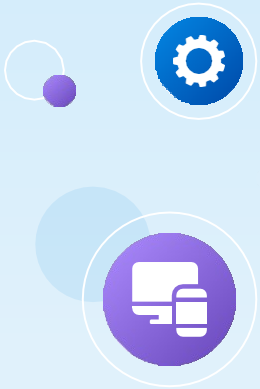
CoverQuick Industry Research

ALY 6080 - Integrated Experiential Learning

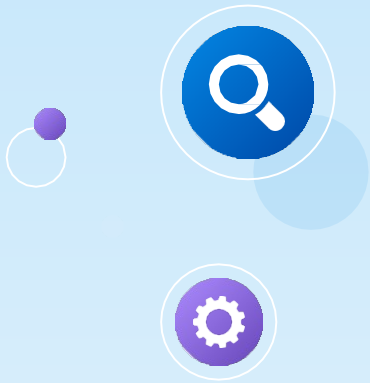
By:
Murtaza Vora

Table of contents



01

Executive Summary



About CoverQuick

- CoverQuick is an AI-powered job application platform that simplifies the process of creating customized resumes and cover letters for job seekers.
- The platform allows users to add up to five work experiences, three educational qualifications, and relevant projects or skills to their resumes.
- It also generates tailored cover letters that match the user's experience and skills to the job description, providing a starting point for the job application process.
- Website Link : www.coverquick.co

Technology Used

Front End
1. Nextjs
2. TailwindCSS

Back End
1. FastAPI
2. Postgres
3. Reddis

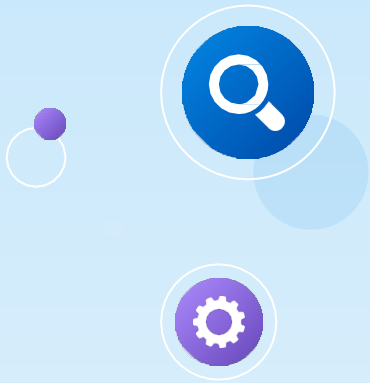
Company's Competition

Currently, we can't find any exact competition for the company but we have nearby competitors which are

1. Resume-Now
2. Zety
3. TopResume
4. Hiration

02

Business Problem

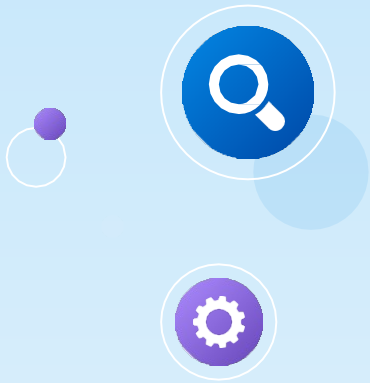


Business Problem

- Making cover letters can be time-consuming, complex, and daunting task for many job seekers
- Struggle to highlight their relevant experiences, skills, and education while aligning them with the specific job requirements.
- Increasing competitiveness of the job market demands a more strategic approach to job applications.
- Designing custom cover letters that can showcase the ideal candidate for the job.

