* 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)

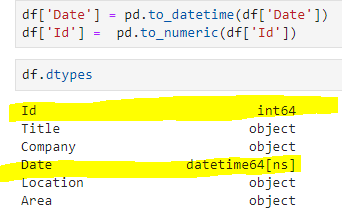
The ‘replace’ method is used to remove all text after the ‘T’ which represents the time in the “Date” column



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

Jupyter Notebook

We are using method “to\_datetime” to change the type of “Date” column to Datime and method “to\_numeric” to change the type of “Id” column to Numeric.



* Are there any duplicate data in the dataset? Provide the way you find them and your solution to fix it.

There is no duplicate in the dataset.

To find duplication in the dataset using this command: df[df.duplicated()].shape[0]

To fix duplication using this command: df.drop\_duplicates(inplace=True)

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

Miss data mostly appears in the “Area” column which 56503 missing values, in the other hand, there is no missing data in these columns: Id, Title, Date, Location, SubClassification, Requirement,

LowestSalary and HighestSalary

