

Supplementary information for

Exploring data analytics climate startups using BERTopic

Supplementary Note 1. Common morphemes removed.

Common morphemes removed from pitch lines, case-insensitive.

'developer', 'develop', 'developing', 'develops', 'developed', 'developers', '
', '</br>',
'company', 'company's', 'company.', 'company,', 'companys', 'solutions', 'platform', 'big',
'data', 'ai', 'ai-based', 'ai-driven', 'ai-powered', 'artificial', 'intelligence', 'technology', 'analytics'

Common morphemes removed from patent abstract, case-insensitive.

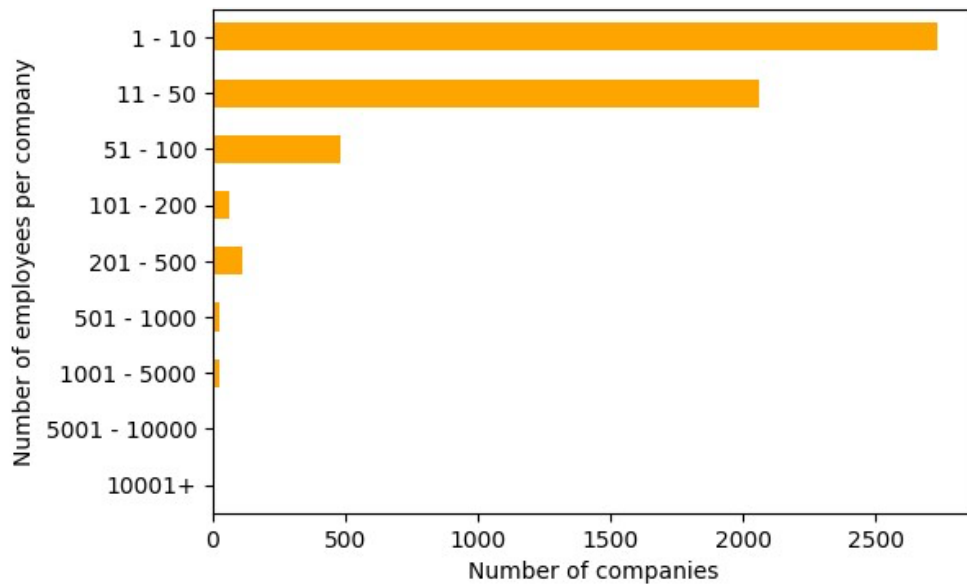
'developer', 'develop', 'developing', 'develops', 'developed', 'developers', '
', '</br>'

Supplementary Methods. The method to determine the state where US companies reside.

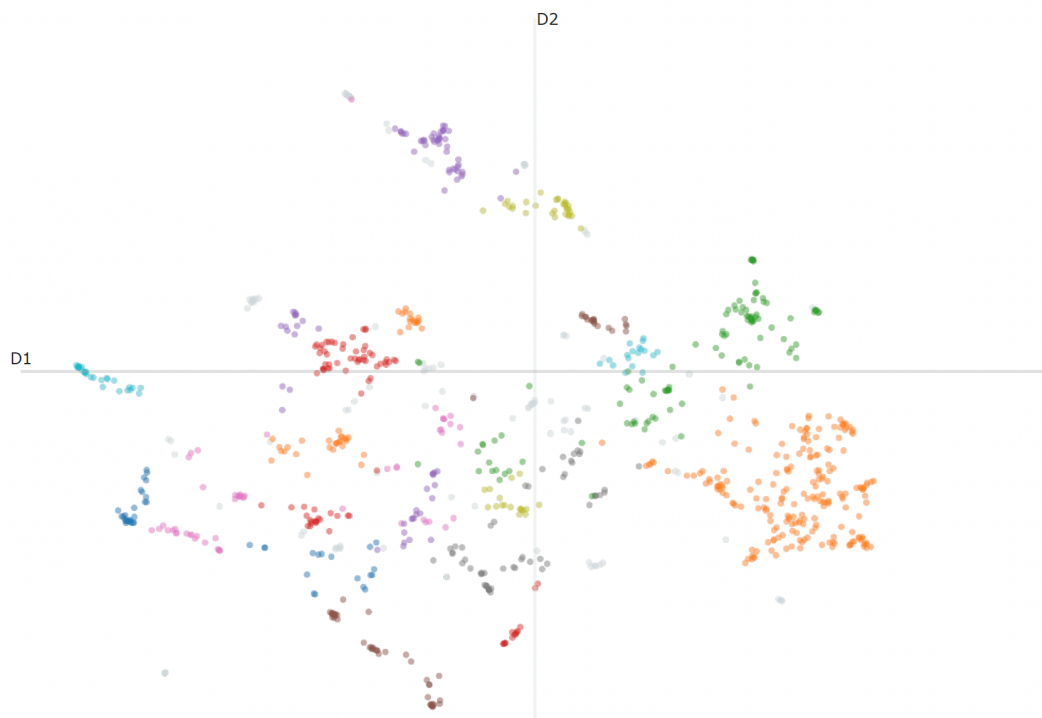
States are inferred from city names using Microsoft Excel Geography data type. When ambiguity occurs, the company's website and/or online database are checked.

