

LinkedIn[®] Job Posting 2023 Dataset

Data Source: This dataset from [Kaggle](#) contains a record of **33,000+ job postings** listed over the course of 2 days.

Motivation: LinkedIn is the most well-known social media site used to post various employment opportunities making it a good source to analyze

- Amount of job postings for a specific job types/industries
- Offered salary ranges for job postings
- Locations with highest job postings for specific job types/industries

This analysis will be covering job postings related specifically to **data analysis jobs**.



Dataset Analysis 1

Original dataset is 33,246 rows and 28 columns.

Filtering specifically for roles containing “**Data Analyst**” in the title brought up 91 job postings.

After some data cleaning (filtering & replacing values), **the top 5 states with the majority of “Data Analyst” job postings were:**

1. CA - California
2. TX - Texas
3. VA - Virginia
4. NC - North Carolina
5. FL - Florida

```
#Since there are just 3 rows with this discrepancy for the job title we are searching for, we can update using loc
df2.loc[df2['city'] == 'California', 'state'] = ' CA'
df2.loc[df2['city'] == 'Hawaii', 'state'] = ' HI'
df2.loc[df2['city'] == 'Maryland', 'state'] = ' MD'
```

```
[ ] # Checking to see if values have been successfully updated
df2.loc[[9858, 16897, 31082]]
```

| | job_id | company_id | title | description | max_salary | med_salary | min_salary | pay_period | formatted_v |
|-------|------------|------------|----------------------------------|--|------------|------------|------------|------------|-------------|
| 9858 | 3756112944 | 1088079.0 | Senior Data Analyst | The Company: Sungrow Power Supply Co., Ltd. ("S... | NaN | NaN | NaN | NaN | |
| 16897 | 3748844912 | 34237.0 | Data Analyst | The position of Data Processing Systems Analys... | NaN | NaN | NaN | NaN | |
| 31082 | 3693051072 | 10577525.0 | Power BI Data Analyst – (Remote) | Power BI Data Analyst – (Remote) CARIAN is gro... | NaN | NaN | NaN | NaN | |

```
[ ] #Filtering top 5 states with the most job postings containing 'Data Analyst' in the title
df2.groupby(['state'])['state'].count().sort_values(ascending=False).head(5)
```

```
state
CA    10
TX     8
VA     6
NC     6
FL     4
Name: state, dtype: int64
```

Dataset Analysis 2

For job postings with “**Data Analysts**” in the title:

- 24 out of the 91 roles indicated that remote work was allowed for hired employees.
- 27 roles made mention of the program “Python” in the job description.
 - For a majority of these roles (17 in total) requested Python for Mid-Senior Level job positions

```
[ ] # Of the 91 job postings, only 24 are remote
df2.groupby(['remote_allowed'])['remote_allowed'].count().sort_values(ascending=False)
```

```
remote_allowed
1.0      24
Name: remote_allowed, dtype: int64
```

```
# Filtering for 'Python' mentioned in job description
df_pyth= df2[df2["description"].str.contains("Python", na=False)]
```

```
df_pyth.shape
```

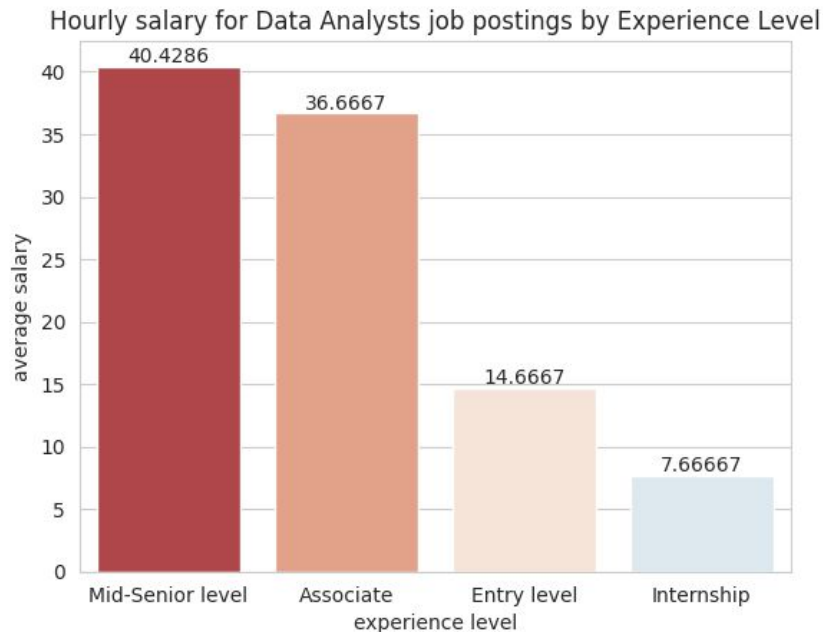
```
(27, 19)
```

```
[ ] #How many of these job descriptions mention 'Python'?
df_pyth.groupby(['experience_level'])['experience_level'].count().sort_values(ascending=False)
```

```
experience level
Mid-Senior level    17
Entry level         5
Associate           1
Name: experience level, dtype: int64
```

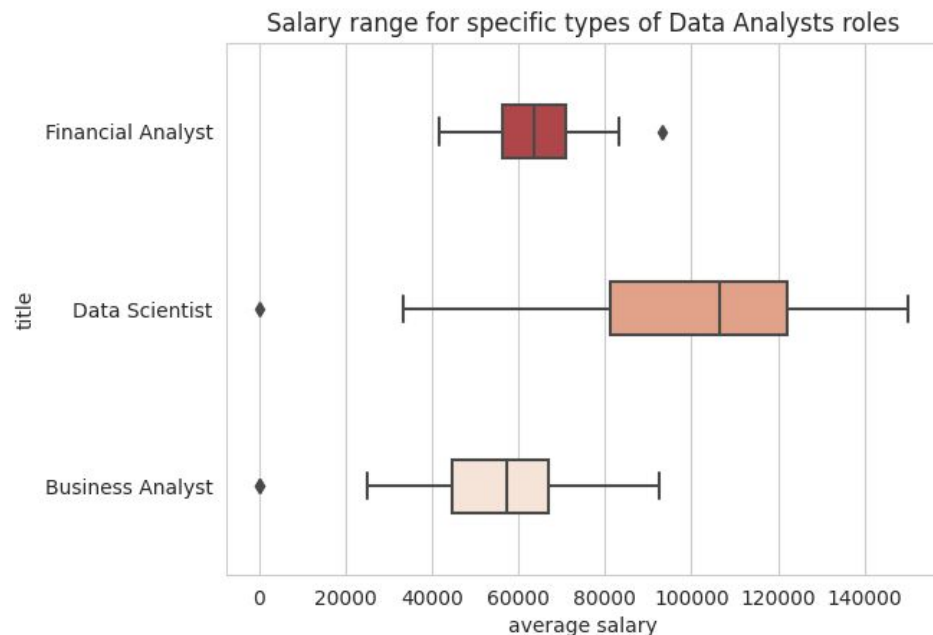
Data Visualization 1

- The dataset included columns for “max_salary”, “min_salary”, and “med_salary”.
- Many values were NaN values that I had to replace with 0.
- I also changed the column type from object to int64 and created a new column which was the average of all 3 numbers for each job posting range.
- The bar chart showcases the average hourly salary for Data Analyst job postings in the data set by experience level.



Data Visualization 2

- The dataset was then filtered for 3 specific types of Data Analyst roles: Financial Analyst, Data Scientist, and Business Analysts.
- The box plot chart showcases the average salary range for all job postings containing these roles in the title column.



The End.