```
1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4 import seaborn as sns
5 sns.set_theme(color_codes=True)
```

```
1 df = pd.read_csv('Jobs.csv')
2 df
```

	Unnamed: 0	title	company	announcement	description
0	0	Senior Analyst, Data Science and Analytics	TransUnion	The Muse	TransUnion's Job Applicant Privacy Notice Wha...
1	1	Senior Data Scientist	Grubhub Holdings, Inc.	ZipRecruiter	About The Opportunity We're all about connect...
2	2	Lead Data Science Analyst	Discover Financial Services	LinkedIn	Discover. A brighter future. With us, you'll ...
3	3	Data Science Intern	AbelsonTaylor	Startup Jobs	Are you a 2023 college graduate or rising coll...
4	4	Data Scientist	NORC at the University of Chicago	SimplyHired	JOB DESCRIPTION: At NORC, Data Scientists pla...

```
1 df.drop(columns='Unnamed: 0', inplace=True)
2 df
```

	title	company	announcement	description
0	Senior Analyst, Data Science and Analytics	TransUnion	The Muse	TransUnion's Job Applicant Privacy Notice Wha...
1	Senior Data Scientist	Grubhub Holdings, Inc.	ZipRecruiter	About The Opportunity We're all about connect...
2	Lead Data Science Analyst	Discover Financial Services	LinkedIn	Discover. A brighter future. With us, you'll ...
3	Data Science Intern	AbelsonTaylor	Startup Jobs	Are you a 2023 college graduate or rising coll...
4	Data Scientist	NORC at the University of Chicago	SimplyHired	JOB DESCRIPTION: At NORC, Data Scientists pla...
...	...	...	...	...
785	Research and Data Specialist	GovernmentJobs.com	Learn4Good.com	Description $3,000 hiring bonus to join the J...

▾ Clean the text

```
 1 #Remove Punctuation
 2 from nltk.stem.porter import *
 3 from nltk.stem.snowball import SnowballStemmer
 4 import string
 5 stemmer = SnowballStemmer("english")
 6
 7 df['description'] = df['description'].str.lower()
 8 def remove_punct(text):
 9     text  = "".join([char for char in text if char not in string.punctuation])
10     text = re.sub('[0-9]+', '', text)
11     return text
12
13 df['description_punct'] = df['description'].apply(lambda x: remove_punct(x))
14 df.drop(columns='description', inplace=True)
15 df
```

	title	company	announcement	description_punct
0	Senior Analyst, Data Science and Analytics	TransUnion	The Muse	transunions job applicant privacy notice what...
1	Senior Data Scientist	Grubhub Holdings, Inc.	ZipRecruiter	about the opportunity were all about connecti...
2	Lead Data Science Analyst	Discover Financial Services	LinkedIn	discover a brighter future with us you'll do ...
3	Data Science Intern	AbelsonTaylor	Startup Jobs	are you a college graduate or rising college ...
4	Data Scientist	NORC at the University of Chicago	SimplyHired	job description at norc data scientists play ...
...	...	...	...	...
785	Research and Data Specialist	GovernmentJobs.com	Learn4Good.com	description hiring bonus to join the judicia...
	Quality Assurance			are you looking for a

```
1 #Remove Stopwords
2 # Import stopwords with nltk.
3 from nltk.corpus import stopwords
4 import nltk
5 nltk.download("stopwords")
6 stop = stopwords.words('english')
7
```

```
 8 # Exclude stopwords
 9 df['desc_without_stopwords'] = df['description_punct'].apply(lambda x: ' '.join([word for word in x.split() if word not in (stop)]))
10 df.drop(columns='description_punct',inplace=True)
11 df
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```
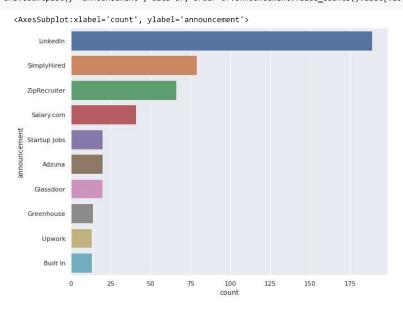
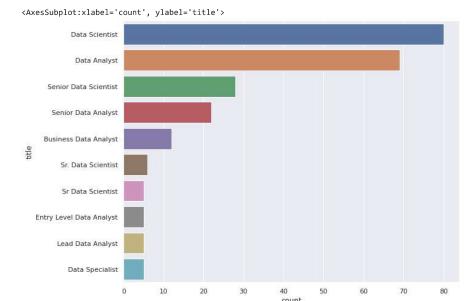|	title	company	announcement	desc_without_stopwords
0	Senior Analyst, Data Science and Analytics	TransUnion	The Muse	transunions job applicant privacy notice well ...
1	Senior Data Scientist	Grubhub Holdings, Inc.	ZipRecruiter	opportunity connecting hungry diners network r...
2	Lead Data Science Analyst	Discover Financial Services	LinkedIn	discover brighter future us you'll meaningful ...
3	Data Science Intern	AbelsonTaylor	Startup Jobs	college graduate rising college senior abelson...
4	Data Scientist	NORC at the University of Chicago	SimplyHired	job description norc data scientists play key ...
...	...	...	...	...
785	Research and Data Specialist	GovernmentJobs.com	Learn4Good.com	description hiring bonus join judicial branch ...
786	Quality Assurance Data	Metrocare Services	Glassdoor	looking purposedriven career metrocare serve n...

▾ Text Analysis

```
 1 from subprocess import check_output
 2 from wordcloud import WordCloud, STOPWORDS
 3
 4 text = df['desc_without_stopwords'].values
 5
 6 wordcloud = WordCloud(collocations = False, background_color = 'white').generate(str(text))
 7 plt.figure(figsize =(10, 15))
 8 plt.imshow(wordcloud, interpolation='bilinear')
 9 plt.axis("off")
10 plt.show()
```



```
1 plt.figure(figsize=(10,8))
2 sns.countplot(y="announcement", data=df, order=df.announcement.value_counts().iloc[:10].index)
```

```
<AxesSubplot:xlabel='count', ylabel='announcement'>
```



```
1 plt.figure(figsize=(10,8))
2 sns.countplot(y="title", data=df, order=df.title.value_counts().iloc[:10].index)
```

```
<AxesSubplot:xlabel='count', ylabel='title'>
```

✓ 0s   completed at 7:58 PM   ● ✕