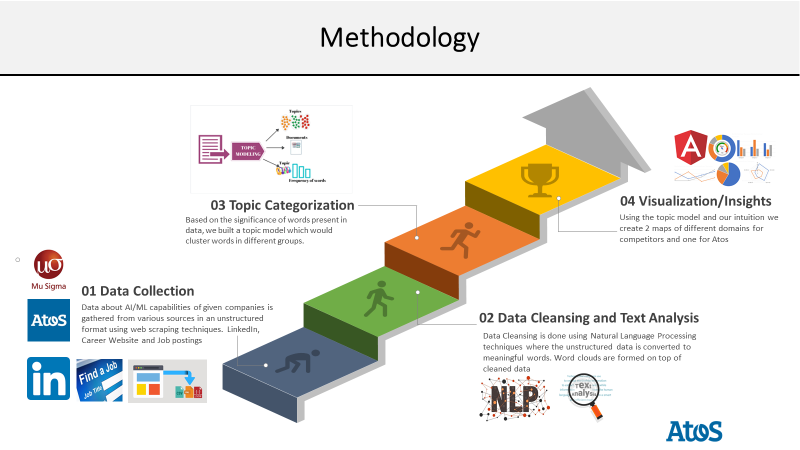
**Comparative Analysis of Data Analytics (including AI/ML) Offerings and Capabilities**

In this analysis of offerings and capabilities of the selected companies with ATOS we are throwing light on Artificial Intelligence, the next step in Human Evolution. The aim of the technical companies, these days, to create technology that allows computers and machines to function in an intelligent manner has made the topic of AI the most coveted one in the world of technology.

After a comprehensive study of the selected companies we have made a comparison of the companies across the industries on their artificial intelligence capabilities.



**(01) Data Collection**

**1. Web Scraping using Python - Company’s Website**

The career page of each company leads us to the findings related to what the firm had done in the past and the areas that they would want to explore in the future. The objective of every firm was very clearly captured. The main industries getting services that were found as a common catch among 5 firms were

1.Banking and Finance

2.Healthcare

3.Energy

4.Manufacturing and Retail

5.Technology and Media

These were the industries where all the firms are trying to conquer with their AI/ML capabilities.

