#### **Final Project: Topic Proposal – TEAM 3 DATS 6101 – Spring 2025**

#### **Project Title: Predicting Data Science Salaries in 2025**

#### **Research Topic:** Our project focuses on forecasting Data Scientist salaries using data from 2020 to 2024. With rising demand for data professionals, predicting 2025 compensation trends is essential for both employers and job seekers. Building on our midterm analysis, we’ll now develop models to estimate the average 2025 salary.

#### **SMART Question:** “How accurately can we predict the 2025 average salary of data science professionals using 2020–2024 data, based on job title, experience level, employment type, and work model?”

#### **Specific:** Focuses on predicting 2025 salaries using four key features in the data science field. **Measurable:** Model accuracy will be assessed using R², RMSE, and MAE in R. **Achievable:** Historical data and R packages like tidymodels, ggplot2, and lm make this feasible. **Relevant:** Addresses real-world salary forecasting needs for professionals and employers. **Time-bound:** Uses data from 2020–2024 to make predictions specifically for 2025.

#### **DATASET INFORMATION:**

#### *Data Science Salaries 2024* from Kaggle, containing ~5,200 observations from 2020–2024. It includes features like job title, experience level, employment type, remote ratio, and salary in USD.

#### **GitHub Repository:** <https://github.com/sameerbatra1/Data-Science-Salaries>

#### 