Xero/Hubdoc Log File Take-home Assessment

Background

Xero's Hubdoc customers interact with the Hubdoc web application in a variety of ways which are captured in user logs. Analyzing these logs can provide us with insights on how our customers use our system as well as instruct actions we can take to improve our user experience and reduce customer churn.

Your task is to extract, transform and load the Canadian 2019 log files provided such that you are able to efficiently analyze the data to provide insight into the customer profiles and behaviors described below. You can accomplish this using any techniques and tools you choose. Some things to think about when you do this exercise.

- The more '*production ready*' your solution is the better.
- Include instructions on how to run your solution.
- Document your assumptions and artifacts so we can understand what you were trying to do.
- Timebox this exercise and if there are things you would like to change or know could be done better and did not have the time then document them. Typically 2 - 4 hours is a reasonable amount of time.

Customer Churn

A customer is deemed to have stopped using Hubdoc if they have not logged into the system within thirty days.

- Provide the total number of customers who have stopped using Hubdoc month by month. If a customer's last login was February 10th, 2019 and we see a thirty day period where no additional logins took place then we would say that customer has stopped using Hubdoc in February. For the purpose of this exercise once a customer has been inactive for 30 days they will not appear in the logs again. Assume that you are doing this reporting as of January 1st 2020.

Customer churn rate is defined as the number of churned customers over a given period divided by the number of active customers at the end of a given period.

- Provide monthly customer churn rate from January 1, 2019 to December 31, 2019. Assume that you are doing this reporting as of January 1st 2020.
    - Example: January:0.03, February: 0.05 ...

Plan Upgrade and Downgrade

It is important to understand what plans our customers are signing up for initially.

- Provide the total number of customers who selected a plan when they signed up and what percentage of total signups that represents. Calculate this for all of 2019.
    - Example: assuming we have 2 plans: X. and Y
        Plan X: 400 signups, represents 40% of total signups.
        Plan Y: 600 signups, represents 60% of total signups.

We also want to understand how customers change plans after initial sign up.

- Provide a breakdown of how many customers changed from their initial signup plan to another plan and the average number of days that customers took to make that change.
    - Example: assuming we have 2 plans: X. and Y
        Plan Y to Plan X: 80 customers. Average time to change: 50 days.
        Plan X to Plan Y: 0 customers. Average time to change: N/A.

Geographic Location

It is important for our sales and marketing teams to know the geological demographic of our customer base as it provides insights on how to advertise and provide support for the product.

- Provide the number of *active* customers in each province at the end of the year.
    - Example: AB : 84, PE : 8, ON : 224...