USA Career Job Listing EDA using Plotly

Import the desired Libraries

In [5]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
import plotly.express as px
import wordcloud
warnings.simplefilter('ignore')
```

In [6]:
```
df=pd.read_csv(r'C:\Users\Ajay Dhariwal\Desktop\Github\usa career building\CareerBuilder_Jobs_2020.csv')
```

In [7]:
```
missing_values=df.isna().sum()
```

In [8]:
```
df_missing_value=pd.DataFrame(missing_values)
df_missing_value.reset_index(inplace=True)
df_missing_value.columns=['Features','missing_count']
df_missing_value['missing_percentage']=(100*df_missing_value['missing_count'])/len(df_missing_value['missing_count'])
```

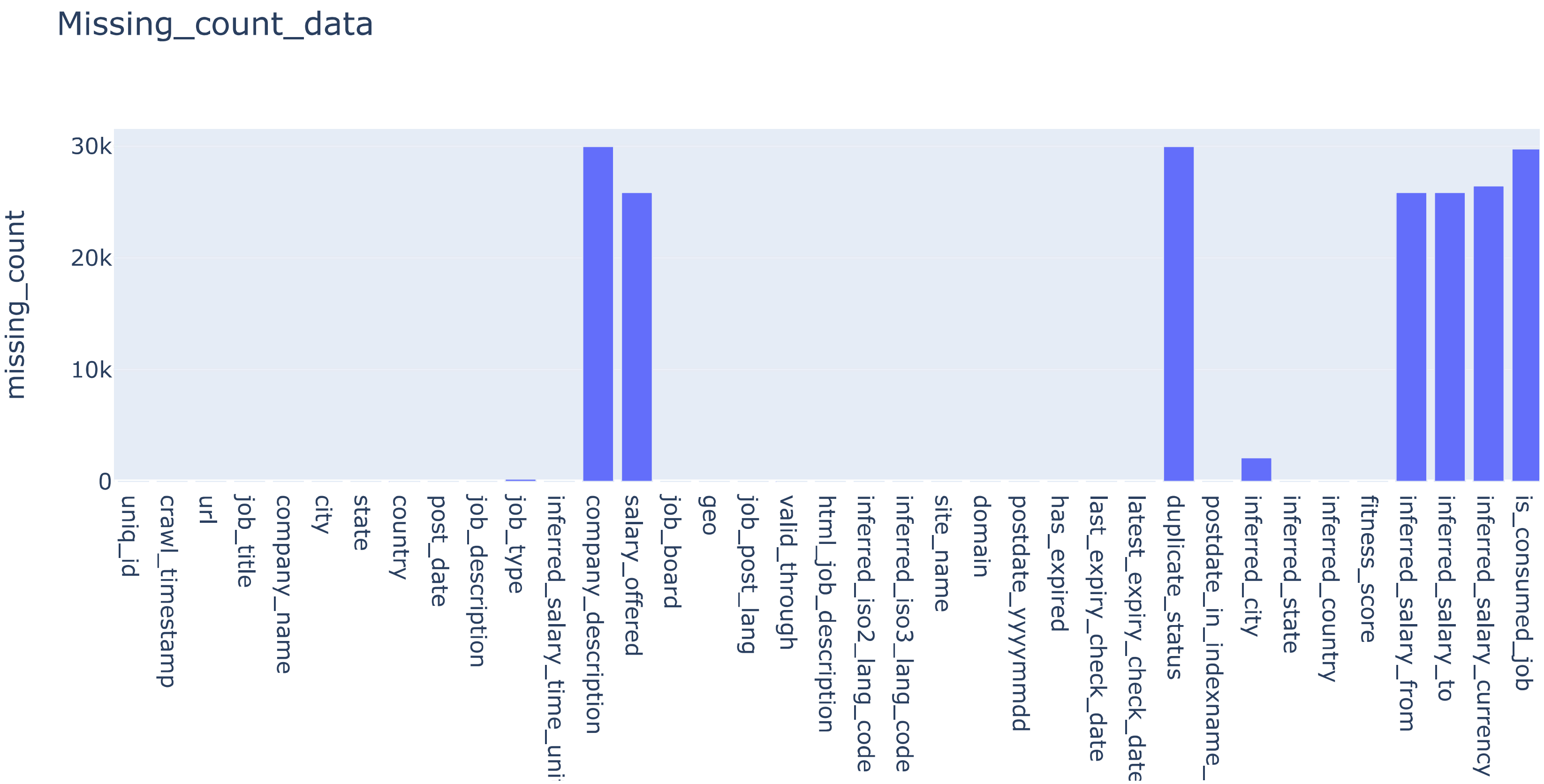
In [9]:

```
df_missing_value.head()
```

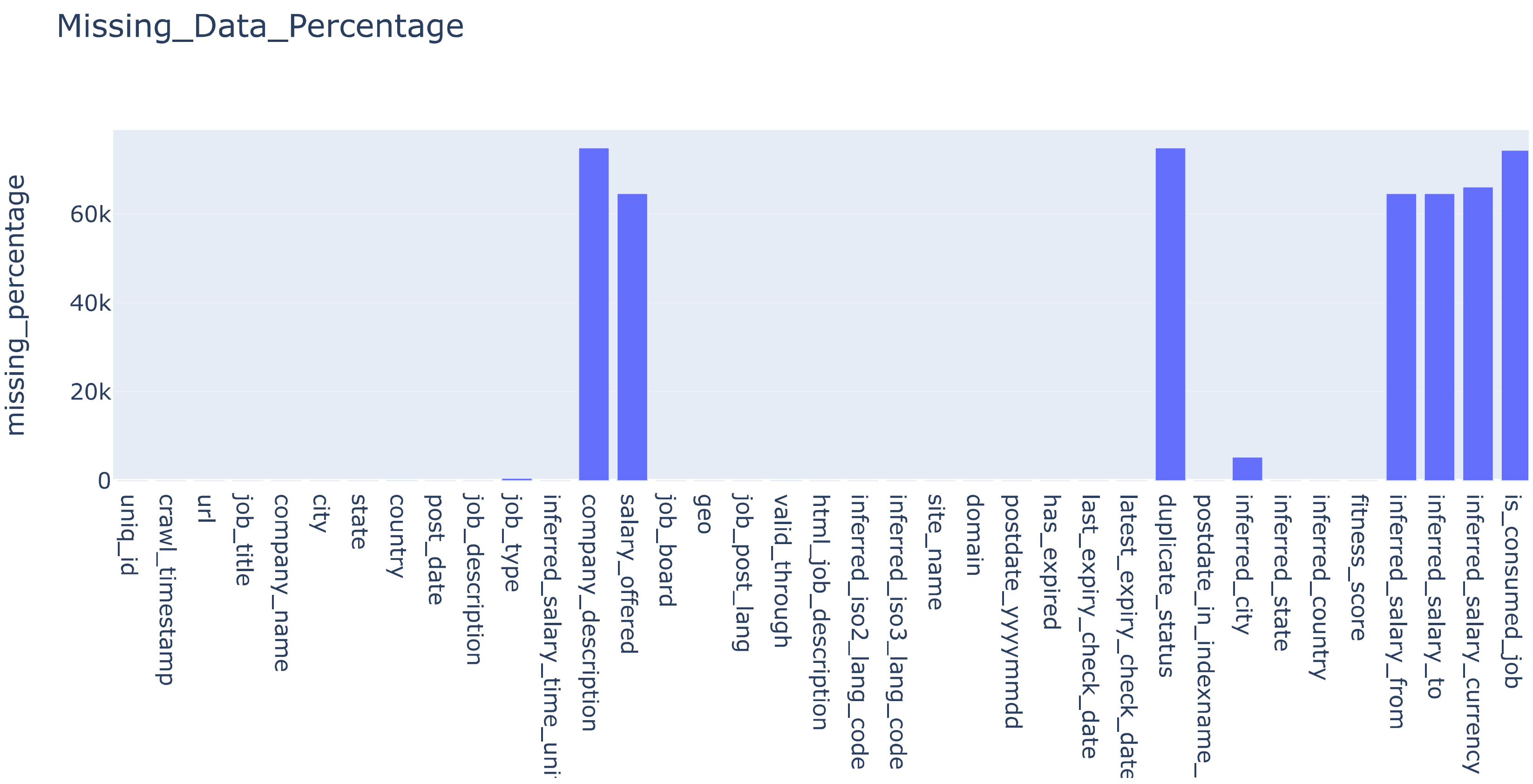
Out[9]:

