ALY 6080:

INTEGRATED EXPERIENTIAL LEARNING

Module 11: XN Project- CoverQuick Final Draft

Submitted To:
Dr. Chinthaka Pathum Dinesh, PhD,
Prof. Herath Gedara, Faculty Lecturer

Submitted By:
Abhilash Dikshit
Kush Patel
Siddharth Alashi

Academic Term: Spring 2023
Graduate Students at Northeastern University, Vancouver, BC, Canada
Master of Professional Studies in Analytics

July 03, 2023

CoverQuick

Title: CoverQuick Data Analysis in Business Performance

I. Executive Summary:

Data preparation is an essential step in building a database that can provide useful insights. In this report, we discuss the process of preparing data for a job applicant database. The data were obtained from a job search website and contained information about job seekers' education, work experience, and skills. We used Python programming language and Panda's library to clean, explore, and transform the data. The tasks completed include removing duplicate and null values, standardizing column names, converting data types, splitting and cleaning the dataset for answering the research questions. Further tasks include analyzing the data and creating visualizations to understand patterns and trends. We used Jupyter Notebook and Tableau for data analysis and visualization. Finally, we will communicate the results to stakeholders through reports and presentations.

II. Introduction

CoverQuick utilizes an AI-based program to generate tailored cover letters and resumes for job applications. The aim is to help job seekers differentiate themselves from the competition by employing the latest AI techniques and creating customized documents that align with the specific job requirements.

One notable feature of CoverQuick is its ability to generate personalized cover letters for each application, ensuring that candidates do not submit generic letters that fail to impress recruiters.

CoverQuick's product offerings include the preparation of resumes and cover letters, application tracking, and a resume grading feature to provide valuable feedback to applicants.

III. Keywords:

Data preparation, data cleaning, data transformation, data analysis, data

visualization, job applicant database.

IV. Business Problem

1. What are the three industries that the majority of CoverQuick's users have applied?

2. Discover trends in demographics and find which industries yield the best and the worst resumes (CoverQuick provides metrics for defining a "Good" resume).

3. Determine the expected age and approximate experience level.

4. Determine trends in experience and skills for these target users.

5. Determine Suggested Skills for respective Candidates as per demographics.

V. Planning and Execution:

1. EDA on job description dataset.
2. Dataset splitting for the respective columns which were in json and nested json format.
3. Identification and visualization for the top 3 industries that most users have applied.
4. Identification and visualization for the approximate age range and experience level.
5. Identification and visualization for the trends in experience and skills for these target users.
6. Identification and visualization to discover the trends in demographics for the number of candidates registering to the website across globe for resume building.

Dataset Provided:

1. With Job Description
2. No Job Description

VI. Exploratory Data Analysis:

We obtained the data from our Industry sponsor "CoverQuick". We used Python programming language and other libraries to clean, explore, and transform the data. The tasks completed include replacing/removing duplicate and null values, standardizing column names, and converting data types. We also created a new column that extracted country names into one field to make it easier to analyze the data by country code; Split the required columns with given Keys and Values to get respective data for in-depth analysis.

CoverQuick No Job Description Dataset:

Display Raw Dataset:

	id	content
0	clezbskz2000ryi2yj9sgidax	{"awards": {"awards": []}, "header": {"role": ...
1	clezeac8k001myg2u6yph9yz0	{"awards": {"awards": []}, "header": {"role": ...
2	clezh9pxu0185yp2yo6ztscpf	{"awards": {"awards": []}, "header": {"role": ...
3	clf06bf5b014dx22ur1bph3sd	{"awards": {"awards": []}, "header": {"role": ...
4	clezdavhy000eyg2uh55ycnb4	{"awards": {"awards": []}, "header": {"role": ...
...	...	...
13193	clbmqeb0p00bowy50puy0zh47	{"awards": {"awards": []}, "header": {"role": ...
13194	cld9mmcjj004mx32l3q1bn6ju	{"awards": {"awards": []}, "header": {"role": ...
13195	cld8csfqd002yyc2m1l7vqsuv	{"awards": {"awards": []}, "header": {"role": ...
13196	clelqs1nd0002x12vg9vjumyg	{"awards": {"awards": []}, "header": {"role": ...
13197	cldy3pez8000oxd2xvvj0fpyk	{"awards": {"awards": []}, "header": {"role": ...

13198 rows × 2 columns

Display Type, Length, Shape about the dataset:

Type	Length	Shape
<class 'pandas.core.frame.DataFrame'>	13198	(13198, 2)

Display datatypes of respective columns in dataset:

	id	content
0	object	object

Showing max, min length and NA values:

Column	Max Length	Min Length	NA Count
id	25	25	0
content	30743	484	0

CoverQuick With Job Description Dataset:

Display Raw Dataset:

	id	content	jobDescription
0	clg43d9an007gx02ug1i694j6	{"awards": {"awards": []}, "header": {"role": ...	Job Posting:\nDo you have a passion for helpin...
1	clg3itetj006jx92tdkcrw195	{"awards": {"awards": []}, "header": {"role": ...	Tasks:\n\nCreation of concepts for dashboard i...
2	clg3iy1sd007rx32utnuhnrgy	{"awards": {"awards": [{"name": "Dean's List",...	Responsibilities:\n\nWork closely with product...
3	clg5j15lz00k3x02uaau7g9z0	{"awards": {"awards": []}, "header": {"role": ...	What is Talentport :\n\nTalentport connects SE...
4	clg43pte600ddya2umakfw3c3	{"awards": {"awards": []}, "header": {"role": ...	Hyperproof is hiring a Product Manager with a ...
...	...	...	...
11971	cleexyzag006ayg2vhr087als	{"awards": {"awards": []}, "header": {"role": ...	Assist with content ideation and creation, inc...
11972	cleec90b0005nyf2tlos9qc95	{"awards": {"awards": [{"name": "Honor Roll ",...	This person must excel in a fast-paced environ...
11973	cleey05qa000exd2up87uehkz	{"awards": {"awards": []}, "header": {"role": ...	In collaboration with the Senior Communication...
11974	cle0edrgo00a5wz2utru0nt5u	{"awards": {"awards": [{"name": "Honor Roll ",...	About the job\nYou've got 52 weeks a year to f...
11975	cleecwhm6006dyf2tsr12f761	{"awards": {"awards": [{"name": "Honor Roll ",...	About InsideTracker\n\nCreated by experts in a...

