PRESENTATION

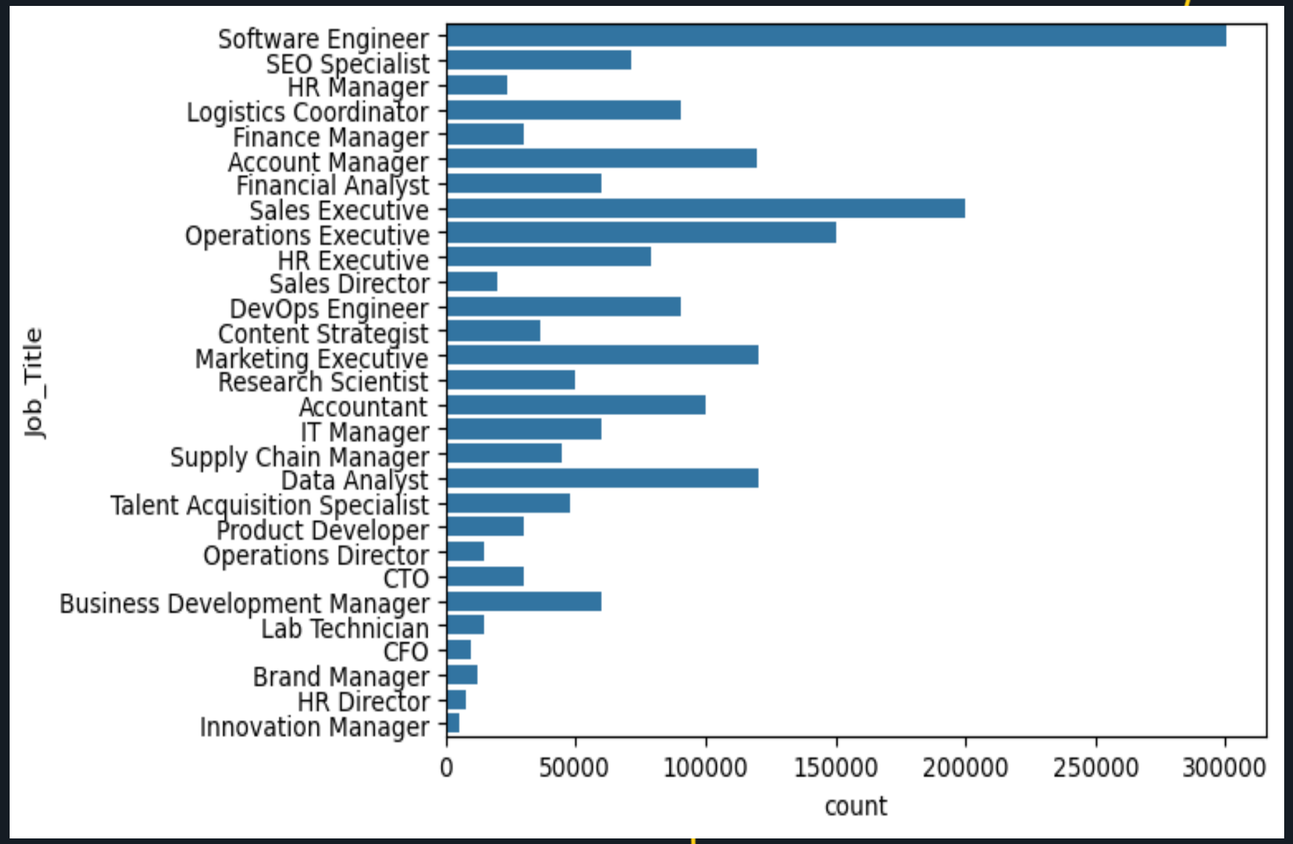
Data Analysis & Visualization

By: Shahd Ahmed Ibrahim

1.HR DATA

1.Distribution of Job Titles:

What the visualization shows:
The frequency of different job titles in the dataset.
Why I used this visualization:
A horizontal bar chart makes it easy to compare the number of employees in each role.
It highlights which job positions are the most common.
Key takeaway:
Software Engineer is by far the most frequent role.
Other roles with high counts include *Sales* Executive and Operations Executive.

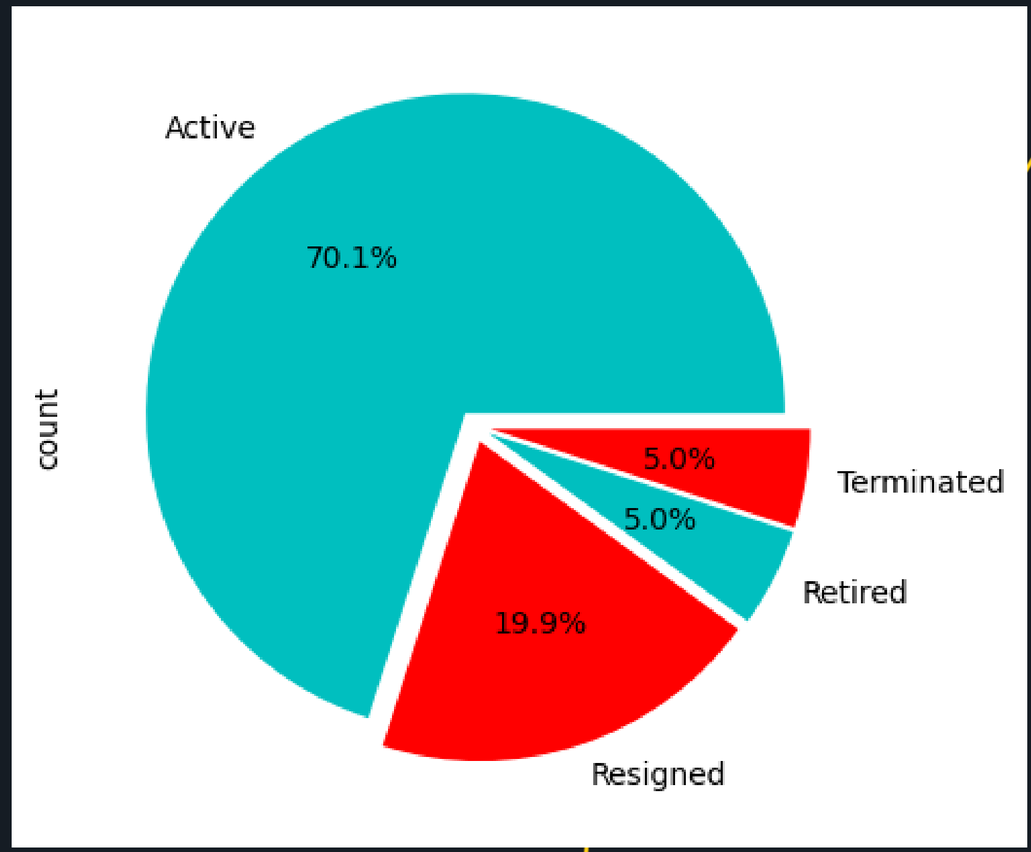


2. Distribution of Employee Status

What the visualization shows:
The proportion of employees by status (e.g., Active, Resigned, Retired).

Why I used this visualization:
A pie chart clearly shows the share of each category and allows quick comparison of employment status across the workforce.

Key takeaway:
The majority of employees are currently active, while only a small fraction are resigned or retired.

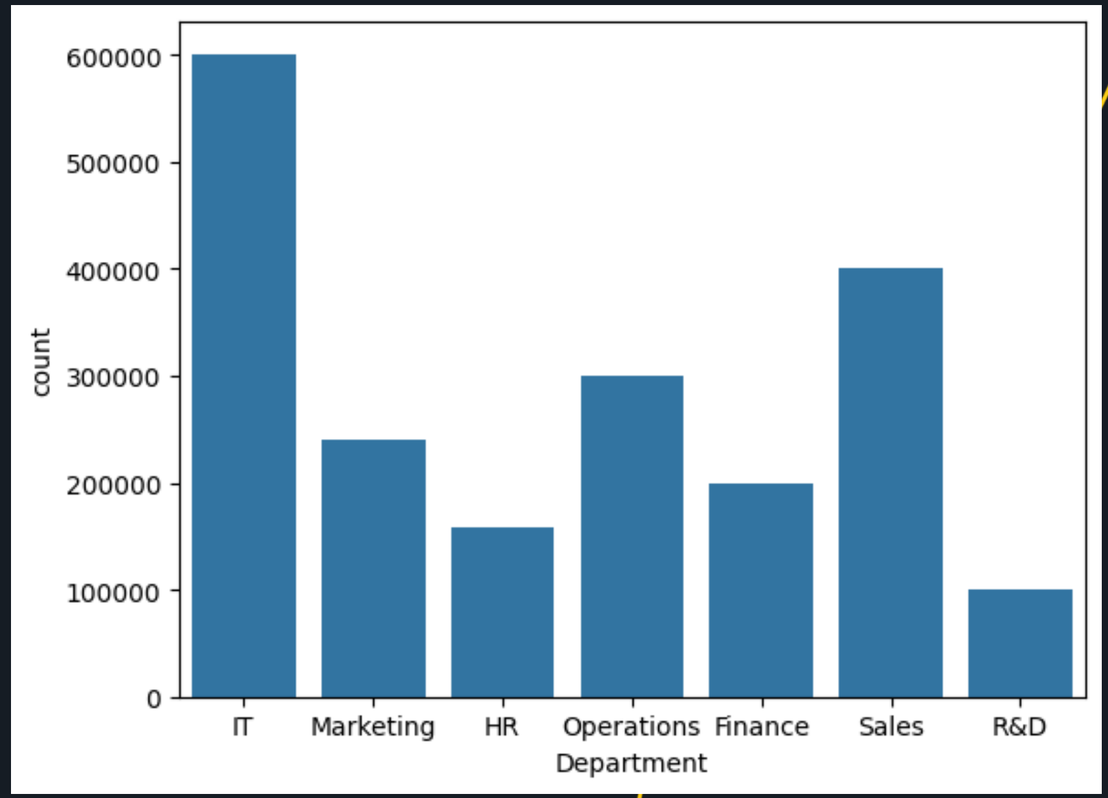


3. Employees by Department

What the visualization shows:
The number of employees in each department.

Why I used this visualization:
A bar chart allows us to compare department sizes and identify where the largest groups of employee's work.

Key takeaway:
IT and Sales departments have the highest employee counts, while smaller departments like HR or R&D employ fewer people.

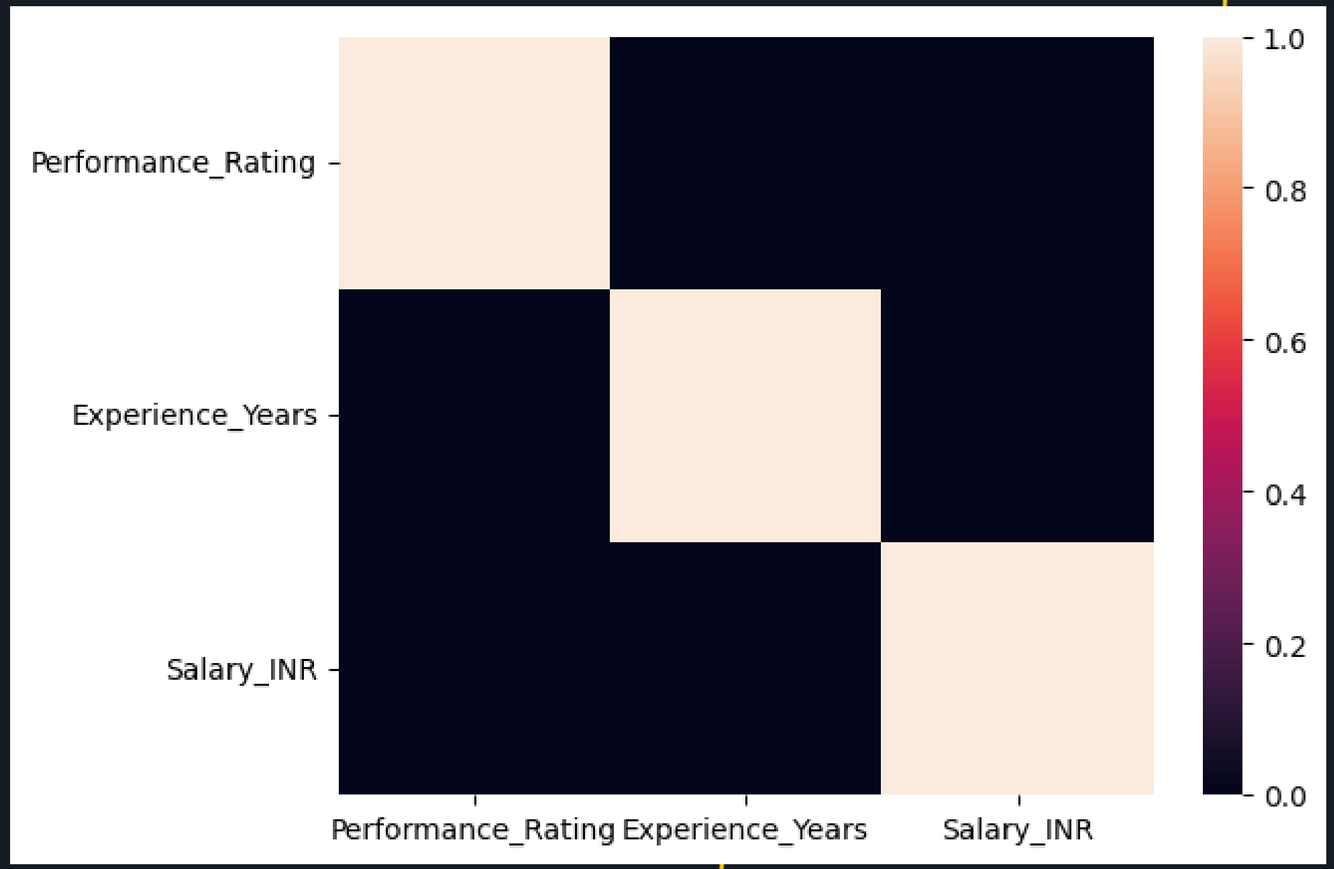


4. Correlation Heatmap

What the visualization shows:
The relationships between numerical variables such as salary, age, performance rating, and years of experience.

Why I used this visualization:
A heatmap helps identify strong positive or negative correlations, which can reveal important patterns such as whether higher experience leads to higher salary.

Key takeaway:
Salary is positively correlated with experience and performance rating, suggesting that both factors influence pay.

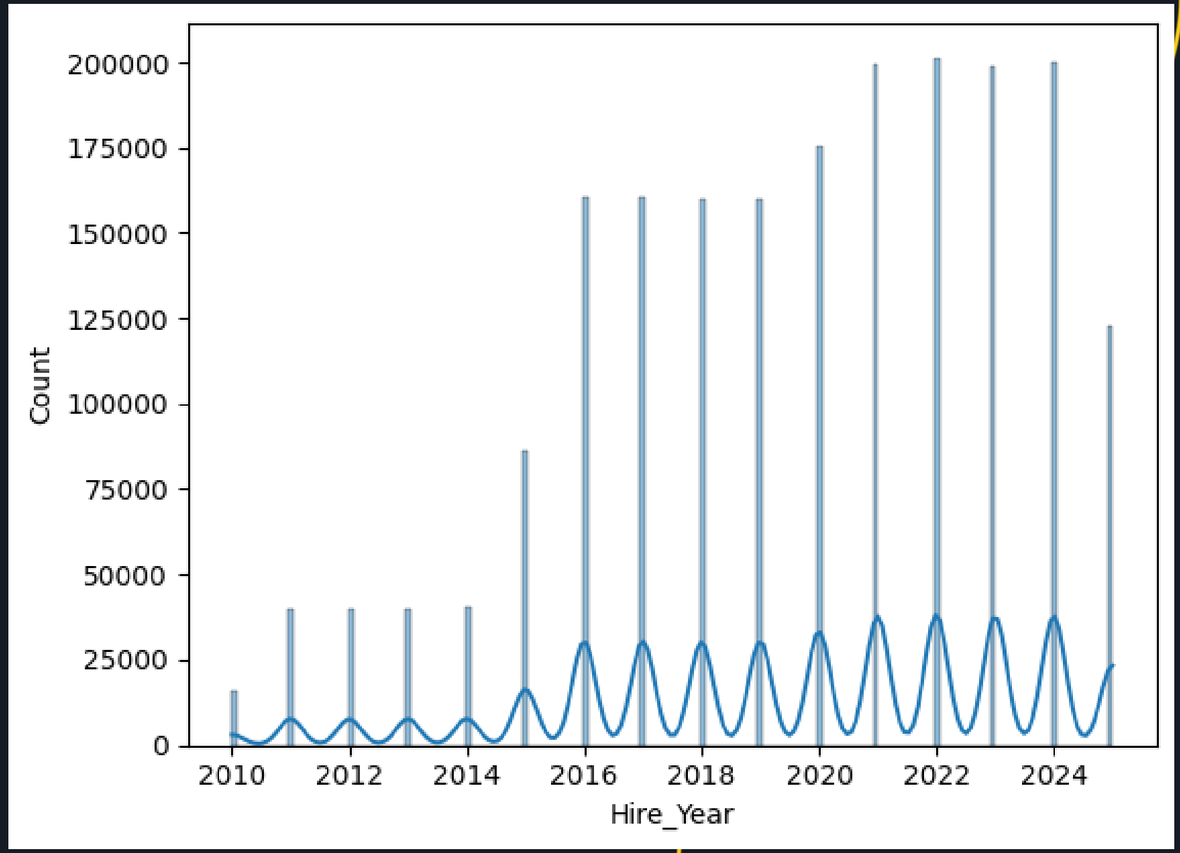


5. Distribution of Hire Years

What the visualization shows:
The frequency of employees hired in each year.