The states were inferred from city names using Microsoft Excel's geography data type and cross-checked with online sources when the inference is ambiguous, where 27 companies miss city data, and 2 companies' states cannot be inferred.

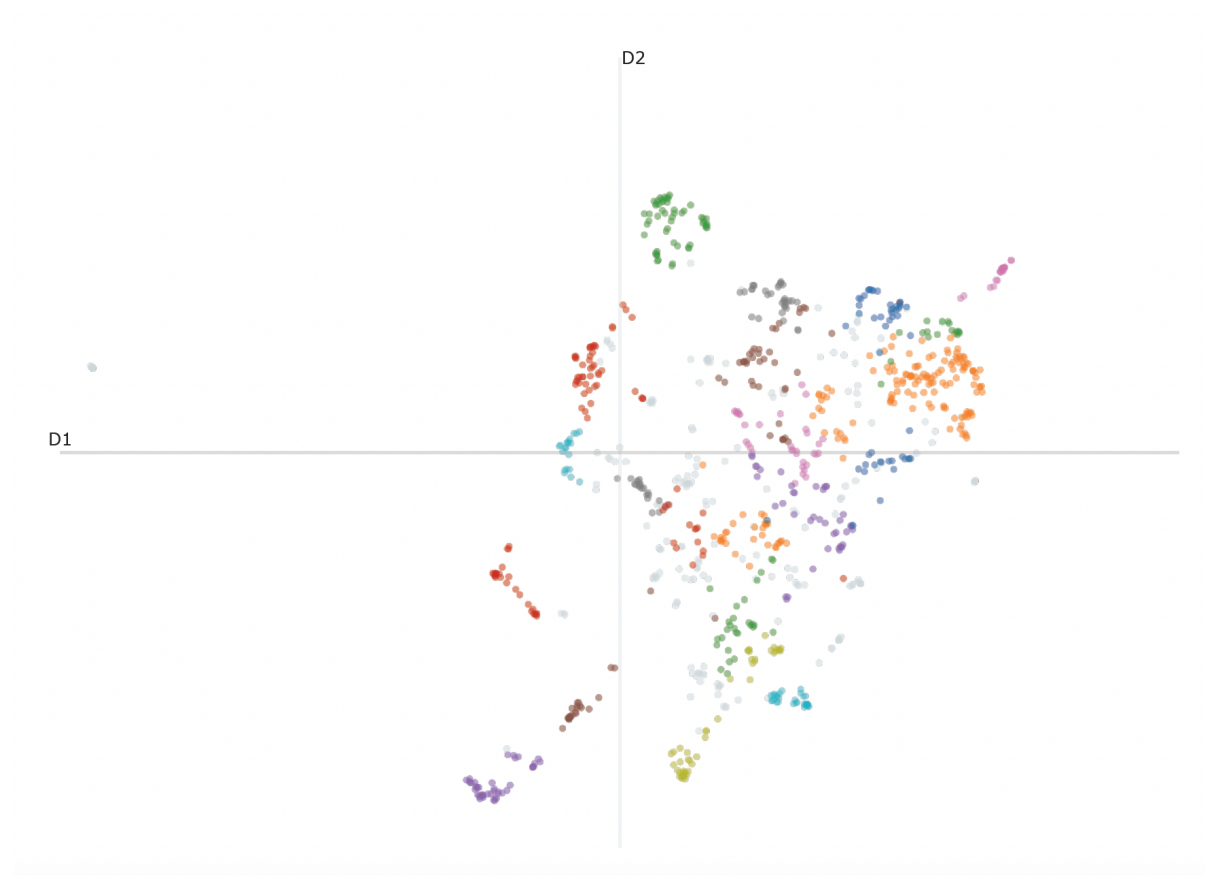
Supplementary Figure 1. The employee number distribution.



Supplementary Figure 2. Visualisation of data analytics climate startup patent clusters based on the abstract, using the all-MiniLM-L6-v2 embedding. The embeddings are recalculated and reduced to 2-dimensional space. Note that some clusters may share the same colour.



Supplementary Figure 3. Visualisation of data analytics climate startup patent clusters based on the abstract, using the PatentSBERTa embedding. The embeddings are recalculated and reduced to 2-dimensional space. Note that some clusters may share the same colour.



Supplementary Tables 1-8.

Attached as another Excel and PDF file.