Features Engineering

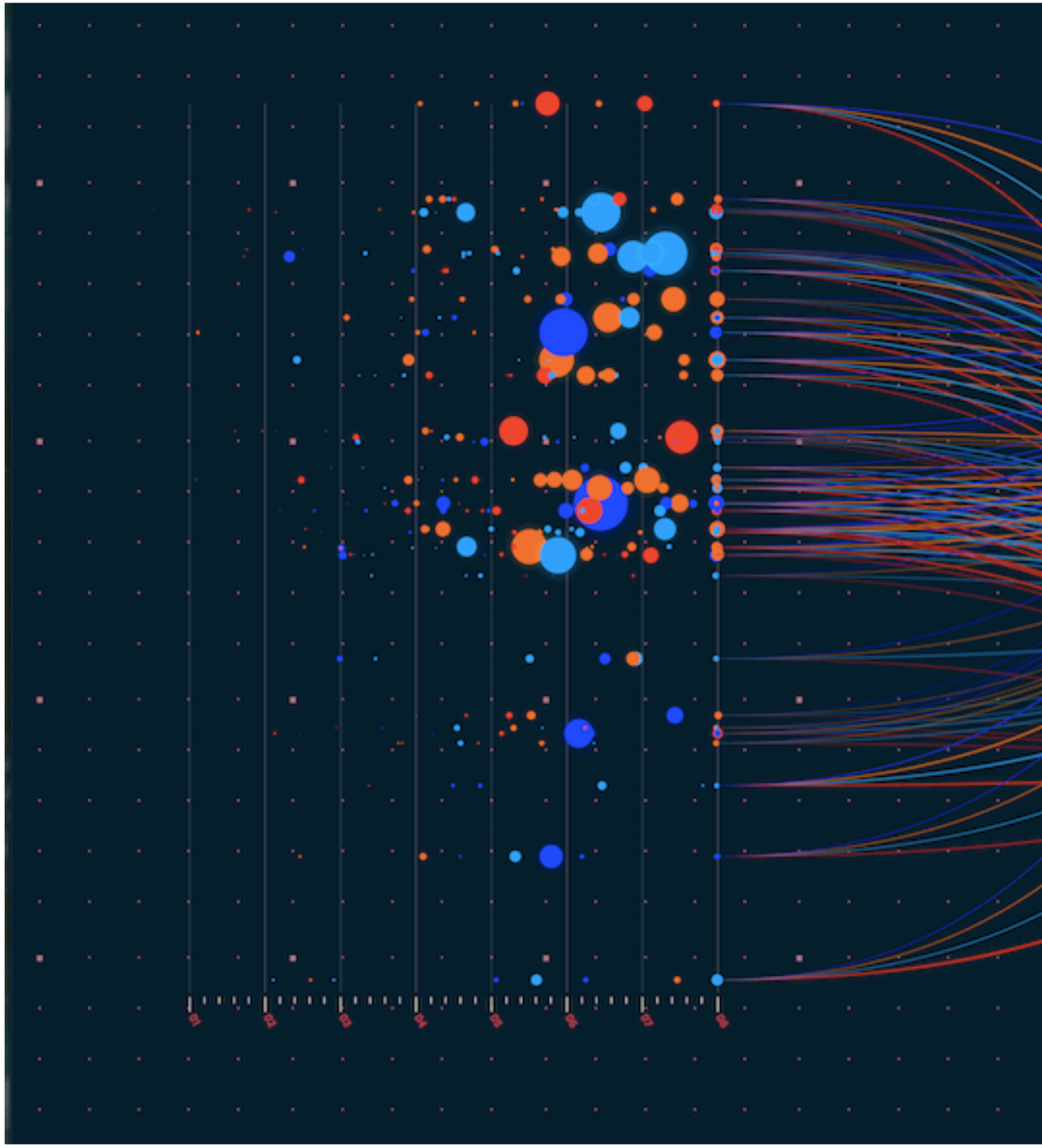
for Workforce Analytics

Michael Ghattas | Lily Chen | Khoi Phan

CSCI 4502

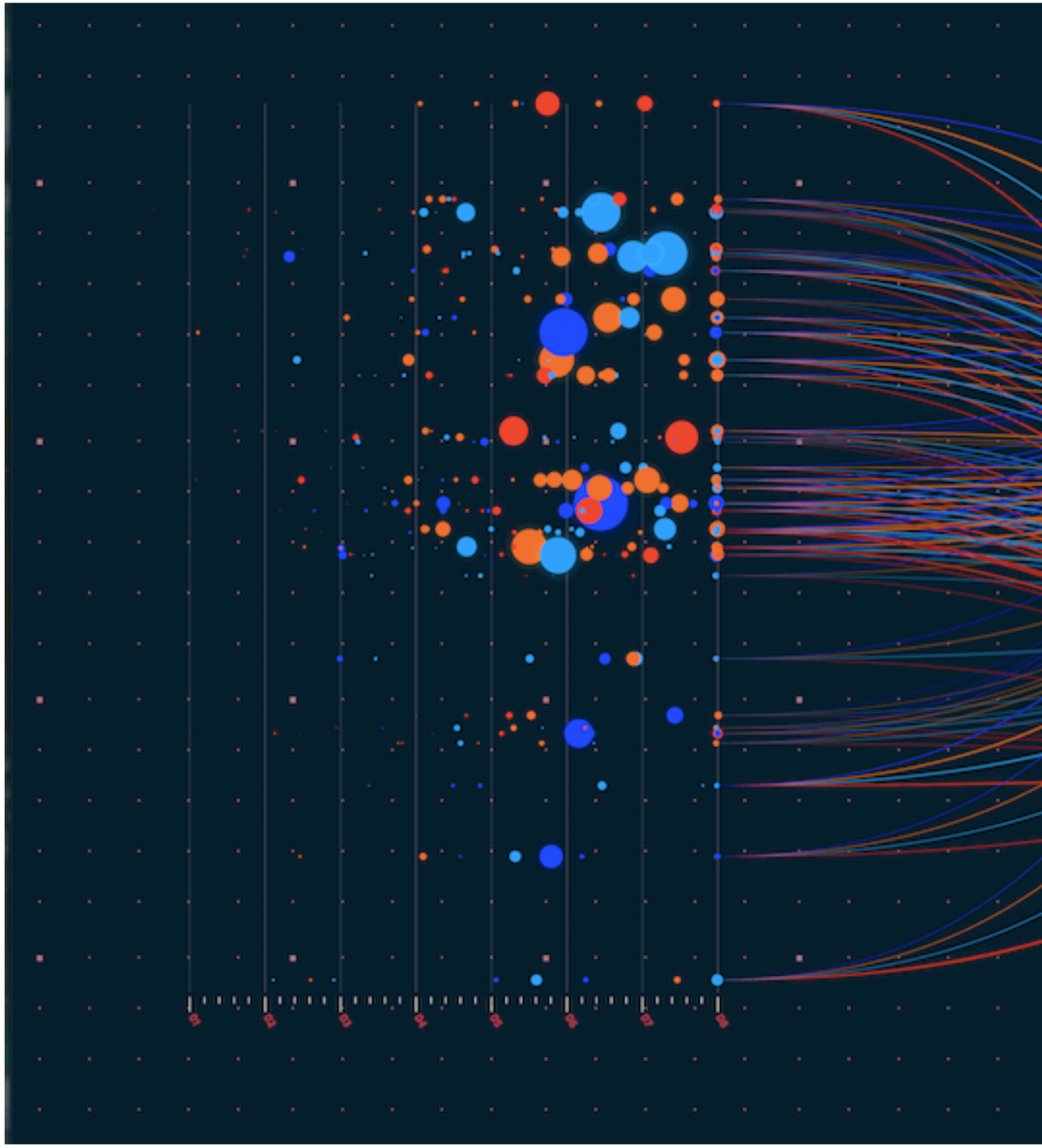
Background

- Data science is a fast-growing field dominating the decision-making process within the business world and supported by the need for operational efficiency and Return On Investment (ROI) growth mindset.
- The driving force behind such initiatives is usually the workforce intelligence teams within the Human Resources department.