	Features	missing_count	missing_percentage
0	uniq_id	0	0.0
1	crawl_timestamp	0	0.0
2	url	0	0.0
3	job_title	0	0.0
4	company_name	0	0.0

```
In [10]: fig = px.bar(df_missing_value, x='Features', y='missing_count',title='Missing_count_data')
         fig.show()
```



```
In [11]: fig=px.bar(df_missing_value,x='Features',y='missing_percentage',title='Missing_Data_Percentage')
         fig.show()
```



In [12]:

```
df.isna().sum()
```

```
Out[12]: uniq_id                             0
         crawl_timestamp                     0
         url                                 0
         job_title                           0
         company_name                        0
         city                                0
         state                               0
         country                             9
         post_date                           0
         job_description                     0
         job_type                          232
         inferred_salary_time_unit           0
         company_description             29982
         salary_offered                  25854
         job_board                           0
         geo                                 0
         job_post_lang                       0
         valid_through                       9
         html_job_description                0
         inferred_iso2_lang_code             0
         inferred_iso3_lang_code             0
         site_name                           0
         domain                              0
         postdate_yyyymmdd                   0
         has_expired                         0
         last_expiry_check_date              0
         latest_expiry_check_date            0
         duplicate_status                29982
         postdate_in_indexname_format        0
         inferred_city                    2129
         inferred_state                      0
         inferred_country                    0
         fitness_score                       0
         inferred_salary_from            25854
         inferred_salary_to              25854
         inferred_salary_currency        26464
         is_consumed_job                 29767
         job_requirements                29979
         contact_email                   29970
         test_contact_email              29978
         dtype: int64
```

In [13]:

```
