# Topic Modeling for Job Matching

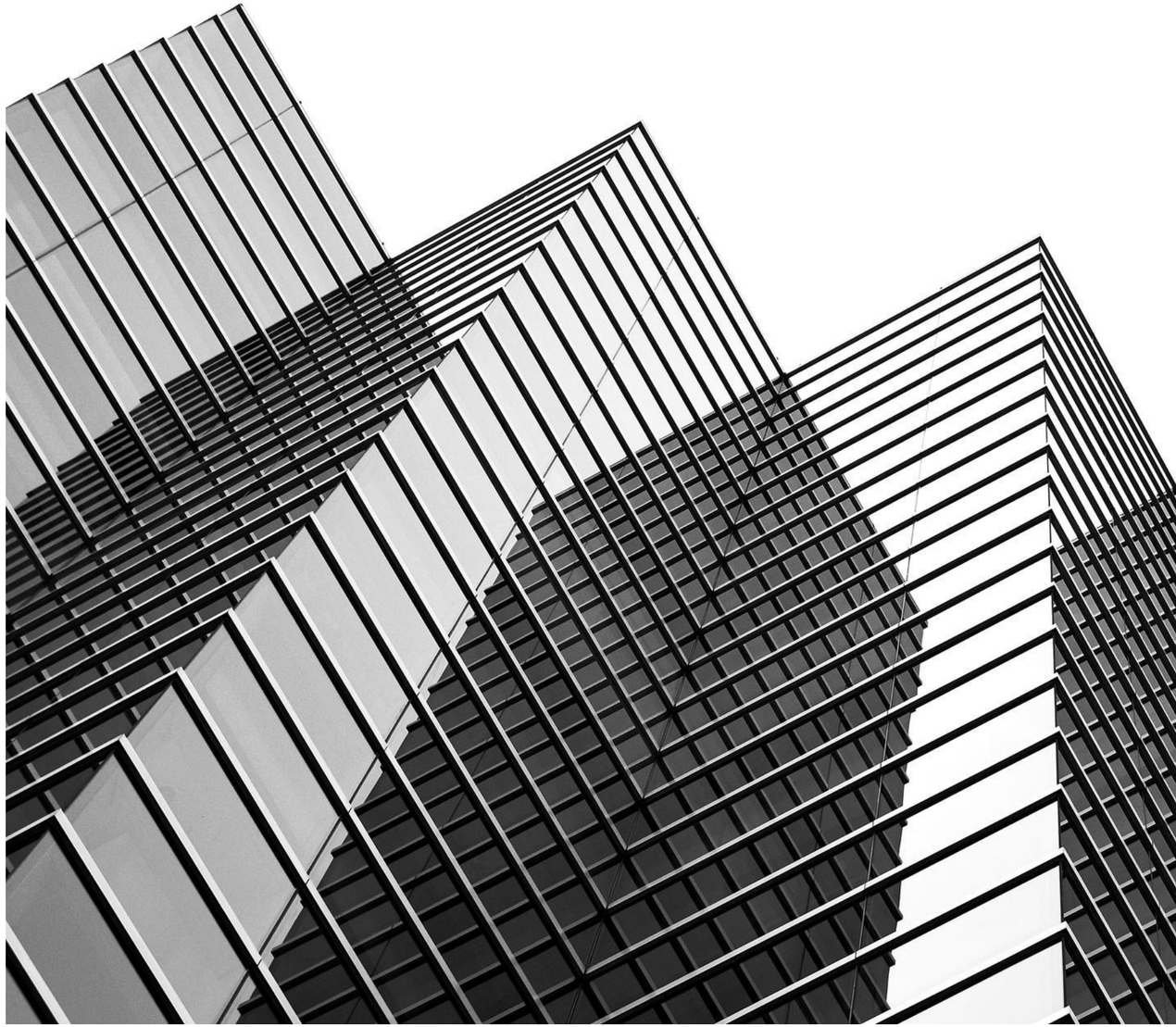
Pang Hong Xiang

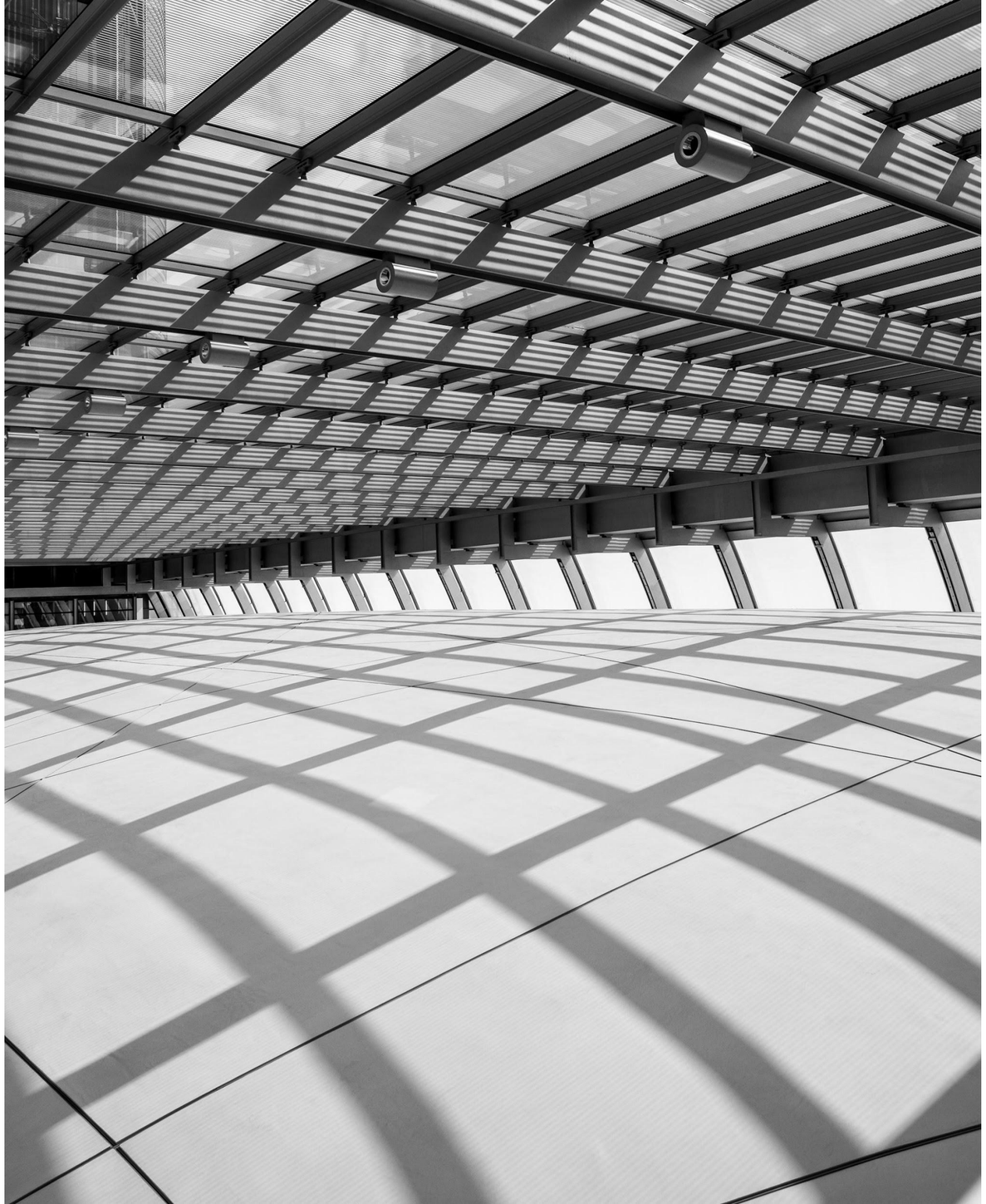
# Problem Statement

Workforce Singapore manages a job matching portal (MyCareerFuture), which serves employers and jobseekers alike.

Employers complain that the applicants on the portal are not qualified for the jobs they are posting for.

