Resumatchmaker

A model helps you find jobs that match your skills and experience
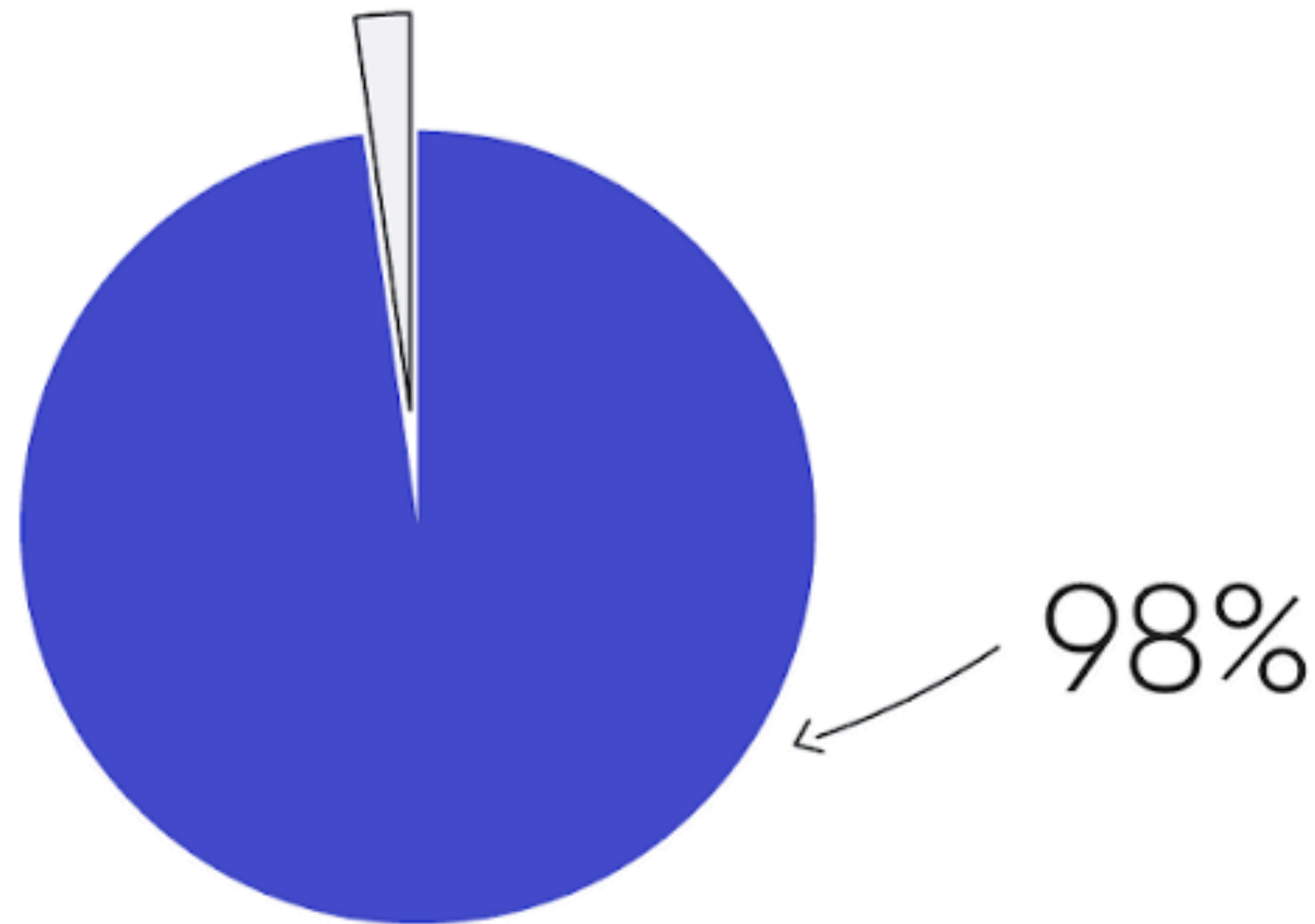
Johana Yuan Feng, 19th May 2021

