

9 Exercises

1. **Data Access:** Please download `Purchase_Records_Dataset.csv` from the [course Github](#) and read this `csv` file to a `DataFrame`.
2. **Dataset Overview:** What are the data types and the count of non-null values in each column?
3. **Summary Statistics:** What are the summary statistics for the numerical columns in the dataset?
4. **Missing Data:** Which columns have missing values, and how many missing values does each column have?
5. What is the total number of missing values in the dataset?
6. **Selecting Specific Data Types:** Select and display only the numerical columns in the dataset.
7. **Data Type Transformation:** Convert the `Age` column to an integer data type. Verify if the conversion was successful.
8. **Unique Values:** What are the unique regions and product categories in the dataset?
9. **Count of Unique Values:** How many unique customers are present in the dataset?
10. **Value Counts:** What are the most frequent product categories?
11. **Querying Data:** Which customers made purchases in the "North" region and paid using a "Credit Card"?
12. **Filtering with Queries:** Find all online orders for electronics with a discount greater than 10%.
13. **Grouping Data:** What is the average price of products purchased in each region?
14. **Aggregating Data:** Calculate the total quantity sold for each product category by region.
15. **Pivot Table:** Create a pivot table showing the average price of products for each payment method by region.
16. **Correlation Analysis:** What is the correlation between `Age`, `Quantity`, and `Price`? Visualize it using `seaborn`.
17. **Mapping Values:** Create a column categorizing `Age` into "Young" (<40), "Middle-aged" (40-59), and "Old" (>=60).
18. **Using Apply:** Create a column calculating the total value of a transaction (`Price * Quantity`) and apply a discount.
19. **Customer Analysis:** Which customers are returning customers (more than one transaction)? What percentage of all customers are returning customers?
20. **Lead Time Analysis:** What is the average lead time for online orders across different product categories?
21. **Region-Specific Trends:** Are discounts more frequent in one region compared to others? What is the average discount by region?
22. **Payment Method Preferences:** Which payment method is most commonly used for online purchases? Is there a significant difference between regions?
23. **Product-Region Analysis:** Which product categories are most popular in each region? Create a visualization to represent this relationship.
24. **Seasonal Trends:** Are there any trends in the number of transactions over different months? Plot the monthly transaction count. You need to use `.dt` to answer this question.
25. **Top Customers:** Identify the top 10 customers based on the total monetary value of their transactions.
26. **In-Store vs. Online Trends:** What is the average discount provided for in-store purchases compared to online purchases? Is one platform more likely to offer higher discounts?