However, they are not able to justify what they mean by “qualified”.



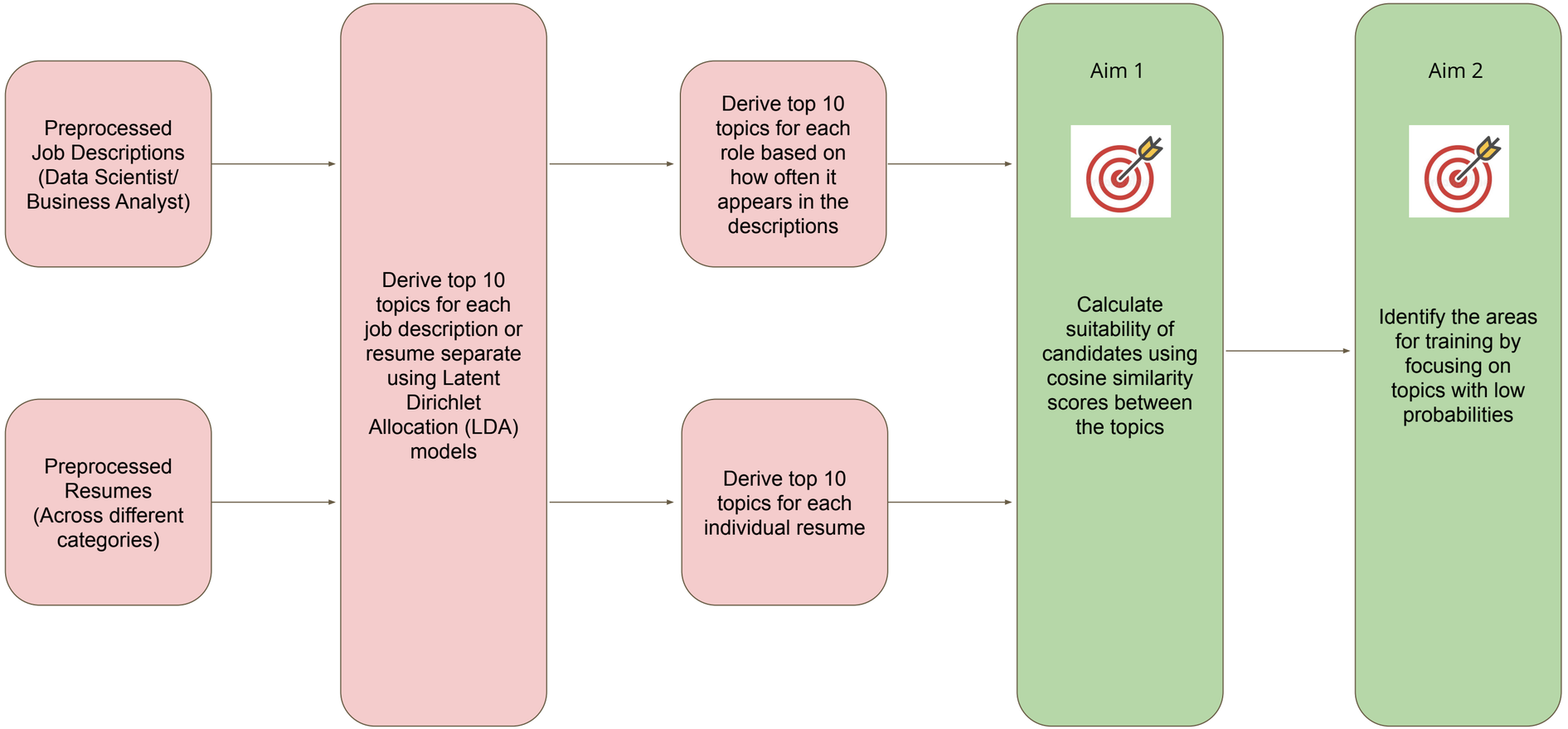


# Aims of Project (Proof-of-Concept)

Quantify the suitability of jobseekers to recommend the most “qualified” candidates to employers.

Identify the gaps in quality of the candidates to recommend employers to focus their efforts on training post-hiring.

