Job Market Trends - Data Analyst Portfolio Project

By Jannath Syed

```
import pandas as pd

df = pd.read_csv('/content/Job_data_project.csv')
df.head()
```

	job_title	location	salary	sector
0	IT Support Technician Job in Madison	Madison, WI 53702	NaN	IT/Software Development
1	Business Reporter/Editor Job in Madison	Madison, WI 53708	NaN	NaN
2	Johnson & Johnson Family of Companies Job Appl...	DePuy Synthes Companies is a member of Johnson...	NaN	NaN
3	Engineer - Quality Job in Dixon	Dixon, CA	NaN	Experienced (Non-Manager)
4	Shift Supervisor - Part-Time Job in Camphill	Camphill, PA	NaN	Project/Program Management

Next steps: Generate code with df | View recommended plots | New interactive sheet

check the number of rows and columns

```
print("Rows and columns:", df.shape)
```

```
Rows and columns: (22000, 4)
```

check the column names

```
print("\ncolumn names:")
print(df.columns)
```

```

column names:
Index(['job_title', 'location', 'salary', 'sector'], dtype='object')
```

see how many missing values are in each column

```
print("\nmissing values:")
print(df.isnull().sum())
```

```

missing values:
job_title        0
location         0
salary       18554
sector        5194
dtype: int64
```

Create a new DataFrame without missing salary or sector

```
clean_df = df.dropna(subset=['salary', 'sector'])
```

show how many rows are left

```
print("cleaned data size:", clean_df.shape)
```

```
cleaned data size: (2612, 4)
```

preview the cleaned data

```
clean_df.head()
```

	job_title	location	salary	sector
13	Primrose Private Preschool Teacher Job in Houston	Houston, TX 77098	9.00 - 13.00 $ /hour	Entry Level
14	Superintendent Job in Houston	Houston, TX	80,000.00 - 95,000.00 $ /year	Building Construction/Skilled Trades
19	Technician - Robot & Multi-Axis CNC Field Serv...	Carter Lake, IA 51510	60,000.00 - 72,000.00 $ /year	Experienced (Non-Manager)
29	Sr. Process Engineer	Sr. Process Engineer, Manufacturing	70,000.00 - 100,000.00 $ /year	Engineering
30	Machine Control Systems Engineer - BSEE Job	Aurora, IL 60502	75,000.00 - 100,000.00 $	Experienced (Non-Manager)

Next steps: Generate code with clean_df | View recommended plots | New interactive sheet