Agenda

- Background story
- Model Building

  A. Data Cleaning

  B. EDA

  C. Tokenization

- Demonstration
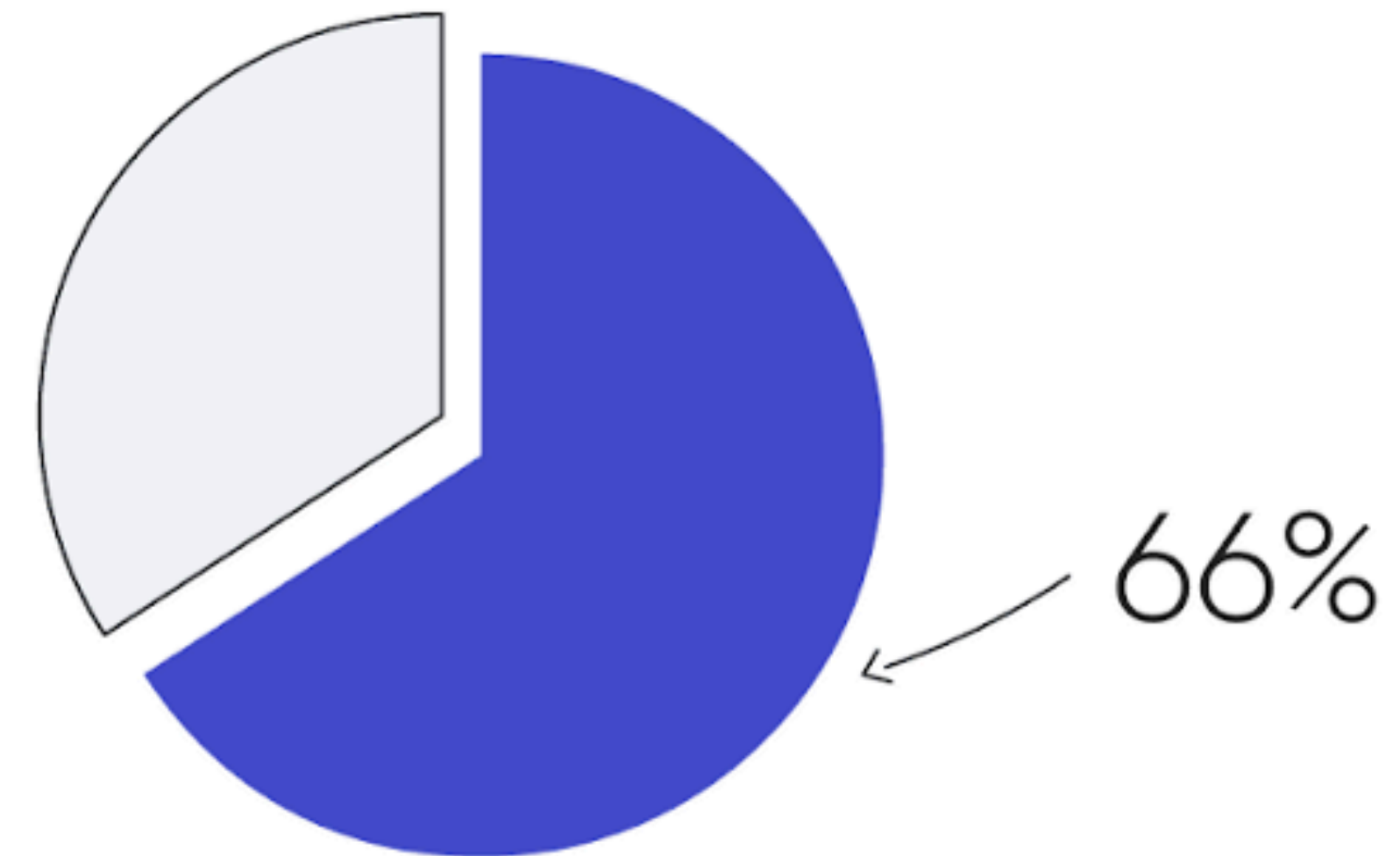- Improvement

Applicant Tracking Systems

Most large corporations utilize applicant tracking systems

98%

of Fortune 500 companies use ATS

66%

of large companies use ATS

Applicant Tracking Systems

- **Resume Pasring**
- **Resume Storage**
- **Keyword Search**
- **Filters**
- **Automated Email Customization**

