

```
In [3]: import pandas as pd
```

```
In [4]: data = pd.read_csv("data_analyst_wanted_job_description.csv", index_col=0)
```

```
In [5]: data.head(5)
```

```
Out[5]:
```

	title	company	link	task	qualification	preferred
0	[쿠팡페이] Principal, Business Analyst (Fintech BA)	쿠팡	https://www.wanted.co.kr/wd/65454	쿠팡 페이의 Business Analyst는 최상의 고객 경험을 창출하기 위해 수백...	• 5년 이상의 데이터 분석 관련 경험 \r\n• 빅데이터 분석 환경에 대한 경험과...	• 물리학, 통계학, 수학, 컴퓨터공학 석사 학위 혹은 해당 분야에서 그에 준하는 ...
1	런드리고 데이터 분석가	의식주컴퍼니	https://www.wanted.co.kr/wd/65726	1. 의사결정 지원을 위한 분석 모델링\r\n- 분석 모델링 수행, 고객 분석주제(...	- IT분야 경력 4년 이상 (3~4년)\r\n- 통계분석 알고리즘 이해 및 활용 능력...	- 빅데이터 분석 프로젝트 수행 경험자 우대 (마케팅, 제조, 물류, 서비스)\r\n...
2	Head of Growth	프로젝트 바닐라	https://www.wanted.co.kr/wd/69930	• 프로젝트바닐라의 서비스 핵심 지표 관리 및 데이터 분석\r\n• 다양한 실험을 ...	• 빠르게 성장하는 조직에서 마케팅, 데이터 분석, 또는 사업 기획 업무를 경험하신...	• B2C 분야 신규 서비스 런칭 경험이 있으신 분 \r\n• 금융 투자 분야에 대한...
3	전략기획 전문가	해준	https://www.wanted.co.kr/wd/49216	• 전사/서비스별 중장기 전략 기획 및 로드맵 수립\r\n• KPI 지표 관리 및 ...	• 유관 경력 7년 이상 \r\n• 시장 조사 및 데이터 분석을 통해 서비스 문제를 ...	• 스타트업, IT업계에서의 전략수립 경험\r\n• 태양광, 전력, 에너지 시장에 ...
4	[우아한청년들] 운영 데이터 분석 및 서비스 개선 담당자(3년 이상)	우아한형제들(배달의민족)	https://www.wanted.co.kr/wd/60917	- 배민라이더스 배달데이터 분석 및 개선/이슈 파악\r\n- 배달품질 모니터링 및 ...	- 데이터 분석을 통해 인사이트를 도출해내고, 결과로 정리가 가능한 분\r\n- ...	- 숫자 기반의 논리적 추론과 결과를 바탕으로 사업적 측면에서 커뮤니케이션 역량을 ...

Perform EDA

```
In [6]: import seaborn as sns
import matplotlib.pyplot as plt
```

```
In [7]: %matplotlib inline
```

```
In [8]: num_position = data.groupby("company")["title"].count().sort_values(ascending=True)
top_25_num_position = data.groupby("company")["title"].count().nlargest(25).sort_values(ascending=True)
```

```
In [9]: plt.rc('font', family='Malgun Gothic')
```

원티드의 Data Analyst 포지션으로 분류된 공고는 총 368개가 있음

```
In [10]: data.shape
```

```
Out[10]: (368, 6)
```

총 131개의 회사가 Data Analyst 포지션 채용 중

이로써 (아주 대략적으로) 각 회사당 평균 2~3개의 Data Analyst 수요가 있다는 것을 확인할 수 있음

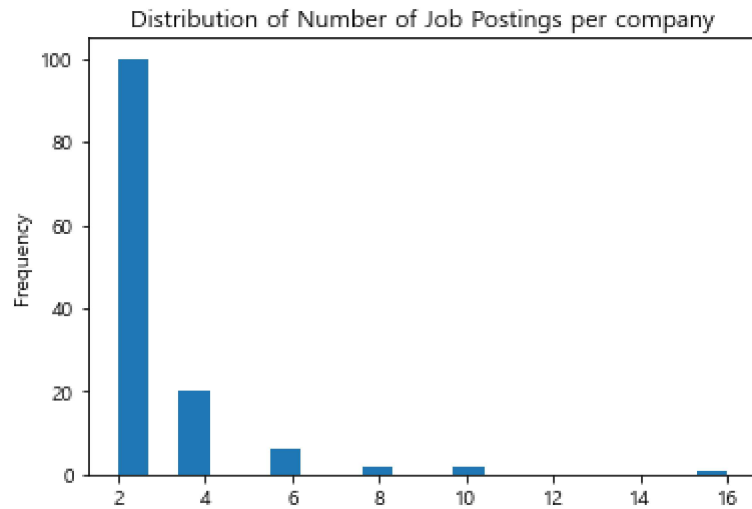
```
In [11]: len(data["company"].unique())
```

```
Out[11]: 131
```