Why I used this visualization:
A histogram highlights hiring trends over time, showing whether recruitment increased or decreased in certain years.

Key takeaway:
Hiring peaked in specific years, indicating periods of major recruitment activity.

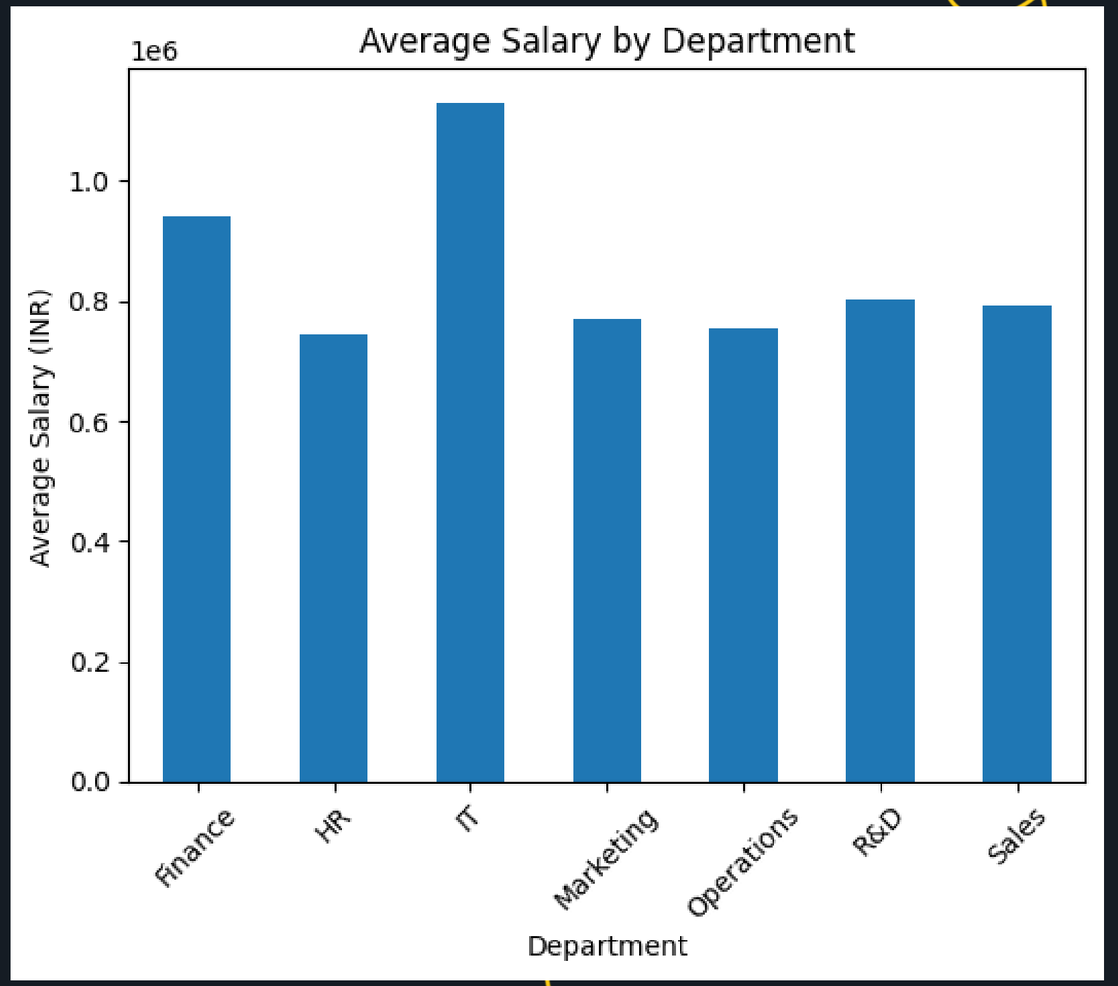


6. Average Salary by Department

What the visualization shows:
The mean salary for employees grouped by department.

Why I used this visualization:
A bar chart makes it easy to compare average salaries across departments and see which fields are more highly paid.

Key takeaway:
Departments such as Finance and R&D have higher average salaries, while departments like Operations and HR have comparatively lower salaries.



7. Salary Distribution

What the visualization shows:
The range of employee salaries in the dataset.

Why I used this visualization:
A box plot shows the salary spread, highlights outliers, and provides insights into pay inequality.

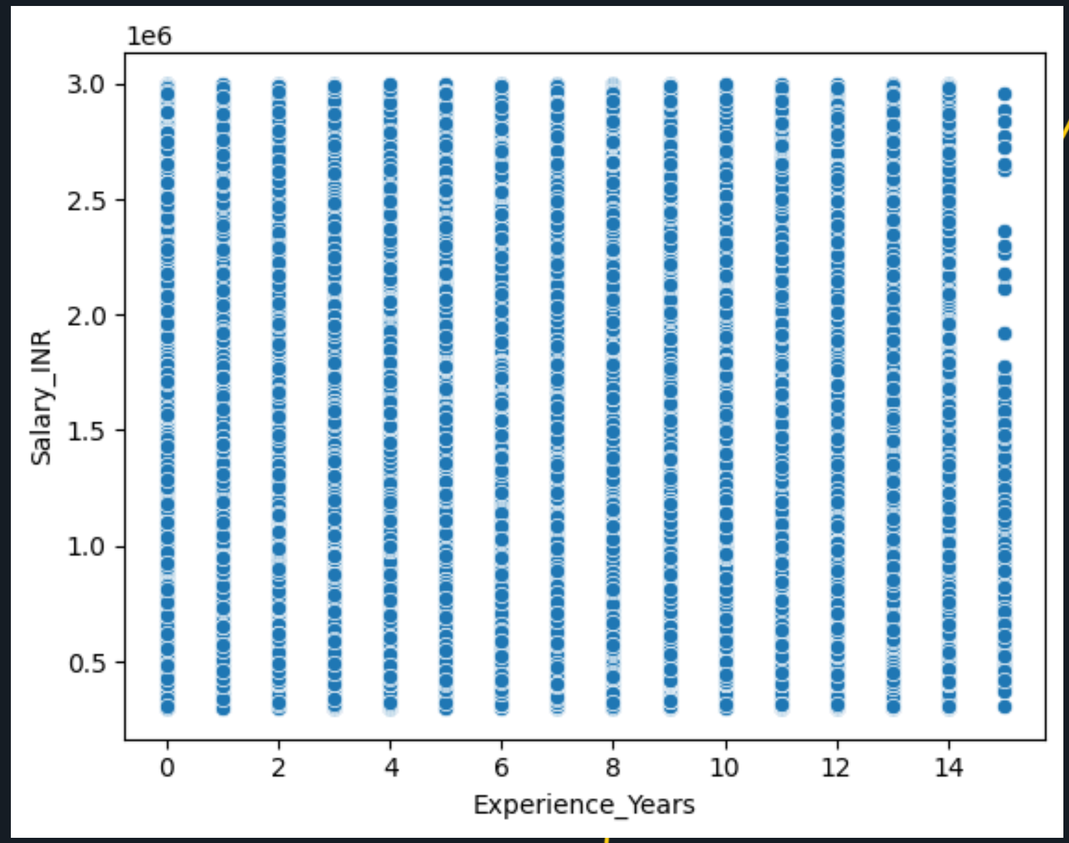
Key takeaway:
There is a wide variation in salaries, with most employees earning in a middle range but a few individuals receiving very high salaries.

8. Salary vs. Experience

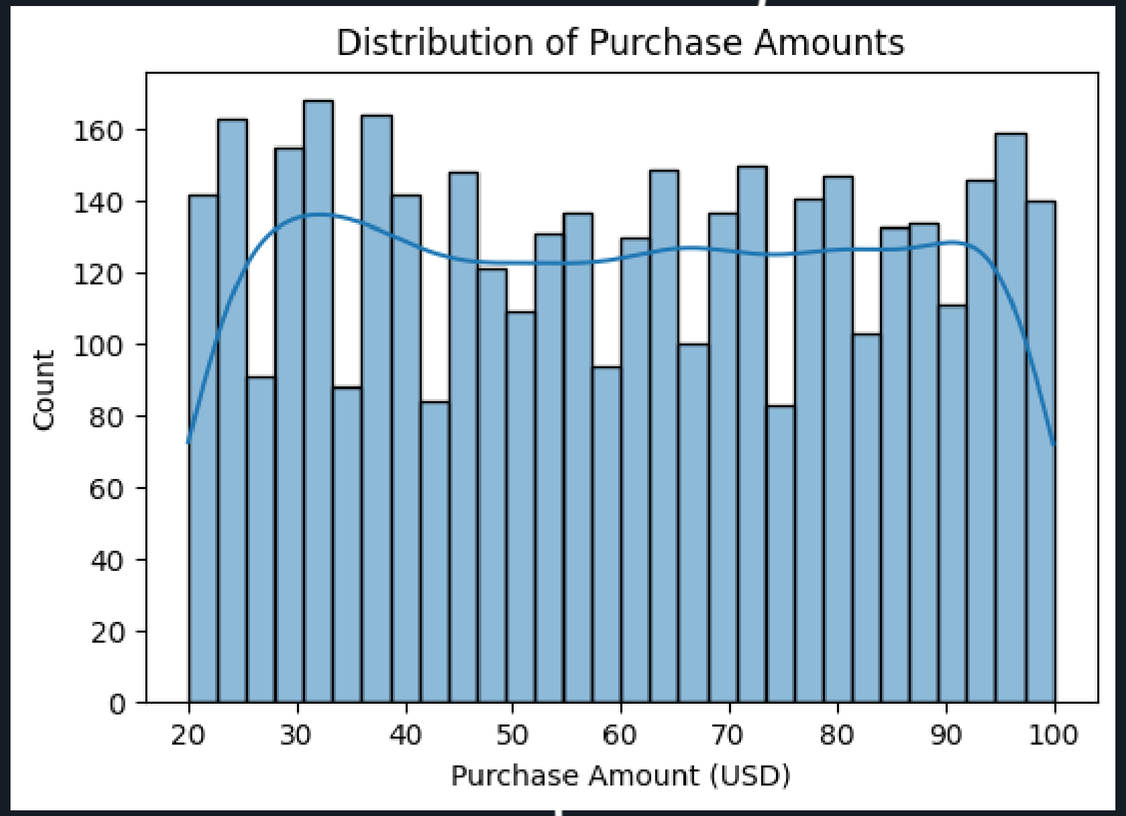
What the visualization shows:
The relationship between years of experience and salary, with a regression line.

Why I used this visualization:
A scatter plot with regression helps demonstrate trends and shows if experience has an impact on salary growth.

Key takeaway:
There is a clear positive relationship: employees with more years of experience generally earn higher salaries.



2.Shopping Trends

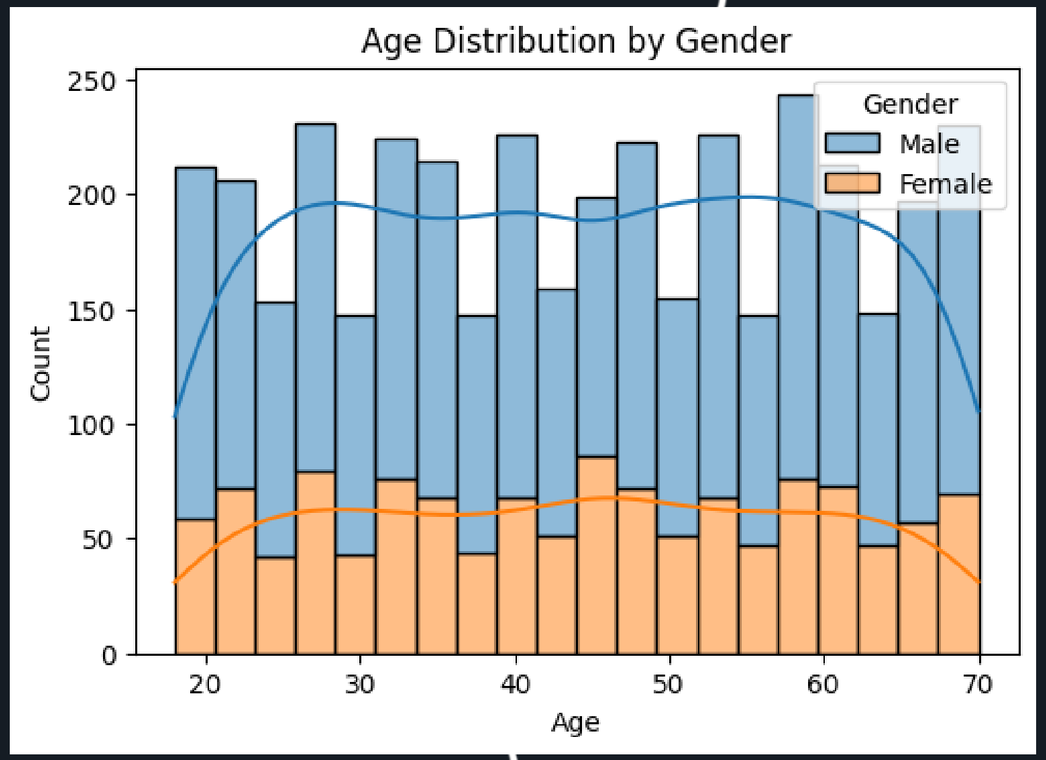


1. Distribution of Purchase Amounts

What the visualization shows:
The spread of customer purchase amounts in USD.

Why I used this visualization:
A histogram helps visualize the overall spending behavior and detect if most customers buy in similar ranges.

Key takeaway:
Most purchases fall within a mid-range amount, while a few customers spend significantly higher, creating outliers.