col_drop=['inferred_salary_from','duplicate_status','inferred_city','salary_offered','company_description','inferred_salary_to','inferred_salary_currency','is_consumed_job','job_requirements','contact_email','test_contact_email']
```

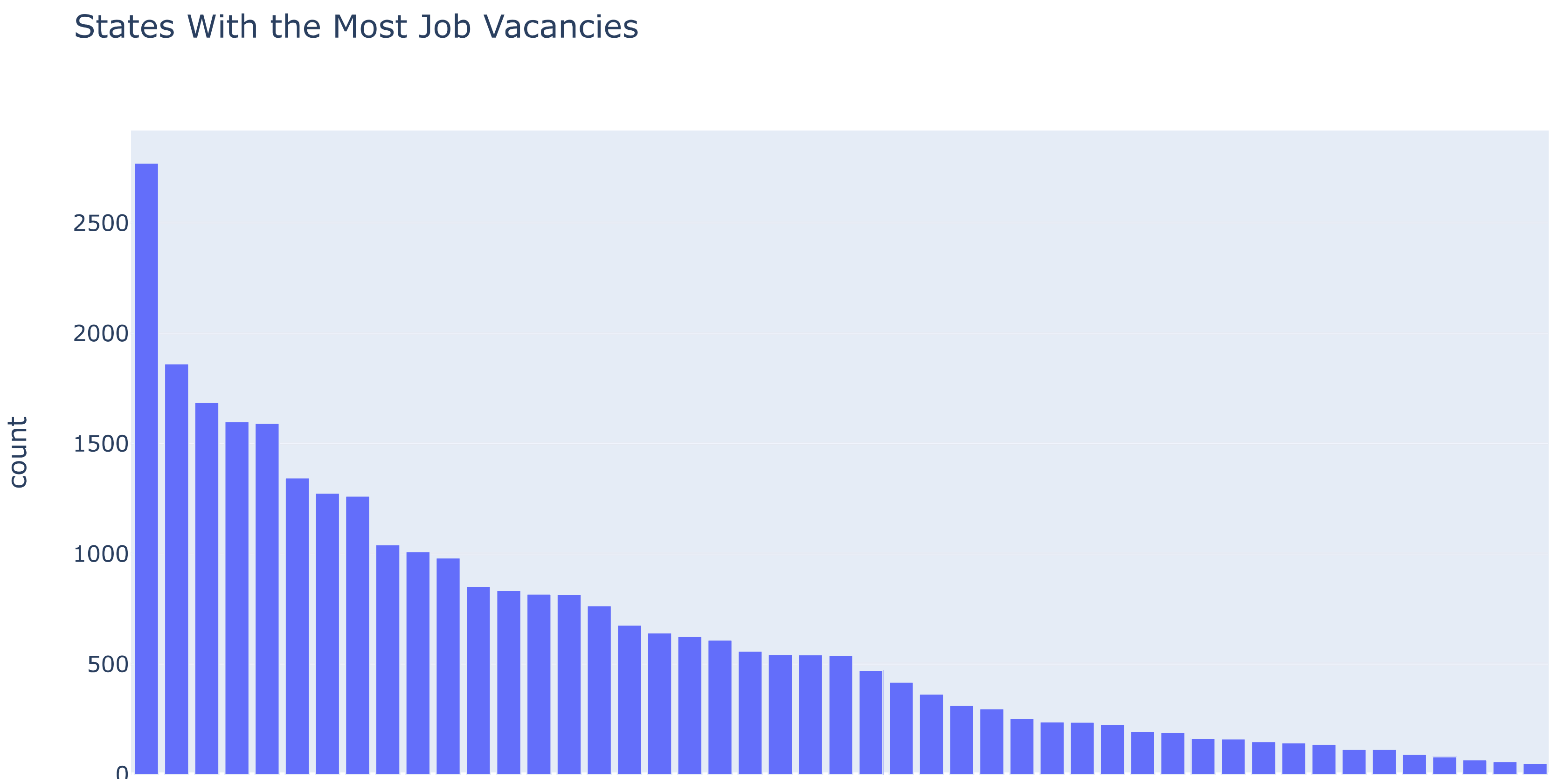
In [14]:

```
df.drop(columns=col_drop,axis=1,inplace=True)
```

In [15]:

```
fig=px.bar(df,x='job_title',y=df['job_title'])
```

In [16]:
```
df_state=df['state'].value_counts()
df_state1=pd.DataFrame(df_state)
df_state1.reset_index(inplace=True)
df_state1.columns=['state','count']
fig1=px.bar(df_state1,x='state',y='count',title='States With the Most Job Vacancies')
fig1.show()
```



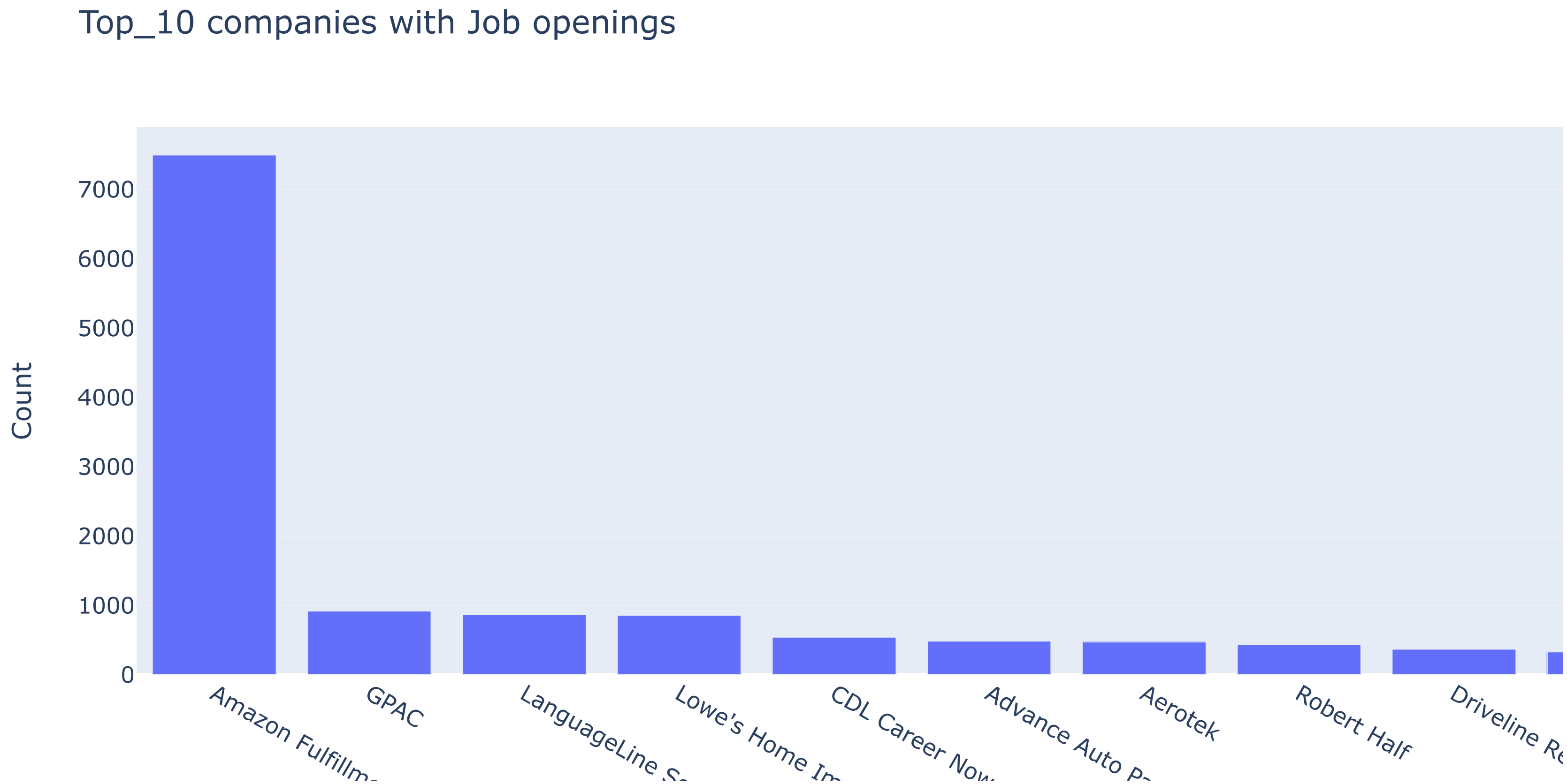
In [17]:
```
df_company=df['company_name'].value_counts().sort_values(ascending=False)[:10]
```

In [18]:
```
df_company1=pd.DataFrame(df_company)
```

In [19]:
```
