Junior Data Analyst @ WhatWapp

Introduction

Hello! Are you eager to be part of our Whatwapp team? This is the first step to show us your knowledge and problem solving skills.

This document is divided into three main sections: technical, analytical and problem-solving. Each section presents you with a different set of problems that you have to solve and to present in a readable format. Info on how the results should be presented are reported in each section.

Before commencing, here's a brief description of the sections:

- Technical: can you understand and write SQL? You will presented with data structures and some questions that you'll have to answer using SQL
- Analytical: data has to be interpreted before digging into it. Using a large set of events you will have to understand the requests and to present your findings
- Problem-solving: what if the solution is not straightforward? Can you find your way between multiple solutions and find the best one?

Technical section

Description

Consider the following table definition

```
CREATE TABLE events (
   event_id VARCHAR(36) NOT NULL PRIMARY KEY,
   event_time TIMESTAMP NOT NULL,
   event_name VARCHAR(50) NOT NULL,
   user_id VARCHAR(50) NOT NULL,
   iap_name VARCHAR(255) NULL,
   iap_price FLOAT NULL,
   game_name VARCHAR(50) NULL
);
```

Consider the following descriptions

Column name	**Column description**	**Example**
event_id	the unique id of the event	9074feb4-ac67-410d-85d7-1b57ff4db189
event_time	when the event occurred	2021-01-01T00:00:00.000Z
event_name	the name of the event	open, purchase, play
user_id	the id of the user that performed the event	176236383372, 443921755658
iap_name	the name of the In-App-Purchase that the user purchased	bundle_1, 30_coins
iap_price	how much the user spent to purchase the iap	2.99, 1.99, 0.99
game_name	the name of the game that the user is playing at	poker, bingo, scopa, belote

Consider the following events

Event name	**Description**
open	This event occurs when the user opens the application. No additional info is provided

purchase	This event occurs when the user buys an In-App-Purchase in the store. The event contains the content purchased and the price paid
play	This event occurs when the user starts a game and begins playing. The event contains the name of the game played

Consider that:

- the table contains all the events generated by the application between 2021-01-01 and 2021-03-31
- every day users open the app, play games and buy IAPs
- IAPs are: bundle_1, bundle_2, coins_1, coins_2
- IAP prices are:
    - bundle_1: 1.99
    - bundle_1: 9.99
    - coins_1: 0.99
    - coins_2: 2.99
- games are: scopa, belote, bingo, burraco

Deliverable

The solutions to the following questions should be provided in a readable format. Any comment useful to understand the query is well accepted.

Questions

Write one SQL query for each of the following questions:

- What is the number of unique users that use the application every day?
- What is the daily number of users that open the app more than 10 times?
- How are the top 10 users that spent the most buying IAPs during the three months of data?
- How many users played in the two most played games?

Analytical section

We are collecting events from one of our apps and we want to understand how much the users are engaged. The app allows the user to play a match with other players and to purchase items from the store.
We track:

- when the user opens the application
- when the user starts or leaves the match
- when the match ends
- when the user performs an in-app-purchase (iap) and what they purchase

Events

Below is the description of the events that we collect.

Name	**Description**
open_app	The user opens the application
start_match	The user starts a match
leave_match	The user leaves the match before it ends
end_match	The match the users is in ends (it might be missed if the user closes the app)
iap	The user successfully performs an iap

The following table reports the properties that events have:

- 4 properties that are present in all events and that can never be null nor empty (user_id, event_id, even_ts, event_name)
- 2 properties that are specific to the "iap" event

Event name	**Property**	**Description**
<all>	user_id	The unique id of the user
<all>	event_id	The unique id of the event
<all>	event_ts	The timestamp of the event
<all>	event_name	The name of the event
iap	sku	The item bought by the user

iap	sku_price	The price of the item

NB: Take into account that the application might send the same event multiple times and that some events might be missing

Datasets

The events are collected into two datasets – retention and purchases:

- retention: contains all the collected events along with only the common properties
- purchases: contains iap events with the event-specific properties

Please consider that the datasets contain all the events collected from the app between September and October.