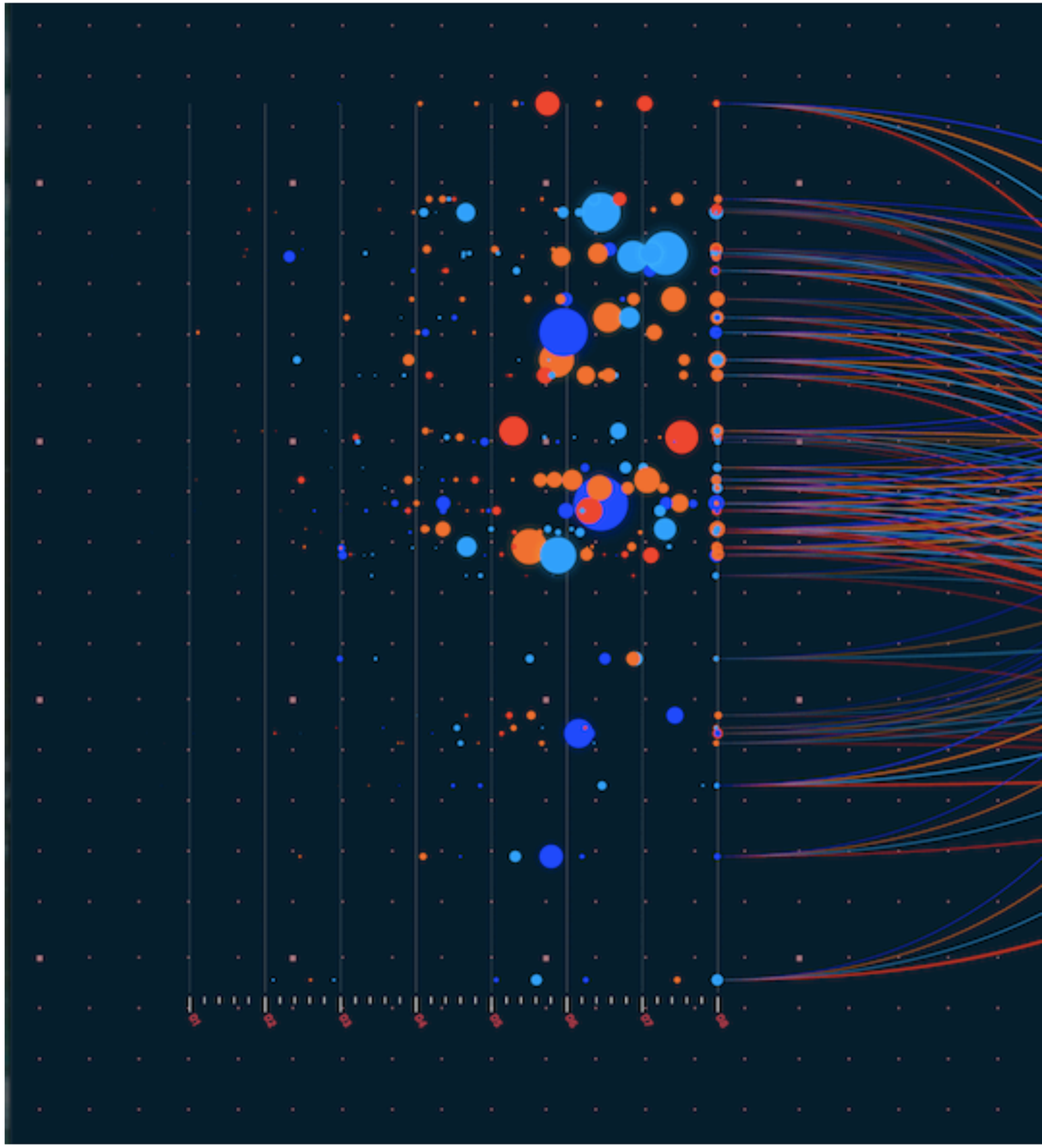
Inspiration

- Determinants of employee engagement and their impact on employee performance
- Factors affecting employee performance of PT.Kiyokuni Indonesia
- Factors affecting employee performance: an empirical approach
- An Effectiveness of Human Resource Management Practices on Employee Retention in Institute of Higher learning



