**Project proposal**

**Class**: DATS\_6103\_DataMining

**Group:** A

**Group Members:**

|  |  |
| --- | --- |
| 1 | Farris Darawsheh |
| 2 | Ghadeer Alqassab |
| 3 | Luangxay Hadsadong |
| 4 | Mina Hanna |

**The dataset:** Anonymized data from profiles scraped on LinkedIn found on Kaggle. It contains data from about 15000 profiles in Australia and US and few other countries.

There are 52 columns for each profile record that contains basic profile data such as employer name and length of the tenures in addition to various facial and emotional analysis on the profile image (i.e. smiling, race and head orientation) .

**Objectives**: Enhance the data with some further fields available through public LinkedIn APIs and perform data cleaning, EDA and preprocessing to build various models to help us with the following analysis:

1. What is the average tenure length for each age group ?
2. Does the younger generation switch jobs faster ?
3. Is there any difference based on the gender ?
4. What is the percentage of employees who switches roles within the same employer ?
5. Do people smile or show a serious face in their Linkedin profile photos ?
6. The average tenure length among different ethnicity groups as well as nationality
7. The tenure trends among different companies
8. The tenure trends among different religions, but I am not sure how good is the race match in identifying individual religion.
9. Does employer industry affect the length of the tenure ?