Project 3 - Data Science Skills

Project Document

3/12/23

Team Members:

* Kory Martin
* Genesis Middleton
* Joe Garcia
* Pei-Ming Chin
* Daniel Craig

Collaboration Tools:

* Github (Code Collaboration)
* Google Drive/Docs (Project Documentation)
* Slack Channel (Communication

Data Sources:

* LinkedIn - Kory
* Simply Hired - Pei-Ming
* Indeed - Daniel
* Monster - Joe
* Glassdoor - Genesis

(Most have API’s, a few will need web scraping, plan is to normalize the data and then upload each dataset into Github repo)

Basic Entity Relationship:

Industry > Company > Job Listing < Job Board

\* Job Skill will be related to Job Listing

Read as:

Industry is related to Company.

Company is related to Job Listing

Job Skill is related to Job Listing

Job Listing is related to Job Board

* Job Board Table
  + Job Board ID
  + Job Board Name
  + Job Board URL
* Company Table
  + Company Name
  + Company ID
  + Industry ID
  + Location
* Job Listing Table
  + Company ID
  + Job Board ID
  + Job Posting ID
  + Data Source ID
  + Position/Title
  + Seniority Level
  + Years of Experience
  + Education
  + Remote vs. On-Location vs. Hybrid
* Job Skill Table:
  + Skill ID
  + Skill Name
  + Skill Type
* Industry Table
  + Industry ID
  + Industry

Rough Timeline

Part 1

* Documenting the process - Kory **(3/11)**
  + Create Private Slack Channel - **Genesis**
  + Create Google Docs - **Kory**
  + Setup GitHub - **Joe**

Part 2

* Web scraping/Data Collection
  + Create a master spreadsheet that is structured like our Datasource - **Kory (3/9)**
  + Each person selects a different data source and scrapes the data for at least 25 job postings **(3/14)**
    - LinkedIn - Kory
    - Simply Hired - Pei-Ming
    - Indeed - Daniel
    - Monster - Joe
    - Glassdoor - Genesis
* Ingesting into DB -
  + Build out database in Amazon AWS - **Genesis (3/15)**
    - Building out the data tables
  + Cleaning of the data in R - **Kory (3/16)**
  + Ingesting data into database - **Joe (3/17)**
* Analysis of Data / Data Visualization - **PeiMing (3/18)**
* Final Code Review/Team Sync-up - **All (3/19)**

Part 3

* Putting together presentation slides - **Daniel (3/21)**
* Identify Presenters/Section
  + Kory
  + Joe
  + Genesis
  + Daniel
* Presentation Sync-up - **All (3/21)**

Team Meeting:

* Thursday 3/16