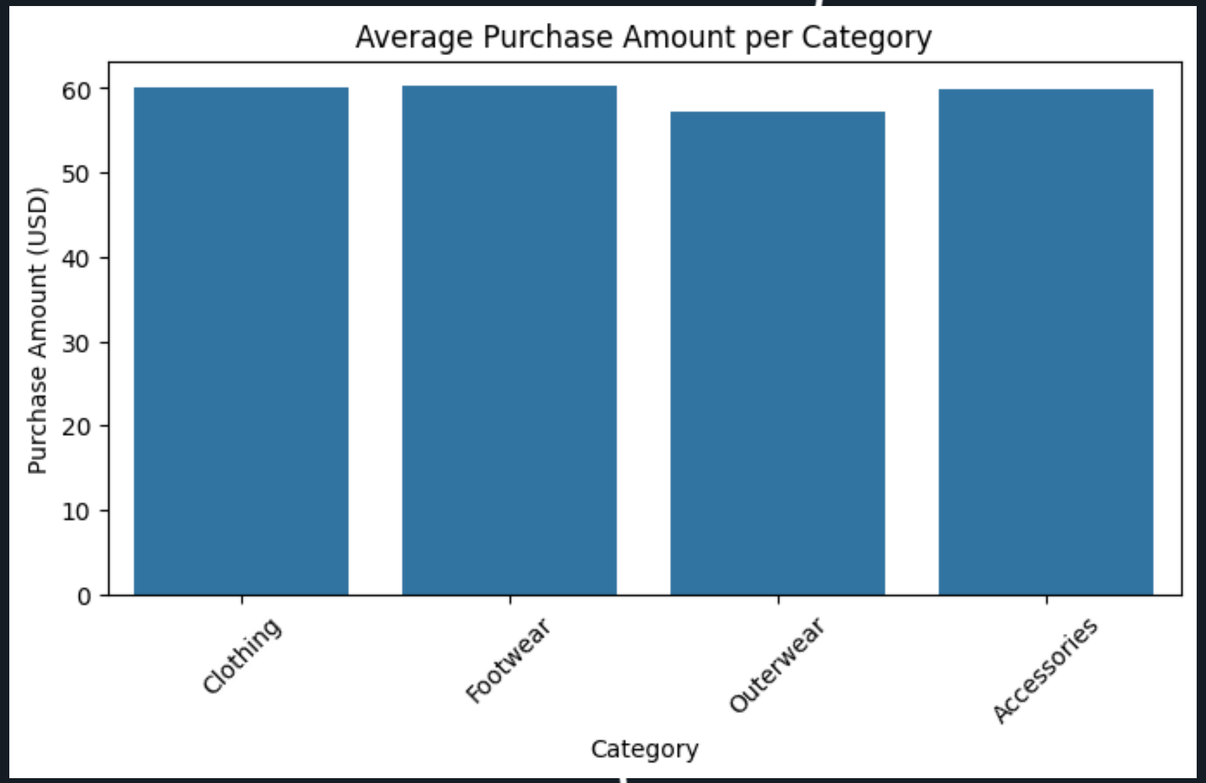


2. Age Distribution by Gender

What the visualization shows:
The distribution of customer ages, separated by gender.

Why I used this visualization:
Stacked histograms allow comparison of male vs. female customers across different age groups.

Key takeaway:
Both genders are represented across all ages, with peaks in younger age groups, showing a strong youth customer base.

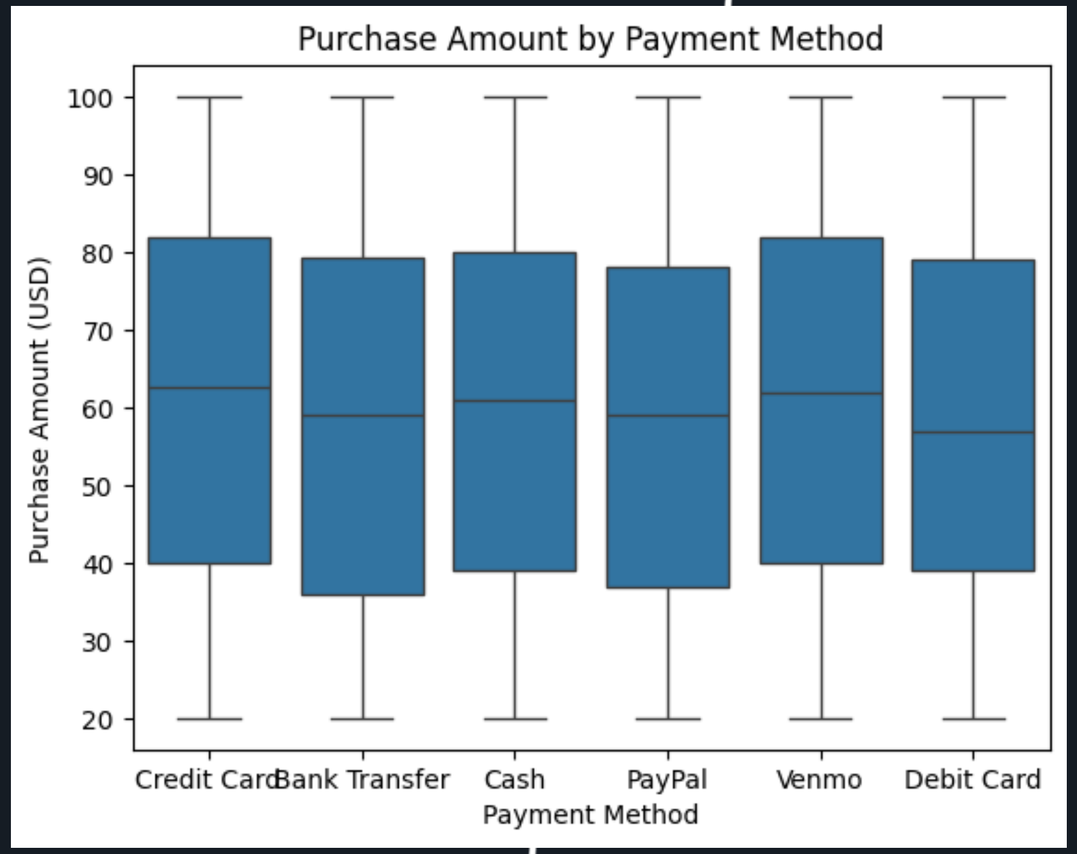


3. Average Purchase Amount per Category

What the visualization shows:
The mean purchase value for each product category.

Why I used this visualization:
A bar chart is effective to compare spending across categories.

Key takeaway:
Some categories (e.g., Electronics, Fashion) have higher purchase amounts, while others remain lower.

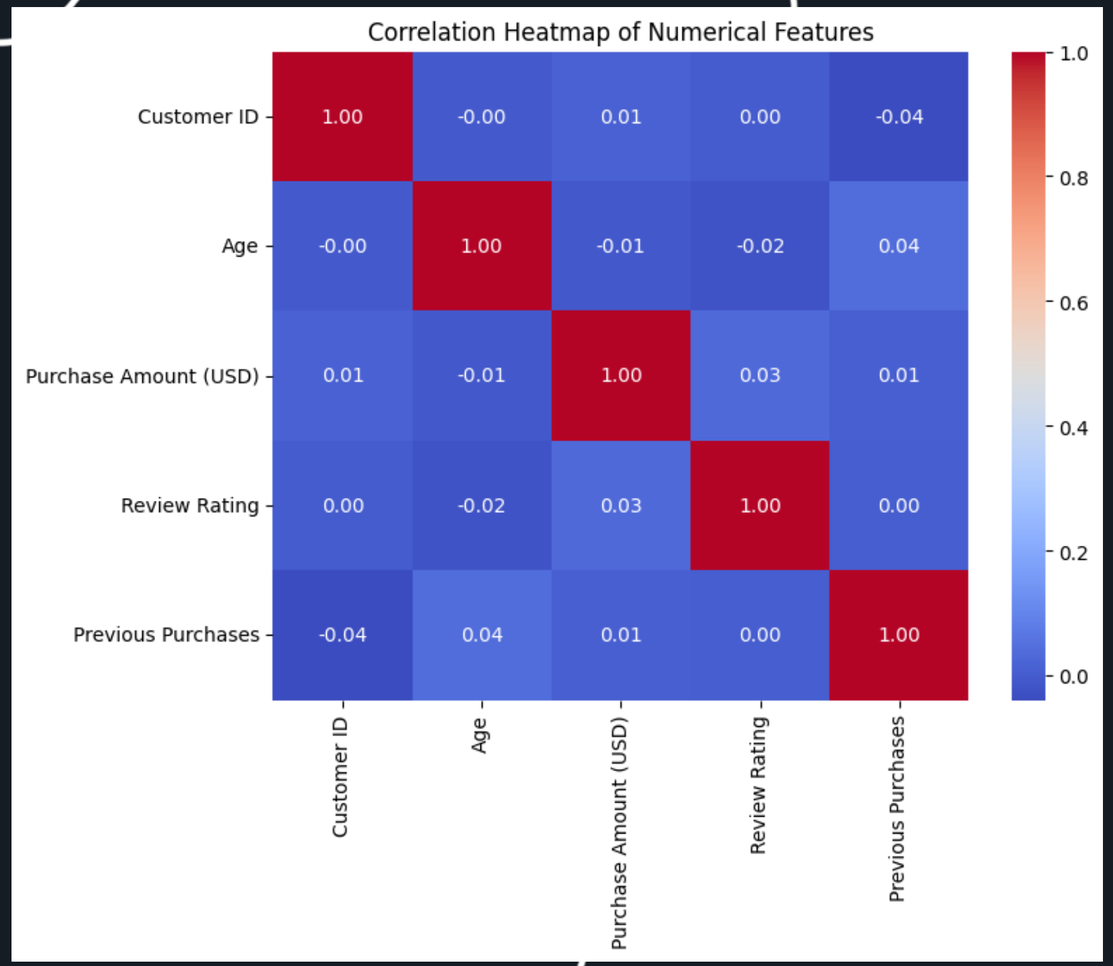


4. Purchase Amount by Payment Method

What the visualization shows:
Variation in purchase values depending on payment method.

Why I used this visualization:
A box plot highlights how different payment types affect spending and reveals outliers.

Key takeaway:
Certain payment methods are linked with higher spending, while others are used mainly for smaller transactions.

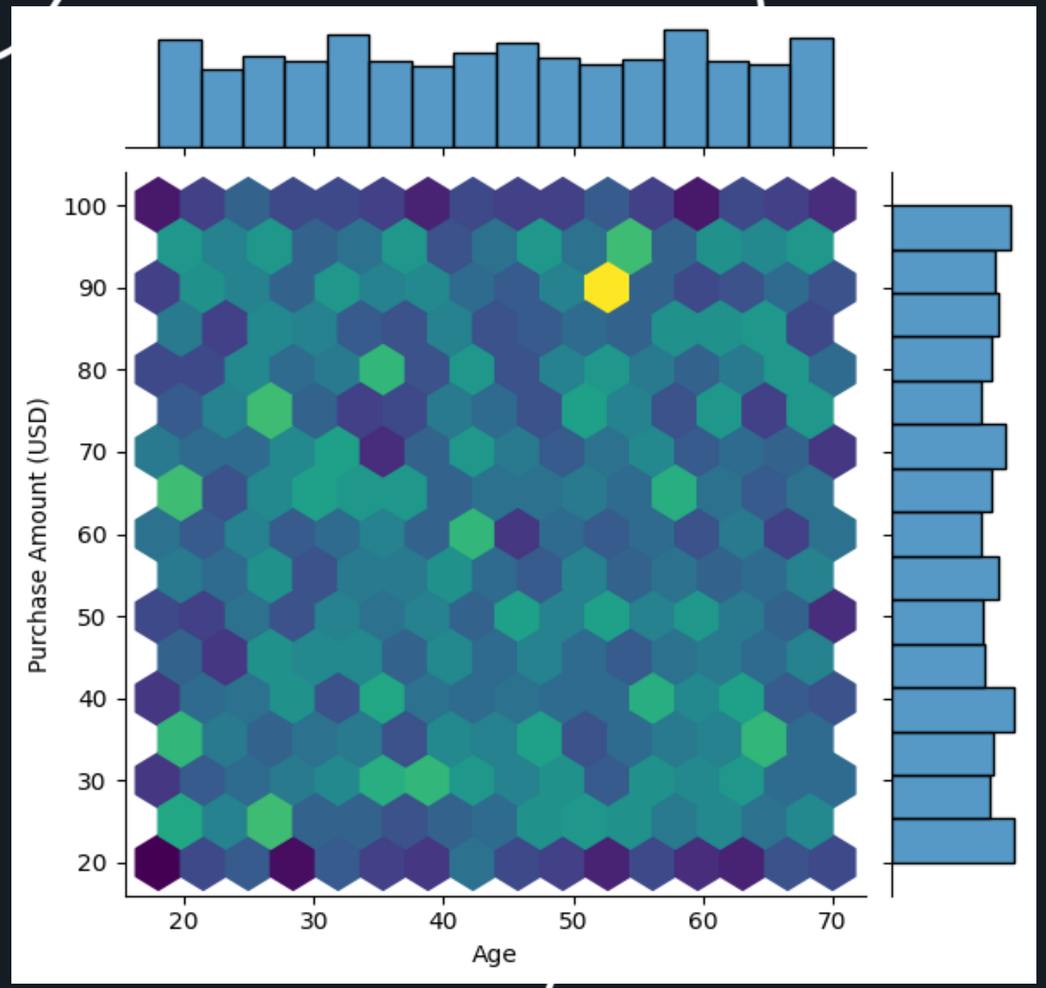


5. Correlation Heatmap of Numerical Features

What the visualization shows:
Relationships between numerical fields such as Age, Purchase Amount, Review Rating, etc.

Why I used this visualization:
A heatmap makes it easy to see which variables are related.

Key takeaway:
Purchase amount has weak-to-moderate relationships with other features, while some variables (like Age and Rating) show minimal correlation.

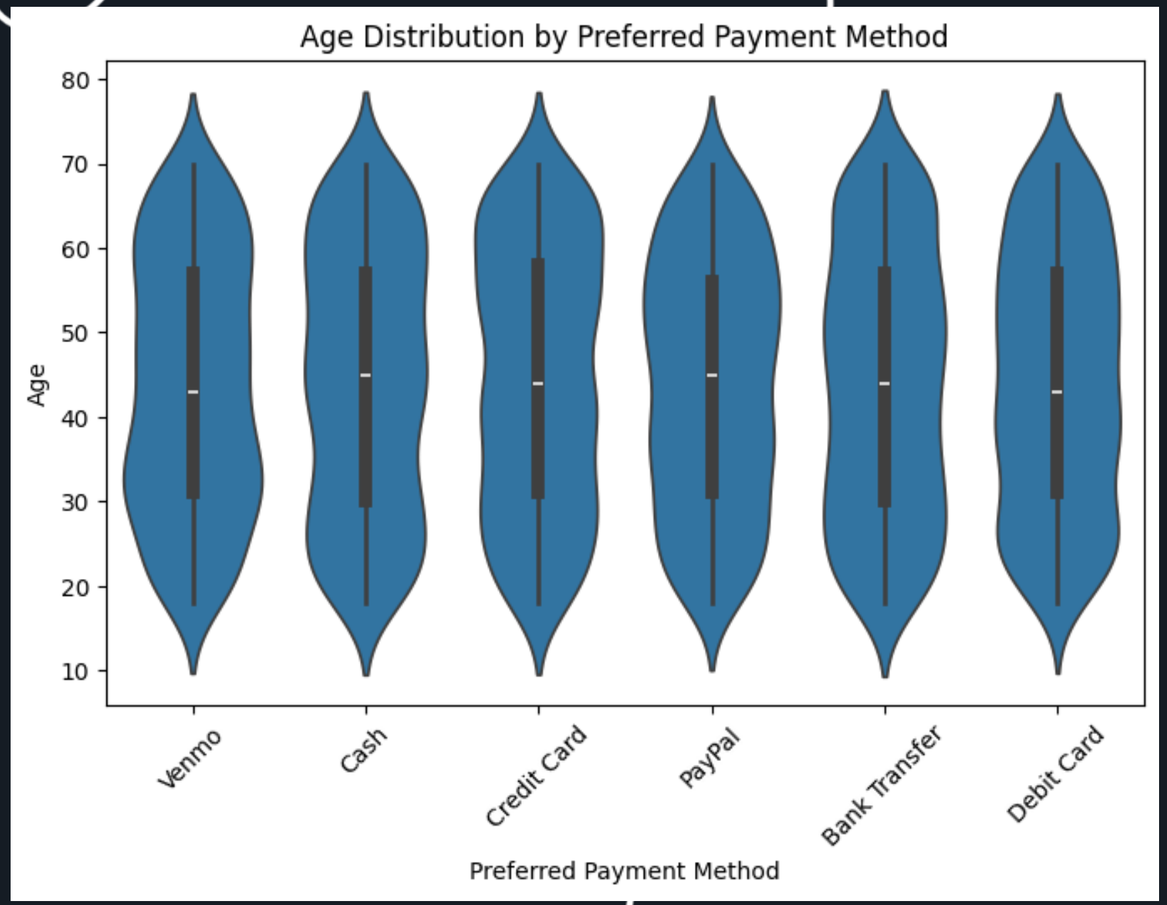


6. Age vs Purchase Amount

What the visualization shows:
A scatter plot with side histograms that shows how customer **age** relates to **purchase amount**.