대부분의 회사는 약 2명의 Data Analyst 포지션을 채용하고 있으며, 분포로 봤을 때는 long tail 형태를 보임

```
In [12]: plt.hist(num_position, bins=20)
plt.gca().set(title='Distribution of Number of Job Postings per company', ylabel='Frequency')
```

```
Out[12]: [Text(0.5, 1.0, 'Distribution of Number of Job Postings per company'),
Text(0, 0.5, 'Frequency')]
```

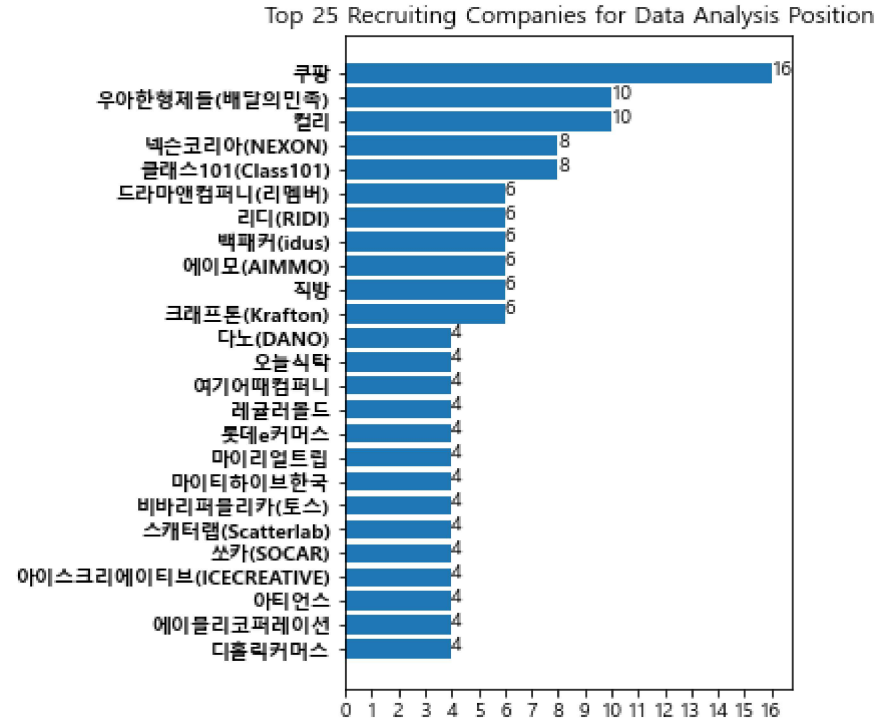


현재 쿠팡이 가장 많은 채용 공고를 올렸다는 것을 알 수 있음

```
In [13]: plt.figure(figsize=(4, 6))
plt.title("Top 25 Recruiting Companies for Data Analysis Position")
plt.yticks(
    fontweight='bold',
    fontsize="10"
)
plt.xticks(
    ticks=range(0,20)
)

plt.barh(y=top_25_num_position.index, width=top_25_num_position)

for index, value in enumerate(top_25_num_position):
    plt.text(value, index, str(value))
```



워드 클라우드로 보기

```
In [42]: import soynlp
import matplotlib.pyplot as plt
from wordcloud import WordCloud
from konlpy.tag import Hannanum
from sklearn.feature_extraction.text import TfidfVectorizer
```

텍스트 전처리

```
In [51]: corpus = " ".join(data["preferred"]).strip()
```

```
In [52]: h = Hannanum()
tokens = h.nouns(phrase=corpus)
```

```
In [53]: noun_corpus = " ".join(tokens)
```

```
In [54]: wordcloud = WordCloud(font_path="C:/Windows/Fonts/malgun.ttf",
background_color='white').generate(noun_corpus)
```

```
In [55]: plt.imshow(wordcloud)
```

```
Out[55]: <matplotlib.image.AxesImage at 0x20d62316cd0>
```

