

```

import pandas as pd
import re
import nltk
import spacy
from num2words import num2words
from nltk import word_tokenize
from nltk.corpus import stopwords
from sklearn.feature_extraction.text import TfidfVectorizer
from nltk import pos_tag
from spacy.pipeline import EntityRecognizer

#from google.colab import drive
#drive.mount('/content/drive')

# Read the JSON file into a DataFrame
#df = pd.read_json('/content/drive/MyDrive/Resume.json', lines=True)
df = pd.read_json('Resume.json', lines=True)

# Save the DataFrame to CSV
df.to_csv('dataframe.csv', index=None)

# Print the shape of the dataset
print(f"Shape of the dataset: {df.shape}")

# Display the first few rows of the DataFrame
df.head()

```

Shape of the dataset: (200, 2)

```

                                content \
0  Govardhana K\nSenior Software Engineer\n\nBeng...
1  Harini Komaravelli\nTest Analyst at Oracle, Hy...
2  Hartej Kathuria\nData Analyst Intern - Oracle ...
3  Ijas Nizamuddin\nAssociate Consultant - State ...
4  Imgeeyaul Ansari\njava developer\n\nPune, Maha...

```

```

                                annotation
0  [{'label': ['Companies worked at'], 'points': ...
1  [{'label': ['Companies worked at'], 'points': ...
2  [{'label': ['Skills'], 'points': [{'start': 22...
3  [{'label': ['Skills'], 'points': [{'start': 46...
4  [{'label': ['Skills'], 'points': [{'start': 18...

```

```
df['content'][0]
```

```

'Govardhana K\nSenior Software Engineer\n\nBengaluru, Karnataka,
Karnataka - Email me on Indeed: indeed.com/r/Govardhana-K/\n
nb2de315d95905b68\n\nTotal IT experience 5 Years 6 Months\nCloud
Lending Solutions INC 4 Month • Salesforce Developer\nOracle 5 Years 2
Month • Core Java Developer\nLanguages Core Java, Go Lang\nOracle PL-
SQL programming,\nSales Force Developer with APEX.\n\nDesignations &

```

Promotions\n\nWilling to relocate: Anywhere\n\nWORK EXPERIENCE\n\nSenior Software Engineer\n\nCloud Lending Solutions - Bangalore, Karnataka -\n\nJanuary 2018 to Present\n\nPresent\n\nSenior Consultant\n\nOracle - Bangalore, Karnataka -\n\nNovember 2016 to December 2017\n\nStaff Consultant\n\nOracle - Bangalore, Karnataka -\n\nJanuary 2014 to October 2016\n\nAssociate Consultant\n\nOracle - Bangalore, Karnataka -\n\nNovember 2012 to December 2013\n\nEDUCATION\n\nB.E in Computer Science Engineering\n\nAdithya Institute of Technology - Tamil Nadu\n\nSeptember 2008 to June 2012\n\nhttps://www.indeed.com/r/Govardhana-K/b2de315d95905b68?isid=rex-download&ikw=download-top&co=IN\n\nhttps://www.indeed.com/r/Govardhana-K/b2de315d95905b68?isid=rex-download&ikw=download-top&co=IN\n\nSKILLS\n\nAPEX. (Less than 1 year), Data Structures (3 years), FLEXCUBE (5 years), Oracle (5 years),\n\nAlgorithms (3 years)\n\nLINKS\n\nhttps://www.linkedin.com/in/govardhana-k-61024944/\n\nADDITIONAL INFORMATION\n\nTechnical Proficiency:\n\nLanguages: Core Java, Go Lang, Data Structures & Algorithms, Oracle\n\nPL-SQL programming, Sales Force with APEX.\n\nTools: RADTool, Jdeveloper, NetBeans, Eclipse, SQL developer,\n\nPL/SQL Developer, WinSCP, Putty\n\nWeb Technologies: JavaScript, XML, HTML, Webservice\n\nOperating Systems: Linux, Windows\n\nVersion control system SVN & Git-Hub\n\nDatabases: Oracle\n\nMiddleware: Web logic, OC4J\n\nProduct FLEXCUBE: Oracle FLEXCUBE Versions 10.x, 11.x and 12.x\n\nhttps://www.linkedin.com/in/govardhana-k-61024944/'

df['annotation'][0]

```
[{'label': ['Companies worked at'],
  'points': [{'start': 1749, 'end': 1754, 'text': 'Oracle'}]},
 {'label': ['Companies worked at'],
  'points': [{'start': 1696, 'end': 1701, 'text': 'Oracle'}]},
 {'label': ['Companies worked at'],
  'points': [{'start': 1417, 'end': 1422, 'text': 'Oracle'}]},
 {'label': ['Skills'],
  'points': [{'start': 1356,
    'end': 1792,
    'text': 'Languages: Core Java, Go Lang, Data Structures & Algorithms, Oracle\nPL-SQL programming, Sales Force with APEX.\nTools: RADTool, Jdeveloper, NetBeans, Eclipse, SQL developer,\nPL/SQL Developer, WinSCP, Putty\nWeb Technologies: JavaScript, XML, HTML, Webservice\n\nOperating Systems: Linux, Windows\nVersion control system SVN & Git-Hub\nDatabases: Oracle\nMiddleware: Web logic, OC4J\nProduct FLEXCUBE: Oracle FLEXCUBE Versions 10.x, 11.x and 12.x'}]},
 {'label': ['Companies worked at'],
  'points': [{'start': 1209, 'end': 1214, 'text': 'Oracle'}]},
 {'label': ['Skills'],
  'points': [{'start': 1136,
    'end': 1247,
    'text': 'APEX. (Less than 1 year), Data Structures (3 years), FLEXCUBE (5 years), Oracle (5 years),\nAlgorithms (3 years)\n'}]},
 {'label': ['Graduation Year'],
```

```

    'points': [{ 'start': 928, 'end': 931, 'text': '2012' } ] },
    { 'label': [ 'College Name' ],
      'points': [{ 'start': 858,
                    'end': 888,
                    'text': 'Adithya Institute of Technology' } ] },
    { 'label': [ 'Degree' ],
      'points': [{ 'start': 821,
                    'end': 855,
                    'text': 'B.E in Computer Science Engineering' } ] },
    { 'label': [ 'Graduation Year' ],
      'points': [{ 'start': 787, 'end': 790, 'text': '2012' } ] },
    { 'label': [ 'Companies worked at' ],
      'points': [{ 'start': 744, 'end': 749, 'text': 'Oracle' } ] },
    { 'label': [ 'Designation' ],
      'points': [{ 'start': 722, 'end': 741, 'text': 'Associate
Consultant' } ] },
    { 'label': [ 'Companies worked at' ],
      'points': [{ 'start': 658, 'end': 663, 'text': 'Oracle' } ] },
    { 'label': [ 'Designation' ],
      'points': [{ 'start': 640, 'end': 655, 'text': 'Staff Consultant' } ] },
    { 'label': [ 'Companies worked at' ],
      'points': [{ 'start': 574, 'end': 579, 'text': 'Oracle' } ] },
    { 'label': [ 'Designation' ],
      'points': [{ 'start': 555, 'end': 572, 'text': 'Senior Consultant\
n' } ] },
    { 'label': [ 'Companies worked at' ],
      'points': [{ 'start': 470, 'end': 492, 'text': 'Cloud Lending
Solutions' } ] },
    { 'label': [ 'Designation' ],
      'points': [{ 'start': 444,
                    'end': 468,
                    'text': 'Senior Software Engineer\n' } ] },
    { 'label': [ 'Companies worked at' ],
      'points': [{ 'start': 308, 'end': 313, 'text': 'Oracle' } ] },
    { 'label': [ 'Companies worked at' ],
      'points': [{ 'start': 234, 'end': 239, 'text': 'Oracle' } ] },
    { 'label': [ 'Companies worked at' ],
      'points': [{ 'start': 175, 'end': 197, 'text': 'Cloud Lending
Solutions' } ] },
    { 'label': [ 'Email Address' ],
      'points': [{ 'start': 93,
                    'end': 136,
                    'text': 'indeed.com/r/Govardhana-K/\nb2de315d95905b68\n' } ] },
    { 'label': [ 'Location' ],
      'points': [{ 'start': 39, 'end': 47, 'text': 'Bengaluru' } ] },
    { 'label': [ 'Designation' ],
      'points': [{ 'start': 13, 'end': 37, 'text': 'Senior Software
Engineer\n' } ] },
    { 'label': [ 'Name' ],
      'points': [{ 'start': 0, 'end': 11, 'text': 'Govardhana K' } ] }

```

Data Cleaning

```
df.isnull().sum()
```

```
content      0
annotation   0
dtype: int64
```

Text Cleaning

```
def clean_text(resume):
    # Convert the text to lowercase
    resume = resume.lower()

    # Remove newlines
    resume = re.sub("\n", ' ', resume)

    # Remove special characters
    resume = re.sub(r'[,•()>□]', ' ', resume)

    # Remove extra whitespaces, dashes, and dots
    resume = re.sub(r'\s\s+|\s-\s|\.\s', ' ', resume)

    # Tokenize the text into words
    tokenized_words = resume.split(" ")

    length = len(tokenized_words)

    # Convert digits to words
    for i in range(length):
        if tokenized_words[i].isdigit():
            tokenized_words[i] = num2words(tokenized_words[i])

    # Remove stopwords
    sw = set(stopwords.words('english'))
    tokens_without_sw = []
    for w in tokenized_words:
        if w not in sw:
            tokens_without_sw.append(w)

    # Join the tokens back into a string
    final_resume = " ".join(tokens_without_sw)

    return final_resume
```

Clean content column

```
content_resumes = df['content']
```

```
for i in range(0,200):
    content_resumes[i] = clean_text(content_resumes[i])
```

```
df['content'][0]
```

'govardhana k senior software engineer bengaluru karnataka karnataka email indeed: indeed.com/r/govardhana-k/ b2de315d95905b68 total experience five years six months cloud lending solutions inc four month salesforce developer oracle five years two month core java developer languages core java go lang oracle pl-sql programming sales force developer apex designations & promotions willing relocate: anywhere work experience senior software engineer cloud lending solutions bangalore karnataka january two thousand and eighteen present present senior consultant oracle bangalore karnataka november two thousand and sixteen december two thousand and seventeen staff consultant oracle bangalore karnataka january two thousand and fourteen october two thousand and sixteen associate consultant oracle bangalore karnataka november two thousand and twelve december two thousand and thirteen education b.e computer science engineering adithya institute technology tamil nadu september two thousand and eight june two thousand and twelve <https://www.indeed.com/r/govardhana-k/b2de315d95905b68?isid=rex-download&ikw=download-top&co=in> <https://www.indeed.com/r/govardhana-k/b2de315d95905b68?isid=rex-download&ikw=download-top&co=in> skills apex less one year data structures three years flexcube five years oracle five years algorithms three years links <https://www.linkedin.com/in/govardhana-k-61024944/> additional information technical proficiency: languages: core java go lang data structures & algorithms oracle pl-sql programming sales force apex tools: radtool jdeveloper netbeans eclipse sql developer pl/sql developer winscp putty web technologies: javascript xml html webservice operating systems: linux windows version control system svn & git-hub databases: oracle middleware: web logic oc4j product flexcube: oracle flexcube versions 10.x 11.x 12.x <https://www.linkedin.com/in/govardhana-k-61024944/>

```
df['content'][1]
```

"harini komaravelli test analyst oracle hyderabad hyderabad telangana email indeed: indeed.com/r/harini-komaravelli/2659eee82e435d1b six yrs experience manual automation testing work experience qa analyst oracle test analyst oracle hyderabad infosys ltd hyderabad telangana november two thousand and eleven february two thousand and sixteen hyderabad nov two thousand and eleven feb17 two thousand and sixteen worked tata consultancy services hyderabad feb twenty-four apr eleven two thousand and seventeen currently working test analyst oracle hyderabad qa analyst six years experience oracle education mca osmania university b.sc computer science osmania university skills functional testing blue prism qtp additional information area expertise: familiar agile methodologies knowledge energy petroleum & health care domains involved preparation test scenarios preparing test data test cases <https://www.indeed.com/r/harini-komaravelli/2659eee82e435d1b?isid=rex-download&ikw=download-top&co=in> <https://www.indeed.com/r/harini-komaravelli/2659eee82e435d1b?isid=rex-download&ikw=download-top&co=in>

experienced development execution test cases effectively experienced functional testing gui testing smoke testing regression testing integration testing experienced accessibility testing application ability understand user requirements functional design specifications good knowledge sdlc stlc processes deciding severity priority bugs experience using microsoft test manager & oracle test manager test management tools good experience testing windows based & web based applications involved client interactions reviews issues clarifications web services testing writing test scripts qtp testcomplete creating object repositories function libraries qtp enhanced qtp scripts using vb script strong experience working blue prism tool worked different environments like windows application & web application technical skills: test automation tools: blue prism qtp 10.0 testcomplete test management tool: microsoft test manager oracle test manager & jira databases: oracle 10g sql server operating systems: windows seven project 1: title: cadence client: baker hughes technologies: microsoft visual studio microsoft team foundation server client background: oilfield services company delivering focused efforts shale gas oilfield services provides services tools software drilling formation evaluation well completion production management seismic data collection interpretation project description: aut application test next generation revolutionary robust easy use scalable well site data acquisition processing interpretation system client's drilling services deliver services meets cross divisional business requirements consistently project 2: description: paragon supports entire care team one tool clinicians need help deliver best patient care designed physicians nurses pharmacists mid level providers first-hand understanding clinical workflow needs paragon clinical applications allow caregivers focus matters most; spending time caring patients since paragon fully-integrated across applications built around single patient database information entered anywhere system immediately available entire care team immediate access helps clinicians make better treatment decisions also helps promote patient safety paragon offers broad suite multidisciplinary clinical software solutions together anytime anywhere access complete patient record responsibilities: performed smoke testing regression testing involved generating executing test script using quick test pro & blue prism usability user interface testing involved defect tracking reporting bugs using tfs participated frequent walk-through meetings internal quality assurance groups development groups participated client calls clarifying doubts at&t sessions involved functional regression smoke testing validate application data changes done windows application certifying build status running scripts part smoke testing project 3: description: food & beverages r&a: easily manage business across multiple locations reducing cost complexity cloud-based point-of-sale pos solutions enable centralized enterprise management lower upfront costs smaller footprint responsibilities: performed functional testing regression testing involved generating executing test scripts using blue prism tool open script involved preparing bots using blue prism tool accessibility testing web

application involved defect tracking reporting bugs using jira
webservices testing calling api's export data"

Word Analysis

*# We use the Tf-Idf vectorizer to gain insights about the
probabilities of all the possible words and collocations
we can find in these resumes.*

*# Create an instance of the Tf-Idf vectorizer with a specified ngram
range*

```
vect = TfidfVectorizer(ngram_range=(1, 3))
```

Apply the Tf-Idf vectorizer to the 'content_resumes' data

```
tf_idf = vect.fit_transform(content_resumes)
```

*# Retrieve the terms (words and collocations) in the same order as
they appear in the tf_idf matrix*

```
terms = vect.get_feature_names_out()
```

Print the tf_idf matrix

```
print(tf_idf)
```

```
(0, 344) 0.0392661003166396
(0, 272) 0.0392661003166396
(0, 137) 0.0392661003166396
(0, 118959) 0.0392661003166396
(0, 43688) 0.0392661003166396
(0, 74443) 0.0392661003166396
(0, 43683) 0.0392661003166396
(0, 82323) 0.0392661003166396
(0, 72530) 0.0392661003166396
(0, 62382) 0.0392661003166396
(0, 120304) 0.0392661003166396
(0, 67734) 0.0392661003166396
(0, 74524) 0.0392661003166396
(0, 28603) 0.0392661003166396
(0, 50074) 0.0392661003166396
(0, 46406) 0.0392661003166396
(0, 105593) 0.0392661003166396
(0, 106190) 0.0392661003166396
(0, 24520) 0.0392661003166396
(0, 118928) 0.036428206199717955
(0, 121189) 0.036428206199717955
(0, 61877) 0.036428206199717955
(0, 106376) 0.03285288244566607
(0, 73835) 0.02873890004625888
(0, 120480) 0.0392661003166396
:
(199, 53218) 0.008534517639336575
(199, 3294) 0.008942340037136233
```

```
(199, 123414) 0.01204278283231081
(199, 73208) 0.010115007446032302
(199, 99480) 0.007442812760637008
(199, 51597) 0.013272708021282891
(199, 20087) 0.013272708021282891
(199, 112776) 0.013272708021282891
(199, 50808) 0.013272708021282891
(199, 34186) 0.026545416042565783
(199, 91831) 0.013272708021282891
(199, 56215) 0.013272708021282891
(199, 123125) 0.013272708021282891
(199, 50031) 0.013272708021282891
(199, 35130) 0.006905887664991431
(199, 80667) 0.00932716400701023
(199, 6070) 0.0068035254194839054
(199, 111028) 0.0068035254194839054
(199, 121600) 0.006636354010641446
(199, 114774) 0.027214101677935622
(199, 123764) 0.02176546014260206
(199, 40348) 0.026545416042565783
(199, 20875) 0.019909062031924336
(199, 52466) 0.026545416042565783
(199, 36365) 0.006636354010641446
```

terms

```
array(['000', '000 servers', '000 servers trouble', ..., 'zxf05u01',
      'zxf05u01 validating', 'zxf05u01 validating vendor'],
      dtype=object)
```

Showing the part of speech that these terms belong to

```
nltk.pos_tag(terms)
```

```
[('000', 'CD'),
 ('000 servers', 'NNS'),
 ('000 servers trouble', 'CD'),
 ('000members', 'NNS'),
 ('000members evaluatedpatientcareneeds', 'CD'),
 ('000members evaluatedpatientcareneeds prioritizedtreatment', 'CD'),
 ('0023411a049a1441', 'CD'),
 ('0023411a049a1441 challenging', 'VBG'),
 ('0023411a049a1441 challenging career', 'CD'),
 ('0023411a049a1441 isid', 'CD'),
 ('0023411a049a1441 isid rex', 'CD'),
 ('005elab800b4cb42', 'CD'),
 ('005elab800b4cb42 isid', 'CD'),
 ('005elab800b4cb42 isid rex', 'CD'),
 ('005elab800b4cb42 work', 'CD'),
 ('005elab800b4cb42 work experience', 'CD'),
 ('00f125c7b9b95a35', 'CD'),
```