Data access

Datasets are stored in an AWS Redshift database and exposed through a web interface at https://analyst.jobs.whatwapp.io that can be accessed using the credentials that you have received. With this tool, you can perform SQL queries on the three datasets in the "SQL Lab" section. You can also export the data in CSV format.

Deliverable

The result of the analysis should be presented in the form of a readable document or a set of slides. The solutions should be provided with a brief explanation of both the request and the result and with anything else that can help to understand the results, such as graphs or comments.
Any assumption or any incongruence you may notice in the data set please report it along with the analysis.

Questions

This test has a total of three main assignments, described in the following sections, that are divided into multiple questions. For each assignment you are provided with the name of the datasets that you can use.

1. User activity

We want to understand how many users are using our app and how they are using it between the beginning of September 2019 and the end of October 2019.

Datasets:

- retention

- purchases

Provide a solution and an interpretation of the results of the following questions:

1. How many users are using the app every day?
2. How much do the users spend buying iap every day?
3. How much did the top 10 spenders spend through both September and October?

2. Retention

We want to understand how long a user keeps using our app since he first joined. The metric we want to use is **retention**.

We consider the retention on day N as the percentage of users that use our app on the Nth day of their life. For example, considering a group of users that joined on the 1st of September, we measure how many of them still use our app on the 2nd of September, on the 3rd of September and so on and we compare it to the users on the 1st of September. As reference, the following table shows an example of retention values of users that joined on a specific day.

Days	**Day 0**	**Day 1**	**Day 2**	**Day 3**	**Day 4**	**Day 5**	**Day 6**
Users	1532	902	635	430	211	199	183
Calculation	1532/1532	902/1532	635/1532	430/1532	211/1532	199/1532	183/1532
Retention	100,00%	58,88%	41,45%	28,07%	13,77%	12,99%	11,95%

Datasets:

- retention

Provide a solution and an interpretation of the results of the following questions:

1. How many users joined on each day of September?
2. How many users that joined on the 15th of September used the app after 7 days?
3. For each day of September how many users joined and how many of these users opened the app again after 7 days?
4. Using the values from the question above, what is the retention rate for each day of September. What's its trend?

3. Purchases

We want to understand the trend of the purchases over September and October, also considering the "subscription_1" IAP. This is a subscription that guarantees some privileges to the user through the current month.

Datasets:

- retention

Provide a solution and an interpretation of the results of the following questions:

1. How many users purchase IAPs and how much do they pay every day in total?
2. What are the top 5 most purchased IAPs in September and in October? How much did the users pay in total to buy these IAPs?
3. Are the users buying more IAPs in September or in October?
4. How much do the users spend in total in their first week of playing the game?

Problem-solving section

The Candy Problem

You are a happy parent of 9 kids: Ander, Bonnie, Elijah, Farley, Freya, Lexi, Roxy, Tristan and Willow. Today all of them had an important exam (what are the chances - I know, right?) and the results were very different between every individual kid. Being a fair and just (but also a very kind and loving) parent you decide to reward their efforts by giving out some candies. Disciplined as you've raised them, the children are standing in line in the order demonstrated in the table below - their exam grade is in the rightmost column.

Order number	**Name**	**Exam grade**
1	Ander	30
2	Farley	46
3	Bonnie	90
4	Tristan	95
5	Lexi	85
6	Willow	70
7	Elijah	50
8	Roxy	46
9	Freya	40

You'd like to distribute the candies in a way to make sure that:

- each child must have at least one candy
- as a form of motivation, children with a higher exam grade should get more candies than neighbors children with a lower grade (for example, Willow must have more candies than Elijah, but less than Lexi - since Willow's exam grade is 70 and 85 > 70 > 50)

The order in which children are staying is fixed and unchangeable.

Question

What is **the minimum amount** of candies that you are going to need to satisfy the above mentioned requirements?
Describe your approach in finding the solution.