# Methodology



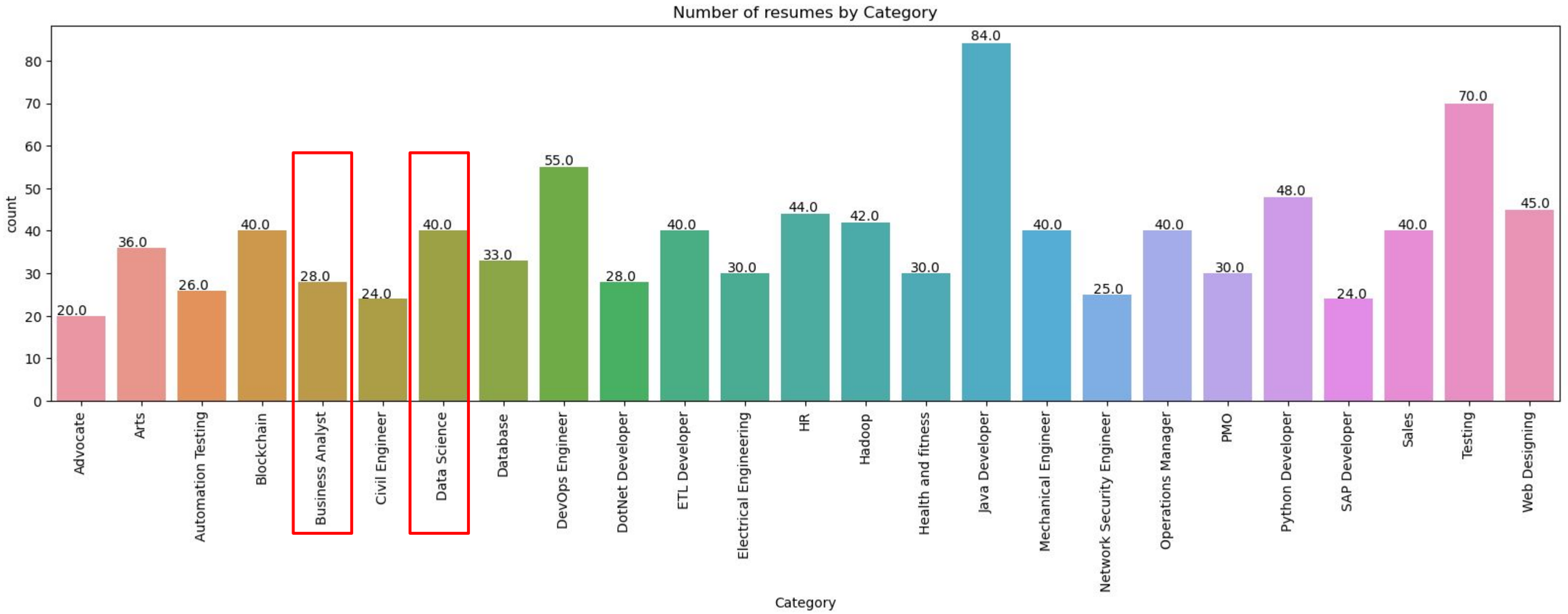
# Datasets

960 resumes across 25 categories (https://www.kaggle.com/datasets/gauravduttakiit/resume-dataset)

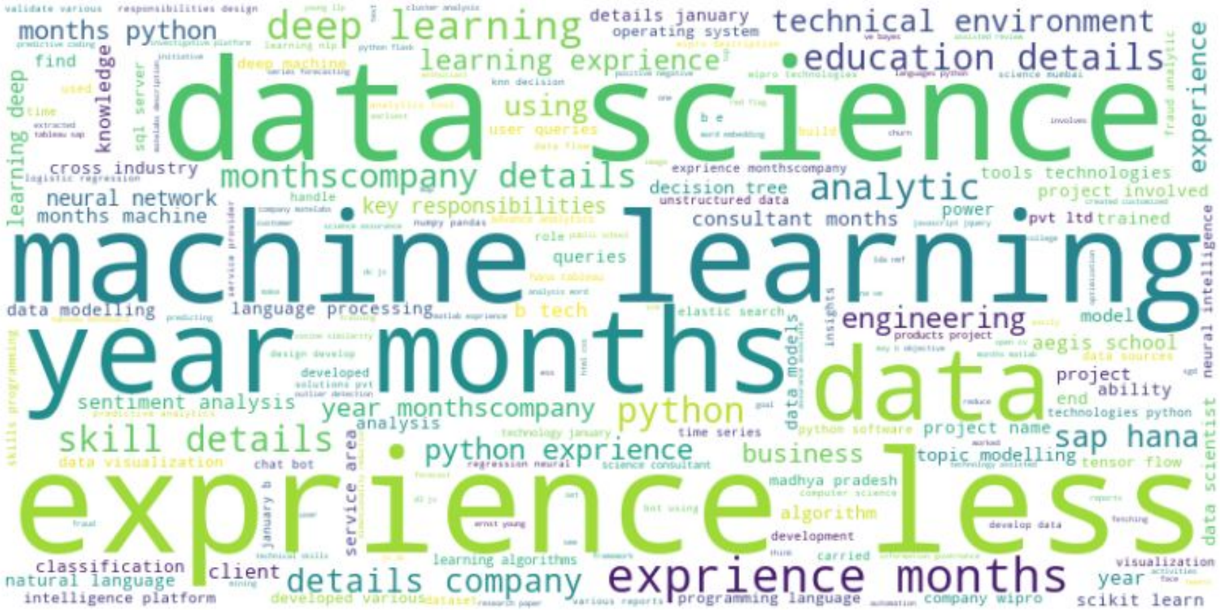
100 job descriptions for Data Scientist role (https://www.kaggle.com/datasets/andrewmvd/data-scientist-jobs)

100 job descriptions for Business Analyst role (https://www.kaggle.com/datasets/andrewmvd/business-analyst-jobs)

