

Data Analyst Job Market Analysis



Steps

- 1.Data Cleaning
- 2.Data Analysis
- 3.Data Visualization

Obective

- Import and clean the Data set using pandas in python
- Analyze data trends and distributions using summary statistics and visualizations in python.
- Create plot like bar plot scatter plot using Matplotlib and Seaborn.
- Extract key trends and patterns, such as top industries paying the highest salary.

Importing Libraries:

```
import pandas as pd  
import numpy as np  
import seaborn as sns  
import matplotlib.pyplot as plt
```

Data Analysis:



```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2253 entries, 0 to 2252
Data columns (total 16 columns):
 #   Column           Non-Null Count  Dtype  
 ---  --  
 0   Unnamed: 0        2253 non-null   int64  
 1   Job Title         2253 non-null   object  
 2   Salary Estimate   2253 non-null   object  
 3   Job Description   2253 non-null   object  
 4   Rating            2253 non-null   float64 
 5   Company Name      2252 non-null   object  
 6   Location           2253 non-null   object  
 7   Headquarters       2253 non-null   object  
 8   Size               2253 non-null   object  
 9   Founded            2253 non-null   int64  
 10  Type of ownership 2253 non-null   object  
 11  Industry           2253 non-null   object  
 12  Sector              2253 non-null   object  
 13  Revenue             2253 non-null   object  
 14  Competitors         2253 non-null   object  
 15  Easy Apply          2253 non-null   object  
dtypes: float64(1), int64(2), object(13)
memory usage: 281.8+ KB
```

```
#Average Salary by industry
df.groupby("Industry")["Average Salary"].mean().sort_values(ascending=False)
```

Industry	Average Salary
Drug & Health Stores	95250.00000
Education Training Services	92833.33333
Health Care Products Manufacturing	89800.00000
Sports & Recreation	88166.66667
Gambling	88000.00000
...	
Casual Restaurants	48300.00000
Oil & Gas Services	46375.00000
Grocery Stores & Supermarkets	41500.00000
Trucking	38500.00000
Audiovisual	36000.00000

```
Name: Average Salary, Length: 89, dtype: float64
```

Data Analysis:

```
#Print first 5 row of the dataset
df.head()
```

	Unnamed: 0	Job Title	Salary Estimate	Job Description	Rating	Company Name	Location	Headquarters	Size	Founded	Type of ownership	Industry
0	0	Data Analyst, Center on Immigration and Justice...	37K–66K (Glassdoor est.)	Are you eager to roll up your sleeves and harn...	3.2	Vera Institute of Justice\n3.2	New York, NY	New York, NY	201 to 500 employees	1961	Nonprofit Organization	Social Assistance
1	1	Quality Data Analyst	37K–66K (Glassdoor est.)	Overview\n\nProvides analytical and technical ...	3.8	Visiting Nurse Service of New York\n3.8	New York, NY	New York, NY	10000+ employees	1893	Nonprofit Organization	Health Care Services & Hospitals
2	2	Senior Data Analyst, Insights & Analytics Team...	37K–66K (Glassdoor est.)	We're looking for a Senior Data Analyst who ha...	3.4	Squarespace\n3.4	New York, NY	New York, NY	1001 to 5000 employees	2003	Company - Private	Internet
3	3	Data Analyst	37K–66K (Glassdoor est.)	Requisition NumberRR-0001939\nRemote:Yes\nWe c...	4.1	Celerity\n4.1	New York, NY	McLean, VA	201 to 500 employees	2002	Subsidiary or Business Segment	IT Services
4	4	Reporting Data Analyst	37K–66K (Glassdoor est.)	ABOUT FANDUEL GROUP\n\nFanDuel Group is a w...	3.9	FanDuel\n3.9	New York, NY	New York, NY	501 to 1000 employees	2009	Company - Private	Sports & Recreation

Data Cleaning:

Remove Null Value and Duplicate

```
#Remove null value from Company Name column  
df['Company Name'] = df['Company Name'].fillna('Unknown')  
  
df.isnull().sum()
```

```
Unnamed: 0          0  
Job Title         0  
Salary Estimate   0  
Job Description   0  
Rating            0  
Company Name      0  
Location          0  
Headquarters      0  
Size              0  
Founded           0  
Type of ownership 0  
Industry          0  
Sector             0  
Revenue            0  
Competitors        0  
Easy Apply        0  
dtype: int64
```

