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

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


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.

4) Enrich Topics: Add metadata (the function, industry, and geography of a document) to allow for application for end-users to easily understand what each topic means, how documents are related, and explore how topics change across time and space.

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.

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

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.

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