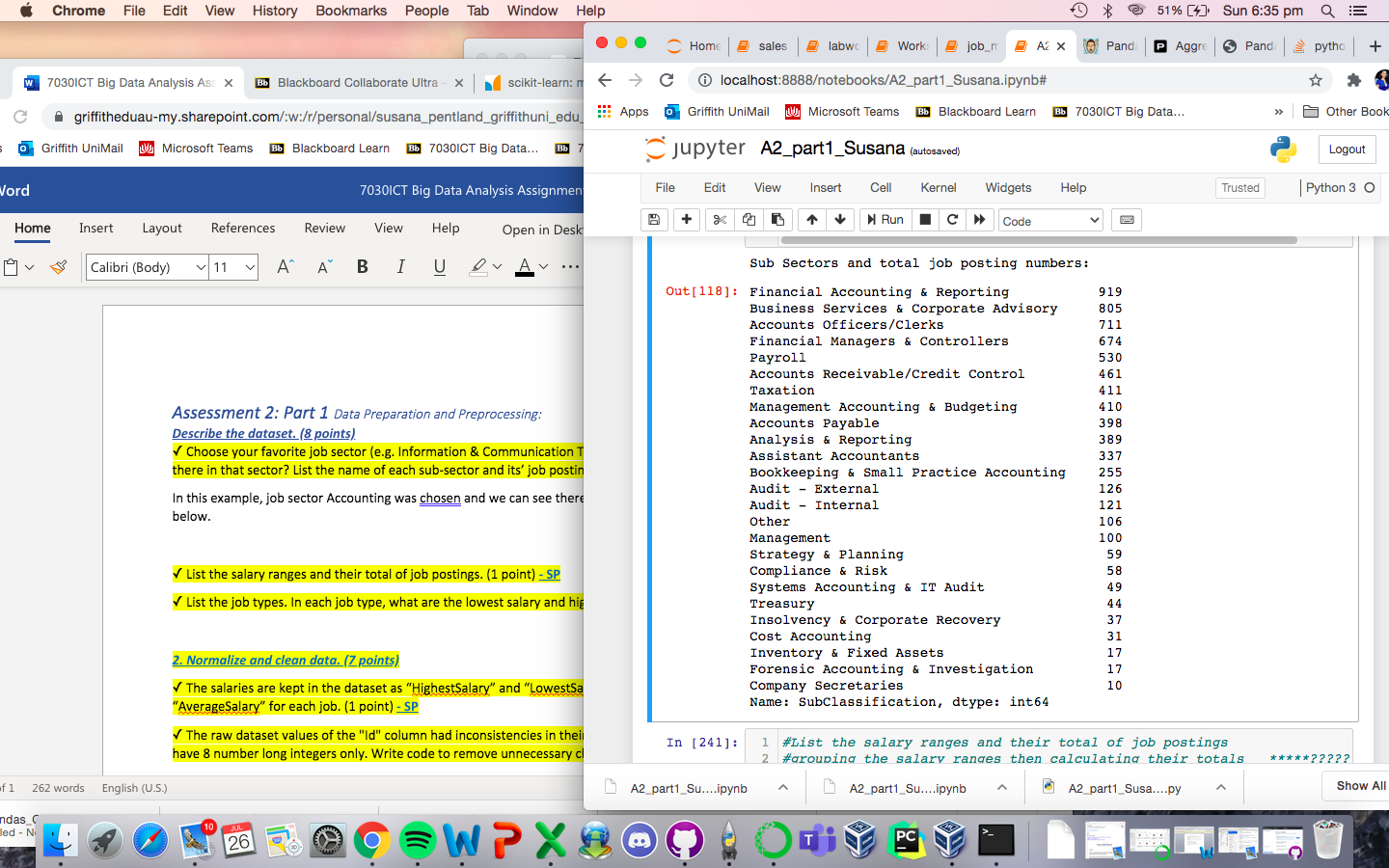
#### Assessment 2: Part 1 Data Preparation and Preprocessing: **Describe the dataset. (8 points)**

✓ 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**

In this example, job sector Accounting was chosen, and we can see there are a total of 25 sub sectors as listed below.



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

...TBC

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

From the dataset given, we identified the lowest and highest salary ranges across 4 different job types.

Casual/Vacation ...

Contract/Temp ...

Full Time ...

Part Time …

…. TBC

***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**

From the minimum and maximum salaries given in the dataset, we could calculate the average salary for each job. …. TBC

✓ 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**

During the data cleaning process, some types of challenges we would expect to see in the dataset include formatting, missing data, erroneous data, irrelevant and especially inconsistent data such as in the ID column. The dataset given is already clean with consistent format in ID column, so no further processing was required.