df_company1.reset_index(inplace=True)
```

In [20]:
```
df_company1.columns=['Company_Name','Count']
```

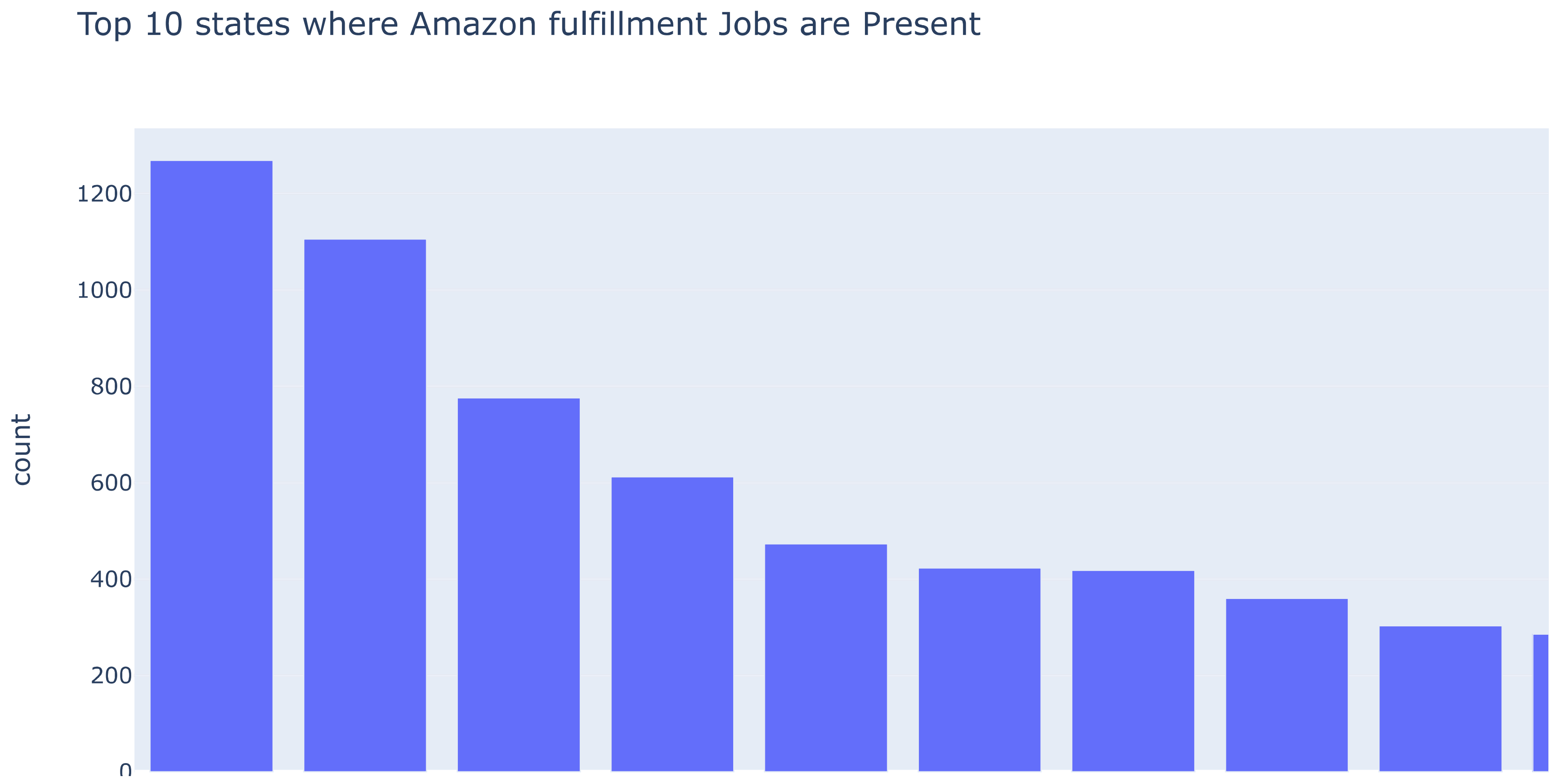
In [21]:
```
fig1=px.bar(df_company1,x='Company_Name',y='Count',title='Top_10 companies with Job openings')
fig1.show()
```



Top_10 companies with Job openings

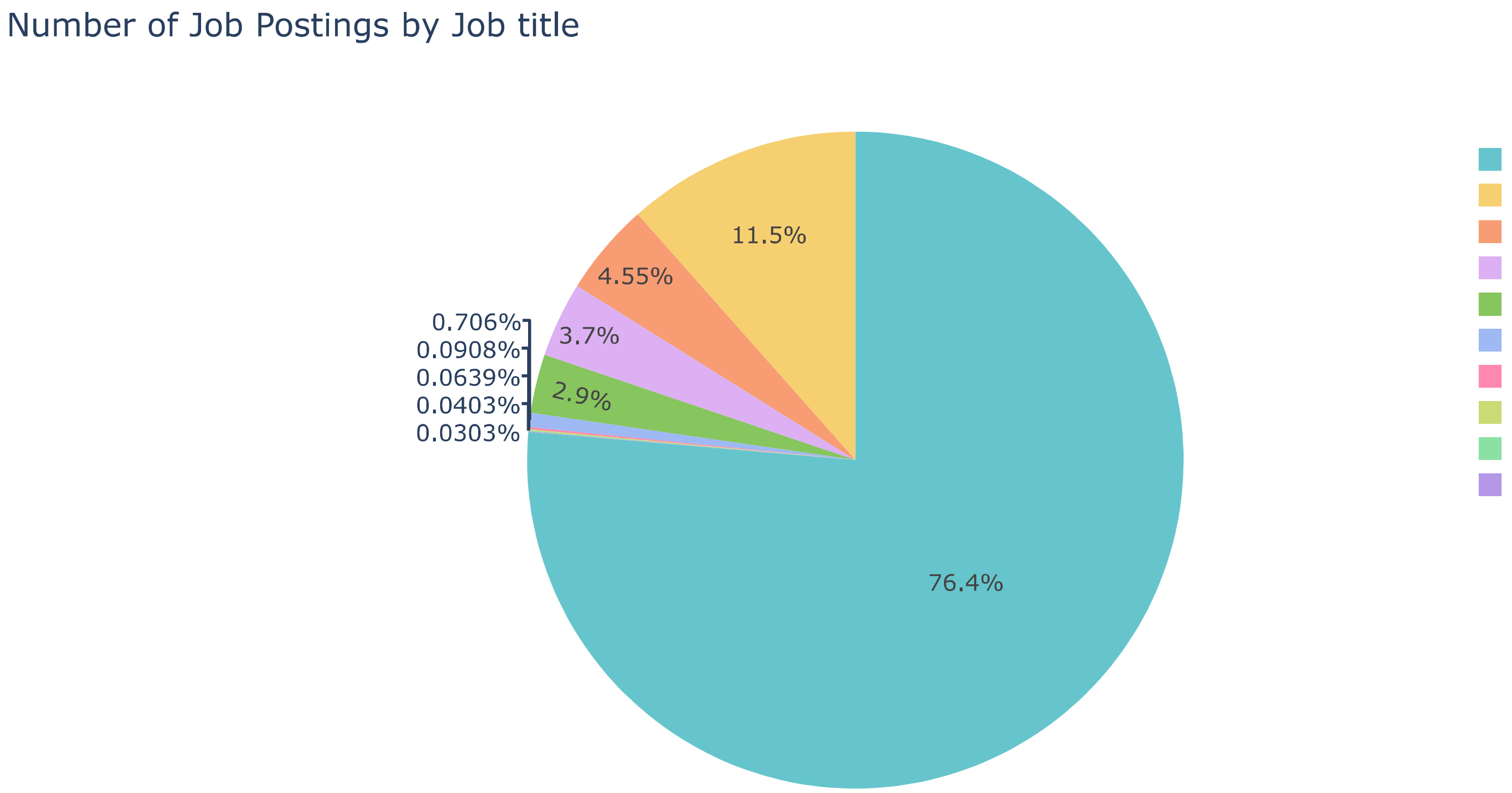
In [22]:
```
amazon=df[df['company_name']=='Amazon Fulfillment']['state'].value_counts().sort_values(ascending=False)[:10]
```

In [23]:
```
df_amazon=pd.DataFrame(amazon)
df_amazon.reset_index(inplace=True)
df_amazon.columns=['state','count']
fig=px.bar(df_amazon,x='state',y='count',title='Top 10 states where Amazon fulfillment Jobs are Present')
fig.show()
```



In [24]:

```
df_jobtypes=df['job_type'].value_counts()