Applicant Tracking Systems

- **Resume Pasring**
- **Resume Storage**
- **Keyword Search**
- **Filters**
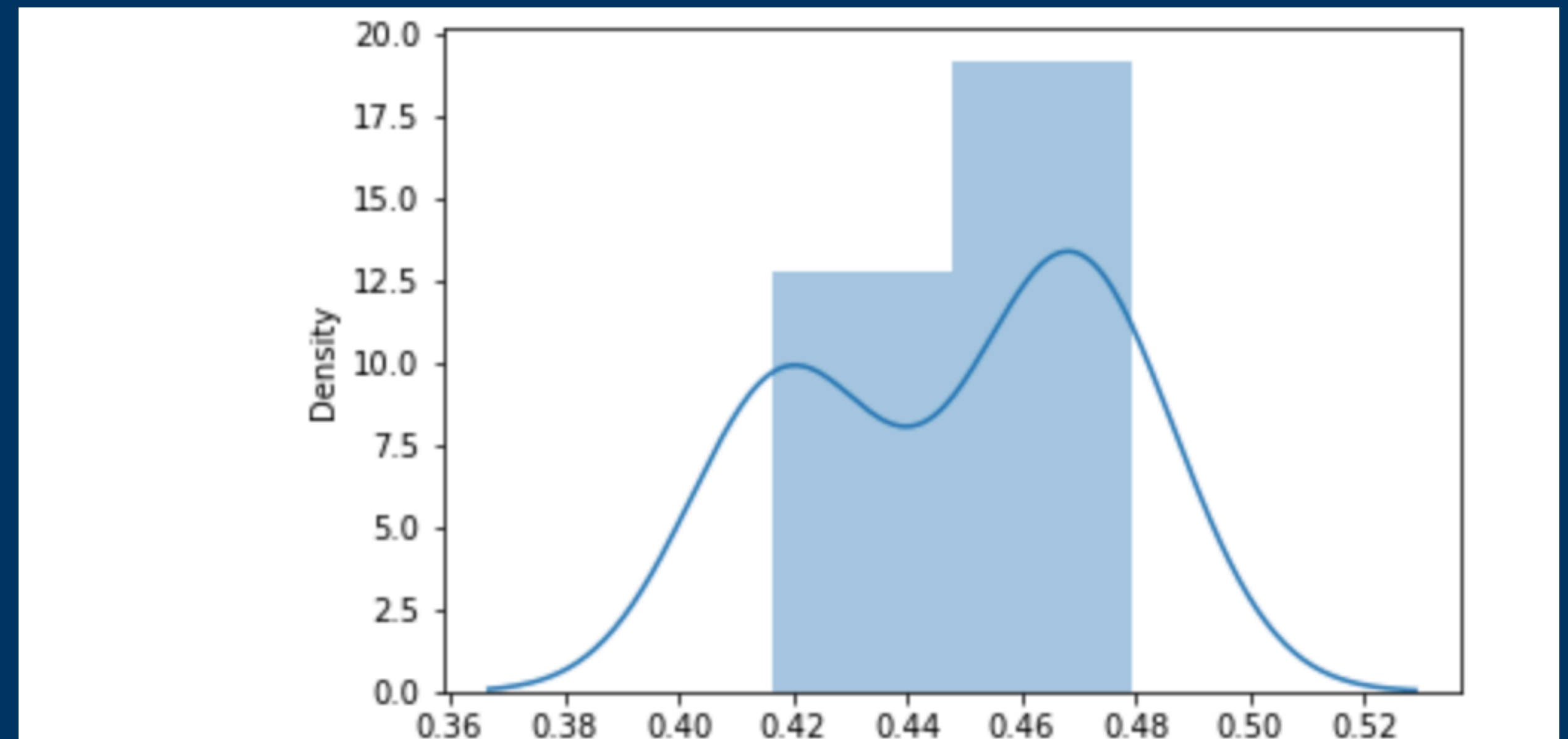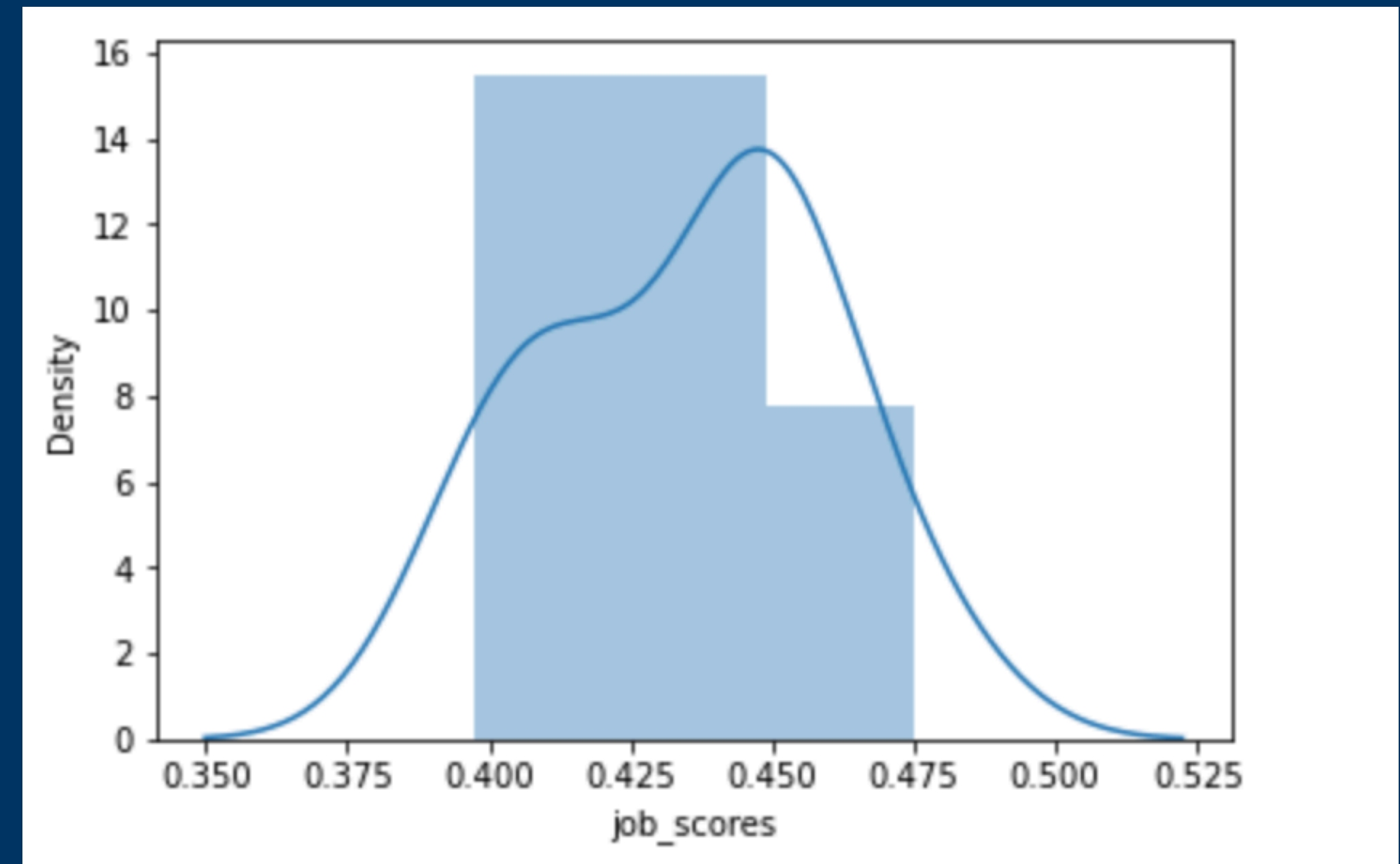- **Automated Email Customization**

