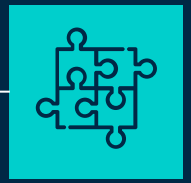
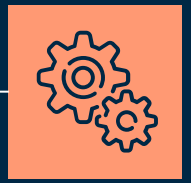
Job Matching Recommendation System

TABLE OF CONTENTS



01 PROBLEM STATEMENT



02 EXPLORATORY DATA ANALYSIS

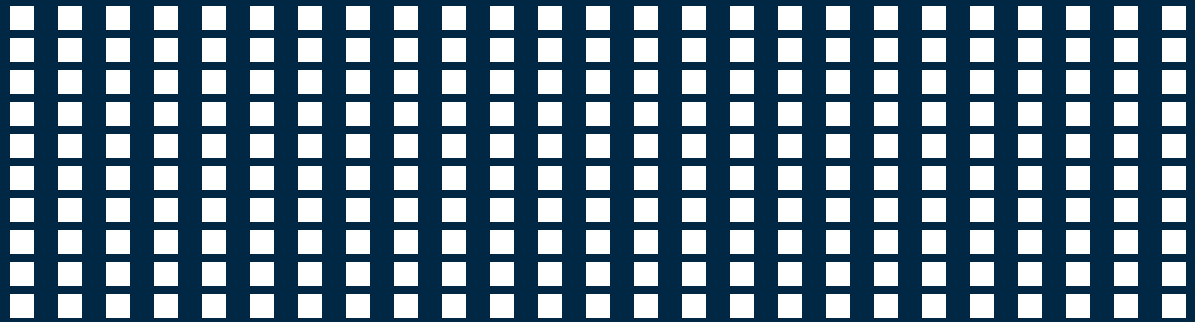
03 RECOMMENDATION SYSTEM

UNDERSTANDING THE PROBLEM

UNDERSTANDING THE PROBLEM

CORPORATE JOB OPENING ATTRACTS
250 RESUMES

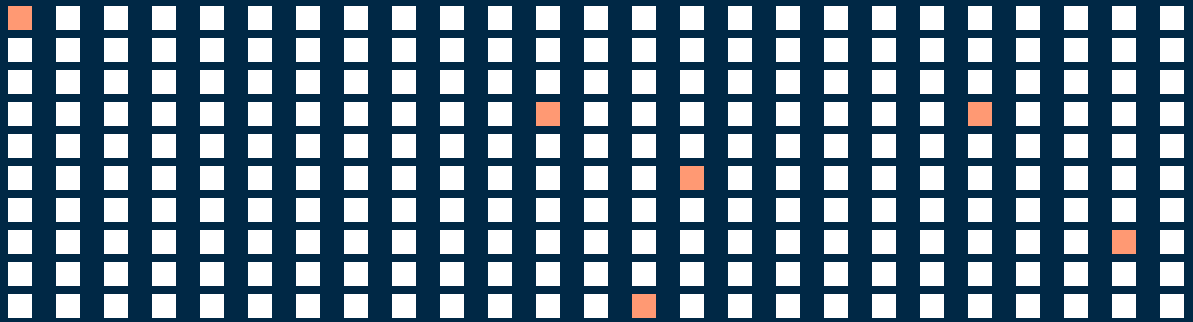
6 SECONDS
SPENT PER RESUME



UNDERSTANDING THE PROBLEM

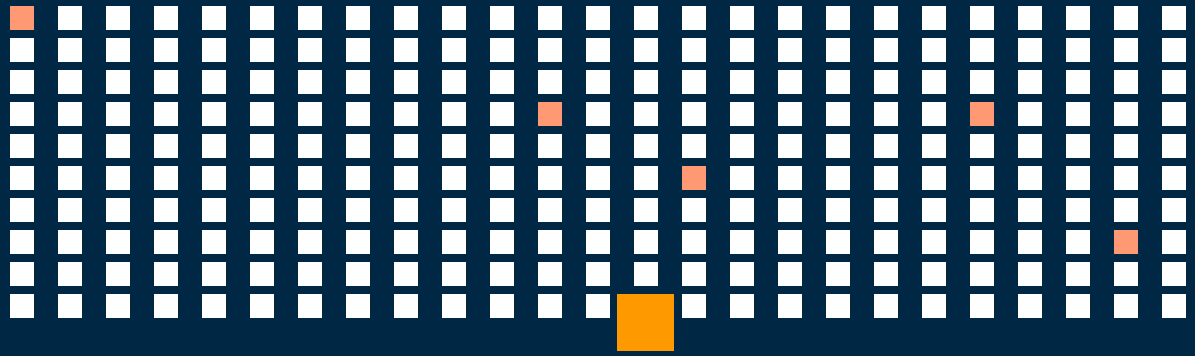
OUT OF THESE CANDIDATES

4-6 WILL BE INTERVIEWED



UNDERSTANDING THE PROBLEM

ONLY **1** WILL GET THE JOB OFFER



Problem Statement

01

PROBLEM STATEMENT

RECOMMENDATION SYSTEM

Matching skill sets of job posts and resumes

IMPROVE EFFICIENCY

Reducing time required to find relevant candidates

COST BENEFIT ANALYSIS

How does this benefit the company

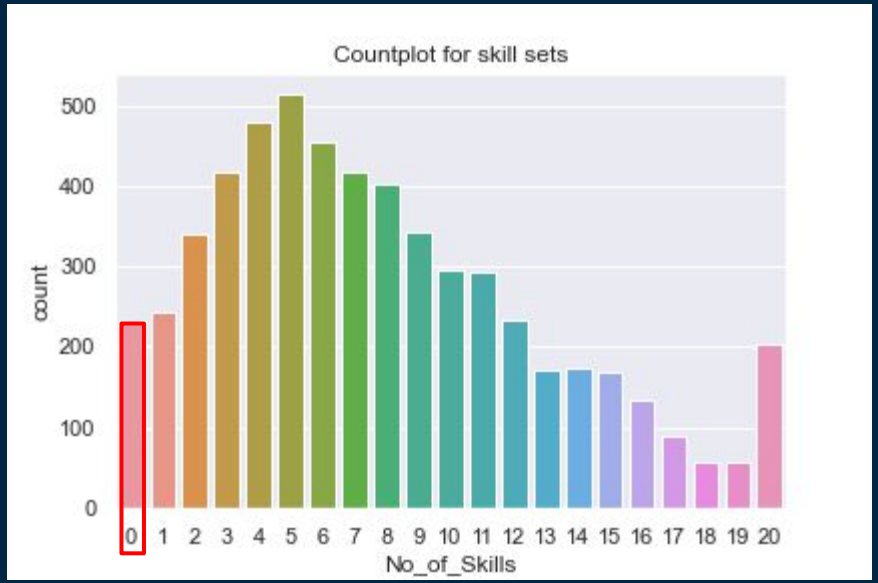
DATASETS

	Column 1	Column 2	Column 3
Job Post	Job Post ID	Job Title	Skills Requirement
Resume	Resume ID	Title applied	Resume Content