03

Exploratory Data Analytics

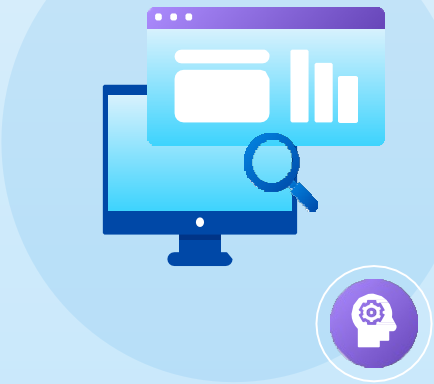


Strengths of Dataset

1. Large Sample Size

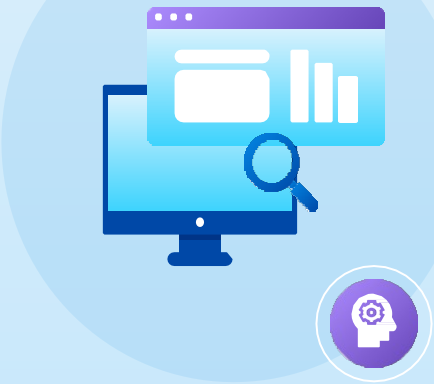
2. Diversity of Variables

3.Real Time High Quality Data



Weakness of Dataset

1. Sampling Bias
2. Missing Data
3. Data Errors



Data Preparation

- We found severe difficulty with missing data while preparing data for analysis, which hampered our ability to extract meaningful insights.
- To overcome this issue, we used a technique known as "Dependable to Relate Variable," which entails focusing on relevant columns and their related fields within the dataset.
- For example, while references are not typically included in a resume, education is an important factor.
- As a result, we treated education as a dependent variable during data preparation, while references were treated as an optional field.
- We used Python as the programming language to accomplish our goal, applying a variety of approaches such as data cleaning and sorting.
- We were able to patch gaps in the data and ensure that variables were leveled and assigned proper formats using these procedures.
- We were able to assure that our subsequent analysis would be trustworthy and useful by using this complete approach to data preparation. This resulted in valuable insights for the CoverQuick team.

Summary of the data

```
data.describe()
```

	id	content	jobDescription
count	39450	39450	13058
unique	39450	28600	12195
top	clf7yo3t70000yk2unt32re3h	{"awards": {"awards": []}, "header": {"role": ...	Who we are\nAbout Stripe\nStripe is a financia...
freq	1	8169	153

This summary provides information about the non-null values, uniqueness, most frequent values, and frequencies of the dataset's columns.

Dataframe View

```
In [51]: TestFrame
```

	city	state	country	KeyWords	awards	Education	Graduation Date	job Title	Previous organization	certifications
0	Nairobi	Nairobi	Kenya	[customer experience, empathy, expertise, “...	[No awards]	[Bachelor Of Commerce]	[June 2013]	[Account Executive, Busines Solutions Executiv...	[Amesi Kenya Limited, Nation Media Group, CMC ...	[Introduction to Data Analytics]
1	Estcourt	KwaZulu Natal	South Africa	[computer databases, keyboard, mouse, track...	[No awards]	[National Senior Certificate,]	[2016,]	[Marketing Agent, Delivery Driver, Educator As...	[Hlalanathi Guest Lodge, Qhakaza Cleaning Solu...	[OSHA-30, Procore Student Certification, Six S...

The given DataFrame, named "TestFrame," displays information about individuals including their city, state, country, keywords, awards, education, graduation date, job titles, previous organizations, and certifications. It consists of multiple rows representing different individuals and their respective details, providing insights into their skills, experiences, and qualifications.

More findings...

```
blank_row
len(blank_row)
```

```
23998
```

```
No_awards=TestFrame['awards']=="[No awards]"
No_awards.count()
```

```
343
```

```
uniqueCountries=TestFrame['country'].value_counts()
print(uniqueCountries)
```

```
not found              1915
United States           886
Kenya                   678
Nigeria                 323
Canada                  281
                       ...
Brunei Darussalam         1
Portugal                  1
Dominican Republic        1
Djibouti                  1
Spain                     1
Name: country, Length: 101, dtype: int64
```

Here, we find the number of blank rows in our dataset followed by any unique countries.

04

Flow of Approach and Analysis (Until Now)

