

```

from google.colab import files
uploaded = files.upload()

import pandas as pd
df = pd.read_csv('extended_dummy_job_postings.csv')
df.head()
df.info()
df.describe()
df.isnull().sum()
df.drop_duplicates(inplace=True)
df['Skills'] = df['Skills'].str.lower().str.strip()
df['Job Title'] = df['Job Title'].str.strip()
df.to_csv('cleaned_jobs.csv', index=False)
import matplotlib.pyplot as plt
import pandas as pd
df = pd.read_csv('cleaned_jobs.csv')
top_companies = df['Company'].value_counts().head(10)
print(top_companies)
top_companies = df['Company'].value_counts().head(10)
print(top_companies)

top_companies.plot(kind='bar', color='orange')
plt.title('Top Hiring Companies for Data Science Roles')
plt.xlabel('Company')
plt.ylabel('Number of Job Postings')
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()

job_counts = df['Job Title'].value_counts().head(10)
job_counts.plot(kind='bar', color='skyblue')
plt.title("Top 10 Job Titles")
plt.ylabel("Count")
plt.xticks(rotation=45)
plt.show()

location_counts = df['Location'].value_counts()
print(location_counts)

import matplotlib.pyplot as plt

location_counts.plot(kind='barh', color='green')
plt.title("Top Job Locations for Data Science Roles")
plt.xlabel("Number of Job Postings")
plt.ylabel("City")
plt.tight_layout()
plt.show()

all_skills = ", ".join(df['Skills'].dropna())
print(all_skills)
from wordcloud import WordCloud

wordcloud = WordCloud(width=800, height=400, background_color='white').generate(all_skills)

```

```
plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.title('Most Common Skills in Job Postings')
plt.show()
```

```
job_levels = ['Entry', 'Mid', 'Entry', 'Senior', 'Mid', 'Mid', 'Entry', 'Mid']
df['Level'] = job_levels
df[['Job Title', 'Level']]
from collections import Counter
levels = df['Level'].unique()
level_skill_counts = {}
```

```
for level in levels:
    skills = df[df['Level'] == level]['Skills'].dropna()
    all_skills = [s.strip() for row in skills for s in row.split(',')]
    level_skill_counts[level] = Counter(all_skills)
```

```
level_df = pd.DataFrame(level_skill_counts).fillna(0).astype(int)
level_df = level_df.sort_values(by='Entry', ascending=False)
level_df.head(10)
level_df.head(10).plot(kind='bar', figsize=(12, 6))
plt.title('Skill Trends by Job Level')
plt.ylabel('Frequency')
plt.xlabel('Skill')
plt.xticks(rotation=45)
plt.legend(title='Job Level')
plt.tight_layout()
plt.show()
```

```
from collections import Counter
```

```
company_skills = {}
```

```
for company in df['Company'].unique():
    skills = df[df['Company'] == company]['Skills'].dropna()
    all_skills = [s.strip() for row in skills for s in row.split(',')]
    company_skills[company] = Counter(all_skills)
```

```
company_df = pd.DataFrame(company_skills).fillna(0).astype(int)
company_df = company_df.loc[company_df.sum(axis=1) > 1] # Show only frequently used skills
company_df.head(10)
```

```
company_df.head(10).plot(kind='bar', figsize=(12, 6))
plt.title('Skill Demand Comparison Across Companies')
plt.ylabel('Frequency')
plt.xlabel('Skill')
plt.xticks(rotation=45)
plt.legend(title='Company')
plt.tight_layout()
```

```
plt.show()
```



Choose Files extended_d...ostings.csv

• **extended_dummy_job_postings.csv**(text/csv) - 822 bytes, last modified: 7/12/2025 - 100% done

Saving extended_dummy_job_postings.csv to extended_dummy_job_postings (3).csv

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 8 entries, 0 to 7

Data columns (total 5 columns):