# Exploratory Data Analysis (Resumes)



# Exploratory Data Analysis (Resumes)




Data Science

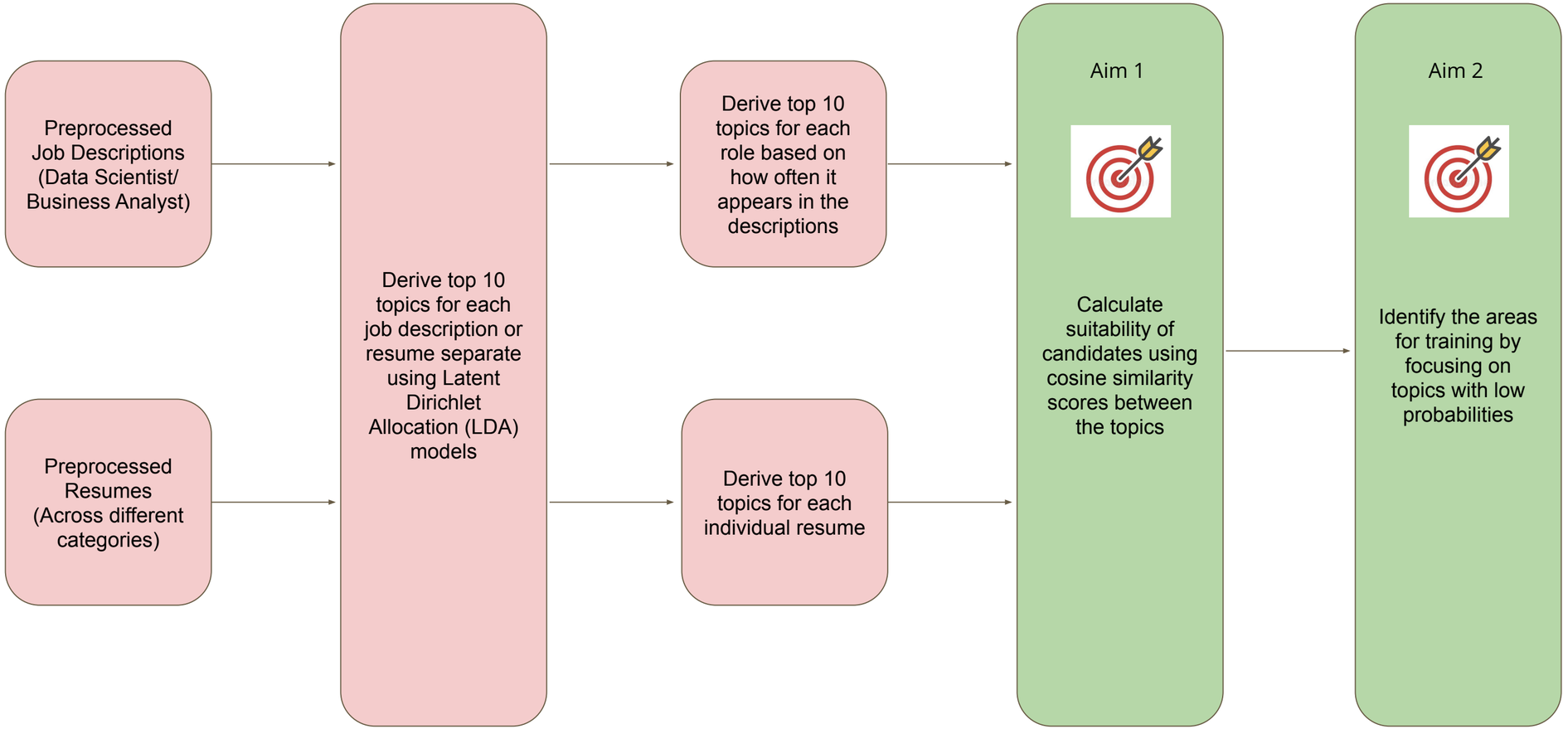
Business Analyst

# Exploratory Data Analysis (Job Descriptions)

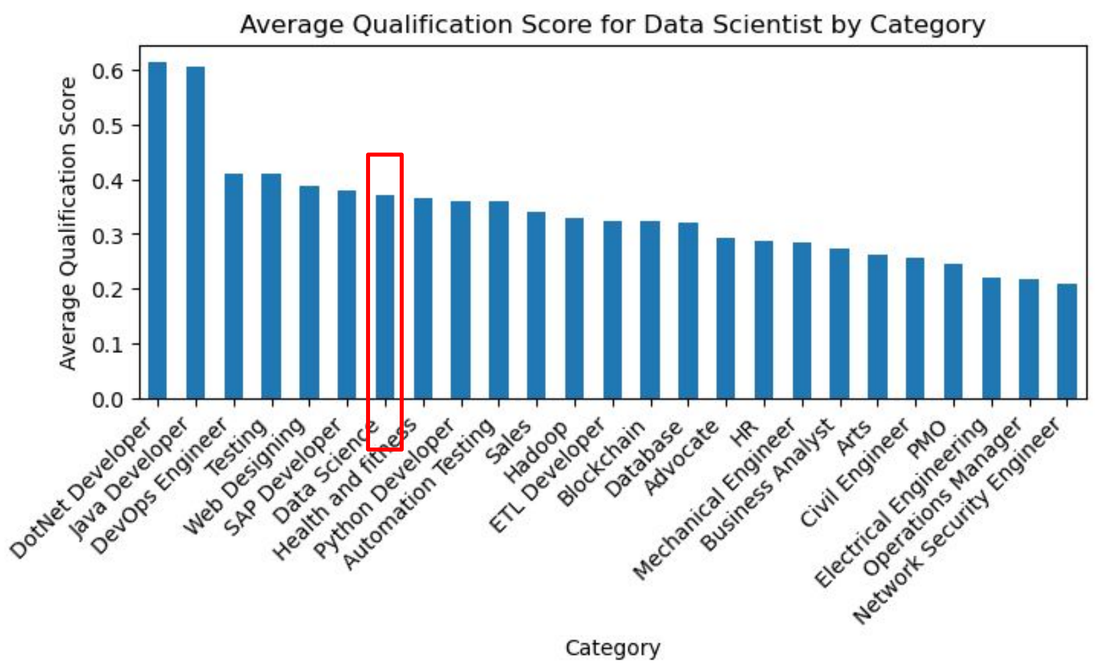
Data Scientist

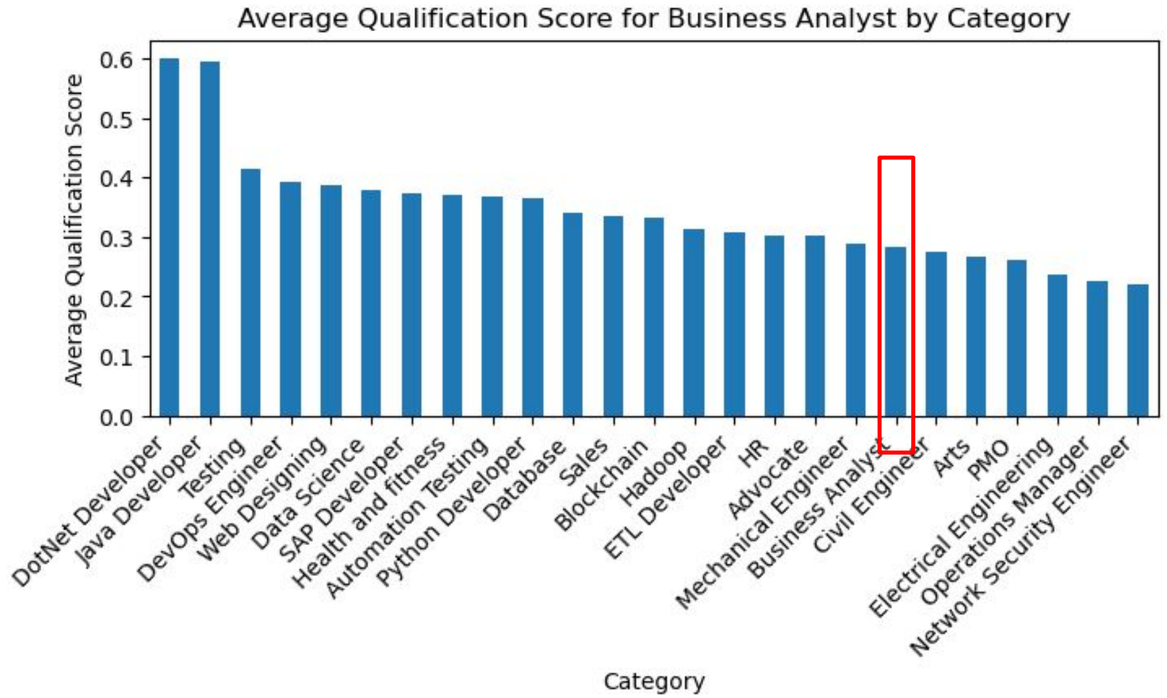
Business Analyst

# Methodology

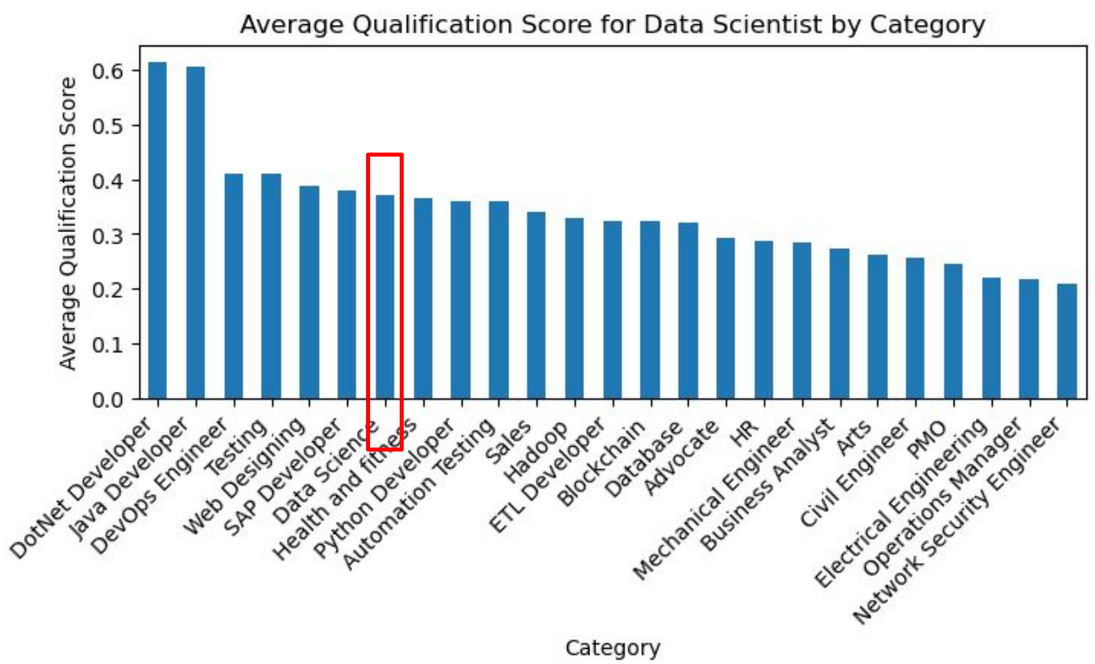


# Results


Average Qualification Score for Data Scientist by Category
Average Qualification Score
0.0
0.1
0.2
0.3
0.4
0.5
0.6
DotNet Developer
Java Developer
DevOps Engineer
Testing
Web Designing
SAP Developer
Data Science
Health and fitness
Python Developer
Automation Testing
Sales
Hadoop
ETL Developer
Blockchain
Database
Advocate
HR
Mechanical Engineer
Business Analyst
Arts
Civil Engineer
PMO
Electrical Engineering
Operations Manager
Network Security Engineer
Category