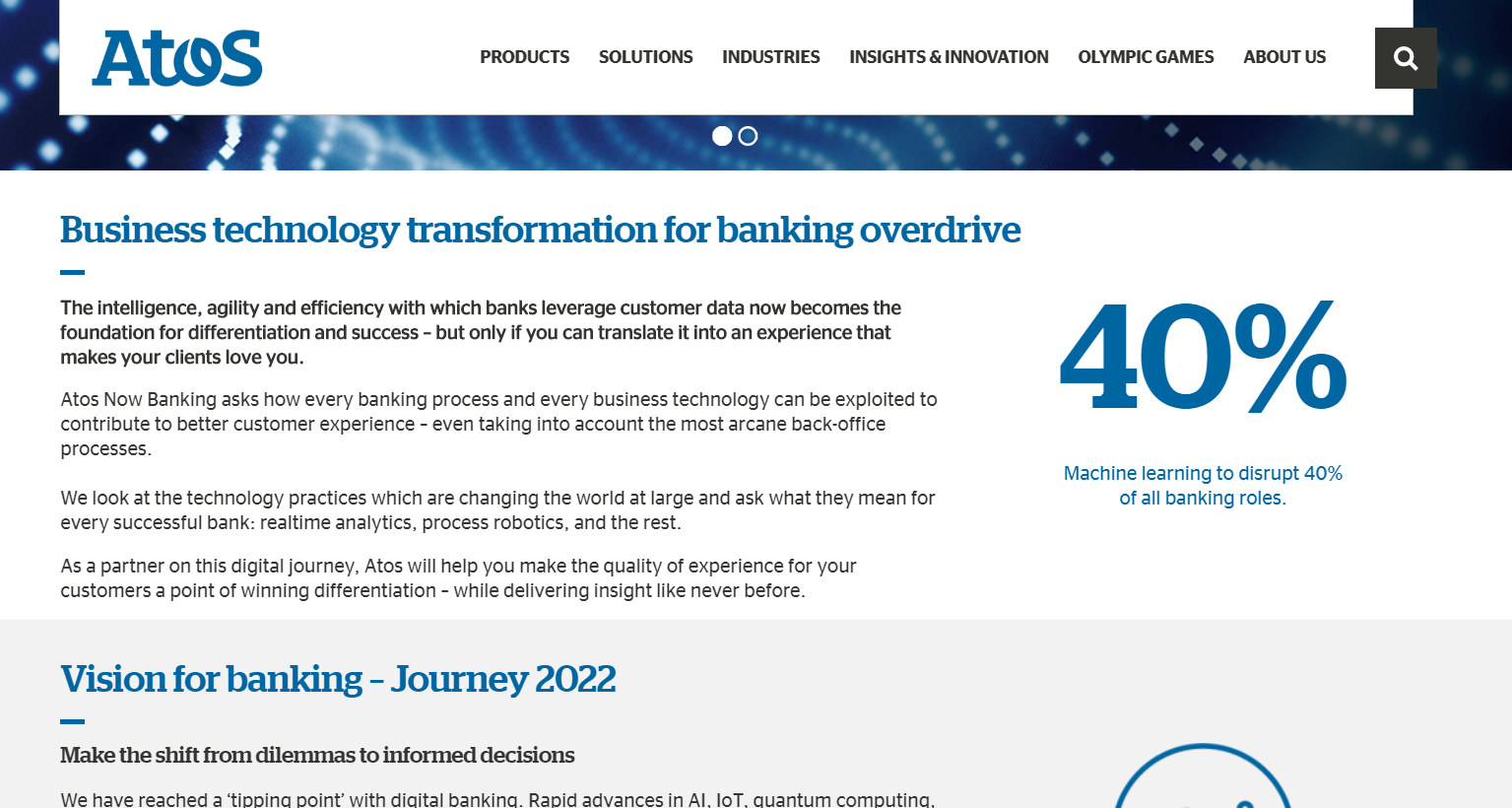


Figure 1

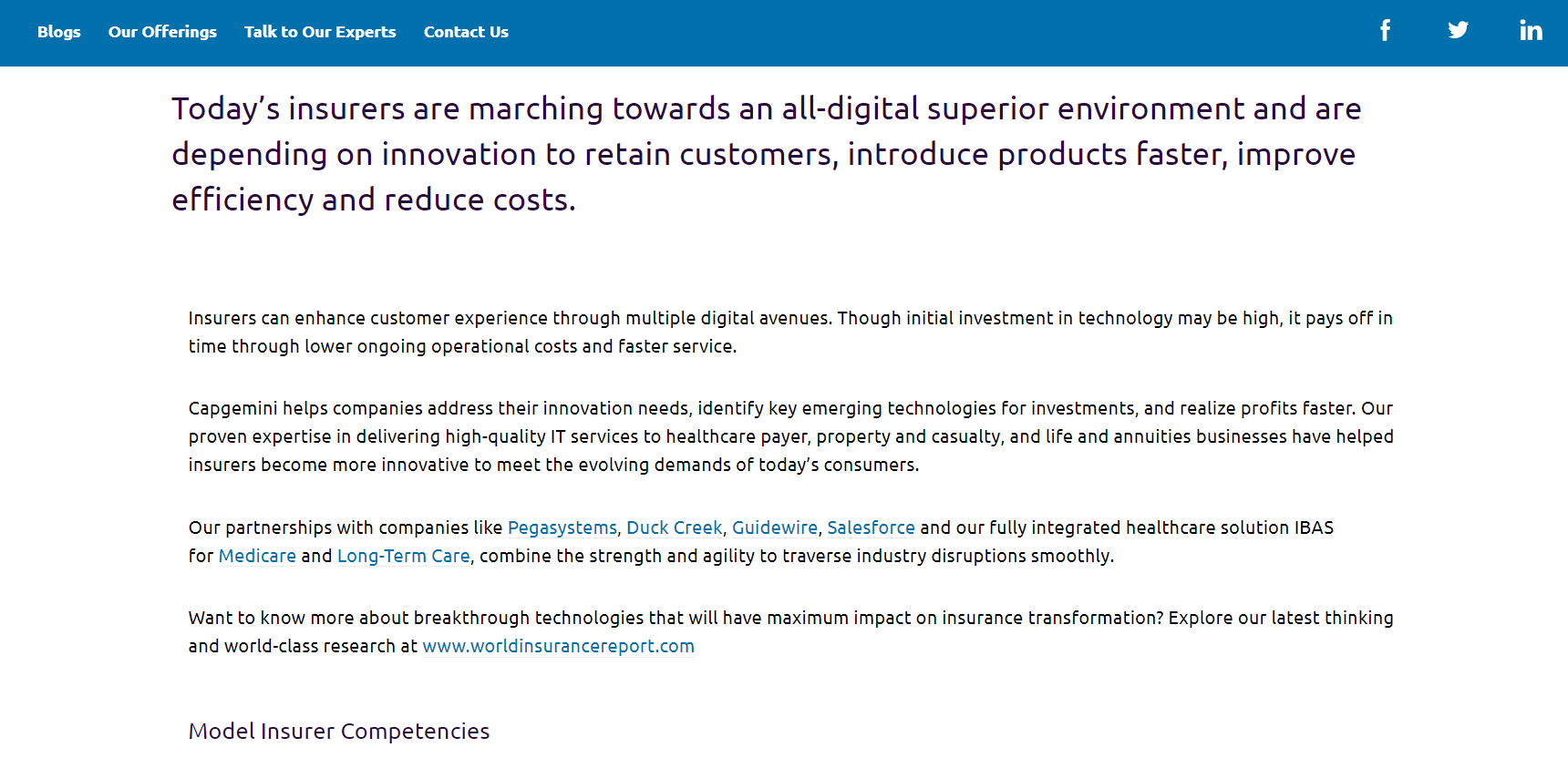


Figure 2

Text from these websites were scraped using a python script using selenium and beautiful soup packages. The data is in the form of flat files (.docx)

**2. LinkedIn**

LinkedIn Job postings related to data science and job description of given companies for the corresponding roles lead us to what the companies expect and the kind of project they are doing. Also, the information related to the tools and technologies that they used were captured

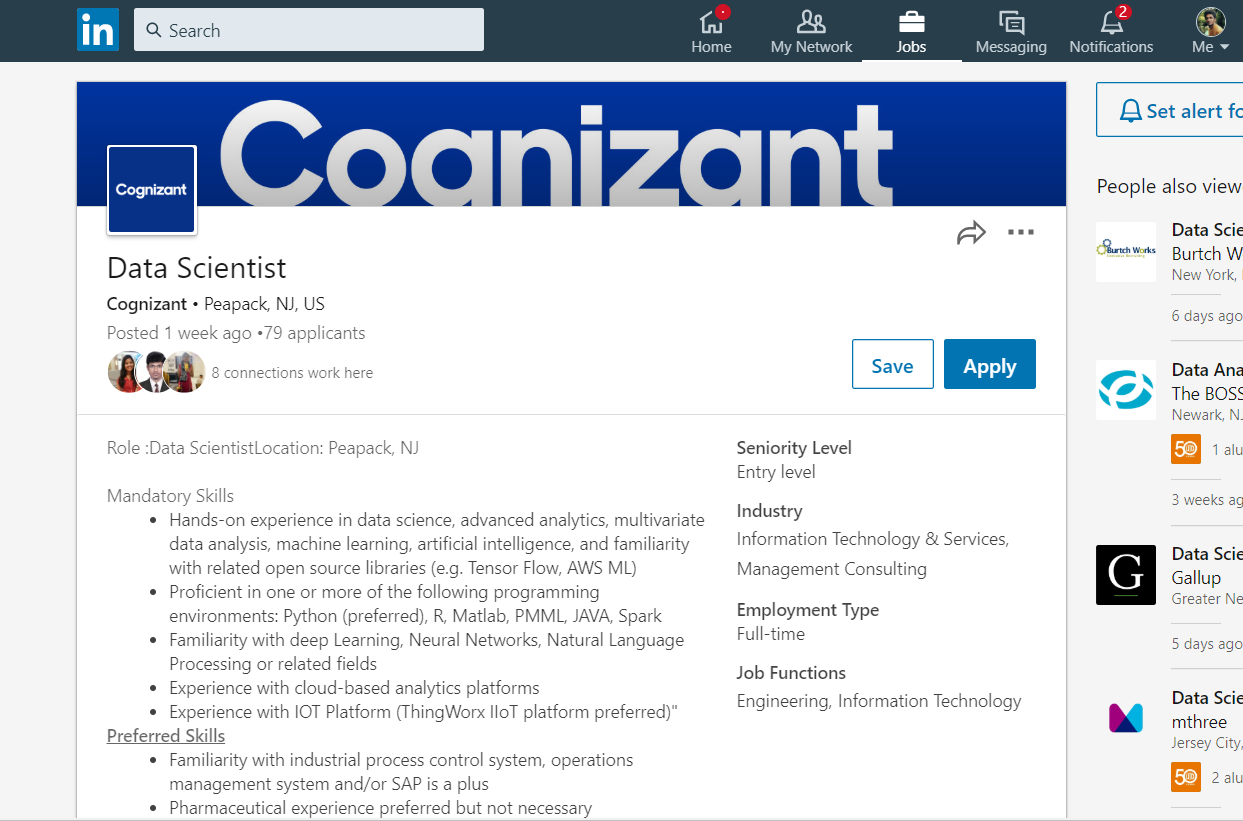


Figure 3

1. **Product names extraction**

The last mode of data collection was done manually. Specific product names and projects the firms had achieved were captured under the 5 mentioned domains to compare what the competitors are doing. The same was done for Atos as well. The end result is stored in an excel spreadsheet

**(02) Data Analysis and cleansing using Text analytics – Natural Language Processing**

1. **Preprocessing Text**

Once the data is collected, it is in an unstructured format. To do an analysis, preprocessing procedures were done using Python