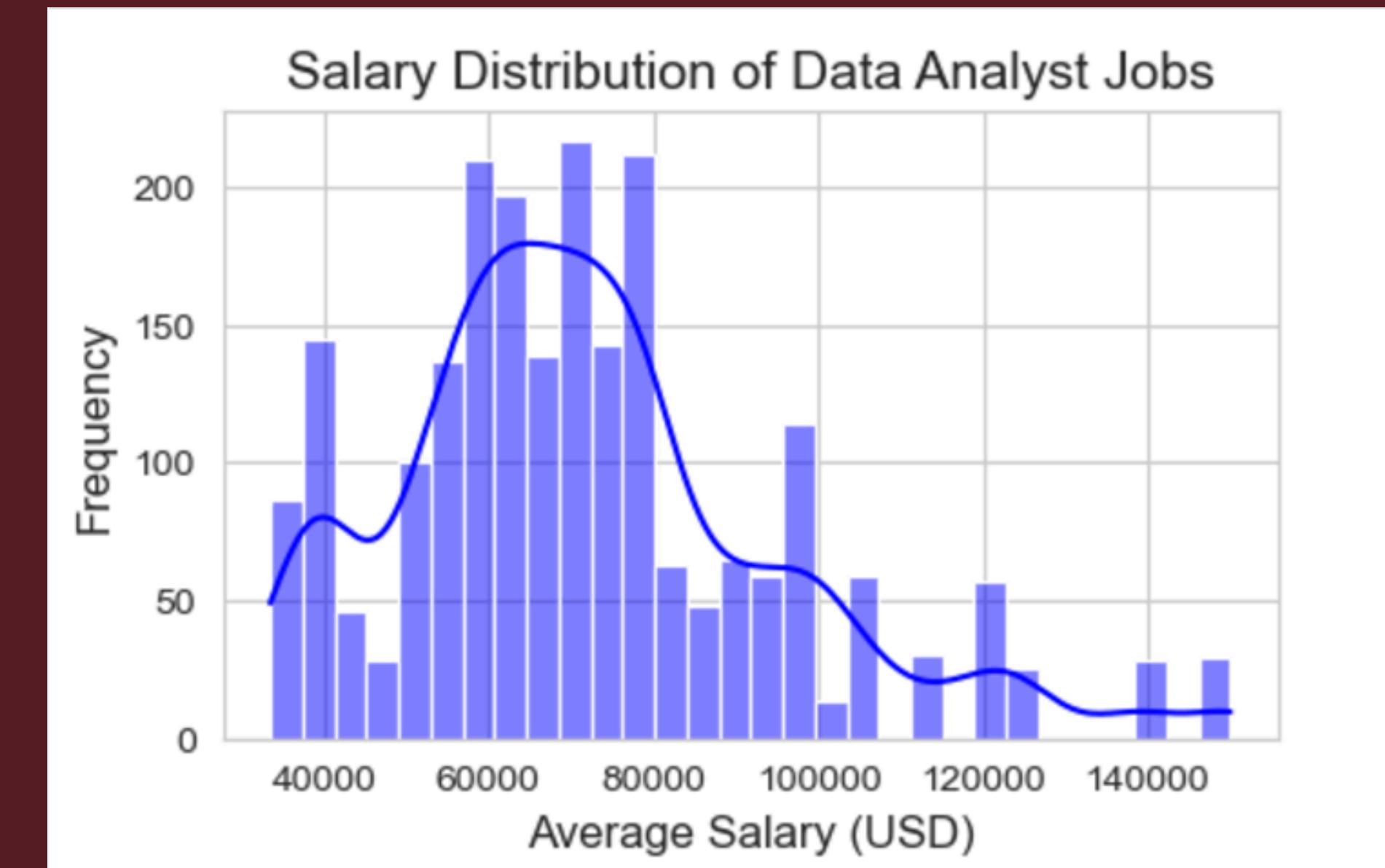
```
#Checking duplicate value  
df.duplicated()
```

```
0      False  
1      False  
2      False  
3      False  
4      False  
...  
2248   False  
2249   False  
2250   False  
2251   False  
2252   False  
Length: 2253, dtype: bool
```

Data Visualization:

```
# 1. Salary Distribution
plt.figure(figsize=(5, 3))
sns.histplot(df["Average Salary"].dropna(), bins=30, kde=True, color="blue")
plt.title("Salary Distribution of Data Analyst Jobs", fontsize=14)
plt.xlabel("Average Salary (USD)", fontsize=12)
plt.ylabel("Frequency", fontsize=12)
plt.show()
```

