**README:**

For the Jobs In Sports database, real data was collected via TeamworkOnline and Wikipedia. LINKEDIN and Indeed will also be scraped once security measures are worked out. For the scraping completed as of 12/05/2022 the following files were developed as outlined below:

Files for Assignment 3:

Database Use Cases Assignment 3 [.docx] – 5 New Use Cases requiring JOINS and the SQL performed to verify.

Updated\_schema\_assignment3 [.docx] – updated schema to reflect changes to this point including the addition of the job\_requirements table and a few column adjustments from job\_postings etc.

Teamwork Online Scrape [IPYNB] – Python Script containing the bulk / if not all of the data scraping and cleaning as well as insert into the database.

Other important files include the following:

DAMG6210\_Assignment2\_JobsInSports [IPYNB] – The same file as previous assignment with one change being editing out quotations with replace so tweets all are included as per assignment 2 feedback.

Notes: Next Steps include the additional scraping and data insertion, Data Normalization insurance, further company info scraping, and finally some more twitter cleaning if time permits.