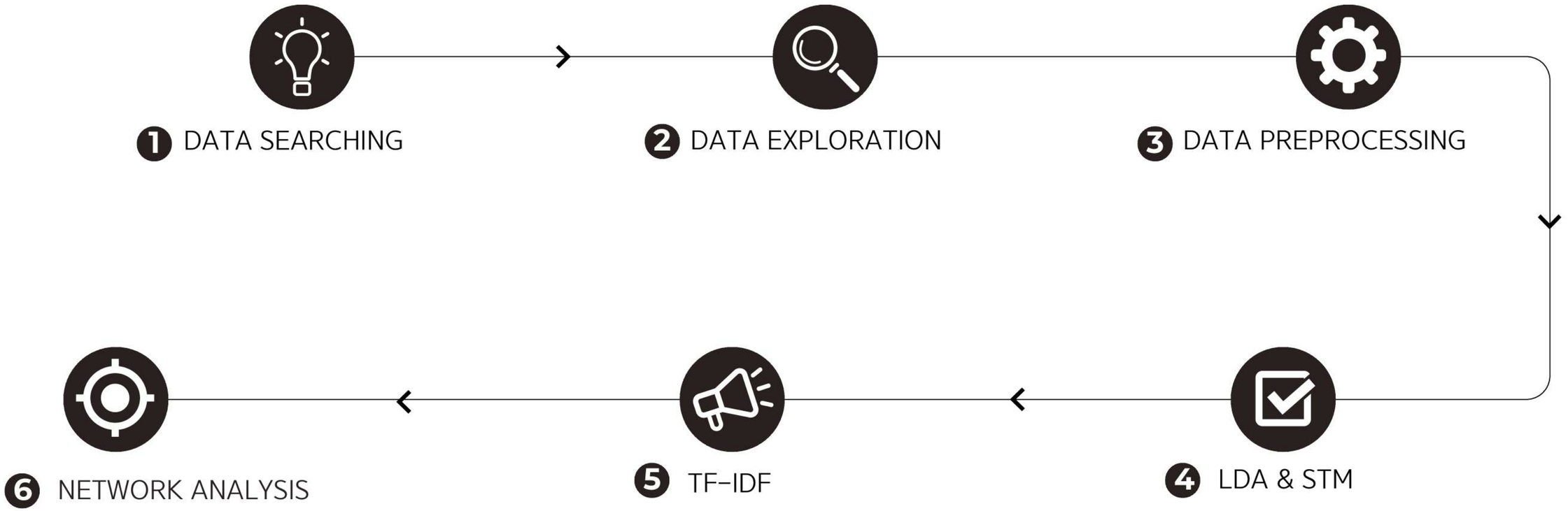
Beyond the Questions

Analyzing Interview Data for Strategic Preparation

Team 5

21900707 YOUNGWOO CHO
22200022 JIHO KANG
22200130 YEWON KIM


1 DATA SEARCHING
2 DATA EXPLORATION
3 DATA PREPROCESSING
4 LDA & STM
5 TF-IDF
6 NETWORK ANALYSIS

Data Source

AI Hub

- Date: 2022
- Type: json
- Number of data: 67,307
- Institution name
  : ㈜사람과숲, ㈜넥스인테크놀로지
- Description: Convert voice to text of job interview questions and answers and summarize the contents

Data Exploration

- **Data Loading:**
  - Select Question, Gender, Experience data in json
  - Dropping 202 data due to lexical error

```
Error occurred: lexical error: invalid character inside string.
          익이란 기업의 단기 순                    서 주주
                      (right here) ------^
```

- **Check Outliers:**
  - No outliers in data

Question	Category	Gender	Experience
0	0	0	0

- **Check Data Count Distribution:**

Design	ICT	Management	PM	PublicService	RND	SalesMarketing
6469	5733	14950	6797	16745	4656	11755

Female	Male
40746	26359

Experienced	New
8991	58114

Risk of result bias due to data imbalance

Data Preprocessing

- **Recheck Data Count Distribution:**

Design	ICT	Management	PM	PublicService	RND	SalesMarketing
1341	1644	2038	1707	1843	1383	1959