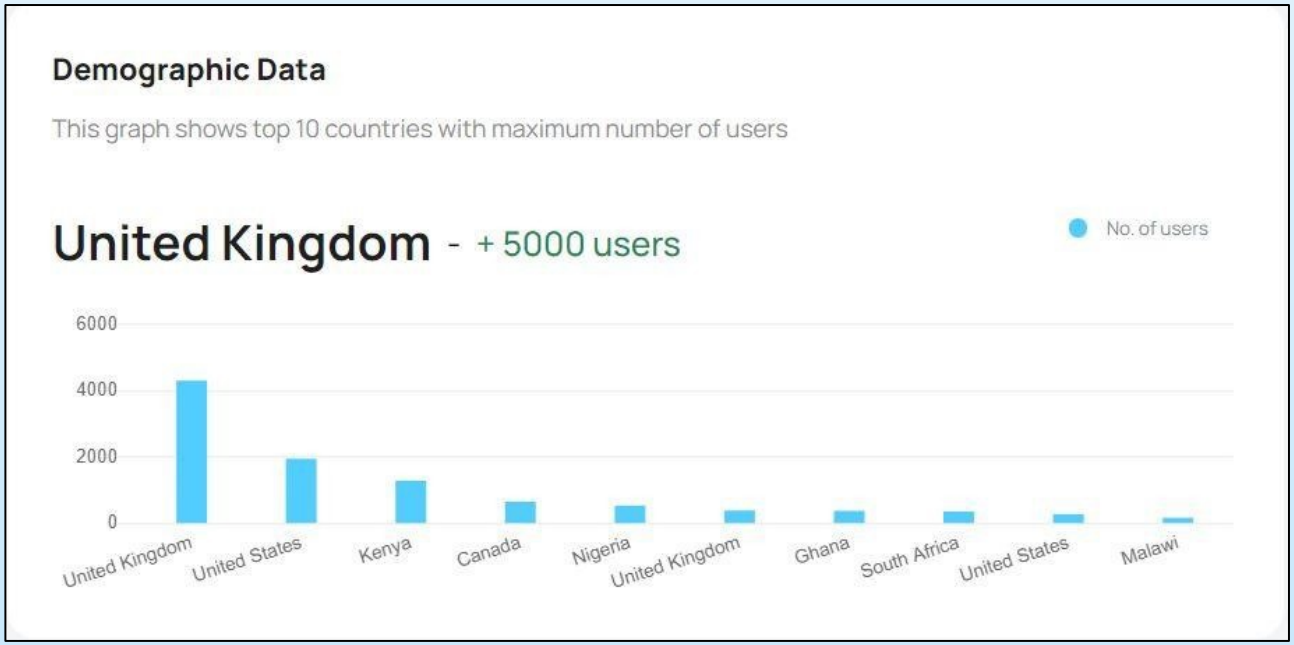
Data Import in Django

CITY	STATE	COUNTRY	KEYWORDS	AWARDS	EDUCATION	GRADUATION DATE	JOB TITLE
London		United Kingdom	['node.js', ' security disclosure', ' 3rd party security audit cycle', ' javascript', ' shell', ' python', ' linux', ' docker', ' networking', ' asynchronous communication', ' kindness']	['No awards']	['AS Science with Specialization in Math', 'AS Engineering', 'Full-stack Web Development']	['January 2022', '', 'November 2022']	['Frontend Developer', Accessibilit 'First Year A Assistant']
London		United Kingdom	['accounting', ' general ledger', ' accounts receivable/billing', ' accounts payable', ' fixed assets', ' period end/year end closing', ' financial reporting', ' internal control', ' accountants', ' general ledger entries', ' accruals', ' auditors', ' year-end audit', ' journal entries', ' subordinates', ' budgets', ' international accounting standards', ' standard costing system', ' revenue', ' expenses', ' budget forecasts', ' variance reports', ' source documents', ' vendors', ' contractors', ' customers.']	['No awards']	['Education Missing']	['Education Missing']	['No experi