Model Building

Data Cleaning

- Web Links
- Email Address
- Non - English
- Punctuationsssss
- Joined words
- Numerical

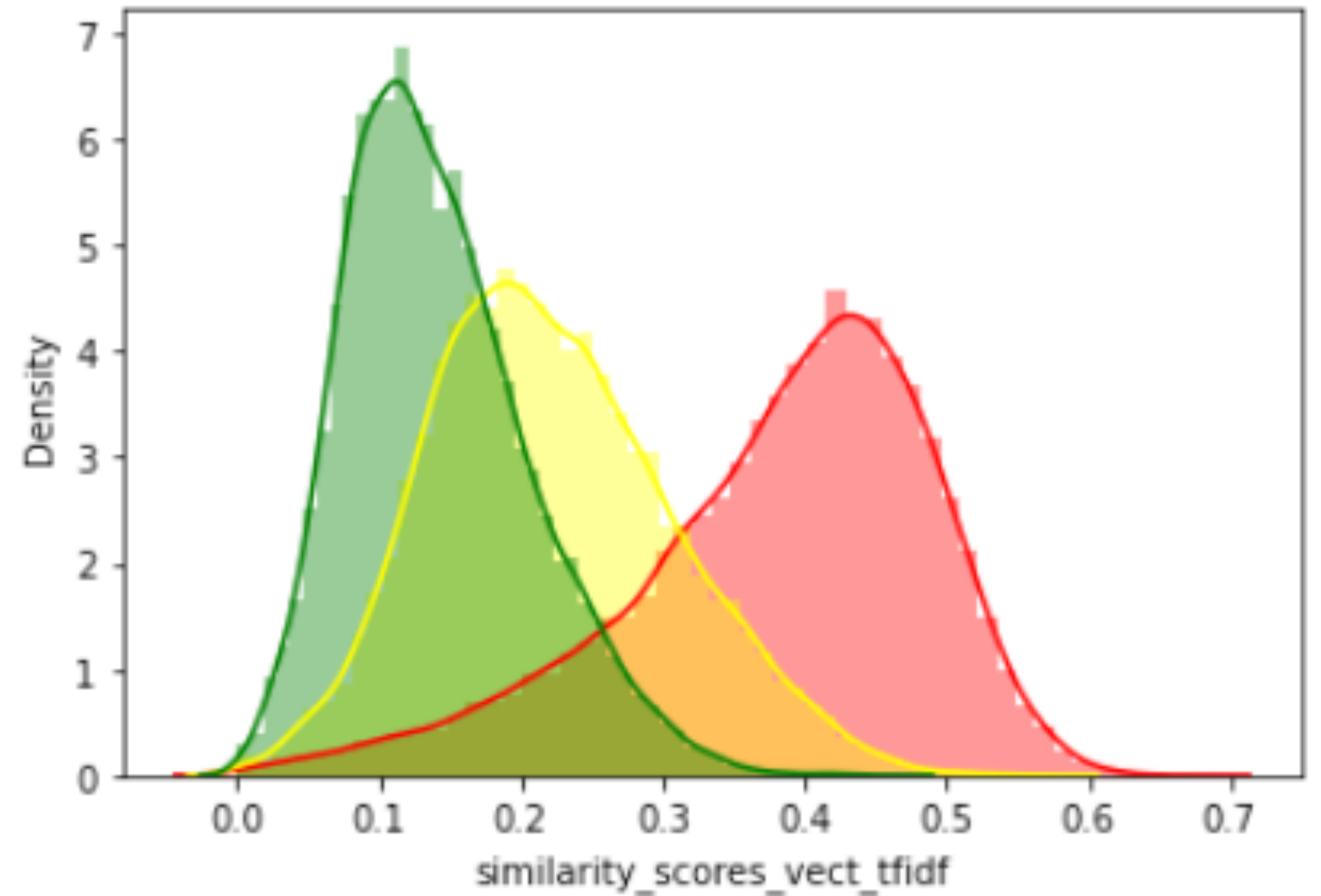Important Matrix:
- Unique words
- Word counts

Libraries:
- Re
- String
- Unidecode
- Wordninja

EDA

```
: ##most used words including stopwords.
  #However, wordcloud has a build-in STOPWORDS list.
  wordcloud(data_clean['jobdescription'])
```

```
##showing the most used words in the dataset.
wordcloud(data_clean_stopwords['jobdescription'])
```

Tokenization

- **Stop words**
- **CountVectorizer**
- **TfidfTransformer**

Demonstration

Improvement

- Paid API
- Put in Function
- Bert
- PDF reader
- CV improvement

Improvement