Visualize most common job sectors

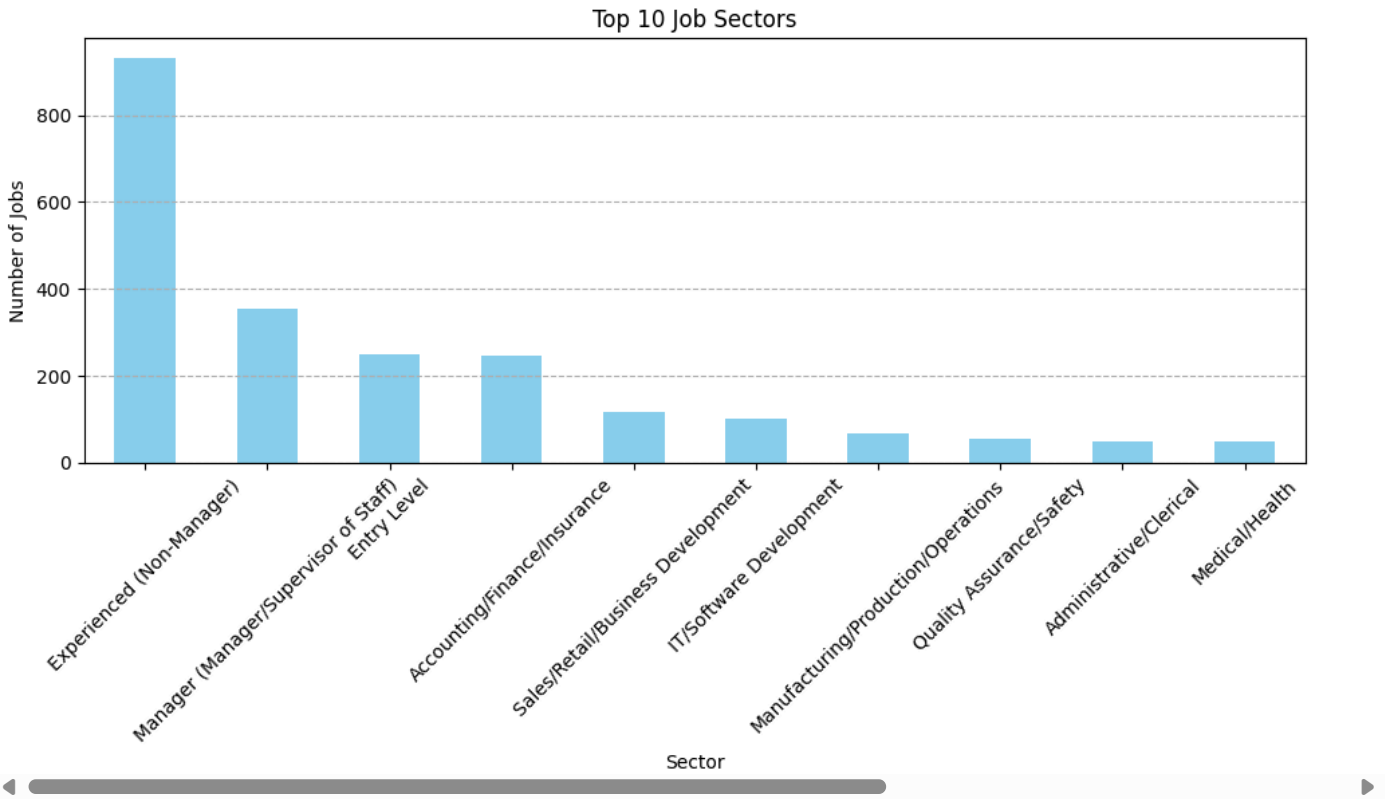
```
import matplotlib.pyplot as plt
```

count job listings by sector

```
sector_counts = clean_df['sector'].value_counts().head(10)
```

Plot the top 10 sectors

```
plt.figure(figsize=(10,6))
sector_counts.plot(kind='bar', color='skyblue')
plt.title('Top 10 Job Sectors')
plt.xlabel('Sector')
plt.ylabel('Number of Jobs')
plt.xticks(rotation=45)
plt.grid(axis='y', linestyle='--')
plt.tight_layout()
plt.show()
```



