A picture containing purple, pink, table

Description automatically generated

**Resume Screening**

**Problem:** Resume screening is still the most time-consuming part of recruiting. When a job opening receives 250 resumes on average 70% of them are unqualified, therefore we need a tool that Screens the most appropriate resumes from that long list on the basis of job description. That Ultimately saves a lot of human efforts and essential hours.

**Task:** Our goal here is to build resume screening model. There is one resume we have in the data you need to extract important information from it like Name, Email address, Phone number and skills.

So, using NLP you need to clean the resume and extract important keywords using regex or any other approach from the resume. Your model will be evaluated based on the following criteria:

1. Data cleaning Techniques using NLTK
2. Regex pattern for extracting important keywords
3. Scope for improvement

**Expected Outputs:**

1. The Final code (Python or R) along with proper comments
2. A summary file (PDF format) stating:
3. Problem Statement
4. Approach Taken
5. Interpretation of Results
6. Minimal but effective Storyboarding - Consider the final audience of this will be the Business team. You need to bring the actionable insights in this section.

**How to Access the Data:** Please use the below drive link to download the Resume in .pdf format and download two excel file for extracting technical and non-technical skills.

[Resume\_data](https://drive.google.com/drive/folders/1wqheqpwz1zypHbbkglfuyaOniV4JnzX2?usp=sharing)

**Hints:**

1. Load the resume data and then convert PDF to text file using “fitz” library in python
2. Apply NLP techniques in order to find the important words from resume
3. Extract phone number and email using Regex pattern.
4. Extract technical and non-technical skills from the resume using the excel sheet provided to you in the data.

**Evaluation:** The project will be evaluated based on requirements and instructions provided above.