11976 rows × 3 columns

Display Type, Length, Shape about the dataset:

Type	Length	Shape
<class 'pandas.core.frame.DataFrame'>	11976	(11976, 3)

Display datatypes of respective columns in dataset:

	id	content	jobDescription
0	object	object	object

Showing max, min length and NA values:

Column	Max Length	Min Length	NA Count
id	25	25	0
content	52580	513	0
jobDescription	22567	1	4

Final Dataset:

	ID	KEYWORDS	SUGGESTEDSKILLS	ROLE	CITY	STATE	SUMMARY	ACCOMPLISHMENTS	COUNTRY_CODE	SKILL_DESCRIPTION	...	PBL_DETAILS	PBL_PUBLISHER	CRT_NAME	CRT_ISSUER
0	CLG43D9AN007GX02UG1I694J6	[admissions representative, admissions, uma,...	[Compliance, Client, Manages, Interaction, Fin...		INDIO	CA	DETAILED AND DRIVEN, I HAVE BUILT STRONG COMMU...		US	VERBAL, WRITTEN, AND VISUAL COMMUNICATION, GOA...	...	NaN	NaN	QUALIFIED APPLICATOR CERTIFICATE	CALIFORNIA RIVERSIDE AGRICULTURE DEPARTMENT
1	CLG3ITETJ006JX92TDKCRW195	[dashboard interfaces, lead generation, mark...	[Analysis, Collection, Research]		ILMENAU	THURINGIA	DETAILED-ORIENTED UI/UX DESIGNER WITH EXPERIEN...		DE	FIGMA, SKETCH, ADOBE XD, FRAMER, MIRO, UXPIN, ...	...	NaN	NaN	VISUAL ELEMENTS OF USER INTERFACE DESIGN	CALIFORNIA INSTITUTE OF THE ARTS
2	CLG3IY1SD007RX32UTNUHNRGY	[product, design, development, business req...	[Vue, DevOps, Delivery]		PEORIA	ARIZONA	AGILE SOFTWARE ENGINEER WITH 2 YEARS OF EXPERI...		US	JIRA, VSC, MYSQL, GIT, BITBUCKET, GITHUB, POS...	...	NaN	NaN	NaN	NaN
3	CLG5J15LZ00K3X02UAAU7G9Z0	[flexibility, international exposure, dream ...	[]		MALANG		INNOVATIVE DIGITAL MARKETING PROFESSIONAL WITH...		GB	MARKETING ANALYTICS, WEBSITE ANALYTICS, PRODUC...	...	NaN	NaN	NaN	NaN
4	CLG43PTE600DDYA2UMAKFW3C3	[product roadmaps, new features, product enh...	[Curiosity]		CALGARY	AB	PASSIONATE JOB SEEKER WITH STRONG ORGANIZATION...		CA	CRITICAL AND ANALYTICAL THINKING, TIME MANAGEM...	...	NaN	NaN	SCRUM MASTER CERTIFICATION	LEARN QUEST
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
11971	CLEEXYZAG006AYG2VHR087ALS	[content, ideation, creation, camera, cont...	[Instagram, Calendar, TikTok]		BROOKLYN	NY	DEPENDABLE VIDEOGRAPHER AND VIDEO EDITOR WITH ...		US	PROJECT MANAGEMENT, SELF-DRIVEN, PRODUCTION PL...	...	NaN	NaN	NaN	NaN
11972	CLEEC90B0005NYF2TLOS9QC95	[adobe premiere, adobe after effects, adobe ...	[Broadcast, Promotional, Broadcast & Promotion...		PEABODY	MA			US	FACEBOOK LIVE, TWITCH, OBS, XSPLIT	...	NaN	NaN	LEARN HTML COURSE	CODECADEMY
11973	CLEEY0SQA000EXD2UPB7UEHKZ	[write, content, graphics, imagery, social...	[]		CRIVITZ	WI	PERSONABLE AND HARDWORKING PROFESSIONAL WITH E...		GB	NaN	...	NaN	NaN	INTERNATIONAL ORGANIZATION MANAGEMENT	UNIVERSITY OF GENEVA
11974	CLE0EDRGO00A5WZ2UTRU0NT5U	[accountable, challenges, social media, bes...	[]		PEABODY	MA	PROFESSIONAL WITH OVER A DECADE OF EXPERIENCE ...		US	FACEBOOK LIVE, TWITCH, OBS, XSPLIT	...	NaN	NaN	LEARN HTML COURSE	CODECADEMY
11975	CLEECWHM6006DYF2TSR12F761	[product, product team, product managers, s...	[Communicate]		PEABODY	MA			US	FACEBOOK LIVE, TWITCH, OBS, XSPLIT	...	NaN	NaN	LEARN HTML COURSE	CODECADEMY

