BI Developer Case Study

Part A

Context

The CRM manager of an online casino has run a CRM campaign and would like a report to assess the results of the campaign. A group of players, **Action Group A**, received a communication through an email while a second group of players, **Action Group B**, received a communication by SMS. A **Control Group** was also included in the campaign. The players in the control group did not receive any communication but are included in the analysis for comparison purposes. The objective of the campaign was to get players to make a deposit **within 12 hours** after the communication is sent.

Data relating to the players targeted in the campaign can be found in the `conversion` database in the tables described below. You can set up the database with all the tables and the data required for this case study by running the `conversion.sql` script on a local MySQL 5.6 server. Alternatively, you can load the data into a local ClickHouse database from the CSV files: `activities_sent.csv`, `deposits.csv`, `user_details.csv`.

Table: `activities_sent` (one row per activity per player)

Field Name	Data Type (MySQL / ClickHouse)	Description
`activity_id`	INT(11) / UInt32	A unique identifier for each activity (campaign). For this case study we are only interested in the rows with `activity_id=156`
`action_group`	VARCHAR(100) / String	The action group that the player being targeted by the activity is part of
`user_id`	INT(11) / UInt32	The unique identifier for the player that was targeted by the activity
`timestamp`	DATETIME / DateTime	The timestamp at which the communication was sent to the player

Table: deposits (one row per deposit made)

Field Name	Data Type	Description
user_id	INT(11) / UInt32	A unique identifier for the player making the deposit
amount	DECIMAL(12,2) / Float32	The amount of the deposit in €
timestamp	DATETIME / DateTime	The timestamp at which the deposit was made

Table: user_details (one row per user)

Field Name	Data Type	Description
user_id	INT(11) / UInt32	A unique identifier for the player (primary key)
country	VARCHAR(100) / String	The country of registration of the player

Requirements

You are required to write an SQL query to generate an aggregated report which summarises the campaign performance per action group.

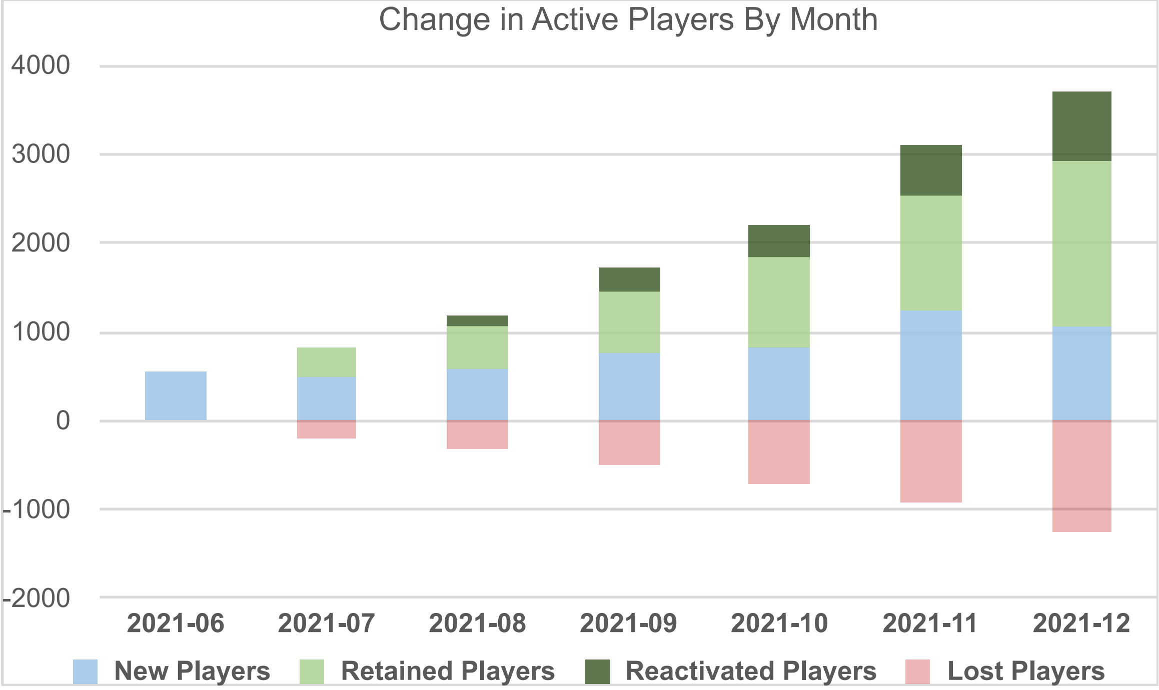
1. The report should only include data for **activity_id = 156**
2. The report should only consider players that are from the following countries**: United Kingdom, Norway, Finland, Sweden**
3. The report should contain the following measures for each action group:
    - **Players Targeted** – This is a count of players that were targeted by the campaign
    - **Conversions** – This is a count of players that made at least 1 deposit **within the first 12 hours after the campaign fired**
    - **Conversion Rate** – This is should be equal to Conversions / Players Targeted
    - **Average First Deposit Amount** – This is the mean of the **amount** of the converted players' **first** deposit (within the first 12 hours after the campaign)
    - **Deposit Count Per Conversion** – Count of **all** deposits made (within the first 12 hours after the campaign fired) / Conversions

- **Deposit Amount Per Conversion** – Sum of the amount of **all** deposits made (within the first 12 hours after the campaign fired) / Conversions
- **Single Biggest Deposit Amount** – The largest amount out of all deposits made (within the first 12 hours after the campaign fired)

Part B

Context

The CRM manager of a new casino would like a chart to show changes in the number of active players in her player base over the last seven months of 2021.



There are 3 possible states that the player can be in at any given point in time: *new, active, inactive*.

When the player registers, the player is given the *new* state. After one month, the player is moved either to the *active* or *inactive* state. The player can then move between the *active* and *inactive* states at any point in time. All the state changes for the player are recorded in the `movements` table. You can load this data into a local MySQL / ClickHouse database from the CSV file `movements.csv`

Table: `movements` (one row for each change in the player's state)

Field Name	**Data Type (MySQL / ClickHouse)**	**Description**
`user_id`	INT(11) / UInt32	The unique identifier for the player
`timestamp`	DATETIME / DateTime	The timestamp when the movement (change in the player's state) happened
`previous_state`	VARCHAR(100) / String	The state that the player was in prior to the movement happening – can be *null* or one of: *new, active, inactive*
`new_state`	VARCHAR(100) / String	The state that the player ended up in after the movement happened – can be one of: *new, active, inactive*

For monthly reporting purposes, the following definitions apply:

- **New Players:** All players that registered in a particular month, i.e they have a record in the movements table with a timestamp in that month and `new_state=new`
- **Retained Players:** All players that have their state being `active` at the end of a particular month and their state being `active` or `new` at the end of the previous month
- **Reactivated Players:** All players that have their state being `active` at the end of a particular month and their state being `inactive` at the end of the previous month
- **Lost Players:** All players that have their state being `inactive` at the end of a particular month and their state being `active` or `new` at the end of the previous month

Requirements

You are required to write an SQL query to get the aggregated data required to create a stacked bar chart similar to the one shown above. When writing your query, you should consider the speed of the query and whether it will scale when the casino grows and has millions of players and hundreds of millions of movements.