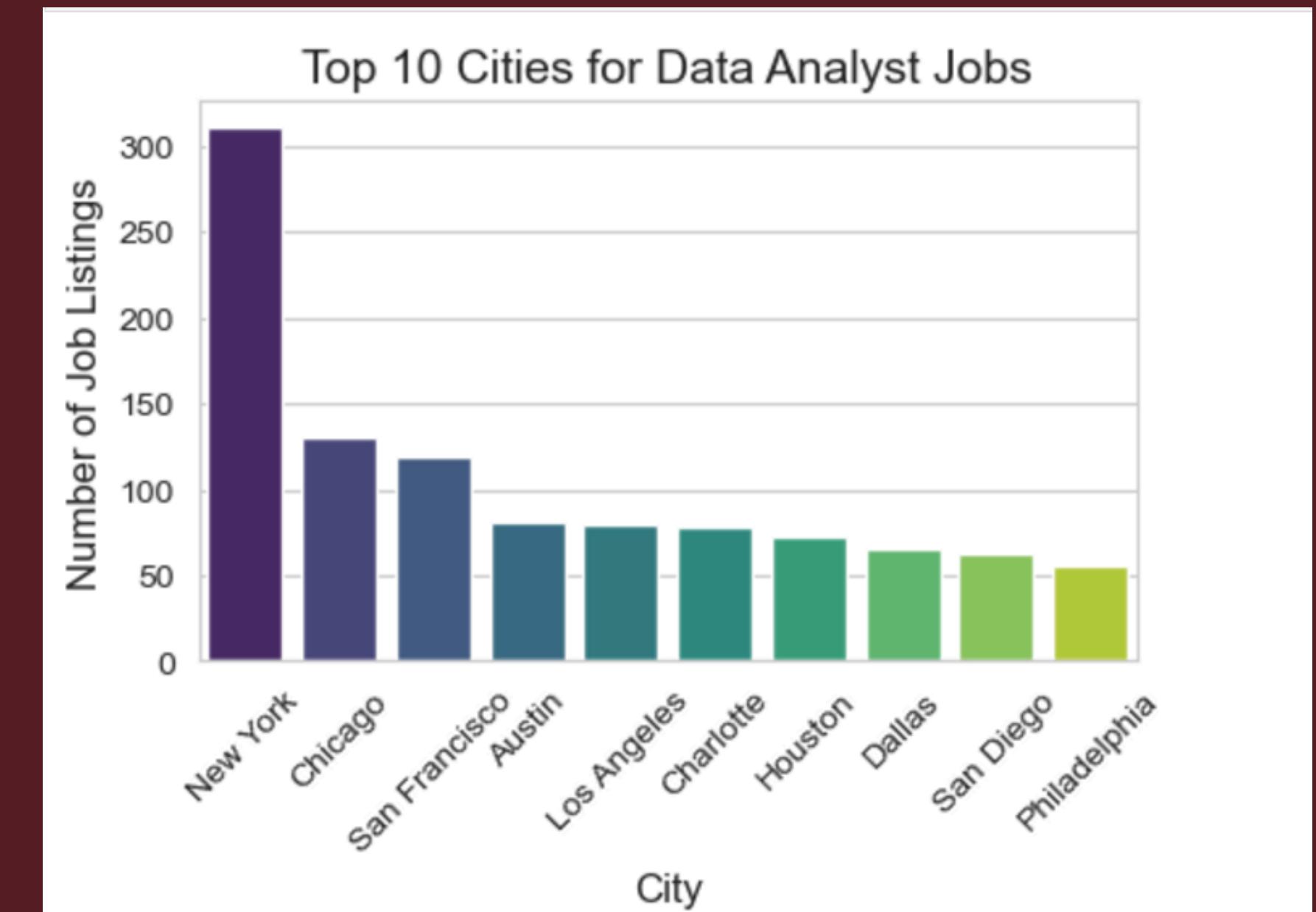
Salary Distribution
of Data Analyst
Jobs



Data Visualization:

