Latihan 1

Cleaning Movie Ticket Sales Dataset

Latihan 2

Cleaning Movie Ticket Sales Dataset

Visualizations	Description
Time Trends: Line Plot	• Show how the number of journeys changes over time (date and time). • Identify peak travel hours and busy days.
Price Distribution: Histogram	• Visualize the distribution of journey prices. • Understand common price ranges and outliers.
Correlation Heatmap	• Display correlation between numerical variables (distance, duration, price, driver_rating, customer_rating). • Visualize relationships and strengths.
Duration vs. Distance: Scatter Plot	• Plot journey duration on one axis and distance on the other. • Explore any relationship between distance and travel time.
Driver vs. Customer Ratings: Side-by-Side Bar Chart	• Compare average driver and customer ratings. • Identify if there's a difference in how they rate each other.
Price vs. Ratings: Scatter Plot	• Show journey price on one axis and driver/customer ratings on the other. • Explore whether higher-priced journeys have higher ratings.
Time Series Analysis: Line Plot	• Plot the average price and ratings over the course of the month. • Identify trends or patterns over time.

Latihan 3

Analyzing Employee Dataset

1. Descriptive Statistics:

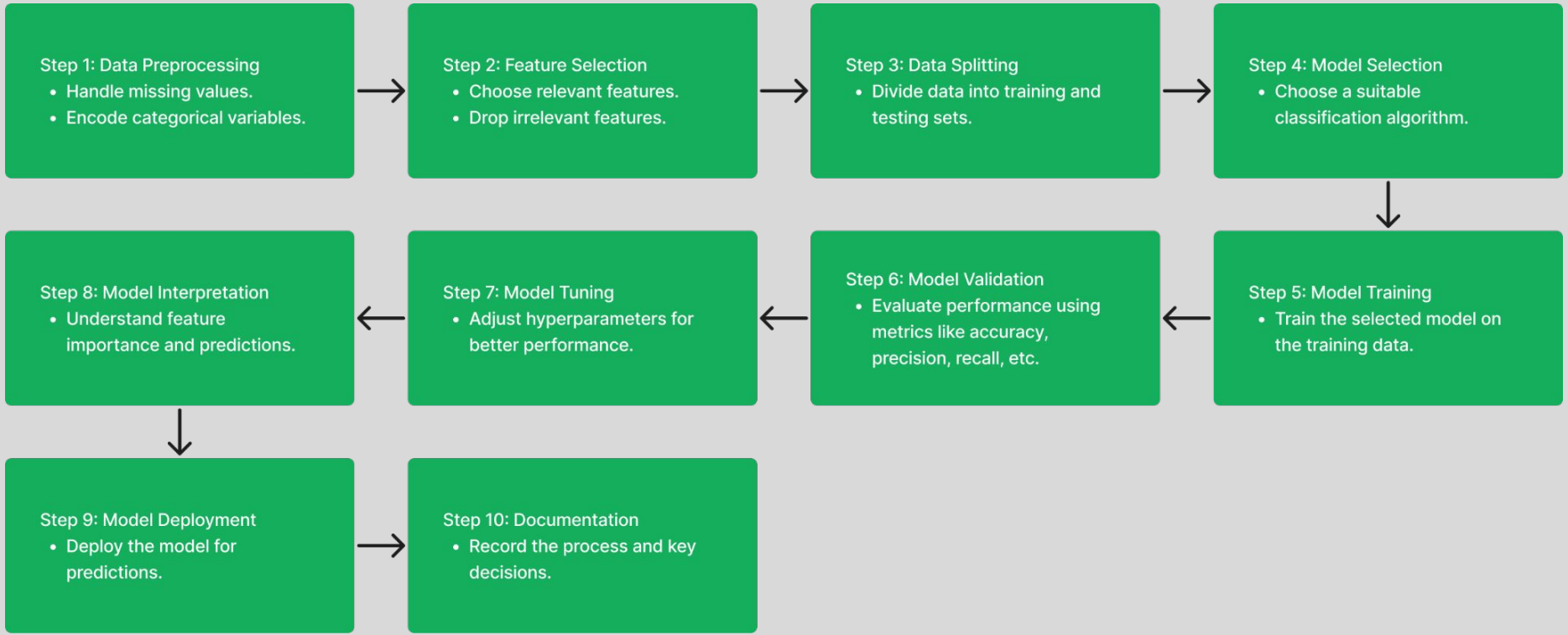
- Central Tendency: Mean, Median, Mode (umur, lama_bekerja, gaji)
- Dispersion: Range, Variance, Std. Deviation (gaji, lama_bekerja)
- Distribution: Histograms (umur, gaji)
- Frequency: Gender Count

