

Automating the Measurement of Cloud Providers' Popularity

Analysing Tech Trends on Twitter and Stack Overflow- Opsmorph

Ana Luísa Pires Fernandes | Ba (Hons) Politics, Philosophy and Economics

Overview of the Data Fellowship

Keeping up with the latest advances in cloud technology is imperative for businesses but this can be a lengthy process which requires expertise and data analysis. During my Data Fellowship I automated the measuring of the popularity of cloud providers by multiple metrics using the public APIs of two diverse data sources: Twitter and Stack Overflow. This data has been processed using Python with the Pandas package and visualised using Plotly. While the focus is on cloud providers, this process can scale to monitor trends in any technology.

Method

To web search Twitter and Stack Overflow we used the requests package and to read the returned text we used the json() method. On Twitter, our queries were based on the number of followers each cloud provider has and their average engagement per tweet which is a weighted average of likes, comments, retweets and quotes. On Stack Overflow, our query returned the total count of tags containing the name of the cloud provider. Accuracy and consistency was verified through the comparison of Twitter and Stack Overflow's query results (Fig.5), and the comparison between these two and an official survey done by Stack Overflow on the popularity of different Cloud Providers (not shown).

Twitter Presence vs Stack Overflow Presence

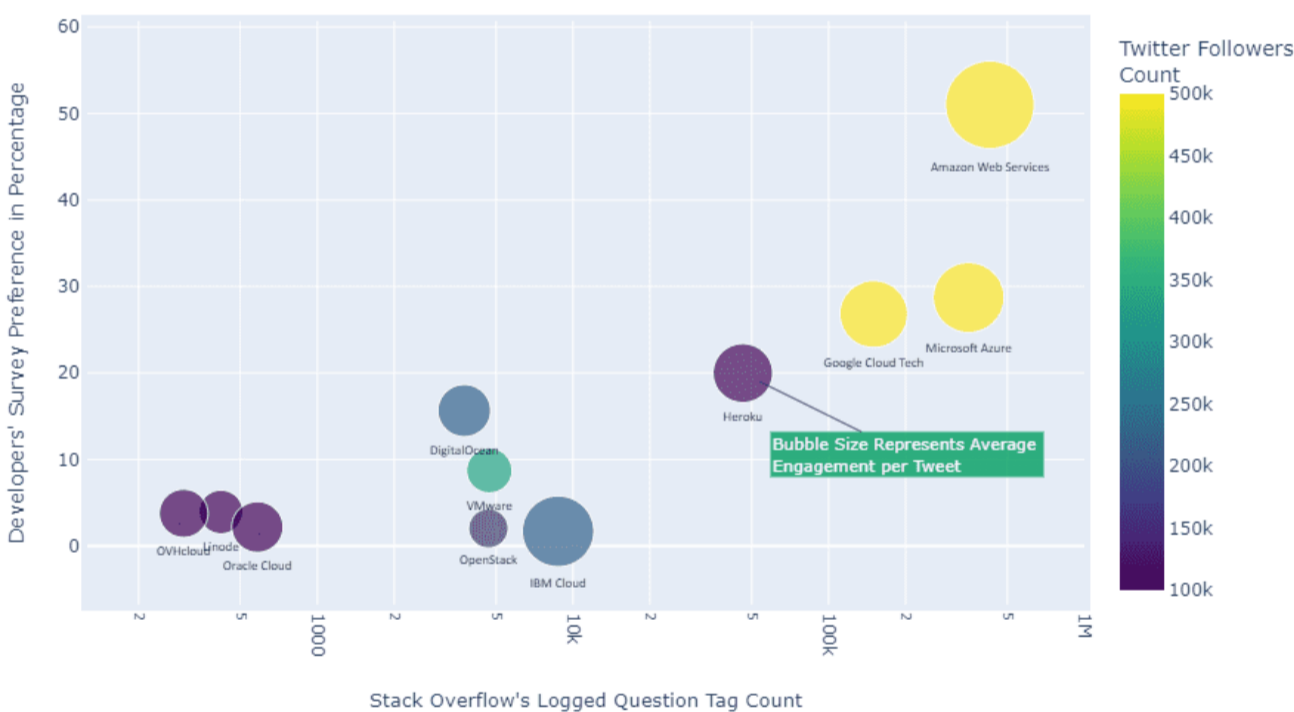


Fig 5. Developers cloud provider preference in percentage plotted against Stack Overflow's question tag count (logarithmic scale). The colour gradient represents the number of followers each cloud provider has in their official verified accounts, and the bubble size represents average engagement per tweet. Source: Twitter API and Stack Overflow API.

Findings

- The leading cloud providers are Amazon Web Services, Microsoft Azure, and Google Cloud Platform, with Amazon Web Services leading by a high margin by all metrics. The smallest providers as measured by all metrics are OVHcloud, Linode and Oracle cloud;
- Besides IBM Cloud's unusually large Tweet engagement, VMwares unusually low tweet engagement, and Heroku's unusually low number of Twitter followers, when compared to their performance on Stack Overflow the data seems to be consistent across cloud providers and data sources;
- There is a clear pattern of darker colours (indicative of a lower number of followers in Twitter) and smaller bubble size (indicative of a lower engagement per tweet) in the bottom left of **Figure 5**. Similarly, there is a pattern of lighter colours (indicative of a higher number of followers in Twitter) and a bigger bubble size (indicative of a higher engagement per tweet) in the top right.

Key Skills Learnt

- **Python Programming:** organising large amounts of data in data frames and dictionaries; manipulating data frames; using the requests package; coding *for loops* and *if else* statements; producing bar plots, scatter plots, line plots, and bubble plots;
- **Data Analysis:** summarising data; standardising data for comparison; finding patterns, dissimilarities and similarities; assessing biasedness of the data collection method;
- **Research:** thorough investigation of cloud providers services and technology; web search for coding errors' solutions; informed selection of representative Twitter accounts and Stack Overflow's tags;
- **Communication and Teamwork:** collaboration throughout the coding, research and production of results; feedback loops; requests for assistance and sharing of ideas;
- **Time Management:** working on different tasks related to the project while meeting deadlines for each part of the process.