EXPLORATORY DATA ANALYSIS

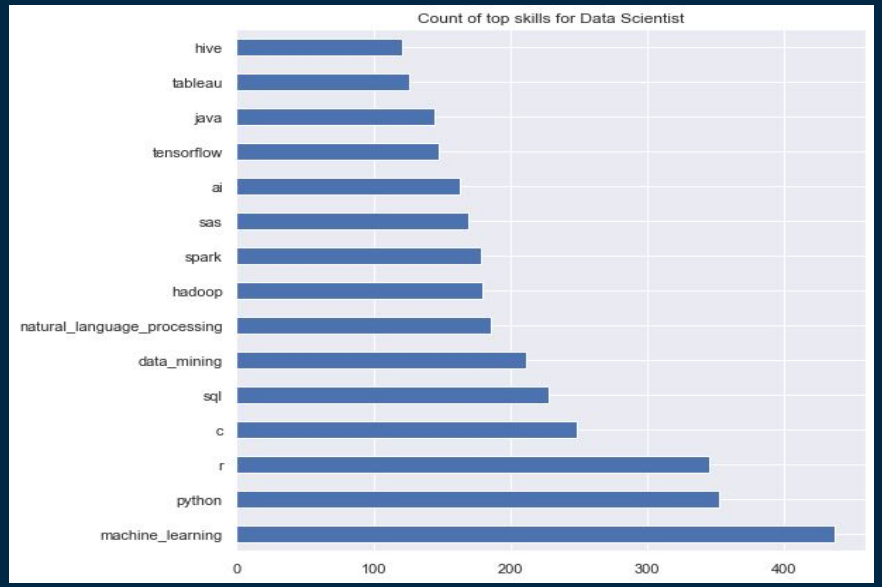
02

EXPLORATORY DATA ANALYSIS



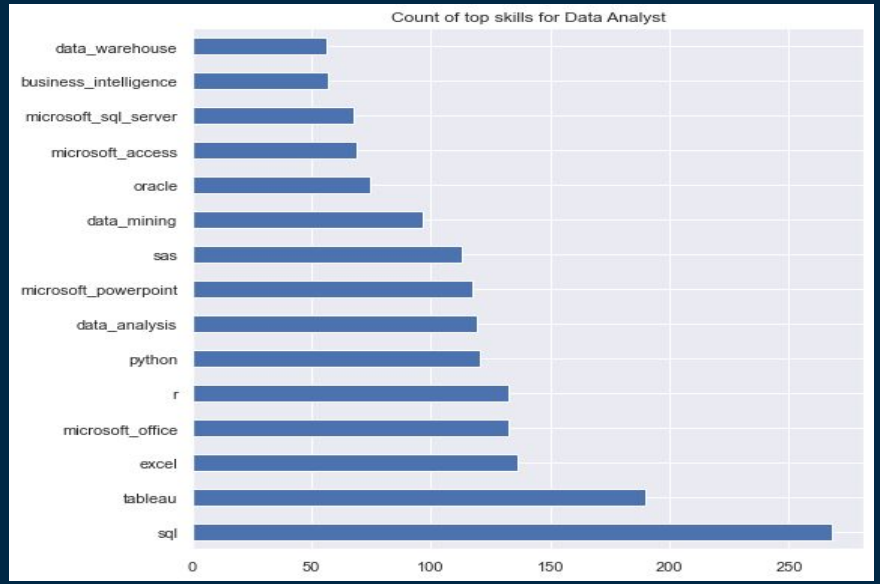
- There are over 200 job posting with no skill set mentioned
- Some of these companies uses a method called “structured interviewing”

EXPLORATORY DATA ANALYSIS



