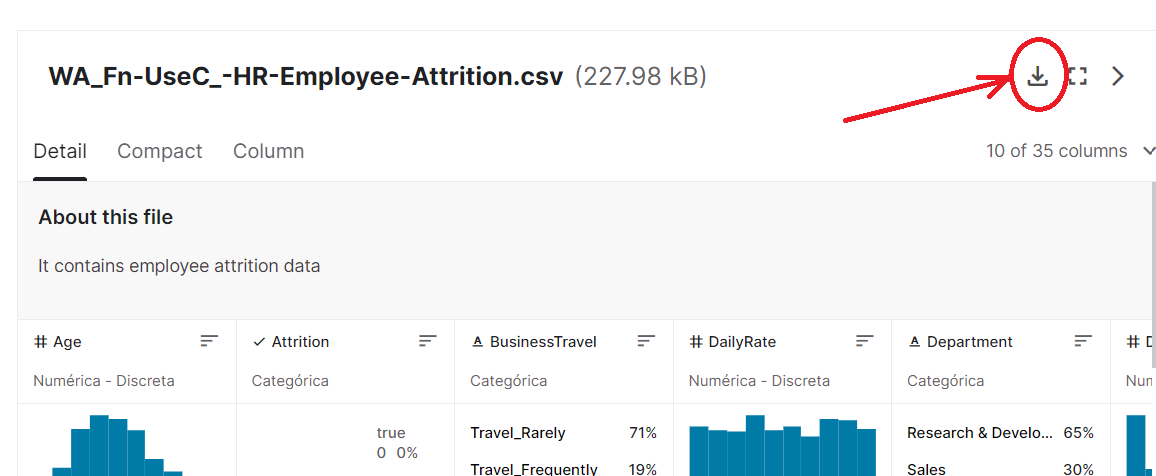
**Task 1**: Data Preprocessing :

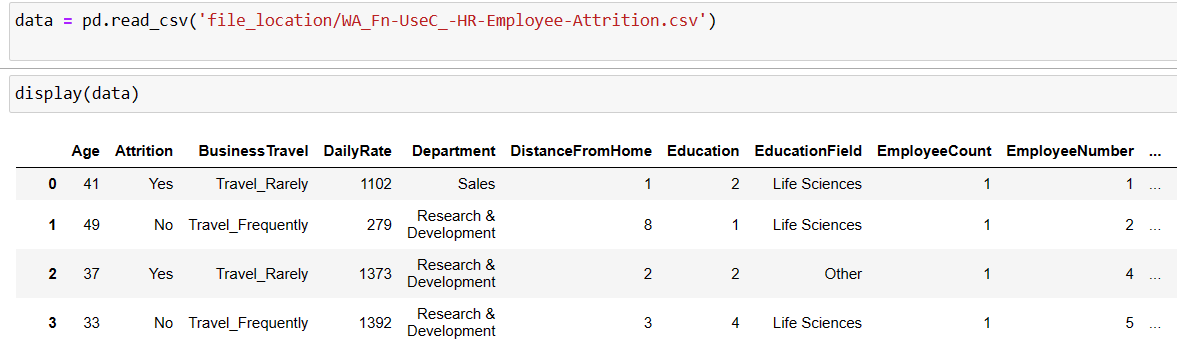
**Step0**: Download the dataset from the kaggle website:



**Step1**: upload the dataset using pandas library.

Une image contenant texte, Police, capture d’écran, ligne

Description générée automatiquement



**Step2.1**: handling missing values:

checks how many missing (null) values are in each column of the DataFrame. If you find any missing values use dropna() function to remove rows with missing values in any column

Une image contenant texte, capture d’écran, Police, nombre

Description générée automatiquement

**Step2.2**: Identify and address out-of-range data: Examine histograms to detect values that fall outside the expected ranges (e.g., non-real values, ages exceeding 100) to detect potential data entry errors.

Une image contenant texte, diagramme, Tracé, capture d’écran

Description générée automatiquement

Upon reviewing the histograms, it appears that the dataset is accurate and does not exhibit any apparent data entry errors.

**Step2.3**: Drop useless columns

we have some columns that contains only one unique value, it is a good choice to drop these columns.

Une image contenant texte, Police, ligne, nombre

Description générée automatiquement

**Step3**: encoding object variables.

**Step3-1**: display the counts of columns with each data type.

Une image contenant texte, Police, capture d’écran, blanc

Description générée automatiquement

* there are 26 columns with integer (int64) data types and 9 columns with object data types in the DataFrame

**step3-2**: separate the columns with object datatypes in a separate dataframe data1

Une image contenant texte, Police, nombre, ligne

Description générée automatiquement

**step3-3**: iterates through the columns of the DataFrame 'data1,' lists the unique values in each column.

Une image contenant texte, capture d’écran, Police, nombre

Description générée automatiquement

**Step3.4**: Define mappings for the columns and use the replace method with the defined mappings for the columns with object datatypes in data dataframe



Une image contenant texte, capture d’écran, Police, nombre

Description générée automatiquement

**Step4**: scaling features

The target should not be scaled.

**Step4.1**: Split data into features and target.

Une image contenant texte, Police, capture d’écran, ligne

Description générée automatiquement

**Step4.2**: determine scaling method.

Une image contenant texte, capture d’écran, Police, Page web

Description générée automatiquement

Une image contenant texte, capture d’écran, Police, nombre

Description générée automatiquement

**Step4.3**: scale features:

Une image contenant texte, capture d’écran, Police, affichage

Description générée automatiquement

Une image contenant texte, capture d’écran, Police, nombre

Description générée automatiquement

**Step5:** save the scaled\_x dataframe and target data to csv files.

Une image contenant texte, Police, capture d’écran, ligne

Description générée automatiquement

**Result**: Cleaned dataset

Below, you'll find two inserted Excel table containing the scaled\_x\_data and the y\_data. To access the Excel table, simply double-click on it.

Scaled\_x\_data:



Y\_data:

