

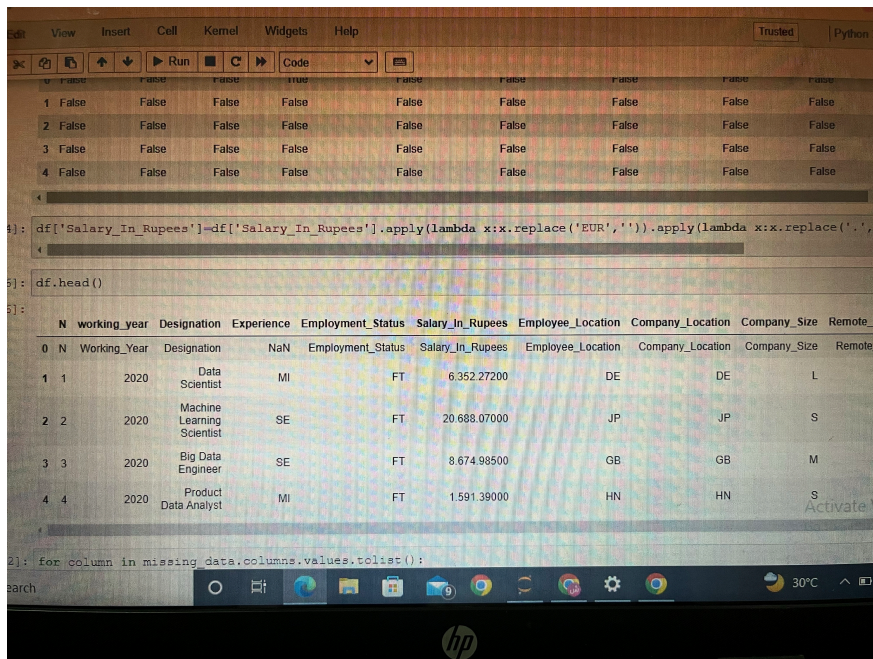
# Data Wrangling

## Steps:

- First we delete some values from our dataset
- Then we started the code by importing libraries (pandas, matplotlib, numpy)
- Replace missing values by doing drop
- Convert object format to another correct form which is float
- Data Normalisation by using min-max
- We Binned data (Remote\_Working\_Ratio column) to (Low,medium,high)

We replace commas by dot and it works then it gave us error

Here is a picture when it worked



```
4: df['Salary_In_Rupees'] = df['Salary_In_Rupees'].apply(lambda x: x.replace('EUR', ''))
5: df.head()
```

	N	working_year	Designation	Experience	Employment_Status	Salary_In_Rupees	Employee_Location	Company_Location	Company_Size	Remote_W
0	N	Working_Year	Designation	NaN	Employment_Status	Salary_In_Rupees	Employee_Location	Company_Location	Company_Size	Remote_Y
1	1	2020	Data Scientist	MI	FT	6.352.27200	DE	DE	L	
2	2	2020	Machine Learning Scientist	SE	FT	20.688.07000	JP	JP	S	
3	3	2020	Big Data Engineer	SE	FT	8.674.98500	GB	GB	M	
4	4	2020	Product Data Analyst	MI	FT	1.591.39000	HN	HN	S	

**Also we tried**

**df['x'].str.replace(',', '.')**

**And**

**df['x'].replace(',', '.')**

**But they didn't work**

**And salary in binning always became Nan so we work in Remote\_Working\_Ratio column instead.**

**\*The group work together at same time so we didnt divide the**