1. Stop word removal
2. Stemming/ Lemmatizing
3. Punctuation removal
4. POS Tagging (Considering only Nouns)

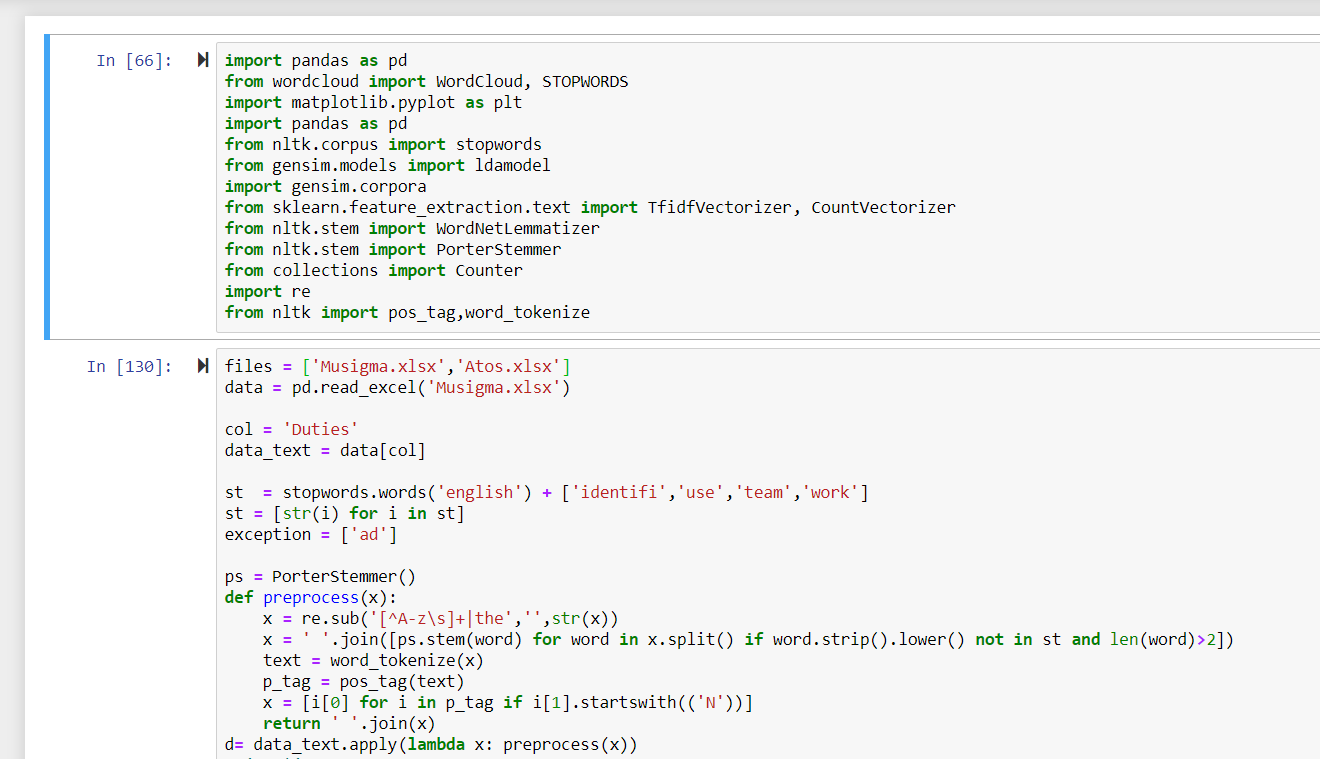
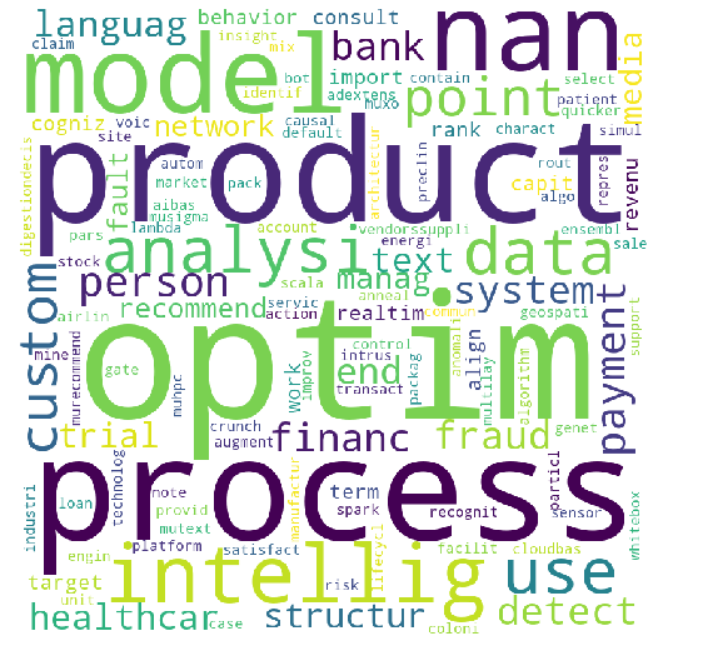


Figure 4

1. **Word Cloud**

Based on the preprocessed text, significant words are collected and a frequency Count vectorizer matrix is created to visualize the results in the form of wordcloud which is a nltk library to do text visualization. The bigger the size of the word, more the frequency. We can get an overall idea of what the company is upto using this.