Average Qualification Score for Business Analyst by Category
Average Qualification Score
0.0
0.1
0.2
0.3
0.4
0.5
0.6
DotNet Developer
Java Developer
Testing
DevOps Engineer
Web Designing
Data Science
SAP Developer
Health and fitness
Automation Testing
Python Developer
Database
Sales
Blockchain
Hadoop
ETL Developer
HR
Advocate
Mechanical Engineer
Business Analyst
Civil Engineer
Arts
PMO
Electrical Engineering
Operations Manager
Network Security Engineer
Category

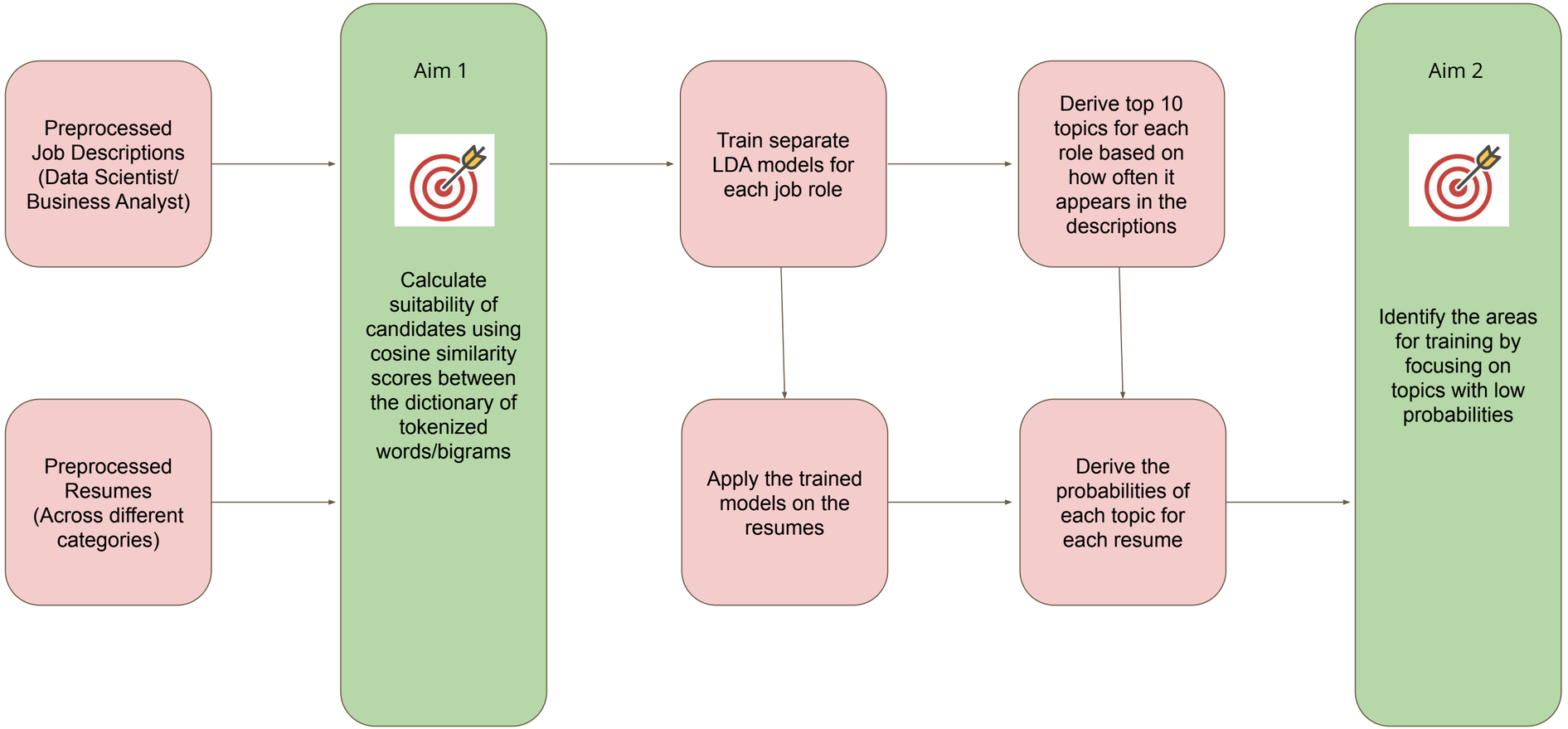
# Results


Average Qualification Score for Data Scientist by Category
Average Qualification Score
0.0
0.1
0.2
0.3
0.4
0.5
0.6
DotNet Developer
Java Developer
DevOps Engineer
Testing
Web Designing
SAP Developer
Data Science
Health and fitness
Python Developer
Automation Testing
Sales
Hadoop
ETL Developer
Blockchain
Database
Advocate
HR
Mechanical Engineer
Business Analyst
Arts
Civil Engineer
PMO
Electrical Engineering
Operations Manager
Network Security Engineer
Category