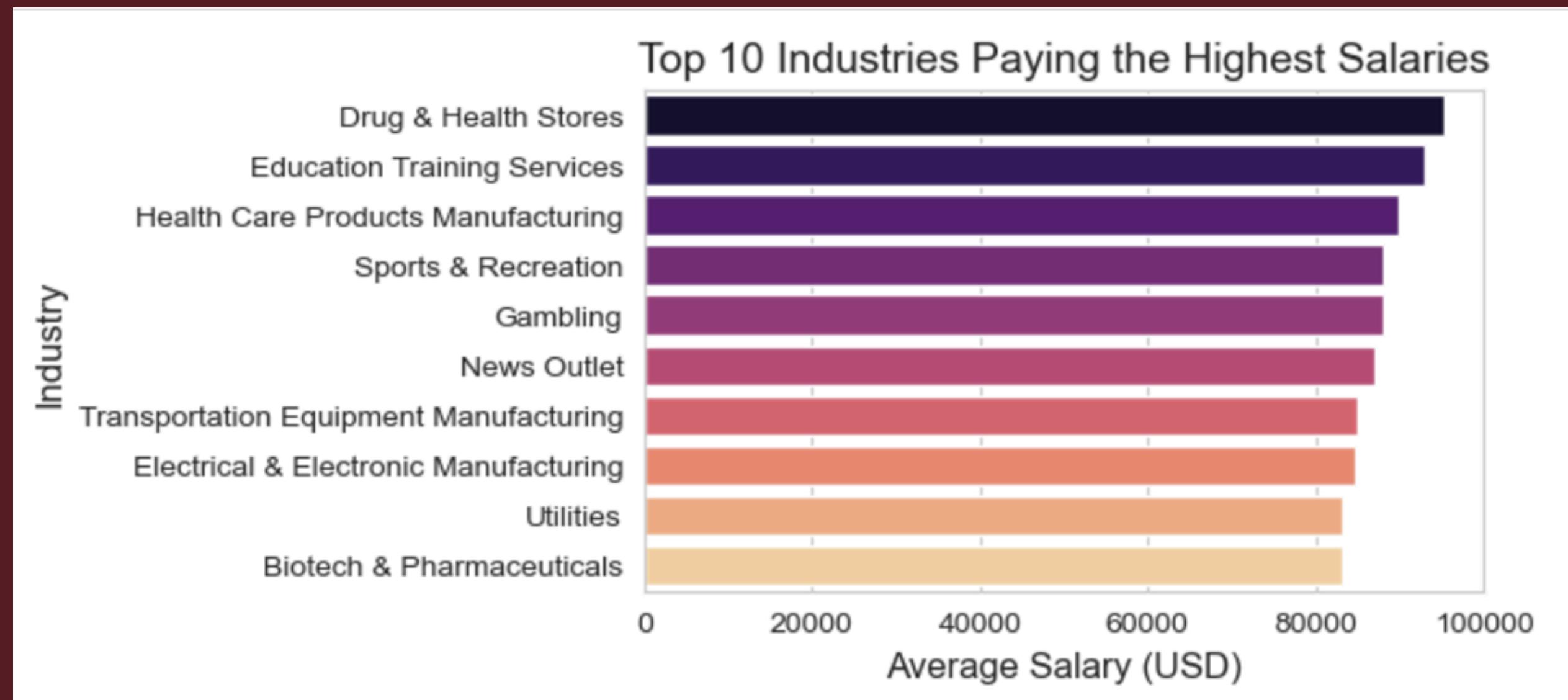
```
# 2. Top 10 Cities with Most Job Listings
plt.figure(figsize=(5, 3))
top_cities = df["City"].value_counts().head(10)
sns.barplot(x=top_cities.index, y=top_cities.values, hue = top_cities.index, palette = 'viridis', legend = False)
plt.title("Top 10 Cities for Data Analyst Jobs", fontsize=14)
plt.xlabel("City", fontsize=12)
plt.ylabel("Number of Job Listings", fontsize=12)
plt.xticks(rotation=45)
plt.show()
```