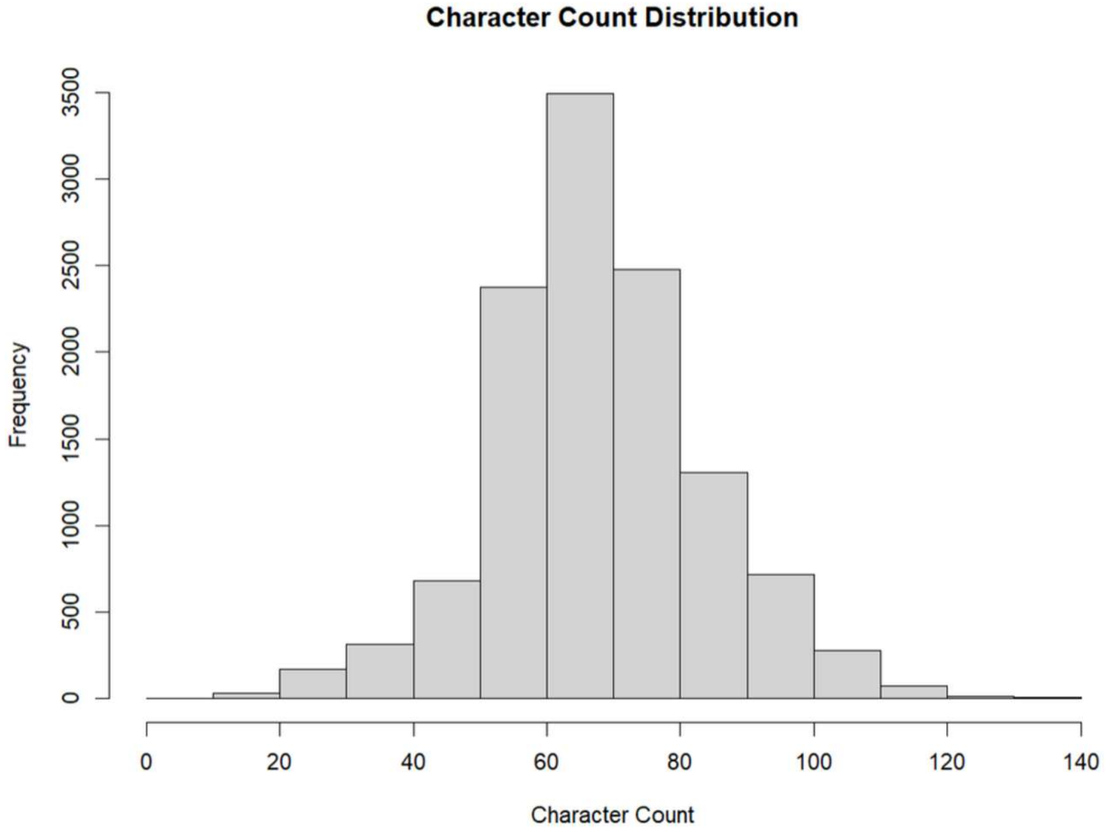
Female	Male
5784	6131

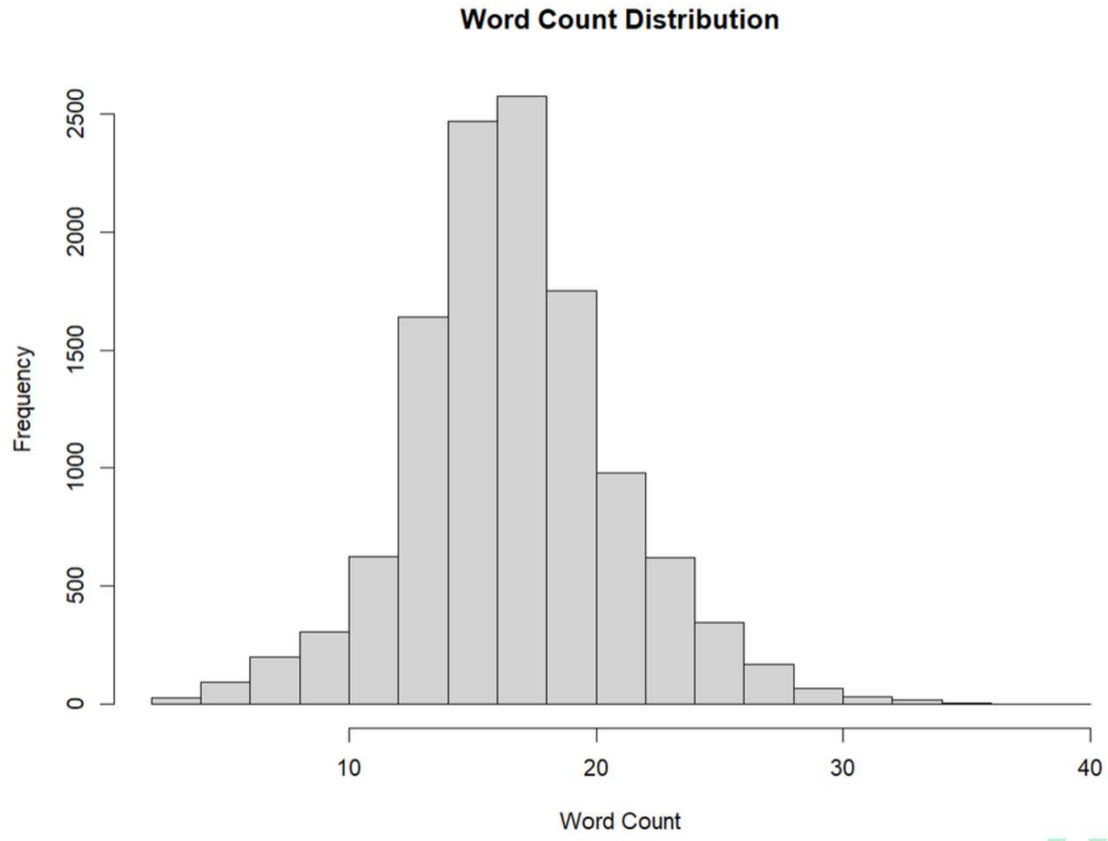
Experienced	New
4897	7018

Resolving data imbalance problem
by considering all data combination

Each combination is around 500

- **Count character and word in each question**





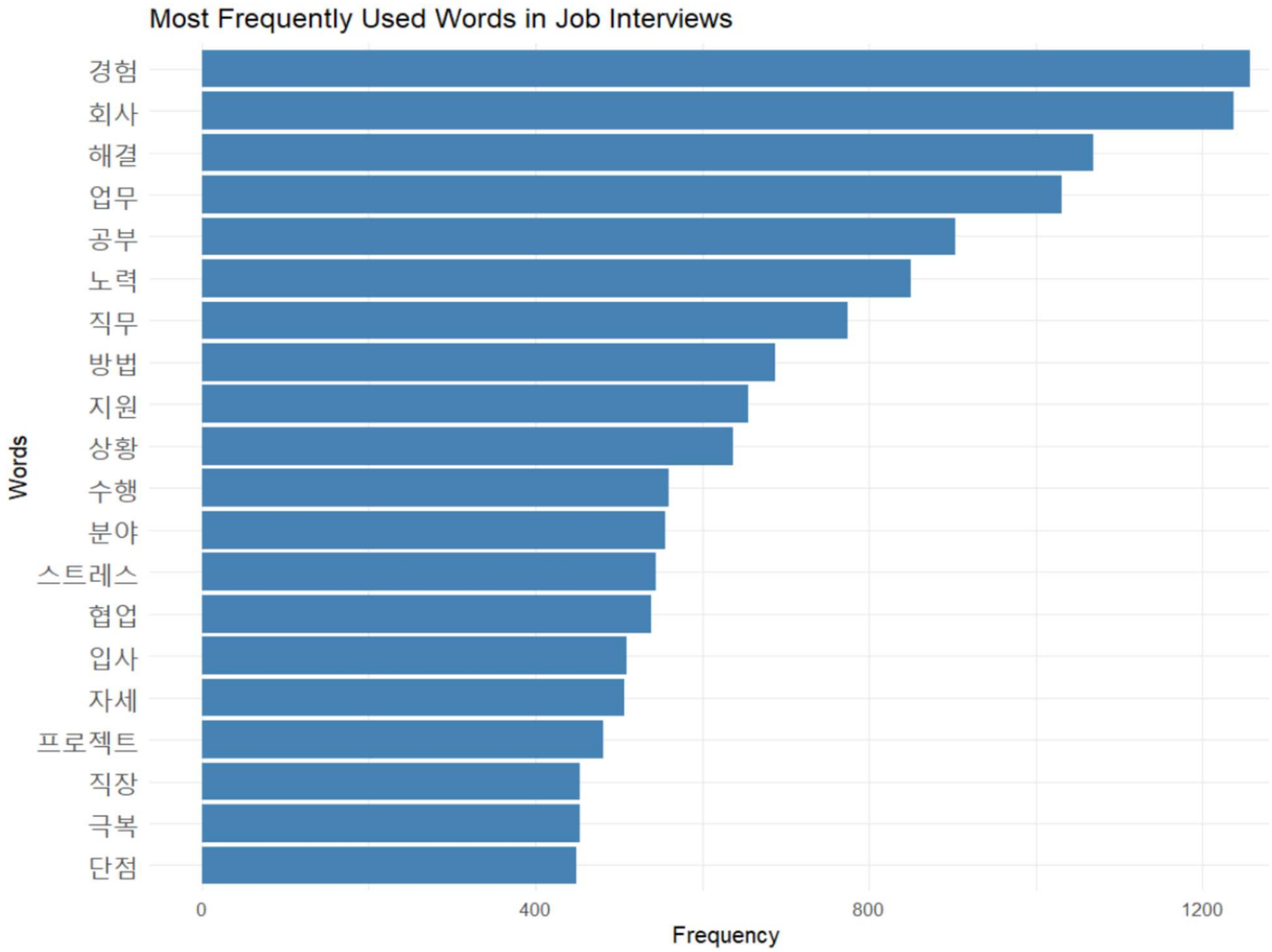
- **Removing Stopwords**

단어	빈도
해	8039
지원자	5905
말씀	5733
님	5522
생각	3925
것	3833
일	3400
한	2965
무엇	2829
본인	2795

Delete one letter

단어	빈도
지원자	5981
말씀	5800
생각	4032
무엇	2882
