

---

---

# Impact of Experience Level And Job Title on Salary

Computer and Technology Group 1

---

---

GurSimran K Sujlana

---

---

# Outline

---

---

1. Source of data set
  2. Problem or challenge
  3. Data dictionary
  4. Descriptive analysis
    - Summary metrics
    - Descriptive plots
    - Use Probability Theory & Bayes Theorem on your data set
  5. Summary
  6. Conclusions and Insights
- 
-

Source of DataSet

kaggle

# Problem / Challenge

**What we want to know:** What influences salary more, Experience level or Job Title

The dataset classifies the experience level of employees ranging from “Entry-Level” to “Executive”.

The dataset also gives specific job titles within the data field such as ‘Data Scientist’ or ‘Data Engineer’.

Our objective is to determine whether experience level or job title has a greater impact on salary.

In the future we hope to be able to leverage predictive modeling to see which feature (experience level or job title) influences the salary the most.

# Data Dictionary

Field Name	Data Type	Field Length	Description
work_year	Integer	4	The year in which the data was recorded.
job_title	String	40	The specific title of the job role.
job_category	String	30	A classification of the job role.
salary_currency	String	3	The currency in which the salary is paid.
salary	Integer	6	The annual gross salary of the role in the local currency.
salary_in_usd	Integer	6	The annual gross salary converted to United States Dollars (USD).
employee_residence	String	24	The country of residence of the employee.
experience_level	String	11	Classifies the professional experience level of the employee.

## Data Dictionary (continued)

Field Name	Data Type	Field Length	Description
employment_type	String	9	Specifies the type of employment.
work_setting	String	9	The work setting or environment.
company_location	String	24	The country where the company is located.
company_size	String	1	The size of the employer company, often categorized into small (S), medium (M), and large (L) sizes.

# Summary Metrics: Experience Level

Summary by Experience Level:

experience_level	Count	Mean	Std	Min	25%	Median	75%	Max	Skewness	Kurtosis
Entry-level	496	88534.8	49102.1	15000	51726	80000	120000	281700	1.03048	1.0189
Executive	281	189463	68793	15000	140000	185000	235000	416000	0.367679	-0.102954
Mid-level	1869	117524	55453.6	15000	75000	110000	149600	450000	1.27149	3.25047
Senior	6709	162356	59523	18381	122600	155000	198800	412000	0.629989	0.607562

The **Mid-level experience** category exhibits the **highest skewness (1.27149)** and **kurtosis (3.25047)** among the groups, indicating that **salaries are largely concentrated on the lower end with a right-skewed distribution**, and there's a pronounced **presence of outliers** with more extreme salary values.

# Summary Metrics: Job Title

Summary by Job Title (Top 10):

job_title	Count	Mean	Std	Min	25%	Median	75%	Max	Skewness	Kurtosis
Analytics Engineer	256	155239	55607.5	37573	116920	149400	185175	430640	0.8867	2.27501
Applied Scientist	272	190172	50196.3	20000	136000	192000	222200	350000	0.155353	0.0788192
Business Intelligence Engineer	144	151405	52944.3	43064	104300	156400	185225	259000	-0.0899781	-1.06411
Data Analyst	1388	109911	42994.1	15000	80000	105320	135000	430967	1.10128	4.15142
Data Architect	213	164061	56105.9	52500	120000	159500	192564	376080	0.90724	1.37404
Data Engineer	2195	146620	56643.6	18000	106800	140000	180000	385000	0.581873	0.324287
Data Scientist	1989	156681	59914.4	16000	120000	154800	190000	412000	0.390596	0.601277
Machine Learning Engineer	991	184786	61760.6	20000	142200	182200	220000	392000	0.21831	0.0371832
Research Engineer	144	182840	68469.4	16455	139750	169056	226250	385000	0.686569	0.618615
Research Scientist	269	184376	68479	23000	144000	175000	220000	450000	0.686175	0.917388

The **Data Scientist** role exhibits a **moderate kurtosis** (0.601277), which is **higher than some other roles like Machine Learning Engineer and Applied Scientist**, indicating a **slightly more peaked distribution** with the potential for more outliers compared to a normal distribution, but less so than the Data Analyst role.