#	Column	Non-Null Count	Dtype
0	Job Title	8 non-null	object
1	Company	8 non-null	object
2	Location	8 non-null	object
3	Skills	8 non-null	object
4	Date Posted	8 non-null	object

dtypes: object(5)

memory usage: 452.0+ bytes

Company

Google 2

Amazon 2

Microsoft 1

OpenAI 1

Meta 1

Flipkart 1

Name: count, dtype: int64

Company

Google 2

Amazon 2

Microsoft 1

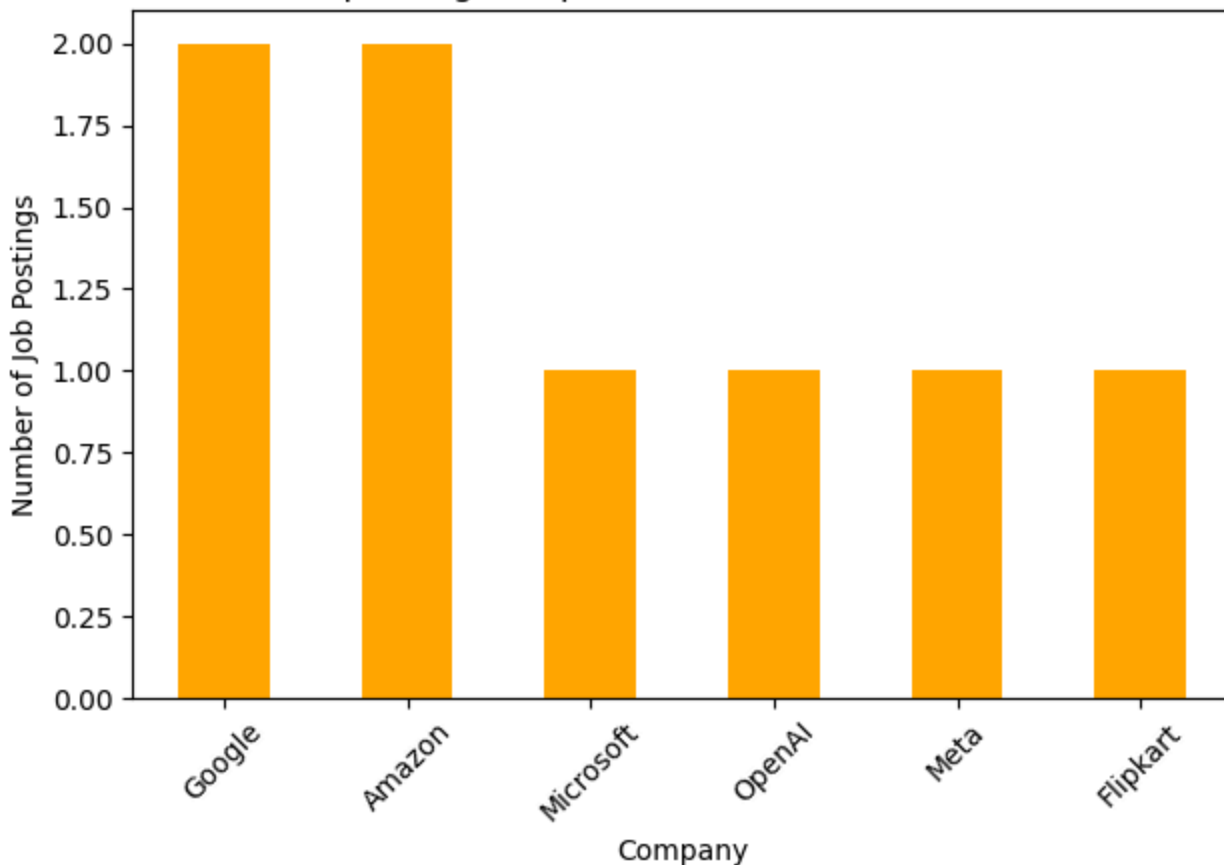
OpenAI 1

Meta 1

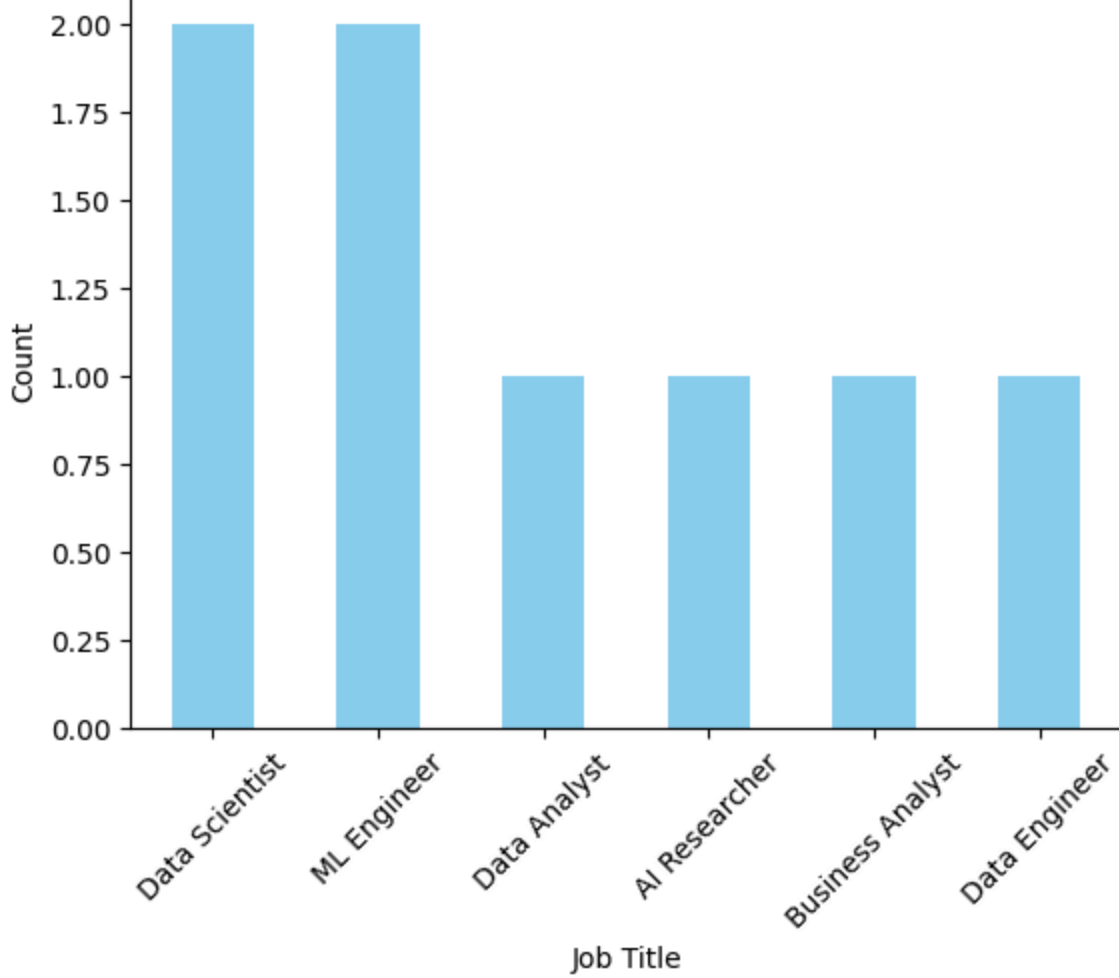
Flipkart 1

Name: count, dtype: int64

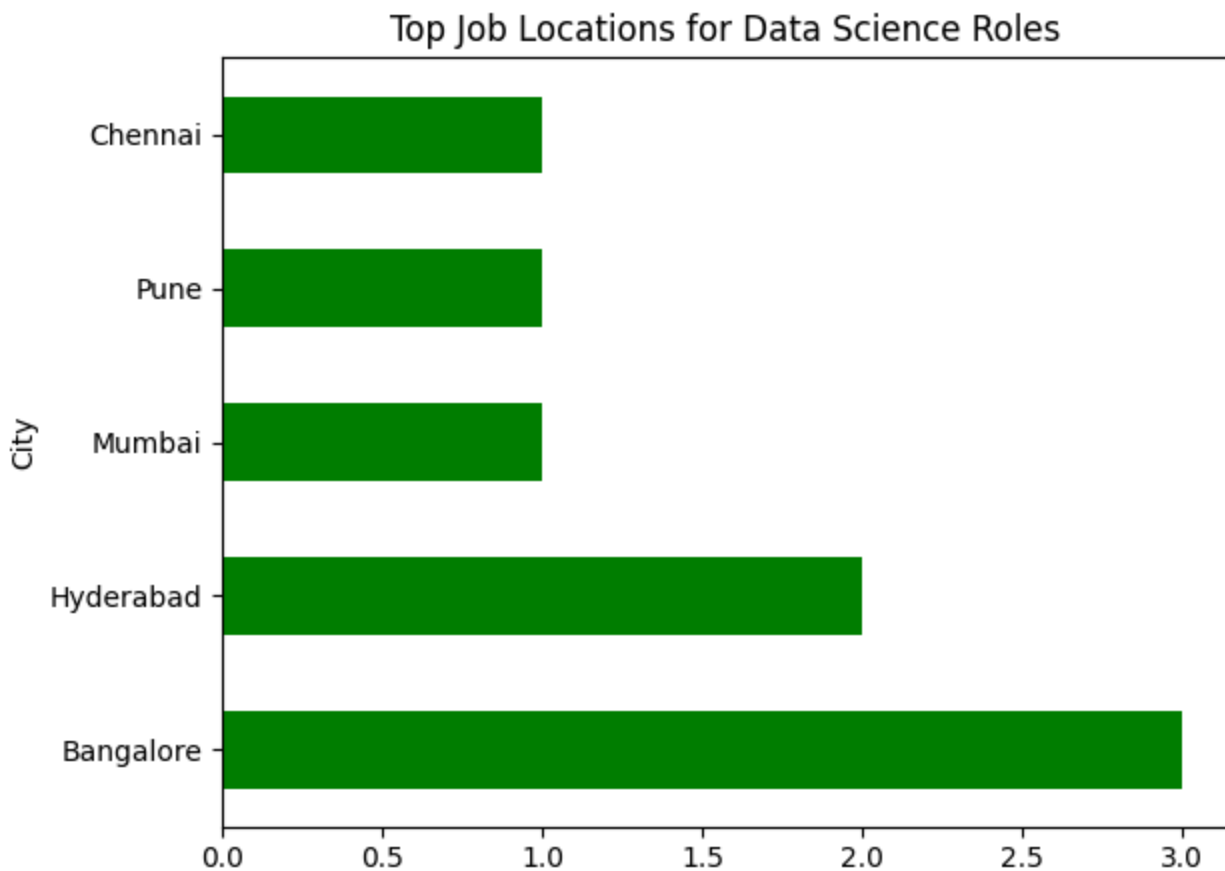
Top Hiring Companies for Data Science Roles