('00f125c7b9b95a35 isid', 'CD'),
('00f125c7b9b95a35 isid rex', 'CD'),
('00f125c7b9b95a35 two', 'CD'),
('00f125c7b9b95a35 two year', 'CD'),
('01', 'CD'),
('01 crm', 'CD'),
('01 crm 02', 'CD'),
('01 dopra', 'CD'),
('01 dopra description', 'CD'),
('01 ethernet', 'CD'),
('01 ethernet routing', 'VBG'),
('01 tnc', 'CD'),
('01 tnc controller', 'CD'),
('01st', 'CD'),
('01st aug', 'CD'),
('01st aug two', 'CD'),
('01st oct', 'CD'),
('01st oct two', 'CD'),
('02', 'CD'),
('02 crm', 'CD'),
('02 crm 03', 'CD'),
('02 fb50', 'CD'),
('02 fb50 gl', 'CD'),
('02 metro', 'CD'),
('02 metro ethernet', 'CD'),
('02 vsm', 'CD'),
('02 vsm ipsec', 'CD'),
('02e488f477e2f5bc', 'CD'),
('02e488f477e2f5bc currently', 'RB'),
('02e488f477e2f5bc currently pursuing', 'VBG'),
('02e488f477e2f5bc isid', 'CD'),
('02e488f477e2f5bc isid rex', 'CD'),
('03', 'CD'),
('03 education', 'CD'),
('03 education computer', 'CD'),
('03 feb', 'CD'),
('03 feb 2011', 'CD'),
('038dfd47a0cf071f', 'CD'),
('038dfd47a0cf071f isid', 'CD'),
('038dfd47a0cf071f isid rex', 'CD'),
('038dfd47a0cf071f serve', 'CD'),
('038dfd47a0cf071f serve organization', 'CD'),
('04181c5962a4af19', 'CD'),
('04181c5962a4af19 isid', 'CD'),
('04181c5962a4af19 isid rex', 'CD'),
('04181c5962a4af19 willing', 'VBG'),
('04181c5962a4af19 willing relocate', 'CD'),
('049577580b3814e6', 'CD'),
('049577580b3814e6 isid', 'CD'),
('049577580b3814e6 isid rex', 'CD'),

('049577580b3814e6 seeking', 'VBG'),
('049577580b3814e6 seeking opportunities', 'CD'),
('04a52a262175111c', 'CD'),
('04a52a262175111c around', 'CD'),
('04a52a262175111c around four', 'CD'),
('04a52a262175111c isid', 'CD'),
('04a52a262175111c isid rex', 'CD'),
('04b36892f9d2e2eb', 'CD'),
('04b36892f9d2e2eb isid', 'CD'),
('04b36892f9d2e2eb isid rex', 'CD'),
('04b36892f9d2e2eb looking', 'VBG'),
('04b36892f9d2e2eb looking challenging', 'CD'),
('04b8914a81df5a81', 'CD'),
('04b8914a81df5a81 isid', 'CD'),
('04b8914a81df5a81 isid rex', 'CD'),
('04b8914a81df5a81 project', 'CD'),
('04b8914a81df5a81 project water', 'CD'),
('066e4d4956f82ee3', 'CD'),
('066e4d4956f82ee3 isid', 'CD'),
('066e4d4956f82ee3 isid rex', 'CD'),
('066e4d4956f82ee3 work', 'CD'),
('066e4d4956f82ee3 work experience', 'CD'),
('06a16ac2d087d3c9', 'CD'),
('06a16ac2d087d3c9 isid', 'CD'),
('06a16ac2d087d3c9 isid rex', 'CD'),
('06a16ac2d087d3c9 look', 'CD'),
('06a16ac2d087d3c9 look forward', 'CD'),
('06ecf59ddac448c7', 'CD'),
('06ecf59ddac448c7 dynamic', 'CD'),
('06ecf59ddac448c7 dynamic hardworking', 'VBG'),
('06ecf59ddac448c7 isid', 'CD'),
('06ecf59ddac448c7 isid rex', 'CD'),
('07', 'CD'),
('07 2002', 'CD'),
('07 2002 till', 'CD'),
('07 asset', 'CD'),
('07 asset balances', 'NNS'),
('0771f65bfa7aff96', 'CD'),
('0771f65bfa7aff96 isid', 'CD'),
('0771f65bfa7aff96 isid rex', 'CD'),
('0771f65bfa7aff96 work', 'CD'),
('0771f65bfa7aff96 work experience', 'CD'),
('08', 'CD'),
('08 years', 'CD'),
('08 years experience', 'CD'),
('08b5b8elacd8cf07', 'CD'),
('08b5b8elacd8cf07 isid', 'CD'),
('08b5b8elacd8cf07 isid rex', 'CD'),
('08b5b8elacd8cf07 willing', 'VBG'),
('08b5b8elacd8cf07 willing relocate', 'CD'),

('0b57f5f9d35b9e5c', 'CD'),
('0b57f5f9d35b9e5c accomplished', 'CD'),
('0b57f5f9d35b9e5c accomplished professional', 'CD'),
('0b57f5f9d35b9e5c isid', 'CD'),
('0b57f5f9d35b9e5c isid rex', 'CD'),
('0cdc3284bf1bbeab', 'CD'),
('0cdc3284bf1bbeab isid', 'CD'),
('0cdc3284bf1bbeab isid rex', 'CD'),
('0cdc3284bf1bbeab work', 'CD'),
('0cdc3284bf1bbeab work experience', 'CD'),
('0da1137537d8b159', 'CD'),
('0da1137537d8b159 isid', 'CD'),
('0da1137537d8b159 isid rex', 'CD'),
('0da1137537d8b159 willing', 'VBG'),
('0da1137537d8b159 willing relocate', 'CD'),
('0in', 'CD'),
('0in fi', 'CD'),
('0in fi module', 'CD'),
('10', 'CD'),
('10 11', 'CD'),
('10 11 12', 'CD'),
('10 15', 'CD'),
('10 15 system', 'CD'),
('10 amp', 'CD'),
('10 amp 2010', 'CD'),
('10 associate', 'CD'),
('10 associate consultant', 'CD'),
('10 b2x', 'CD'),
('10 b2x process', 'CD'),
('10 banking', 'CD'),
('10 banking software', 'CD'),
('10 classroom', 'CD'),
('10 classroom job', 'CD'),
('10 database', 'CD'),
('10 database oracle', 'CD'),
('10 expertise', 'CD'),
('10 expertise role', 'CD'),
('10 four', 'CD'),
('10 four years', 'CD'),
('10 implementation', 'CD'),
('10 implementation department', 'CD'),
('10 jdbc', 'CD'),
('10 jdbc data', 'CD'),
('10 jdk1', 'CD'),
('10 jdk1 oracle', 'CD'),
('10 linux', 'CD'),
('10 linux environment', 'CD'),
('10 member', 'CD'),
('10 member software', 'CD'),
('10 module', 'CD'),

('10 module supervising', 'CD'),
('10 ms', 'CD'),
('10 ms dos', 'CD'),
('10 technical', 'JJ'),
('10 technical experience', 'CD'),
('10 testcomplete', 'CD'),
('10 testcomplete test', 'CD'),
('10 time', 'CD'),
('10 time migration', 'CD'),
('10 websphere', 'CD'),
('10 websphere apache', 'CD'),
('10 years', 'CD'),
('10 years additional', 'CD'),
('10 years agile', 'CD'),
('10 years appraisal', 'CD'),
('10 years automation', 'CD'),
('10 years aws', 'CD'),
('10 years business', 'CD'),
('10 years cloud', 'CD'),
('10 years customer', 'CD'),
('10 years database', 'CD'),
('10 years experience', 'CD'),
('10 years java', 'CD'),
('10 years links', 'CD'),
('10 years operations', 'NNS'),
('10 years product', 'CD'),
('10 years program', 'CD'),
('10 years rich', 'CD'),
('10 years risk', 'CD'),
('10 years sap', 'CD'),
('10 years scheduling', 'CD'),
('10 years scripting', 'CD'),
('10 years selenium', 'CD'),
('10 years shell', 'CD'),
('10 years sql', 'CD'),
('10 years testing', 'CD'),
('10 years ui', 'CD'),
('10 years user', 'CD'),
('10 years virtualization', 'CD'),
('100', 'CD'),
('100 annual', 'CD'),
('100 annual quota', 'CD'),
('100 bp', 'CD'),
('100 bp product', 'CD'),
('100 closure', 'CD'),
('100 closure customer', 'CD'),
('100 club', 'CD'),
('100 club award', 'CD'),
('100 effectiveness', 'CD'),
('100 effectiveness resolving', 'CD'),

('100 projects', 'NNS'),
('100 projects cd', 'CD'),
('100 renewals', 'NNS'),
('100 renewals existing', 'CD'),
('100 revenue', 'CD'),
('100 revenue target', 'CD'),
('1000', 'CD'),
('1000 cases', 'CD'),
('1000 cases snmp', 'CD'),
('101', 'CD'),
('101 kc', 'CD'),
('101 kc narayana', 'CD'),
('108623501355423636575', 'CD'),
('108623501355423636575 https', 'CD'),
('108623501355423636575 https www', 'CD'),
('10days', 'CD'),
('10days https', 'CD'),
('10days https www', 'CD'),
('10g', 'CD'),
('10g 11g', 'CD'),
('10g 11g 11g', 'CD'),
('10g 11g 12c', 'CD'),
('10g bi', 'CD'),
('10g bi publisher', 'CD'),
('10g discoverer', 'CD'),
('10g discoverer admin', 'CD'),
('10g ide', 'CD'),
('10g ide eclipse', 'CD'),
('10g mysql', 'CD'),
('10g mysql languages', 'CD'),
('10g oracle', 'CD'),
('10g oracle reports', 'CD'),
('10g oracle university', 'CD'),
('10g platform', 'CD'),
('10g platform linux', 'CD'),
('10g sql', 'CD'),
('10g sql pl', 'CD'),
('10g sql server', 'CD'),
('10g unix', 'CD'),
('10g windows', 'CD'),
('10g windows seven', 'CD'),
('10th', 'CD'),
('10th february', 'CD'),
('10th february two', 'CD'),
('10th school', 'CD'),
('10th school https', 'CD'),
('10th september', 'CD'),
('10th september two', 'CD'),
('10th std', 'CD'),
('10th std angels', 'CD'),

('10years', 'CD'),
('10years g4s', 'CD'),
('10years g4s security', 'CD'),
('11', 'CD'),
('11 12', 'CD'),
('11 12 https', 'CD'),
('11 data', 'CD'),
('11 data radio', 'CD'),
('11 describes', 'CD'),
('11 describes mac', 'CD'),
('11 java', 'CD'),
('11 java python', 'CD'),
('11 mac', 'CD'),
('11 mac split', 'CD'),
('11 operating', 'CD'),
('11 operating systems', 'CD'),
('11 oracle', 'CD'),
('11 oracle applications', 'NNS'),
('11 qos', 'CD'),
('11 qos requirements', 'NNS'),
('11 split', 'CD'),
('11 split ap', 'CD'),
('11 successfully', 'RB'),
('11 successfully completed', 'CD'),
('11g', 'CD'),
('11g 11g', 'CD'),
('11g 11g 12c', 'CD'),
('11g 12c', 'CD'),
('11g 12c database', 'CD'),
('11g 12c db', 'CD'),
('11g 12c sql', 'CD'),
('11g 12c version', 'CD'),
('11g apache', 'CD'),
('11g apache http', 'CD'),
('11g configuring', 'CD'),
('11g configuring deploying', 'CD'),
('11g database', 'CD'),
('11g database connectivity', 'CD'),
('11g devops', 'CD'),
('11g devops process', 'CD'),
('11g extensive', 'CD'),
('11g extensive knowledge', 'CD'),
('11g oracle', 'CD'),
('11g oracle fusion', 'CD'),
('11g platform', 'CD'),
('11g platform linux', 'CD'),
('11g release', 'CD'),
('11g release tools', 'CD'),
('11g years', 'CD'),
('11g years professional', 'CD'),

('11i', 'CD'),
('11i broad', 'CD'),
('11i broad knowledge', 'CD'),
('11i environment', 'CD'),
('11i environment auto', 'CD'),
('11i r12', 'CD'),
('11i r12 accounts', 'NNS'),
('11i r12 baselines', 'CD'),
('11i r12 customer', 'CD'),
('11i r12 database', 'CD'),
('11i r12 migration', 'CD'),
('11i r12 per', 'CD'),
('11i r12 upgrades', 'CD'),
('11i responsibilities', 'NNS'),
('11i responsibilities amat', 'CD'),
('11i xml', 'CD'),
('11i xml eclipse', 'CD'),
('12', 'CD'),
('12 1994', 'CD'),
('12 1994 gender', 'CD'),
('12 5m', 'CD'),
('12 5m clients', 'NNS'),
('12 baselines', 'CD'),
('12 baselines responsible', 'JJ'),
('12 formulated', 'CD'),
('12 formulated sales', 'CD'),
('12 https', 'CD'),
('12 https www', 'CD'),
('12 upgrades', 'CD'),
('12 upgrades 11i', 'CD'),
('12 years', 'CD'),
('12 years experience', 'CD'),
('12c', 'CD'),
('12c database', 'CD'),
('12c database level', 'CD'),
('12c db', 'CD'),
('12c db awarded', 'CD'),
('12c oracle', 'CD'),
('12c oracle saas', 'CD'),
('12c sql', 'CD'),
('12c sql pl', 'CD'),
('12c version', 'CD'),
('12c version adapters', 'NNS'),
('12th', 'CD'),
('12th indian', 'CD'),
('12th indian academy', 'CD'),
('13', 'CD'),
('13 years', 'CD'),
('13 years experience', 'CD'),
('135176103', 'CD'),

('135176103 certifications', 'NNS'),
('135176103 certifications licenses', 'CD'),
('135176103 https', 'CD'),
('135176103 https www', 'CD'),
('1366179051f145eb', 'CD'),
('1366179051f145eb establish', 'CD'),
('1366179051f145eb establish sincere', 'CD'),
('1366179051f145eb isid', 'CD'),
('1366179051f145eb isid rex', 'CD'),
('13th', 'CD'),
('13th february', 'CD'),
('13th february two', 'CD'),
('1404633e9449f641', 'CD'),
('1404633e9449f641 isid', 'CD'),
('1404633e9449f641 isid rex', 'CD'),
('1404633e9449f641 system', 'CD'),
('1404633e9449f641 system engineering', 'CD'),
('140749dace5dc26f', 'CD'),
('140749dace5dc26f 10', 'CD'),
('140749dace5dc26f 10 years', 'CD'),
('140749dace5dc26f isid', 'CD'),
('140749dace5dc26f isid rex', 'CD'),
('145ebled39df317c', 'CD'),
('145ebled39df317c isid', 'CD'),
('145ebled39df317c isid rex', 'CD'),
('145ebled39df317c work', 'CD'),
('145ebled39df317c work experience', 'CD'),
('1498048ada755ac3', 'CD'),
('1498048ada755ac3 cisco', 'CD'),
('1498048ada755ac3 cisco certified', 'CD'),
('1498048ada755ac3 isid', 'CD'),
('1498048ada755ac3 isid rex', 'CD'),
('14th', 'CD'),
('14th july', 'CD'),
('14th july two', 'CD'),
('15', 'CD'),
('15 higher', 'CD'),
('15 higher response', 'CD'),
('15 release', 'CD'),
('15 release enable', 'JJ'),
('15 scope', 'CD'),
('15 scope project', 'CD'),
('15 system', 'CD'),
('15 system maintenance', 'CD'),
('15 years', 'CD'),
('15 years competent', 'CD'),
('15 years experience', 'CD'),
('15 yrs', 'CD'),
('15 yrs 4yrs', 'CD'),
('152269fb5b986c26', 'CD'),

('152269fb5b986c26 isid', 'CD'),
('152269fb5b986c26 isid rex', 'CD'),
('152269fb5b986c26 secure', 'CD'),
('152269fb5b986c26 secure challenging', 'CD'),
('15906b55159d4088', 'CD'),
('15906b55159d4088 isid', 'CD'),
('15906b55159d4088 isid rex', 'CD'),
('15906b55159d4088 willing', 'CD'),
('15906b55159d4088 willing relocate', 'CD'),
('16', 'CD'),
('16 range', 'CD'),
('16 range obh2', 'CD'),
('16 role', 'CD'),
('16 role sap', 'CD'),
('16d5fa56f8c19eb6', 'CD'),
('16d5fa56f8c19eb6 isid', 'CD'),
('16d5fa56f8c19eb6 isid rex', 'CD'),
('16d5fa56f8c19eb6 work', 'CD'),
('16d5fa56f8c19eb6 work experience', 'CD'),
('16m', 'CD'),
('16m gx', 'CD'),
('16m gx developer', 'CD'),
('16th', 'CD'),
('16th august', 'CD'),
('16th august 7th', 'CD'),
('16th august two', 'CD'),
('16th june', 'CD'),
('16th june seventeen', 'CD'),
('17', 'CD'),
('17 scope', 'CD'),
('17 scope project', 'CD'),
('17thnov', 'CD'),
('17thnov 2016', 'CD'),
('17thnov 2016 member', 'CD'),
('18', 'CD'),
('18 dlink', 'CD'),
('18 dlink kvm', 'CD'),
('18b373e3b03b371f', 'CD'),
('18b373e3b03b371f isid', 'CD'),
('18b373e3b03b371f isid rex', 'CD'),
('18b373e3b03b371f technocrat', 'CD'),
('18b373e3b03b371f technocrat ten', 'CD'),
('18th', 'CD'),
('18th february', 'CD'),
('18th february two', 'CD'),
('1961c4eff806e6f4', 'CD'),
('1961c4eff806e6f4 isid', 'CD'),
('1961c4eff806e6f4 isid rex', 'CD'),
('1961c4eff806e6f4 willing', 'CD'),
('1961c4eff806e6f4 willing relocate', 'CD'),