Major Industries People Apply To

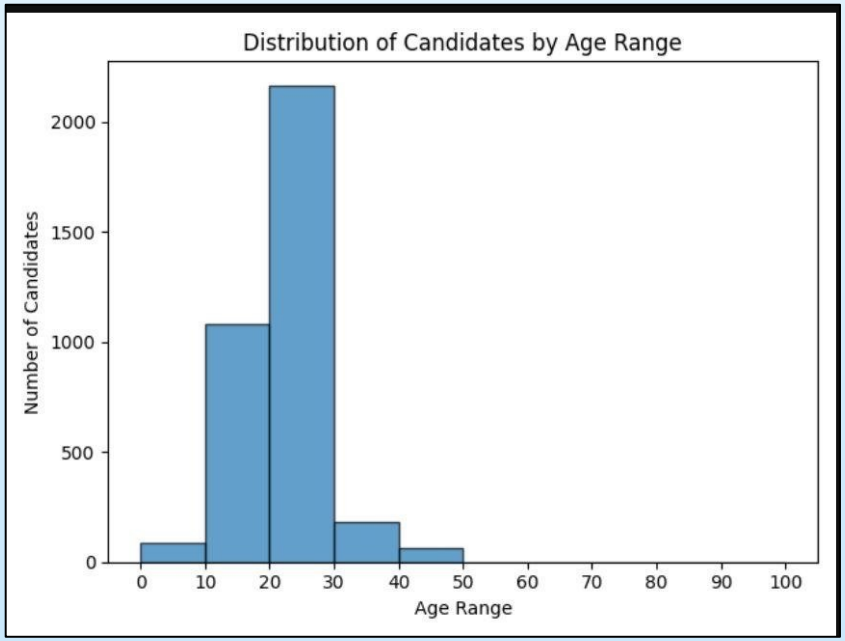
```
def display_graph_top_industries(request):
    # Retrieve all job titles
    job_titles = Candidate.objects.values_list('job_title', flat=True)

    # Extract keywords from job titles
    keywords = []
    for title in job_titles:
        if title:
            words = re.findall(r'\w+', title.lower())
            keywords.extend(words)

    # Calculate keyword counts
    keyword_counts = dict(Counter(keywords))

    # Calculate industry keyword counts
    industry_keyword_counts = {industry: sum(keyword_counts.get(keyword, 0) for keyword in keywords) for industry, keywords in industry_keywords.items()}

    # Sort the industries based on their keyword counts in descending order
    sorted_industries = sorted(industry_keyword_counts.items(), key=lambda x: x[1], reverse=True)

    # Take the top ten industries
    top_industries = sorted_industries[:10]

    industry_names = [industry for industry, count in top_industries]
    industry_counts = [count for industry, count in top_industries]

    # Set the figure size
    plt.figure(figsize=(10, 6))

    # Generate the graph
    plt.bar(industry_names, industry_counts)
    plt.xlabel('Industry')
    plt.ylabel('Keyword Count')
    plt.title('Top Ten Industries based on Job Titles')
    plt.xticks(rotation=30, fontsize=8, ha='right')

    # Save the graph to a buffer
    buffer = BytesIO()
    plt.savefig(buffer, format='png')
    buffer.seek(0)

    # Set the appropriate response headers
    response = HttpResponse(content_type='image/png')
    response['Content-Disposition'] = 'inline; filename=graph.png'

    # Send the buffer content as the HTTP response
    response.write(buffer.getvalue())
    return response
```

Trends in Demographics



Approximate Age Range of Candidates

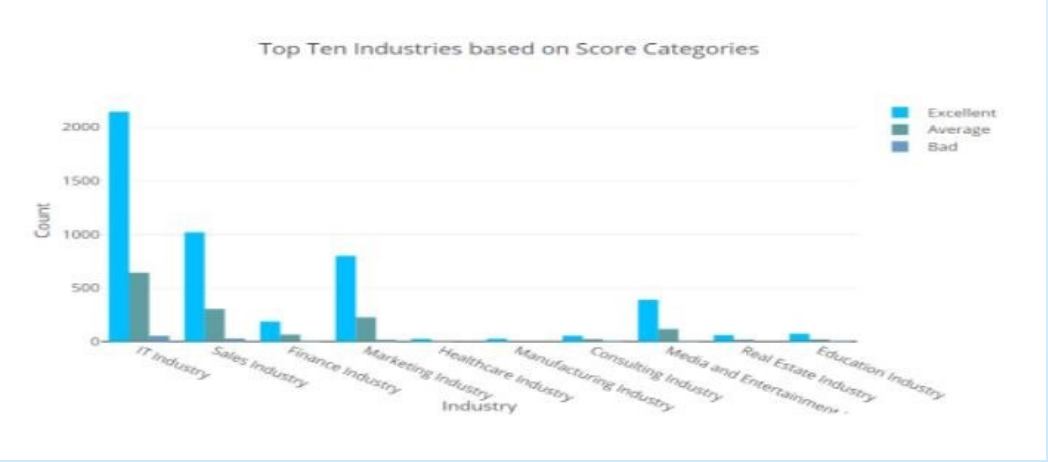
```
from datetime import datetime

from django.db import models

class Candidate(models.Model):
    city = models.CharField(max_length=255)
    state = models.CharField(max_length=255)
    country = models.CharField(max_length=255)
    keywords = models.TextField()
    awards = models.TextField()
    education = models.TextField()
    graduation_date = models.CharField(max_length=255)
    job_title = models.TextField()
    previous_organization = models.TextField()
    certifications = models.TextField()

    def calculate_age(self):
        if self.graduation_date:
            try:
                graduation_year = int(self.graduation_date.split()[-1])
                current_year = datetime.now().year
                age = current_year - graduation_year
                return age
            except ValueError:
                pass
        return None

    def calculate_experience(self):
        if self.job_title:
            job_titles = eval(self.job_title)
            return len(job_titles)
        return None
```



