Dataset:** [TMS daily traffic counts.](#)
- **Problem:** Determine the total traffic per road.
- **Example:**
 - **Input:** Counts for Road X [100, 200, 150]
 - **Output:** {"Road X": 450}

8. Network Traffic Analysis

- **Dataset:** [IP Network Traffic Flows](#)
 - **Problem:** Count the number of requests made by each IP address.
 - **Example:**
 - **Input:** Log entries with IP addresses.
 - **Output:** {"192.168.1.1": 120, "192.168.1.2": 98}
-

9. Hashtag Frequency in Tweets

- **Dataset:** [Twitter Hashtags](#)
 - **Problem:** Count occurrences of each hashtag.
 - **Example:**
 - **Input:** Tweets with hashtags.
 - **Output:** {"#AI": 300, "#MachineLearning": 250}
-

10. Airline Delay Analysis

- **Dataset:** [Airline flight data.](#)
 - **Problem:** Calculate the average delay per airline.
 - **Example:**
 - **Input:** Flight delays for "Airline A" [10, 5, -3]
 - **Output:** {"Airline A": 4.0}
-

11. Network Traffic by Protocol Type

- **Dataset:** [Network traffic logs with protocol details \(e.g., TCP, UDP\).](#)
 - **Problem:** Count the number of packets per protocol type.
 - **Example:**
 - **Input:** Network traffic logs
 - **Output:** {"TCP": 10, "UDP": 5}
-

12. Energy Consumption per Device

- **Dataset:** [Smart home energy logs with device types.](#)
 - **Problem:** Calculate the total energy consumption per device.
 - **Example:**
 - **Input:** Energy usage for "Fridge" [100, 120, 130]
 - **Output:** {"Fridge": 350}
-

13. Network Social Connections

- **Dataset:** [SNAP Facebook Network Data](#)
 - **Problem:** Find the degree of connections for each user.
 - **Example:**
 - **Input:** Connections between users.
 - **Output:** {"User1": 30, "User2": 50}
-

14. COVID-19 Case Analysis by Region

- **Dataset:** [COVID-19 Data from Johns Hopkins](#)
 - **Problem:** Calculate the total number of cases per region.
 - **Example:**
 - **Input:** Case counts by region.
 - **Output:** {"Region A": 15000, "Region B": 20000}
-

15. Movie Genre Popularity

- **Dataset:** [IMDB Datasets](#)
 - **Problem:** Count the number of movies per genre.
 - **Example:**
 - **Input:** Movie data with genres.
 - **Output:** {"Drama": 500, "Comedy": 300, "Action": 200}
-

16. Taxi Trip Analysis

- **Dataset:** [NYC Taxi Trip Data](#)
 - **Problem:** Calculate the total distance traveled per day.
 - **Example:**
 - **Input:** Trip records with distances.
 - **Output:** {"2024-01-01": 1000 miles, "2024-01-02": 1200 miles}
-

17. Daily Steps by User

- **Dataset:** [Fitness tracker logs with user IDs and steps count.](#)
 - **Problem:** Calculate the total steps taken by each user.
 - **Example:**
 - **Input:** Steps for "User123" [5000, 8000, 7000]
 - **Output:** {"User123": 20000}
-

18. Customer Service Call Duration

- **Dataset:** [Customer service call logs with timestamps, call duration.](#)
 - **Problem:** Find the average call duration per agent.
 - **Example:**
 - **Input:** Call durations for Agent A [10, 15, 12], Agent B [8, 10, 7]
 - **Output:** {"Agent A": 12.33, "Agent B": 8.33}
-

19. Email Spam Detection Analysis

- **Dataset:** [Ling-Spam Dataset](#)
 - **Problem:** Count the number of spam and non-spam emails.
 - **Example:**
 - **Input:** Email metadata.
 - **Output:** {"Spam": 450, "Non-Spam": 1050}
-

20. Food Delivery Times by Restaurant

- **Dataset:** [Food delivery records including restaurant names, delivery times, and orders.](#)
- **Problem:** Find the average delivery time per restaurant.
- **Example:**
 - **Input:** Delivery times for "Restaurant A" [30, 35, 40], "Restaurant B" [25, 20, 30]
 - **Output:** {"Restaurant A": 35, "Restaurant B": 25}