### Topic 3: Evaluation of the trends in jobs and skill-sets using data analytics: a case study

#### Introduction & Objective

* + Understanding the transitions in the job market to proactively identify the skill sets required.
  + Identifying trending jobs through a case study in the oil and gas industry by The proposed approach leverages a range of data analytics tools, including Latent Semantic Indexing (LSI), Latent Dirichlet Allocation (LDA), Factor Analysis and Non-Negative Matrix Factorization (NMF),
  + To study changes in the market by identifying the unequal between skills from the educational system and the skills that are required in the job market.

#### About Dataset

we focused on the title and description of the job from July 2015 to June 2018. The data collected were for the oil & gas sector in the GCC countries consisting of the United Arab Emirates (UAE), Saudi Arabia, Oman, Qatar, Bahrain, and Kuwait and the USA (Texas, California, Louisiana, Oklahoma, and Pennsylvania) by using O\*NET is one of the primary sources of occupational information for researches involving job analysis.

#### Methodology

* + **Data mining** is used to analyze unstructured textual data by using the Text-Mining Method of NLP to derive extracted text.
  + **Data Collection:** they scrape data from the O*NET website job posting (because the O*NET is a primary source of occupational information for research involving job analysis and its database is a US-based program. Ensuring that the O\*NET database is a good representation of not only US data but also GCC job market data is critical.). with data collected, they focus on Job titles and Job descriptions with 1110 occupations.

#### Data preprocessing

* + - tokenize Data: split the text into significant tokens and convert our data into single-word tokens by counting the frequencies of the words.
    - Removal of English Stop-Words: they remove words such as and, or, to, as, to reduce the matrix size.
    - Additional Filtering step: To avoid the issues of overfitting and underfitting, we removed all words that appeared in more than 90% of the occupations.
    - Stemming: The process of removing morphological affixes from words such that only the word stem is left, is called stemming. For example, a Stemming algorithm reduces the words ‘working’, ‘works’, and ‘worker’ to the root word, ‘work’.

#### Model

* + - **intensive change analysis (changes in the task content within an occupation.)**
      * **Factor analysis (FA):** describing the variability and EDA method used to search influential underlying factors or latent variables from a set of observed variables by reducing the number of variables called factors. It has the detriment known as split loadings and low correlation of such skills with multiple factors
      * **Non-negative matrix factorization (NMF):** handles polysemy which is the issue of a single word having multiple meanings
    - **Extensive change analysis(changes *across* occupations or analyzing the distribution of demand for the different occupation classes)**
      * **Latent Semantic Indexing (LSI):** a \*\*\**mathematical method for finding conceptually related words in textual data by dealing with synonymy and polysemy. In this case, O*NET dataset occupation descriptions are transformed into vector space where each element in the vector represents a word.
      * **Latent Dirichlet Allocation (LDA):** is a generative probabilistic model for collections of discrete data, such as text corpora.

#### Result

* + **LSI model results—job market data**
    - Mechanical Engineers have a higher demand than Civil Engineers in the GCC oil and gas industry.
    - There are overlapping high-demand jobs between the oil & gas industries in the USA and the GCC.
    - Engineers are in greater demand in the GCC, while manager and HR positions are more prevalent in the USA.
    - HR positions may be seen as a cost rather than a valued resource in GCC countries, leading to lower demand.
    - The model provides valuable insights for analyzing job market trends in the oil & gas industry.
  + **LDA oil & gas job market data**
    - LDA analysis of the oil & gas job market data reveals the top 10 occupations in the GCC and USA.
    - The demand for certain occupations, such as "Intelligence Analysts," is higher in the GCC, indicating a greater need for specific skills in the region.
    - It brings unique results that the LSI model did not find
    - Some occupations, like "Curators" and "Statisticians," may not seem directly related to the industry but have connections through job descriptions and related words.
    - There are overlapping high-demand jobs, such as "Chief Sustainability Officers," between the USA and GCC oil & gas industries.
    - The USA job market shows a trend for managerial and HR positions, while the GCC market has a different occupational demand pattern.
  + **Results of the factor analysis (FA) and non-negative matrix (NMF) factorization**
    - Using to analyze the intensive margin changes in occupational requirements.
    - NMF extracted six skill factors: Management, Quality Control, Analytics, Equipment, Manual, and Cognitive. The importance of manual skills decreased over time, while equipment skills increased. Cognitive skills showed slight negative changes.
    - FA identified five factors: Supervision, Mathematics, Perception, Equipment, and Problem Solving. The importance of supervision and perception skills decreased, while equipment skills showed a stronger demand. Mathematics and problem-solving skills remained relatively unchanged.
    - When weights from LSI and LDA models were introduced, the sides of demand did not change significantly, only the intensity of demand was strengthened.