ETL Project – Data Related Job Market

Team Members: Kelvin Deng, Thao Hoang, Yijing Su, May Ang

# Project Purpose:

The project will explore the current job market for data related roles. This includes Data Scientist, Data Analyst, and Data Engineer opportunities. The focus will be documenting the Extraction, Transformation, and Loading into databases.

# Data Sources:

1. Data Scientist Jobs (for US job market) [https://www.kaggle.com/andrewmvd/data-scientist-jobs]
2. Web scrapping from Indeed (for Canada job market) [ca.indeed.com]
3. Job applied using Linkedin Bot [https://www.kaggle.com/manuelalfredocollado/jobs-applied-in-linkedin-with-easyapplybot?select=jobs.csv]

# Process steps:

## Extraction

* Load data from data sources into Jupyter Notebook

## Transformation

* Use Jupyter Notebook (Pandas) to clean up data (i.e. remove NA values/ inconsistent values)
* Transform data into data frame
* Create any foreign keys

## Loading

* Using pgAdmin to create tables (US job market, CA job market) to database
* Commit to cloud database

# Report

* Discuss complications and challenges
* Document into report