(a)Word cloud of Competitor’s data



(b) Word Cloud of COMPANY X



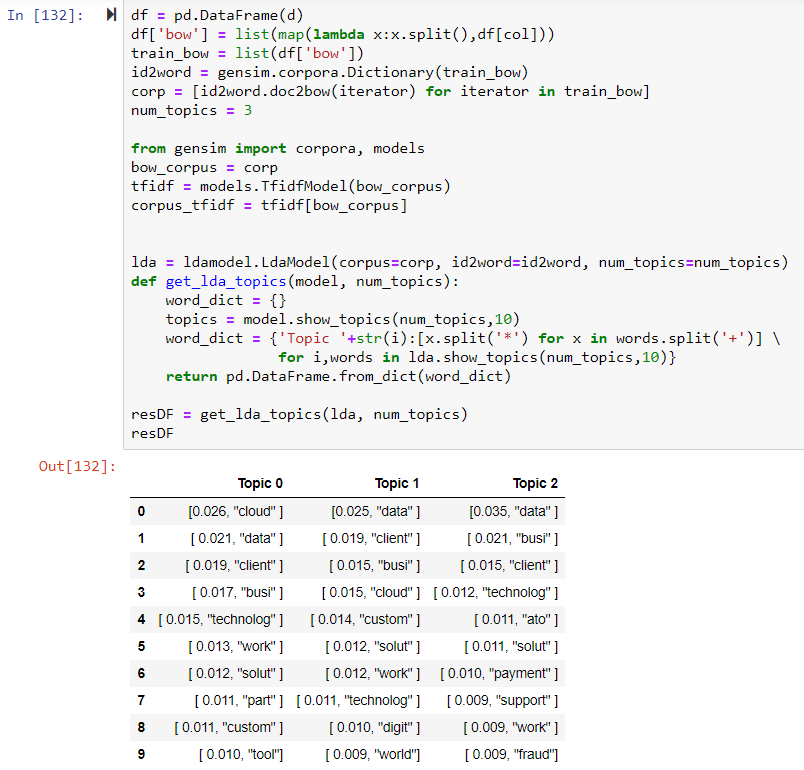
**(03) Topic Modeling**

Since the collected data is unstructured, we need to follow an unsupervised approach to arrive at insights. The data is converted to

1. Count vectorizer – The created Bag of words is used to pass it on to LDA topic model where it clusters the words into groups. As a result, we get 3 different topics and we get a set of 10 words under each topic.
2. TF-IDF vectorizer – The disadvantage of count vectorizer is overcome by tf-idf vectorizer. Here Inverse document Frequency matrix is calculated where significant words are given more weightage before passing on to LDA Model.

The output of topic model gives us 3 different topic for both Atos and Competitor’s data. It gives a generic idea of what the companies are doing

