# Ranking Resumes based on Job Requirements

... using NLP and other cool stuff

# In the next 10 minutes..

01

## Problem

What is the business problem?

02

## Solution

How did I plan to solve the problem?

03

## Project

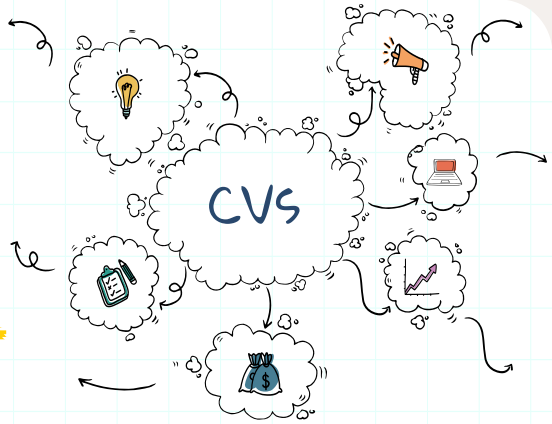
How I actually (*semi-*)solved it..

04

## Takeaways

Limitations, learnings and next steps

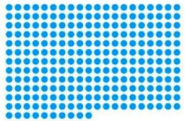
# Problem




Imagine you are a **recruiter** who receives many low quality applications on a daily. How do you filter?

Corporate job openings attract

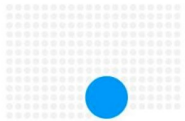
**250** resumes



Out of these candidates

**4-6** will be interviewed

Only **1** will get a job offer



Source: Glassdoor

| Source Name | Hire Count | Hire Percentage | Applications | Application Percentage | Effectiveness |
|---|---|---|---|---|---|
| Internal Hire | 37463 | 0.077 | 145516 | 0.004 | 18.183 |
| Referral | 181268 | 0.374 | 2298343 | 0.067 | 5.57 |
| Federated Search | 557 | 0.001 | 12283 | 0 | 3.203 |
| Agency | 23380 | 0.048 | 603749 | 0.018 | 2.735 |
| Resume Search | 424 | 0.001 | 12657 | 0 | 2.366 |
| Email Campaign | 2463 | 0.005 | 97877 | 0.003 | 1.777 |
| Custom Campaign | 26279 | 0.054 | 1062531 | 0.031 | 1.747 |
| Career Sites | 106049 | 0.219 | 11714598 | 0.342 | 0.639 |
| Job Board | 86618 | 0.179 | 14793496 | 0.432 | 0.414 |
| Social Media Shares | 522 | 0.001 | 111201 | 0.003 | 0.332 |
| Notifications | 817 | 0.002 | 176367 | 0.005 | 0.327 |

# Solution: Ranking CVs based on job requirements

### Time Saver

Reduce time spent on applications for recruiters

### Talent First

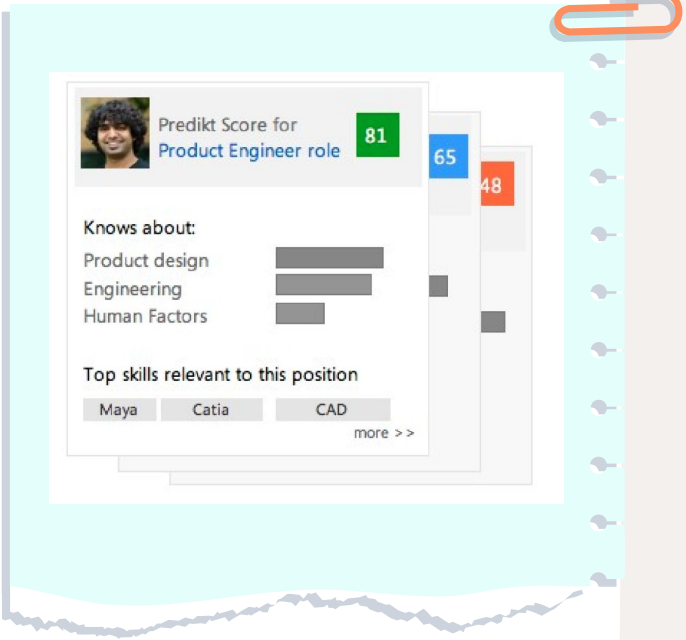
Connect with top talent faster

### Efficiency

Reduce time-to-fill

### Fairness?!

Can we reduce some biases?