('1962', 'CD'),
('1962 mormugao', 'CD'),
('1962 mormugao port', 'CD'),
('1994', 'CD'),
('1994 gender', 'CD'),
('1994 gender male', 'CD'),
('19d56a964e37fala', 'CD'),
('19d56a964e37fala goal', 'CD'),
('19d56a964e37fala goal oriented', 'CD'),
('19d56a964e37fala isid', 'CD'),
('19d56a964e37fala isid rex', 'CD'),
('19ff20f4b8552375', 'CD'),
('19ff20f4b8552375 isid', 'CD'),
('19ff20f4b8552375 isid rex', 'CD'),
('19ff20f4b8552375 work', 'CD'),
('19ff20f4b8552375 work experience', 'CD'),
('1b4436206cf5871b', 'CD'),
('1b4436206cf5871b isid', 'CD'),
('1b4436206cf5871b isid rex', 'CD'),
('1b4436206cf5871b work', 'CD'),
('1b4436206cf5871b work experience', 'CD'),
('1b83bc42482ed5a0', 'CD'),
('1b83bc42482ed5a0 isid', 'CD'),
('1b83bc42482ed5a0 isid rex', 'CD'),
('1b83bc42482ed5a0 prove', 'CD'),
('1b83bc42482ed5a0 prove worth', 'CD'),
('1e4b62ea17458993', 'CD'),
('1e4b62ea17458993 isid', 'CD'),
('1e4b62ea17458993 isid rex', 'CD'),
('1e4b62ea17458993 work', 'CD'),
('1e4b62ea17458993 work experience', 'CD'),
('1e7d0305af766bf6', 'CD'),
('1e7d0305af766bf6 isid', 'CD'),
('1e7d0305af766bf6 isid rex', 'CD'),
('1e7d0305af766bf6 look', 'CD'),
('1e7d0305af766bf6 look forward', 'CD'),
('1f27a18d2e4b1948', 'CD'),
('1f27a18d2e4b1948 isid', 'CD'),
('1f27a18d2e4b1948 isid rex', 'CD'),
('1f27a18d2e4b1948 quality', 'CD'),
('1f27a18d2e4b1948 quality education', 'CD'),
('1m', 'CD'),
('1m pipeline', 'CD'),
('1m pipeline ranked', 'CD'),
('1month', 'CD'),
('1month 10days', 'CD'),
('1month 10days https', 'CD'),
('1q', 'CD'),
('1q dotlad', 'CD'),
('1q dotlad based', 'CD'),

('1q dotlad tested', 'CD'),
('1q mac', 'CD'),
('1q mac aging', 'CD'),
('1st', 'CD'),
('1st aug', 'CD'),
('1st aug two', 'CD'),
('1st august', 'CD'),
('1st august two', 'CD'),
('1st february', 'CD'),
('1st february two', 'CD'),
('1st iete', 'CD'),
('1st iete student', 'CD'),
('1st july', 'CD'),
('1st july two', 'CD'),
('1st level', 'CD'),
('1st level troubleshooting', 'CD'),
('1st march', 'CD'),
('1st march two', 'CD'),
('1st year', 'CD'),
('1st year pracheen', 'CD'),
('20', 'CD'),
('20 direct', 'CD'),
('20 direct reportees', 'CD'),
('20 revenues', 'CD'),
('20 revenues rose', 'CD'),
('200', 'CD'),
('200 cases', 'CD'),
('200 cases various', 'JJ'),
('2000', 'CD'),
('2000 2003', 'CD'),
('2000 2003 two', 'CD'),
('2000 unix', 'CD'),
('2000 unix sun', 'CD'),
('2002', 'CD'),
('2002 till', 'CD'),
('2002 till date', 'CD'),
('2003', 'CD'),
('2003 r2', 'CD'),
('2003 r2 microsoft', 'CD'),
('2003 server', 'CD'),
('2003 server vista', 'CD'),
('2003 two', 'CD'),
('2003 two thousand', 'CD'),
('2004', 'CD'),
('2004 april', 'CD'),
('2004 april 2005', 'CD'),
('2005', 'CD'),
('2005 mysql', 'CD'),
('2005 mysql ms', 'CD'),
('2005 oracle', 'CD'),

('2005 oracle 11g', 'CD'),
('2005 programmer', 'CD'),
('2005 programmer analyst', 'CD'),
('2007', 'CD'),
('2007 feb', 'CD'),
('2007 feb two', 'CD'),
('2008', 'CD'),
('2008 2005', 'CD'),
('2008 2005 oracle', 'CD'),
('2008 2010', 'CD'),
('2008 2010 jr', 'CD'),
('2008 may', 'CD'),
('2008 may 2016', 'CD'),
('2008 quality', 'CD'),
('2008 quality assurance', 'CD'),
('2008 r2', 'CD'),
('2008 r2 2003', 'CD'),
('2008environment', 'CD'),
('2008environment installation', 'CD'),
('2008environment installation server', 'CD'),
('2008r2', 'CD'),
('2008r2 server', 'CD'),
('2008r2 server two', 'CD'),
('2009', 'CD'),
('2009 10', 'CD'),
('2009 10 amp', 'CD'),
('2010', 'CD'),
('2010 11', 'CD'),
('2010 11 successfully', 'CD'),
('2010 dec', 'CD'),
('2010 dec 2011', 'CD'),
('2010 jr', 'CD'),
('2010 jr developer', 'CD'),
('2010 october', 'CD'),
('2010 october two', 'CD'),
('2011', 'CD'),
('2011 june', 'CD'),
('2011 june two', 'CD'),
('2011 project', 'CD'),
('2011 project description', 'CD'),
('2012', 'CD'),
('2012 june', 'CD'),
('2012 june two', 'CD'),
('2012 microsoft', 'CD'),
('2012 microsoft sql', 'CD'),
('2012 microsoft system', 'CD'),
('2012 r2', 'CD'),
('2012 r2 2012', 'CD'),
('2012 r2 microsoft', 'CD'),
('2012r2', 'CD'),

('2012r2 exchange', 'CD'),
('2012r2 exchange two', 'CD'),
('2013', 'CD'),
('2013 december', 'CD'),
('2013 december 2013', 'CD'),
('2013 team', 'CD'),
('2013 team size', 'CD'),
('2013 two', 'CD'),
('2013 two thousand', 'CD'),
('2014', 'CD'),
('2014 jun', 'CD'),
('2014 jun 2015', 'CD'),
('2015', 'CD'),
('2015 2016', 'CD'),
('2015 2016 30', 'CD'),
('2015 present', 'CD'),
('2015 present selected', 'CD'),
('2015 team', 'CD'),
('2015 team size', 'CD'),
('2016', 'CD'),
('2016 30', 'CD'),
('2016 30 extensive', 'CD'),
('2016 may', 'CD'),
('2016 may two', 'CD'),
('2016 member', 'CD'),
('2016 member discipline', 'CD'),
('2016 recently', 'CD'),
('2016 recently technologies', 'CD'),
('2016 work', 'CD'),
('2016 work experience', 'CD'),
('2017', 'CD'),
('2017 chartered', 'CD'),
('2017 chartered level', 'CD'),
('2017 till', 'CD'),
('2017 till date', 'CD'),
('2018h2', 'CD'),
('2018h2 microsoft', 'CD'),
('2018h2 microsoft two', 'CD'),
('20th', 'JJ'),
('20th february', 'CD'),
('20th february two', 'CD'),
('22', 'CD'),
('22 billion', 'CD'),
('22 billion annual', 'CD'),
('22 concepts', 'CD'),
('22 concepts oop', 'CD'),
('23cle4e94779b465', 'CD'),
('23cle4e94779b465 isid', 'CD'),
('23cle4e94779b465 isid rex', 'CD'),
('23cle4e94779b465 willing', 'CD'),

('23cle4e94779b465 willing relocate', 'CD'),
('23rd', 'CD'),
('23rd july', 'CD'),
('23rd july two', 'CD'),
('23rd march', 'CD'),
('23rd march two', 'CD'),
('24', 'CD'),
('24 basis', 'CD'),
('24 basis deployment', 'CD'),
('24 basis involved', 'CD'),
('24 customer', 'CD'),
('24 customer service', 'CD'),
('24 inc', 'CD'),
('24 inc hr', 'CD'),
('24 office', 'CD'),
('24 office working', 'CD'),
('24 production', 'CD'),
('24 production support', 'CD'),
('24 support', 'CD'),
('24 support model', 'CD'),
('24 support resolved', 'CD'),
('24 team', 'CD'),
('24 team size', 'CD'),
('24th', 'CD'),
('24th june', 'CD'),
('24th june two', 'CD'),
('24th march', 'CD'),
('24th march two', 'CD'),
('24x7', 'CD'),
('24x7 explanation', 'CD'),
('24x7 explanation support', 'CD'),
('24x7 infrastructure', 'CD'),
('24x7 infrastructure services', 'CD'),
('25', 'CD'),
('25 involved', 'CD'),
('25 involved setup', 'CD'),
('25 posted', 'CD'),
('25 posted results', 'CD'),
('250', 'CD'),
('250 250', 'CD'),
('250 250 optik', 'CD'),
('250 optik', 'CD'),
('250 optik one', 'CD'),
('250 partners', 'CD'),
('250 partners dealing', 'CD'),
('256d6054d852b2a7', 'CD'),
('256d6054d852b2a7 isid', 'CD'),
('256d6054d852b2a7 isid rex', 'CD'),
('256d6054d852b2a7 seeking', 'CD'),
('256d6054d852b2a7 seeking challenging', 'CD'),

('25th', 'CD'),
('25th june', 'CD'),
('25th june two', 'CD'),
('25th two', 'CD'),
('25th two thousand', 'CD'),
('26', 'CD'),
('26 designed', 'CD'),
('26 designed events', 'CD'),
('260nos', 'CD'),
('260nos lab', 'CD'),
('260nos lab client', 'CD'),
('2659eee82e435d1b', 'CD'),
('2659eee82e435d1b isid', 'CD'),
('2659eee82e435d1b isid rex', 'CD'),
('2659eee82e435d1b six', 'CD'),
('2659eee82e435d1b six yrs', 'CD'),
('26f392ec8251143b', 'CD'),
('26f392ec8251143b isid', 'CD'),
('26f392ec8251143b isid rex', 'CD'),
('26f392ec8251143b willing', 'CD'),
('26f392ec8251143b willing relocate', 'CD'),
('27', 'CD'),
('27 12', 'CD'),
('27 12 1994', 'CD'),
('27 33', 'CD'),
('27 33 million', 'CD'),
('27001', 'CD'),
('27001 two', 'CD'),
('27001 two thousand', 'CD'),
('277a11151', 'CD'),
('277a11151 additional', 'CD'),
('277a11151 additional information', 'CD'),
('27b31f359c52ef76', 'CD'),
('27b31f359c52ef76 isid', 'CD'),
('27b31f359c52ef76 isid rex', 'CD'),
('27b31f359c52ef76 organized', 'CD'),
('27b31f359c52ef76 organized independent', 'CD'),
('27th', 'CD'),
('27th february', 'CD'),
('27th february two', 'CD'),
('28', 'CD'),
('28 description', 'CD'),
('28 description jabil', 'CD'),
('283106d88eb4649c', 'CD'),
('283106d88eb4649c isid', 'CD'),
('283106d88eb4649c isid rex', 'CD'),
('283106d88eb4649c willing', 'CD'),
('283106d88eb4649c willing relocate', 'CD'),
('29', 'CD'),
('29 oct', 'CD'),

('29 oct two', 'CD'),
('2bd46ce0f01fad54', 'CD'),
('2bd46ce0f01fad54 isid', 'CD'),
('2bd46ce0f01fad54 isid rex', 'CD'),
('2bd46ce0f01fad54 willing', 'CD'),
('2bd46ce0f01fad54 willing relocate', 'CD'),
('2d', 'CD'),
('2d dynamic', 'CD'),
('2d dynamic animation', 'CD'),
('2d picture', 'CD'),
('2d picture within', 'CD'),
('2d20204ef7c22049', 'CD'),
('2d20204ef7c22049 consultant', 'CD'),
('2d20204ef7c22049 consultant years', 'CD'),
('2d20204ef7c22049 isid', 'CD'),
('2d20204ef7c22049 isid rex', 'CD'),
('2d6f2e970b9a7ff6', 'CD'),
('2d6f2e970b9a7ff6 isid', 'CD'),
('2d6f2e970b9a7ff6 isid rex', 'CD'),
('2d6f2e970b9a7ff6 work', 'CD'),
('2d6f2e970b9a7ff6 work experience', 'CD'),
('2d9f28ccfa115f79', 'CD'),
('2d9f28ccfa115f79 isid', 'CD'),
('2d9f28ccfa115f79 isid rex', 'CD'),
('2d9f28ccfa115f79 willing', 'CD'),
('2d9f28ccfa115f79 willing relocate', 'CD'),
('2ed7aded59ecf425', 'CD'),
('2ed7aded59ecf425 isid', 'CD'),
('2ed7aded59ecf425 isid rex', 'CD'),
('2ed7aded59ecf425 total', 'CD'),
('2ed7aded59ecf425 total experience', 'CD'),
('2k12', 'CD'),
('2k12 attended', 'CD'),
('2k12 attended three', 'CD'),
('2months', 'CD'),
('2months 4days', 'CD'),
('2months 4days bangalore', 'CD'),
('2nd', 'CD'),
('2nd level', 'CD'),
('2nd level support', 'CD'),
('2nd sap', 'CD'),
('2nd sap server', 'CD'),
('2nd year', 'CD'),
('2nd year completed', 'CD'),
('2nd year may', 'CD'),
('30', 'CD'),
('30 07', 'CD'),
('30 07 2002', 'CD'),
('30 40', 'CD'),
('30 40 wpm', 'CD'),

('30 cost', 'CD'),
('30 cost reduction', 'CD'),
('30 extensive', 'CD'),
('30 extensive increase', 'CD'),
('30 jan', 'CD'),
('30 jan two', 'CD'),
('30 projects', 'CD'),
('30 projects within', 'CD'),
('300', 'CD'),
('300 applications', 'NNS'),
('300 applications using', 'CD'),
('300 assets', 'CD'),
('300 assets successfully', 'RB'),
('300 crm', 'CD'),
('300 crm mobile', 'CD'),
('300 simatic', 'JJ'),
('300 simatic manager', 'CD'),
('300 users', 'NNS'),
('300 users environment', 'CD'),
('30th', 'CD'),
('30th april', 'CD'),
('30th april worked', 'CD'),
('30th july', 'CD'),
('30th july two', 'CD'),
('30th june', 'CD'),
('30th june two', 'CD'),
('31', 'CD'),
('31 crm', 'CD'),
('31 crm 31', 'CD'),
('31 dell', 'CD'),
('31 dell boomi', 'CD'),
('31 eclipse', 'CD'),
('31 eclipse citrix', 'CD'),
('31 jan', 'CD'),
('31 jan two', 'CD'),
('31 period', 'CD'),
('31 period july', 'CD'),
('31 period may', 'CD'),
('31st', 'CD'),
('31st feb', 'CD'),
('31st feb two', 'CD'),
('31th', 'CD'),
('31th july', 'CD'),
('31th july seventeen', 'CD'),
('32', 'CD'),
('32 64', 'CD'),
('32 64 platforms', 'CD'),
('32 64 windows', 'CD'),
('322', 'CD'),
('322 awarded', 'CD'),

('322 awarded thrice', 'CD'),
('32472fc557546084', 'CD'),
('32472fc557546084 isid', 'CD'),
('32472fc557546084 isid rex', 'CD'),
('32472fc557546084 willing', 'VBG'),
('32472fc557546084 willing relocate', 'CD'),
('33', 'CD'),
('33 million', 'CD'),
('33 million tonnes', 'CD'),
('3573e36088ddc073', 'CD'),
('3573e36088ddc073 isid', 'CD'),
('3573e36088ddc073 isid rex', 'CD'),
('3573e36088ddc073 years', 'CD'),
('3573e36088ddc073 years professional', 'CD'),
('357ea77b3b002be6', 'CD'),
('357ea77b3b002be6 isid', 'CD'),
('357ea77b3b002be6 isid rex', 'CD'),
('357ea77b3b002be6 work', 'CD'),
('357ea77b3b002be6 work experience', 'CD'),
('3714fe32f98b03a9', 'CD'),
('3714fe32f98b03a9 isid', 'CD'),
('3714fe32f98b03a9 isid rex', 'CD'),
('3714fe32f98b03a9 quality', 'CD'),
('3714fe32f98b03a9 quality analyst', 'CD'),
('386', 'CD'),
('386 provide', 'CD'),
('386 provide pi', 'CD'),
('39c80e42cb6bc97f', 'CD'),
('39c80e42cb6bc97f isid', 'CD'),
('39c80e42cb6bc97f isid rex', 'CD'),
('39c80e42cb6bc97f test', 'CD'),
('39c80e42cb6bc97f test manager', 'CD'),
('3a382a7b7296a764', 'CD'),
('3a382a7b7296a764 4yrs', 'CD'),
('3a382a7b7296a764 4yrs solid', 'CD'),
('3a382a7b7296a764 isid', 'CD'),
('3a382a7b7296a764 isid rex', 'CD'),
('3bd9e5de546cc3c8', 'CD'),
('3bd9e5de546cc3c8 bachelor', 'CD'),
('3bd9e5de546cc3c8 bachelor computer', 'CD'),
('3bd9e5de546cc3c8 isid', 'CD'),
('3bd9e5de546cc3c8 isid rex', 'CD'),
('3c6042bd141ad353', 'CD'),
('3c6042bd141ad353 isid', 'CD'),
('3c6042bd141ad353 isid rex', 'CD'),
('3c6042bd141ad353 years', 'CD'),
('3c6042bd141ad353 years experience', 'CD'),
('3c79ad143578c3f2', 'CD'),
('3c79ad143578c3f2 isid', 'CD'),
('3c79ad143578c3f2 isid rex', 'CD'),

