Kahoot!

Data & Analytics Team

Data Scientist (Analytics) Challenge

This challenge has two parts. In part 1, we would like you to write SQL queries to answer some business questions. In part 2 we would like you to analyse some user data and create a short presentation to explain your findings and conclusions.

Part 1

The below database table schema captures the historical state for every change a user makes to their profile. The current user profile can be found in the record with the latest created timestamp for any given user_id, but there can also be duplicated records.

Table: user_changes	
Column	**Description**
uuid	String: The unique record id
user_id	String: The unique id of the user
user_created	Timestamp: When the user was created
created	Timestamp: When this record was created
name	String: The full name of the user
email	String: The email address of the user
account_type	String: The type of account (teacher, student, business etc)

Write an SQL query to answer:

- ➔ *How many users signed up each month in each account type?*
- ➔ *How many business account users might be working at Apple?*

Please provide your queries in a plain text file.

Part 2

Please prepare a presentation which covers your responses to part 2. You should be prepared to present this to a small group of business and product stakeholders who you can assume have very little prior knowledge about this challenge and the data.

A: Who uses 'presentation mode'?

Attached is a CSV data extract (play_vs_fullscreen_by_primary_usage.csv) and a description of the data (Play vs Fullscreen Data Fields.pdf), containing 'Enter Fullscreen' and 'Play' event counts for teacher and social user account types by day. The 'Enter Fullscreen' event occurs when a user is hosting a Kahoot! and puts their browser into presentation mode. The 'Play' event occurs when a user hosts a game and the first question ends. Typically, there are zero or more 'Enter Fullscreen' events for each 'Play' event.

- ➔ *What is the average daily number of events in the period i) for teachers, ii) for social users?*
- ➔ *Is there any trend in the total daily number of games played over the period? If so, what do you think might explain that?*
- ➔ *Is there any significant difference between social users and teachers number of games played in the period? How do you know if it's significant or not?*
- ➔ *Is there any significant difference between social users and teachers proportion of fullscreen events to games played in the period? How do you know if it's significant or not?*
- ➔ *How would you design a controlled experiment to find out whether changing the design of the fullscreen button results in an increase in its use? How long should we run it for?*

B: User Segmentation

Attached is a CSV data extract (users.csv) and a description of the data (User Data Fields.pdf), which contains an anonymised random sample* of users that created a kahoot or launched a game of kahoot during one week in May 2017.

- ➔ *What can we learn about our user segments in this data?*
- ➔ *What problems could there be with basing an analysis on a small sample from a specific week?*
- ➔ *What improvements would you make to the analysis or the data used to help find important user segments?*

Please include in your presentation a section that summarises your findings and conclusions.

* The sample rate is about 1%

Notes

- ➔ Make it clear if you have made any assumptions about the data/problems
- ➔ Please share your slides in PDF format