11976 rows × 57 columns

VII. Analytics / Visuals:

Distribution of Country Codes with respect to ID

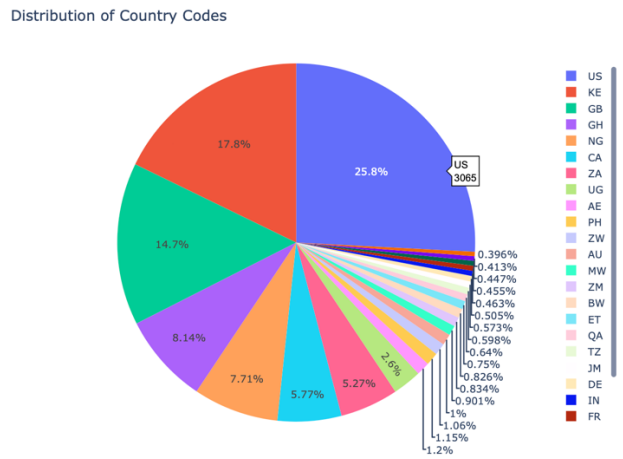


Fig: No Job Description Dataset

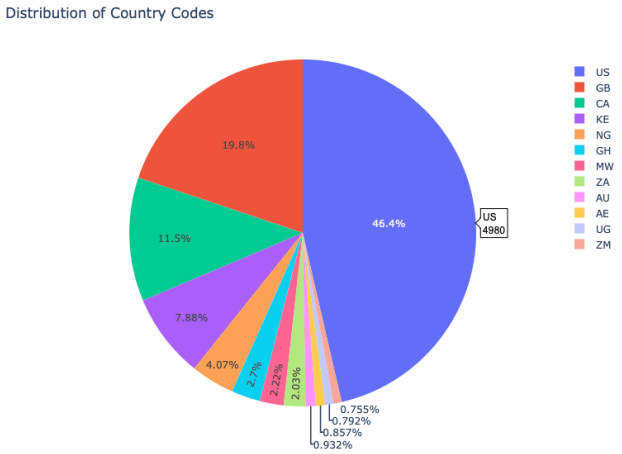


Fig: With Job Description Dataset

Top 3 countries as per number of users.

1. United States has maximum user: 25.8% of total user: 3065 applicants.
2. Kenya (KE) has 17.8% :2113 applicants.
3. Great Britain (GB): 14.7%: 1714 number of applicants

Top 3 countries as per number of users.

1. United States has maximum user: 46.4% of total user: 4980 applicants.
2. Great Britain (GB): 19.8%
3. Kenya (KE): 11.5%

Number of Candidates per City in respective Countries US and CA

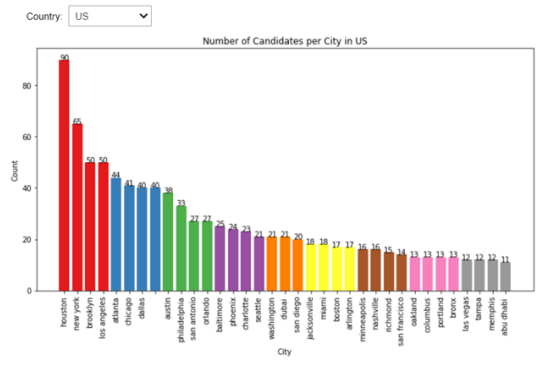


Fig: No Job Description Dataset

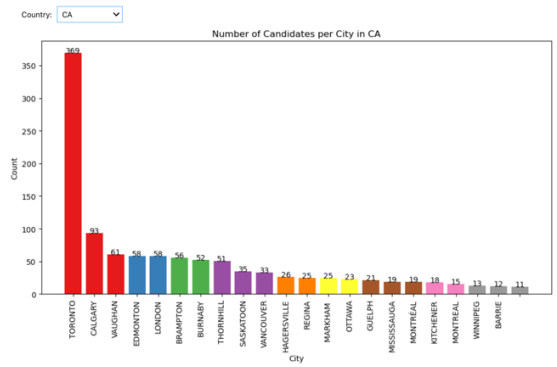


Fig: With Job Description Dataset

USA:

1. Houston with 90 applicants
2. New York: 65 applicants
3. Brooklyn, Los Angeles: 50
4. Least users: Las Vegas, Memphis: 12 and 11 applicants

CANADA:

1. TORONTO: 369 users
2. CALGARY: 93 users
3. Least number of users in CA MONTREAL, WINNIPEG, BARRIE: 15, 13, and 12

Determine the expected age and experience level?