Why I used this visualization:
Combines two views ,overall distributions and the relationship between them.

Key takeaway:
Middle-aged customers tend to spend more consistently, while younger groups show more variation.



7. Age by Purchase Method

What the visualization shows:
Each violin shows the distribution of customer ages for a specific purchase method (e.g., Credit Card, Cash, Online).

Why I used this visualization:
Highlights which age groups prefer which payment types, and how spread out or concentrated they are.

Key takeaway:
Younger customers may favor modern payment methods (like online or card), while older customers cluster around traditional methods.

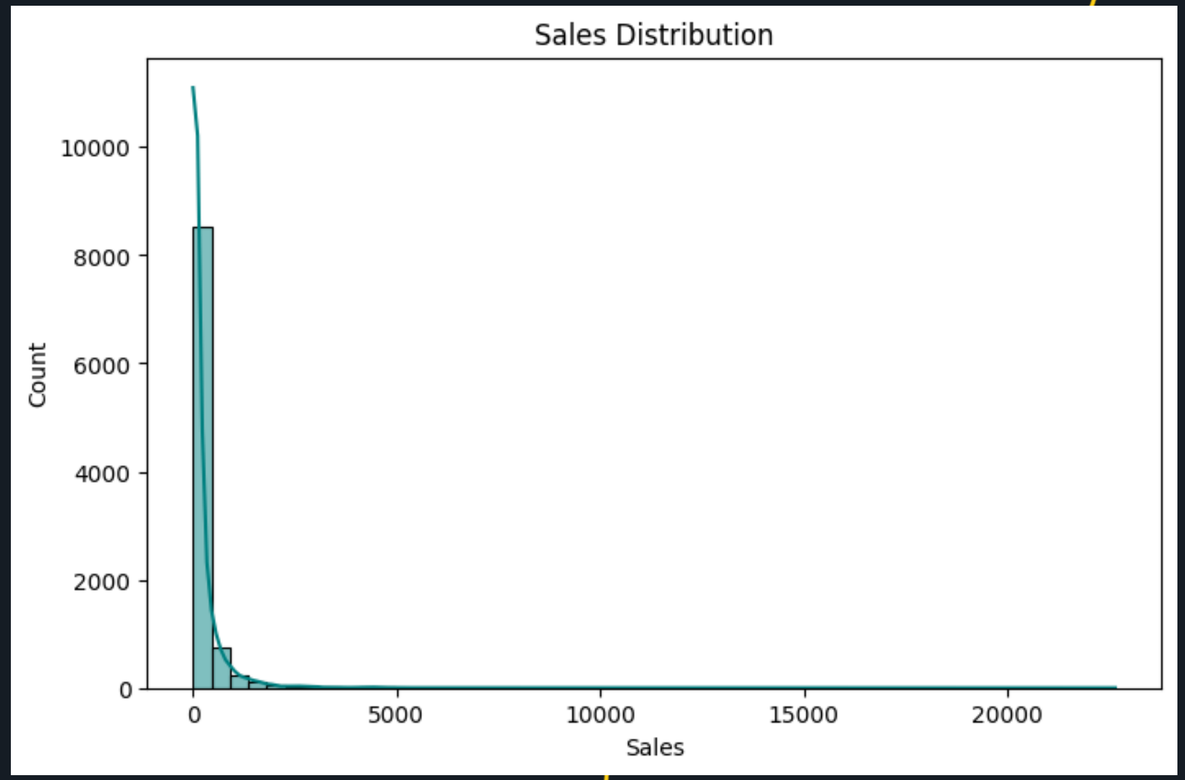
3.Supermarket

1. Sales Distribution

What the visualization shows:
The spread of sales amounts across all transactions.

Why I used this visualization:
A histogram is useful for spotting common sales values and detecting skewed distributions.

Key takeaway:
Most sales transactions cluster in a middle range, with fewer very high or very low sales.

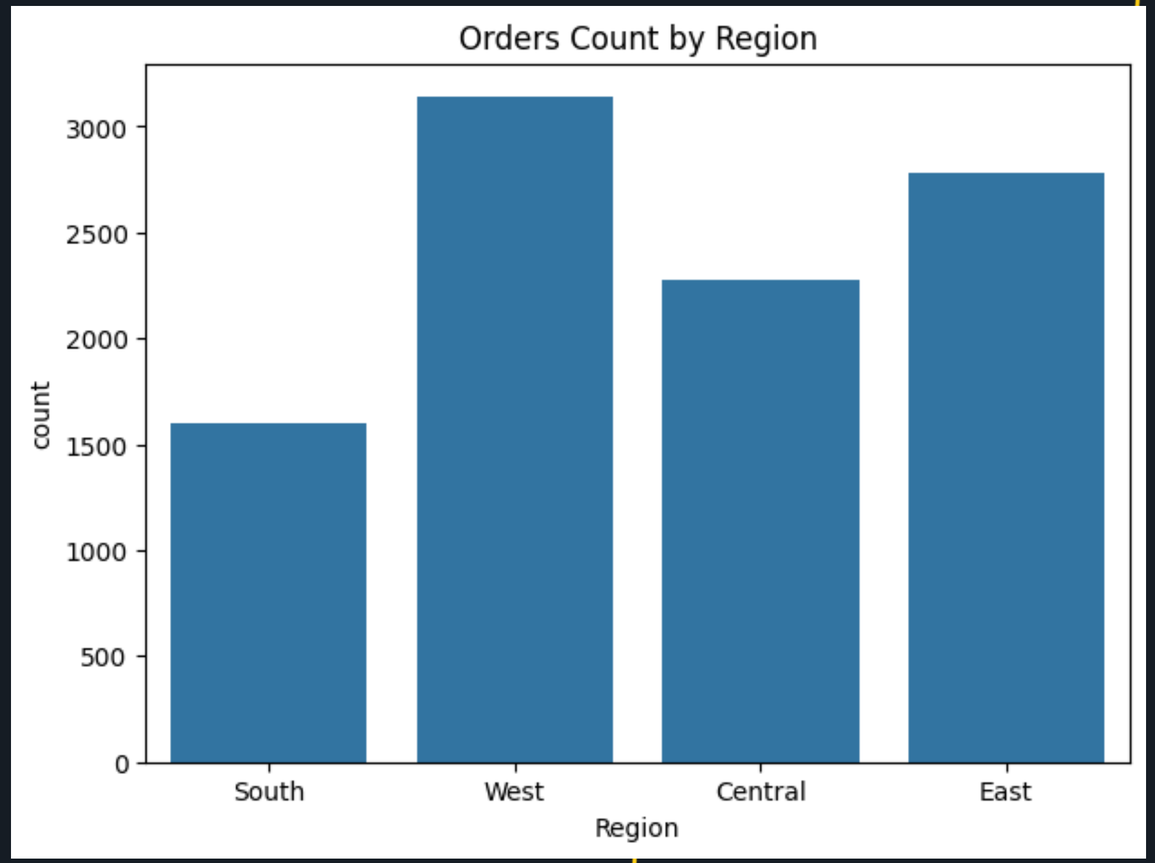


2. Orders Count by Region

What the visualization shows:
The number of orders placed in each region.

Why I used this visualization:
A count plot helps identify the busiest regions in terms of customer orders.

Key takeaway:
Certain regions dominate order counts, highlighting strong customer demand there.

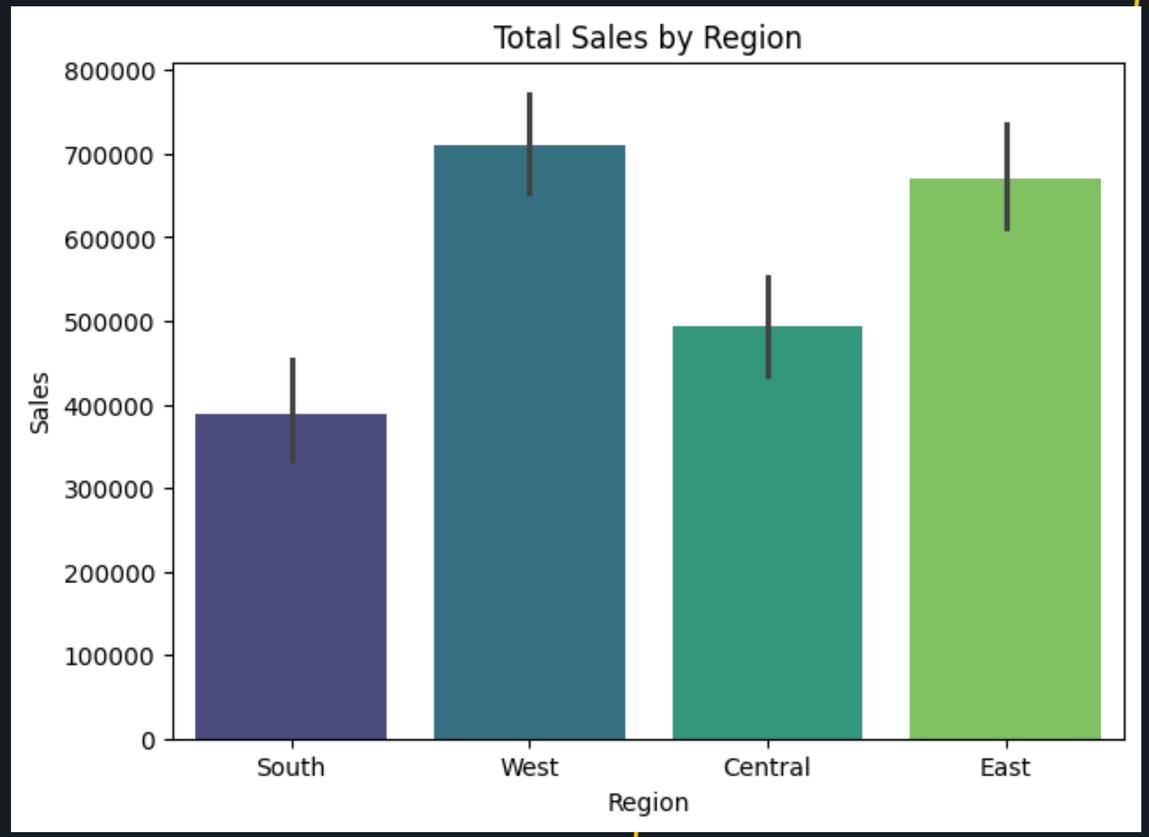


3. Total Sales by Region

What the visualization shows:
Total revenue generated from each region.

Why I used this visualization:
A bar chart allows comparison of how much revenue different regions contribute.

Key takeaway:
Some regions not only have more orders but also generate higher total sales, making them top-performing markets.

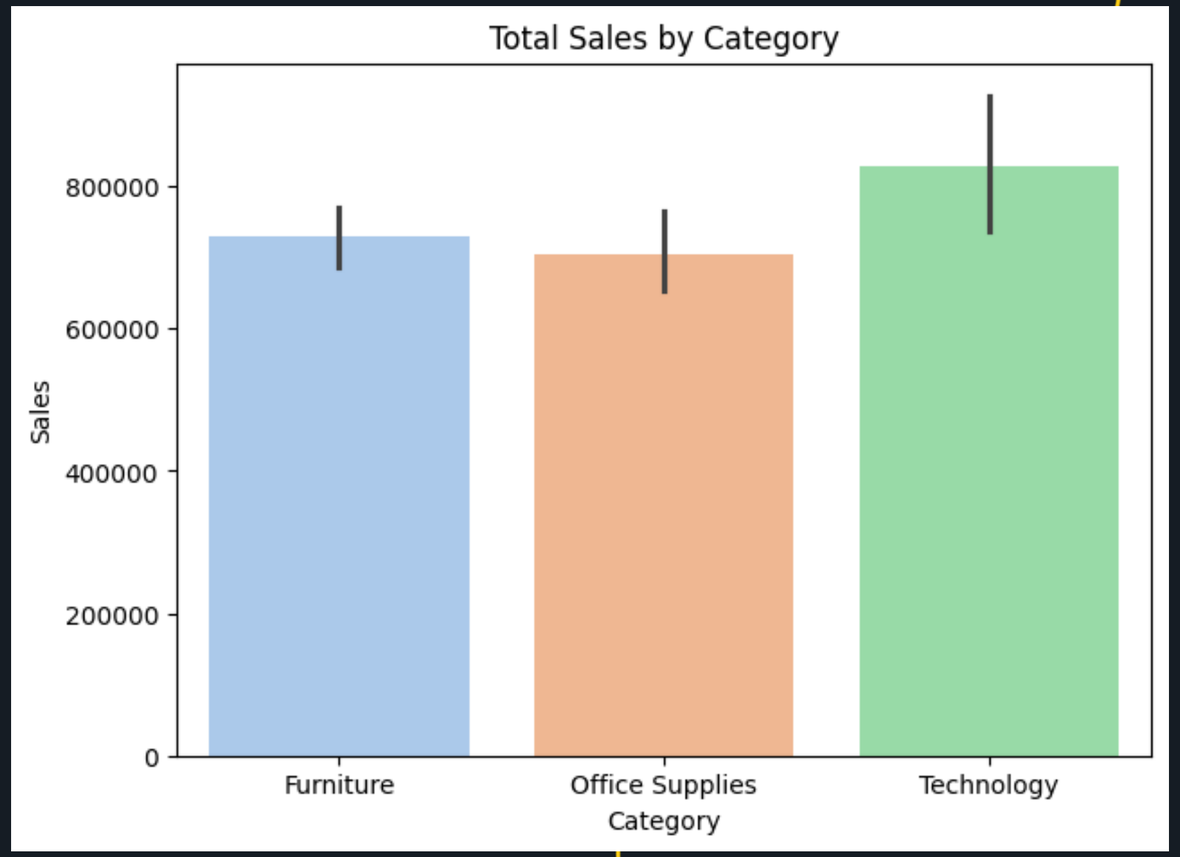


4. Total Sales by Category

What the visualization shows:
The overall sales split across product categories.

Why I used this visualization:
Category analysis highlights which products bring the most revenue.

Key takeaway:
A few categories dominate total sales, suggesting where the business should focus.