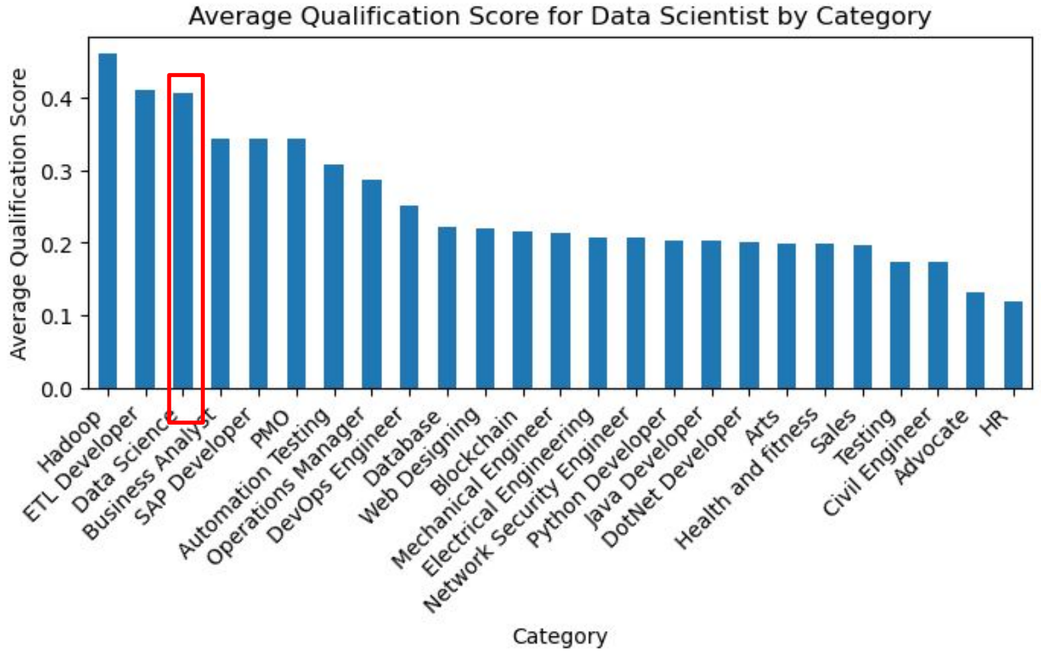
Average Qualification Score for Business Analyst by Category
Average Qualification Score
0.0
0.1
0.2
0.3
0.4
0.5
0.6
DotNet Developer
Java Developer
Testing
DevOps Engineer
Web Designing
Data Science
SAP Developer
Health and fitness
Automation Testing
Python Developer
Database
Sales
Blockchain
Hadoop
ETL Developer
HR
Advocate
Mechanical Engineer
Business Analyst
Civil Engineer
Arts
PMO
Electrical Engineering
Operations Manager
Network Security Engineer
Category

