Vimeo – Data Analyst, Product – Take Home Exercise

Overview

The 2 datasets below represent visit and viewing data for a small sample of Vimeo users in a week. Download the raw data files here:

Sessions: https://dl.dropboxusercontent.com/u/42522/sessions.csv?dl=1
Plays: https://dl.dropboxusercontent.com/u/42522/plays.csv?dl=1

Descriptions and definitions for each dataset are below. Please review the data and provide responses for the following questions. Analysis should be done using SQL or other data analysis language (R, Python, etc), and code should be included with responses. Please contact andy@vimeo.com if you need any clarifications as you work.

Questions

1. What is the median viewing time (in seconds) for desktop users?
2. What percent of total viewing time do the top 10% of viewers comprise?
3. What are notable behavioral differences between users based on device type?
4. How would you define heavy/medium/light viewer segments based on device, viewing time, number of views, and/or number of categories viewed?
5. Describe the process by which you would identify any outliers in these datasets.

Definitions

sessions.csv – each record represents one session (i.e visit) on vimeo.com

vuid : unique user ID
session_id : unique session ID
device : device type

plays.csv – each record represents one video playback event

vuid : unique user ID
seconds_viewed : total seconds viewed during playback
category : video category name