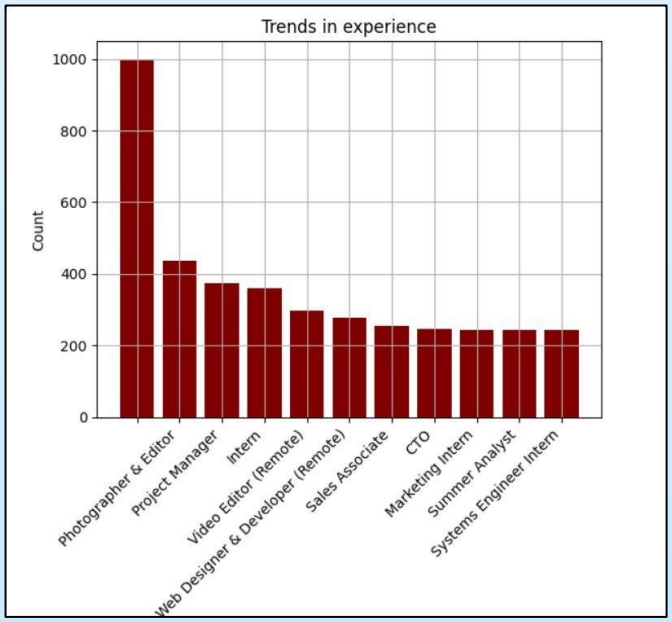
Industry Specific resume Categories(excellent ,average ,bad)

Industries based on Score Categories

Industry	Excellent Count	Average Count	Bad Count
IT Industry	2146	643	53
Sales Industry	1019	303	29
Finance Industry	186	64	3
Marketing Industry	799	226	12
Healthcare Industry	25	3	0
Manufacturing Industry	26	0	0
Consulting Industry	53	23	1
Media and Entertainment Industry	390	117	3
Real Estate Industry	60	18	0
Education Industry	72	20	2



