For the Format. Json and CSV File we it should be dynamic can take in any format.

How to organise

Final output: Left

1st Process Data

Accountant ()

Clean Job Title (Removed in bracket)

Sales,

Treat them separate job titles

Typo in job title,

String match, distance…

Approximate match to mathe the title

Job Description itself

Compare words and factorise.

Focus 20% that is able to give 80%

Sentence Structure, Remove Stop words

Focus on important words, sometimes needs to factorise

Pull up the things we want ONLY

Remove and Replace

Careful on the Encoding

Roles and Responsibilities and Requirements for JD

Identify different Segments

Find Regular expressions

TO standardize, looking at 1 sentence and tokenization token but sentence.

1 Row 1 Sentence…

Rows and Responsibilities Sentence 1

Requirements..

Data Exploration

Iterations for data set for modelling

Can check on Transformer HuggingFace

A good model should be Hybrid

Simple part , build basic foundation and based of the

Should have plenty Roles and Responsibilities

What kind of input, what type of generate output? (Necessary skills varies)

After clean data…

Step 1: Job Title and sentences

Step 2: Aggregate, Look at keywords, out of 1000 JD, this particular row should be 80%.

Highest frequency. Groupby, what are the job title want to work with. Smaller subset 1st is ok.

Remove the stopwords. Those with the same important words and group together.

Group in similar meaning. Retain the keywords in separate text. Sort by descending count.

**Basic JD Recommender.**

**Linked**IN

Can focus on each industry as well.

Use that as a base.

Ideal JD for Comparisons

Slowly move to skills framework.

1. narrow down to a fixed list of job titles
2. find a way to group similar sentences together (remove stopwords, keyword extraction, vectorization etc.)
3. group and rank the similar sentences based on counts
4. recommend the highest count sentences by job title

Different Components each one can work independently.

**Next week Deliverables:**

Narrow down Job Industry…

Find Similarity…

Clean Datas whatever we can….