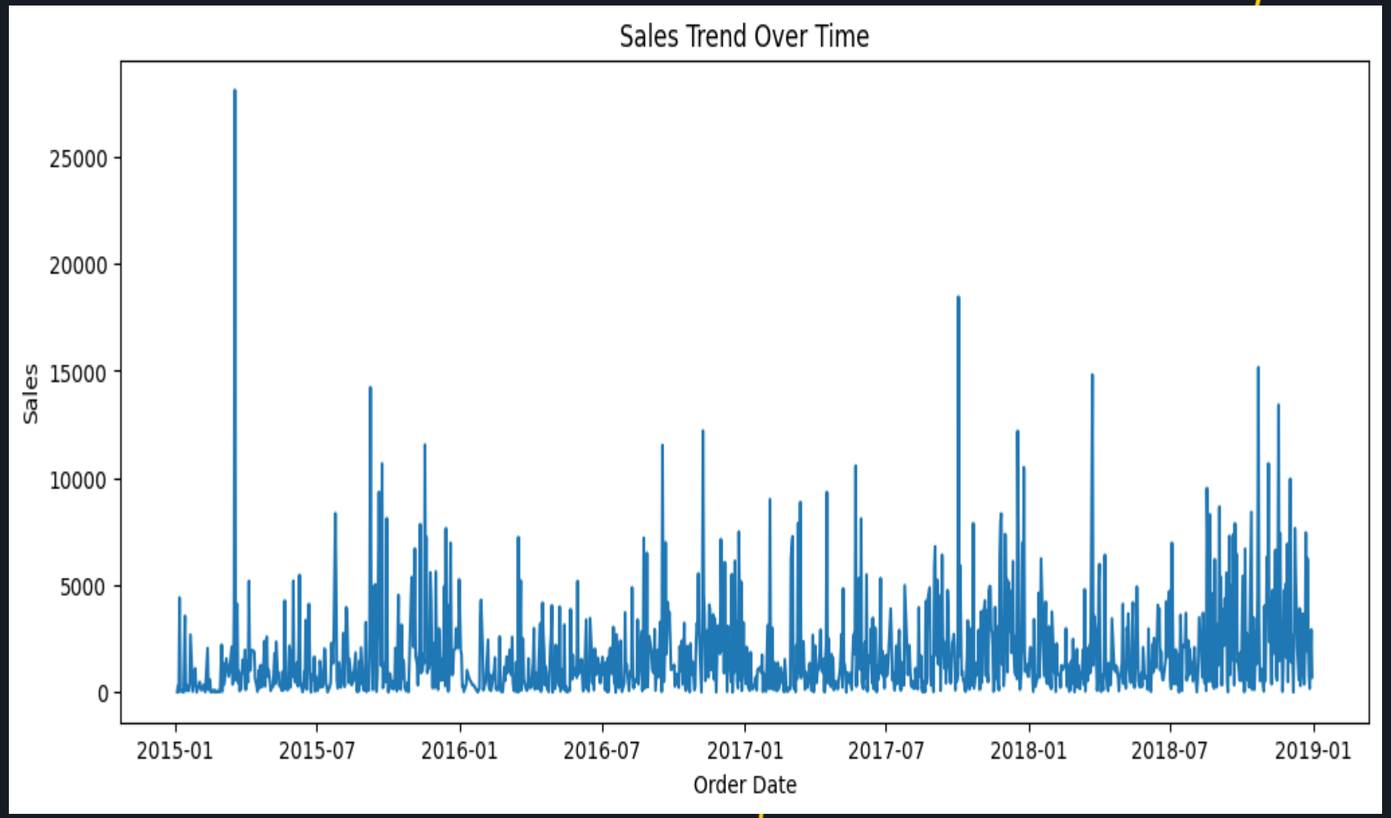


5. Sales Trend Over Time

What the visualization shows:
The movement of total sales over time.

Why I used this visualization:
A line plot shows seasonal peaks, growth patterns, or drops in sales.

Key takeaway:
Sales increase steadily with clear peaks in certain months, suggesting strong seasonal demand.

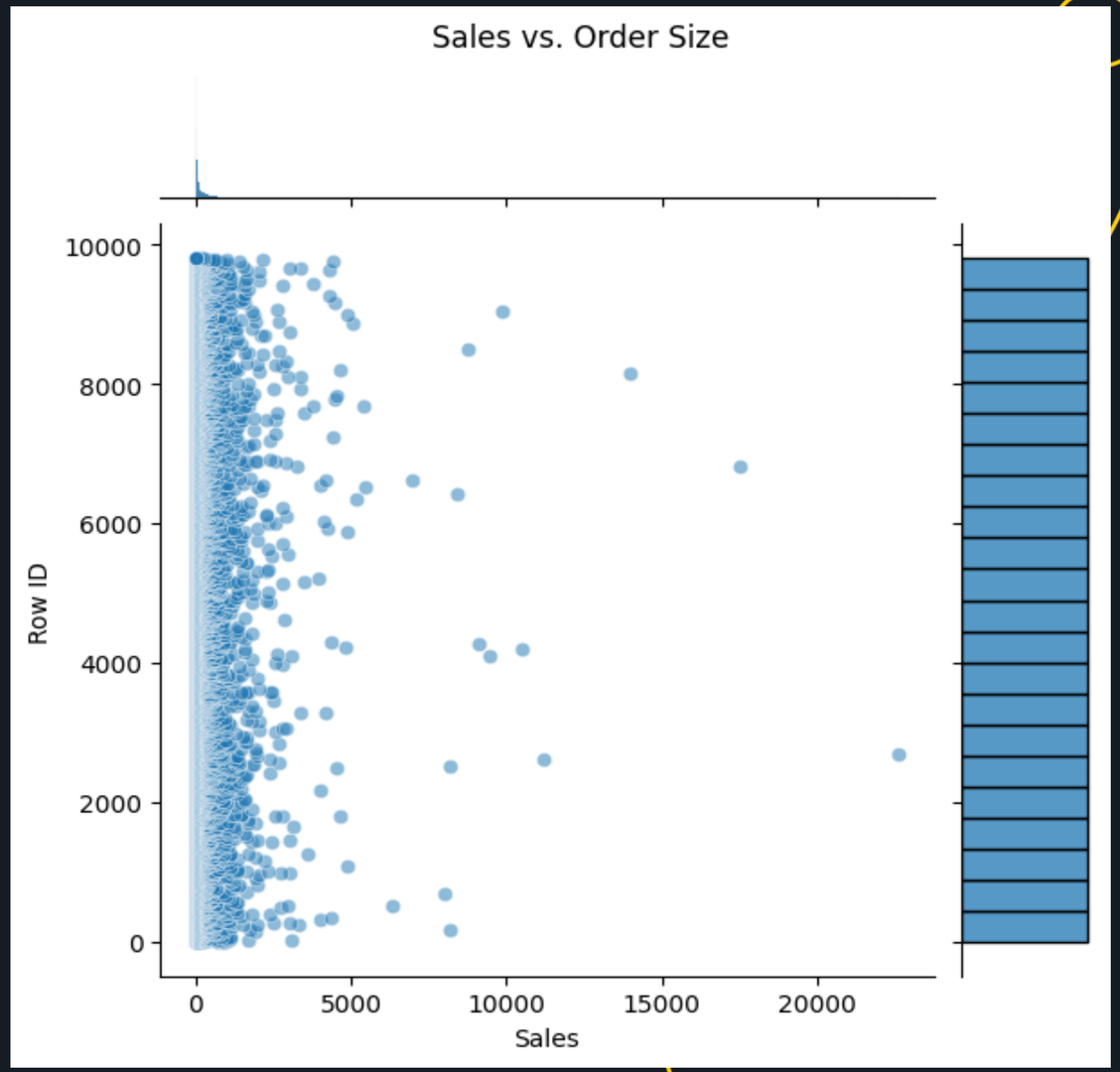


6. Sales vs Order Size

What the visualization shows:
The relationship between the size of an order (number of items) and the total sales value.

Why I used this visualization:
To check if larger orders always lead to higher revenue, or if small orders can also contribute significantly.

Key takeaway:
Larger order sizes generally bring more sales, but there are some small orders with high sales due to expensive items.

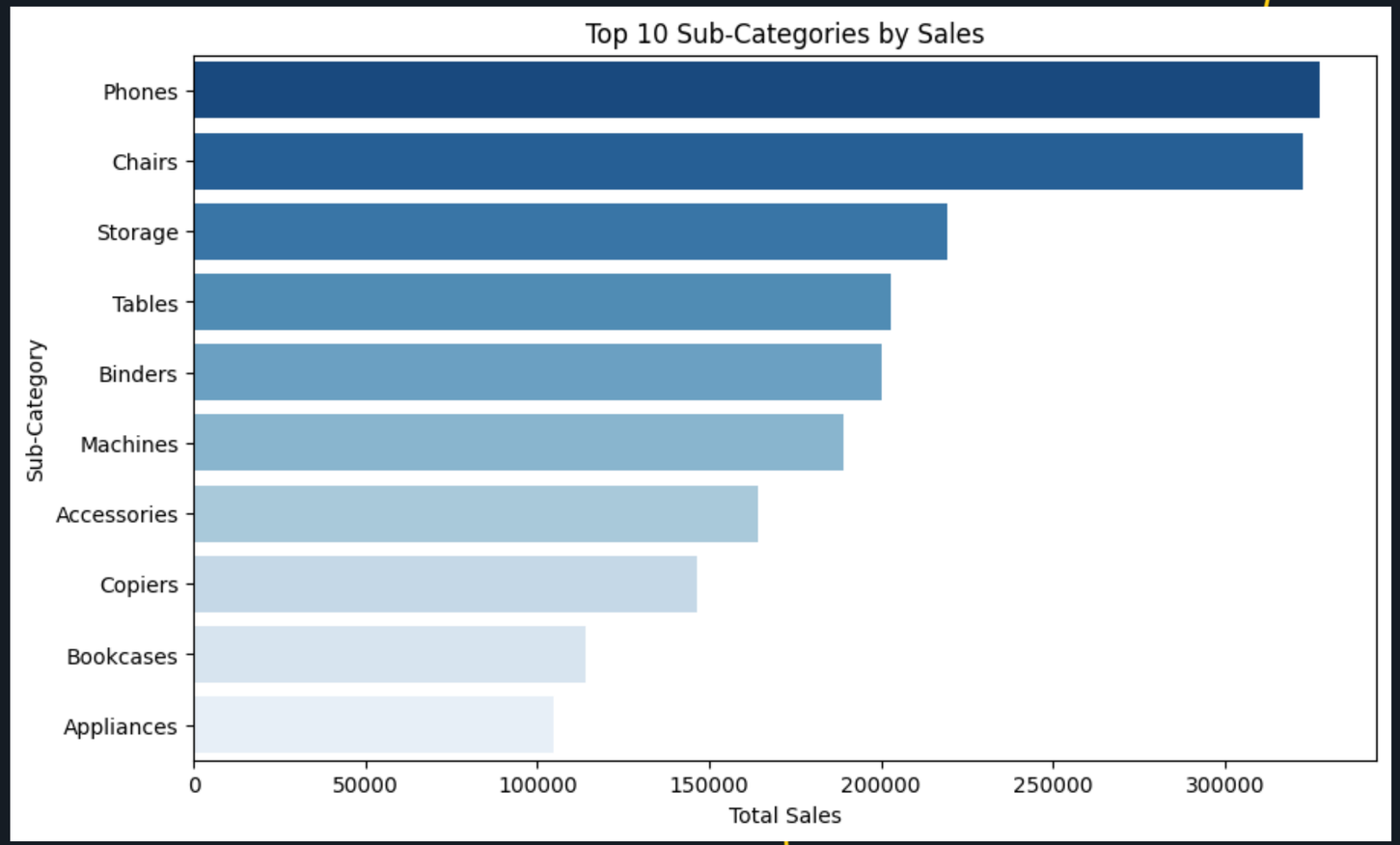


7. Top 10 Sub-Categories by Sales

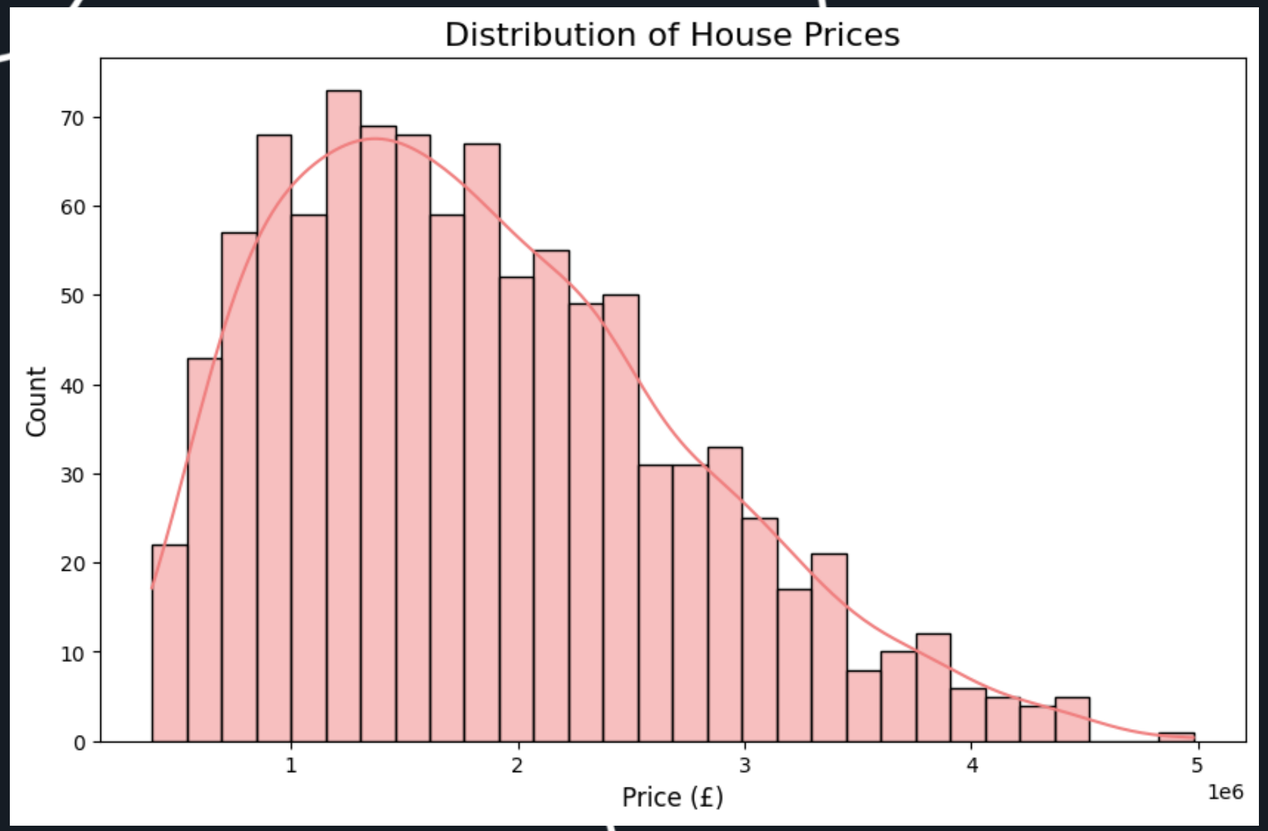
What the visualization shows:
Total or average sales broken down by grouped categories, such as **Customer Segment** or **Product Category**.

Why I used this visualization:
Grouped bar charts make it easier to compare how different segments contribute to overall sales.

Key takeaway:
Certain groups (e.g., Corporate customers or Technology products) stand out as stronger revenue drivers compared to others.



4.London Houses

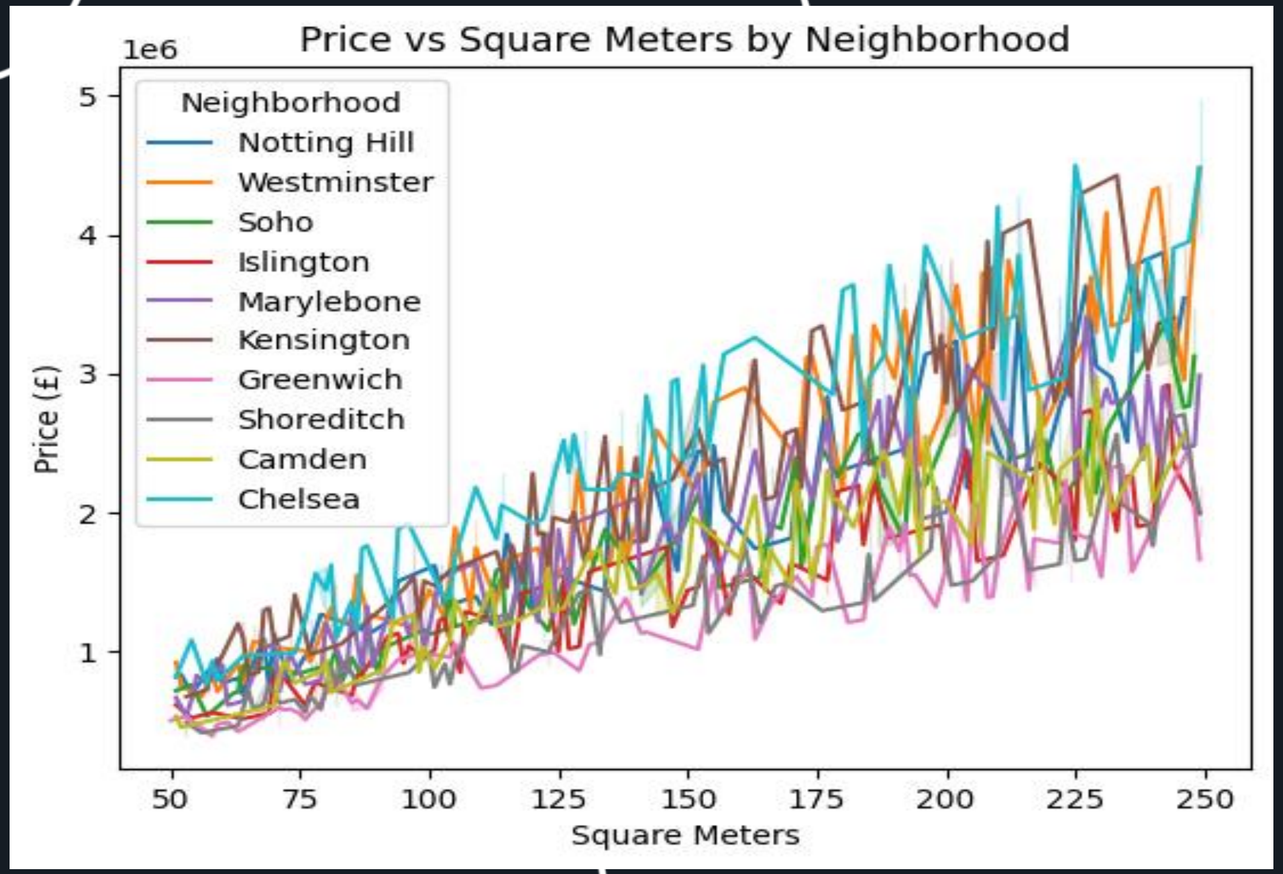


1. Distribution of House Prices

What the visualization shows:
How house prices are distributed across different ranges.