본인	2841
이유	2358
설명	2200
궁금	1699
하시	1639
그것	1284

stopwords.txt, additional words (면접자, 지원자, 말씀, 하시...)


Most Frequently Used Words in Job Interviews
경험
회사
해결
업무
공부
노력
직무
방법
지원
상황
수행
분야
스트레스
협업
입사
자세
프로젝트
직장
극복
단점
Words
0
400
800
1200
Frequency

Data Analysis

Developing Interview Preparation Strategies through Question Analysis

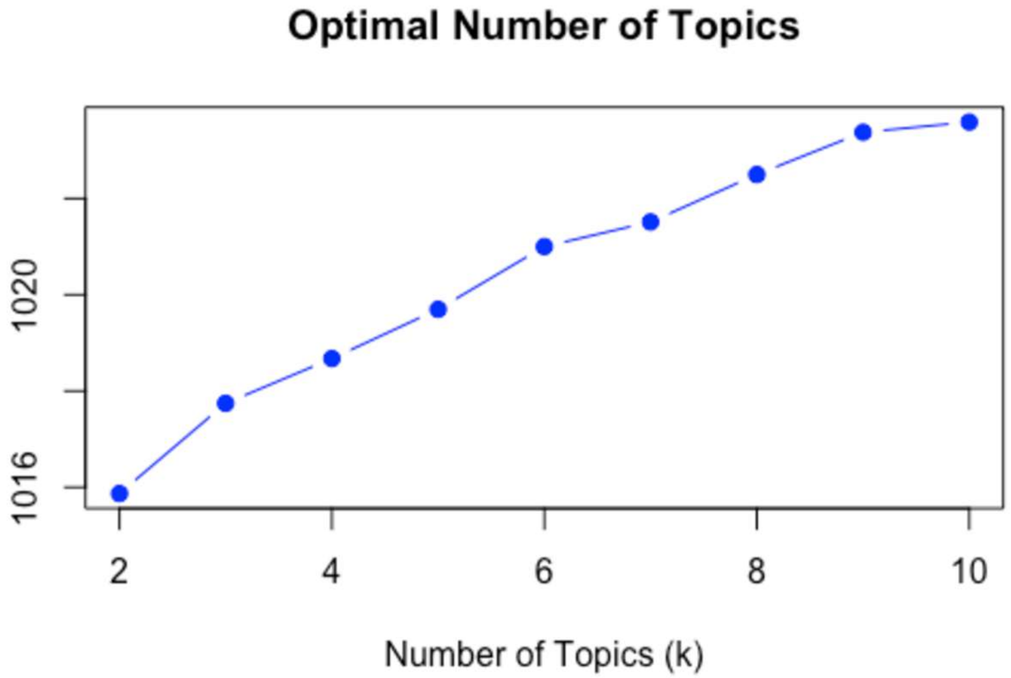
Uncovering *Hidden Topics* and *Evaluation Criteria* with LDA and STM


LDA and STM
FOR IN-DEPTH ANALYSIS

✓ DTM

✓ K = 2 ?