Trends in Experience and Skill Count



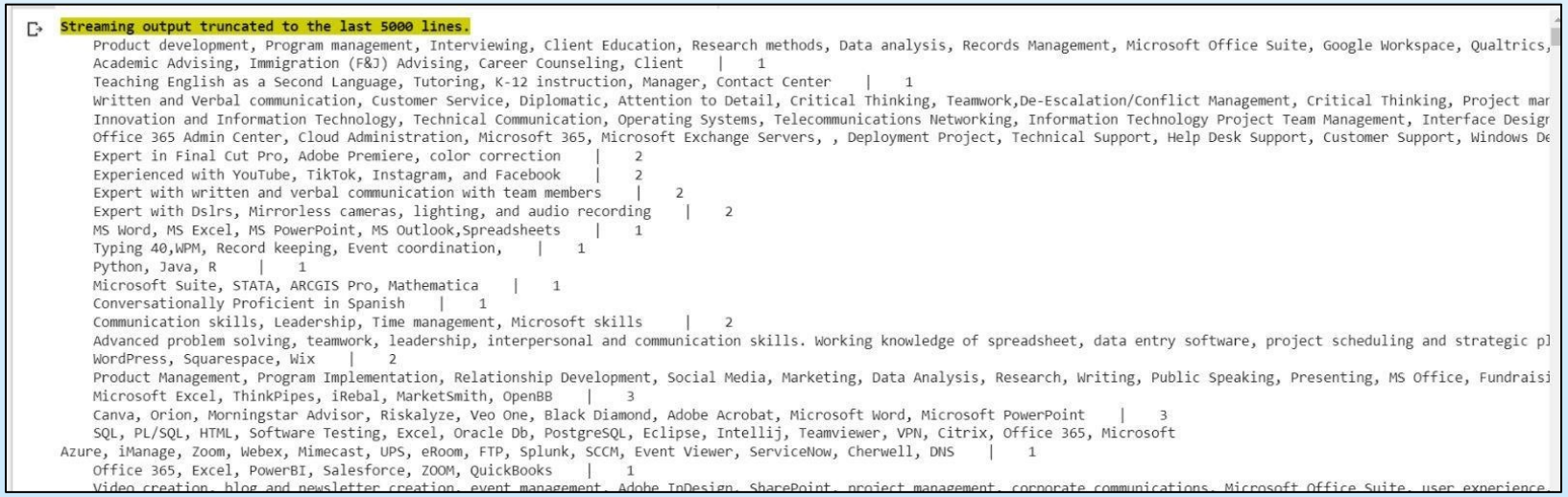
	Skil	Count
6		3825
452	Adobe Audition, Pro Tools, Logic Pro	251
450	Blender, 3ds Max	250
451	Adobe After Effects, Cinema 4D	250
454	Lightroom Classic, Lightroom CC, Bridge	249
456	Dropbox, Google Drive, OneDrive	247
453	Figma, Adobe XD	241
993	Pytorch, Tensorflow, Huggingface, Numpy, Pandas	240
457	Filezilla	240
992	React, React Native, Node	239
449	Final Cut, Davinci Resolve, Premiere Composer,...	238
462	Facebook Live, Twitch, OBS, Xsplit	234
448	Photoshop, Illustrator, InDesign, Dreamweaver,...	233
459	WordPress, Squarespace	230
455	Microsoft Office Suite, Word, Excel, PowerPoin...	227
458	Monday.com, Trello, Notion.so	227
991	Python, C++, C, Rust, Go, HTML/CSS, SASS	225
995	Kubernetes, Docker	222
994	Microsoft Office Suite, Adobe Photoshop & Illu...	220

Trends in Experience and Skill Count

```
#a=trends_in_experience.value_counts()
arr=trends_in_experience

title=[]
countArr=[]

fr = [None] * len(arr);
visited = -1;

for i in range(0, len(arr)):
    count = 1;
    for j in range(i+1, len(arr)):
        if(arr[i] == arr[j]):
            count = count + 1;
            #To avoid counting same element again
            fr[j] = visited;

    if(fr[i] != visited):
        fr[i] = count;

#Displays the frequency of each element present in array
#print("---------------------");
#print(" Element | Frequency");
#print("---------------------");
for i in range(0, len(fr)):
    if(fr[i] != visited):
        print("    " + str(arr[i]) + "    |    " + str(fr[i]));
        title.append(arr[i])
        countArr.append(fr[i])


TestFrameExpTrand=pd.DataFrame(list(zip(title,countArr)),columns=['Name','Count'])
```

```
Streaming output truncated to the last 5000 lines.
    Fraud Claim Analyst     |    7
    Junior Web/Mobile App Developer    |    1
    Frontend Developer    |    2
    Shaker, Cashier, Server     |    2
    Security Guard     |    11
    Quality Control Analyst    |    1
    Radio Presenter    |    1
    Computer Engineering Co-op    |    6
    Undergraduate Research Assisstant    |    6
    Math Domain Tutor    |    6
    Customer Service     |    2
    Student Social Media Contributor     |    2
    Fufilment Associate (Part-time)    |    1
    IT Business Analyst / User Support    |    1
    CONTRACTOR    |    1
    SUMMER ENFORCEMENT OFFICER     |    1
    SECURITY GUARD     |    1
    Senior Project Manager    |    23
    Test Lead    |    1
    Senior LQA Engineer    |    1
    LQA Engineer    |    1
    LQA Tester    |    1
    Fieldwork Supervisor    |    3
    Patient Service Agent    |    1
    Residental Crisis Counselor    |    1
    Scheduling and Pre Authorization Coordinator    |    1
    Medical Practice Representative    |    1
    Access Specialist I    |    1
    Eligibility Office Assistant    |    1
    Salesforce Consultant    |    1
    Managing Director & Salesforce Administrator    |    1
    Social Media Specialist     |    4
    Student Affairs Specialist          |    4
```