Why I used this visualization:
A histogram with KDE reveals whether prices are concentrated in certain ranges or widely spread.

Key takeaway:
Most houses fall into middle price ranges, with fewer very high-priced properties.

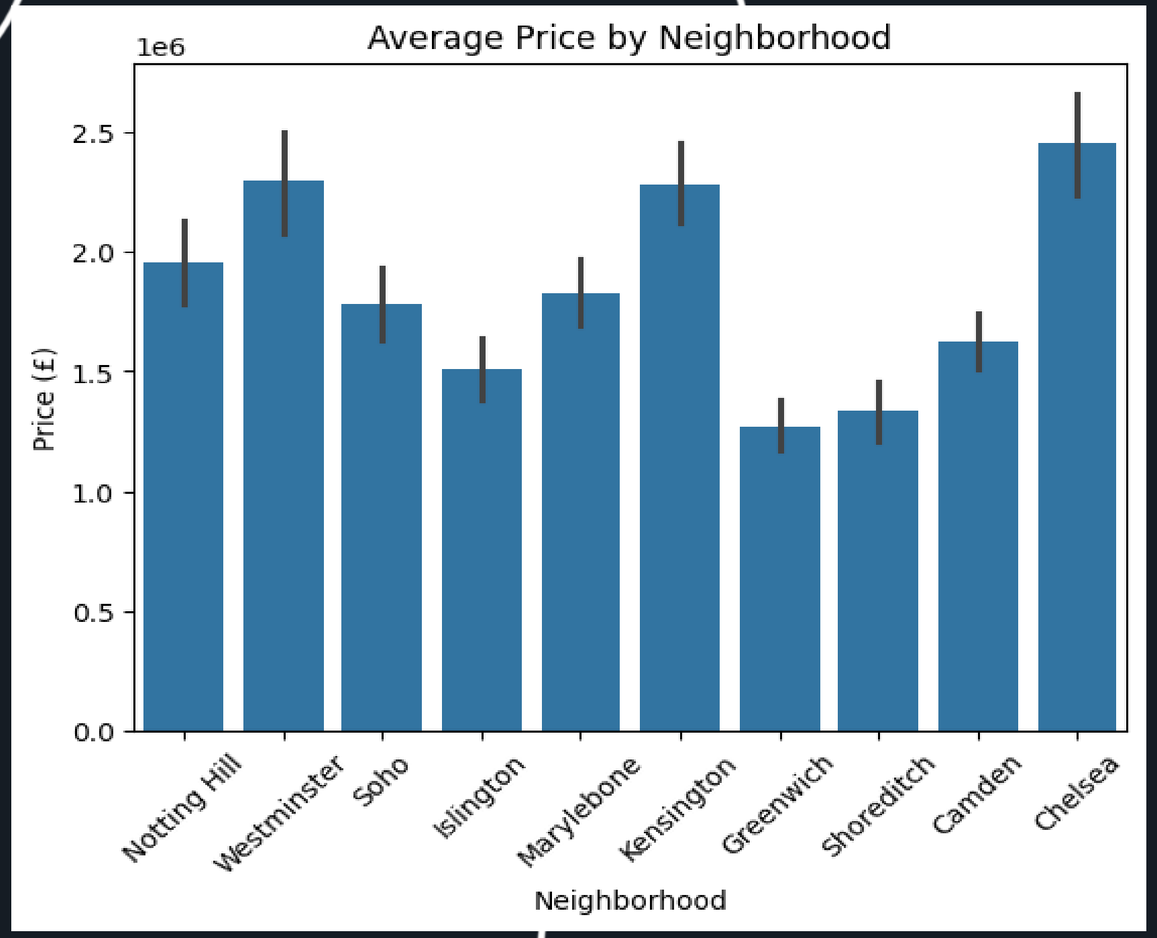


2. Price vs Square Meters by Neighborhood

What the visualization shows:
The relationship between property size and price, separated by neighborhood.

Why I used this visualization:
To see how price grows with size and whether the trend differs across neighborhoods.

Key takeaway:
Larger properties generally cost more, but some neighborhoods are more expensive regardless of size.

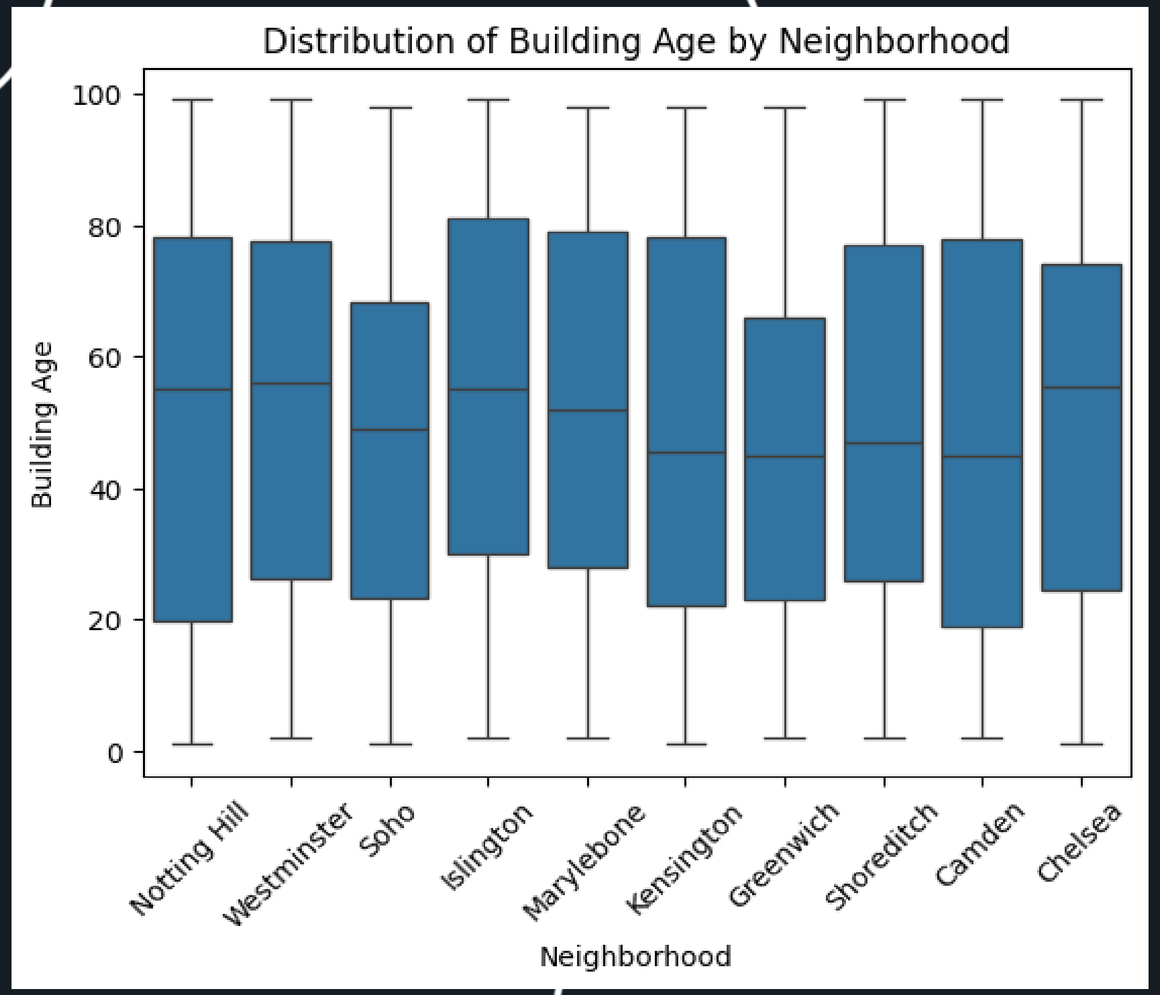


3. Average Price by Neighborhood

What the visualization shows:
The mean house price in each neighborhood.

Why I used this visualization:
Averages give a quick comparison of neighborhood affordability.

Key takeaway:
Some neighborhoods consistently have higher prices, standing out as more premium locations.

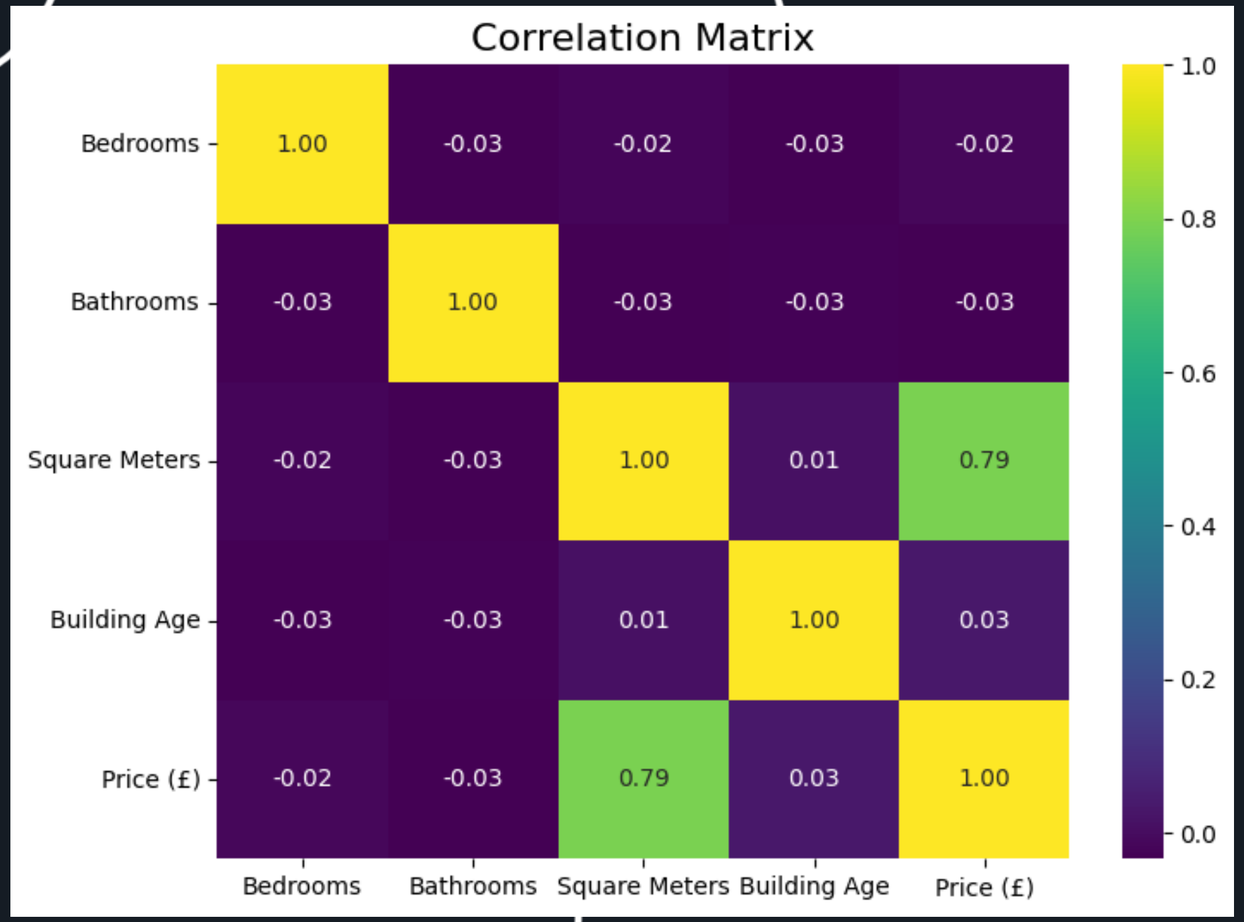


4. Distribution of Building Age by Neighborhood

What the visualization shows:
How old the buildings are in different neighborhoods.

Why I used this visualization:
A box plot shows medians, ranges, and outliers, highlighting age variation.

Key takeaway:
Certain neighborhoods have much older housing stock, while others feature newer developments.

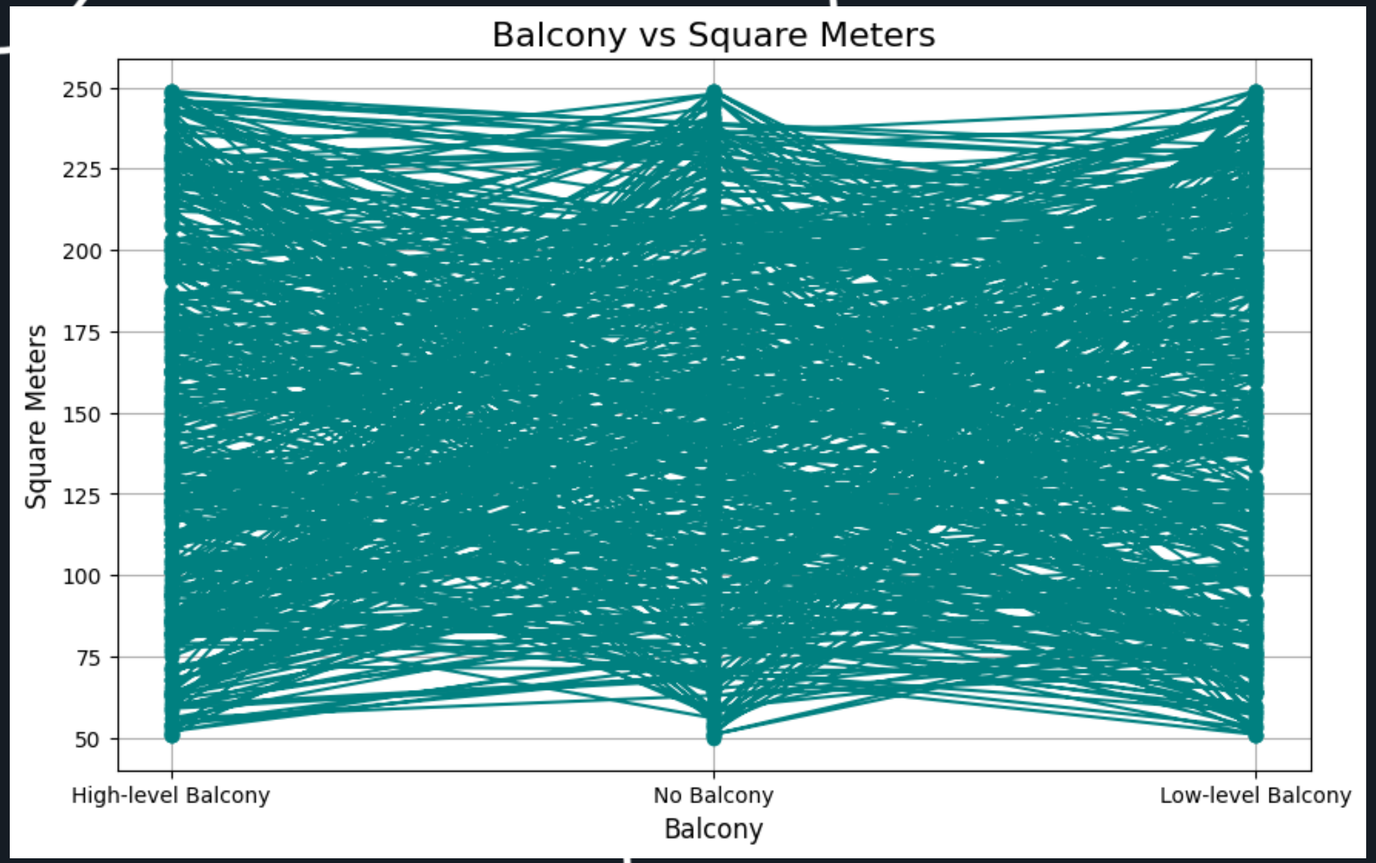


5. Correlation Matrix

What the visualization shows:
The strength of relationships between numeric variables like bedrooms, size, age, and price.

