**PROJECT ONE: PROPOSAL**

**Project Title:** Data Science Job Salaries Around the World

**Team Members:** Carl, Prathima, Shaun

**Project Description/Outline:**

* To find correlations between data science job salaries and employee situations.
* We have received data from 600+ individuals on an international scale who are currently working in a data science role within their company. We would like to find out the key drivers that influence the salary of these workers.

**Research Questions to Answer:**

1. **(CARL)** We would like to find out whether there is a difference in salaries between those employees working from home and those employees working from the office. Would employees sacrifice their salary to be able to 100% work from home? How do data science role salaries vary by employee experience level?
2. **(PRATHIMA)** How do data science role salaries vary by geographic location?
3. **(SHAUN)** How do data science role salaries vary by role type? Out of all the Data Science Job Titles, which Job Title has the highest average salary?

**Datasets to be used:**

<https://www.kaggle.com/datasets/ruchi798/data-science-job-salaries>

**Rough breakdown of Tasks:**

* Together we will put csv into DataFrame and clean DataFrame with relevant columns according to your questions
* Select 3 questions from above – (1 person for each question)
* Use bargraphs, scatterplots, boxplots to visualise data and find relationships between the categories.
* Use relevant testing such as summary statistics tables, correlation coefficient, hypothesis testing to come up with conclusions for data.
* Push files to **individual workspace branches** in GitHub repo. Merge files to **devtesting** branch to test the codes are working fine. Merge and push final files to **main** branch once everyone is happy with file.
* Create slides for Presentation.