2. Inferential Statistics:

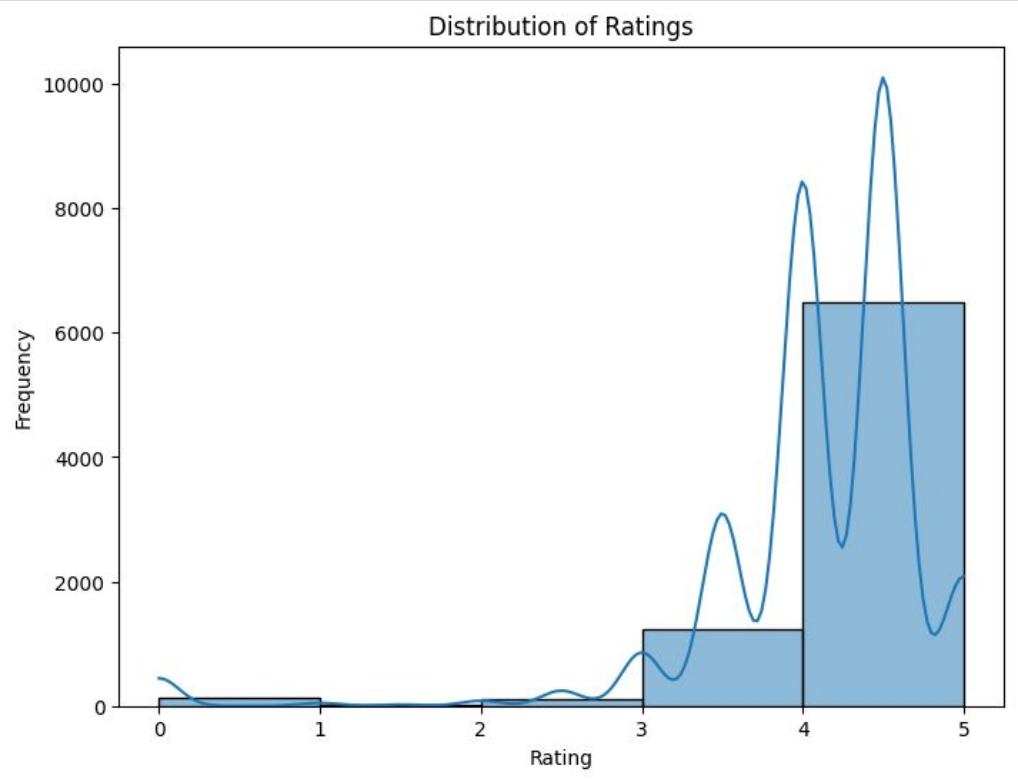
- Correlation: umur, lama_bekerja, gaji
- Hypothesis Testing: T-tests (gaji berdasarkan jenis kelamin, pendidikan)
- Regression: Predict Salary (umur, lama_bekerja)
- ANOVA (Analysis of Variance): gaji berdasarkan pendidikan