Top 10 Job Titles



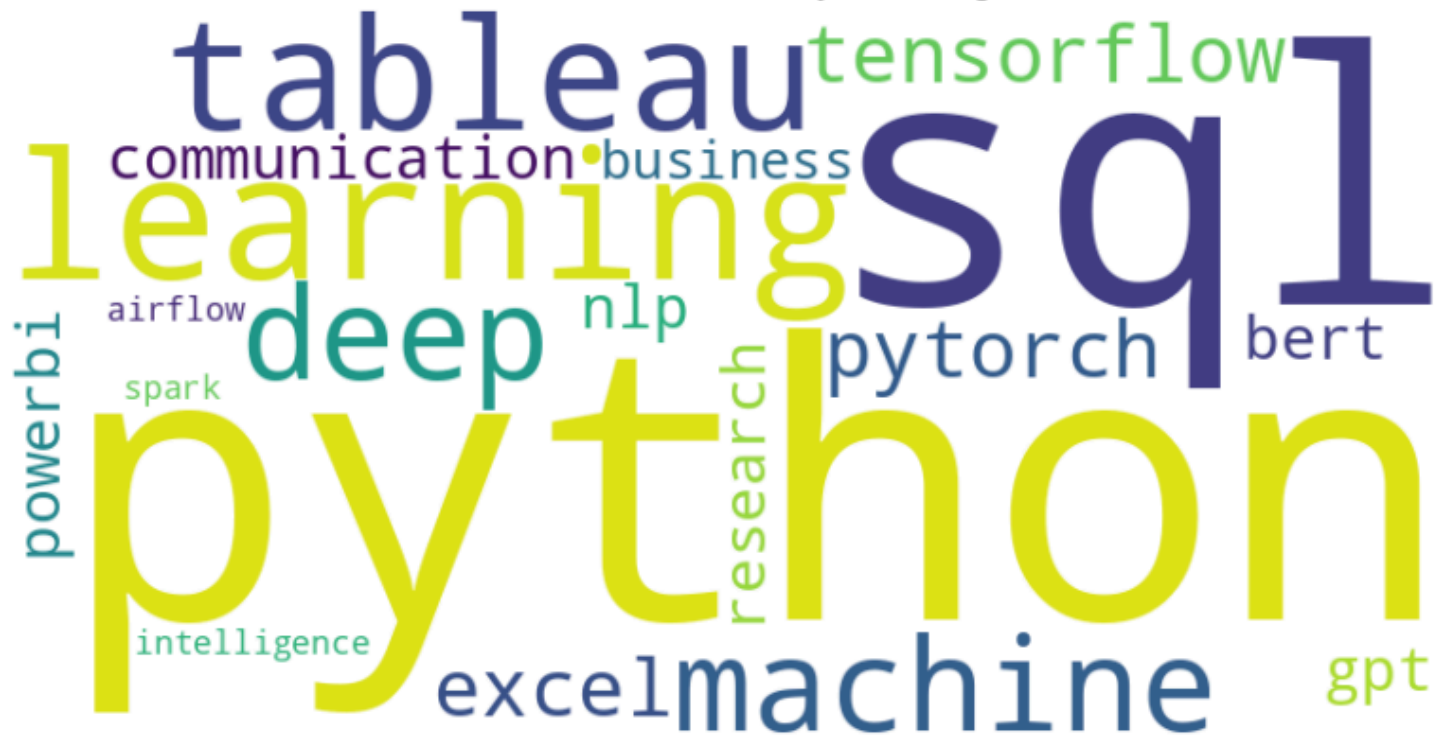
```
Location
Bangalore    3
Hyderabad    2
Mumbai       1
Pune         1
Chennai      1
Name: count, dtype: int64
```