Top 10 Cities For Data Analyst Jobs



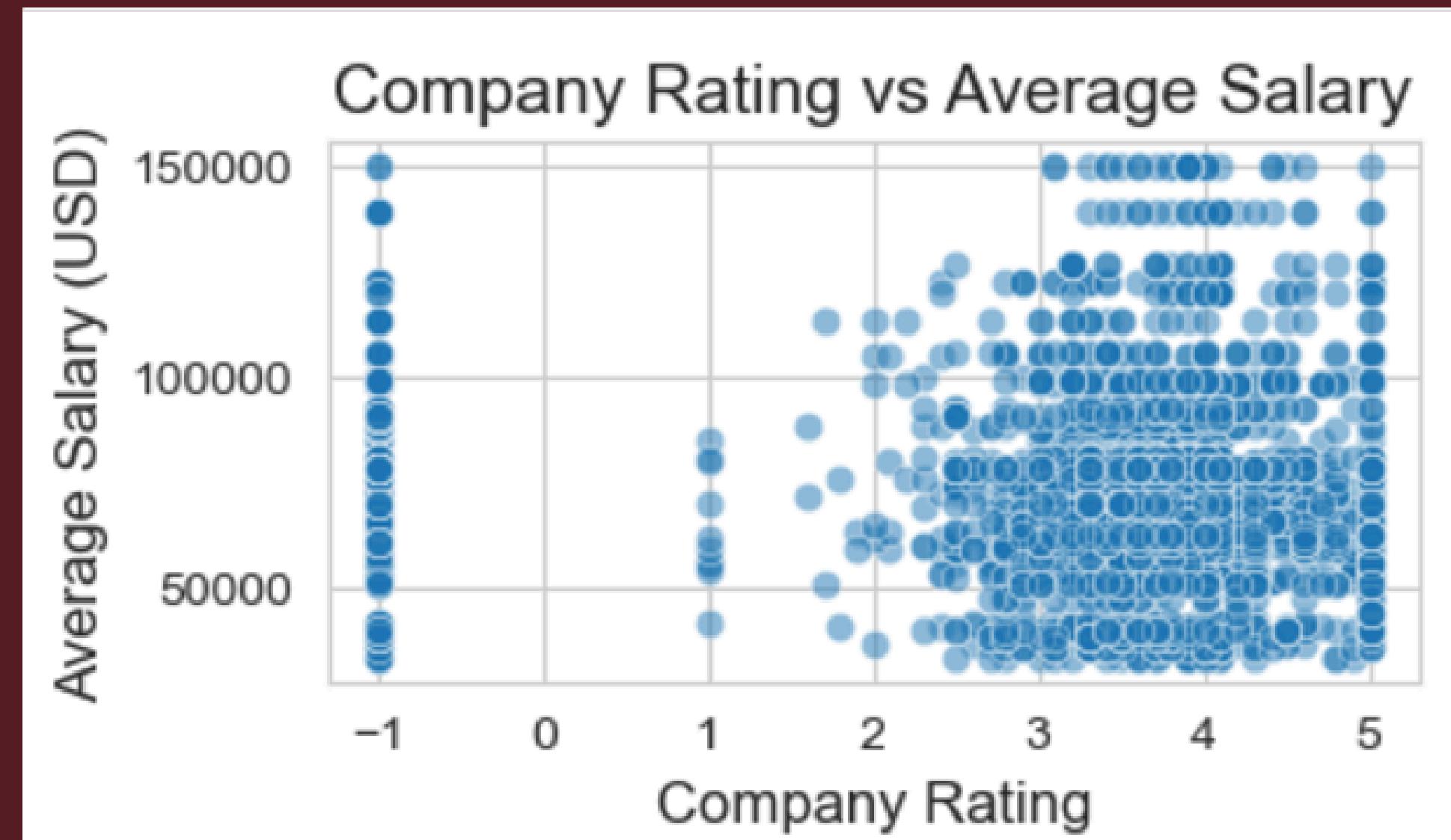
Data Visualization:

```
# 3. Average Salary by Industry (Top 10)
plt.figure(figsize=(5, 3))
top_industries = df.groupby("Industry")["Average Salary"].mean().sort_values(ascending=False).head(10)
sns.barplot(x=top_industries.values, y=top_industries.index, hue = top_industries.index, palette="magma")
plt.title("Top 10 Industries Paying the Highest Salaries", fontsize=14)
plt.xlabel("Average Salary (USD)", fontsize=12)
plt.ylabel("Industry", fontsize=12)
plt.show()
```



Data Visualization:

```
# 5. Relationship Between Company Rating & Salary
plt.figure(figsize=(4, 2))
sns.scatterplot(data=df, x="Rating", y="Average Salary", alpha=0.5)
plt.title("Company Rating vs Average Salary", fontsize=14)
plt.xlabel("Company Rating", fontsize=12)
plt.ylabel("Average Salary (USD)", fontsize=12)
plt.show()
```



Key Takeaways:

1. The average salary distribution shows most data analyst positions pay between 50,000 - 100,000.
2. New York dominates as the city with the most data analyst job posting.
3. The Biotech and pharmaceutical sector offers the highest average salaries.
4. Many top rated companies are technical service and recruiting firms.