Name :Mariam Gaber Mohamed Elmahdy
Email : mariamgaber168@gmail.com

Report (WUZZUF JOBS PROJECT)

Our 3 datasets are :

- **2 of them about software and IT jobs .**
- **1 of them have a lot of categories not only about software /it**
- **First I will merge the first 2 datasets and make processing to take insights from them.**
- **Then I will merge the last one with them .**

---

in first data set (wuzzuf_2020), i took the important features that will help me in analysis :
Title, Company_Name, Description
i renamed the column to match the new data set names >>it helps me to merge them .

Second source of data is wuzzuf site,
i made scrubbing to get links of IT/Software jobs from the site and divide every job to :
job title, description, company name
Then I read it as a csv file and I used index_col = false to force pandas to not use the first column as the index.

Then i merged the 2 datasets

Now i want to make processing in the merged dataset:

1. **i checked null values**
2. **i know that a lot of companies don't want to write their names so i fill them by 'Hidden_Name'**
3. **make bar plot to see (The Top 10 Companies that offers IT/Software JOBS)**
4. **i Separated Levels from Description in separate column**
5. **Separate type of JOB in seperate column**
6. **Make plot that answer >> What is the most offered jobs?**
7. **Divide all jobs to 2 categories in Category column (IT, Software Engineering)**
8. **i made plot that show ('TOP CATEGORIES IN 2020-2021')**
9. **made a plot show that (what is the most levels of job are required)**
10. **finally i divide description to > Skills and Years of experience columns.**
11. **Save the data set as 'JOBS.csv' file**

Now i will take the last dataset and make processing in it then i will merge it with JOBS dataset:

1. first i dropped all columns that is not important in analysis
2. i renamed columns so that match the jobs dataset columns.
3. i combine the 3 columns of categories in one column .
4. check null values and fill them
5. Because most of opportunities is full Time

will fill nan values with full Time

6. All_Jobs['Type']= All_Jobs['Type'].fillna('Full Time')
7.