WHAT ARE LARGE ORGANISATIONS HUNGRY FOR?

MIT MBAn CAPSTONE (Sponsor: McKinsey & Company)
Benjamin Lim, Rita Yuan | Mentored by Carine Simon, Chris McCord

About Us

We are the world's first data-science restaurant run by recovering consultants hailing from China, Germany, USA, and Singapore. Disclaimer: Our food does not contain any HiPPOs*.

Highest Paid Person's Opinion

Our Mission

To identify what is top-of-mind for large organizations using **topic modelling**, so as to lead knowledge acquisition efforts within McKinsey. Finding out what organizations care about helps us to **highlight knowledge gaps**. We also model relationships between different topics to **uncover cross-functional synergies** within the firm. To date, we have partnered with two Practices to derive insights using our tool.

The Ingredients

McKinsey-authored proposals

Searches and Downloads

Emails

Google Trends

Internal Knowledge Articles

External Publications

Our Recipe

Weeks 1-2

1) Processing: Melt documents to boil off any uninformative words and confidential information.
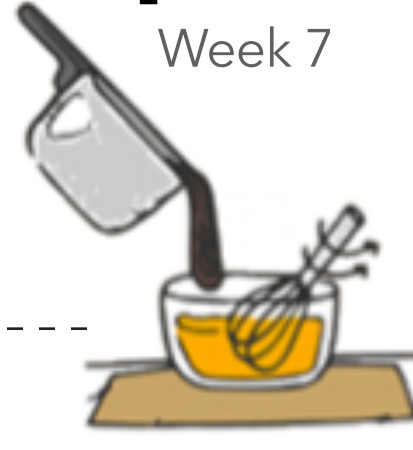
Weeks 3-4

2) Model Topics: Train and compare Latent Dirichlet Allocation, Biterm Topic Models and Correlated Topic Models.

Weeks 5-6

3) Label Topics: Apply auto-labelling algorithms to derive labels for topics and quantify their quality. Topics with low-quality auto-labels are manually labelled.

Week 7

4) Enrich Topics: Add metadata (the function, industry, and geography of a document) to allow for tailored analyses and cross-functional comparisons.
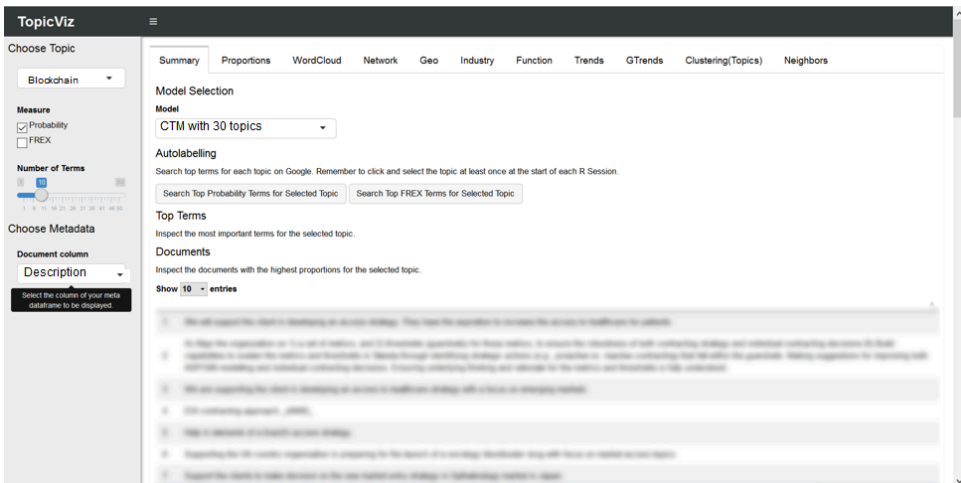
Weeks 7-8

5) Visualize Models: Build application for end-users to easily understand what each topic means, how documents are related, and explore how topics change across time and space.

Weeks 9-10

6) Derive Insights: Partner with specific Practices to build custom models and generate actionable insights.