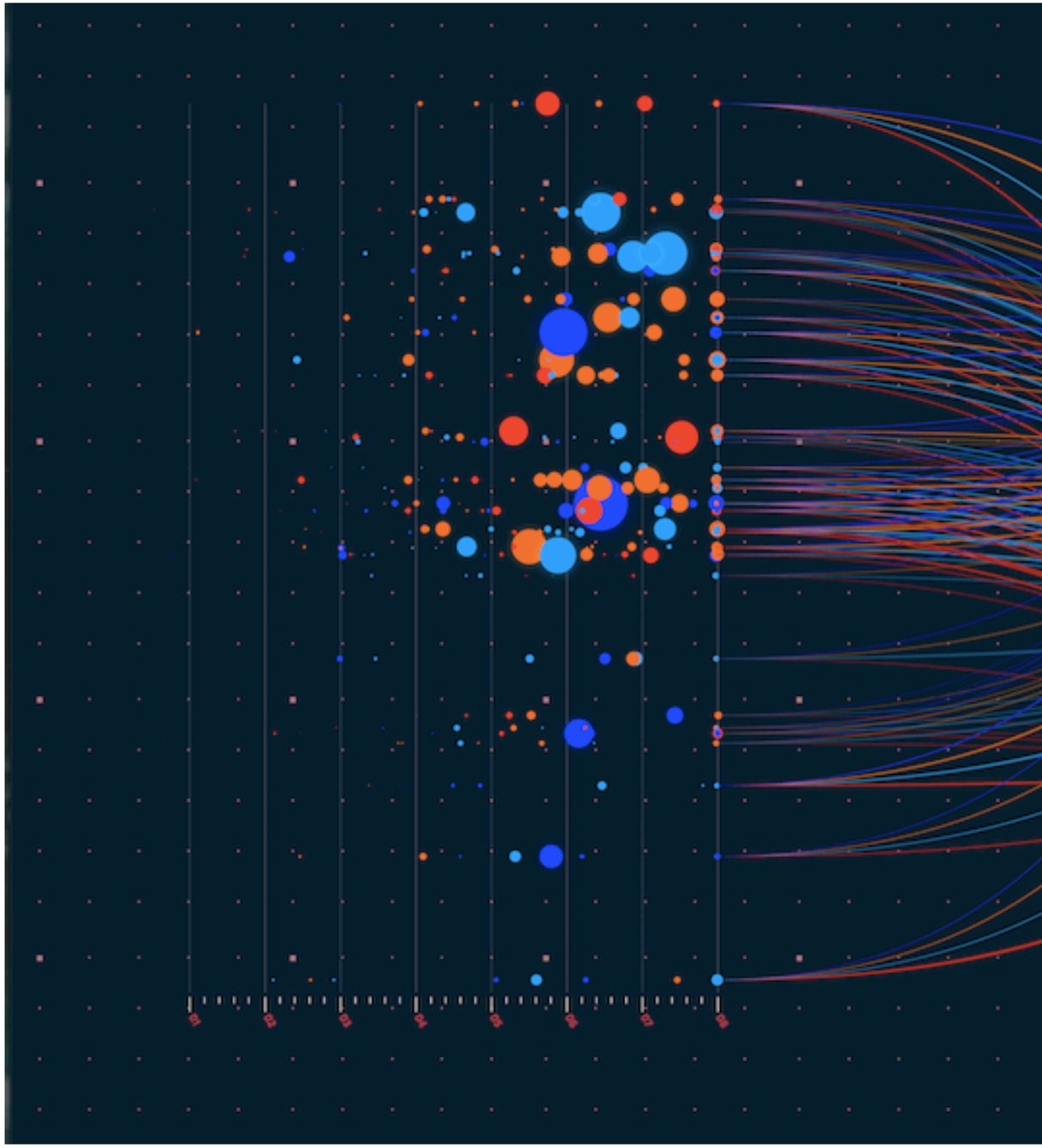
Description

The project aims to develop a methodology for a supervised ML algorithm utilizing a well-defined series of techniques. The goal is to classify, transform and interpret the aggregated data, and extract the datasets valuable as predictors for a workforce analytics model to measure candidate and employee metrics.