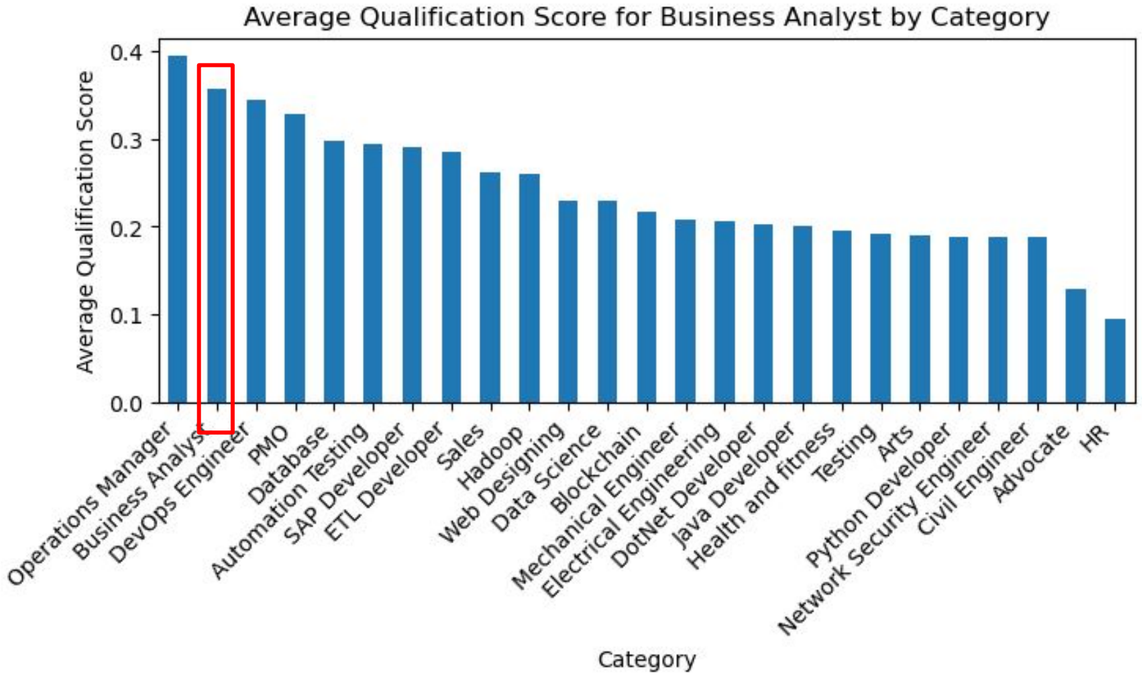
REJECTED

# Methodology (Revised)

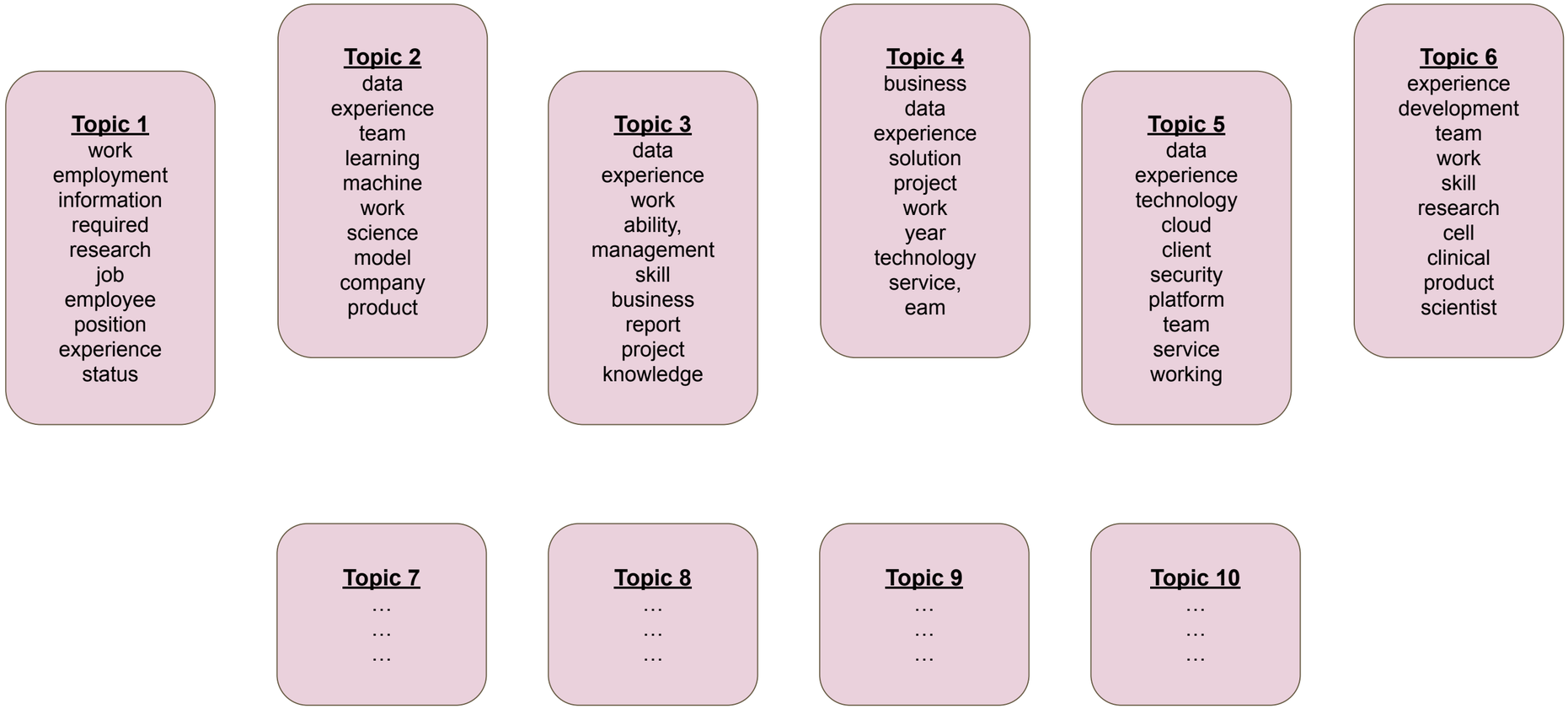


# Results (Revised Methodology)

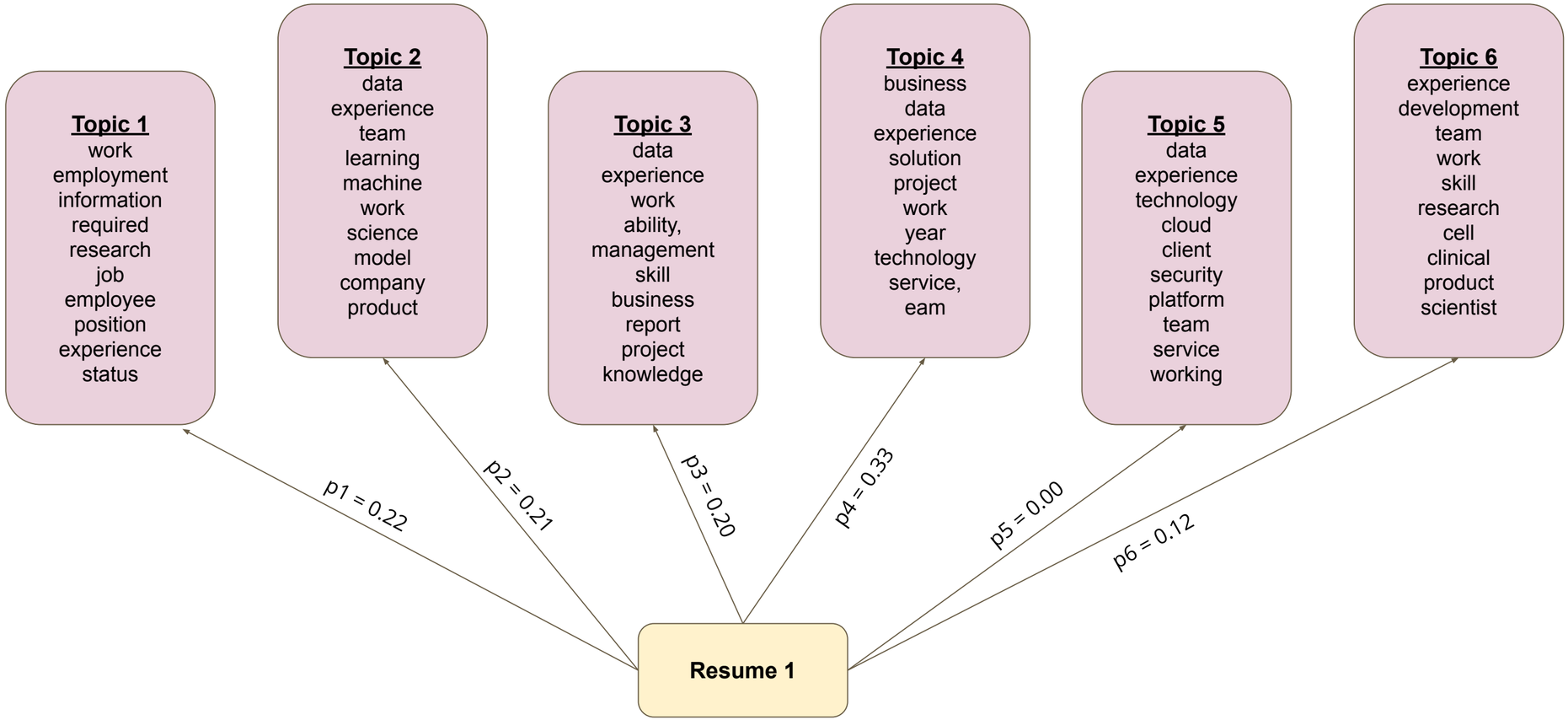


# Topic Modeling (Data Scientist)



# Topic Modeling (Data Scientist)



# Next Steps..

**Limitations of POC**

"Qualified" status using tokenized words and cosine similarity does not consider the position of the words in the resumes.

Topic modeling using LDA not effective, as it relies on bag-of-words approach without consideration for context.

The model captures topics but still require manual assessment of the bag-of-words to identify the nature of the topic.

**Approach for actual project**

Explore assigning different weightage to words based on recency of work experience

Explore using RoBERTa model to capture more complex language patterns and semantic relationships.

Explore mapping to the Technical Skills and Competencies listed in Skillsfuture Framework, which will allow us to identify and propose the post-hire training areas for employers to focus on.

# Thank You