Perplexity
k = 2 ~ 10



K=2	k=3	k=4
topic 1 **problem-solving.** 'solution', 'explain' **topic 2** **conversational questions.** 'say', 'nowadays', 'curious'	**topic 1** **problem-solving and career related.** 'solution', 'experience' **topic 2** **observational.** 'topic', 'curious' **topic 3** **motivation / enthusiasm.** 'challenge', 'wish', 'passion'	**topic 1** **problem- solving** 'solution' 'support', 'problem' **topic 2** **conversational** 'nowadays', 'say' **topic 3** **career and background** 'experience', 'resume' **topic 4** **attitudes and aspirations** 'passion', 'challenge', 'wish'

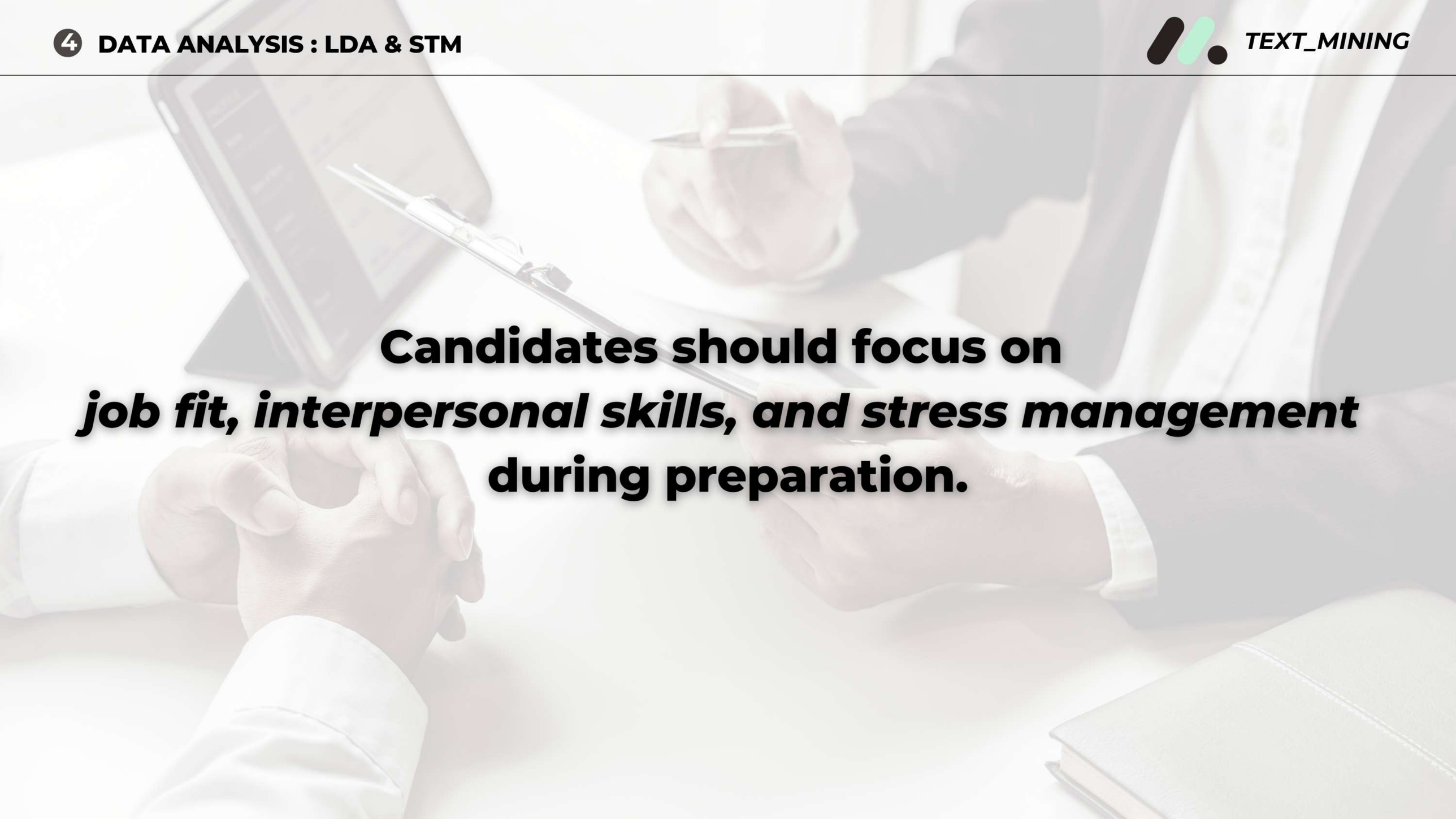
K = 3 ,

- **Topic 1 : <u>Problem solving and career related.</u>**
  - ‘solution’ , ‘experience’.

  Deep understanding of job roles.
- **Topic 2 : <u>Observational.</u>**
  - ‘topic’, ‘curious’.

  Coping with challenges and interpersonal skills.
- **Topic 3 : <u>Motivation and Enthusiams.</u>**
  - ‘challenge’, ‘wish’, ‘passion’.

  Stress management and achieving goals.



Candidates should focus on
job fit, interpersonal skills, and stress management
during preparation.

THE LDA
***LIMITATION** OF*
EXTRACTING TOPICS
BASED SOLELY ON
***CO-OCCURRENCE** IS*
SUPPLEMENTED
WITH STM.

LDA and STM

FOR IN-DEPTH ANALYSIS

Enables deeper analysis of
question evaluation patterns by including metadata.

✓ **Used metadata (Category, Experience, Gender)**

✓ **Tested multiple k values (3~10) and selected 5.**

K = 3 , 4 (normal)
K > 5 (too complicated)

STM K=5 RESULTS

- **Topic 1 : Job fit and company understanding.**
- **Topic 2 : Values and problem-solving skills.**
- **Topic 3 : Self-motivation and goal setting.**
- **Topic 4 : Collaboration skills.**
- **Topic 5 : Strengths, weaknesses, and improvement plans.**

Document Proportion Analysis :

Topic 1 has the ***highest proportion***

→ **Key** evaluation focus.

Topic 2 has a lower proportion

→ Secondary evaluation focus.

LDA and STM's Conclusion

- Candidates should **focus on key areas** such as job fit, company understanding, and personal values.
- Comprehensive preparation for **Topic1-related questions is critical.**

Future Applications

- Develop customized **question banks** by job role.
- **Create interview preparation guidelines.**
- Improve company recruitment evaluation processes.

