Trial Activation

Analytics Engineering Task

Business Problem

The product team wants to improve the efficiency of converting trialists into paying customers by creating better user experiences. To measure success, they aim to track the "activation" of trialists based on specific actions that indicate a successful trial. Initial analyses have identified the following:

Trial Activation Goals

- At least 2 *Shift created*
- At least 1 *Employee invited*
- At least 1 *Punch-In*
- At least 1 *Punch-In entry approved*
- At least 2 advanced features viewed (*Revenue*, *Integrations*, *Absence* or *Availability*)

An account that completes all *Trial Activation Goals* has achieved a *Trial Activation*.

Data

You've been provided with a raw dataset of behavioral events from users that started their trial between January and March of this year.
The dataset has the following table schema:

organization_id	**activity_name**	**activity_detail**	**timestamp**
int32	*varchar*	*varchar*	*timestamp*

The trial goals map to the following activity event structure:

Goals	**Activities**
At least 2 Shifts Created	*Shift.Created*
At least 1 Employee invited	*Hr.Employee.Invited*
At least 1 Punch-In	*PunchClock.PunchedIn*
At least 1 Punch-In entry approved	*PunchClock.Approvals.EntryApproved*
At least 2 advanced features viewed (Revenue, Integrations, Absence or Availability)	*Page.Viewed*

Task

Given the information provided to you above, please do the following:

1. Using data warehousing best practices, build an SQL-based model. Consider layers typically used in data warehousing. Please provide us with two marts sources:
    a. Trial Goals: Tracks whether an organization has completed each of the trial goals listed above.
    b. Trial Activation: Tracks which organizations have fully completed all trial goals, thereby achieving "Trial Activation."
2. Using the sources you built in (1), run some **basic** descriptive analyses using Python to inform the product team's decision-making. Consider commonly used product metrics you can derive from the data.
3. Suggest how you would test the data transformations to ensure data quality.

Please provide your solution in a **git repository** with your full code.