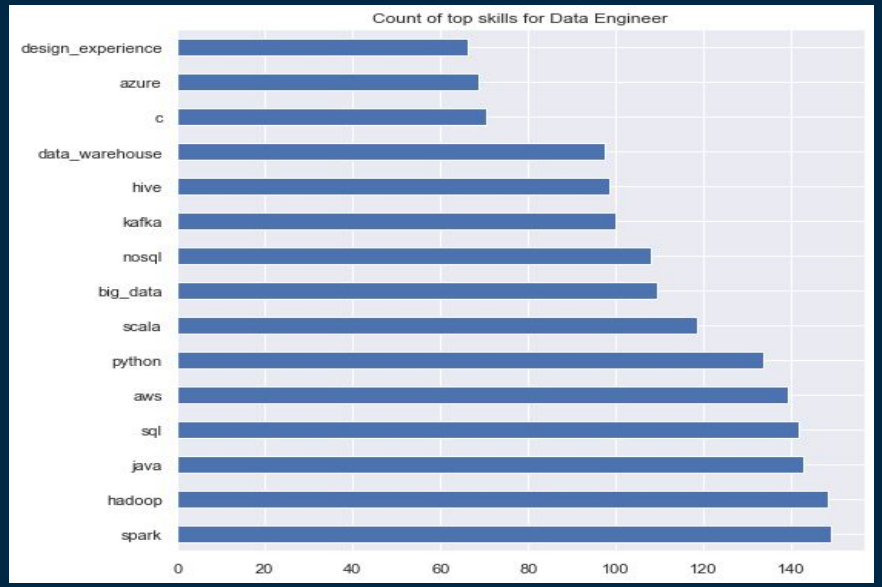
- Includes Machine Learning, python, r

EXPLORATORY DATA ANALYSIS



- Includes SQL, tableau, excel

EXPLORATORY DATA ANALYSIS



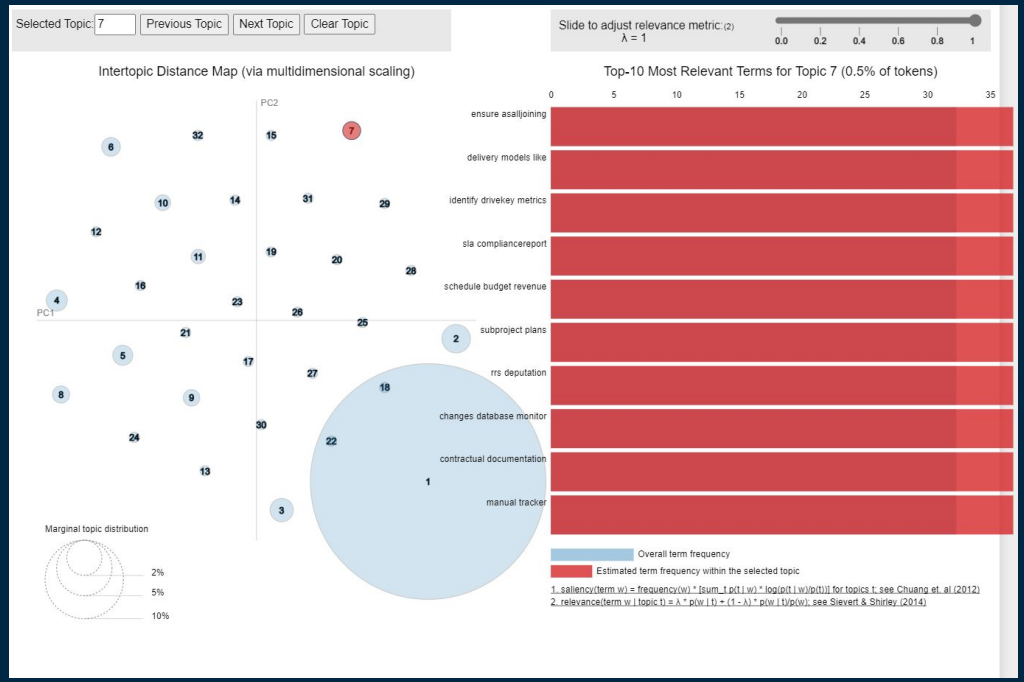
- Includes Spark, Hadoop and Java

EXPLORATORY DATA ANALYSIS

Topic modelling using Latent Dirichlet Allocation (LDA)