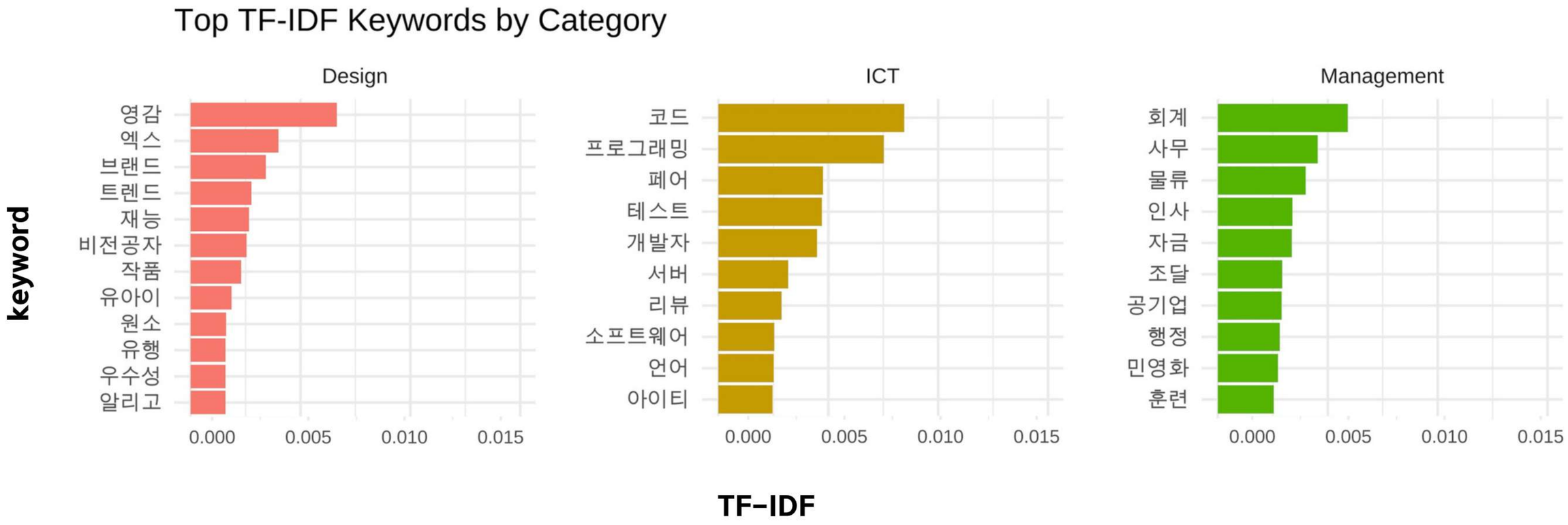
TF-IDF

- *KEYWORD ANALYSIS BY JOB CATEGORY*
- *KEYWORD ANALYSIS BY GENDER*
- *KEYWORD ANALYSIS BY EXPERIENCED*

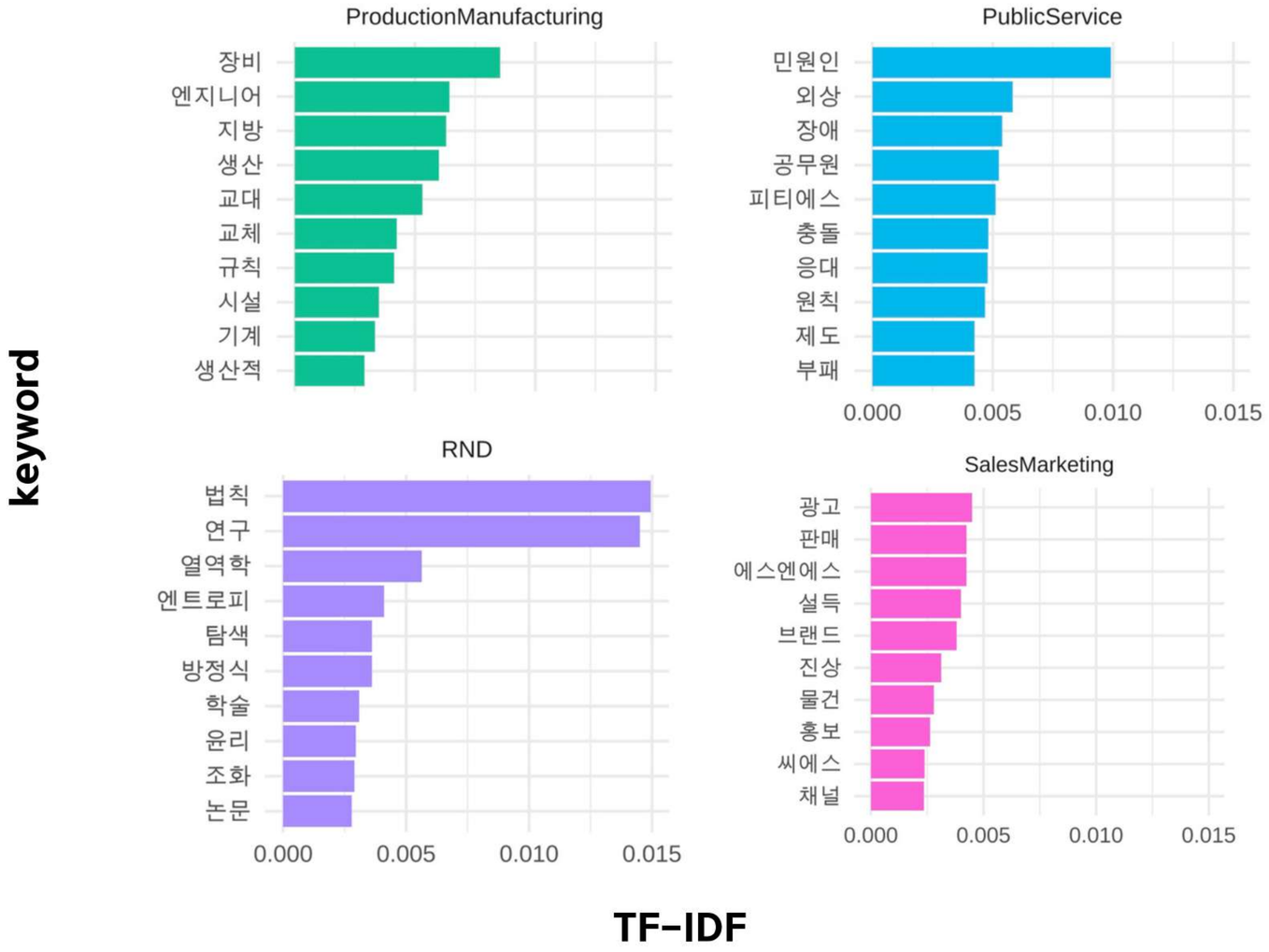
TF-IDF STEP

- **WORD FREQUENCY CALCULATION**
- **DELETE ADDITIONAL STOPWORD**
- **CALCULATE TF-IDF**
- **VISUALIZATION: TOP 10 KEYWORDS BY TF-IDF**