# RESUME RANKING USING MACHINE LEARNING — IMPLEMENTATION

Synerzip Aug 3, 2018 · 5 min read

Author: Vinayak Joglekar, Co-Founder & CTO, Synerzip.

# How I used NLP (Spacy) to screen Data Science Resume

Position your Data Science Resume better through NLP

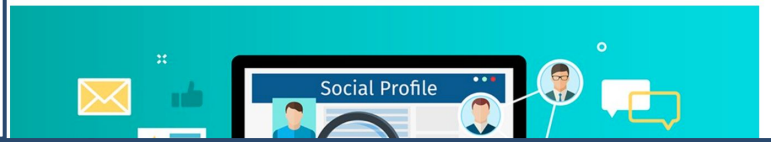
Venkat Raman Jan 14, 2019 · 6 min read


# Ranking resumes for a given job description using Natural Language Processing

A Toy project

Vishwanath Beena Follow

Jul 27 · 7 min read



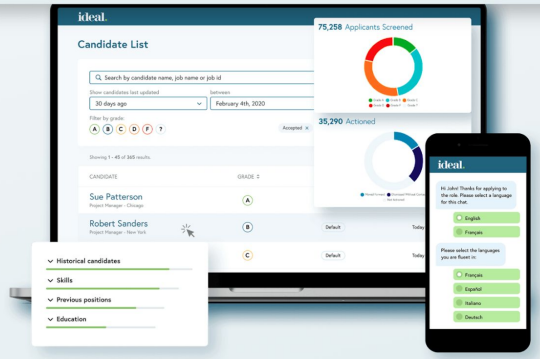

Of course, many people have already thought of this

Resources Request a Demo

## Screen Thousands of Candidates in Seconds.

Ideal is an AI-powered talent screening & matching system that helps enterprise teams make more accurate, fair, and efficient talent decisions.

Request a Demo Learn More




# "Resume Ranking using NLP and Machine Learning"

Project Report

Submitted in fulfillment of the requirements for the degree of

**Bachelor of Engineering**

by

**Juneja Afzal Ayub Zubeda** (12CO32)
**Momin Adnan Ayyas Shaheen**(12CO46)
**Gunduka Rakesh Narsayya Godavari**(12CO29)
**Sayed ZainulAbideen Mohd Sadiq Naseem** (13CO72)

Supervisor

**Prof. Tabrez Khan**

Co-Supervisor

**Prof. Irfan Jamkhandikar**

# In the end, I tried lots of things...

## Cosine Similarity

The CV ranking is based on calculating cosine similarity between the job description and the CVs.

## Job Skills Matching

The ranking is based on concrete job requirements and if those have been mentioned in the CV.