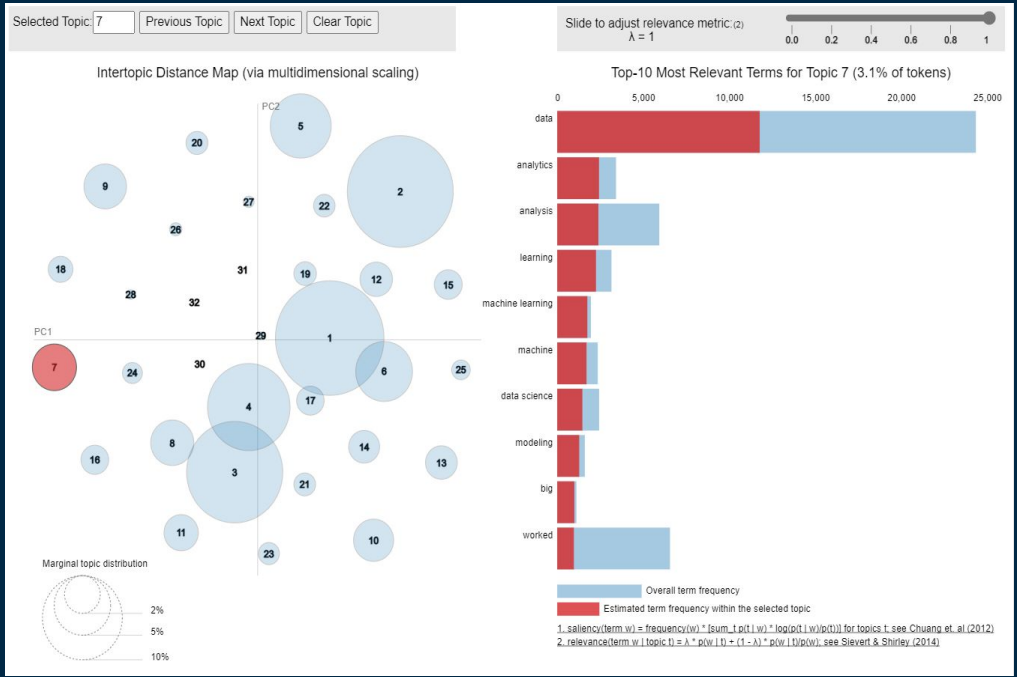
- Discover topics in a collection of documents
- Inform the algorithm how many topics you think there are
- The algorithm will assign every word to a temporary topic
- The algorithm will check and update topic assignments

EXPLORATORY DATA ANALYSIS



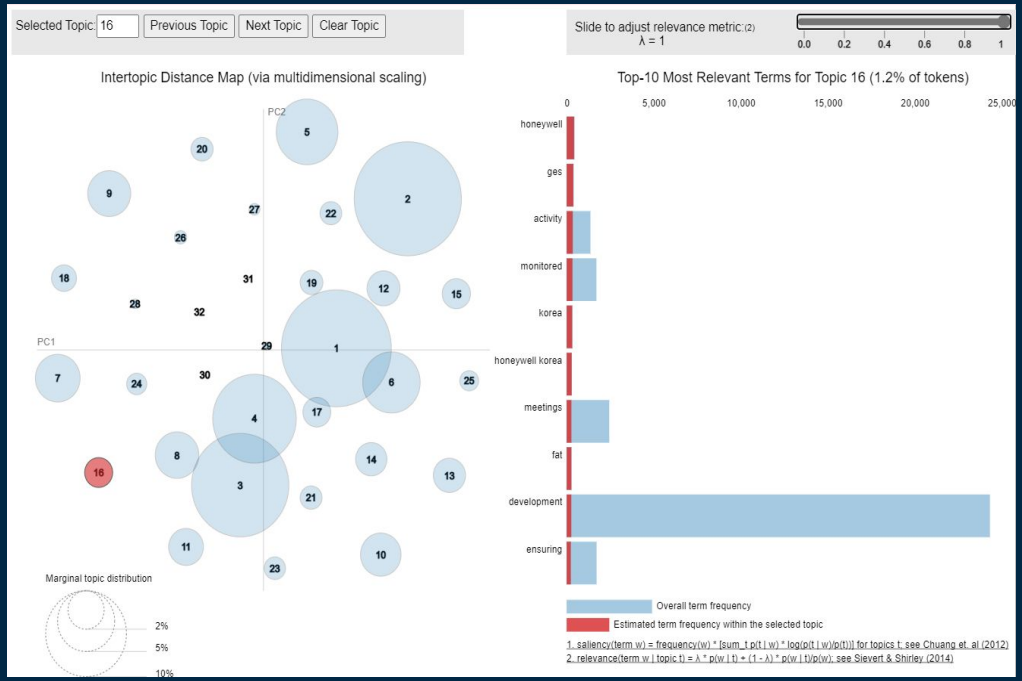
- Used TF-IDF Vectorizer
- Majority of the words are represented in topic 1
- Unable to identify a topic for topic 7