('3c79ad143578c3f2 years', 'CD'),
('3c79ad143578c3f2 years professional', 'CD'),
('3d', 'CD'),
('3d components', 'NNS'),
('3d components rendered', 'CD'),
('3d models', 'CD'),
('3d models pcb', 'CD'),
('3d viewer', 'CD'),
('3d viewer provided', 'CD'),
('3dgs', 'CD'),
('3dgs opengl', 'CD'),
('3dgs opengl ibm', 'CD'),
('3ecdecbcba549e21', 'CD'),
('3ecdecbcba549e21 isid', 'CD'),
('3ecdecbcba549e21 isid rex', 'CD'),
('3ecdecbcba549e21 offering', 'VBG'),
('3ecdecbcba549e21 offering twenty', 'CD'),
('3f560fd91275495b', 'CD'),
('3f560fd91275495b erudite', 'CD'),
('3f560fd91275495b erudite professional', 'CD'),
('3f560fd91275495b isid', 'CD'),
('3f560fd91275495b isid rex', 'CD'),
('3g', 'CD'),
('3g mobile', 'CD'),
('3g mobile license', 'CD'),
('3rd', 'CD'),
('3rd parties', 'NNS'),
('3rd parties third', 'CD'),
('3rd party', 'CD'),
('3rd party blogging', 'VBG'),
('3rd party service', 'CD'),
('3rd party services', 'CD'),
('3years', 'CD'),
('3years maintained', 'CD'),
('3years maintained superior', 'CD'),
('40', 'CD'),
('40 associates', 'CD'),
('40 associates working', 'CD'),
('40 wpm', 'CD'),
('400k', 'CD'),
('400k pipeline', 'CD'),
('400k pipeline account', 'CD'),
('4018c67548312089', 'CD'),
('4018c67548312089 associate', 'CD'),
('4018c67548312089 associate organization', 'CD'),
('4018c67548312089 isid', 'CD'),
('4018c67548312089 isid rex', 'CD'),
('405nos', 'CD'),
('405nos individual', 'CD'),
('405nos individual pc', 'CD'),

```
( '40m', 'CD'),
( '40m behalf', 'CD'),
( '40m behalf leading', 'CD'),
( '42', 'CD'),
( '42 mitsubishi', 'CD'),
( '42 mitsubishi fx3u', 'CD'),
( '4270d63f03e71ee8', 'CD'),
( '4270d63f03e71ee8 isid', 'CD'),
( '4270d63f03e71ee8 isid rex', 'CD'),
( '4270d63f03e71ee8 willing', 'CD'),
( '4270d63f03e71ee8 willing relocate', 'CD'),
( '42aa9e8655a5f7a3', 'CD'),
( '42aa9e8655a5f7a3 isid', 'CD'),
( '42aa9e8655a5f7a3 isid rex', 'CD'),
( '42aa9e8655a5f7a3 total', 'CD'),
( '42aa9e8655a5f7a3 total thirty', 'CD'),
( '4350', 'CD'),
( '4350 386', 'CD'),
( '4350 386 provide', 'CD'),
( '445cbf3eb0a361cd', 'CD'),
( '445cbf3eb0a361cd isid', 'CD'),
( '445cbf3eb0a361cd isid rex', 'CD'),
( '445cbf3eb0a361cd willing', 'CD'),
( '445cbf3eb0a361cd willing relocate', 'CD'),
( '445e6b4cb0b43094', 'CD'),
( '445e6b4cb0b43094 isid', 'CD'),
( '445e6b4cb0b43094 isid rex', 'CD'),
( '445e6b4cb0b43094 succeed', 'CD'),
( '445e6b4cb0b43094 succeed environment', 'CD'),
( '45', 'CD'),
( '45 spent', 'CD'),
( '45 spent data', 'CD'),
( '464', 'CD'),
...]
```

Displaying the features table, where columns are the possible mono-, bi- and tri-grams in all of the resumes.
 ### We try to identify the most common words and collocations to use in our NER model, later.

```
pd.DataFrame.sparse.from_spmatrix(tf_idf, index = content_resumes,
columns=terms)[0:2]
```

	000	000
servers \		
content		
govardhana k senior software engineer bengaluru...	0.0	0.0
harini komaravelli test analyst oracle hyderaba...	0.0	0.0

trouble \	000 servers
content	
govardhana k senior software engineer bengaluru...	
0.0	
harini komaravelli test analyst oracle hyderaba...	
0.0	
	000members \
content	
govardhana k senior software engineer bengaluru...	0.0
harini komaravelli test analyst oracle hyderaba...	0.0
	000members
evaluatedpatientcareneeds \	
content	
govardhana k senior software engineer bengaluru...	
0.0	
harini komaravelli test analyst oracle hyderaba...	
0.0	
	000members
evaluatedpatientcareneeds prioritizedtreatment \	
content	
govardhana k senior software engineer bengaluru...	
0.0	
harini komaravelli test analyst oracle hyderaba...	
0.0	
0023411a049a1441 \	
content	
govardhana k senior software engineer bengaluru...	0.0
harini komaravelli test analyst oracle hyderaba...	0.0
	0023411a049a1441
challenging \	
content	
govardhana k senior software engineer bengaluru...	
0.0	

harini komaravelli test analyst oracle hyderaba...
0.0

0023411a049a1441

challenging career \
content

govardhana k senior software engineer bengaluru...
0.0
harini komaravelli test analyst oracle hyderaba...
0.0

0023411a049a1441

isid \
content

govardhana k senior software engineer bengaluru...
0.0
harini komaravelli test analyst oracle hyderaba...
0.0

... zoom zoom

knowledge \
content

...

govardhana k senior software engineer bengaluru...
0.0
harini komaravelli test analyst oracle hyderaba...
0.0

... 0.0

... 0.0

zoom knowledge

webrtc \
content

govardhana k senior software engineer bengaluru...
0.0
harini komaravelli test analyst oracle hyderaba...
0.0

zoom supporting \
content

govardhana k senior software engineer bengaluru...
harini komaravelli test analyst oracle hyderaba...

0.0

0.0

zoom supporting

bada2 \
content

govardhana k senior software engineer bengaluru...

```
0.0
harini komaravelli test analyst oracle hyderaba...
0.0
```

```
                                zoom text  zoom
text size \
content
```

```
govardhana k senior software engineer bengaluru...    0.0
0.0
harini komaravelli test analyst oracle hyderaba...    0.0
0.0
```

```
                                zxf05u01 \
content
govardhana k senior software engineer bengaluru...    0.0
harini komaravelli test analyst oracle hyderaba...    0.0
```

```
                                zxf05u01
validating \
content
```

```
govardhana k senior software engineer bengaluru...
0.0
harini komaravelli test analyst oracle hyderaba...
0.0
```

```
                                zxf05u01
validating vendor
content
```

```
govardhana k senior software engineer bengaluru...
0.0
harini komaravelli test analyst oracle hyderaba...
0.0
```

```
[2 rows x 124345 columns]
```

```
### Final Clean Resume Sample
```

```
print(content_resumes[1])
```

```
harini komaravelli test analyst oracle hyderabad hyderabad telangana
email indeed: indeed.com/r/harini- komaravelli/2659eee82e435d1b six
yrs experience manual automation testing work experience qa analyst
oracle test analyst oracle hyderabad infosys ltd hyderabad telangana
november two thousand and eleven february two thousand and sixteen
hyderabad nov two thousand and eleven feb17 two thousand and sixteen
worked tata consultancy services hyderabad feb twenty-four apr eleven
two thousand and seventeen currently working test analyst oracle
```

hyderabad qa analyst six years experience oracle education mca osmania university b.sc computer science osmania university skills functional testing blue prism qtp additional information area expertise: familiar agile methodologies knowledge energy petroleum & health care domains involved preparation test scenarios preparing test data test cases <https://www.indeed.com/r/harini-komaravelli/2659eee82e435d1b?isid=rex-download&ikw=download-top&co=in> <https://www.indeed.com/r/harini-komaravelli/2659eee82e435d1b?isid=rex-download&ikw=download-top&co=in> experienced development execution test cases effectively experienced functional testing gui testing smoke testing regression testing integration testing experienced accessibility testing application ability understand user requirements functional design specifications good knowledge sdlc stlc processes deciding severity priority bugs experience using microsoft test manager & oracle test manager test management tools good experience testing windows based & web based applications involved client interactions reviews issues clarifications web services testing writing test scripts qtp testcomplete creating object repositories function libraries qtp enhanced qtp scripts using vb script strong experience working blue prism tool worked different environments like windows application & web application technical skills: test automation tools: blue prism qtp 10.0 testcomplete test management tool: microsoft test manager oracle test manager & jira databases: oracle 10g sql server operating systems: windows seven project 1: title: cadence client: baker hughes technologies: microsoft visual studio microsoft team foundation server client background: oilfield services company delivering focused efforts shale gas oilfield services provides services tools software drilling formation evaluation well completion production management seismic data collection interpretation project description: aut application test next generation revolutionary robust easy use scalable well site data acquisition processing interpretation system client's drilling services deliver services meets cross divisional business requirements consistently project 2: description: paragon supports entire care team one tool clinicians need help deliver best patient care designed physicians nurses pharmacists mid level providers first-hand understanding clinical workflow needs paragon clinical applications allow caregivers focus matters most; spending time caring patients since paragon fully-integrated across applications built around single patient database information entered anywhere system immediately available entire care team immediate access helps clinicians make better treatment decisions also helps promote patient safety paragon offers broad suite multidisciplinary clinical software solutions together anytime anywhere access complete patient record responsibilities: performed smoke testing regression testing involved generating executing test script using quick test pro & blue prism usability user interface testing involved defect tracking reporting bugs using tfs participated frequent walk-through meetings internal quality assurance groups development groups participated client calls clarifying doubts at&t sessions involved functional regression smoke testing validate application data changes

done windows application certifying build status running scripts part
smoke testing project 3: description: food & beverages r&a: easily
manage business across multiple locations reducing cost complexity
cloud-based point-of-sale pos solutions enable centralized enterprise
management lower upfront costs smaller footprint responsibilities:
performed functional testing regression testing involved generating
executing test scripts using blue prism tool open script involved
preparing bots using blue prism tool accessibility testing web
application involved defect tracking reporting bugs using jira
webservices testing calling api's export data

NLTK with word tokenization results in seperating urls

```
tokenized_words = word_tokenize(content_resumes[1])
```

```
nltk.pos_tag(tokenized_words)
```

```
[('harini', 'NN'),  
 ('komaravelli', 'JJ'),  
 ('test', 'NN'),  
 ('analyst', 'NN'),  
 ('oracle', 'NN'),  
 ('hyderabad', 'NN'),  
 ('hyderabad', 'NN'),  
 ('telangana', 'NN'),  
 ('email', 'VBP'),  
 ('indeed', 'RB'),  
 (':', ':'),  
 ('indeed.com/r/harini-', 'JJ'),  
 ('komaravelli/2659eee82e435d1b', 'NN'),  
 ('six', 'CD'),  
 ('yrs', 'NN'),  
 ('experience', 'NN'),  
 ('manual', 'JJ'),  
 ('automation', 'NN'),  
 ('testing', 'VBG'),  
 ('work', 'NN'),  
 ('experience', 'NN'),  
 ('qa', 'JJ'),  
 ('analyst', 'NN'),  
 ('oracle', 'NN'),  
 ('test', 'NN'),  
 ('analyst', 'NN'),  
 ('oracle', 'NN'),  
 ('hyderabad', 'NN'),  
 ('infosys', 'NN'),  
 ('ltd', 'NN'),  
 ('hyderabad', 'NN'),  
 ('telangana', 'JJ'),  
 ('november', 'RB'),
```

('two', 'CD'),
('thousand', 'NN'),
('and', 'CC'),
('eleven', 'RB'),
('february', 'JJ'),
('two', 'CD'),
('thousand', 'NN'),
('and', 'CC'),
('sixteen', 'JJ'),
('hyderabad', 'NN'),
('nov', 'RB'),
('two', 'CD'),
('thousand', 'NN'),
('and', 'CC'),
('eleven', 'RB'),
('feb17', 'JJ'),
('two', 'CD'),
('thousand', 'NN'),
('and', 'CC'),
('sixteen', 'JJ'),
('worked', 'VBN'),
('tata', 'NN'),
('consultancy', 'NN'),
('services', 'NNS'),
('hyderabad', 'VBP'),
('feb', 'JJ'),
('twenty-four', 'JJ'),
('apr', 'NNS'),
('eleven', 'RB'),
('two', 'CD'),
('thousand', 'NN'),
('and', 'CC'),
('seventeen', 'JJ'),
('currently', 'RB'),
('working', 'VBG'),
('test', 'NN'),
('analyst', 'NN'),
('oracle', 'NN'),
('hyderabad', 'NN'),
('qa', 'JJ'),
('analyst', 'NN'),
('six', 'CD'),
('years', 'NNS'),
('experience', 'RB'),
('oracle', 'VBP'),
('education', 'NN'),
('mca', 'NN'),
('osmania', 'VBP'),
('university', 'NN'),
('b.sc', 'NN'),

('computer', 'NN'),
('science', 'NN'),
('osmania', 'CD'),
('university', 'NN'),
('skills', 'NNS'),
('functional', 'JJ'),
('testing', 'VBG'),
('blue', 'JJ'),
('prism', 'NN'),
('qtp', 'VBD'),
('additional', 'JJ'),
('information', 'NN'),
('area', 'NN'),
('expertise', 'NN'),
(':', ':'),
('familiar', 'JJ'),
('agile', 'IN'),
('methodologies', 'NNS'),
('knowledge', 'VBP'),
('energy', 'NN'),
('petroleum', 'NN'),
('&', 'CC'),
('health', 'NN'),
('care', 'NN'),
('domains', 'VBZ'),
('involved', 'VBN'),
('preparation', 'NN'),
('test', 'NN'),
('scenarios', 'NNS'),
('preparing', 'VBG'),
('test', 'NN'),
('data', 'NNS'),
('test', 'NN'),
('cases', 'NNS'),
('https', 'NN'),
(':', ':'),
('//www.indeed.com/r/harini-komaravelli/2659eee82e435d1b', 'JJ'),
('?', '.'),
('isid=rex-download', 'JJ'),
('&', 'CC'),
('ikw=download-top', 'JJ'),
('&', 'CC'),
('co=in', 'JJ'),
('https', 'NN'),
(':', ':'),
('//www.indeed.com/r/harini-komaravelli/2659eee82e435d1b', 'JJ'),
('?', '.'),
('isid=rex-download', 'JJ'),
('&', 'CC'),
('ikw=download-top', 'JJ'),

('&', 'CC'),
('co=in', 'NN'),
('experienced', 'VBD'),
('development', 'NN'),
('execution', 'NN'),
('test', 'NN'),
('cases', 'NNS'),
('effectively', 'RB'),
('experienced', 'VBD'),
('functional', 'JJ'),
('testing', 'VBG'),
('gui', 'JJ'),
('testing', 'VBG'),
('smoke', 'NN'),
('testing', 'VBG'),
('regression', 'NN'),
('testing', 'VBG'),
('integration', 'NN'),
('testing', 'VBG'),
('experienced', 'JJ'),
('accessibility', 'NN'),
('testing', 'VBG'),
('application', 'NN'),
('ability', 'NN'),
('understand', 'VBP'),
('user', 'NN'),
('requirements', 'NNS'),
('functional', 'JJ'),
('design', 'NN'),
('specifications', 'NNS'),
('good', 'JJ'),
('knowledge', 'NN'),
('sdlc', 'NN'),
('stlc', 'NN'),
('processes', 'VBZ'),
('deciding', 'VBG'),
('severity', 'NN'),
('priority', 'NN'),
('bugs', 'NNS'),
('experience', 'NN'),
('using', 'VBG'),
('microsoft', 'JJ'),
('test', 'NN'),
('manager', 'NN'),
('&', 'CC'),
('oracle', 'NN'),
('test', 'NN'),
('manager', 'NN'),
('test', 'NN'),
('management', 'NN'),

('tools', 'RB'),
('good', 'JJ'),
('experience', 'NN'),
('testing', 'VBG'),
('windows', 'NNS'),
('based', 'VBN'),
('&', 'CC'),
('web', 'NNS'),
('based', 'VBN'),
('applications', 'NNS'),
('involved', 'VBN'),
('client', 'NN'),
('interactions', 'NNS'),
('reviews', 'VBP'),
('issues', 'NNS'),
('clarifications', 'NNS'),
('web', 'VBP'),
('services', 'NNS'),
('testing', 'VBG'),
('writing', 'VBG'),
('test', 'NN'),
('scripts', 'NNS'),
('qtp', 'VBP'),
('testcomplete', 'JJ'),
('creating', 'VBG'),
('object', 'JJ'),
('repositories', 'NNS'),
('function', 'NN'),
('libraries', 'NNS'),
('qtp', 'VBP'),
('enhanced', 'VBN'),
('qtp', 'JJ'),
('scripts', 'NNS'),
('using', 'VBG'),
('vb', 'JJ'),
('script', 'NN'),
('strong', 'JJ'),
('experience', 'NN'),
('working', 'VBG'),
('blue', 'JJ'),
('prism', 'NN'),
('tool', 'NN'),
('worked', 'VBD'),
('different', 'JJ'),
('environments', 'NNS'),
('like', 'IN'),
('windows', 'NNS'),
('application', 'NN'),
('&', 'CC'),
('web', 'JJ'),

('application', 'NN'),
('technical', 'JJ'),
('skills', 'NNS'),
(':', ':'),
('test', 'NN'),
('automation', 'NN'),
('tools', 'NNS'),
(':', ':'),
('blue', 'JJ'),
('prism', 'NN'),
('qtp', 'VBD'),
('10.0', 'CD'),
('testcomplete', 'JJ'),
('test', 'NN'),
('management', 'NN'),
('tool', 'NN'),
(':', ':'),
('microsoft', 'JJ'),
('test', 'NN'),
('manager', 'NN'),
('oracle', 'VBP'),
('test', 'NN'),
('manager', 'NN'),
('&', 'CC'),
('jira', 'NN'),
('databases', 'NNS'),
(':', ':'),
('oracle', 'NN'),
('10g', 'CD'),
('sql', 'NN'),
('server', 'NN'),
('operating', 'VBG'),
('systems', 'NNS'),
(':', ':'),
('windows', 'NNS'),
('seven', 'CD'),
('project', 'NN'),
('1', 'CD'),
(':', ':'),
('title', 'NN'),
(':', ':'),
('cadence', 'NN'),
('client', 'NN'),
(':', ':'),
('baker', 'NN'),
('hughes', 'VBZ'),
('technologies', 'NNS'),
(':', ':'),
('microsoft', 'JJ'),
('visual', 'JJ'),

