

Hacker Rank Content Analysis

Download: [Jupyter Notebook file, ReadMe, hacker_news.csv](#)

Abstract: 20,000 rows of Hacker News dataset from [Kaggle](#), reduced from almost 300,000 rows (by removing all submissions that did not receive any comments, and then randomly sampling from the remaining submissions) that includes 7 features.

Data Set Characteristics:	Multivariate	Number of Instances:	20000	Area:	Technology
Attribute Characteristics:	N/A	Number of Attributes:	7	Missing Values:	0

Source:

<https://www.kaggle.com/hacker-news/hacker-news-posts>

Data Set Information:

"This data set is Hacker News posts from the last 12 months (up to September 26 2016). It includes the following columns:

- title: title of the post (self explanatory)
- url: the url of the item being linked to
- num_points: the number of upvotes the post received
- num_comments: the number of comments the post received
- author: the name of the account that made the post
- created_at: the date and time the post was made (the time zone is Eastern Time in the US)

To view a copy of this license: <https://creativecommons.org/publicdomain/zero/1.0/>".

Data variables include:

App Store DataSet

Label	Data Type	Description
'title'	String	title of the post
'url'	String	the url of the item being linked to
'num_points'	String	the number of upvotes the post received

'num_comments'	String	the number of comments the post received
'author'	String	the name of the account that made the post
'created_at'	String	the date and time the post was made (the time zone is Eastern Time in the US)

Data Summaries Calculated:

- Mean
- Frequency

Units for Data Summaries and Variables:

- Number of Comments
- Hours (24-Hour Time)

Data Transformations:

- Data was downloaded from <https://www.kaggle.com/hacker-news/hacker-news-posts>
- File "hacker_news.csv" was imported into Jupyter Notebook.
- Removed headers.
- Converted values to lowercase
- Converted Date/Time String to datetime objects.
- Formatted time objects.
- Identified post type that receives the most comments on average.
- Identified top five hours to make comments on posts.