DSA Take Home Project

David Karp, Tumblr's CEO, has asked you, the newest member of our Data Science Analytics team for your help in understanding user retention. To help explore this question, we have provided a sample dataset of users who signed up for Tumblr in early September 2015. A few days later, we pulled another activity dataset to see if those users who had previously signed up were had returned and logged in on that day. If so, we consider them as "retained users". See below for a detailed description of the dataset:

Dataset: User Registration <link here>

regi_ts	unix timestamp when users signed up for Tumblr
user_id	unique identifier of a Tumblr user
regi_geo	the geo in which a user registered
is_verified	1 if the user verified their email address, otherwise 0
pageviews	number of pageviews the user had during their first 24 hours
follows	number of blogs the user followed during their first 24 hours
likes	number of likes the user issued during their first 24 hours
reblogs	number of reblogs the user created during their first 24 hours
original_posts	number of original posts the user made during their first 24 hours
searches	number of searches the user performed during their first 24 hours
unfollows	number of blogs the user unfollowed during their first 24 hours
received_engagments	number of likes and reblogs the user received during their first 24 hours

regi_device	the device the user used for signup: iphone, android, or web
regi_source	the marketing campaign the user saw before sign up

Dataset: Active Users

ts	unix timestamp marking the first time a user logged in with a device on a given day
user_id	unique identifier for a user
active_device	the device on which the user was active

We would like you to use these two datasets to help understand the important factors for our user retention by responding to these prompts:

1. Perform whatever exploratory analysis you think is most appropriate to understanding the data and gathering an initial set of insights from it. Please be sure to include these data exploration steps in your code.
2. What percentage of registered users were retained? What are the most important factors that appear to be affecting user retention? What are the implications of those factors affecting retention? What actionable insights can be derived from this exercise?
3. Prepare a slide presentation to share your results as if you were presenting the results of your analysis to Tumblr's executive team and senior product managers. Feel free to include charts, tables, or any visualizations that would help convey the results of your analysis.

Once you have had the chance to complete this assignment, please submit both your code you've used for your analysis as well as the presentation you've prepared via the Greenhouse link, and also in an email to data-science-analytics@tumblr.com with a subject line of "Take Home Submission -- {Your Name Here}".