Latihan 4

Creating a Machine Learning Model for Customer Credit Card Interest Prediction



Data Visualization

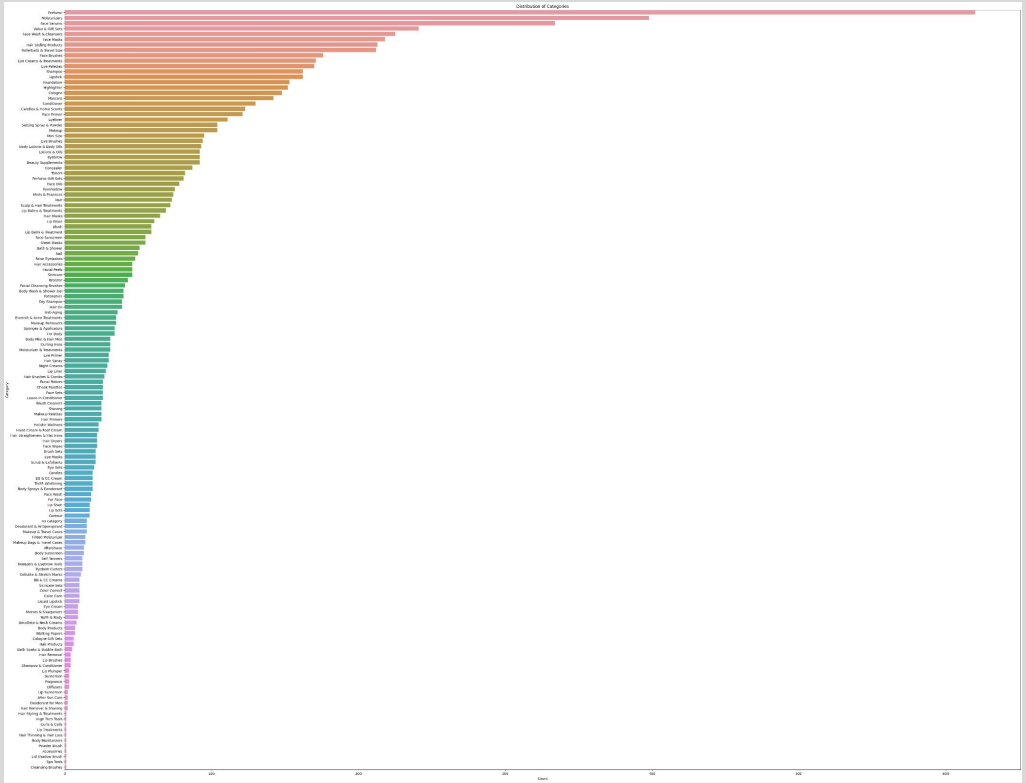


- Peaks or clusters in the histogram indicate the most common rating values, helping to identify trends in customer sentiment.
- This visualization offers insights into whether customers tend to rate products more positively (higher ratings) or negatively (lower ratings) and whether any specific rating values dominate.

Latihan 5

Machine Learning Models and Data Visualization

Data Visualization

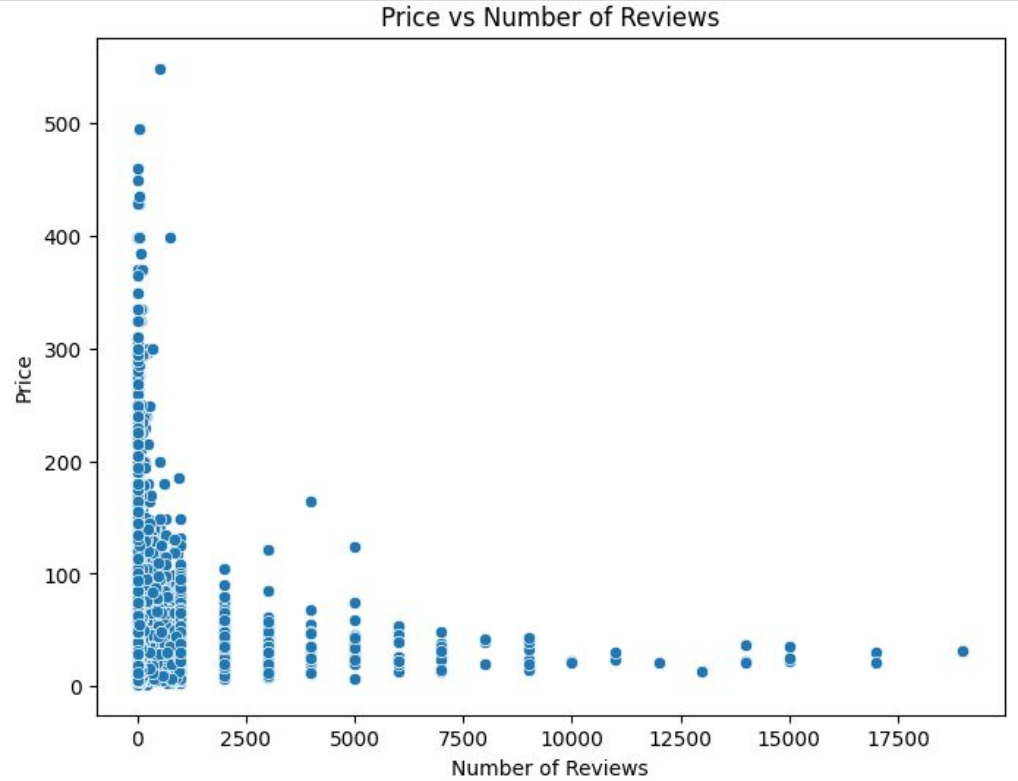


[For bigger resolution of the picture]

- The count plot displays the frequency of products in each category, providing a visual summary of category popularity.
- The order of categories can provide insights into which categories are most common or least common.
- This visualization assists in identifying which categories have more product offerings, informing marketing strategies and inventory management.

Latihan 5

Data Visualization



- The scatter plot illustrates the relationship between two numeric variables: 'price' and 'number_of_reviews'.
- Patterns or trends in the scatter plot reveal potential connections between price and customer engagement (reviews).
- The plot may show whether higher-priced products tend to attract more or fewer reviews, helping to understand the impact of pricing on customer interactions.

Latihan 5

Source Code:

https://github.com/andrewsihotang/pre_test/blob/main/exercise5.ipynb