In this essay, I have researched several factors affecting wages and explained them through the data visualizations I used. I would like to start by explaining the importance of the dataset I have chosen. By analysing this dataset, we can understand the dynamics affecting wage levels for both individuals and organizations.

It details the methodological analysis of a data set consisting of variables such as Age, Gender, Education Level, Job Title, Years of Experience, and Salary. The main goal here is to find out which factors have a significant impact on salary and how this information can be communicated effectively by visualizing it.

I started the analysis by determining the structure and types of data it contains. There are both numerical variables (Age, Years of Experience and Salary) and categorical variables (Sex, Education Level, Job Title).

```
[ ] import pandas as pd
    import matplotlib.pyplot as plt

    data = pd.read_csv('/content/Salary Data.csv')
```

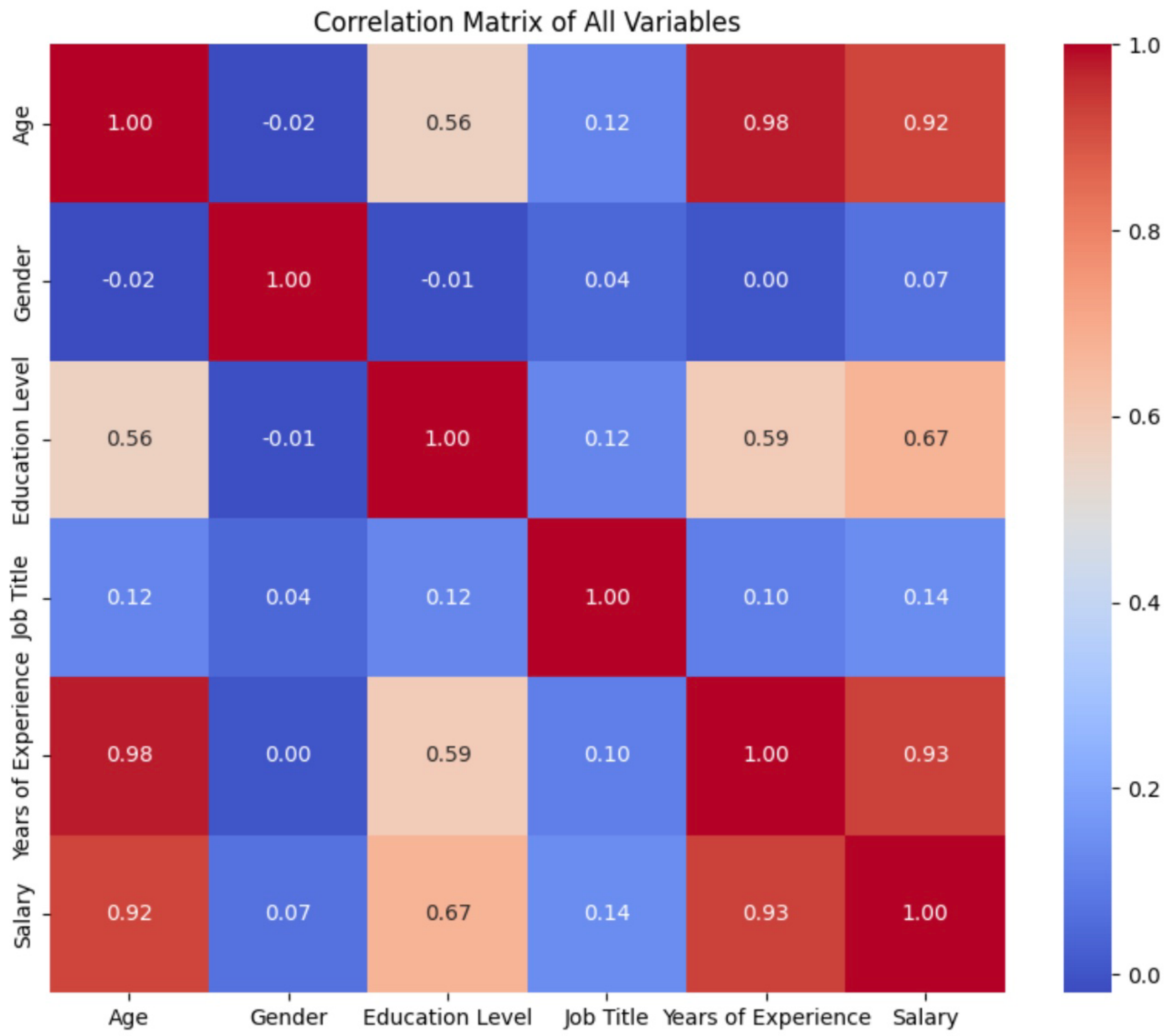
```
[ ] data
```

	Age	Gender	Education Level	Job Title	Years of Experience	Salary
0	32.0	Male	Bachelor's	Software Engineer	5.0	90000.0
1	28.0	Female	Master's	Data Analyst	3.0	65000.0
2	45.0	Male	PhD	Senior Manager	15.0	150000.0
3	36.0	Female	Bachelor's	Sales Associate	7.0	60000.0
4	52.0	Male	Master's	Director	20.0	200000.0
...	...	...	...	...	...	...