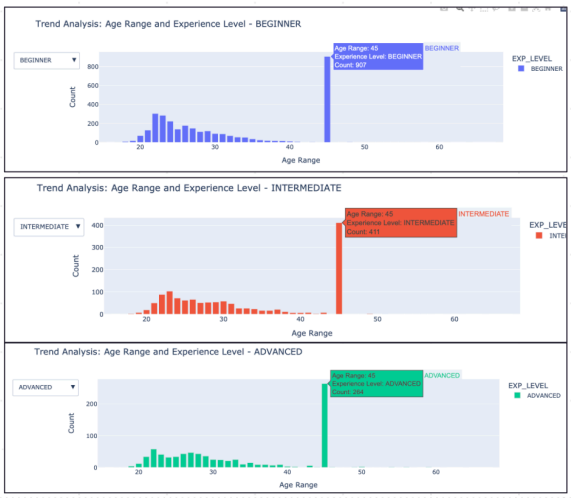


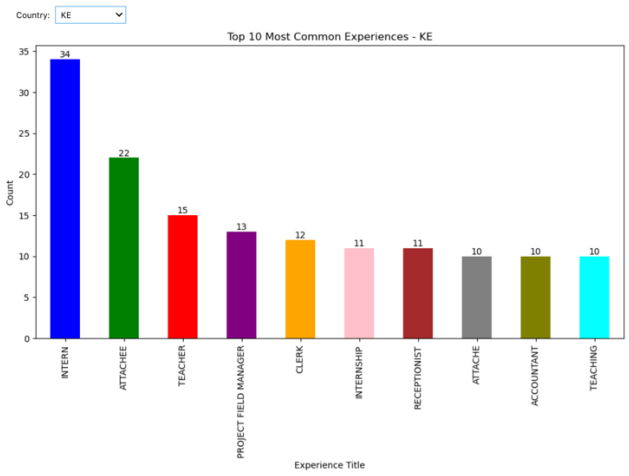
Fig: No Job Description Dataset

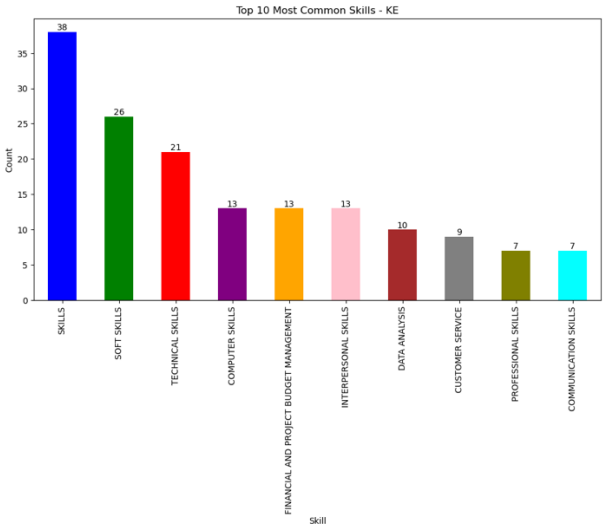
Fig: With Job Description Dataset

By selecting different experience levels from the dropdown menu, the chart dynamically updates to show the trend analysis specifically for that experience level. The title of the chart also changes accordingly to provide focused insights.

The x-axis represents the age ranges, ranging from 18-24, 25-34, 35-44, 45-54, and 55+. The y-axis represents the count of candidates. Each bar in the chart is segmented into different colors representing different experience levels.

Determine trends in experience and skills for the target users?





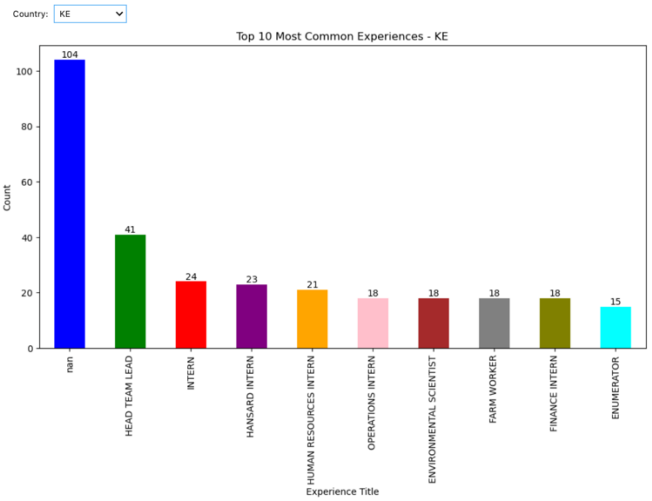
Top 10 Most Common Skills:

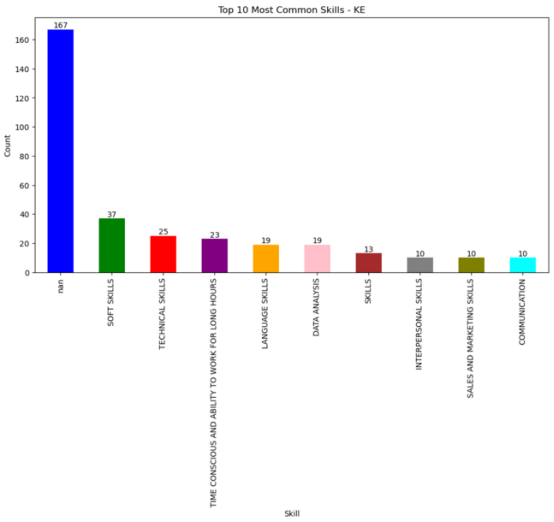
Skill	Count
SKILLS	38
SOFT SKILLS	26
TECHNICAL SKILLS	21
COMPUTER SKILLS	13
FINANCIAL AND PROJECT BUDGET MANAGEMENT	13
INTERPERSONAL SKILLS	13
DATA ANALYSIS	10
CUSTOMER SERVICE	9
PROFESSIONAL SKILLS	7
COMMUNICATION SKILLS	7

Top 10 Most Common Experiences:

Experience	Count
INTERN	34
ATTACHEE	22
TEACHER	15
PROJECT FIELD MANAGER	13
CLERK	12
INTERNSHIP	11
RECEPTIONIST	11
ATTACHE	10
ACCOUNTANT	10
TEACHING	10

Fig: No Job Description Dataset





Top 10 Most Common Skills:

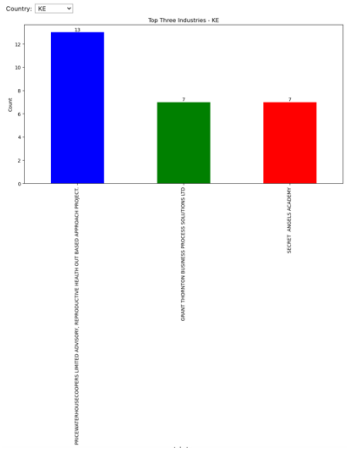
Skill	Count
nan	167
SOFT SKILLS	37
TECHNICAL SKILLS	25
TIME CONSCIOUS AND ABILITY TO WORK FOR LONG HOURS	23
LANGUAGE SKILLS	19
DATA ANALYSIS	19
SKILLS	13
INTERPERSONAL SKILLS	10
SALES AND MARKETING SKILLS	10
COMMUNICATION	10

Top 10 Most Common Experiences:

Experience	Count
nan	104
HEAD TEAM LEAD	41
INTERN	24
HANSARD INTERN	23
HUMAN RESOURCES INTERN	21
OPERATIONS INTERN	18
ENVIRONMENTAL SCIENTIST	18
FARM WORKER	18
FINANCE INTERN	18
ENUMERATOR	15