Why I used this visualization:
Helps identify which features strongly influence price.

Key takeaway:
Square meters and number of bedrooms are highly correlated with price, while building age has a weaker effect.

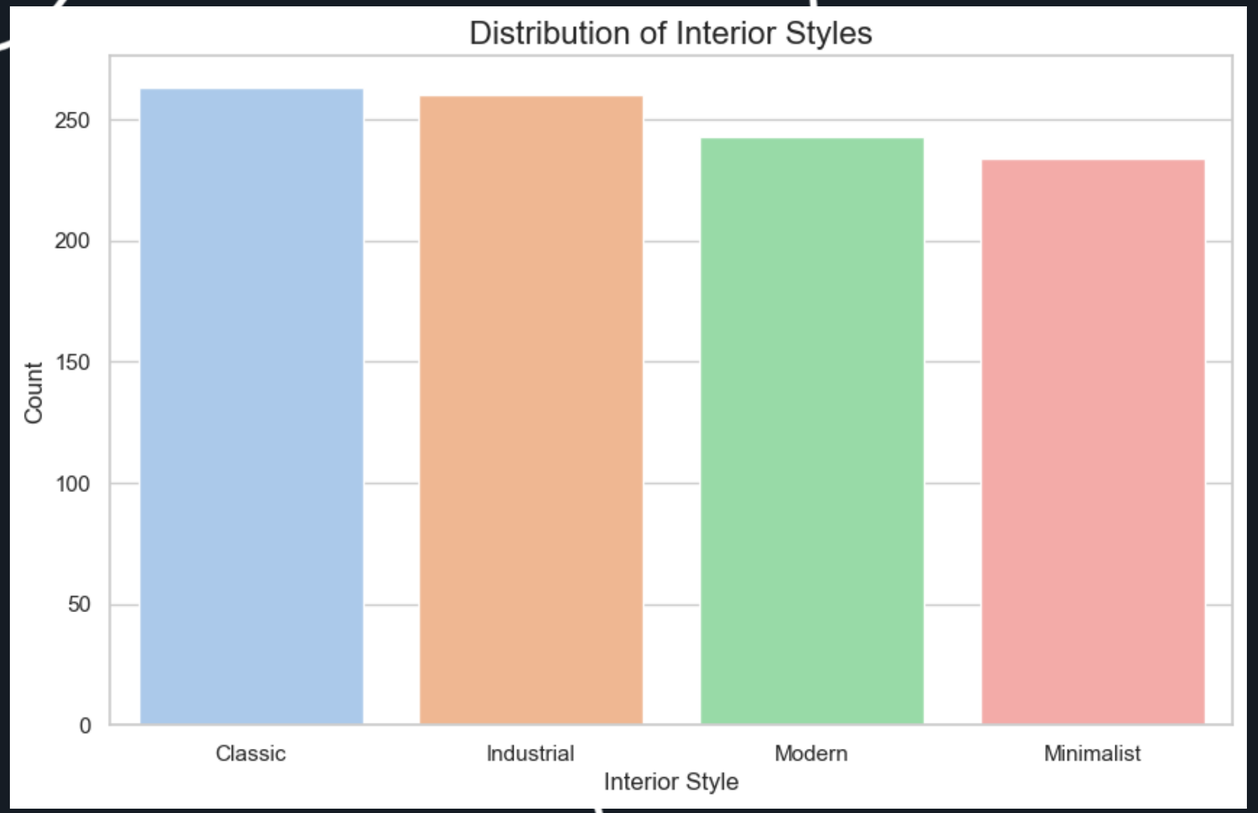


6. Balcony vs Square Meters

What the visualization shows:
The relationship between balcony size/availability and property size.

Why I used this visualization:
To see whether bigger houses are more likely to include balconies.

Key takeaway:
Larger properties usually include balconies, but smaller ones rarely do.



7. Distribution of Interior Styles

What the visualization shows:
How many properties fall under each interior style (e.g., Modern, Classic).

Why I used this visualization:
Count plots are great for comparing categorical popularity.

Key takeaway:
Modern styles dominate the dataset, while other styles are less common.

8. Building Age vs House Price

What the visualization shows:
The relationship between building age and price, with neighborhoods and property types highlighted.

Why I used this visualization:
Scatter plots reveal patterns and clustering that indicate how age impacts price.

Key takeaway:
Newer buildings often sell at higher prices, but location (neighborhood) also strongly influences value.

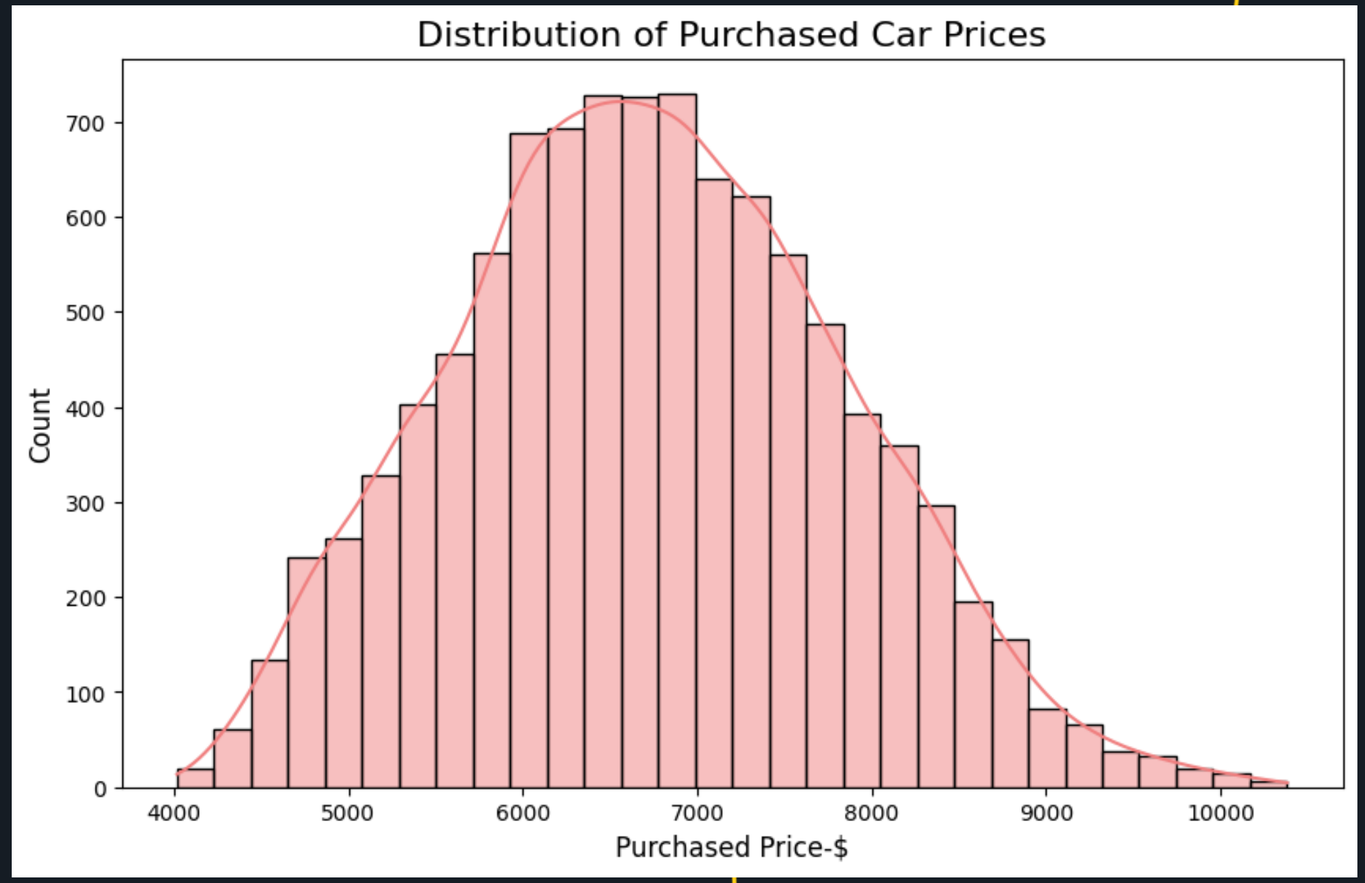
5.Used Cars

1. Distribution of Purchased Car Prices

What the visualization shows:
The spread of car purchase prices, with frequency counts and an overlaid smooth curve (KDE).

Why I used this visualization:
A histogram with KDE highlights the overall distribution, central tendency, and whether prices are skewed.

Key takeaway:
Most cars fall within a certain purchase price range, with fewer very cheap or very expensive cars.

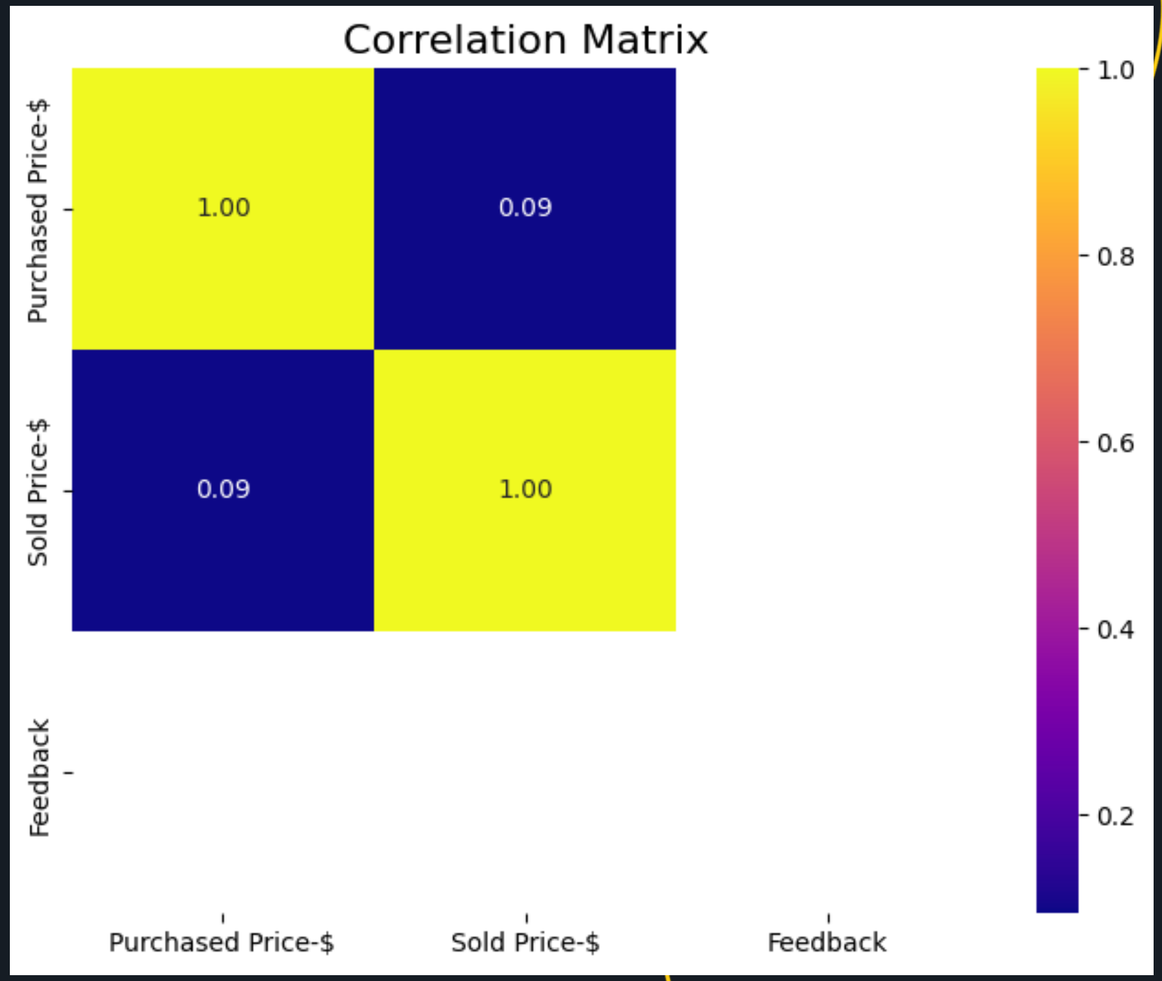


2. Correlation Matrix

What the visualization shows:
Correlations between numerical features like purchased price, sold price, and feedback.

Why I used this visualization:
A heatmap makes it easy to see strong positive or negative relationships at a glance.

Key takeaway:
Purchased price and sold price are strongly correlated, while feedback may have weaker relationships.

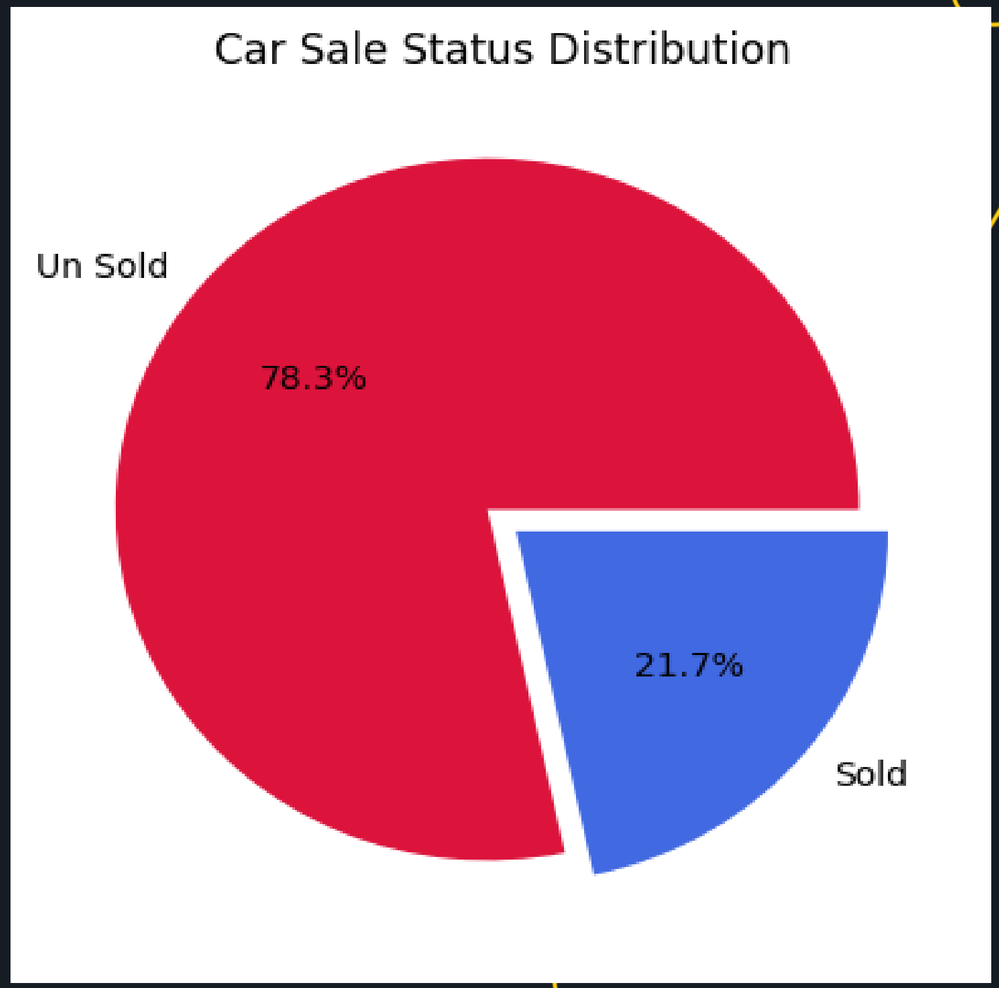


3. Car Sale Status Distribution

What the visualization shows:
The proportion of cars in different sale statuses (e.g., sold, pending, etc.).

