

Answers

(1) What is your one sentence executive summary?

- People read most on Entrepreneurship, Growth Hacking, Web Marketing, Motivational and Economic History and have interest in content types like News, List, Summary, Blog Post and How-To.

(2) What is your detailed assessment (for a technical audience)? Please quantify, use technical jargon.

- Users like to read short articles like a list, summary, how-to, etc and stuff which is currently trending like news. Also, they like to read on topics from which they can gain something like entrepreneurship, growth hacking, motivation, etc.
- All the bar charts further assure my assumption.

(3) What tools did you use?

- I used SQLite, IPython Notebook, Pandas library for data processing and Matplotlib for visualizing the data in form of bar charts.

(4) What techniques did you try?

- First, I created 6 different csv files from each of the tables in the given dataset and the log data.
- Then I used IPython and its library Pandas to load these csv files into Dataframes which are very easy to manipulate. Also, performing on the view is faster than performing on the real copy of the records.
- After slicing the Dataframes into individual Series and calculating the frequency count on them I was able find the top n-most valuable data from them. (I have shown 10-most valuable data in my code)
- With the help of Matplotlib and n-most valuable data, I plotted a bar chart using them to have a clear picture.

(5) What three plots did you make to best explain the data?

- Top 10 topics getting highest number of clicks vs # of clicks per each topic (Bar chart)
- Top 10 types getting highest number of clicks vs # of clicks per each type (Bar chart)
- Top 10 users contributing highest clickable content vs # of clicks per each user (Bar chart)

(6) What is your commercial recommendation for business unit heads who are nontechnical?

- Maximum users are likely to read on topics given below. My advice is to include these topics in the email digest.

Growth Hacking
Web Marketing
Motivational
Economic History
Advertising
Leadership
Budgeting
Biography
Consumer Behavior

- Maximum users read links of the types mentioned below. My advice is to have the content of the following type.

News
List
Summary
Blog Post
How-To
Podcast
White paper
Comment
Webinar
Data and Statistics

- Below users contribute rich content which other users are likely to read. My advice is to include as much content from these users as possible.

kevin.montez@icedh.com
dennis.hooker@fvrya.com
jason.williams@ygfbg.com
charlie.hansen@juzai.com
clifford.faust@xnoid.com
maximo.lindsey@eytkr.com
misty.labelle@crorv.com
kathryn.myers@qbmso.com
rebecca.matthews@xdata.com
joseph.gallagher@phjue.com

(7) What other data would you like to see about the platform? What questions would this additional data help you answer?

- I would like to have a more accurate log file if possible. Because, I had to manually exclude all the links with article id greater than 17029(maximum id of an article). I assume all these links are erroneous.
- I would also like to have the name of articles in addition to article id so that I can find the top articles read by users and reference it by names in future.