Pandas Assignment: Analyzing Used Bikes Dataset

Mustafa AbdulRazek

September 13, 2024

Task 1: Basic Exploration

Objective: Understand the structure of the dataset.

Task Description

In this task, you will perform basic exploratory operations to familiarize yourself with the dataset, including displaying the initial rows and understanding the data types.

Task

- Load the dataset using pandas.
- Display the first 10 rows of the dataset.
- Print the information of the dataset using `info()`.
- Find the total number of rows and columns in the dataset.

Task 2: Descriptive Statistics

Objective: Generate basic statistics about the dataset.

Task Description

You will use pandas functions to extract statistical insights from numerical columns and explore unique values in categorical data.

Task

- Use the `describe()` function to get statistical information about numerical columns.
- Find the unique values in the `city` column.
- Calculate the number of unique bike brands.

Task 3: Value Counts and Duplicates

Objective: Identify common values and duplicate entries.

Task Description

Learn how to find the most common entries in the dataset and identify duplicate rows.

Task

- Find the most common bike brand using `value_counts()`.
- Check for duplicates in the dataset and count how many duplicate rows exist.

Task 4: Basic Filtering

Objective: Filter data based on conditions.

Task Description

Practice filtering rows in the dataset based on simple conditions.

Task

- Filter and display bikes priced below 50,000 units.
- Find bikes that have been driven less than 10,000 kilometers.

Task 5: Data Slicing and Indexing

Objective: Practice selecting data using indexing and slicing.

Task Description

Explore data slicing techniques to extract specific rows and columns from the dataset.

Task

- Select and display only the `bike_name`, `price`, and `city` columns.
- Display data for the first 20 bikes with prices above 100,000 units.

Task 6: Grouping and Aggregation

Objective: Analyze the dataset using grouping and aggregation.

Task Description

Learn how to group data and perform aggregation to extract meaningful insights.

Task

- Group the bikes by `brand` and find the average price for each brand.
- Count the number of bikes available in each city.

Task 7: Advanced Filtering

Objective: Use multiple conditions to filter data.

Task Description

Filter the dataset using complex conditions involving multiple columns.

Task

- Filter bikes that are less than 5 years old and have more than 15 horsepower.
- Find all first-owner bikes priced above 75,000 units.

Task 8: Cleaning Operations

Objective: Identify and handle missing or irrelevant data.

Task Description

Learn how to check for and handle missing data within a dataset.

Task

- Check for missing values in the dataset.
- If there are any, fill missing values in the `kms_driven` column with the mean of the column.

Task 9: Sorting Data

Objective: Sort data based on specific criteria.

Task Description

Use sorting operations to organize data based on various columns.

Task

- Sort the dataset by `price` in descending order and display the top 10 most expensive bikes.
- Sort by `age` and `kms_driven` to find the oldest bike with the least kilometers driven.