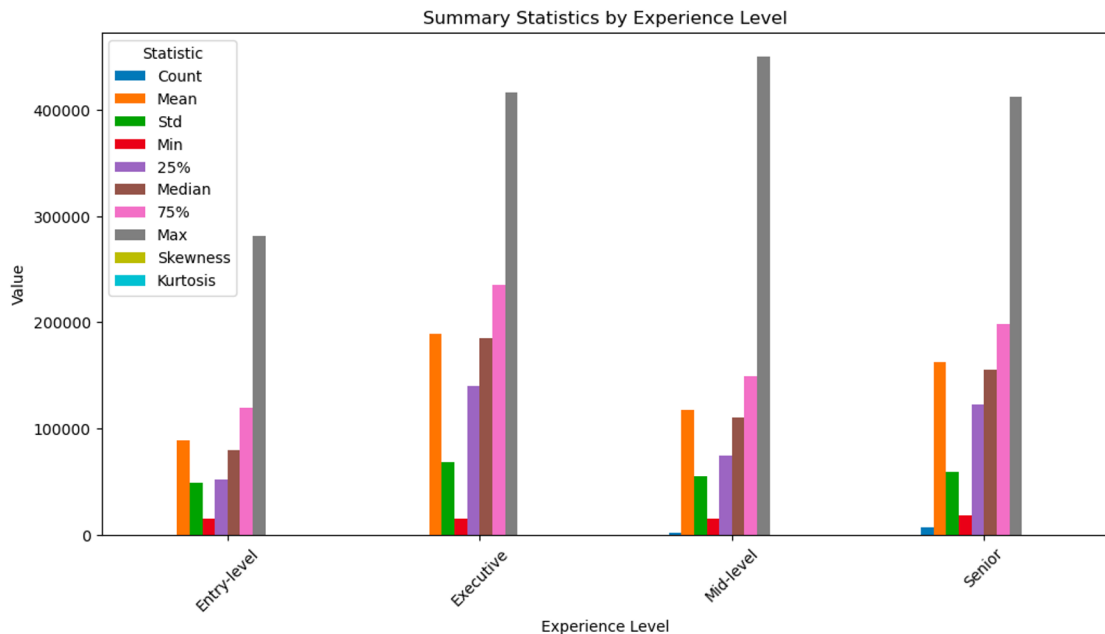


# Summary Metric: Top 10 Salary Comparisons

	Executive	Mid-level	Senior	salary_in_usd
job_title				
AI Architect	1	0	11	250328.000000
AWS Data Architect	0	1	0	258000.000000
Analytics Engineering Manager	0	0	1	399880.000000
Cloud Data Architect	0	0	1	250000.000000
Data Science Tech Lead	0	0	1	375000.000000
Director of Data Science	24	2	3	221365.034483
Head of Data	27	1	5	211186.545455
Head of Machine Learning	2	0	0	259000.000000
Machine Learning Software Engineer	0	2	11	212907.692308
Managing Director Data Science	1	0	0	300000.000000

- No entry level positions
- Only 6 mid level
- 55 exec, 33 senior
- Highest paid jobs have limited data, discrepancies

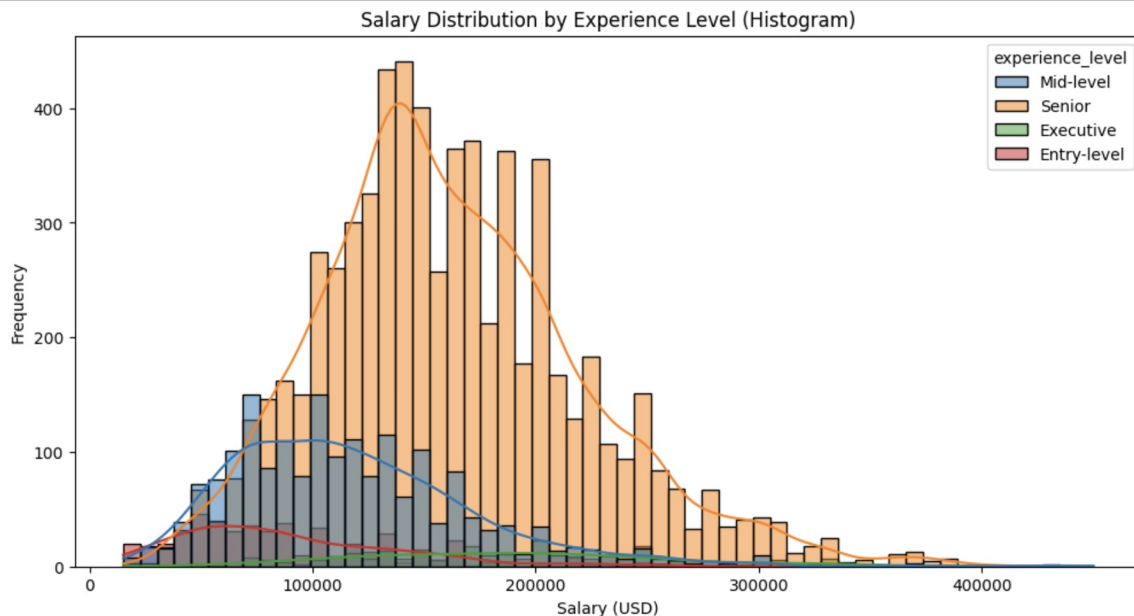
# Descriptive Plots: Experience Level vs Value



- **Salary Progression:** **Entry-level** roles typically have the **lowest salary** statistics, while **mid-level** roles show the **highest**, suggesting a clear salary progression with experience.
- **Statistical Spread:** The spread of salaries (as indicated by standard deviation bars) **increases** with **higher experience levels**, particularly for **Executives**, which have the widest spread.