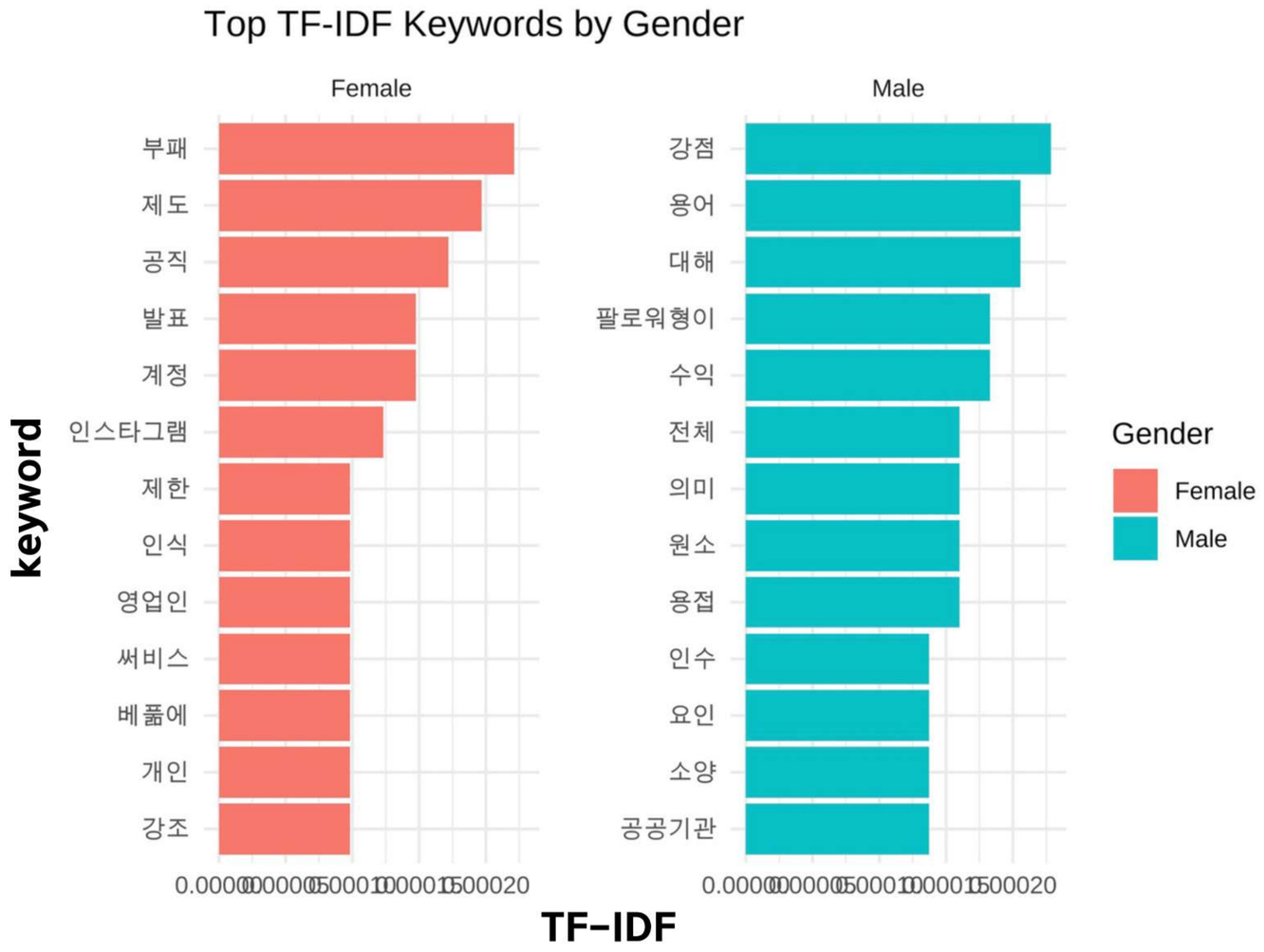
KEYWORD ANALYSIS BY JOB CATEGORY



KEYWORD ANALYSIS BY JOB CATEGORY



KEYWORD ANALYSIS BY GENDER

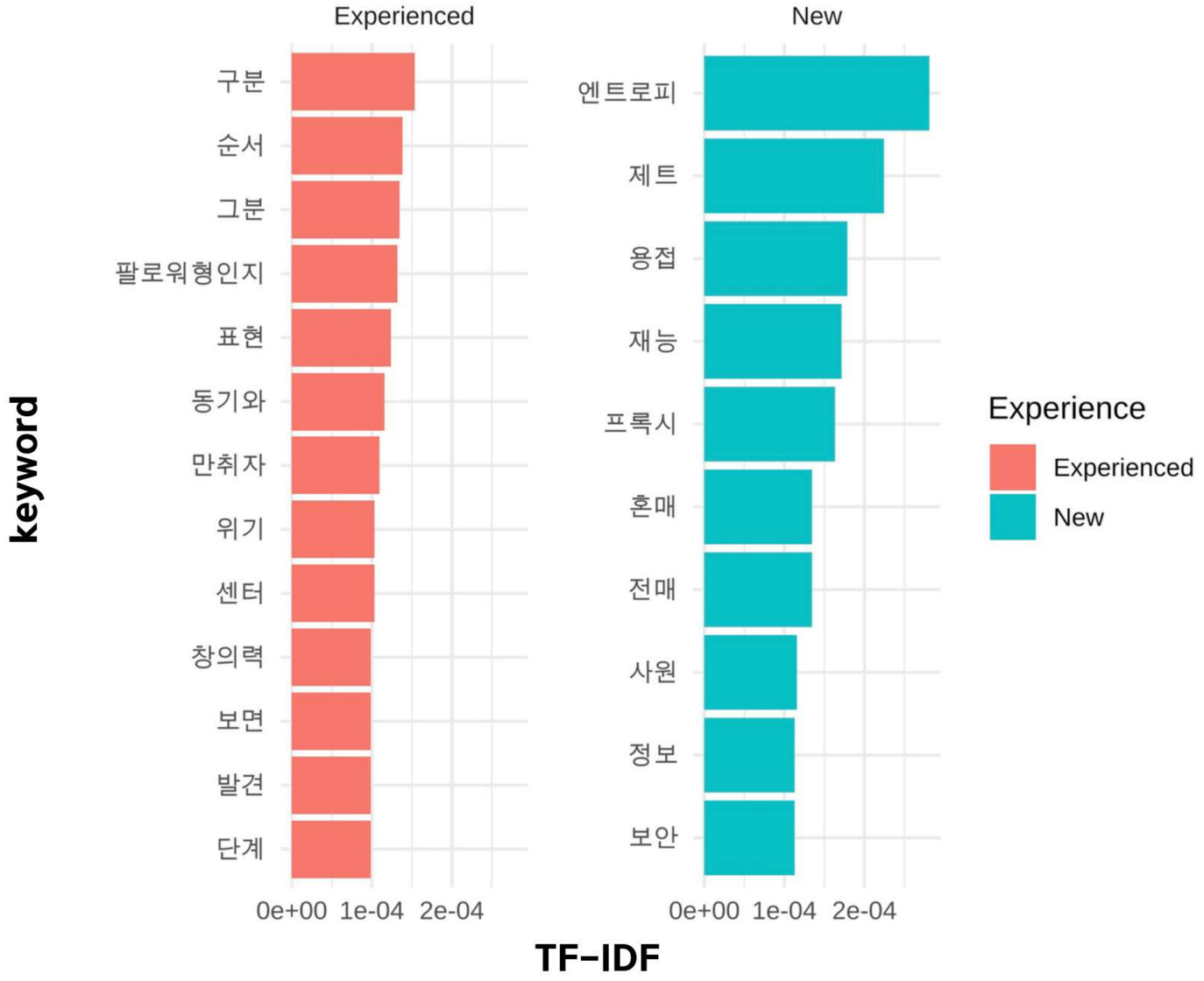


- Female candidates can prepare to emphasize social responsibility, interpersonal skills, and digital media expertise.

- Male candidates should focus on showcasing technical expertise, economic achievements, and practical problem-solving abilities.

KEYWORD ANALYSIS BY EXPERIENCED