Visualize most common job titles

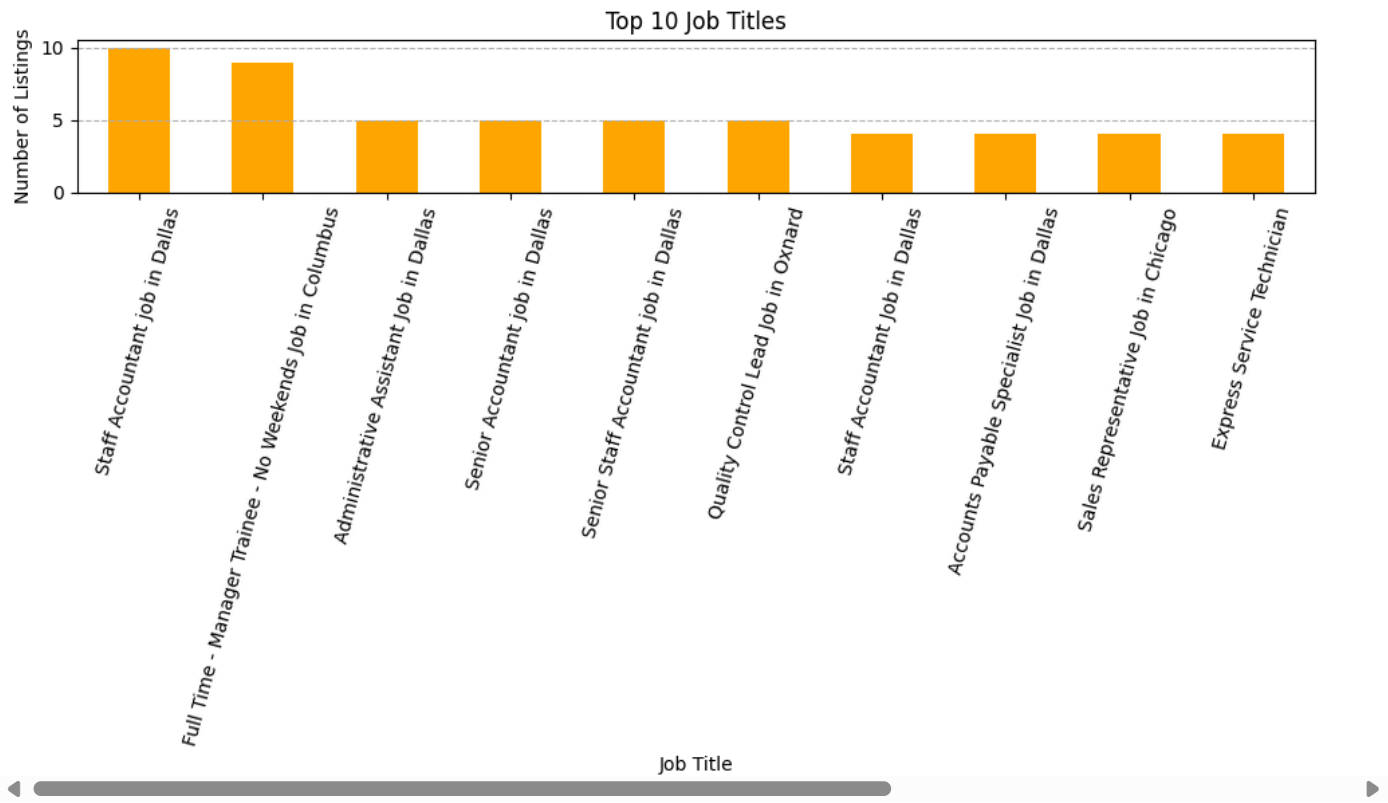
```
import matplotlib.pyplot as plt
```

count most common job titles

```
job_counts = clean_df['job_title'].value_counts().head(10)
```

plot the top 10 job titles

```
plt.figure(figsize=(10,6))
job_counts.plot(kind='bar', color='orange')
plt.title('Top 10 Job Titles')
plt.xlabel('Job Title')
plt.ylabel('Number of Listings')
plt.xticks(rotation=75)
plt.grid(axis='y', linestyle='--')
plt.tight_layout()
plt.show()
```



Analyze Salary Data

let's check if we can work with the salary column (it might have text like $60,000$ — 80,000 /year). We'll first look at what kind of data it has.

Show 10 sample salary values

```
clean_df['salary'].head(10)
```

	salary
13	9.00 - 13.00 $ /hour
14	80,000.00 - 95,000.00 $ /year
19	60,000.00 - 72,000.00 $ /year
29	70,000.00 - 100,000.00 $ /year
32	75,000.00 - 100,000.00 $ /year
36	68,000.00 - 72,000.00 $ /year
41	58,000.00 - 65,000.00 $ /year
42	Up to $32000.00
58	Salary, plus commission
61	45,000.00 - 100,000.00 $ /yearBonus, Benefits,...

dtype: object

Check Salary Data Format

Show 10 random salary values

```
clean_df['salary'].sample(10)
```

	salary
4861	42,000.00 - 45,000.00 $ /year
17417	16.00+ /hour
903	16.00 - 18.00 $ /hour
15852	70,000.00 - 90,000.00 $ /year
13104	competitive pay
5737	10.00 - 20.00 $ /hourHealth insurance available
21814	70,000.00 - 110,000.00 $ /year
13814	11.00 - 16.00 $ /hour
10605	95,000.00 - 100,000.00 $ /year
4834	82,064.00 - 106,683.00 $ /year

dtype: object

Save a cleaned data to a CSV file

```
clean_df.to_csv('cleaned_job_data.csv', index=False)
```

Download it to your computer

```
from google.colab import files
files.download('cleaned_job_data.csv')
```

Summary

and **Description of the Project**

Job Market Analysis Using Python

By Jannath Syed

Objective:

To analyze a dataset of 22,000+ U.S. job listings, identify trends in job titles and sectors, and build insights using Python and Excel.

Tools Used:

- Python (Pandas, Matplotlib)
- Google Colab
- Excel (for initial filtering)
- GitHub (for version control)
- CSV data from Kaggle

Steps Taken:

- Downloaded a real-world job listings dataset from Kaggle
- Opened and explored the data using Excel
- Cleaned the dataset by removing rows with missing salary and sector values
- Analyzed the structure and contents of the dataset using Python

- Visualized:
    - Top 10 most common *job titles*
    - Top 10 most common *job sectors*
- Investigated the salary column and explained why it could not be analyzed due to inconsistent formatting
- Exported a cleaned version of the dataset into a new CSV file