

UML & Java Programming Course – Final Exercise

Exercise Description

Data Set:

Wuzzuf jobs in Egypt data set at Kaggle

<https://www.kaggle.com/omarhanyy/wuzzuf-jobs>

Task:

- Build all java needed classes (POJO , DAO, web service and a tester client for the web service)
- Make a web service to get the following from the data set:
 1. Read data set and convert it to dataframe or Spark RDD and display some from it.
 2. Display structure and summary of the data.
 3. Clean the data (null, duplications)
 4. Count the jobs for each company and display that in order (What are the most demanding companies for jobs?)
 5. Show step 4 in a pie chart
 6. Find out What are it the most popular job titles?
 7. Show step 6 in bar chart
 8. Find out the most popular areas?
 9. Show step 8 in bar chart

10. Print skills one by one and how many each repeated and order the output to find out the most important skills required?
11. Factorize the YearsExp feature and convert it to numbers in new col. (Bounce)
12. Apply K-means for job title and companies (Bounce)

Team:

Group of three students.

Deliverables:

- Each team must share with us a git hub link for a maven EE application.
- Each team must be ready to present his work.