- Experienced candidates should highlight examples of **problem-solving and strengths** demonstrated in **real-world scenarios**.
- Inexperienced candidates should showcase **adaptability and learning potential** through relevant examples.



Network Analysis

NETWORK ANALYSIS STEP

Step 1: Create Edge List

- Iterating Over Pairs of Sentences
- Building Edges

Step 2: Process Data for Each Category

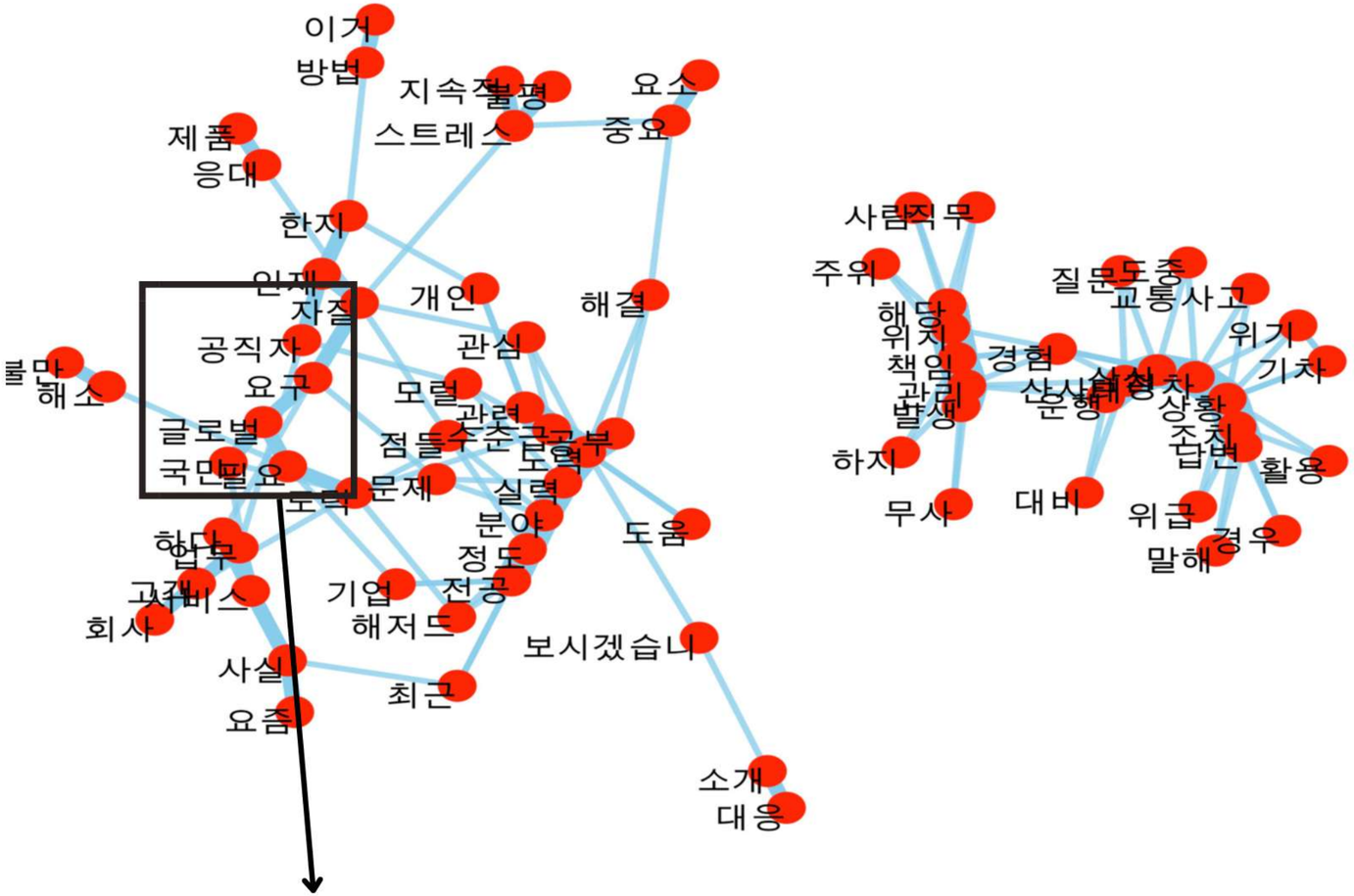
- Filtering Data
- Tokenizing Text
- Selecting Word Lists