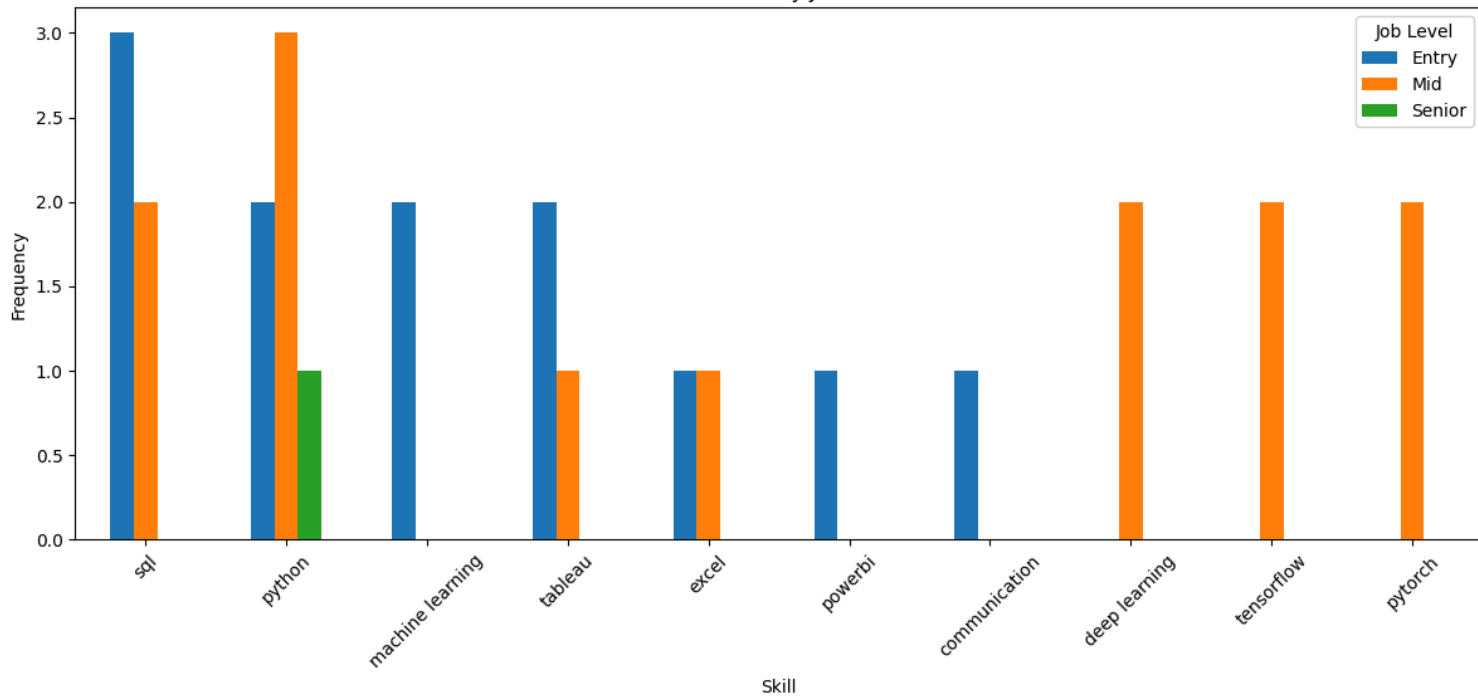
Number of Job Postings

python, sql, machine learning, tableau, python, deep learning, tensorflow, pytorch, excel, sql, p

Most Common Skills in Job Postings



Skill Trends by Job Level



Skill Demand Comparison Across Companies