# Our starting point

## CVs

- Database + 2500 CVs
- Focused on 10 CVs

## Job Description

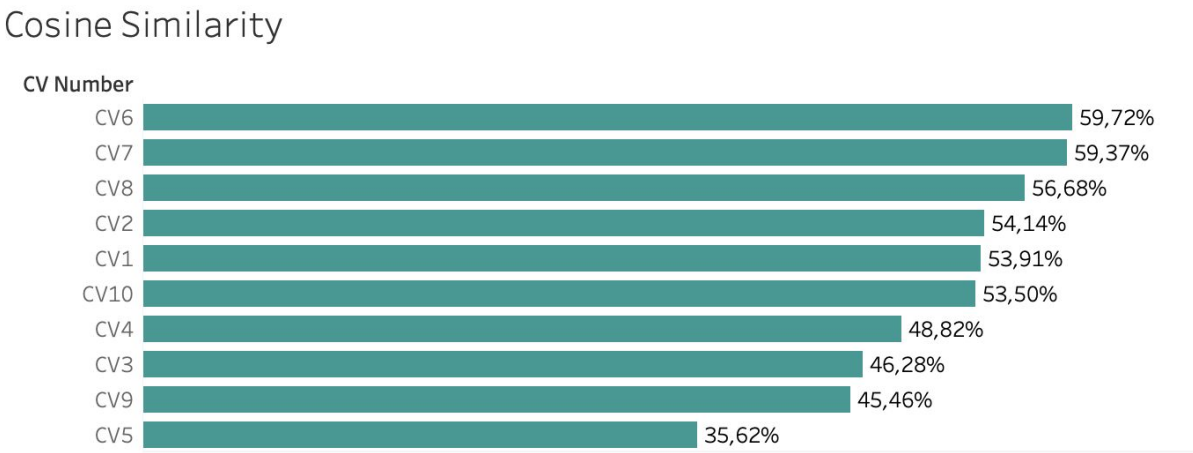
- Focused on one: Data Analyst

## Job Skills

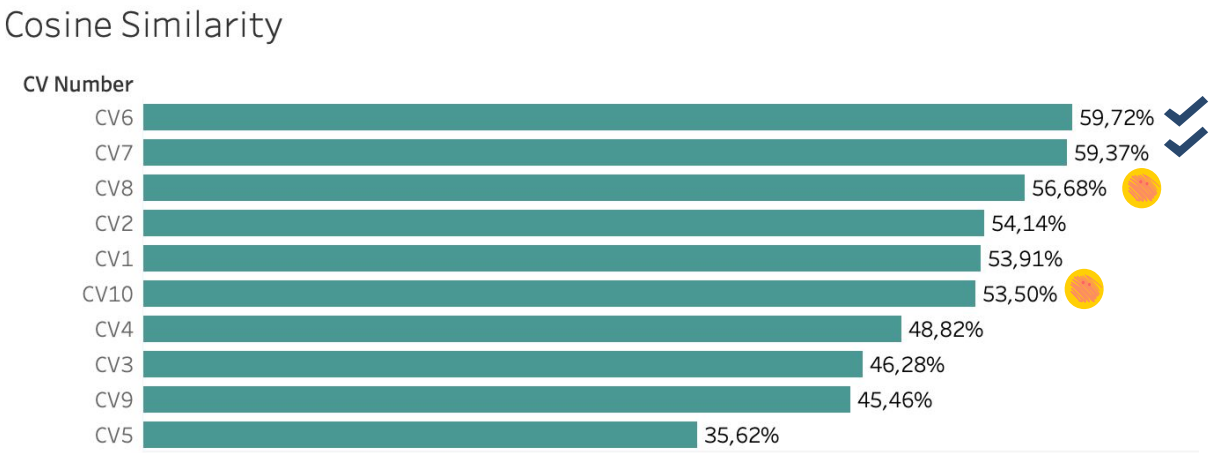
- csv file with job skills requirements based on the job posting

| Profile | Last Job Title |
|---|---|
| CV1 | Accountant |
| CV2 | Assistant Manager – Finance |
| CV3 | Senior Audit Associate |
| CV4 | Assistant Fund Manager |
| CV5 | Intern Sales |
| **CV6** | **Data Analyst** |
| **CV7** | **Data Warehouse Analyst** |
| CV8 | Assistant, Finance |
| CV9 | Senior Analyst, Finance & Treasury |
| **CV10** | **Data Analyst** |

# Method #1

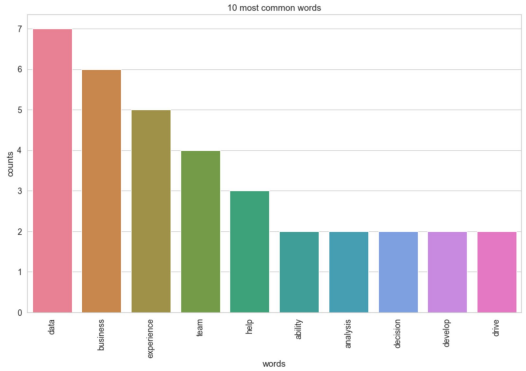

Cosine Similarity
CV Number
CV6
59,72%
CV7
59,37%
CV8
56,68%
CV2
54,14%
CV1
53,91%
CV10
53,50%
CV4
48,82%
CV3
46,28%
CV9
45,46%
CV5
35,62%

# Method #1




Seems it didn't tell the whole story ..

# Method #2-...100




```
Topics via LDA:

Topic #0:
experience ability skill acquire user metric forecasting notebook solving engineering

Topic #1:
product analysis feature hive spark strong analytics decide big engaged

Topic #2:
data sale deliver identify present tour ticket highly looker key

Topic #3:
plus finding knowledge measurement dashboard minimal relevant international interpreting issue

Topic #4:
business help problem intelligence package engaged requirement make implement engineering

Topic #5:
meaning science insight report quantitative communication great environment explore direction

Topic #6:
develop year range statistical getyourguide opportunity demonstrated tactical company new

Topic #7:
decision large success improve individual proven python responsibility serve solve

Topic #8:
drive tool strategy tech reas

Topic #9:
team understand number report
```

| | keyword | score |
|---|---|---|
| 0 | data | 0.311103 |
| 1 | decision | 0.255610 |
| 2 | decisions | 0.255610 |
| 3 | experience | 0.221352 |
| 4 | team | 0.195705 |
| 5 | teams | 0.195705 |
| 6 | business | 0.174313 |
| 7 | product | 0.151129 |
| 8 | develop | 0.148329 |
| 9 | skills | 0.138911 |