px.pie(df_jobtypes,values=df_jobtypes.values,names=df_jobtypes.index,title='Number of Job Postings by Job title',
       color_discrete_sequence = px.colors.qualitative.Pastel)
```



In [25]:

```
df_jobtype=df['job_type'].value_counts().sort_values(ascending=False)
df_job=pd.DataFrame(df_jobtype)
df_job.reset_index(inplace=True)
df_job.columns=['job_type','count']
```

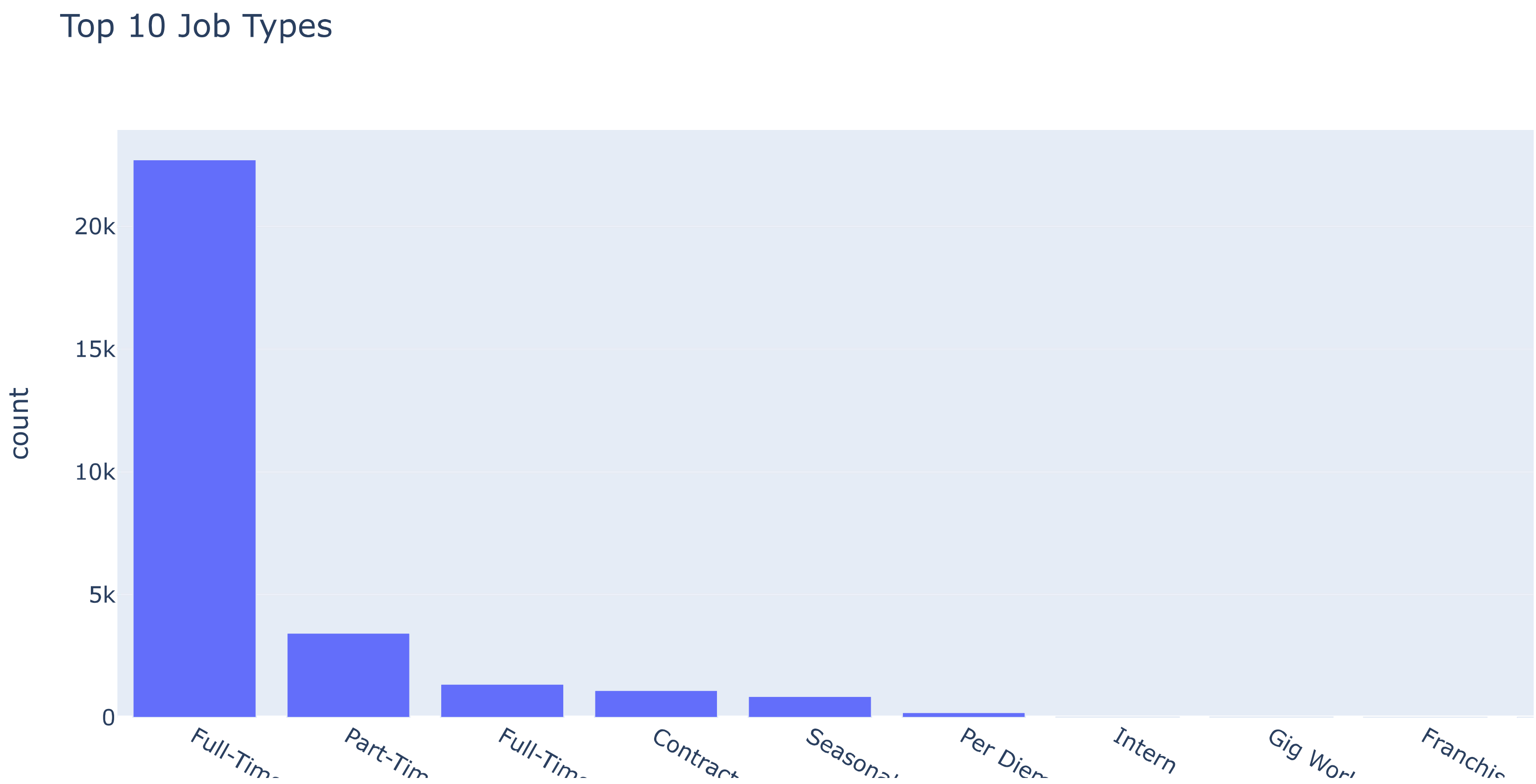
In [26]:

```
df_company1.head()
```

Out[26]:

	Company_Name	Count
0	Amazon Fulfillment	7501
1	GPAC	924
2	LanguageLine Solutions	871
3	Lowe's Home Improvement	861
4	CDL Career Now	545

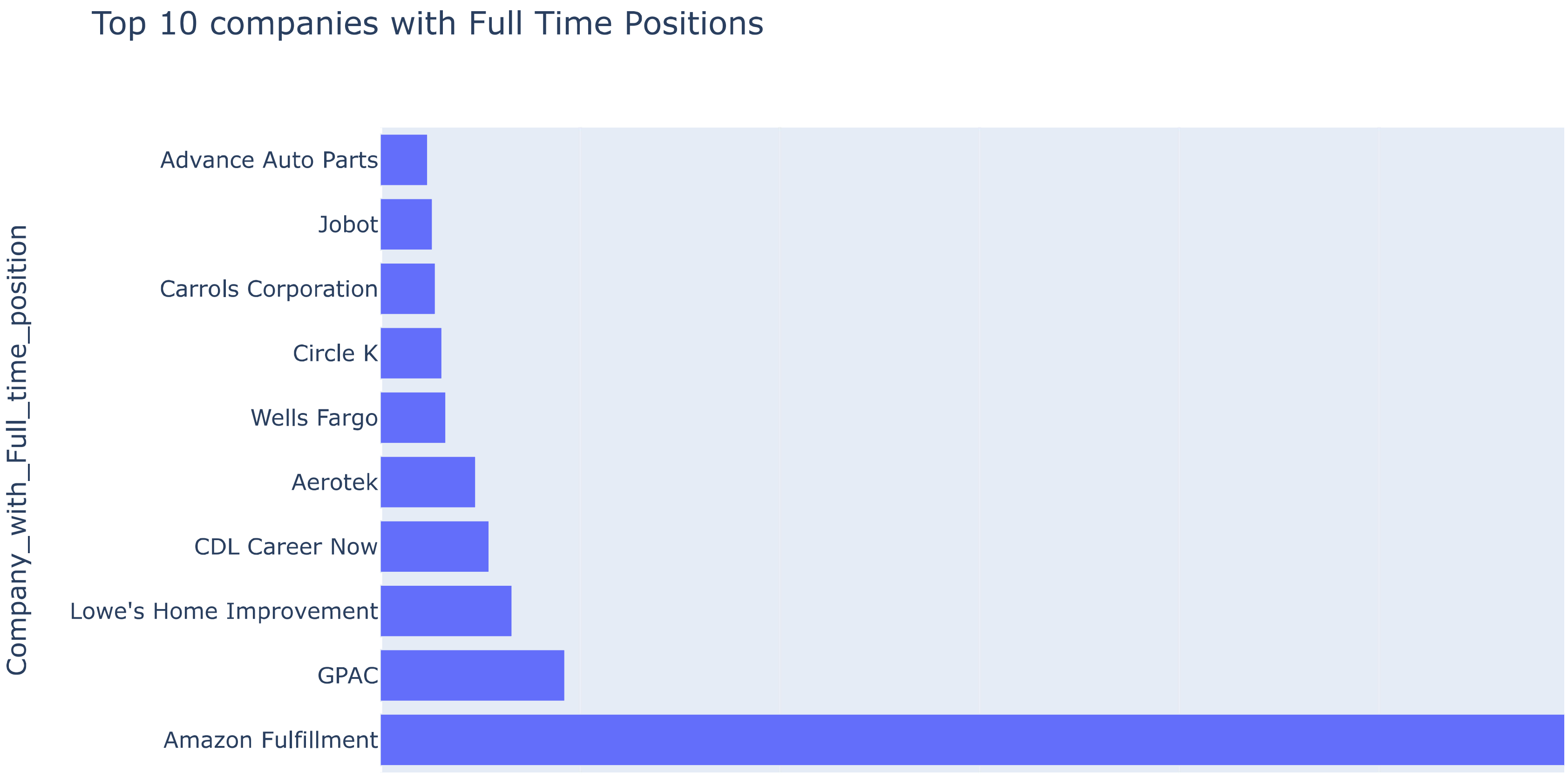
In [27]:
```
fig=px.bar(df_job,x='job_type',y='count',title='Top 10 Job Types')
fig.show()
```



In [28]:
```
job_type=df[df['job_type']=='Full-Time'] [['company_name', 'state', 'city', 'job_title']]
```