Why I used this visualization:
A pie chart is straightforward for showing relative proportions across categories.

Key takeaway:
Certain statuses dominate (e.g., most cars are sold), while others are less frequent.

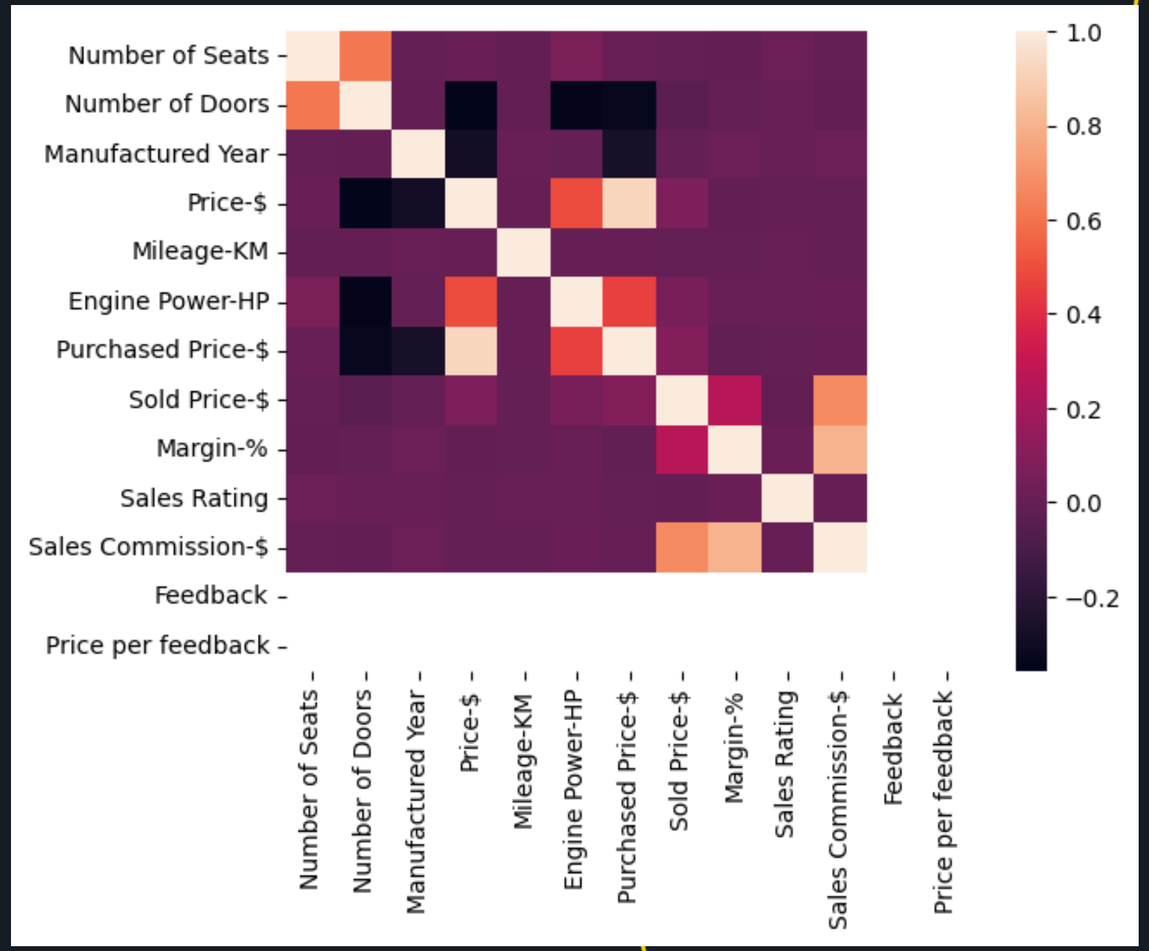


4. Full Correlation Heatmap

What the visualization shows:
Pairwise correlations across all numeric features in the dataset.

Why I used this visualization:
To explore broader relationships beyond the selected features in Visualization 2.

Key takeaway:
Some features show multicollinearity, while others are mostly independent.

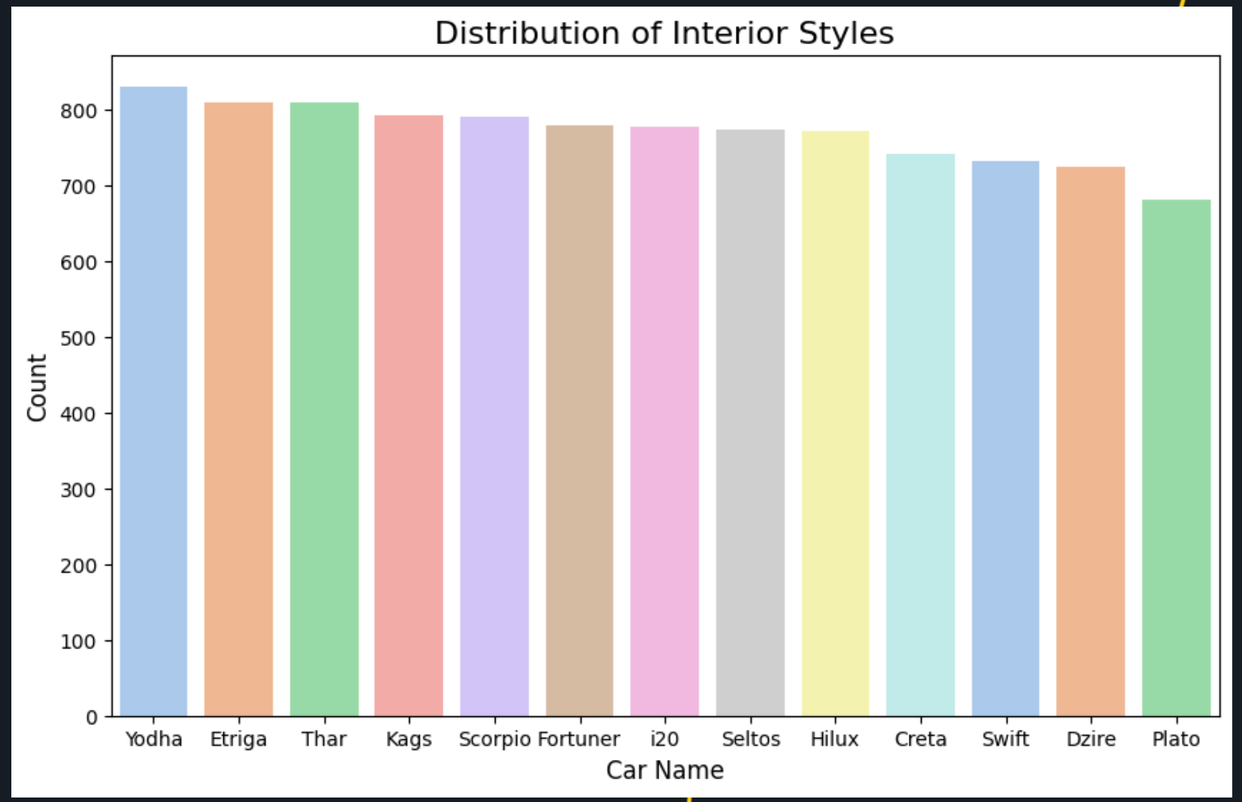


5. Distribution of Car Names

What the visualization shows:
How many times each car model/name appears in the dataset.

Why I used this visualization:
A count plot clearly shows which car names dominate the dataset.

Key takeaway:
A few car names are very common, while many appear only rarely.

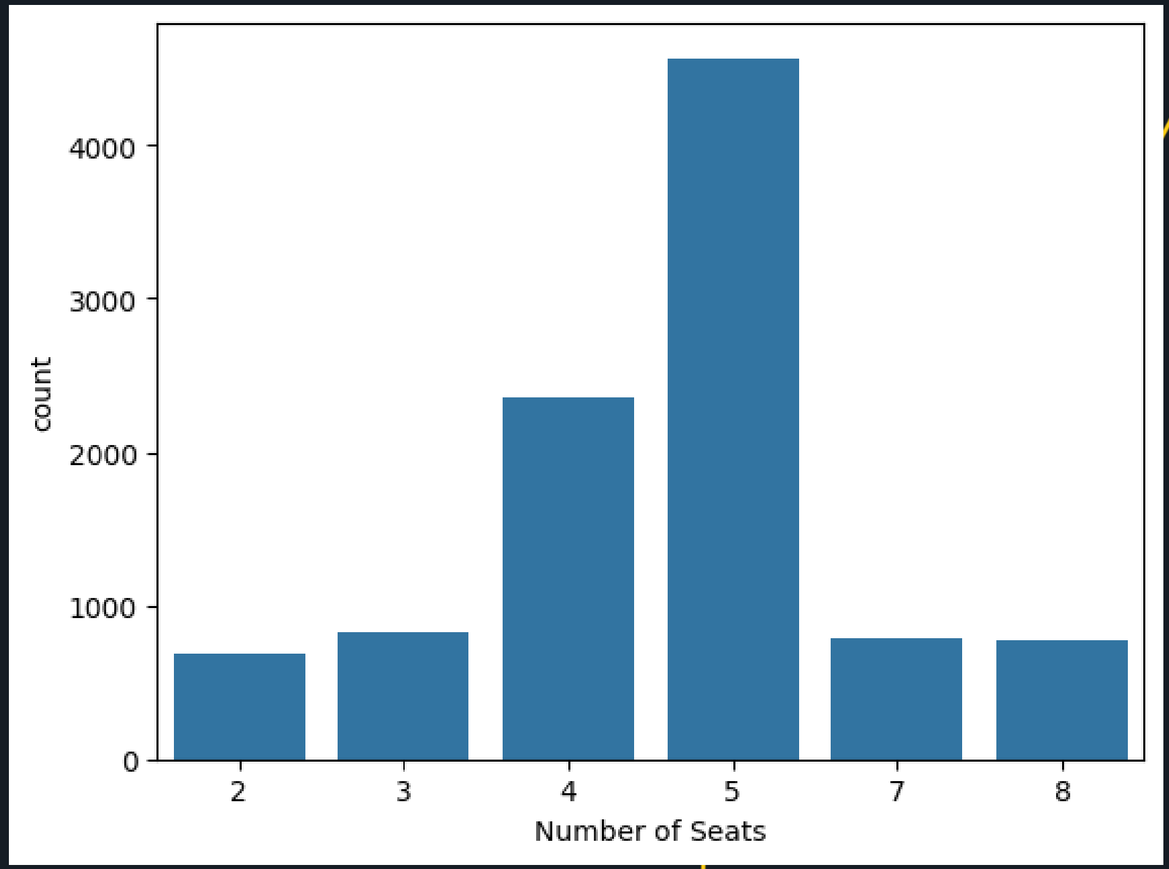


6. Distribution of Number of Seats

What the visualization shows:
The frequency of cars by their seating capacity.

Why I used this visualization:
Count plots are effective for categorical or discrete numerical data.

Key takeaway:
Most cars have a standard seating capacity (like 4–5), with fewer cars outside this range.



7. Purchased Price vs. Sold Price

What the visualization shows:
The relationship between car purchase price and its sold price.

Why I used this visualization:
A line plot with markers highlights individual transactions and trends.

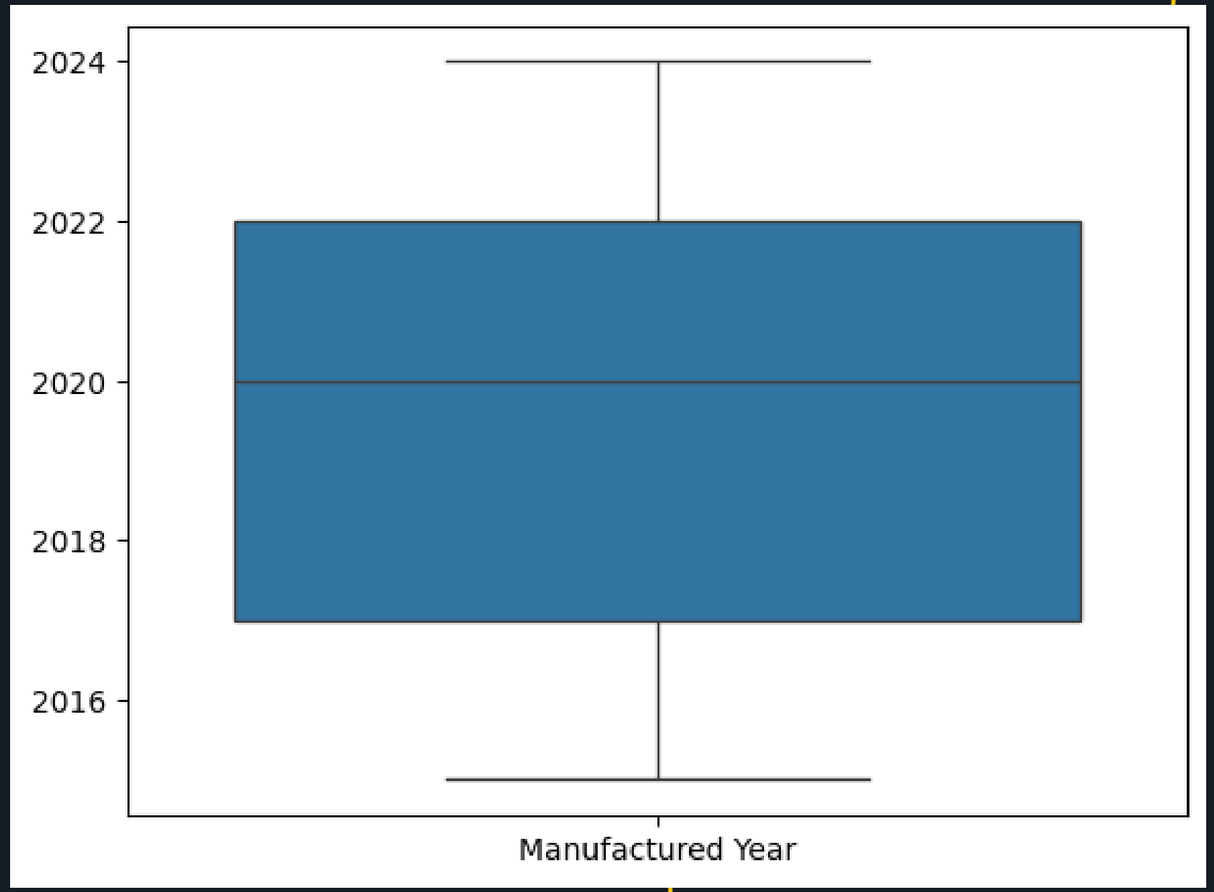
Key takeaway:
Sold prices generally follow purchased prices, but some cars resell at higher or lower values than expected.

8. Distribution of Manufactured Year

What the visualization shows:
The spread of car manufacturing years, including the median, quartiles, and outliers.

Why I used this visualization:
A box plot is ideal for detecting outliers and seeing how recent or old the cars in the dataset are.

Key takeaway:
Most cars were manufactured within a certain modern time range, but there are some outliers representing much older vehicles.



Thank You