Here’s a detailed documentation of the task we performed, including an explanation of the results:

Cohort Analysis Documentation

Task Overview

The goal was to perform a **Cohort Analysis** to understand user retention based on their participation in group deals over time. This analysis groups users by the month they signed up and calculates retention percentages for their subsequent monthly activity. The results are presented in a CSV file for easy visualization and reporting.

Step-by-Step Execution

1. Understanding the Data

We identified the following tables and columns from the PostgreSQL database:

- **`users` table**:
  - **`id`**: Unique identifier for each user.
  - **`created_at`**: The signup date for each user.
- **`groups` table**:
  - **`id`**: Unique identifier for each group.
  - **`group_deals_id`**: Foreign key linking to the `group_deals` table.
  - **`created_at`**: The date users participated in group deals.
  - **`created_by`**: Foreign key linking to the `users` table.
- **`group_deals` table**:
  - **`id`**: Unique identifier for each group deal.

2. Cohort Analysis Methodology

Cohort analysis groups users by their signup month (`cohort_month`) and tracks their participation in group deals over subsequent months. This helps identify user retention trends and the effectiveness of group deals in engaging users.

3. SQL Query

The data was fetched using the following SQL query:

```
SELECT
```

```
    users.id AS user_id,
    users.created_at AS signup_date,
    groups.created_at AS participation_date
FROM
    users
JOIN
    groups ON users.id = groups.created_by
WHERE
    users.created_at IS NOT NULL
    AND groups.created_at IS NOT NULL;
```

4. Python Script

A Python script was developed to:

1. **Fetch Data**: Connect to the PostgreSQL database and execute the SQL query.
2. **Preprocess Data**: Convert dates into usable formats and calculate the number of months between signup and participation.
3. **Cohort Analysis**: Group users by their signup month and calculate retention percentages for up to 12 months.
4. **Save Results**: Export the final cohort analysis to a CSV file.

Results Explanation

1. Output Format

The results are saved in a CSV file with the following structure:

cohort_month	Month 1	Month 2	Month 3	...	Month 12
2024-01	100%	80%	70%	...	20%
2024-02	100%	75%	60%	...	10%

Columns Explanation

- **`cohort_month`**: The month users signed up (e.g., January 2024).
- **`Month 1`** **to** **`Month 12`**: Each column represents a specific month after the signup month.
    - **`Month 1`**: Percentage of users who participated in a group deal in the same month they signed up.
    - **`Month 2`**: Percentage of users who participated in a group deal in the second month after signup, and so on.

2. Key Insights

- The **retention rate** trends downward, indicating that fewer users participate in deals as time passes from their signup date.
- Retention rates for the **earliest cohorts** (e.g., January 2024) may highlight how new user onboarding and engagement campaigns have performed.
- **Comparison across cohorts** (e.g., January 2024 vs. February 2024) can show seasonal trends or the impact of marketing efforts.

Benefits of the Analysis

1. **Identify Retention Trends**: Understand how long users stay active after signing up.
2. **Evaluate Engagement Strategies**: Measure the effectiveness of group deals in retaining users.
3. **Plan Targeted Campaigns**: Focus on re-engaging cohorts with declining retention rates.