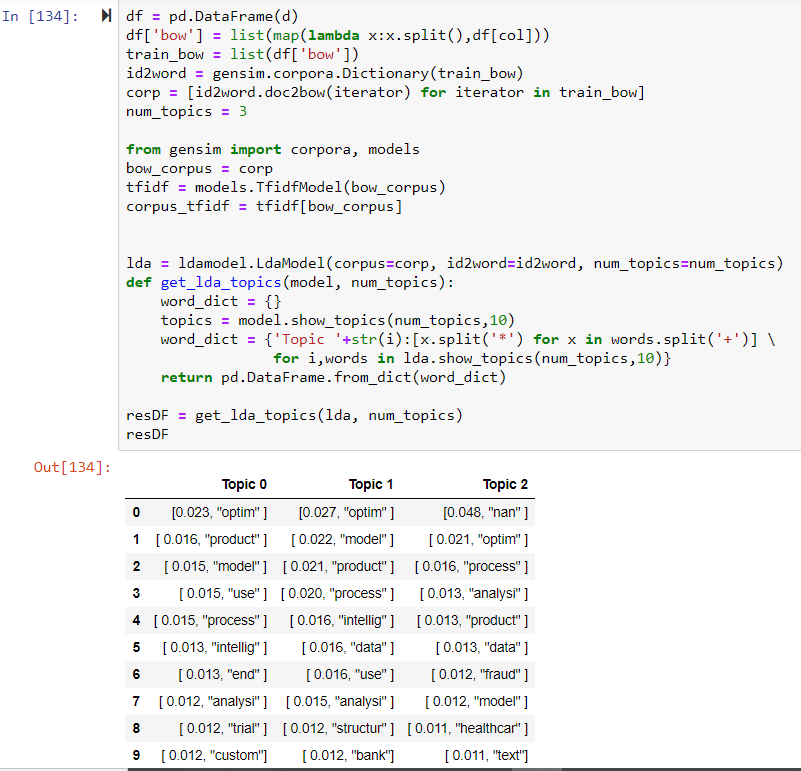
Topic Model Result of Company X



From the output we can see that Company X is centered around

* Cloud
* Digital solutions
* Fraud Detection
* Clients business
* Data related service providers

Topic model Result of Competitors data



From the competitor’s topic model output, we can see that they are all centered around

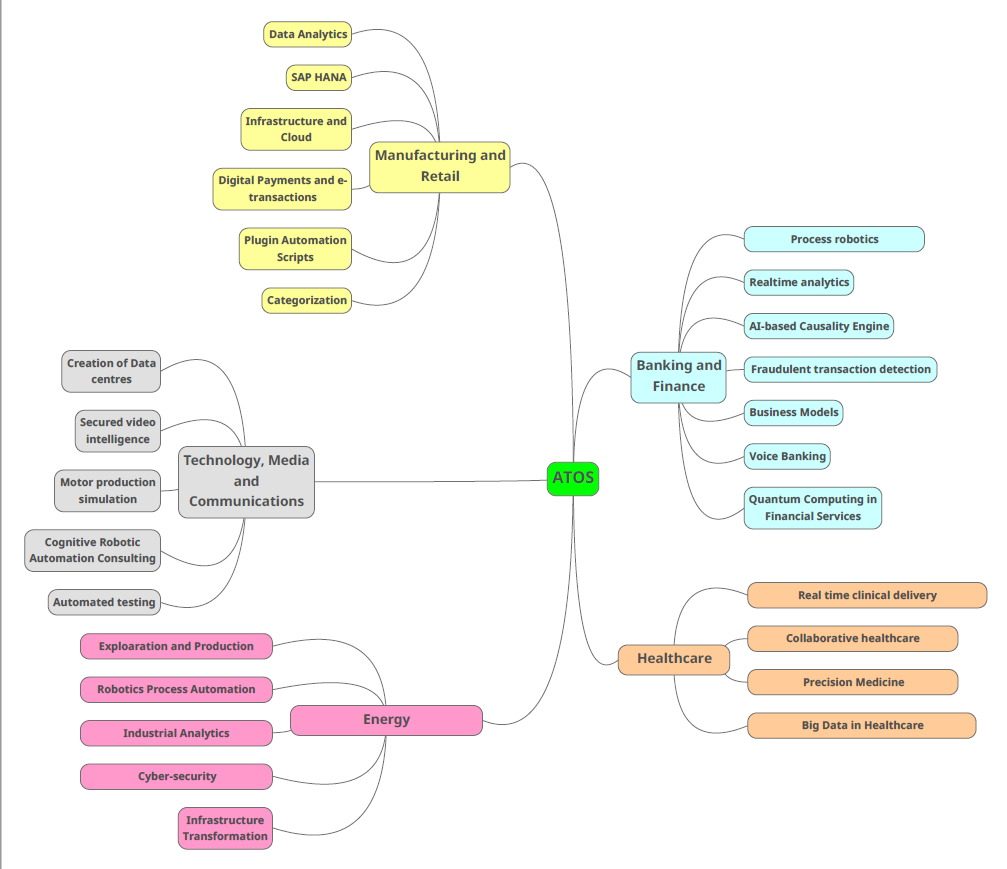
-Healthcare

-optimization related analytics

-banking models

-Text and semantic related services

**(04) Data Visualization**



Company X