370	35.0	Female	Bachelor's	Senior Marketing Analyst	8.0	85000.0
371	43.0	Male	Master's	Director of Operations	19.0	170000.0
372	29.0	Female	Bachelor's	Junior Project Manager	2.0	40000.0
373	34.0	Male	Bachelor's	Senior Operations Coordinator	7.0	90000.0
374	44.0	Female	PhD	Senior Business Analyst	15.0	150000.0

375 rows × 6 columns

I first calculated a correlation matrix to determine how each variable correlated with salary. My first data visualization, the matrix, is the best way to select the dependent factors that influence the main factor. Because here I can easily choose the most influencing variable by clearly seeing the correlation level of each one. This approach is very important to determine which factors have the greatest influence on wages. Here, "Years of Experience" and "Age" were found to have a strong positive correlation with salary. And accordingly, I took them as the main factors to investigate.

```
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Matrix of All Variables')
plt.show()
```

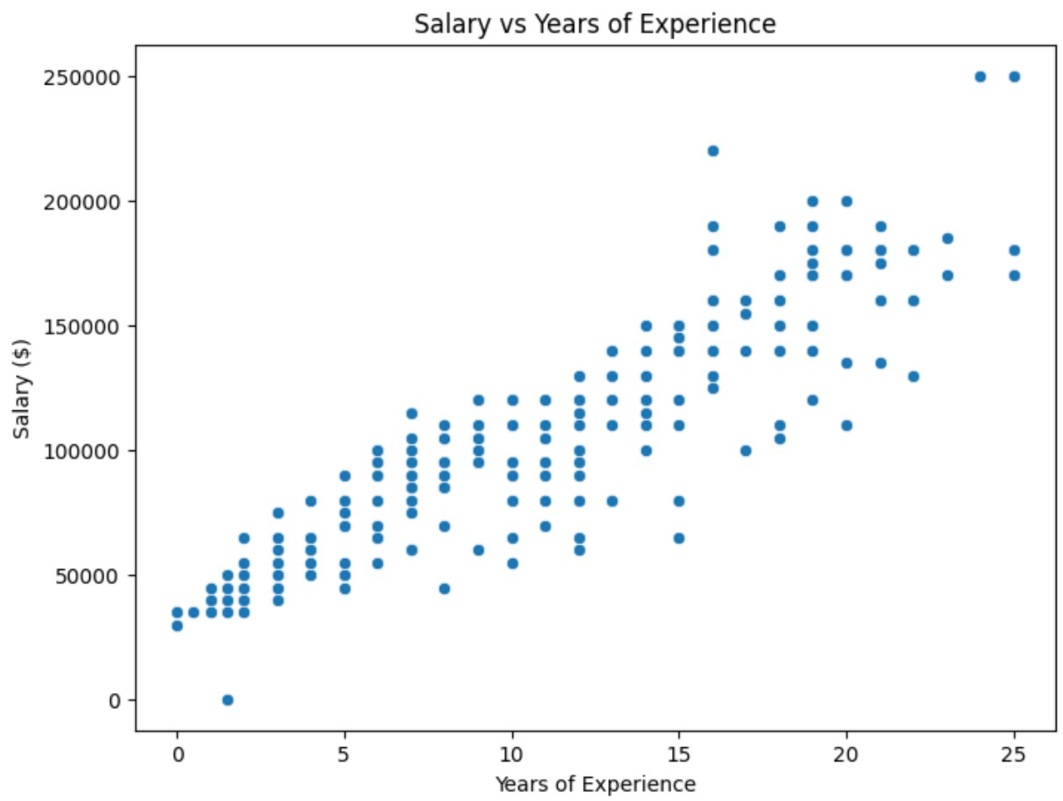


Then, given the 2 strong correlations I selected based on matrix, I chose scatterplots as the ideal way to visualize their relationship with salary. Scatter plots are commonly used to describe the relationship between continuous variables, effectively displaying trends, patterns, and outliers.