Data

- **Absenteeism:** https://www.kaggle.com/code/hypnobear/absenteeism-at-work-dataset/data
- **Human Resources:** https://www.kaggle.com/datasets/rhuebner/human-resources-data-set
- **Turnover:** https://www.kaggle.com/datasets/davinwijaya/employee-turnover
- **Job Classification:** https://www.aihr.com/blog/hr-data-sets-people-analytics/
- **IBM-HR:** https://www.aihr.com/blog/hr-data-sets-people-analytics/



Methods

Languages:

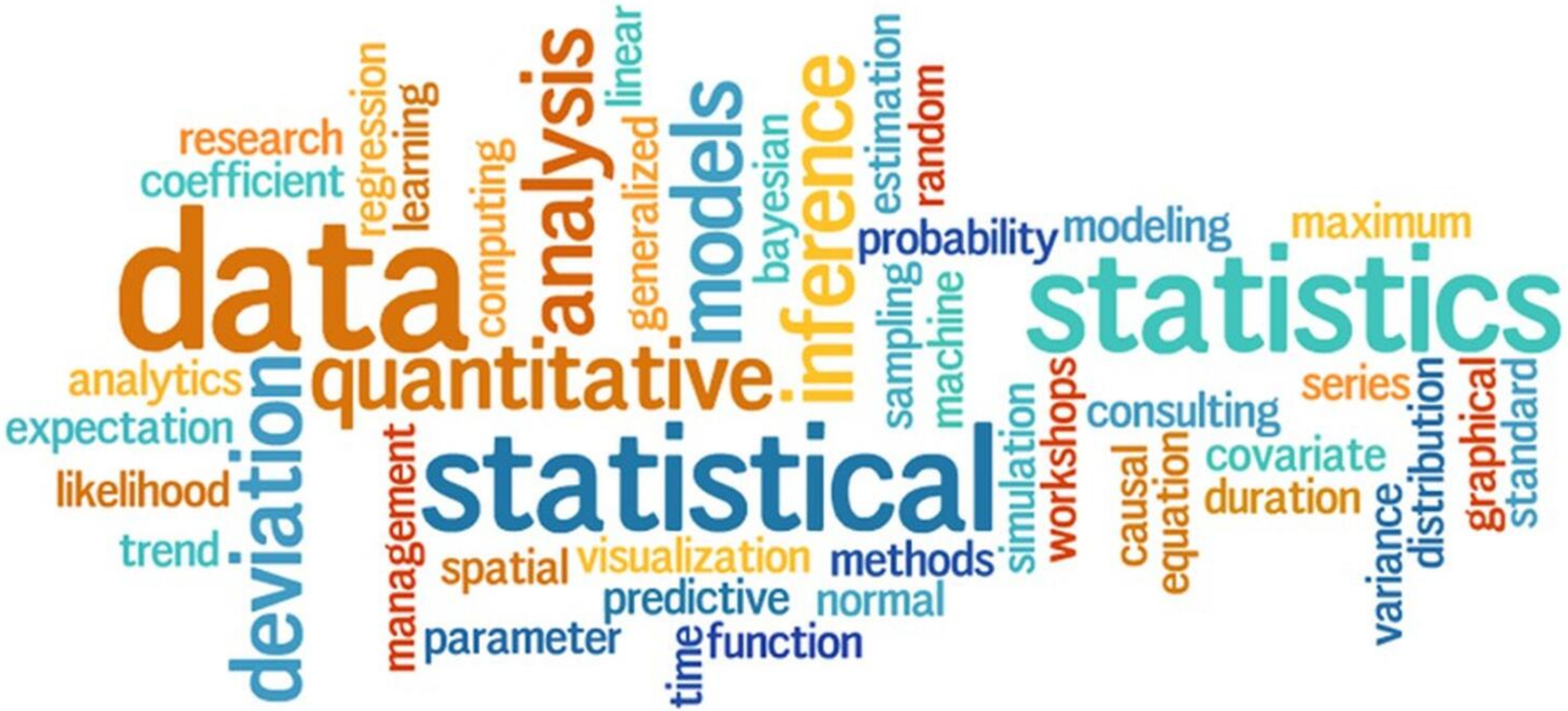
- R
- Python

Environments:

- JupyterLab
- VSCode

Possible Tools:

- Dynamic Discretization
- Clustering
- Deviation
- Multidimensional Analysis
- Imbalance Ratio
- Hypothesis Testing
- Akaike information criterion (AIC)
- Bayesian information criterion (BIC)
- K - Nearest Neighbor (KNN)
- Monte-Carlo Markov Chains (MCMC)