('studio', 'NN'),
('microsoft', 'NN'),
('team', 'NN'),
('foundation', 'NN'),
('server', 'NN'),
('client', 'NN'),
('background', 'NN'),
(':', ':'),
('oilfield', 'NN'),
('services', 'NNS'),
('company', 'NN'),
('delivering', 'VBG'),
('focused', 'VBD'),
('efforts', 'NNS'),
('shale', 'NN'),
('gas', 'NN'),
('oilfield', 'NN'),
('services', 'NNS'),
('provides', 'VBZ'),
('services', 'NNS'),
('tools', 'NNS'),
('software', 'NN'),
('drilling', 'VBG'),
('formation', 'NN'),
('evaluation', 'NN'),
('well', 'RB'),
('completion', 'NN'),
('production', 'NN'),
('management', 'NN'),
('seismic', 'JJ'),
('data', 'NNS'),
('collection', 'NN'),
('interpretation', 'NN'),
('project', 'NN'),
('description', 'NN'),
(':', ':'),
('aut', 'JJ'),
('application', 'NN'),
('test', 'NN'),
('next', 'JJ'),
('generation', 'NN'),
('revolutionary', 'JJ'),
('robust', 'JJ'),
('easy', 'JJ'),
('use', 'NN'),
('scalable', 'JJ'),
('well', 'RB'),
('site', 'NN'),
('data', 'NNS'),
('acquisition', 'NN'),

('processing', 'VBG'),
('interpretation', 'NN'),
('system', 'NN'),
('client', 'NN'),
('s', 'POS'),
('drilling', 'NN'),
('services', 'NNS'),
('deliver', 'NN'),
('services', 'NNS'),
('meets', 'VBZ'),
('cross', 'JJ'),
('divisional', 'NN'),
('business', 'NN'),
('requirements', 'NNS'),
('consistently', 'RB'),
('project', 'VBP'),
('2', 'CD'),
(':', ':'),
('description', 'NN'),
(':', ':'),
('paragon', 'NN'),
('supports', 'NNS'),
('entire', 'JJ'),
('care', 'NN'),
('team', 'NN'),
('one', 'CD'),
('tool', 'NN'),
('clinicians', 'NNS'),
('need', 'VBP'),
('help', 'VB'),
('deliver', 'VB'),
('best', 'JJS'),
('patient', 'NN'),
('care', 'NN'),
('designed', 'VBN'),
('physicians', 'NNS'),
('nurses', 'NNS'),
('pharmacists', 'NNS'),
('mid', 'VBP'),
('level', 'JJ'),
('providers', 'NNS'),
('first-hand', 'VBP'),
('understanding', 'JJ'),
('clinical', 'JJ'),
('workflow', 'NN'),
('needs', 'NNS'),
('paragon', 'VBP'),
('clinical', 'JJ'),
('applications', 'NNS'),
('allow', 'VBP'),

('caregivers', 'NNS'),
('focus', 'VBP'),
('matters', 'NNS'),
('most', 'RBS'),
(';', ':'),
('spending', 'NN'),
('time', 'NN'),
('caring', 'VBG'),
('patients', 'NNS'),
('since', 'IN'),
('paragon', 'NN'),
('fully-integrated', 'JJ'),
('across', 'IN'),
('applications', 'NNS'),
('built', 'VBN'),
('around', 'IN'),
('single', 'JJ'),
('patient', 'NN'),
('database', 'NN'),
('information', 'NN'),
('entered', 'VBD'),
('anywhere', 'RB'),
('system', 'NN'),
('immediately', 'RB'),
('available', 'JJ'),
('entire', 'JJ'),
('care', 'NN'),
('team', 'NN'),
('immediate', 'JJ'),
('access', 'NN'),
('helps', 'VBZ'),
('clinicians', 'NNS'),
('make', 'VBP'),
('better', 'JJR'),
('treatment', 'NN'),
('decisions', 'NNS'),
('also', 'RB'),
('helps', 'VBZ'),
('promote', 'VB'),
('patient', 'JJ'),
('safety', 'NN'),
('paragon', 'NN'),
('offers', 'NNS'),
('broad', 'JJ'),
('suite', 'NN'),
('multidisciplinary', 'JJ'),
('clinical', 'JJ'),
('software', 'NN'),
('solutions', 'NNS'),
('together', 'RB'),

('anytime', 'RB'),
('anywhere', 'JJ'),
('access', 'NN'),
('complete', 'JJ'),
('patient', 'NN'),
('record', 'NN'),
('responsibilities', 'NNS'),
(':', ':'),
('performed', 'VBN'),
('smoke', 'NN'),
('testing', 'VBG'),
('regression', 'NN'),
('testing', 'VBG'),
('involved', 'VBN'),
('generating', 'VBG'),
('executing', 'VBG'),
('test', 'NN'),
('script', 'NN'),
('using', 'VBG'),
('quick', 'JJ'),
('test', 'NN'),
('pro', 'NN'),
('&', 'CC'),
('blue', 'JJ'),
('prism', 'NN'),
('usability', 'NN'),
('user', 'JJ'),
('interface', 'NN'),
('testing', 'VBG'),
('involved', 'VBN'),
('defect', 'JJ'),
('tracking', 'VBG'),
('reporting', 'NN'),
('bugs', 'NNS'),
('using', 'VBG'),
('tfs', 'NN'),
('participated', 'VBD'),
('frequent', 'JJ'),
('walk-through', 'JJ'),
('meetings', 'NNS'),
('internal', 'JJ'),
('quality', 'NN'),
('assurance', 'NN'),
('groups', 'NNS'),
('development', 'NN'),
('groups', 'NNS'),
('participated', 'VBD'),
('client', 'NN'),
('calls', 'NNS'),
('clarifying', 'VBG'),

('doubts', 'NNS'),
('at', 'IN'),
('&', 'CC'),
('t', 'JJ'),
('sessions', 'NNS'),
('involved', 'VBN'),
('functional', 'JJ'),
('regression', 'NN'),
('smoke', 'VBD'),
('testing', 'VBG'),
('validate', 'NN'),
('application', 'NN'),
('data', 'NNS'),
('changes', 'NNS'),
('done', 'VBN'),
('windows', 'VBZ'),
('application', 'NN'),
('certifying', 'VBG'),
('build', 'JJ'),
('status', 'NN'),
('running', 'VBG'),
('scripts', 'NNS'),
('part', 'NN'),
('smoke', 'VBD'),
('testing', 'VBG'),
('project', 'NN'),
('3', 'CD'),
(':', ':'),
('description', 'NN'),
(':', ':'),
('food', 'NN'),
('&', 'CC'),
('beverages', 'NNS'),
('r', 'VBP'),
('&', 'CC'),
('a', 'DT'),
(':', ':'),
('easily', 'RB'),
('manage', 'NN'),
('business', 'NN'),
('across', 'IN'),
('multiple', 'JJ'),
('locations', 'NNS'),
('reducing', 'VBG'),
('cost', 'NN'),
('complexity', 'NN'),
('cloud-based', 'JJ'),
('point-of-sale', 'JJ'),
('pos', 'NN'),
('solutions', 'NNS'),

('enable', 'JJ'),
('centralized', 'JJ'),
('enterprise', 'NN'),
('management', 'NN'),
('lower', 'JJR'),
('upfront', 'NN'),
('costs', 'NNS'),
('smaller', 'JJR'),
('footprint', 'JJ'),
('responsibilities', 'NNS'),
(':', ':'),
('performed', 'VBN'),
('functional', 'JJ'),
('testing', 'VBG'),
('regression', 'NN'),
('testing', 'VBG'),
('involved', 'VBN'),
('generating', 'VBG'),
('executing', 'VBG'),
('test', 'NN'),
('scripts', 'NNS'),
('using', 'VBG'),
('blue', 'JJ'),
('prism', 'NN'),
('tool', 'NN'),
('open', 'JJ'),
('script', 'NN'),
('involved', 'VBN'),
('preparing', 'VBG'),
('bots', 'NNS'),
('using', 'VBG'),
('blue', 'JJ'),
('prism', 'NN'),
('tool', 'NN'),
('accessibility', 'NN'),
('testing', 'VBG'),
('web', 'JJ'),
('application', 'NN'),
('involved', 'VBN'),
('defect', 'JJ'),
('tracking', 'VBG'),
('reporting', 'NN'),
('bugs', 'NNS'),
('using', 'VBG'),
('jira', 'NN'),
('webservices', 'NNS'),
('testing', 'VBG'),
('calling', 'VBG'),
('api', 'NN'),
('"s"', 'POS'),

```
('export', 'NN'),  
('data', 'NNS')]
```

Word Tagging

```
def tag_words(text):  
    # Load the English language model in spaCy  
    nlp = spacy.load("en_core_web_sm")  
  
    # Process the text with the language model  
    words = nlp(text)  
  
    tagged = []  
  
    # Iterate over each word in the processed text  
    for word in words:  
        # Print the word and its part-of-speech tag  
        print(word, word.pos_)  
  
        # Append the word and its part-of-speech tag to the tagged  
        list tagged.append((word.text, word.pos_))  
  
    return tagged
```

Using spacy tagging is prone to errors

```
tag_words(content_resumes[1])  
  
harini PROPN  
komaravelli PROPN  
test PROPN  
analyst PROPN  
oracle PROPN  
hyderabad PROPN  
hyderabad PROPN  
telangana PROPN  
email NOUN  
indeed ADV  
: PUNCT  
indeed.com/r/harini- PROPN  
komaravelli/2659eee82e435d1b PROPN  
six NUM  
yrs NOUN  
experience NOUN  
manual ADJ  
automation NOUN  
testing VERB  
SPACE  
work NOUN  
experience NOUN
```

qa NOUN
analyst NOUN
oracle PROPN
test PROPN
analyst NOUN
oracle PROPN
hyderabad PROPN
infosys PROPN
ltd PROPN
SPACE
hyderabad PROPN
telangana PROPN
SPACE
november PROPN
two NUM
thousand NUM
and CONJ
eleven NUM
february NOUN
two NUM
thousand NUM
and CONJ
sixteen NUM
hyderabad NOUN
nov PROPN
two NUM
thousand NUM
and CONJ
eleven NUM
feb17 NUM
two NUM
thousand NUM
and CONJ
sixteen NUM
worked VERB
tata NOUN
consultancy NOUN
services NOUN
hyderabad PROPN
feb PROPN
twenty PROPN
- PUNCT
four NUM
apr NOUN
eleven NUM
two NUM
thousand NUM
and CONJ
seventeen NUM
currently ADV

working VERB
test NOUN
analyst NOUN
oracle PROPN
hyderabad PROPN
qa PROPN
analyst NOUN
six NUM
years NOUN
experience NOUN
oracle PROPN
education PROPN
mca PROPN
osmania PROPN
university PROPN
b.sc PROPN
computer PROPN
science PROPN
osmania PROPN
university PROPN
skills VERB
functional ADJ
testing NOUN
blue ADJ
prism NOUN
qtp VERB
additional ADJ
information NOUN
area NOUN
expertise NOUN
: PUNCT
familiar ADJ
agile ADJ
methodologies NOUN
SPACE
knowledge NOUN
energy PROPN
petroleum PROPN
& CCONJ
health NOUN
care NOUN
domains NOUN
SPACE
involved VERB
preparation NOUN
test NOUN
scenarios NOUN
SPACE
preparing VERB
test NOUN

data NOUN
test NOUN
cases NOUN
SPACE
<https://www.indeed.com/r/harini-komaravelli/2659eee82e435d1b?isid=rex-download&ikw=download-top&co=in> NOUN
<https://www.indeed.com/r/harini-komaravelli/2659eee82e435d1b?isid=rex-download&ikw=download-top&co=in> NOUN
experienced VERB
development NOUN
execution NOUN
test NOUN
cases NOUN
effectively ADV
SPACE
experienced VERB
functional ADJ
testing NOUN
gui NOUN
testing NOUN
smoke NOUN
testing NOUN
regression NOUN
testing NOUN
integration NOUN
testing NOUN
experienced VERB
accessibility NOUN
testing NOUN
application NOUN
ability NOUN
understand VERB
user NOUN
requirements NOUN
functional ADJ
design NOUN
specifications NOUN
SPACE
good ADJ
knowledge NOUN
sdlc NOUN
stlc NOUN
processes NOUN
SPACE
deciding VERB
severity NOUN
priority NOUN
bugs NOUN
SPACE
experience NOUN

using VERB
microsoft PROPN
test NOUN
manager NOUN
& CONJ
oracle PROPN
test NOUN
manager NOUN
test PROPN
management PROPN
tools VERB
SPACE
good ADJ
experience NOUN
testing VERB
windows NOUN
based VERB
& CONJ
web NOUN
based VERB
applications NOUN
SPACE
involved VERB
client NOUN
interactions NOUN
reviews VERB
issues NOUN
clarifications NOUN
SPACE
web NOUN
services NOUN
testing NOUN
writing VERB
test NOUN
scripts NOUN
qtp VERB
testcomplete ADV
SPACE
creating VERB
object ADJ
repositories NOUN
function NOUN
libraries NOUN
qtp VERB
SPACE
enhanced VERB
qtp NOUN
scripts NOUN
using VERB
vb PROPN

script NOUN
SPACE
strong ADJ
experience NOUN
working VERB
blue ADJ
prism NOUN
tool NOUN
worked VERB
different ADJ
environments NOUN
like ADP
windows PROP
application NOUN
& CONJ
web NOUN
application NOUN
technical ADJ
skills NOUN
: PUNCT
test NOUN
automation NOUN
tools NOUN
: PUNCT
blue ADJ
prism NOUN
qtp VERB
10.0 NUM
testcomplete ADJ
test NOUN
management NOUN
tool NOUN
: PUNCT
microsoft PROP
test NOUN
manager NOUN
oracle PROP
test PROP
manager PROP
& CONJ
jira PROP
databases VERB
: PUNCT
oracle NOUN
10 NUM
g NOUN
sql ADJ
server NOUN
SPACE
operating NOUN

systems NOUN
: PUNCT
windows VERB
seven NUM
project NOUN
1 NUM
: PUNCT
title NOUN
: PUNCT
cadence NOUN
client NOUN
: PUNCT
baker NOUN
hughes VERB
technologies NOUN
: PUNCT
microsoft PROPN
visual PROPN
studio NOUN
microsoft PROPN
team PROPN
foundation PROPN
server NOUN
client NOUN
background NOUN
: PUNCT
oilfield NOUN
services NOUN
company NOUN
delivering VERB
focused VERB
efforts NOUN
shale NOUN
gas NOUN
oilfield NOUN
services NOUN
provides VERB
services NOUN
tools NOUN
software NOUN
drilling NOUN
formation NOUN
evaluation NOUN
well NOUN
completion NOUN
production NOUN
management NOUN
seismic ADJ
data NOUN
collection NOUN

interpretation NOUN
SPACE
project NOUN
description NOUN
: PUNCT
aut NOUN
application NOUN
test NOUN
next ADJ
generation NOUN
revolutionary ADJ
robust ADJ
easy ADJ
use NOUN
scalable ADJ
well ADJ
site NOUN
data NOUN
acquisition NOUN
processing VERB
interpretation NOUN
system NOUN
client NOUN
's PART
drilling NOUN
services NOUN
deliver VERB
services NOUN
meets VERB
cross VERB
divisional ADJ
business NOUN
requirements NOUN
consistently ADV
SPACE
project NOUN
2 NUM
: PUNCT
description NOUN
: PUNCT
paragon PROPN
supports VERB
entire ADJ
care NOUN
team NOUN
one NUM
tool NOUN
clinicians NOUN
need VERB
help NOUN

deliver VERB
best ADJ
patient NOUN
care NOUN
designed VERB
physicians NOUN
nurses VERB
pharmacists NOUN
mid ADJ
level NOUN
providers NOUN
first ADJ
- PUNCT
hand NOUN
understanding NOUN
clinical ADJ
workflow NOUN
needs VERB
paragon PROPN
clinical ADJ
applications NOUN
allow VERB
caregivers NOUN
focus VERB
matters NOUN
most ADV
; PUNCT
spending VERB
time NOUN
caring VERB
patients NOUN
since SCONJ
paragon ADJ
fully ADV
- PUNCT
integrated VERB
across ADP
applications NOUN
built VERB
around ADP
single ADJ
patient NOUN
database NOUN
information NOUN
entered VERB
anywhere ADJ
system NOUN
immediately ADV
available ADJ
entire ADJ

care NOUN
team NOUN
immediate ADJ
access NOUN
helps VERB
clinicians NOUN
make VERB
better ADJ
treatment NOUN
decisions NOUN
also ADV
helps VERB
promote VERB
patient ADJ
safety NOUN
paragon NOUN
offers VERB
broad ADJ
suite NOUN
multidisciplinary ADJ
clinical ADJ
software NOUN
solutions NOUN
together ADV
anytime ADV
anywhere ADV
access NOUN
complete ADJ
patient NOUN
record NOUN
SPACE
responsibilities NOUN
: PUNCT
performed VERB
smoke NOUN
testing NOUN
regression NOUN
testing NOUN
SPACE
involved VERB
generating VERB
executing VERB
test NOUN
script NOUN
using VERB
quick ADJ
test NOUN
pro ADJ
& CCONJ
blue ADJ

prism PROP
usability NOUN
user NOUN
interface NOUN
testing NOUN
SPACE
involved VERB
defect NOUN
tracking NOUN
reporting VERB
bugs NOUN
using VERB
tfs PROP
participated VERB
frequent ADJ
walk NOUN
- PUNCT
through ADP
meetings NOUN
internal ADJ
quality NOUN
assurance NOUN
groups NOUN
development NOUN
groups NOUN
SPACE
participated VERB
client NOUN
calls VERB
clarifying VERB
doubts PROP
at&t PROP
sessions NOUN
involved VERB
functional ADJ
regression NOUN
smoke NOUN
testing VERB
validate NOUN
application NOUN
data NOUN
changes NOUN
done VERB
windows NOUN
application NOUN
certifying VERB
build NOUN
status NOUN
running VERB
scripts NOUN

part NOUN
smoke NOUN
testing NOUN
project NOUN
3 NUM
: PUNCT
description NOUN
: PUNCT
food NOUN
& CCONJ
beverages NOUN
r&a PROPN
: PUNCT
easily ADV
manage VERB
business NOUN
across ADP
multiple ADJ
locations NOUN
reducing VERB
cost NOUN
complexity NOUN
cloud NOUN
- PUNCT
based VERB
point NOUN
- PUNCT
of ADP
- PUNCT
sale NOUN
pos NOUN
solutions NOUN
enable VERB
centralized VERB
enterprise NOUN
management NOUN
lower ADJ
upfront ADJ
costs NOUN
smaller ADJ
footprint NOUN
SPACE
responsibilities NOUN
: PUNCT
performed VERB
functional ADJ
testing NOUN
regression NOUN
testing NOUN
SPACE

involved VERB
generating VERB
executing VERB
test NOUN
scripts NOUN
using VERB
blue ADJ
prism NOUN
tool NOUN
open ADJ
script NOUN
involved VERB
preparing VERB
bots NOUN
using VERB
blue ADJ
prism NOUN
tool NOUN
SPACE
accessibility NOUN
testing NOUN
web NOUN
application NOUN
involved VERB
defect NOUN
tracking NOUN
reporting VERB
bugs NOUN
using VERB
jira PROP
webservices NOUN
testing NOUN
calling VERB
api NOUN
's PART
export NOUN
data NOUN