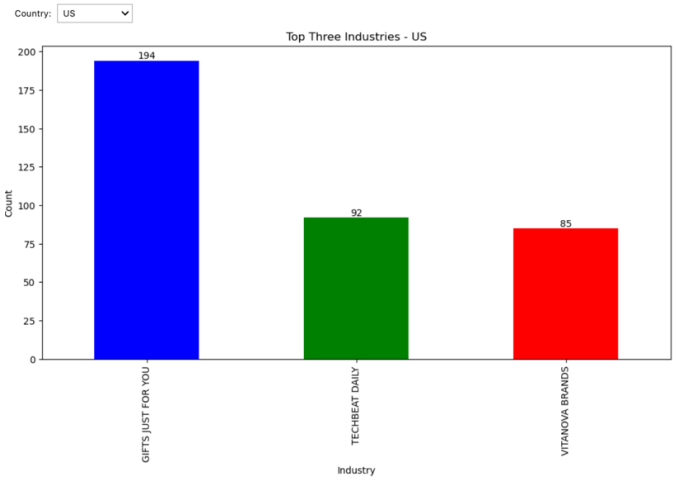
Fig: With Job Description Dataset

Top three industries that the majority of CoverQuick's users have applied?



```
Top Three Industries:
FREELANCE                          33
AMAZON                             21
GHANA EDUCATION SERVICE            17
```

Fig: No Job Description Dataset




```
Top Three Industries:
GIFTS JUST FOR YOU                          194
LIVINGSTONIA SYNOD AIDS PROGRAMME (LISAP)   163
EDMONTON FIRE RESCUE                        159
```

Fig: With Job Description Dataset

VIII. Clear Concise Flow

Discover trends in demographics and find which industries yield the best and the worst resumes:

Resume Optimality Criteria:

1. Important Sections: This may include and not be limited to: work experience, education, projects, as the most important and relevant sections.

EXP_DURATION	EXP_LEVEL	RES_LEN	ACTN_VERB	SPLNG_MSTK	IMP_SEC
nan	NaN	188	47	1	1
699.0	INTERMEDIATE	259	58	1	1
672.0	INTERMEDIATE	188	56	1	1
122.0	BEGINNER	161	49	1	1
92.0	BEGINNER	136	41	1	1
...	...	...	...	...	...
519.0	INTERMEDIATE	145	40	1	1
nan	NaN	484	116	1	1
730.0	INTERMEDIATE	155	36	1	1
nan	NaN	349	89	1	1
nan	NaN	451	109	1	1

We included and not be limited to: work experience, education.

We are checking both the conditions and mapping value to 1 in IMP_SEC

```
imp_sec.loc[~imp_sec['EXP_DURATION'].isna(
) & (imp_sec['EDU_GRAD_YEAR'] != 1900),
'IMP_SEC'] = 1
```

```
1    8486
0    3490
```

2. Resume Length: The solid resume length may be between 300-500 words, however; if the length is outside this range, it may not mean a resume is poor.

EXP_DURATION	EXP_LEVEL	RES_LEN
nan	NaN	188
699.0	INTERMEDIATE	259
672.0	INTERMEDIATE	188
122.0	BEGINNER	161
92.0	BEGINNER	136
...	...	...
519.0	INTERMEDIATE	145
nan	NaN	484
730.0	INTERMEDIATE	155
nan	NaN	349
nan	NaN	451

```
RES_LEN Non-Null and Null Count:
==================================
   Total Count  Non-Null Count  Null Count
0        11976           11976           0
```

We iterate over the specified columns and count the total number of words. The word count is then added as a new column 'RES_LEN' to the DataFrame.

Finally, we mapped it to 'POOR' if it's below 300 and 'GOOD' if it's greater than or equal to 300 using our scorecard.

3. Use of action verbs: Direct use of action verbs in the bullets of a resume will ensure a resume will perform better.

EXP_DURATION	EXP_LEVEL	RES_LEN	ACTN_VERB
nan	NaN	188	47
699.0	INTERMEDIATE	259	58
672.0	INTERMEDIATE	188	56
122.0	BEGINNER	161	49
92.0	BEGINNER	136	41
...	...	...	...
519.0	INTERMEDIATE	145	40
nan	NaN	484	116
730.0	INTERMEDIATE	155	36
nan	NaN	349	89
nan	NaN	451	109

```
import nltk
from nltk import word_tokenize
from nltk.corpus import stopwords
from nltk.corpus import wordnet

# Download necessary NLTK resources if not already downloaded
nltk.download('punkt')
nltk.download('stopwords')
nltk.download('wordnet')
nltk.download('omw-1.4')
```

```
ACTN_VERB Non-Null and Null Count:
====================================
   Total Count  Non-Null Count  Null Count
0        11976           11976           0
```

4. No use of pronouns: Resumes should not contain pronouns such as I, we or me written in the document.

```
pronoun = imp_sec
```

```
pronoun_columns = ['ROLE', 'CITY', 'STATE', 'SUMMARY',
                   'ACCOMPLISHMENTS', 'SKILL_DESCRIPTION', 'SKILL',
                   'EDU_MINOR', 'EDU_AWARDS', 'EDU_SCHOOL', 'EDU_PROGRAM', 'EDU_COURSEWORK',
                   'VLNTR_TITLE', 'VLNTR_DESCRIPTION', 'VLNTR_ORGANIZATION',
                   'EXP_TITLE', 'EXP_COMPANY', 'EXP_DESCRIPTION',
                   'PRJ_LINK', 'PRJ_TITLE', 'PRJ_SKILLS', 'PRJ_DESCRIPTION',
                   'REF_NAME', 'REF_EMAIL', 'REF_RELATIONSHIP',
                   'CRT_NAME', 'CRT_ISSUER',
                   'AWD_NAME', 'AWD_ISSUER', 'AWD_DETAILS', 'AWD_DESCRIPTION']

pronouns = ['I', 'we', 'me']
pattern = fr'\b(?:{"|".join(pronouns)})\b'

df['PRONOUN_CNT'] = df[pronoun_columns].apply(lambda x: x.str.contains(pattern, case=False, regex=True).sum(), axis=1)

display(df)
```

```
# Count the occurrences in PRONOUN_CNT column
count = df['PRONOUN_CNT'].value_counts()

# Display the count
print(count)
```

```
0    9411
1    2171
2     303
3      88
4       3
Name: PRONOUN_CNT, dtype: int64
```