I would like to explain why a scatterplot is the best type of visualization for my database, with detailed explanations and examples:

First, it allows us to see everyone's salary relative to their years in the workforce. For example, if most points move straight along the x-axis and gradually move up the y-axis, we can see that salary increases with more years of experience.

In addition to this, it is also effective in showing correlations.
Plotting a scatterplot for "Age" and "Salary," we see a trend in which scores tend to trend upward as age increases—this visual alignment of data points along an upward trajectory clearly indicates a positive correlation. That is, as the age increases, the salary generally increases, and this is probably due to the accumulation of years of experience and the promotion of the employee to higher positions in the company.



One of the reasons why this visualization is the best in my opinion is that it is easy to read. Without needing to understand complex statistical terms, one can observe that more experienced individuals tend to have higher salaries, aiding in straightforward decision-making regarding career planning or compensation strategies from this plot. Because it's important to consider that the dataset I use targets stakeholders from different backgrounds - be it HR managers, job seekers, or executives.

In addition, in the scatter plot we can also identify outliers and anomalies. For example, consider a scatter plot of "Age" versus "Salary" where one data point deviates dramatically from others—perhaps an individual in their early 30s earning significantly more than peers in the same age group. This outlier might prompt further investigation. Perhaps this individual has accelerated their career through exceptional performance or a rare skill set, or they work in a high-paying industry like technology, which pays premium salaries even to relatively younger professionals.

Audience for this data visualization

Now I want to explain who can benefit from these visualizations. First of all, I want to mention HR Professionals and Executives, because this analysis is especially useful for HR managers, compensation analysts and corporate executives involved in developing salary structures, developing talent acquisition strategies and planning workforce development.

In addition, career counselors and educational institutions can also benefit from these visualizations. Because the information gained from this analysis can help guide students and professionals toward promising career paths and educational endeavors that potentially offer higher financial rewards.

The results of the visualization are also of interest to labor market researchers and economists interested in trends affecting wage inequality and employment patterns.

Also, employees and job seekers can use these insights to make detailed decisions about salary negotiations, job opportunities, and career advancement.

This analysis of the salary data set used a correlation matrix and scatterplots to identify key factors influencing wage levels. First, the correlation matrix highlighted that "Years of Experience" and "Age" had the strongest positive correlation with "Salary". These insights guided the later use of scatterplots, which clearly demonstrate how these variables affect wages. The reason for using each visualization method and why it is considered the best visualization method for this sample is explained in detail above. These plots provided an intuitive visual confirmation that both age and professional experience were significant predictors of wage growth.

These visualizations serve as critical tools for a range of stakeholders, including HR professionals, corporate executives and career advisors, who rely on understanding salary dynamics for strategic decision-making and advice. In addition, they offer valuable insights for employees and job seekers planning their career trajectories. Overall, this analysis not only deepens our understanding of the effects of pay, but also demonstrates the power of data visualization in strategic business applications and personal career development.

Appendixe

```
import pandas as pd
import matplotlib.pyplot as plt

data = pd.read_csv('/content/Salary Data.csv')

data['Gender'] = data['Gender'].astype('category').cat.codes
data['Job Title'] = data['Job Title'].astype('category').cat.codes
data['Education Level'] = data['Education
Level'].astype('category').cat.codes


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

```
correlation_matrix = data.corr()
```

```
# Now, plotting the heatmap
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Matrix of All Variables')
plt.show()
```

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

# Load your data
data = pd.read_csv('/content/Salary Data.csv')

plt.figure(figsize=(8, 6))
sns.scatterplot(data=data, x='Years of Experience', y='Salary')
plt.title('Salary vs Years of Experience')
plt.xlabel('Years of Experience')
plt.ylabel('Salary ($)')
plt.show()
```

```
plt.figure(figsize=(8, 6))
sns.scatterplot(data=data, x='Age', y='Salary')
plt.title('Salary vs Age')
plt.xlabel('Age')
plt.ylabel('Salary ($)')
plt.show()
```