Step 3: Build Networks

Step 4: Visualize Networks

- Edges
- Nodes
- Labels

NETWORK ANALYSIS PUBLIC SERVICE



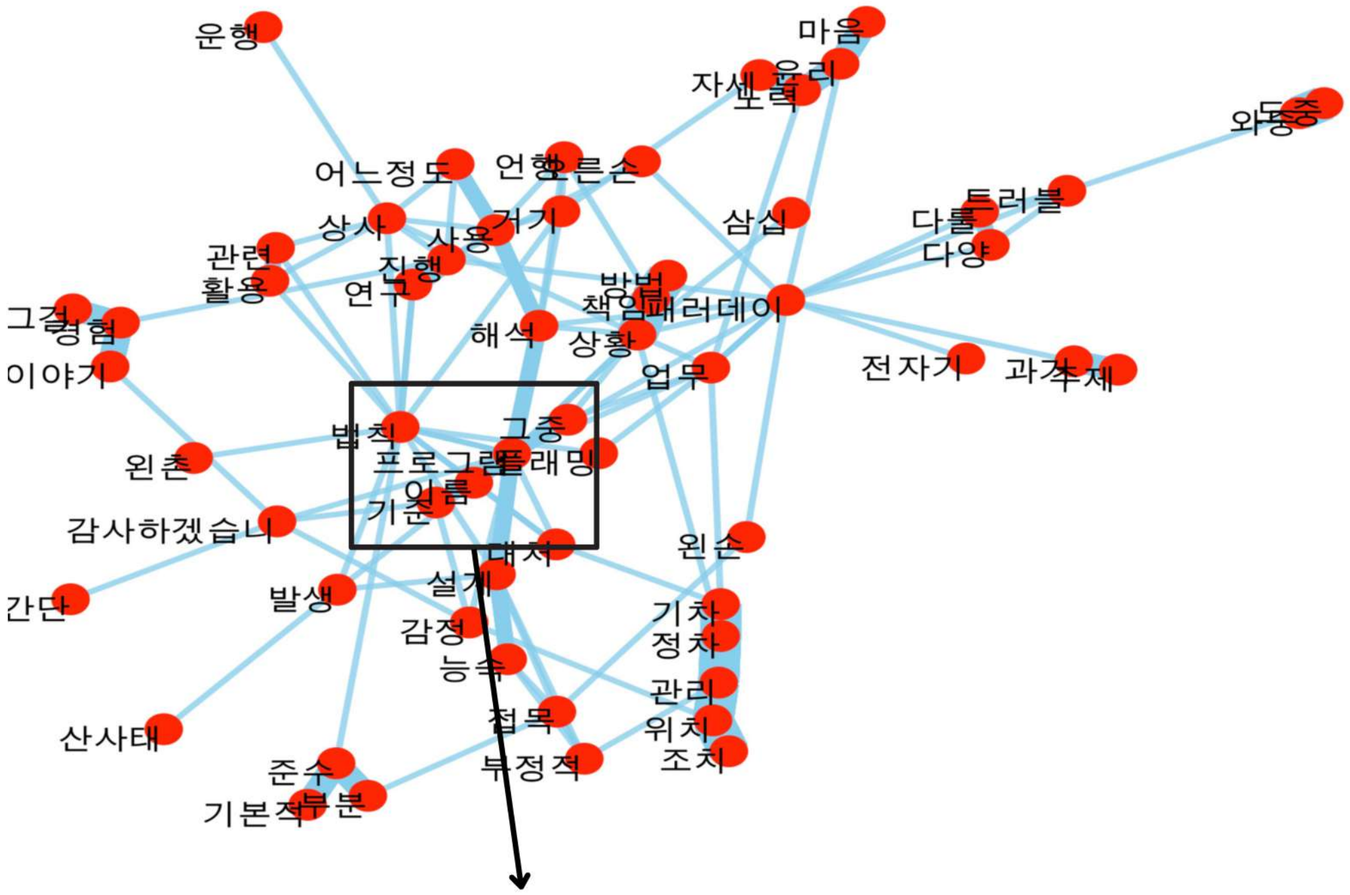
{Public, Need, Global ,Citizen}

EXPECTED QUESTION

- "How would you balance global standards with citize needs in public service? Can you explain how you would apply international practices while considering local context?"

->**Topic1 : Job fit and company understanding**

NETWORK ANALYSIS □RND



{□Principle, Fleming}

EXPECTED QUESTION

- "How have you applied fundamental principles such as Faraday's Law or Fleming's Laws in your work? Could you share an example where these principles were practically applied in your projects?"

->**Topic 2 : Values and problem-solving skills**

- By using various analysis approaches, you can find out important interview keywords for each job type and create expected questions based on these.
- While we haven’t reached the stage of generating interview questions, we were able to discover the possibility of connecting with a model that generates questions using keywords.
- This will help simulate real interview situations more realistically in the future.

THANK YOU