5. Excessive bullet points: A resume experience or section should not have an excessive number of bullet points. If a section has over 10 bullet points, it is looked upon unfavorably.

```python
import re

# Define the columns to check for excessive bullet points
bullet_columns = ['ROLE', 'CITY', 'STATE', 'SUMMARY',
                  'ACCOMPLISHMENTS', 'SKILL_DESCRIPTION', 'SKILL',
                  'EDU_MINOR', 'EDU_AWARDS', 'EDU_SCHOOL', 'EDU_PROGRAM', 'EDU_COURSEWORK',
                  'VLNTR_TITLE', 'VLNTR_DESCRIPTION', 'VLNTR_ORGANIZATION',
                  'EXP_TITLE', 'EXP_COMPANY', 'EXP_DESCRIPTION',
                  'PRJ_LINK', 'PRJ_TITLE', 'PRJ_SKILLS', 'PRJ_DESCRIPTION',
                  'REF_NAME', 'REF_EMAIL', 'REF_RELATIONSHIP',
                  'CRT_NAME', 'CRT_ISSUER',
                  'AWD_NAME', 'AWD_ISSUER', 'AWD_DETAILS', 'AWD_DESCRIPTION']

# Define the pattern to match bullet points
bullet_pattern = r'\n\s*[-•*]'

# Function to count the number of bullet points in a string
def count_bullet_points(text):
    if pd.isnull(text):
        return 0
    return len(re.findall(bullet_pattern, text))

# Add the "BULLET_CNT" column to the DataFrame
df['BULLET_CNT'] = df[bullet_columns].applymap(count_bullet_points).sum(axis=1)

# Display the updated DataFrame
display(df)
```

```python
# Count the occurrences in BULLET_CNT column
count = df['BULLET_CNT'].value_counts()

# Display the count
print(count)
```

```
4     3426
2     1985
3     1702
0     1561
5      668
1      525
6      513
7      409
8      406
12     172
10     133
9      115
11      70
13      44
16      40
15      39
14      31
17      25
26      13
20      13
25      13
24      13
18      11
21       9
19       8
23       6
22       6
35       2
40       2
45       2
27       2
34       2
31       2
46       2
36       2
30       2
38       1
86       1
Name: BULLET_CNT, dtype: int64
```

6. Spelling Mistakes: A resume with spelling errors is immediately penalized against.

EXP_DURATION	EXP_LEVEL	RES_LEN	ACTN_VERB	SPLNG_MSTK
nan	NaN	188	47	1
699.0	INTERMEDIATE	259	58	1
672.0	INTERMEDIATE	188	56	1
122.0	BEGINNER	161	49	1
92.0	BEGINNER	136	41	1
...	...	...	...	...
519.0	INTERMEDIATE	145	40	1
nan	NaN	484	116	1
730.0	INTERMEDIATE	155	36	1
nan	NaN	349	89	1
nan	NaN	451	109	1

```
!pip install spellchecker
```

```python
import pandas as pd
from spellchecker import SpellChecker
splng_mstk.SPLNG_MSTK.unique()
```

```
array([ 1,  0,  2, 24,  9,  3,  5,  4])
```

```
SPLNG_MSTK Non-Null and Null Count:
====================================
   Total Count  Non-Null Count  Null Count
0        11976           11976           0
```

7. Excessive sentence or bullet length

```python
# Define the columns to check for excessive sentence or bullet length
bullet_columns = ['ROLE', 'CITY', 'STATE', 'SUMMARY',
                  'ACCOMPLISHMENTS', 'SKILL_DESCRIPTION', 'SKILL',
                  'EDU_MINOR', 'EDU_AWARDS', 'EDU_SCHOOL', 'EDU_PROGRAM', 'EDU_COURSEWORK',
                  'VLNTR_TITLE', 'VLNTR_DESCRIPTION', 'VLNTR_ORGANIZATION',
                  'EXP_TITLE', 'EXP_COMPANY', 'EXP_DESCRIPTION',
                  'PRJ_LINK', 'PRJ_TITLE', 'PRJ_SKILLS', 'PRJ_DESCRIPTION',
                  'REF_NAME', 'REF_EMAIL', 'REF_RELATIONSHIP',
                  'CRT_NAME', 'CRT_ISSUER',
                  'AWD_NAME', 'AWD_ISSUER', 'AWD_DETAILS', 'AWD_DESCRIPTION']

# Define the maximum length threshold for excessive length
max_length = 100

# Function to count the number of sentences or bullet points with excessive length
def count_excessive_length(text):
    if pd.isnull(text):
        return 0
    sentences = re.split(r'(?<!\w\.\w.)(?<![A-Z][a-z]\.)(?<=\.|\?|\!)\s', text)
    return sum(len(sentence) > max_length for sentence in sentences)

# Add the "BULLET_LEN" column to the DataFrame
df['BULLET_LEN'] = df[bullet_columns].applymap(count_excessive_length).sum(axis=1)

# Display the updated DataFrame
display(df)
```

```python
# Count the occurrences in BULLET_LEN column
count = df['BULLET_LEN'].value_counts()

# Display the count
print(count)
```

```
3     1802
5     1771
4     1643
6     1286
2     1186
7      895
1      857
8      602
0      531
9      427
10     360
11     165
12     101
13      79
14      64
16      51
15      38
19      29
18      28
20      24
17      18
21       6
25       3
22       2
23       2
32       1
31       1
26       1
34       1
24       1
44       1
Name: BULLET_LEN, dtype: int64
```

