# Bayes Assignment 3 of 2025

Sean van der Merwe

2025-02-26

## Instructions

The dataset is the 100 AI Companies of 2024 dataset from Kaggle. It is provided on the learning management system for convenience. The last 3 columns are the most interesting and what you will focus on.

Note that the data is raw, and slightly corrupted. You must first clean up the data and transform the variables. All such steps must be done using R code. You may not make any alterations to the data set. Your code must run correctly if someone downloads the data from the source again. For example, there are scores that Excel converted to dates at some point, transform them back intelligently with code.

**Goal:** To implement a robust Bayesian regression model on a dataset to test for basic trends and make a prediction with uncertainty.

After data cleaning, drop the two companies with missing scores. Then drop the company that corresponds to your position on the class list, leaving 97 companies.

Explain, based on statistics, whether you think the variables are related (before or after transformations). For the company that corresponds to your position on the class list, predict their revenue **distribution** for 2025 (one year older) assuming that their Glassdoor score drops by 0.5/5.

Your marks will be based how well you explain your approach and how sensible your reasoning is (all your steps, and especially your predicted distribution, must make sense given the constraints of the data).

Also, consider using transformed variables in your regressions, such as log annual revenue and log age, instead of the raw values.

For this assignment, submit Word, PDF, and Rmd/qmd files in one submission on the learning management system, in that order.