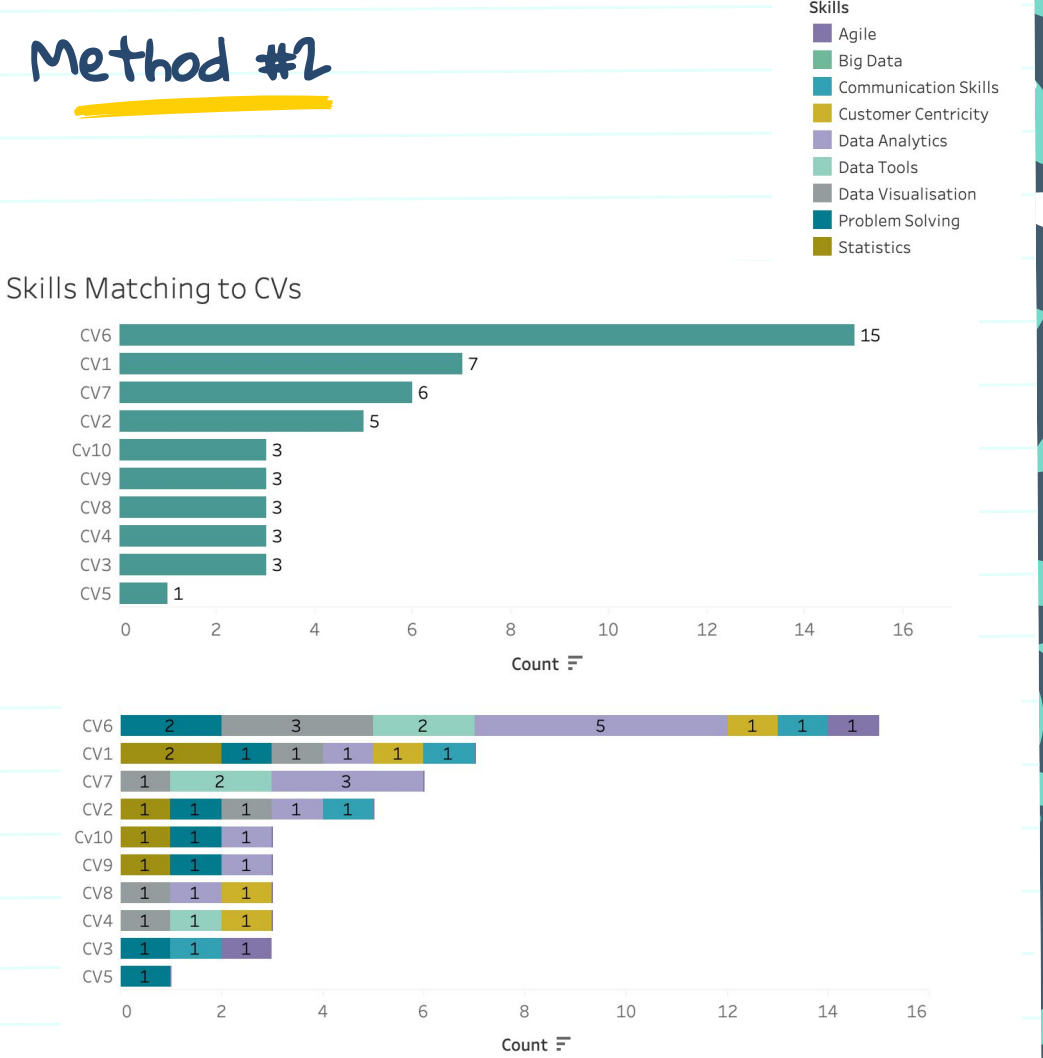
| | score | Phrase |
|---|---|---|
| 0 | 46.500000 | highly engaged individuals great communication... |
| 1 | 23.500000 | product decisions explore large datasets |
| 2 | 22.400000 | business intelligence role strong knowledge |
| 3 | 16.500000 | make impactful decisions design |
| 4 | 14.750000 | improve customer service experience |
| 5 | 14.071429 | data engineering team requirements |
| 6 | 13.321429 | relevant experience interpreting data |
| 7 | 10.000000 | new product features |
| 11 | 9.000000 | making within getyourguide |
| 12 | 9.000000 | develop quantitative analysis |
| 8 | 9.000000 | tech company familiarity |
| 10 | 9.000000 | sales team decide |
| 9 | 9.000000 | stakeholders deliver analysis |
| 13 | 8.333333 | help develop strategies |
| 14 | 8.071429 | big data tools |
| 15 | 8.000000 | understand user behavior |
| 16 | 7.833333 | help drive measurement |
| 17 | 7.000000 | plus demonstrated ability |
| 18 | 6.500000 | reasoning skills |
| 19 | 4.900000 | drive business |

