Test Tool

System Requirements Document

Contents

1. Introduction

1.1 Objectives

A role of a job-agency, which receives thousands of resumes every month, but does not have time to classify these.

Resumes come in, but we don't know exactly how these should be assigned. Is this applicant an engineer, HR or warehouse worker?

You are expected to help in reading the resumes and provide a revision for all of them about the position, experience, skills, courses …. In a data frame and then doing some analysis.

Outlines:

- Data Wrangling
- Data Cleaning
- Visualizations.

1.2 Data Description

Context

A collection of Resume Examples taken from livecareer.com for categorizing a given resume into any of the labels defined in the dataset.

Content

Contains 2400+ Resumes in string as well as PDF format.
PDF stored in the data folder differentiated into their respective labels as folders with each resume residing inside the folder in pdf form with filename as the id defined in the csv.

Inside the CSV:

- ID: Unique identifier and file name for the respective pdf.
- Resume_str : Contains the resume text only in string format.
- Resume_html : Contains the resume data in html format as present while web scrapping.
- Category : Category of the job the resume was used to apply.

Present categories are
HR, Designer, Information-Technology, Teacher, Advocate, Business-Development, Healthcare, Fitness, Agriculture, BPO, Sales, Consultant, Digital-Media, Automobile, Chef, Finance, Apparel, Engineering, Accountant, Construction, Public-Relations, Banking, Arts, Aviation

2. Rules

- Keep the rules of the clean code.

- Do not cheat.

- Save the project with your name.

- Have fun.

3. Functional Requirements

3.1 Task 1:

Create a descriptive csv file from the plain-text samples with the following columns:

- POSITION_TITLE: Job title.
- QUALIFICATIONS: The qualifications that each applicant had for the job
- EDUCATION _TYPE: type of education that the applicant had (master, bachelor, diploma….)
- SCHOOL_TYPE: The type of school the applicant study in (university, college, high school)
- EDUCATION_MAJOR: The education major you study in this school.
- EXPERIENCE_LENGTH: The number of experience years needed to apply for this job.
- AWORDS: If the applicant had an award or not.
- COMMUNITY SERVICE: If the applicant had a community service work or not.
- VOLUNTEERING: If the applicant had any volunteering or not.
- DRIVERS_LICENSE_AVAILABILITY: Is driver license available or not for this applicant.
- SKILLS_COUNT: number of skills that the applicant had.
- SKILLS: the skill that applicant had.
- Languages: the number of languages for each applicant.

3.2. Task 2:

Assume you are a DA in a company, and you are about to prepare a report to the job agency, by answering some questions and give insight about the data provided.

- Q1 Review the number of applicants for each position.

- Q2 A concerned parents contacts you to tell them if there is a specific school type that would guaranteed for their child a good future if you can tell, and which is it ?

- Q3 what is the most common skills among all applications?

- Q4 Based on the data that you have, could you tell the most active and community helpful applicants are for what position?

- Q5 could you nominate the best job application for each position?

- Q6 Do you think learning another language could effect the application, support your opinion.

- Q7 Give a full statistical description for all numerical data columns including all insights and needed figures to visualize them.

- Q8 Give a full statistical description for the categorical data columns that can be described including all insights and needed figures to visualize them.

4. Notes

- **PROVE ALL YOUR ANSWERS WITH GRAPHS IF AVAILABLE.**
- All problems **MUST** be solved.
- Read the problems well, before you start solving.
- The solution **MUST** pass all the test cases.
- Feel free to add more cells to develop the program, **BUT** make sure you gather them in ONE adequately working code cell .

5. Project Programming Language and Technologies

- Python.

- Jupiter Notebook.

  You can also use any other console that you are familiar with but save the

  file with .py extent.

6. Project Evaluation

- 52% Task1.

- 40% Task2.

- 8% Overall appearance and code clearance.

7. Deadline

24/12/2024-02/01/2025