In [29]:
```
j_df=job_type['company_name'].value_counts().sort_values(ascending=False)[:10]
j_df_=pd.DataFrame(j_df)
j_df_.reset_index(inplace=True)
j_df_.columns=['Company_with_Full_time_position','count']
```

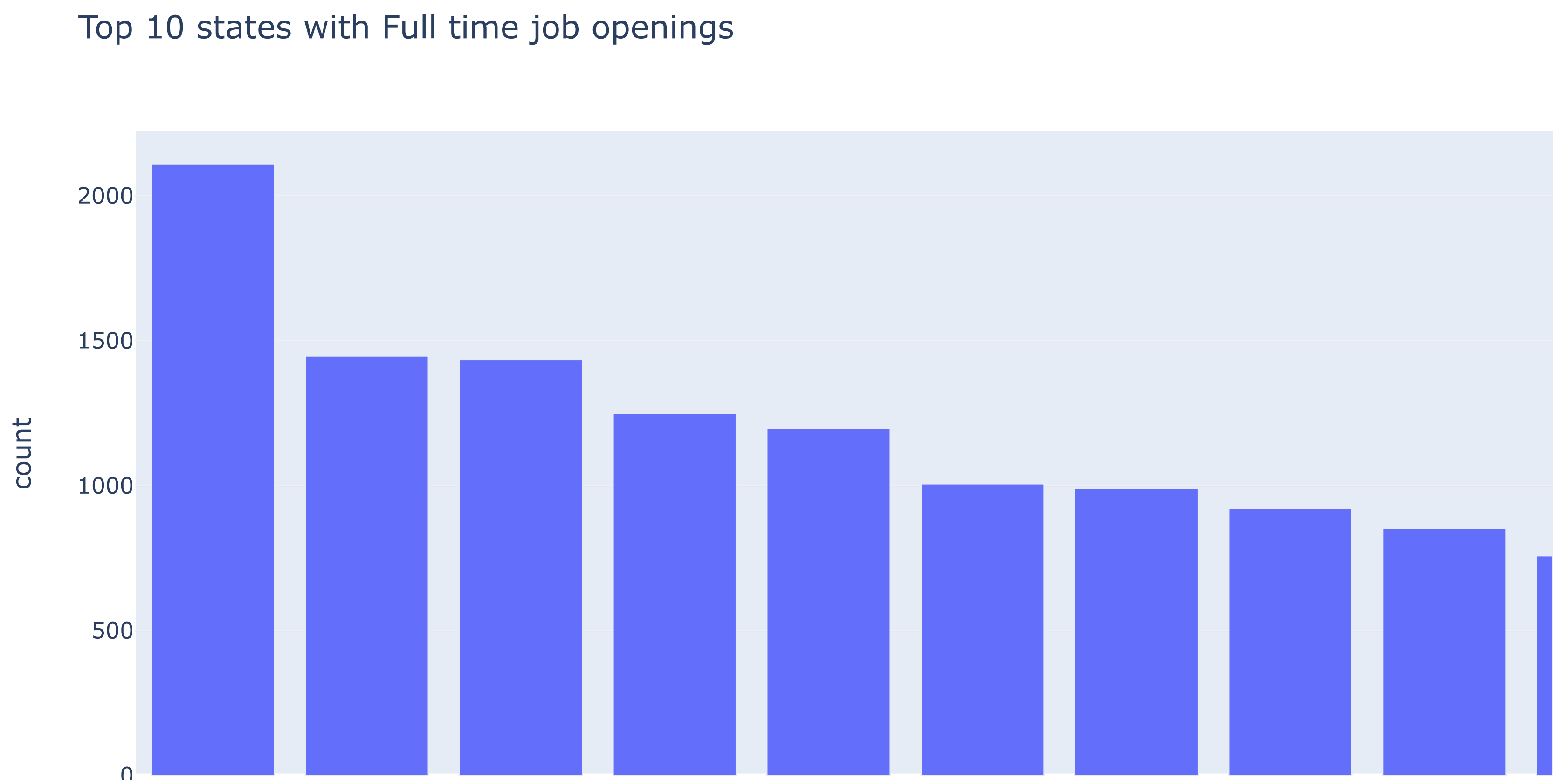
In [30]:
```
fig=px.bar(j_df_,y='Company_with_Full_time_position',x='count',orientation='h',title='Top 10 companies with F
ull Time Positions')
fig.show()
```



In [31]:
```
s_df=job_type['state'].value_counts().sort_values(ascending=False)[:10]
s_df_=pd.DataFrame(s_df)
s_df_.reset_index(inplace=True)
s_df_.columns=['Company_with_Full_time_position','count']
```

In [32]:

```
fig=px.bar(s_df_,x='Company_with_Full_time_position',y='count',title='Top 10 states with Full time job openings')
fig.show()
```


Top 10 states with Full time job openings
count