IX. Analysis and synthesis of the data

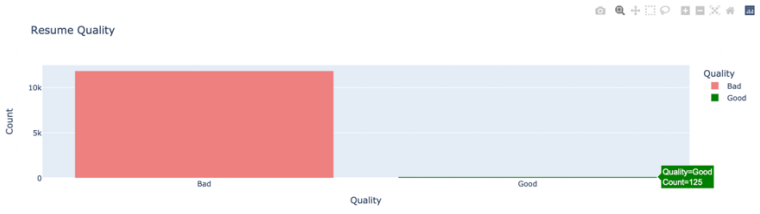
SCORE CARD CALCULATOR

Creating a score-card calculator column with name "RES_QUALITY" with value Good and Bad from data frame df using columns 'RES_LEN', 'ACTN_VERB','SPLNG_MSTK','IMP_SEC','PRONOUN_CNT', 'BULLET_CNT', 'BULLET_LEN' using the below logic.

Check the values in the respective columns using the following conditions to evaluate Good or Bad resume.

- If RES_LEN is more than 300 then score it as 1 or else 0.
- If ACTN_VERB is more than 20 then score it as 1 or else 0.
- If SPLNG_MSTK is 0 then score it as 1 or else 0 for anything greater than that.
- If IMP_SEC is 1 then score it as 1 or else 0.
- If PRONOUN_CNT is less than 1 than score it as 1 or else 0 for anything greater than or equal to that.
- If BULLET_CNT is less than 7 than score it as 1 or else 0 for anything greater than or equal to that.
- If BULLET_LEN is less than 20 than score it as 1 or else 0 for anything greater than or equal to that.

If all the above conditions meet, then assign value "Good" to column "RES_QUALITY" or else "Bad" for the rest.




```
# Count the occurrences in BULLET_LEN column
count = df['RES_QUALITY'].value_counts()

# Display the count
print(count)
```

```
Bad     11851
Good      125
Name: RES_QUALITY, dtype: int64
```

X. Recommendations And Findings

TABLEAU: MAIN DASHBOARD

We have presented a comprehensive and detailed examination on our Tableau dashboard, illustrating the differentiation between well-crafted and subpar resumes. Additionally, we have provided an analysis of

recommended skills categorized by country, state, and city. The abundance of information available offers valuable insights into our dataset.

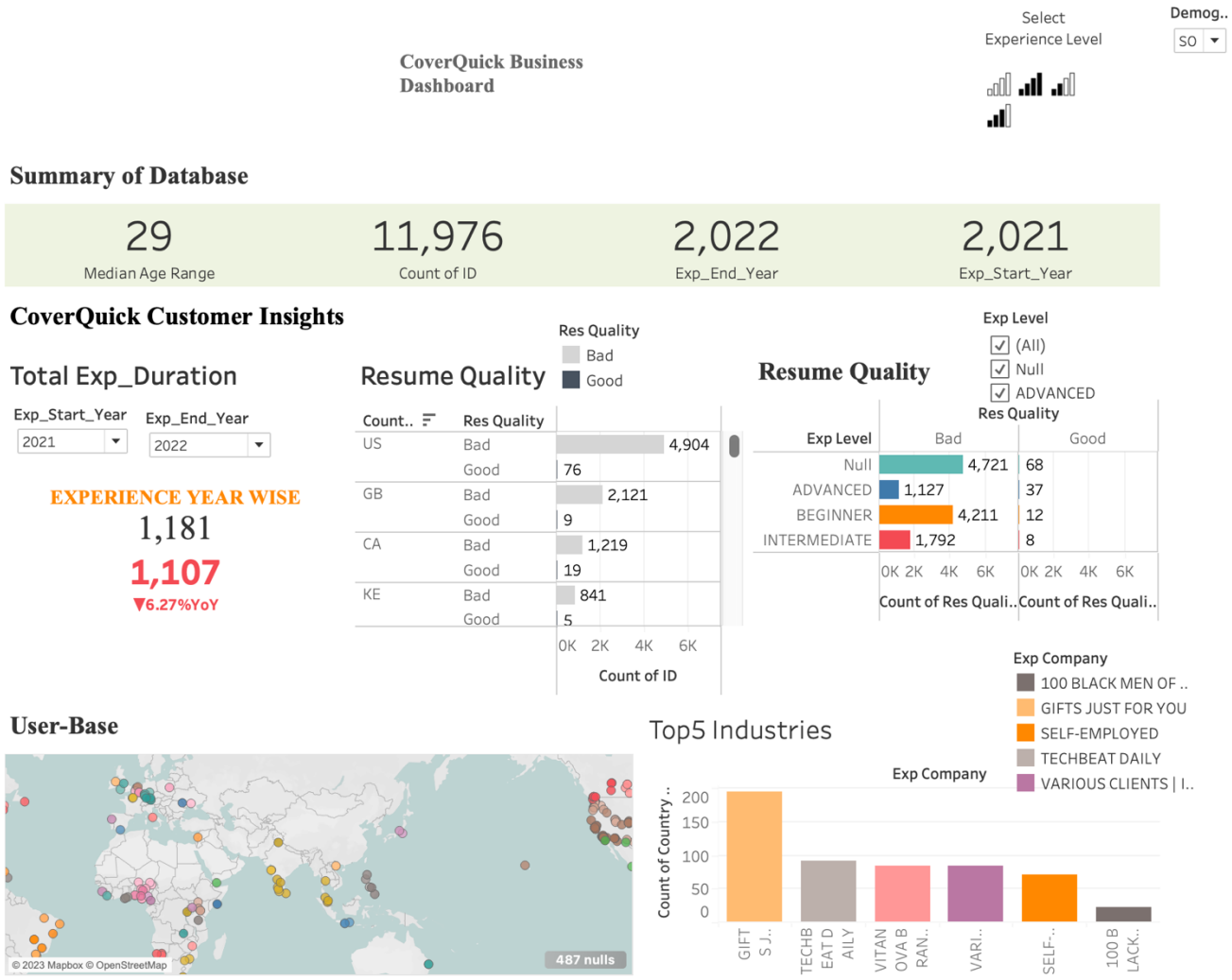
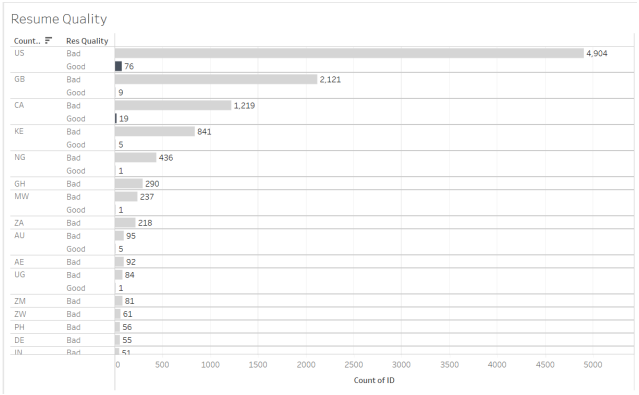


