INSIGHTS INTERPRETATION SUMMARY

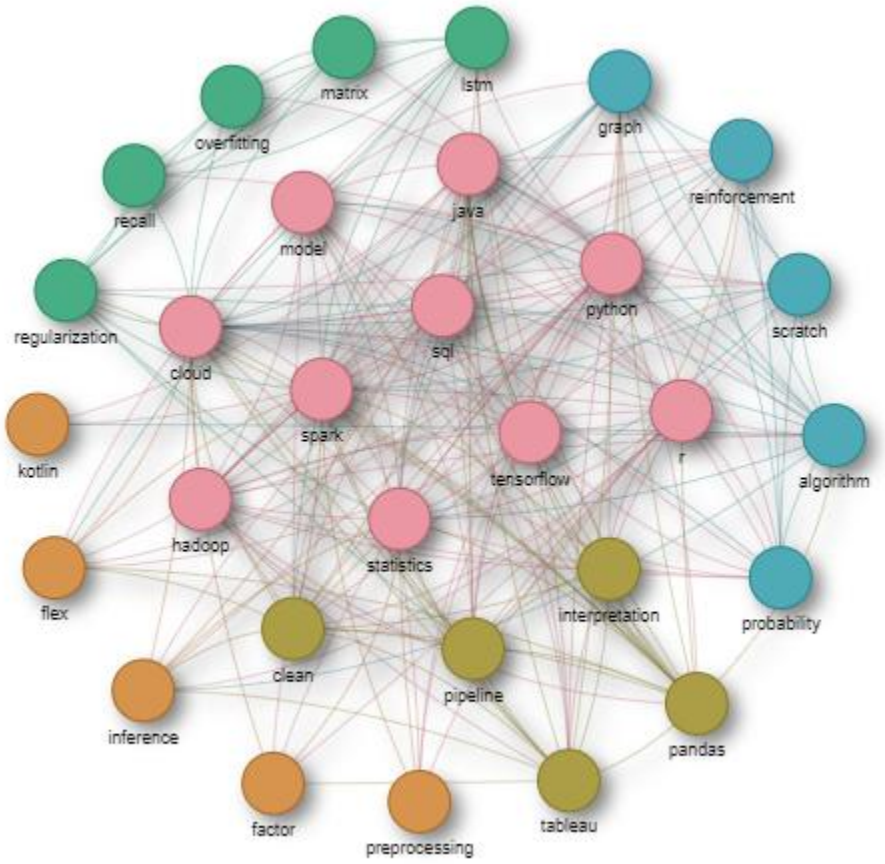

Fig 1.1: Hard Skills Dependency Graph of Records with Job Title of 'Data Scientist' [1, 2, 3]

[1] based on a sample of 163 job postings scraped from Indeed.ca between Nov. 2020 – Feb. 2021.
[2] nodes in pink illustrate the 10 most commonly occurring hard skills. Other node colors represents the clusters or groupings of job postings, as segregated by the spectral clustering algorithm. Refer to code for details on hyperparameters and other algorithm implementation details.
[3] Cluster node labels are the top 5 terms that best characterize a cluster, ranked based on Laplace Smoothed Positive Pointwise Mutual Information (PPMI)

As per Fig 1.1., current industry needs can largely be classified into **3 types of data scientists**. These are inferred from the figure as below:

1. **The "full-stack" data scientist** roles with wide-ranging skills involving data **cleaning** (using **pandas**), creating **pipelines** and **interpreting** findings using visualization tools such as **tableau**
2. **Roles that recognize use of Kotlin** for performing **preprocessing** and **inference**. Kotlin is a Java based programming language, growing rapidly to improve upon the current deficiencies of the programming language Scala.
3. **Two more research-based roles** found to be distinctly separated into two categories:
    a. **Graph** and **Reinforcement** learning focussed roles with ability to develop **algorithms** from **scratch**
    b. **Deep Learning** focussed roles with skills in developing deep neural networks (with **LSTM** found to be most sought after) along with **other** machine learning related know-how