EXPLORATORY DATA ANALYSIS



- Used CountVectorizer
- Topic 7 for this is data scientist

EXPLORATORY DATA ANALYSIS

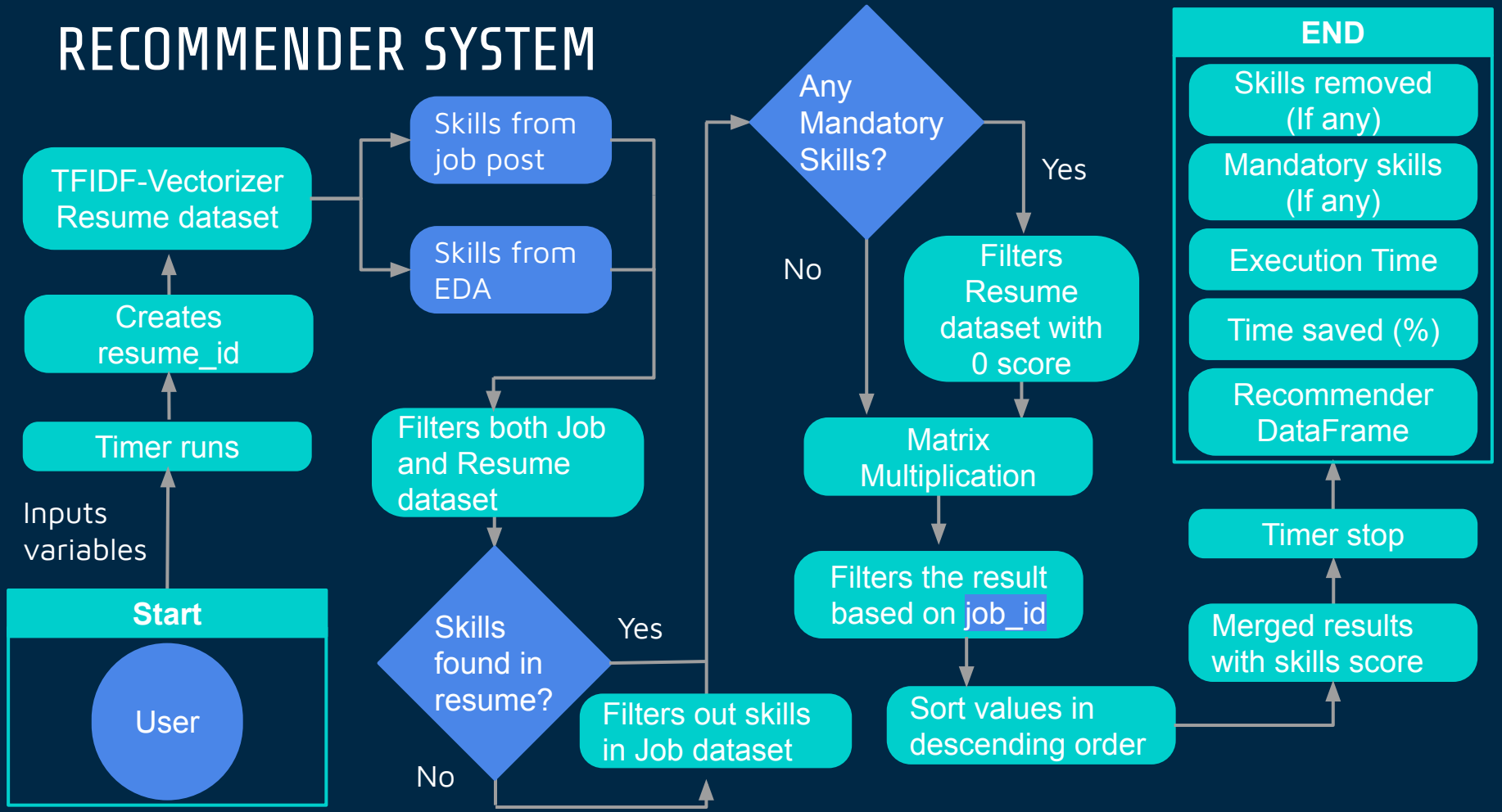


- Used CountVectorizer
- Still challenging to identify a topic for some of the bubbles