[('harini', 'PROP'),
('komaravelli', 'PROP'),
('test', 'PROP'),
('analyst', 'PROP'),
('oracle', 'PROP'),
('hyderabad', 'PROP'),
('hyderabad', 'PROP'),
('telangana', 'PROP'),
('email', 'NOUN'),
('indeed', 'ADV'),
(':', 'PUNCT'),
('indeed.com/r/harini-', 'PROP'),

('komaravelli/2659eee82e435d1b', 'PROPN'),
('six', 'NUM'),
('yrs', 'NOUN'),
('experience', 'NOUN'),
('manual', 'ADJ'),
('automation', 'NOUN'),
('testing', 'VERB'),
(' ', 'SPACE'),
('work', 'NOUN'),
('experience', 'NOUN'),
('qa', 'NOUN'),
('analyst', 'NOUN'),
('oracle', 'PROPN'),
('test', 'PROPN'),
('analyst', 'NOUN'),
('oracle', 'PROPN'),
('hyderabad', 'PROPN'),
('infosys', 'PROPN'),
('ltd', 'PROPN'),
(' ', 'SPACE'),
('hyderabad', 'PROPN'),
('telangana', 'PROPN'),
(' ', 'SPACE'),
('november', 'PROPN'),
('two', 'NUM'),
('thousand', 'NUM'),
('and', 'CCONJ'),
('eleven', 'NUM'),
('february', 'NOUN'),
('two', 'NUM'),
('thousand', 'NUM'),
('and', 'CCONJ'),
('sixteen', 'NUM'),
('hyderabad', 'NOUN'),
('nov', 'PROPN'),
('two', 'NUM'),
('thousand', 'NUM'),
('and', 'CCONJ'),
('eleven', 'NUM'),
('feb17', 'NUM'),
('two', 'NUM'),
('thousand', 'NUM'),
('and', 'CCONJ'),
('sixteen', 'NUM'),
('worked', 'VERB'),
('tata', 'NOUN'),
('consultancy', 'NOUN'),
('services', 'NOUN'),
('hyderabad', 'PROPN'),
('feb', 'PROPN'),

('twenty', 'PROPN'),
('-', 'PUNCT'),
('four', 'NUM'),
('apr', 'NOUN'),
('eleven', 'NUM'),
('two', 'NUM'),
('thousand', 'NUM'),
('and', 'CCONJ'),
('seventeen', 'NUM'),
('currently', 'ADV'),
('working', 'VERB'),
('test', 'NOUN'),
('analyst', 'NOUN'),
('oracle', 'PROPN'),
('hyderabad', 'PROPN'),
('qa', 'PROPN'),
('analyst', 'NOUN'),
('six', 'NUM'),
('years', 'NOUN'),
('experience', 'NOUN'),
('oracle', 'PROPN'),
('education', 'PROPN'),
('mca', 'PROPN'),
('osmania', 'PROPN'),
('university', 'PROPN'),
('b.sc', 'PROPN'),
('computer', 'PROPN'),
('science', 'PROPN'),
('osmania', 'PROPN'),
('university', 'PROPN'),
('skills', 'VERB'),
('functional', 'ADJ'),
('testing', 'NOUN'),
('blue', 'ADJ'),
('prism', 'NOUN'),
('qtp', 'VERB'),
('additional', 'ADJ'),
('information', 'NOUN'),
('area', 'NOUN'),
('expertise', 'NOUN'),
(':', 'PUNCT'),
('familiar', 'ADJ'),
('agile', 'ADJ'),
('methodologies', 'NOUN'),
(' ', 'SPACE'),
('knowledge', 'NOUN'),
('energy', 'PROPN'),
('petroleum', 'PROPN'),
('&', 'CCONJ'),
('health', 'NOUN'),

('care', 'NOUN'),
('domains', 'NOUN'),
(' ', 'SPACE'),
('involved', 'VERB'),
('preparation', 'NOUN'),
('test', 'NOUN'),
('scenarios', 'NOUN'),
(' ', 'SPACE'),
('preparing', 'VERB'),
('test', 'NOUN'),
('data', 'NOUN'),
('test', 'NOUN'),
('cases', 'NOUN'),
(' ', 'SPACE'),
('https://www.indeed.com/r/harini-komaravelli/2659eee82e435d1b?isid=rex-download&ikw=download-top&co=in',
'NOUN'),
('https://www.indeed.com/r/harini-komaravelli/2659eee82e435d1b?isid=rex-download&ikw=download-top&co=in',
'NOUN'),
('experienced', 'VERB'),
('development', 'NOUN'),
('execution', 'NOUN'),
('test', 'NOUN'),
('cases', 'NOUN'),
('effectively', 'ADV'),
(' ', 'SPACE'),
('experienced', 'VERB'),
('functional', 'ADJ'),
('testing', 'NOUN'),
('gui', 'NOUN'),
('testing', 'NOUN'),
('smoke', 'NOUN'),
('testing', 'NOUN'),
('regression', 'NOUN'),
('testing', 'NOUN'),
('integration', 'NOUN'),
('testing', 'NOUN'),
('experienced', 'VERB'),
('accessibility', 'NOUN'),
('testing', 'NOUN'),
('application', 'NOUN'),
('ability', 'NOUN'),
('understand', 'VERB'),
('user', 'NOUN'),
('requirements', 'NOUN'),
('functional', 'ADJ'),
('design', 'NOUN'),
('specifications', 'NOUN'),
(' ', 'SPACE'),

('good', 'ADJ'),
('knowledge', 'NOUN'),
('sdlc', 'NOUN'),
('stlc', 'NOUN'),
('processes', 'NOUN'),
(' ', 'SPACE'),
('deciding', 'VERB'),
('severity', 'NOUN'),
('priority', 'NOUN'),
('bugs', 'NOUN'),
(' ', 'SPACE'),
('experience', 'NOUN'),
('using', 'VERB'),
('microsoft', 'PROPN'),
('test', 'NOUN'),
('manager', 'NOUN'),
('&', 'CCONJ'),
('oracle', 'PROPN'),
('test', 'NOUN'),
('manager', 'NOUN'),
('test', 'PROPN'),
('management', 'PROPN'),
('tools', 'VERB'),
(' ', 'SPACE'),
('good', 'ADJ'),
('experience', 'NOUN'),
('testing', 'VERB'),
('windows', 'NOUN'),
('based', 'VERB'),
('&', 'CCONJ'),
('web', 'NOUN'),
('based', 'VERB'),
('applications', 'NOUN'),
(' ', 'SPACE'),
('involved', 'VERB'),
('client', 'NOUN'),
('interactions', 'NOUN'),
('reviews', 'VERB'),
('issues', 'NOUN'),
('clarifications', 'NOUN'),
(' ', 'SPACE'),
('web', 'NOUN'),
('services', 'NOUN'),
('testing', 'NOUN'),
('writing', 'VERB'),
('test', 'NOUN'),
('scripts', 'NOUN'),
('qtp', 'VERB'),
('testcomplete', 'ADV'),
(' ', 'SPACE'),

('creating', 'VERB'),
('object', 'ADJ'),
('repositories', 'NOUN'),
('function', 'NOUN'),
('libraries', 'NOUN'),
('qtp', 'VERB'),
(' ', 'SPACE'),
('enhanced', 'VERB'),
('qtp', 'NOUN'),
('scripts', 'NOUN'),
('using', 'VERB'),
('vb', 'PROPN'),
('script', 'NOUN'),
(' ', 'SPACE'),
('strong', 'ADJ'),
('experience', 'NOUN'),
('working', 'VERB'),
('blue', 'ADJ'),
('prism', 'NOUN'),
('tool', 'NOUN'),
('worked', 'VERB'),
('different', 'ADJ'),
('environments', 'NOUN'),
('like', 'ADP'),
('windows', 'PROPN'),
('application', 'NOUN'),
('&', 'CCONJ'),
('web', 'NOUN'),
('application', 'NOUN'),
('technical', 'ADJ'),
('skills', 'NOUN'),
(':', 'PUNCT'),
('test', 'NOUN'),
('automation', 'NOUN'),
('tools', 'NOUN'),
(':', 'PUNCT'),
('blue', 'ADJ'),
('prism', 'NOUN'),
('qtp', 'VERB'),
('10.0', 'NUM'),
('testcomplete', 'ADJ'),
('test', 'NOUN'),
('management', 'NOUN'),
('tool', 'NOUN'),
(':', 'PUNCT'),
('microsoft', 'PROPN'),
('test', 'NOUN'),
('manager', 'NOUN'),
('oracle', 'PROPN'),
('test', 'PROPN'),

('manager', 'PROPN'),
('&', 'CCONJ'),
('jira', 'PROPN'),
('databases', 'VERB'),
(':', 'PUNCT'),
('oracle', 'NOUN'),
('10', 'NUM'),
('g', 'NOUN'),
('sql', 'ADJ'),
('server', 'NOUN'),
(' ', 'SPACE'),
('operating', 'NOUN'),
('systems', 'NOUN'),
(':', 'PUNCT'),
('windows', 'VERB'),
('seven', 'NUM'),
('project', 'NOUN'),
('1', 'NUM'),
(':', 'PUNCT'),
('title', 'NOUN'),
(':', 'PUNCT'),
('cadence', 'NOUN'),
('client', 'NOUN'),
(':', 'PUNCT'),
('baker', 'NOUN'),
('hughes', 'VERB'),
('technologies', 'NOUN'),
(':', 'PUNCT'),
('microsoft', 'PROPN'),
('visual', 'PROPN'),
('studio', 'NOUN'),
('microsoft', 'PROPN'),
('team', 'PROPN'),
('foundation', 'PROPN'),
('server', 'NOUN'),
('client', 'NOUN'),
('background', 'NOUN'),
(':', 'PUNCT'),
('oilfield', 'NOUN'),
('services', 'NOUN'),
('company', 'NOUN'),
('delivering', 'VERB'),
('focused', 'VERB'),
('efforts', 'NOUN'),
('shale', 'NOUN'),
('gas', 'NOUN'),
('oilfield', 'NOUN'),
('services', 'NOUN'),
('provides', 'VERB'),
('services', 'NOUN'),

('tools', 'NOUN'),
('software', 'NOUN'),
('drilling', 'NOUN'),
('formation', 'NOUN'),
('evaluation', 'NOUN'),
('well', 'NOUN'),
('completion', 'NOUN'),
('production', 'NOUN'),
('management', 'NOUN'),
('seismic', 'ADJ'),
('data', 'NOUN'),
('collection', 'NOUN'),
('interpretation', 'NOUN'),
(' ', 'SPACE'),
('project', 'NOUN'),
('description', 'NOUN'),
(':', 'PUNCT'),
('aut', 'NOUN'),
('application', 'NOUN'),
('test', 'NOUN'),
('next', 'ADJ'),
('generation', 'NOUN'),
('revolutionary', 'ADJ'),
('robust', 'ADJ'),
('easy', 'ADJ'),
('use', 'NOUN'),
('scalable', 'ADJ'),
('well', 'ADJ'),
('site', 'NOUN'),
('data', 'NOUN'),
('acquisition', 'NOUN'),
('processing', 'VERB'),
('interpretation', 'NOUN'),
('system', 'NOUN'),
('client', 'NOUN'),
('s', 'PART'),
('drilling', 'NOUN'),
('services', 'NOUN'),
('deliver', 'VERB'),
('services', 'NOUN'),
('meets', 'VERB'),
('cross', 'VERB'),
('divisional', 'ADJ'),
('business', 'NOUN'),
('requirements', 'NOUN'),
('consistently', 'ADV'),
(' ', 'SPACE'),
('project', 'NOUN'),
('2', 'NUM'),
(':', 'PUNCT'),

('description', 'NOUN'),
(':', 'PUNCT'),
('paragon', 'PROPN'),
('supports', 'VERB'),
('entire', 'ADJ'),
('care', 'NOUN'),
('team', 'NOUN'),
('one', 'NUM'),
('tool', 'NOUN'),
('clinicians', 'NOUN'),
('need', 'VERB'),
('help', 'NOUN'),
('deliver', 'VERB'),
('best', 'ADJ'),
('patient', 'NOUN'),
('care', 'NOUN'),
('designed', 'VERB'),
('physicians', 'NOUN'),
('nurses', 'VERB'),
('pharmacists', 'NOUN'),
('mid', 'ADJ'),
('level', 'NOUN'),
('providers', 'NOUN'),
('first', 'ADJ'),
('-', 'PUNCT'),
('hand', 'NOUN'),
('understanding', 'NOUN'),
('clinical', 'ADJ'),
('workflow', 'NOUN'),
('needs', 'VERB'),
('paragon', 'PROPN'),
('clinical', 'ADJ'),
('applications', 'NOUN'),
('allow', 'VERB'),
('caregivers', 'NOUN'),
('focus', 'VERB'),
('matters', 'NOUN'),
('most', 'ADV'),
(';', 'PUNCT'),
('spending', 'VERB'),
('time', 'NOUN'),
('caring', 'VERB'),
('patients', 'NOUN'),
('since', 'SCONJ'),
('paragon', 'ADJ'),
('fully', 'ADV'),
('-', 'PUNCT'),
('integrated', 'VERB'),
('across', 'ADP'),
('applications', 'NOUN'),

('built', 'VERB'),
('around', 'ADP'),
('single', 'ADJ'),
('patient', 'NOUN'),
('database', 'NOUN'),
('information', 'NOUN'),
('entered', 'VERB'),
('anywhere', 'ADJ'),
('system', 'NOUN'),
('immediately', 'ADV'),
('available', 'ADJ'),
('entire', 'ADJ'),
('care', 'NOUN'),
('team', 'NOUN'),
('immediate', 'ADJ'),
('access', 'NOUN'),
('helps', 'VERB'),
('clinicians', 'NOUN'),
('make', 'VERB'),
('better', 'ADJ'),
('treatment', 'NOUN'),
('decisions', 'NOUN'),
('also', 'ADV'),
('helps', 'VERB'),
('promote', 'VERB'),
('patient', 'ADJ'),
('safety', 'NOUN'),
('paragon', 'NOUN'),
('offers', 'VERB'),
('broad', 'ADJ'),
('suite', 'NOUN'),
('multidisciplinary', 'ADJ'),
('clinical', 'ADJ'),
('software', 'NOUN'),
('solutions', 'NOUN'),
('together', 'ADV'),
('anytime', 'ADV'),
('anywhere', 'ADV'),
('access', 'NOUN'),
('complete', 'ADJ'),
('patient', 'NOUN'),
('record', 'NOUN'),
(' ', 'SPACE'),
('responsibilities', 'NOUN'),
(':', 'PUNCT'),
('performed', 'VERB'),
('smoke', 'NOUN'),
('testing', 'NOUN'),
('regression', 'NOUN'),
('testing', 'NOUN'),

(' ', 'SPACE'),
('involved', 'VERB'),
('generating', 'VERB'),
('executing', 'VERB'),
('test', 'NOUN'),
('script', 'NOUN'),
('using', 'VERB'),
('quick', 'ADJ'),
('test', 'NOUN'),
('pro', 'ADJ'),
('&', 'CCONJ'),
('blue', 'ADJ'),
('prism', 'PROPN'),
('usability', 'NOUN'),
('user', 'NOUN'),
('interface', 'NOUN'),
('testing', 'NOUN'),
(' ', 'SPACE'),
('involved', 'VERB'),
('defect', 'NOUN'),
('tracking', 'NOUN'),
('reporting', 'VERB'),
('bugs', 'NOUN'),
('using', 'VERB'),
('tfs', 'PROPN'),
('participated', 'VERB'),
('frequent', 'ADJ'),
('walk', 'NOUN'),
('-', 'PUNCT'),
('through', 'ADP'),
('meetings', 'NOUN'),
('internal', 'ADJ'),
('quality', 'NOUN'),
('assurance', 'NOUN'),
('groups', 'NOUN'),
('development', 'NOUN'),
('groups', 'NOUN'),
(' ', 'SPACE'),
('participated', 'VERB'),
('client', 'NOUN'),
('calls', 'VERB'),
('clarifying', 'VERB'),
('doubts', 'PROPN'),
('at&t', 'PROPN'),
('sessions', 'NOUN'),
('involved', 'VERB'),
('functional', 'ADJ'),
('regression', 'NOUN'),
('smoke', 'NOUN'),
('testing', 'VERB'),