Appetizers

Document Exploration
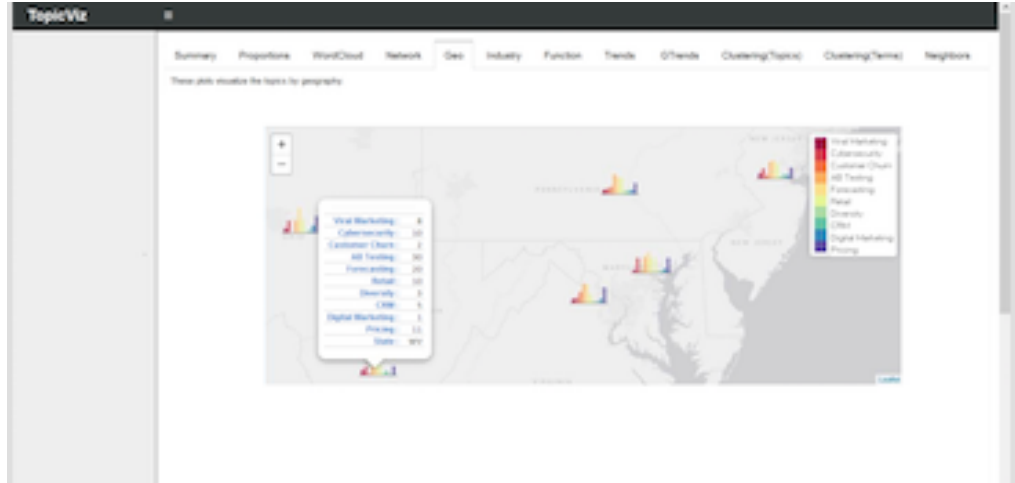We display the documents most representative of each topic.