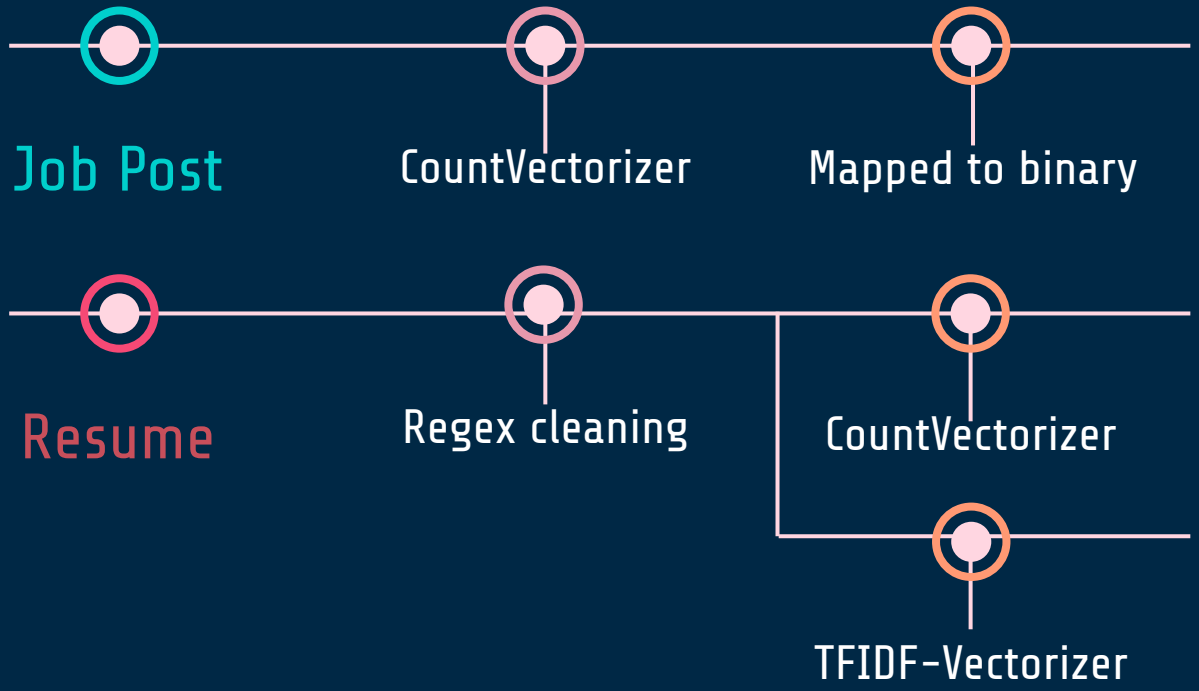
RECOMMENDER SYSTEM

03

RECOMMENDER SYSTEM



PREPROCESSING



RECOMMENDER SYSTEM

Resume_Job_id	data_scientist_2409	python	machine_learning	r	sql	hadoop	spark	data_mining	java	sas	natural_language_processing
Python Developer_1952	151	125	0	1	23	3	11	1	4	2	0
Database Administrator_2602	123	0	0	1	123	0	0	0	0	0	0
Python Developer_1931	120	106	3	0	9	2	7	0	10	0	0
Python Developer_1852	115	57	18	10	29	8	15	1	5	0	3
Python Developer_1898	110	76	6	1	24	2	3	2	1	3	2
Python Developer_1864	107	81	0	0	26	0	0	0	8	1	0
Python Developer_1818	107	82	3	0	18	3	7	1	17	0	1
Python Developer_1783	107	90	0	0	15	2	8	0	18	0	0
Python Developer_1836	106	79	0	0	22	4	1	0	11	0	1
Python Developer_1823	102	84	0	0	18	0	0	0	1	0	0

- Issue: keyword stuffing

RECOMMENDER SYSTEM

$$\mathrm{wf}_{t,d} = \begin{cases} 1 + \log \mathrm{tf}_{t,d} & \text{if } \mathrm{tf}_{t,d} > 0 \\ 0 & \text{otherwise} \end{cases}.$$

$$\text{wf-idf}_{t,d} = \mathrm{wf}_{t,d} \times \mathrm{idf}_t.$$

- Parameter: Sublinear_tf = True
- modifies the formula for Term frequency by assigning a weight

RECOMMENDER SYSTEM

Time taken to run: 8.67 seconds
Time saved: 85.55%
Top 10 applicants recommended for **data_scientist_2409**