**Pro tip:**
Keywords **do NOT** equal specific skills.

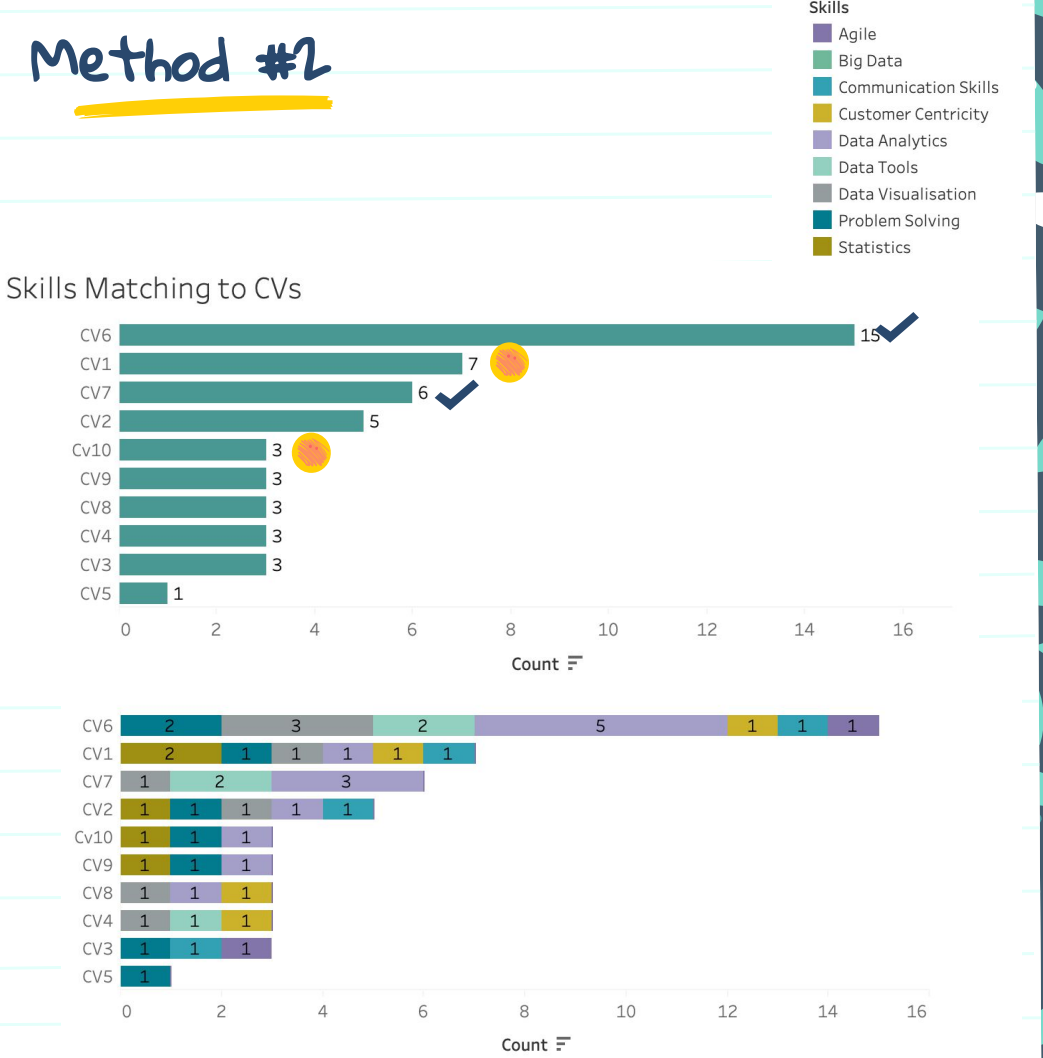
# Job Skills Required

| Data Analytics | Statistics | Big Data | Data Visualization | Data Tools | Problem Solving | Customer Centricity | Adaptability | Communication Skills |
|---|---|---|---|---|---|---|---|---|
| business analytics | statistical models | Spark | dashboards | Looker | solution-oriented | user behaviour | independent | interpersonal skills |
| quantitative analysis | prediction | Hive | reports | Jupyter Notebook | problem analysis | user experience | autonomous | presentation skills |
| data reporting | forecasting | large datasets | metrics | SQL | analytical | customers | curiosity | Communication |
| qualitative analysis | machine learning | hadoop | storytelling | Python | detailed | usability | open minded | written communication |
| predictive analytics | statistical consulting | cloudera | KPI | business intelligenc | systematic | prototyping | perseverance | |
| predictive modeling | logistical regression | mapreduce | KPIs | Tableau | logical | | flexibility | |
| customer analysis | bayesian statistics | impala | | MySQL | | | | |
| marketing mix modeling | probability | sqoop | | numpy | | | | |
| database marketing | variance analysis | | | pandas | | | | |
| data analysis | hypothesis testing | | | matplotlib | | | | |
| data structures | ab testing | | | scipy | | | | |
| data modeling | | | | nltk | | | | |
| deep learning | | | | sklearn | | | | |

# Method #2


Skills
Agile
Big Data
Communication Skills
Customer Centricity
Data Analytics
Data Tools
Data Visualisation
Problem Solving
Statistics
Skills Matching to CVs
CV6 15
CV1 7
CV7 6
CV2 5
Cv10 3
CV9 3
CV8 3
CV4 3
CV3 3
CV5 1
0 2 4 6 8 10 12 14 16
Count
CV6 2 3 2 5 1 1 1
CV1 2 1 1 1 1 1
