#### Assessment 2: Part 1 Data Preparation and Preprocessing: **Describe the dataset. (8 points)**

✓ Describe the dataset (e.g.: type of column, value range). (1 point) **- CC**

✓ How many records are there in the dataset? (1 point) **- CC**

✓ Which period does it cover? How many different dates have job postings? (1 point) **- CC**

✓ How many locations does the dataset have? Which location has the most job postings? (1 point) **- CC**

✓ How many job sectors(job classifications) are there in the dataset? List the name of each sector and its’ total of job postings. (1 point) **- CC**

✓ Choose your favorite job sector (e.g. Information & Communication Technology), how many sub-sectors are there in that sector? List the name of each sub-sector and its’ job posting number. (1 point) **- SP**

✓ List the salary ranges and their total of job postings. (1 point) **- SP**

✓ List the job types. In each job type, what are the lowest salary and highest salary? (1 point) **- SP**

***2. Normalize and clean data. (7 points)***

✓ The salaries are kept in the dataset as “HighestSalary” and “LowestSalary”. You should calculate the “AverageSalary” for each job. (1 point) **- SP**

✓ The raw dataset values of the "Id" column had inconsistencies in their representation. The Id values should have 8 number long integers only. Write code to remove unnecessary characters. (1 point) **- SP**

✓ The "Date" column is represented in a format that contained both date and time information. However, the time is not correct and should be removed. (1 point) **- HN**

✓ Change type of “Id” column to numeric and change type of “Date” column to DateTime. (1 point) **- HN**

✓ Are there any duplicate data in the dataset? Provide the way you find them and your solution to fix it. \*\* (1.5 point) **- HN**

✓ Check missing data and visualize them in a corresponding chart.\*\* (1.5 point) **- HN**