

## **Big Data: Group Project Proposal**

Team 9: Xiao Han, Jent Lapalm, Aoran Wang, Yushan Yang

### **Project Goal:**

Analyze the *Human Resources Dataset* to ask, answer, and visualize questions concerning 'people analytics.' Proposed questions (subject to change as we explore):

- Is there any relationship between who a person works for and their performance score?
- What is the overall diversity profile of the organization?
- What are our best recruiting sources if we want to ensure a diverse organization?
- Can we predict who is going to terminate and who isn't? What level of accuracy can we achieve on this?
- Are there areas of the company where pay is not equitable?

### **Dataset:**

*Human Resources Dataset*: <https://www.kaggle.com/rhuebner/human-resources-data-set>

This teaching dataset is designed for querying, predicting, and visualizing HR information.

### **Database Format & Technologies:**

We plan on accessing, querying, predicting, and visualizing the dataset using MySQL, R, Python, and Tableau. Our hope is to use AWS to store the raw and processed data, similar to our pre-fall *tech bootcamp project*.

### **Potential Organizational Value:**

Many companies lack data maturity in their HR departments - so called 'people analytics.' ([Mckinsey](#)). Addressing this is multiplicative in value, from simple questions such as 'which interview questions predict our top performers' to more complex questions like 'should we use an internal team or 3rd-party vendor for this project?'

Our current questions add value by increasing worker effectiveness (manager-worker performance-score patterns), driving strategy (recruiting sources to meet diversity goals), and lowering costs (predicting employee churn).