CV7 1 2 3
CV2 1 1 1 1 1
Cv10 1 1 1
CV9 1 1 1
CV8 1 1 1
CV4 1 1 1
CV3 1 1 1
CV5 1
0 2 4 6 8 10 12 14 16
Count

# Method #2


Skills
Agile
Big Data
Communication Skills
Customer Centricity
Data Analytics
Data Tools
Data Visualisation
Problem Solving
Statistics
Skills Matching to CVs
CV6 15
CV1 7
CV7 6
CV2 5
Cv10 3
CV9 3
CV8 3
CV4 3
CV3 3
CV5 1
0 2 4 6 8 10 12 14 16
Count
CV6 2 3 2 5 1 1 1
CV1 2 1 1 1 1 1
CV7 1 2 3
CV2 1 1 1 1 1
Cv10 1 1 1
CV9 1 1 1
CV8 1 1 1
CV4 1 1 1
CV3 1 1 1
CV5 1
0 2 4 6 8 10 12 14 16
Count

# Let's summarize!

| Profile | Last Job Title | Cosine Similarity Output | Skills Matching Output | Recruiter Assessment | Ranking Cosine Similarity | Ranking Skills Matching | Delta |
|---|---|---|---|---|---|---|---|
| CV1 | Accountant | 53.91% | 7 | No | 5 | 2 | 3 |
| CV2 | Assistant Manager – Finance | 54.14% | 5 | No | 4 | 4 | 0 |
| CV3 | Senior Audit Associate | 46.28% | 3 | No | 8 | 9 | -1 |
| CV4 | Assistant Fund Manager | 48.82% | 3 | No | 7 | 8 | -1 |
| CV5 | Intern Sales | 35.62% | 1 | No | 10 | 10 | 0 |
| CV6 | Data Analyst | 59.72% | 15 | Yes | 1 | 1 | 0 |
| CV7 | Data Warehouse Analyst | 59.37% | 6 | Yes | 2 | 3 | -1 |
| CV8 | Assistant, Finance | 56.68% | 3 | No | 3 | 7 | -4 |
| CV9 | Senior Analyst, Finance & Treasury | 45.46% | 3 | No | 9 | 6 | 3 |
| CV10 | Data Analyst | 53.50% | 3 | No | 6 | 5 | 1 |

Overall, both models have **similar results**. Still, both models **can be used only for ranking** and not a solid decision-making.

# Final words

## My To-Do

- [ ] Explore more NLP tools!
- [ ] Scalable coding
- [ ] Try the model with an API

## Your To-Do

- [ ] Take the time to customize your CV for the roles you want to apply
- [ ] If using NLP- make sure you explore first the topic and then coding.
- [x] Open that champagne..


IRON
HACK
Thanks!