TABLEAU IN DEPTH ANALYSIS: GOOD AND BAD RESUME



Total Resume Distribution: Out of the total resumes analyzed, we observe the following distribution:

- Bad Resumes: 11,851
- Good Resumes: 125

Resume Quality in the United States: Among the resumes analyzed from the United States, the distribution is as follows:

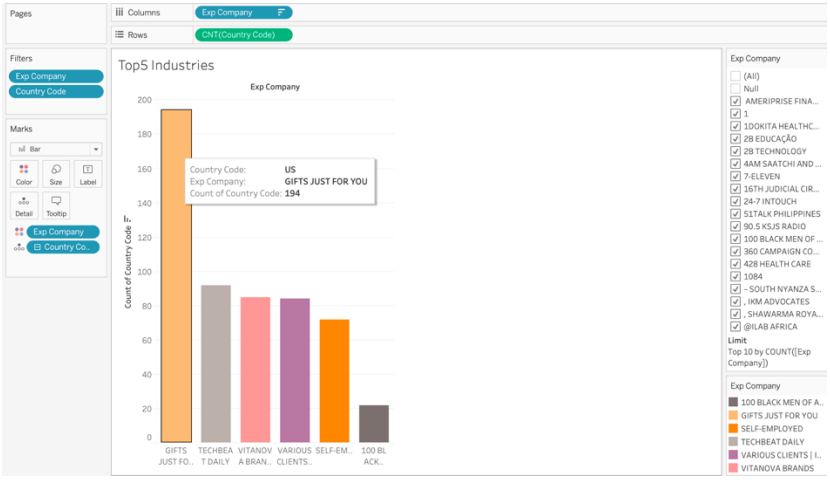
- Bad Resumes: 4,904

- Good Resumes: 76

Resume Quality in Canada: Among the resumes analyzed from Canada, the distribution is as follows:

- Bad Resumes: Majority
- Good Resumes: 19

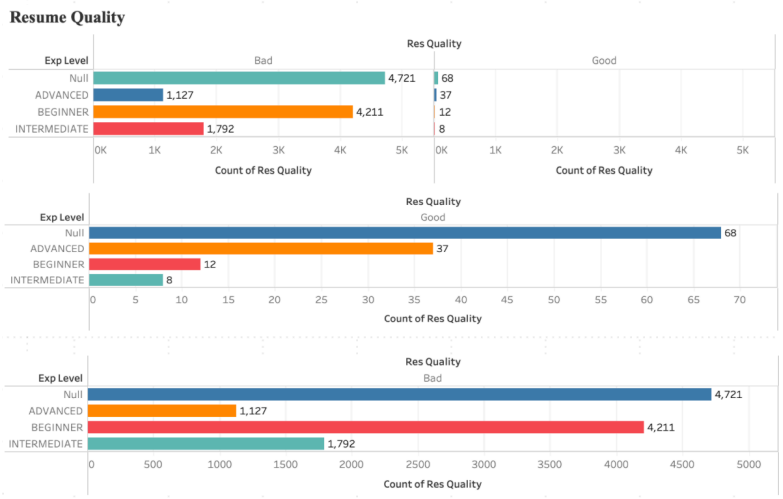
TABLEAU: TOP INDUSTRIES



Based on our data visualization, Top industries from overall countries are shown in the figure.

US has the highest number of candidates where they have applied for job in Gifts Just For You.

TABLEAU: GOOD/BAD RESUME ANALYSIS BASED ON EXPERIENCE LEVEL



- What we have identified, for GOOD resume, there are 37 candidates who have ADVANCED experience level.
- 12 candidates with BEGINNER and 8 candidates with INTERMEDIATE experience level.

XI. Conclusion:

Based on the dataset provided, we successfully addressed the research questions assigned to us for analysis. However, we went beyond the initial expectations and conducted additional analysis using Tableau. This analysis delved into the demographics, revealing the number of candidates applying for each role, their experience levels, age ranges (both minimum and maximum), and the recommended skills for specific regions.

Furthermore, we depicted the disparities in resume quality among different countries, highlighting strengths and weaknesses through our scorecard calculator.

Utilizing artificial intelligence, we can delve even deeper by modifying resumes for candidates applying to specific job roles. This involves suggesting alternative approaches based on market demand and addressing company requirements.

Lastly, we express our gratitude to our sponsors for providing us with an exceptional dataset that facilitated valuable learning. We are confident that our analysis will offer valuable insights to our sponsors for further action.

XII. References

Connect with CoverQuick:
Email: support@coverquick.co
Discord: https://discord.com/invite/2gfk9Yyv2b

XIII. Other relevant information

Connect with Us:
GitHub: Abhilash Dikshit, Kush Patel, Siddharth Alashi
LinkedIn: Abhilash Dikshit, Kush Patel, Siddharth Alashi