('validate', 'NOUN'),
('application', 'NOUN'),
('data', 'NOUN'),
('changes', 'NOUN'),
('done', 'VERB'),
('windows', 'NOUN'),
('application', 'NOUN'),
('certifying', 'VERB'),
('build', 'NOUN'),
('status', 'NOUN'),
('running', 'VERB'),
('scripts', 'NOUN'),
('part', 'NOUN'),
('smoke', 'NOUN'),
('testing', 'NOUN'),
('project', 'NOUN'),
('3', 'NUM'),
(':', 'PUNCT'),
('description', 'NOUN'),
(':', 'PUNCT'),
('food', 'NOUN'),
('&', 'CCONJ'),
('beverages', 'NOUN'),
('r&a', 'PROPN'),
(':', 'PUNCT'),
('easily', 'ADV'),
('manage', 'VERB'),
('business', 'NOUN'),
('across', 'ADP'),
('multiple', 'ADJ'),
('locations', 'NOUN'),
('reducing', 'VERB'),
('cost', 'NOUN'),
('complexity', 'NOUN'),
('cloud', 'NOUN'),
('-', 'PUNCT'),
('based', 'VERB'),
('point', 'NOUN'),
('-', 'PUNCT'),
('of', 'ADP'),
('-', 'PUNCT'),
('sale', 'NOUN'),
('pos', 'NOUN'),
('solutions', 'NOUN'),
('enable', 'VERB'),
('centralized', 'VERB'),
('enterprise', 'NOUN'),
('management', 'NOUN'),
('lower', 'ADJ'),
('upfront', 'ADJ'),

```
('costs', 'NOUN'),
('smaller', 'ADJ'),
('footprint', 'NOUN'),
(' ', 'SPACE'),
('responsibilities', 'NOUN'),
(':', 'PUNCT'),
('performed', 'VERB'),
('functional', 'ADJ'),
('testing', 'NOUN'),
('regression', 'NOUN'),
('testing', 'NOUN'),
(' ', 'SPACE'),
('involved', 'VERB'),
('generating', 'VERB'),
('executing', 'VERB'),
('test', 'NOUN'),
('scripts', 'NOUN'),
('using', 'VERB'),
('blue', 'ADJ'),
('prism', 'NOUN'),
('tool', 'NOUN'),
('open', 'ADJ'),
('script', 'NOUN'),
('involved', 'VERB'),
('preparing', 'VERB'),
('bots', 'NOUN'),
('using', 'VERB'),
('blue', 'ADJ'),
('prism', 'NOUN'),
('tool', 'NOUN'),
(' ', 'SPACE'),
('accessibility', 'NOUN'),
('testing', 'NOUN'),
('web', 'NOUN'),
('application', 'NOUN'),
('involved', 'VERB'),
('defect', 'NOUN'),
('tracking', 'NOUN'),
('reporting', 'VERB'),
('bugs', 'NOUN'),
('using', 'VERB'),
('jira', 'PROPN'),
('webservices', 'NOUN'),
('testing', 'NOUN'),
('calling', 'VERB'),
('api', 'NOUN'),
("'s", 'PART'),
('export', 'NOUN'),
('data', 'NOUN')]
```

```

import pandas as pd
import numpy as np
import re
import string
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
import nltk
nltk.download('stopwords')
nltk.download('wordnet')
from nltk.corpus import stopwords
from nltk.stem import WordNetLemmatizer

[nltk_data] Downloading package stopwords to C:\Users\Nishchay
[nltk_data]   Vaid\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package wordnet to C:\Users\Nishchay
[nltk_data]   Vaid\AppData\Roaming\nltk_data...
[nltk_data]   Package wordnet is already up-to-date!

content_resumes

0      govardhana k senior software engineer bengalur...
1      harini komaravelli test analyst oracle hyderab...
2      hartej kathuria data analyst intern oracle ret...
3      ijas nizamuddin associate consultant state str...
4      imgeeyaul ansari java developer pune maharasht...
...
195     mansi thanki student jamnagar gujarat email in...
196     anil kumar microsoft azure basic management de...
197     siddharth choudhary microsoft office suite exp...
198     valarmathi dhandapani investment banking opera...
199     pradeep kumar security analyst infosys career ...
Name: content, Length: 200, dtype: object

len(content_resumes)

200

content_resumes.describe()

count      200
unique      199
top      sameer kujur orrisha email indeed: indeed.com/...
freq      2
Name: content, dtype: object

content_resumes.duplicated().sum()
#There is one duplicated row

1

```

```
content_resumes = content_resumes.drop_duplicates()
```

```
content_resumes.duplicated().sum()  
#Now we don't have any duplicate rows
```

```
0
```

```
final_df = content_resumes.to_frame().copy()
```

```
final_df
```

```
                                content  
0    govardhana k senior software engineer bengalur...  
1    harini komaravelli test analyst oracle hyderab...  
2    hartej kathuria data analyst intern oracle ret...  
3    ijas nizamuddin associate consultant state str...  
4    imgeeyaul ansari java developer pune maharash...  
..                                     ...  
195  mansi thanki student jamnagar gujarat email in...  
196  anil kumar microsoft azure basic management de...  
197  siddharth choudhary microsoft office suite exp...  
198  valarmathi dhandapani investment banking opera...  
199  pradeep kumar security analyst infosys career ...
```

```
[199 rows x 1 columns]
```

```
#FINDING AVERAGE WORD LENGTH, TOTAL WORD COUNT, TOTAL CHARACTERS IN  
THE CONTENT COLUMN
```

```
#Function calculates the average length of each word in content column  
#Apply the get_avg_word_len function to each element in the 'content'  
column of the DataFrame  
#Convert x to a string to handle cases where x might be NaN or non-  
string values  
#Assign the resulting average word lengths to a new column named  
'avg_word_len'
```

```
def get_avg_word_len(x):  
    words = x.split()  
    word_len = 0  
    for word in words:  
        word_len += len(word)  
    return word_len / len(words)
```

```
final_df['avg_word_len'] = final_df['content'].apply(lambda x:  
get_avg_word_len(str(x)))
```

```
#Apply the lambda function to each element in the 'content' column of  
the DataFrame  
#Convert x to a string to handle cases where x might be NaN or non-  
string values
```

```
#Split the string into individual words using the split() method
#Calculate the length of the resulting list of words using len()
#Assign the total word count to a new column named 'total_word_count'
```

```
final_df['total_word_count'] = final_df['content'].apply(lambda x:
len(str(x).split()))
```

```
#Apply the lambda function to each element in the 'content' column of
the DataFrame
```

```
#Convert x to a string to handle cases where x might be NaN or non-
string values
```

```
#Calculate the length of the resulting string using len()
```

```
#Assign the total character count to a new column named
'total_characters'
```

```
final_df['total_characters'] = final_df['content'].apply(lambda x:
len(str(x)))
```

```
final_df.head(5)
```

	content	avg_word_len \
0	govardhana k senior software engineer bengalur...	7.755760
1	harini komaravelli test analyst oracle hyderab...	7.383764
2	hartej kathuria data analyst intern oracle ret...	6.975439
3	ijas nizamuddin associate consultant state str...	7.262729
4	imgeeyaul ansari java developer pune maharash...	6.503788

	total_word_count	total_characters
0	217	1910
1	542	4570
2	285	2275
3	491	4068
4	264	1995

Exploratory Data Analysis

```
#EDA NISHCHAY VAID
```

Parts of Speech Analysis

```
nltk.download('punkt')
nltk.download('averaged_perceptron_tagger')
nltk.download('tagsets')
from textblob import TextBlob
```

```
[nltk_data] Downloading package punkt to C:\Users\Nishchay
[nltk_data] Vaid\AppData\Roaming\nltk_data...
[nltk_data] Package punkt is already up-to-date!
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data] C:\Users\Nishchay Vaid\AppData\Roaming\nltk_data...
[nltk_data] Package averaged_perceptron_tagger is already up-to-
```



```

[nltk_data]      date!
[nltk_data] Downloading package tagsets to C:\Users\Nishchay
[nltk_data]      Vaid\AppData\Roaming\nltk_data...
[nltk_data]   Package tagsets is already up-to-date!

nltk.help.upenn_tagset()

$: dollar
  $ -$ --$ A$ C$ HK$ M$ NZ$ S$ U.S.$ US$
': closing quotation mark
  ' '
(: opening parenthesis
  ( [ {
): closing parenthesis
  ) ] }
,: comma
  ,
--: dash
  --
.: sentence terminator
  . ! ?
:: colon or ellipsis
  : ; ...
CC: conjunction, coordinating
  & 'n and both but either et for less minus neither nor or plus so
  therefore times v. versus vs. whether yet
CD: numeral, cardinal
  mid-1890 nine-thirty forty-two one-tenth ten million 0.5 one
  forty-
  seven 1987 twenty '79 zero two 78-degrees eighty-four IX '60s .025
  fifteen 271,124 dozen quintillion DM2,000 ...
DT: determiner
  all an another any both del each either every half la many much
  nary
  neither no some such that the them these this those
EX: existential there
  there
FW: foreign word
  gemeinschaft hund ich jeux habeas Haementeria Herr K'ang-si vous
  lutihaw alai je jour objets salutaris fille quibusdam pas trop
Monte
  terram fiche oui corporis ...
IN: preposition or conjunction, subordinating
  astride among upon whether out inside pro despite on by
  throughout
  below within for towards near behind atop around if like until
  below
  next into if beside ...
JJ: adjective or numeral, ordinal
  third ill-mannered pre-war regrettable oiled calamitous first
  separable

```

ectoplasmic battery-powered participatory fourth still-to-be-named
 multilingual multi-disciplinary ...
 JJR: adjective, comparative
 bleaker braver breezier briefer brighter brisker broader bumper
 busier
 calmer cheaper choosier cleaner clearer closer colder commoner
 costlier
 cozier creamier crunchier cuter ...
 JJS: adjective, superlative
 calmest cheapest choicest classiest cleanest clearest closest
 commonest
 corniest costliest crassest creepiest crudest cutest darkest
 deadliest
 dearest deepest densest dinkiest ...
 LS: list item marker
 A A. B B. C C. D E F First G H I J K One SP-44001 SP-44002 SP-
 44005
 SP-44007 Second Third Three Two * a b c d first five four one six
 three
 two
 MD: modal auxiliary
 can cannot could couldn't dare may might must need ought shall
 should
 shouldn't will would
 NN: noun, common, singular or mass
 common-carrier cabbage knuckle-duster Casino afghan shed
 thermostat
 investment slide humour falloff slick wind hyena override
 subhumanity
 machinist ...
 NNP: noun, proper, singular
 Motown Venneboerger Czestochwa Ranzer Conchita Trumplane Christos
 Oceanside Escobar Kreisler Sawyer Cougar Yvette Ervin ODI Darryl
 CTCA
 Shannon A.K.C. Meltex Liverpool ...
 NNPS: noun, proper, plural
 Americans Americas Amharas Amityvilles Amusements Anarcho-
 Syndicalists
 Andalusians Andes Andruses Angels Animals Anthony Antilles
 Antiques
 Apache Apaches Apocrypha ...
 NNS: noun, common, plural
 undergraduates scotches bric-a-brac products bodyguards facets
 coasts
 divestitures storehouses designs clubs fragrances averages
 subjectivists apprehensions muses factory-jobs ...
 PDT: pre-determiner
 all both half many quite such sure this
 POS: genitive marker
 ' 's

PRP: pronoun, personal
 hers herself him himself hisself it itself me myself one oneself
 ours
 ourselves ownself self she thee theirs them themselves they thou
 thy us
 PRP\$: pronoun, possessive
 her his mine my our ours their thy your
 RB: adverb
 occasionally unabatingly maddeningly adventurously professedly
 stirringly prominently technologically magisterially predominately
 swiftly fiscally pitilessly ...
 RBR: adverb, comparative
 further gloomier grander graver greater grimmer harder harsher
 healthier heavier higher however larger later leaner lengthier
 less-
 perfectly lesser lonelier longer louder lower more ...
 RBS: adverb, superlative
 best biggest bluntest earliest farthest first furthest hardest
 heartiest highest largest least less most nearest second tightest
 worst
 RP: particle
 aboard about across along apart around aside at away back before
 behind
 by crop down ever fast for forth from go high i.e. in into just
 later
 low more off on open out over per pie raising start teeth that
 through
 under unto up up-pp upon whole with you
 SYM: symbol
 % & ' ' ' ' ' .)). * + , . < = > @ A[fj] U.S U.S.S.R * ** ***
 TO: "to" as preposition or infinitive marker
 to
 UH: interjection
 Goodbye Goody Gosh Wow Jeepers Jee-sus Hubba Hey Kee-reist Oops
 amen
 huh howdy uh dammit whammo shucks heck anyways whodunnit honey
 golly
 man baby diddle hush sonuvabitch ...
 VB: verb, base form
 ask assemble assess assign assume atone attention avoid bake
 balkanize
 bank begin behold believe bend benefit bevel beware bless boil
 bomb
 boost brace break bring broil brush build ...
 VBD: verb, past tense
 dipped pleaded swiped regummed soaked tidied convened halted
 registered
 cushioned exacted snubbed strode aimed adopted belied figgered
 speculated wore appreciated contemplated ...
 VBG: verb, present participle or gerund

telegraphing stirring focusing angering judging stalling lactating
 hankerin' alleging veering capping approaching traveling besieging
 encrypting interrupting erasing wincing ...
 VBN: verb, past participle
 multihulled dilapidated aerosolized chaired languished panelized
 used
 experimented flourished imitated reunified factored condensed
 sheared
 unsettled primed dubbed desired ...
 VBP: verb, present tense, not 3rd person singular
 predominate wrap resort sue twist spill cure lengthen brush
 terminate
 appear tend stray glisten obtain comprise detest tease attract
 emphasize mold postpone sever return wag ...
 VBZ: verb, present tense, 3rd person singular
 bases reconstructs marks mixes displeases seals carps weaves
 snatches
 slumps stretches authorizes smolders pictures emerges stockpiles
 seduces fizzes uses bolsters slaps speaks pleads ...
 WDT: WH-determiner
 that what whatever which whichever
 WP: WH-pronoun
 that what whatever whatsoever which who whom whosoever
 WP\$: WH-pronoun, possessive
 whose
 WRB: Wh-adverb
 how however whence whenever where whereby wherever wherein
 whereof why
 ``: opening quotation mark
 \`: opening quotation mark

```

blob = TextBlob(str(final_df['content']))
blob.tags

```

```

[('0', 'CD'),
 ('govardhana', 'NN'),
 ('k', 'NN'),
 ('senior', 'JJ'),
 ('software', 'NN'),
 ('engineer', 'NN'),
 ('bengaluru', 'NN'),
 ('1', 'CD'),
 ('harini', 'NN'),
 ('komaravelli', 'JJ'),
 ('test', 'NN'),
 ('analyst', 'NN'),
 ('oracle', 'NN'),
 ('hyderabad', 'NN'),
 ('2', 'CD'),
 ('hartej', 'NN'),
 ('kathuria', 'NNS'),

```

('data', 'NNS'),
('analyst', 'NN'),
('intern', 'JJ'),
('oracle', 'NN'),
('ret', 'NN'),
('3', 'CD'),
('ijas', 'JJ'),
('nizamuddin', 'JJ'),
('associate', 'NN'),
('consultant', 'NN'),
('state', 'NN'),
('str', 'NN'),
('4', 'CD'),
('imgeeyaul', 'NN'),
('ansari', 'NN'),
('java', 'NN'),
('developer', 'NN'),
('pune', 'NN'),
('maharashtr', 'NN'),
('195', 'CD'),
('mansi', 'NN'),
('thanki', 'NN'),
('student', 'NN'),
('jamnagar', 'NN'),
('gujarat', 'NN'),
('email', 'NN'),
('in', 'IN'),
('196', 'CD'),
('anil', 'JJ'),
('kumar', 'NN'),
('microsoft', 'JJ'),
('azure', 'NN'),
('basic', 'JJ'),
('management', 'NN'),
('de', 'NN'),
('197', 'CD'),
('siddharth', 'JJ'),
('choudhary', 'JJ'),
('microsoft', 'JJ'),
('office', 'NN'),
('suite', 'NN'),
('exp', 'NN'),
('198', 'CD'),
('valarmathi', 'NN'),
('dhandapani', 'NN'),
('investment', 'NN'),
('banking', 'NN'),
('opera', 'NN'),
('199', 'CD'),
('pradeep', 'JJ'),

```
( 'kumar', 'NN'),
( 'security', 'NN'),
( 'analyst', 'NN'),
( 'infosys', 'NN'),
( 'career', 'NN'),
( 'Name', 'NN'),
( 'content', 'NN'),
( 'Length', 'NNP'),
( '199', 'CD'),
( 'dtype', 'NN'),
( 'object', 'NN')]
```

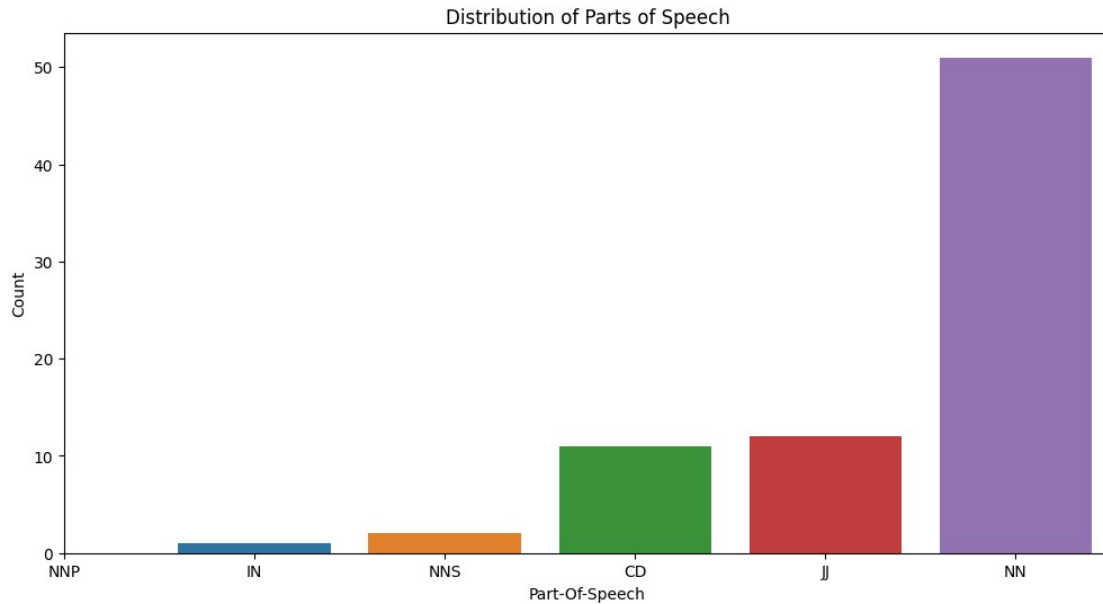
```
pos_df = pd.DataFrame(blob.tags, columns = ['words', 'pos'])
pos_df.head()
```

```
      words pos
0         0  CD
1  govardhana  NN
2          k   NN
3    senior   JJ
4  software   NN
```

```
pos_df = pos_df['pos'].value_counts()
pos_df
```

```
pos
NN      51
JJ      12
CD      11
NNS      2
IN       1
NNP      1
Name: count, dtype: int64
```

```
plt.figure(figsize = (12, 6))
sns.barplot(data = pos_df, y = pos_df.index, x = pos_df.values);
labels = list(pos_df.index)
labels.reverse()
plt.xticks(np.arange(6) - 1.0, labels)
plt.title('Distribution of Parts of Speech')
plt.ylabel('Count')
plt.xlabel('Part-Of-Speech')
plt.show();
```



CONCLUSION

- Nouns occur the most in the resume content

```
%matplotlib notebook
from nltk import FreqDist
from typing import Iterable
import matplotlib.pyplot as plt

# Function to flatten a nested list
def flatten(lis):
    for item in lis:
        if isinstance(item, Iterable) and not isinstance(item, str):
            for x in flatten(item):
                yield x
        else:
            yield item

# Get the tokenized resumes
tokenized_resumes = list(final_df['content'])

# Split each resume into a list of words
for i in range(len(tokenized_resumes)):
    tokenized_resumes[i] = str(tokenized_resumes[i]).split(" ")

# Flatten the list of lists into a single list
tokenized_resumes = list(flatten(tokenized_resumes))

# Compute the frequency distribution
freq_dist = FreqDist(tokenized_resumes)
```

```

# Select the 100 most common words
fd_common100 = freq_dist.most_common(100)

# Extract the words and frequencies
words = []
frequency = []

# Iterate over the range from 1 to 100
for i in range(1, 100):
    words.append(fd_common100[i][0])
    frequency.append(fd_common100[i][1])

# Create a figure with a size of 9x9 inches
plt.figure(figsize=(9, 9))

# Plot the bar chart
plt.bar(words, frequency, width=0.6)

# Rotate the x-axis labels by 90 degrees for better readability
plt.xticks(rotation=90, fontsize='small')

# Adjust the layout of the plot elements
plt.tight_layout()

# Show the plot
plt.show()

```

<IPython.core.display.Javascript object>

<IPython.core.display.HTML object>

CONCLUSION

Some useful insights which can be derived from the above bar graph are as follows :

- The term "experience" ranked as the fifth most commonly used word, highlighting the significant value placed on experience compared to education within the job market.
- Management experience is the fourth most sought after skill looked in the job being applied
- Microsoft and Oracle emerged as the most frequently cited companies on the applicants' resumes.