Word Cloud
Words most representative of each topic are shown in a word cloud.
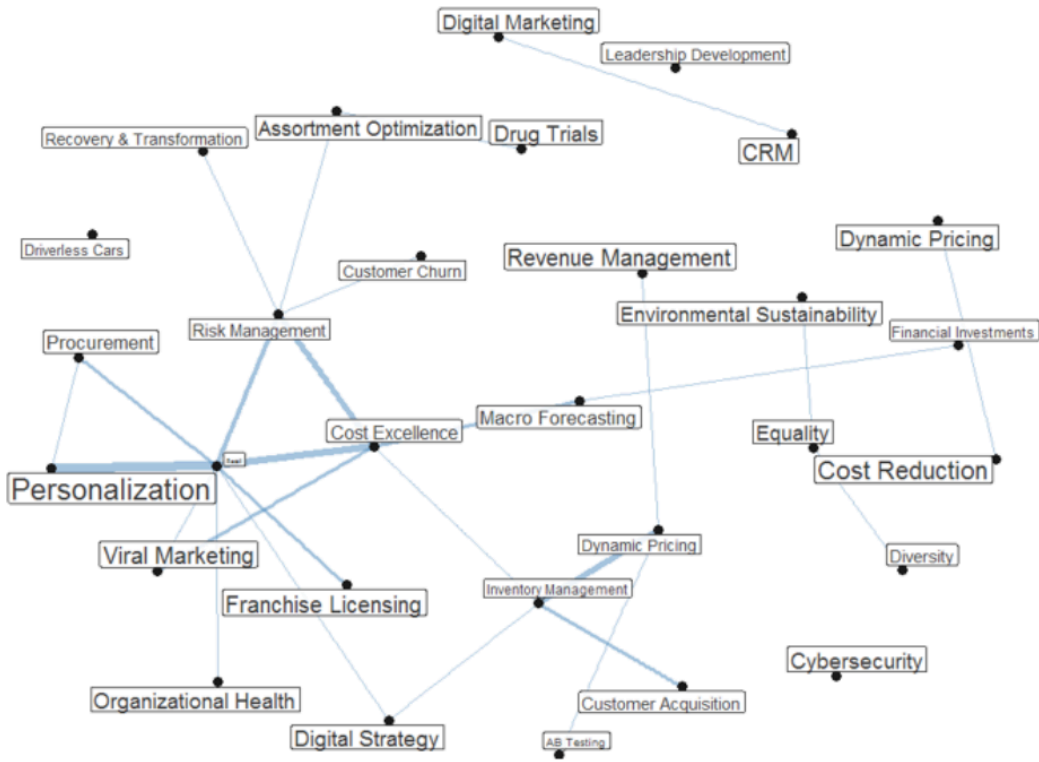
Network Analysis
Each document is a node, and the edge widths represent similarities between documents.
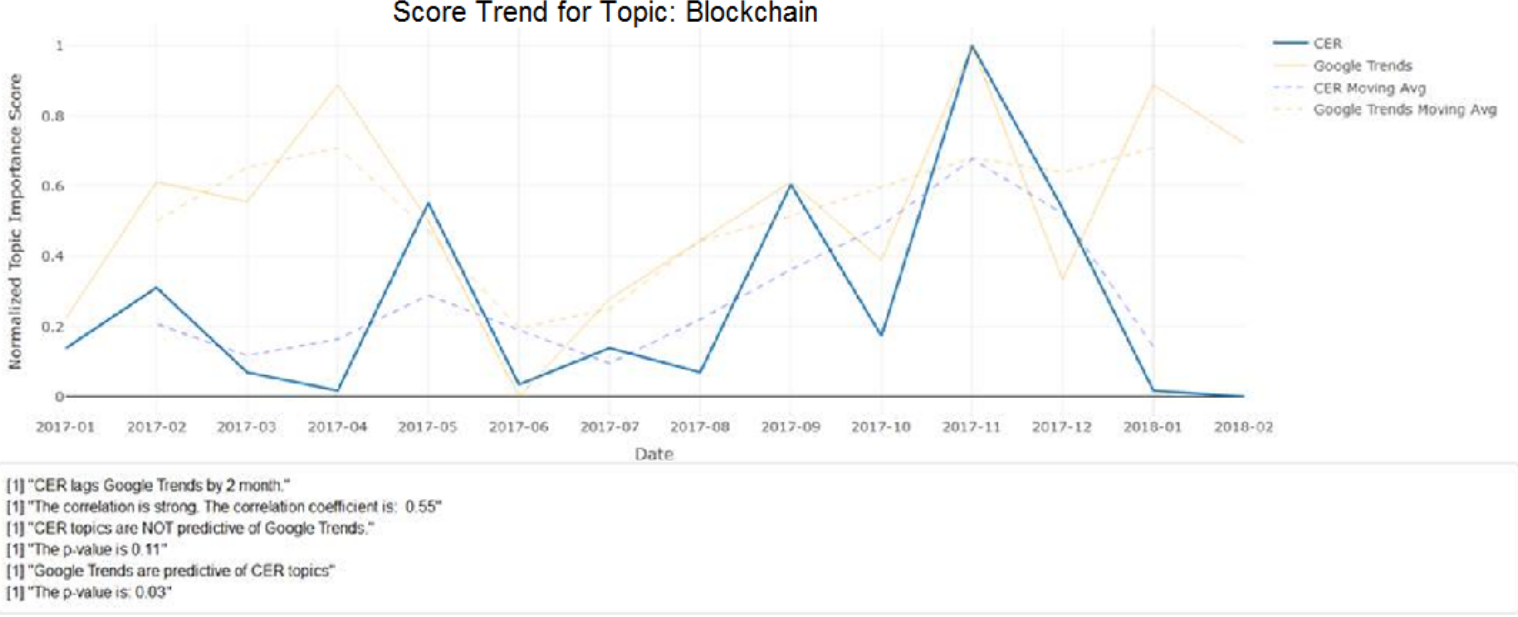
Geospatial Analysis
Interactive map showing how the composition of topics vary by region.
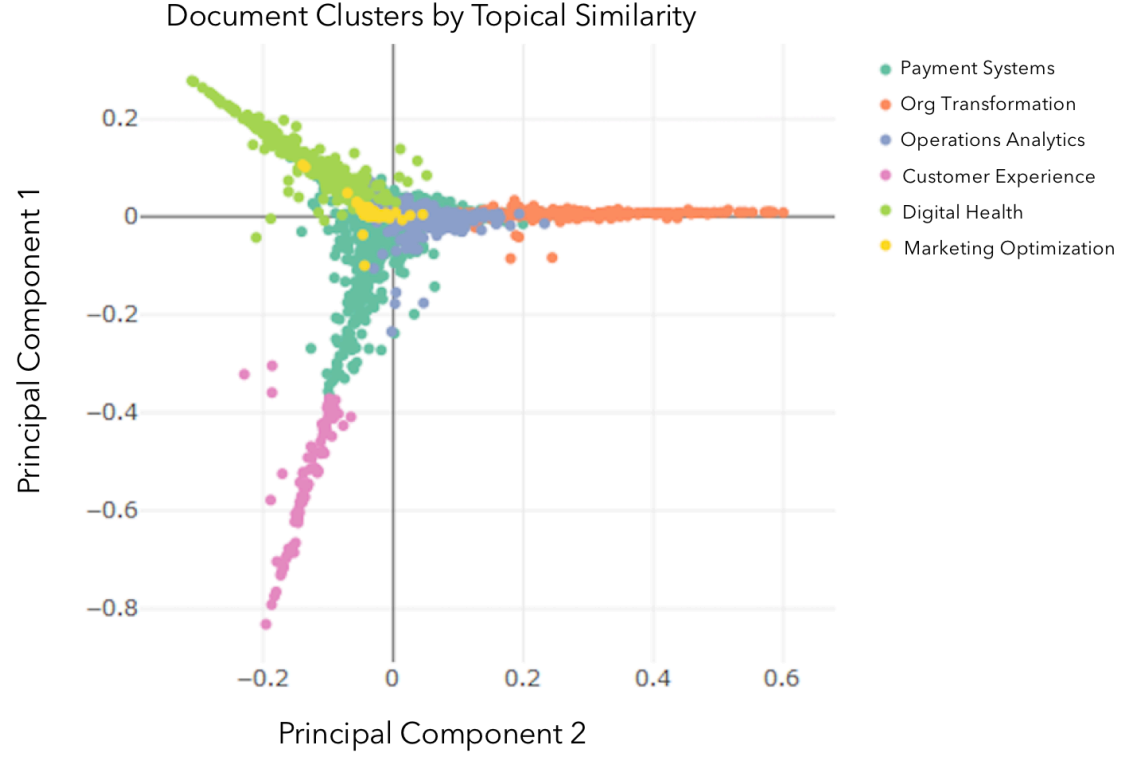
Chef's Recommendations