Resume_Job_id	Total Score	python	machine_learning	r	sql	hadoop	spark	data_mining	java	sas	natural_language_processing
Python Developer_1886	0.3331	0.0868	0.1004	0.0647	0.0433	0.1025	0.1444	0.0000	0.0672	0.0	0.0000
Python Developer_1792	0.2849	0.0715	0.0790	0.0000	0.0463	0.0881	0.0637	0.0000	0.0529	0.0	0.0000
Python Developer_1794	0.2808	0.0710	0.0821	0.0000	0.0440	0.0837	0.0605	0.0000	0.0550	0.0	0.0000
Python Developer_1929	0.2323	0.0489	0.0950	0.0000	0.0000	0.0000	0.0000	0.0551	0.0000	0.0	0.0884
Data Scientist_9	0.2314	0.0816	0.1018	0.0000	0.0480	0.0000	0.0000	0.0000	0.0000	0.0	0.0000
Data Scientist_29	0.2314	0.0816	0.1018	0.0000	0.0480	0.0000	0.0000	0.0000	0.0000	0.0	0.0000
Data Scientist_39	0.2314	0.0816	0.1018	0.0000	0.0480	0.0000	0.0000	0.0000	0.0000	0.0	0.0000
Data Scientist_19	0.2314	0.0816	0.1018	0.0000	0.0480	0.0000	0.0000	0.0000	0.0000	0.0	0.0000
Python Developer_1920	0.2293	0.0641	0.0874	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0	0.0778
Data Scientist_2962	0.2232	0.0374	0.0471	0.0414	0.0334	0.0425	0.0181	0.0000	0.0381	0.0	0.0628

- Using Top Skills Identified during EDA
- Top recommended candidates may have been carried by other skills

$$timesave = \frac{(6seconds * NumOfCandidates) - (RuntimeOfFunction)}{6seconds * NumOfCandidates} * 100\%$$

RECOMMENDER SYSTEM

```
Required skills: ['machine_learning', 'natural_language_processing']
Time taken to run: 8.3 seconds
Time saved: 86.16%
Top 10 applicants recommended for data_scientist_2409
```

Resume_Job_id	Total Score	python	machine_learning	r	sql	hadoop	spark	data_mining	java	sas	natural_language_processing
Python Developer_1929	0.2323	0.0489	0.0950	0.0000	0.0000	0.0000	0.0000	0.0551	0.0000	0.0000	0.0884
Python Developer_1920	0.2293	0.0641	0.0874	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0778
Data Scientist_2962	0.2232	0.0374	0.0471	0.0414	0.0334	0.0425	0.0181	0.0000	0.0381	0.0000	0.0628
Data Scientist_2793	0.2196	0.0517	0.0706	0.0674	0.0326	0.0360	0.0366	0.0000	0.0000	0.0516	0.0287
Data Scientist_2834	0.2148	0.0371	0.0487	0.0484	0.0202	0.0520	0.0304	0.0440	0.0253	0.0654	0.0569
Data Scientist_2844	0.2117	0.0345	0.0403	0.0437	0.0236	0.0526	0.0500	0.0333	0.0000	0.0000	0.0607
Data Scientist_2937	0.2058	0.0288	0.0544	0.0427	0.0292	0.0513	0.0000	0.0458	0.0288	0.0000	0.0420
Data Scientist_2846	0.2028	0.0267	0.0576	0.0000	0.0000	0.0476	0.0555	0.0425	0.0267	0.0000	0.0709
Data Scientist_2909	0.1954	0.0504	0.0613	0.0540	0.0291	0.0232	0.0236	0.0714	0.0000	0.0751	0.0314
Python Developer_1835	0.1924	0.0530	0.0307	0.0378	0.0286	0.0527	0.0348	0.0357	0.0000	0.0000	0.0273

- Using Top Skills Identified during EDA and established required skills from the list
- Better recommendation

RECOMMENDER SYSTEM

```
The following skill(s) have been removed as they are unavailable in our database: ['teaching_experience']
Time taken to run: 8.91 seconds
Time saved: 85.14%
Top 10 applicants recommended for data_scientist_2409
```

Resume_Job_id	Total Score	data_science	hadoop	hive	machine_learning	natural_language_processing	python	sql
Python Developer_1886	0.3808	0.0000	0.1025	0.0478	0.1004	0.000	0.0868	0.0433
Data Scientist_40	0.3557	0.1191	0.0602	0.0395	0.0760	0.000	0.0609	0.0000
Data Scientist_20	0.3557	0.1191	0.0602	0.0395	0.0760	0.000	0.0609	0.0000
Data Scientist_10	0.3557	0.1191	0.0602	0.0395	0.0760	0.000	0.0609	0.0000
Data Scientist_30	0.3557	0.1191	0.0602	0.0395	0.0760	0.000	0.0609	0.0000
Python Developer_1792	0.3545	0.0000	0.0881	0.0696	0.0790	0.000	0.0715	0.0463
Python Developer_1794	0.3469	0.0000	0.0837	0.0661	0.0821	0.000	0.0710	0.0440
Data Scientist_2945	0.3469	0.0763	0.0588	0.0922	0.0527	0.000	0.0386	0.0283
Data Scientist_2937	0.3162	0.0616	0.0513	0.0488	0.0544	0.042	0.0288	0.0292
Data Scientist_2870	0.3141	0.0496	0.0665	0.0702	0.0483	0.000	0.0474	0.0320