#EDA NILSU BOZAN

```

import matplotlib.pyplot as plt
import seaborn as sns

final_df.head(1)

```


	content	avg_word_len	\
0	govardhana k senior software engineer bengalur...	7.75576	

	total_word_count	total_characters
0	217	1910

```
pyplt.figure(figsize=(8, 6))
sns.histplot(final_df['avg_word_len'], bins=15);
pyplt.title('Distribution of Avg Word Length')
pyplt.show()
```

<IPython.core.display.Javascript object>

<IPython.core.display.HTML object>

CONCLUSION

- Bar plot shows that most of the resumes in our dataframe have average word length 7
- More than 70 rows(resumes) have average word length 7

```
pyplt.figure(figsize=(8, 6))
sns.histplot(final_df['total_word_count'], bins=15);
pyplt.title('Distribution of total word count')
pyplt.show()
```

<IPython.core.display.Javascript object>

<IPython.core.display.HTML object>

CONCLUSION

- More than 120 resumes have total word count less than 500.
- There are less resumes that have more than 500 words
- Number of resumes which contains more than 1000 word is very few.

```
pyplt.figure(figsize=(8, 6))
sns.histplot(final_df['total_characters'], bins=40);
pyplt.title('Distribution of total characters count')
pyplt.show()
```

<IPython.core.display.Javascript object>

<IPython.core.display.HTML object>

CONCLUSION

- Most of the resumes have total characters less than 5000.

#FINDING MOST COMMON STOPWORDS IN THE CONTENT COLUMN

```
from collections import Counter
from wordcloud import STOPWORDS
```

#The code analyzes the 'content' column in a DataFrame and extracts

```

individual words.
#It counts the frequency of each word and filters out common stop
words using a predefined set.
#The most frequent stop words are then printed.
words = []

for val in final_df['content']:
    words.extend(str(val).split())

#Create a Counter object to count the frequency of each word
word_counts = Counter(words)
stopword_counts = {word: count for word, count in word_counts.items()
if word in STOPWORDS}

sorted_stopwords = sorted(stopword_counts.items(), key=lambda x: x[1],
reverse=True)

#LIST
common_stopwords = [word for word in STOPWORDS if word in
word_counts.keys()]
common_stopwords.append('https')

commonnums=['one', 'two', 'three', 'four', 'five', 'six', 'seven',
'eight', 'nine', 'ten','thousand']
for i in commonnums:
    common_stopwords.append(i)
#Print the most common stop words and their frequencies
for word, count in sorted_stopwords[:25]:
    print(f'{word}: {count}')

and: 1619
like: 161
also: 79
get: 25
r: 17
since: 16
http: 11
would: 10
i'm: 6
k: 4
hence: 3
else: 2
that's: 2
could: 2
i've: 2
otherwise: 2
ever: 1
however: 1
what's: 1

```

```
why's: 1
shall: 1
```

CONCLUSION

- Most common stopwords used in the resumes are 'and', 'like', 'also'

#CREATING WORD CLOUD(FINDING MOST FREQUENT WORDS IN RESUMES)

#The code generates a word cloud visualization based on the 'content' column of a DataFrame.

#It removes common stop words and converts the text to lowercase before creating the word cloud.

#The word cloud represents the most frequent words in the text, with larger words indicating higher frequency.

#The resulting word cloud is displayed.

```
import matplotlib.pyplot as plt
from wordcloud import WordCloud, STOPWORDS
```

```
cleaned_text = ""
stopwords = set(STOPWORDS)
stopwords.update(common_stopwords)
```

```
for value in final_df.content:
    value = str(value)
    tokens = value.split()
    tokens = [token.lower() for token in tokens if token.lower() not
in stopwords]
    cleaned_text += " ".join(tokens) + " "
```

```
max_words = 50
wordcloud = WordCloud(width=800, height=800, background_color='white',
stopwords=stopwords, min_font_size=10).generate(cleaned_text)
```

```
plt.figure(figsize=(8, 8), facecolor=None)
plt.imshow(wordcloud)
plt.axis("off")
plt.tight_layout(pad=0)
```

```
plt.show()
```

<IPython.core.display.Javascript object>

<IPython.core.display.HTML object>

CONCLUSION

- Word cloud emphasize that most common words in the resumes excluding stopwords are 'application', 'team', 'management', 'year', 'project', 'client', 'service'

#N-GRAM ANALYSIS

```
import nltk
from nltk import ngrams
from collections import Counter
```

#The code utilizes NLTK to perform n-gram analysis on the 'content' column of a DataFrame.

#It converts the content into a list of strings, joins them together, and tokenizes the resulting text into individual words.

#It then calculates the counts of unigrams, bigrams, and trigrams using the ngrams function and stores the counts in separate Counter objects.

#This allows for the analysis of the frequency of n-grams in the text data.

```
import nltk
from nltk import ngrams
from collections import Counter
```

```
content_series = final_df['content']
```

```
content_list = content_series.astype(str).tolist()
```

```
all_content = ' '.join(content_list)
```

```
words = nltk.word_tokenize(all_content)
```

```
unigrams = list(ngrams(words, 1))
unigram_counts = Counter(unigrams)
```

```
bigrams = list(ngrams(words, 2))
bigram_counts = Counter(bigrams)
```

```
trigrams = list(ngrams(words, 3))
trigram_counts = Counter(trigrams)
```

```
print("Most common unigrams:")
for unigram, count in unigram_counts.most_common(60):
    print(unigram, count)
```

Most common unigrams:

```
(':',) 2135
('two',) 1741
('and',) 1621
('thousand',) 1620
('&',) 1089
('management',) 621
('experience',) 520
('project',) 492
('team',) 484
('skills',) 435
('testing',) 419
('work',) 412
('one',) 412
('microsoft',) 407
('data',) 380
('test',) 369
('using',) 365
('years',) 356
('client',) 324
('https',) 322
('support',) 313
('application',) 304
('oracle',) 299
('year',) 294
('*',) 294
('technical',) 282
('business',) 282
('system',) 280
('?',) 277
('isid=rex-download',) 276
('ikw=download-top',) 276
('co=in',) 276
('software',) 266
('sap',) 253
(';',) 252
('development',) 248
('information',) 234
('process',) 229
('engineer',) 228
('infosys',) 228
('knowledge',) 227
('server',) 227
```

```

('customer',) 227
('email',) 225
('worked',) 223
('education',) 220
('new',) 214
('tools',) 212
('sql',) 203
('windows',) 202
('indeed',) 200
('systems',) 199
('karnataka',) 196
('less',) 194
('issues',) 194
('sixteen',) 191
('working',) 189
('services',) 182
('service',) 182
('applications',) 181

print("Most common bigrams:")
for bigram, count in bigram_counts.most_common(60):
    print(bigram, count)

```

```

Most common bigrams:
('two', 'thousand') 1557
('thousand', 'and') 1536
('https', ':') 316
('?', 'isid=rex-download') 276
('isid=rex-download', '&') 276
('&', 'ikw=download-top') 276
('ikw=download-top', '&') 276
('&', 'co=in') 276
('one', 'year') 262
('work', 'experience') 211
('email', 'indeed') 199
('indeed', ':') 199
('less', 'one') 192
('and', 'sixteen') 184
('&', 'amp') 177
('amp', ';') 177
('and', 'fifteen') 158
('and', 'thirteen') 145
('and', 'seventeen') 144
('additional', 'information') 140
('and', 'fourteen') 134
('and', 'twelve') 129
('june', 'two') 120
('and', 'ten') 112
('responsibilities', ':') 108
('bengaluru', 'karnataka') 107
('and', 'eleven') 104

```

```

('and', 'eight') 98
('july', 'two') 87
('willing', 'relocate') 86
('infosys', 'limited') 86
('tamil', 'nadu') 83
('hundred', 'and') 83
('may', 'two') 82
('two', 'years') 82
('co=in', 'https') 80
('december', 'two') 77
('september', 'two') 72
('january', 'two') 71
('october', 'two') 69
('and', 'seven') 68
('test', 'cases') 64
('client', ':') 64
('august', 'two') 64
('march', 'two') 62
('thousand', ',') 62
('february', 'two') 61
('and', 'five') 60
('engineer', 'infosys') 60
('sql', 'server') 59
('tools', ':') 57
('hyderabad', 'telangana') 57
('pune', 'maharashtra') 56
('role', ':') 55
('karnataka', 'email') 53
('november', 'two') 53
('april', 'two') 53
('server', 'two') 53
('c', '#') 53
('description', ':') 51

print("Most common trigrams:")
for trigram, count in trigram_counts.most_common(60):
    print(trigram, count)

```

Most common trigrams:

```

('two', 'thousand', 'and') 1532
('?', 'isid=rex-download', '&') 276
('isid=rex-download', '&', 'ikw=download-top') 276
('&', 'ikw=download-top', '&') 276
('ikw=download-top', '&', 'co=in') 276
('email', 'indeed', ':') 199
('less', 'one', 'year') 192
('thousand', 'and', 'sixteen') 184
('&', 'amp', ';') 177
('thousand', 'and', 'fifteen') 157
('thousand', 'and', 'thirteen') 145
('thousand', 'and', 'seventeen') 144

```

('thousand', 'and', 'fourteen') 134
('thousand', 'and', 'twelve') 128
('june', 'two', 'thousand') 120
('thousand', 'and', 'ten') 108
('thousand', 'and', 'eleven') 103
('thousand', 'and', 'eight') 98
('july', 'two', 'thousand') 87
('may', 'two', 'thousand') 82
('&', 'co=in', 'https') 80
('co=in', 'https', ':') 80
('december', 'two', 'thousand') 77
('september', 'two', 'thousand') 72
('january', 'two', 'thousand') 71
('october', 'two', 'thousand') 69
('thousand', 'and', 'seven') 68
('august', 'two', 'thousand') 64
('march', 'two', 'thousand') 62
('february', 'two', 'thousand') 61
('thousand', 'and', 'five') 59
('karnataka', 'email', 'indeed') 53
('november', 'two', 'thousand') 53
('april', 'two', 'thousand') 53
('server', 'two', 'thousand') 53
('willing', 'relocate', 'to') 48
('relocate', 'to', ':') 48
('thousand', 'and', 'nine') 47
('thousand', 'and', 'eighteen') 46
('one', 'year', 'additional') 46
('year', 'additional', 'information') 46
('engineer', 'infosys', 'limited') 41
('additional', 'information', 'technical') 40
('thousand', ',', 'nine') 40
(',', 'nine', 'hundred') 40
('one', 'thousand', ',') 39
('nine', 'hundred', 'and') 39
('thousand', 'and', 'four') 38
('willing', 'relocate', ':') 37
('relocate', ':', 'anywhere') 37
('and', 'sixteen', 'present') 37
('chennai', 'tamil', 'nadu') 37
(':', 'anywhere', 'work') 36
('anywhere', 'work', 'experience') 36
('and', 'seventeen', 'present') 36
('bengaluru', 'karnataka', 'email') 34
('and', 'fifteen', 'present') 31
('information', 'technical', 'skills') 28
('maharashtra', 'email', 'indeed') 27
('thousand', 'and', 'six') 27


```
labels = [trigram[0] + ' ' + trigram[1] + ' ' + trigram[2] for
trigram, count in trigram_counts.most_common(60)]
counts = [count for trigram, count in trigram_counts.most_common(60)]
```

```
pyplt.figure(figsize=(12, 6))
pyplt.bar(labels, counts)
pyplt.xticks(rotation=90)
pyplt.xlabel('Trigram')
pyplt.ylabel('Count')
pyplt.title('Most Common Trigrams')
pyplt.tight_layout()
pyplt.show()
```

<IPython.core.display.Javascript object>

<IPython.core.display.HTML object>

```
labels = [bigram[0] + ' ' + bigram[1] for bigram, count in
bigram_counts.most_common(60)]
counts = [count for bigram, count in bigram_counts.most_common(60)]
```

```
pyplt.figure(figsize=(12, 6))
pyplt.bar(labels, counts)
pyplt.xticks(rotation=90)
pyplt.xlabel('Bigram')
pyplt.ylabel('Count')
pyplt.title('Most Common Bigrams')
pyplt.tight_layout()
pyplt.show()
```

<IPython.core.display.Javascript object>

<IPython.core.display.HTML object>

```
most_common_unigrams = unigram_counts.most_common(60)
labels = [' '.join(unigram) for unigram, count in
most_common_unigrams]
counts = [count for unigram, count in most_common_unigrams]
```

```
pyplt.figure(figsize=(12, 6))
pyplt.bar(labels, counts)
pyplt.xticks(rotation=90)
pyplt.xlabel('Unigram')
pyplt.ylabel('Count')
pyplt.title('Most Common Unigrams')
pyplt.tight_layout()
pyplt.show()
```

<IPython.core.display.Javascript object>

<IPython.core.display.HTML object>

#EDA ANISH MITRA

df

```

                                content \
0    govardhana k senior software engineer bengalur...
1    harini komaravelli test analyst oracle hyderab...
2    hartej kathuria data analyst intern oracle ret...
3    ijas nizamuddin associate consultant state str...
4    imgeeyaul ansari java developer pune maharasht...
..
195  mansi thanki student jamnagar gujarat email in...
196  anil kumar microsoft azure basic management de...
197  siddharth choudhary microsoft office suite exp...
198  valarmathi dhandapani investment banking opera...
199  pradeep kumar security analyst infosys career ...

                                annotation
0    [{'label': ['Companies worked at'], 'points': ...
1    [{'label': ['Companies worked at'], 'points': ...
2    [{'label': ['Skills'], 'points': [{'start': 22...
3    [{'label': ['Skills'], 'points': [{'start': 46...
4    [{'label': ['Skills'], 'points': [{'start': 18...
..
195  [{'label': ['College Name'], 'points': [{'star...
196  [{'label': ['Location'], 'points': [{'start': ...
197  [{'label': ['Skills'], 'points': [{'start': 78...
198  [{'label': ['Skills'], 'points': [{'start': 92...
199  [{'label': ['Skills'], 'points': [{'start': 58...
```

[200 rows x 2 columns]

```
my_dict = {}
key = []
value = []
for x in df['annotation']:
    for y in x:
        if y['label'] == ['Name']:
            for z in y['points']:
                key.append(z['text'])
        if y['label'] == ['Location']:
            for z in y['points']:
                value.append(z['text'])
    keytuple = tuple(key)
    valuetuple = tuple(value)
    for (a,b) in zip(keytuple,valuetuple):
```

```

        my_dict[a] = b
print(my_dict)

locationslist = list(my_dict.values())
print(locationslist)
len(locationslist)

locationsset = set(my_dict.values())
len(locationsset)
print(locationsset)
locationsseries = pd.Series(locationslist)
locationsdataframe = pd.DataFrame(locationsseries.value_counts())
locationsdataframe.head()
locationsdataframe.to_csv('locationscount.csv')

#Importing matplotlib package again
import matplotlib as plt
import geopandas as gpd
# Load the shapefile
shapefile_path = r"C:\Users\gbore\Downloads\india_state.geojson"
# Replace with the actual path to the shapefile
india_map = gpd.read_file(shapefile_path)
#Viewing dataframe, plot and type of the Indian states map
print(india_map)
india_map.plot()
type(india_map)

#Importing necessary packages again
import geopandas as gpd
import pandas as pd
import matplotlib.pyplot as plt
# Reading in data on how many workers are in each state
finalcountdf = pd.read_csv('readlocationscount.csv')
#Viewing the data
finalcountdf.head()
#Merging data
merged_data = finalcountdf.merge(india_map, on='NAME_1', how='outer')
# Replace NaN values with 0s in Count column to make sure all states display themselves
merged_data['Count'].fillna(0, inplace=True)
#Getting an accurate count of applicants in each state
raw_series = merged_data.groupby('NAME_1')['Count'].sum()
#Converting this series to a dataframe
raw_df = pd.DataFrame(raw_series)
#Looking at merged_data and its type
print(merged_data)
type(merged_data)

#Looking at raw_data and its type
print(raw_df)
type(raw_df)

```

```
#Merging data
final_geodf = raw_df.merge(india_map, on='NAME_1', how='outer')
print(final_geodf)
type(final_geodf)

#Importing necessary packages again
import geopandas as gpd
import pandas as pd
import matplotlib.pyplot as plt

#Converting to GeoDataFrame and plotting
gdf = gpd.GeoDataFrame(final_geodf)
gdf.plot(column='Count', cmap='YlOrRd', linewidth=0.8,
edgecolor='Black', vmin =0, vmax=70, legend=True)

# Add a title and display the plot
plt.title('Number of Values by State')
plt.show()
```