Milestones

Project Development

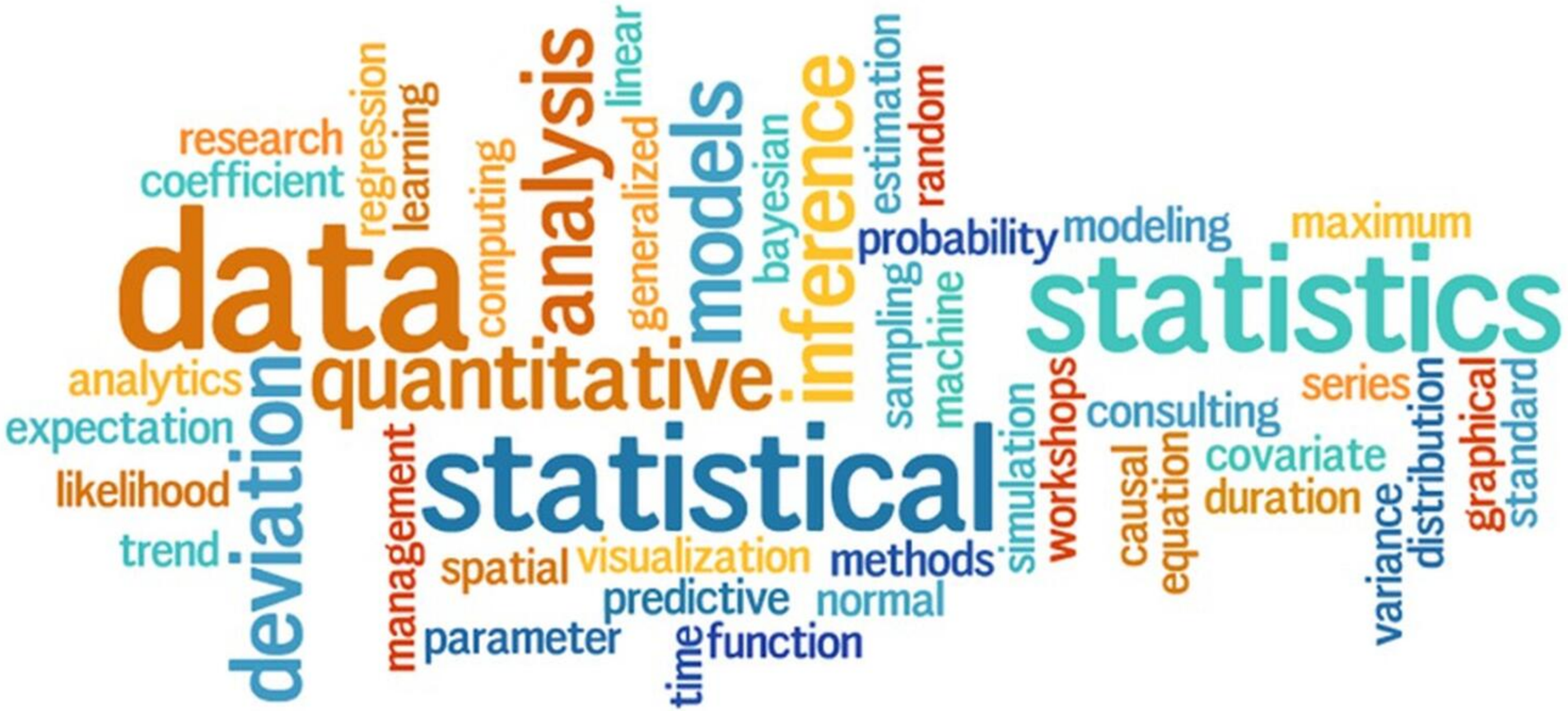
Idea generation & conceptualization
Subject research
Data research & collection
Write & review our proposal

Project Implementation

Develop & finalize our plan
Data exploration & assessment
Data aggregation & transformation
Data analysis & interpretation

Project Delivery

Document our results & findings
Write & review our report
Design & review our presentation
Rehearse team presentation



Thank you!

Please let us know if you have any questions.

Michael Ghattas | Lily Chen | Khoi Phan