# Descriptive Plots: Experience Level Salary vs Frequency

- **Diverse Salary Ranges:** Salaries for Mid-level, Senior, and Executive experience levels show **overlapping distributions**, indicating variability within these groups.
- **Senior Prevalence:** Senior roles appear to be the most prevalent in the dataset, with a **wide range of salaries** extending into higher brackets.



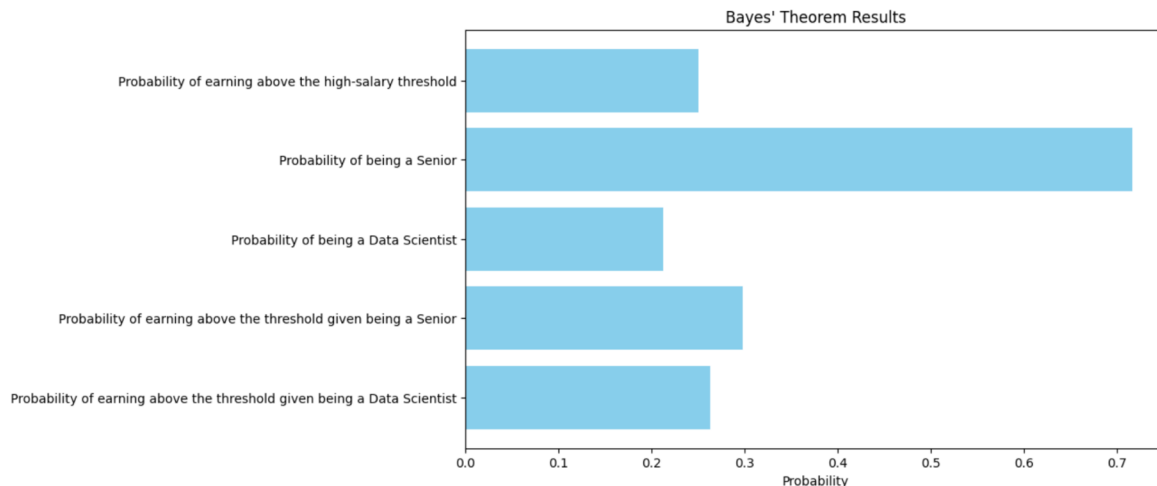
# Baye's Theorm

Bayes' Theorem Results:

	Description	Value
P(A)	Probability of earning above the high-salary threshold	0.250027
P(B1)	Probability of being a Senior	0.717157
P(B2)	Probability of being a Data Scientist	0.212614
P(A B1)	Probability of earning above the threshold given being a Senior	0.297511
P(A B2)	Probability of earning above the threshold given being a Data Scientist	0.262946

- **Earning Above Threshold:** There's a 25% chance overall of earning above the high-salary threshold.
- **Senior Likelihood:** Seniors are more likely to earn above this threshold, with a probability of approximately 30%.
- **Data Scientist Potential:** Data Scientists have around a 26% probability of earning above the threshold.

# Baye's Theorm



**Senior Dominance in High Salary Bracket:** This indicates that in this dataset, experience level (being a Senior) is a strong predictor of high salary.

**Data Scientist Representation:** While Data Scientists have a lower overall representation in the dataset, they also have a noteworthy probability of earning above the high-salary threshold.

**Comparative Analysis of Roles:** When comparing the role of a Senior to that of a Data Scientist, being a Senior is associated with a slightly higher likelihood of earning above the threshold. This supports the hypothesis that experience level may have a more substantial impact on salary than having a specific job title within this dataset.

# Insights

- Applied Scientists are the most likely to earn more than other jobs
- Executive level positions are the most likely to earn more but mid-level positions hold the most potential for the highest individual salary
  - Executive positions have the widest spread of salaries
- Senior level positions are more likely to have a high salary compared to data scientists
- The executive level and senior level average salaries have the smallest difference of around \$27,000

# Conclusion

In summary, a senior level position is more likely to achieve a high salary compared to the growing data scientist job.

If you want to be in the top earners, becoming an experienced applied scientist would be the most effective way of achieving this. This goal can be difficult for executive level positions depending on the job, since it is difficult to predict the salary due to the widespread of values.

Reaching the higher levels of experience are more likely to obtain higher salaries compared to job types. Progressing to the executive position will not have a significant increase on salary compared to the senior level, which is most likely one of the reasons why the it is the most represented position.

Thank you,  
Questions?