Topic Network
Topical relationships are shown in a network, where highly correlated topics have a thick edge.

Internal vs. External Signals
We run statistical tests to see if topical trends within the firm lead or lag topical trends from external sources.

Document Clusters
We perform K-means, Hierarchical and DBSCAN clustering on the documents to uncover tribes within the firm.

" The topic network helped my team **deliver better expertise** to my client by identifying correlated topics. For example, I found out that clients seeking solutions for Revenue Management were often want to better understand Personalised Advertising services.

Finding out that Google Trends closely tracked our internally trending topics allowed our Practice to use it as an indicator of **when and where to grow knowledge acquisition efforts**.

The clustering analysis was helpful in **facilitating knowledge-sharing**. It enabled me to find colleagues who worked on similar topics and allowed me to tap into their expertise. "

All graphics and quotes are purely illustrative for confidentiality reasons

Our Contributions

🍓 Designed **robust text cleaning** procedures that preserve topics while protecting client confidentiality

🍐 Built **reproducible topic models** for diverse data sources and defined methods for evaluating them

🍉 Created an **original heuristic that finds the optimal number of topics** for any topic modelling algorithm

🥝 Implemented **auto-labelling algorithms** that reduce the need for manual labelling by up to 45 percent

🫐 Developed an **app that facilitates easy topic analysis** across a wide range of business use cases

🍊 Partnered with two Practices within the firm to **operationalise our tool and derive actionable insights**