Trends in Experience and Skill Count



```
Streaming output truncated to the last 5000 lines.
    Product development, Program management, Interviewing, Client Education, Research methods, Data analysis, Records Management, Microsoft Office Suite, Google Workspace, Qualtrics,
    Academic Advising, Immigration (F&J) Advising, Career Counseling, Client    |    1
    Teaching English as a Second Language, Tutoring, K-12 instruction, Manager, Contact Center    |    1
    Written and Verbal communication, Customer Service, Diplomatic, Attention to Detail, Critical Thinking, Teamwork,De-Escalation/Conflict Management, Critical Thinking, Project mar
    Innovation and Information Technology, Technical Communication, Operating Systems, Telecommunications Networking, Information Technology Project Team Management, Interface Desigr
    Office 365 Admin Center, Cloud Administration, Microsoft 365, Microsoft Exchange Servers, , Deployment Project, Technical Support, Help Desk Support, Customer Support, Windows De
    Expert in Final Cut Pro, Adobe Premiere, color correction    |    2
    Experienced with YouTube, TikTok, Instagram, and Facebook    |    2
    Expert with written and verbal communication with team members    |    2
    Expert with Dslrs, Mirrorless cameras, lighting, and audio recording    |    2
    MS Word, MS Excel, MS PowerPoint, MS Outlook,Spreadsheets    |    1
    Typing 40,WPM, Record keeping, Event coordination,    |    1
    Python, Java, R    |    1
    Microsoft Suite, STATA, ARCGIS Pro, Mathematica    |    1
    Conversationally Proficient in Spanish    |    1
    Communication skills, Leadership, Time management, Microsoft skills    |    2
    Advanced problem solving, teamwork, leadership, interpersonal and communication skills. Working knowledge of spreadsheet, data entry software, project scheduling and strategic pl
    WordPress, Squarespace, Wix    |    2
    Product Management, Program Implementation, Relationship Development, Social Media, Marketing, Data Analysis, Research, Writing, Public Speaking, Presenting, MS Office, Fundraisi
    Microsoft Excel, ThinkPipes, iRebal, MarketSmith, OpenBB    |    3
    Canva, Orion, Morningstar Advisor, Riskalyze, Veo One, Black Diamond, Adobe Acrobat, Microsoft Word, Microsoft PowerPoint    |    3
    SQL, PL/SQL, HTML, Software Testing, Excel, Oracle Db, PostgreSQL, Eclipse, Intellij, Teamviewer, VPN, Citrix, Office 365, Microsoft
Azure, iManage, Zoom, Webex, Mimecast, UPS, eRoom, FTP, Splunk, SCCM, Event Viewer, ServiceNow, Cherwell, DNS    |    1
    Office 365, Excel, PowerBI, Salesforce, ZOOM, QuickBooks    |    1
    Video creation, blog and newsletter creation, event management, Adobe InDesign, SharePoint, project management, corporate communications, Microsoft Office Suite, user experience,
```

Criteria for Determining Good and Bad Resumes

- Length of Resume shouldn’t be more than 800 words (2 Pages).
- Action Verbs shouldn’t be used more often throughout the resume.
- Spelling mistakes should be minimum. (20 Spelling Mistakes are allowed).
- Pronouns shouldn’t be used in the resume in a frequent manner.
- All basic fields in the resume such as Education, Work Experience and more should always be there.

The following conditions are used to determine a resume is Excellent, Average or Bad.

Determining Best and Worst Resumes

Score category:

Here the scores are divided in 3 categories:

- Excellent- 8 & above
- Average - 5-7
- Bad - less than 4

Candidate Score Summary

#	User ID	Score	Resume Status
1	clf7f9tuw00tau12vb5ntt5cq	8.5	Excellent
2	clf72epej00gdyo2vylqxfvrz	8.5	Excellent
3	clf6y3unr00odxa2yi0284vgc	6.5	Average
4	clf7bw7ty003uxc2ujvpvk8th	8.5	Excellent
5	clf6talqd00bvu12vnk9l6nc4	8.5	Excellent

Thanks!