Trends in Job Market

Problem 01. Identify the top 5 most frequently required skills in Amazon job postings

```
df=data.copy()
df = df.groupby('Required Skills') .count()['Job ID'].sort_values(ascending=False)
df
```

Required Skills	Job ID
Java, Python, SQL, AWS	7
Python, TensorFlow, Keras, AI	5
AWS, Docker, Kubernetes, Terraform	4
Python, SQL, Hadoop, Spark	4
React, JavaScript, HTML, CSS	4
AWS, Azure, Docker, Kubernetes	3
Python, R, SQL, Machine Learning	3

✓ 0s completed at 12:13 AM

Problem 02. Determine the top 3 locations with the highest number of job postings

```
[12] df=data.copy()
     df = df.groupby('Location') .count()['Job ID'].sort_values(ascending=False)
     df
```

Location	Job ID
Remote	52
San Francisco, CA	26
Seattle, WA	22

dtype: int64

Trends in Job Market

Problem 03. Find the job title with the highest number of postings and its count

```
df=data.copy()
df = df.groupby('Job Title') .count()['Job ID'].sort_values(ascending=False)
df
```

Job Title	Job ID
Backend Developer	8
Data Scientist	7
Security Engineer	7
Security Analyst	6
Product Manager	6
Product Designer	6
Frontend Developer	6
Machine Learning Engineer	5

✓ 0s completed at 12:19 AM

Problem 04. Analyze the month with the highest number of job postings

```
df=data.copy()
df['month'] = pd.DatetimeIndex(df['Posting Date']).month
df = df.groupby('month') .count()['Job ID'].sort_values(ascending=False)
df
```

month	Job ID
1	31
3	31
2	27
4	11

dtype: int64

Trends in Job Market

Problem 05. Filter the job postings requiring Python as a skill and count them

```
python_job = df[df['Required Skills'].str.contains('Python', na=False)]
python_job_count = python_job.shape[0]
print(f"\nNumber of Jobs Requiring Python: {python_job_count}")
```

```
Number of Jobs Requiring Python: 50
```

Problem 06. Analyze the distribution of job postings by company department

```
department_distri = df['Department'].value_counts()
print("\nDistribution of Job Postings by Company Department:")
print(department_distri)
```

```
Distribution of Job Postings by Company Department:
Department
Engineering         45
Security            13
Data Science        11
Design              10
Cloud Computing      8
Product              6
HR                   6
Marketing            1
Name: count, dtype: int64
```

Problem 07. Find the average salary range for each job title in the dataset

```
df['Salary Range'] = pd.to_numeric(df['Salary Range'], errors='coerce')
avg_salary = df.groupby('Job Title')['Salary Range'].mean()
print("\nAverage Salary Range for Each Job Title:")
print(avg_salary)
```

```
Average Salary Range for Each Job Title:
Job Title
AI Engineer                   NaN
AI Researcher                 NaN
Backend Developer             NaN
Backend Engineer              NaN
Cloud Architect               NaN
Cloud Engineer                NaN
Cloud Solutions Architect     NaN
Data Analyst                  NaN
Data Engineer                 NaN
Data Scientist                NaN
DevOps Engineer               NaN
Frontend Developer            NaN
Frontend Engineer             NaN
HR Manager                    NaN
HR Specialist                 NaN
Machine Learning Engineer     NaN
Mobile App Developer          NaN
Product Designer              NaN
Product Manager               NaN
Product Marketing Manager     NaN
Security Analyst              NaN
Security Engineer             NaN
Software Engineer             NaN
UI Designer                   NaN
UI/UX Designer                NaN
Name: Salary Range, dtype: float64
```

Trends in Job Market

Problem 08. Identify job titles that are most commonly associated with remote work

```
[32] remote_job = df[df['Remote'] == 'Yes']
     remote_job_titles = remote_job['Job Title'].value_counts()
     print("\nJob Titles Most Commonly Associated with Remote Work:")
     print(remote_job_titles)
```

```
Job Titles Most Commonly Associated with Remote Work:
Series([], Name: count, dtype: int64)
```

Problem 09. Calculate the correlation between job location and required skills

```
[33] location_encoded = pd.get_dummies(df['Location'])
     skill_encoded = pd.get_dummies(df['Required Skills'])
     correlation_df = pd.concat([location_encoded, skill_encoded], axis=1)
     correlation_matrix = correlation_df.corr()
     print("\nCorrelation Matrix:")
     print(correlation_matrix)
```

```
Correlation Matrix:
                                                Remote  San Francisco, CA  \
Remote                                        1.000000          -0.616953
San Francisco, CA                            -0.616953           1.000000
Seattle, WA                                  -0.552771          -0.314880
AWS, Azure, Docker, Kubernetes                0.051628           0.029402
AWS, Azure, GCP, Terraform                    0.096561          -0.059574
AWS, Azure, Kubernetes, Terraform            -0.104608          -0.059574
AWS, Azure, Terraform, Kubernetes            -0.104608           0.169555
AWS, Docker, Kubernetes, Terraform           -0.212459           0.111687
AWS, Docker, Terraform, Kubernetes            0.096561          -0.059574
AWS, GCP, Azure, Kubernetes                  -0.148690          -0.084678
AWS, Python, Security, Encryption            -0.104608          -0.059574
AWS, Security, Python, Encryption            -0.005719           0.078165
Agile, Scrum, Data Analysis, Product Management -0.104608       -0.059574
Agile, Scrum, Data Analytics, Product Management  0.096561      -0.059574
Agile, Scrum, Product Management, Data Analysis   0.051628      -0.104243
Figma, Adobe XD, UI Design, UX Design         0.096561          -0.059574
Figma, Sketch, UI Design, UX Design          -0.104608           0.169555
HR, Employee Relations, Recruitment           0.096561          -0.059574
HR, Leadership, Communication                 0.137253          -0.084678
HR, Recruitment, Communication                0.137253          -0.084678
HR, Recruitment, Employee Relations           0.096561          -0.059574
Java, Python, AWS, SQL                        0.096561          -0.059574
Java, Python, Ruby, SQL                       0.096561          -0.059574
Java, Python, SQL, AWS                        0.185140          -0.073269
JavaScript, React, HTML, CSS                 -0.065708           0.029402
Marketing, Digital Marketing, SEO, SEM        0.096561          -0.059574
Product Management, Data Analysis, Agile      0.096561          -0.059574
Python, AWS, Network Security, Encryption    -0.104608           0.169555
```

✓ 0s completed at 12:42 AM

Trends in Job Market

Problem 10. Predict job posting trends for the upcoming quarter using time-series analysis

```
time_series = df['Posting Date'].value_counts().sort_index()
time_series.plot(title='Job Postings Over Time', figsize=(10, 6))
plt.xlabel('Date')
plt.ylabel('Number of Postings')
plt.show()
```