- Using Skills mentioned in the job requirements

RECOMMENDER SYSTEM

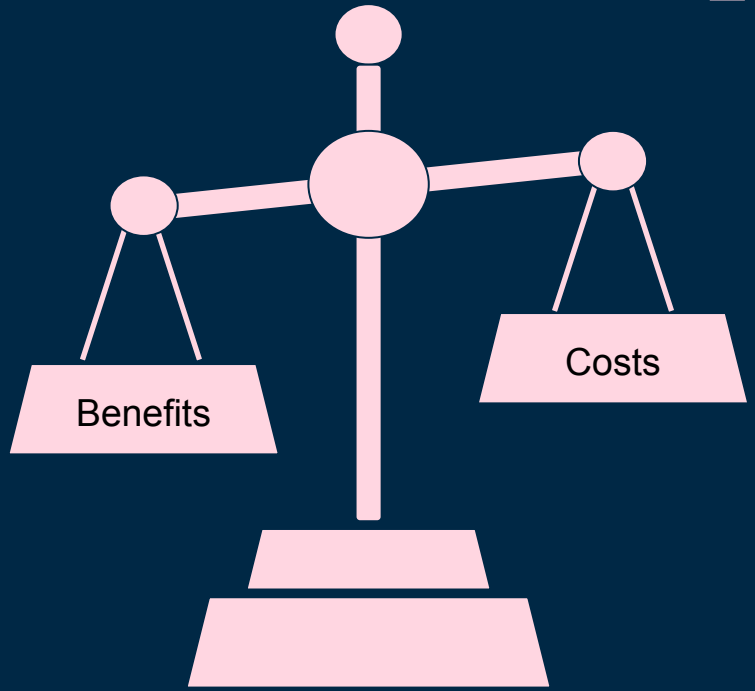
```
The following skill(s) have been removed as they are unavailable in our database: ['teaching_experience']
Required skills: ['natural_language_processing']
Time taken to run: 8.78 seconds
Time saved: 85.37%
Top 10 applicants recommended for data_scientist_2409
```

Resume_Job_id	Total Score	data_science	hadoop	hive	machine_learning	natural_language_processing	python	sql
Data Scientist_2937	0.3162	0.0616	0.0513	0.0488	0.0544	0.0420	0.0288	0.0292
Data Scientist_2844	0.2981	0.0364	0.0526	0.0500	0.0403	0.0607	0.0345	0.0236
Data Scientist_2962	0.2889	0.0458	0.0425	0.0198	0.0471	0.0628	0.0374	0.0334
Data Scientist_2793	0.2798	0.0602	0.0360	0.0000	0.0706	0.0287	0.0517	0.0326
Data Scientist_2834	0.2798	0.0649	0.0520	0.0000	0.0487	0.0569	0.0371	0.0202
Python Developer_1868	0.2687	0.0422	0.0252	0.0432	0.0426	0.0480	0.0400	0.0274
Data Scientist_2846	0.2630	0.0602	0.0476	0.0000	0.0576	0.0709	0.0267	0.0000
Python Developer_1934	0.2608	0.0287	0.0418	0.0494	0.0434	0.0370	0.0358	0.0248
Data Scientist_2928	0.2569	0.0481	0.0000	0.0496	0.0536	0.0659	0.0208	0.0189
Data Scientist_2905	0.2561	0.0790	0.0000	0.0000	0.0721	0.0595	0.0455	0.0000

- Using Skills mentioned in the job requirements and established required skills from the list

COST BENEFIT ANALYSIS

- Reduce the time to hire
- Improves quality of hire



CONCLUSION

- Able to reduce hiring time
- Soft skills are out of reach
- Users have to understand what does the role requires

RECOMMENDATION

- Improve with user feedback
- A/B testing on the skill set option or on different industry
- Retrieve more information to further specialize the recommendation

Do you have any questions?

Leeyunchang03@gmail.com

THANKS

REFERENCE LIST

- Slide 3 : Corporate.io
- Slide 4-6: Corporate.io & Glassdoor.com
- Slide 15: https://nlp.stanford.edu/IR-book/html/htmledition/sublinear-tf-scaling-1.html
- Slide 20: https://www.smallbizdaily.com/importance-of-ats-to-improve-time-cost-quality-hiring/ & https://www.skeeled.com/blog/applicant-tracking-system-main-advantages-of-using-an-ats