CS23334 -Fundamental Of Data Science

240701463

EDA - Quantitative and Qualitative Analysis

Exp : 06

Date: 02-09-2025

Aim:

To understand the importance of **Exploratory Data Analysis (EDA)** by performing both **Quantitative** (numerical) and **Qualitative** (categorical) analyses on a dataset using Python and the **Seaborn** visualization library.

Algorithm:

1. **Load Data:** Load the `tips` dataset into a DataFrame.

2. **Univariate Quantitative EDA:** Use displot to visualize the distribution of total_bill.

3. **Bivariate Quantitative EDA:** Use jointplot (scatter, reg, hex) to analyze the relationship between tip and total_bill.

4. **Multivariate EDA:** Use pairplot to view relationships across all numerical columns, optionally coloring by categorical features (time, day).

5. **Correlation Analysis:** Use tips.corr() and heatmap to visualize linear correlations between numerical variables.

6. **Univariate Qualitative EDA:** Use countplot for frequencies of categorical features (day, sex) and value_counts().plot() for pie/bar charts.

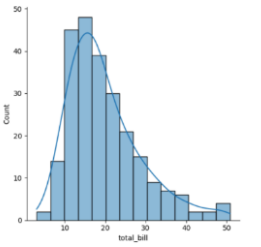
7. **Outlier Check:** Use boxplot to identify the distribution and outliers in numerical features (total_bill, tip).

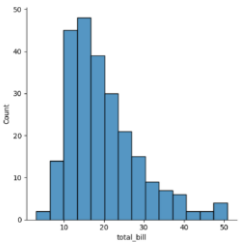
Code:

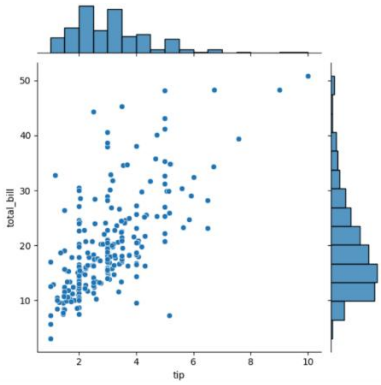
```
import seaborn as sns
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
tips = sns.load_dataset('tips')
tips.head()
sns.displot(tips.total_bill, kde=True)
sns.displot(tips.total_bill, kde=False)
sns.jointplot(x=tips.tip, y=tips.total_bill)
# sns.jointplot(x=tips.tip, y=tips.total_bill, kind="reg")
sns.jointplot(x=tips.tip, y=tips.total_bill, kind="hex")
sns.pairplot(tips)
sns.pairplot(tips, hue='time')
sns.pairplot(tips, hue='day')
tips.time.value_counts()
sns.countplot(tips.day)
sns.countplot(tips.sex)
tips.sex.value_counts().plot(kind='pie')
tips.sex.value_counts().plot(kind='bar')
sns.countplot(tips[tips.time=='Dinner']['day'])
sns.heatmap(tips.corr(numeric_only=True), annot=True)
sns.boxplot(tips.total_bill)
sns.boxplot(tips.tip)
```

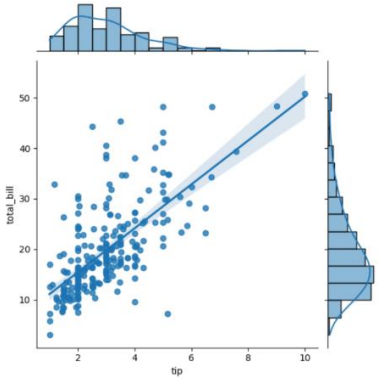
Output:

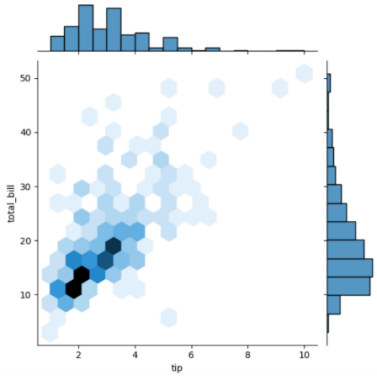
	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3
3	23.68	3.31	Male	No	Sun	Dinner	2
4	24.59	3.61	Female	No	Sun	Dinner	4

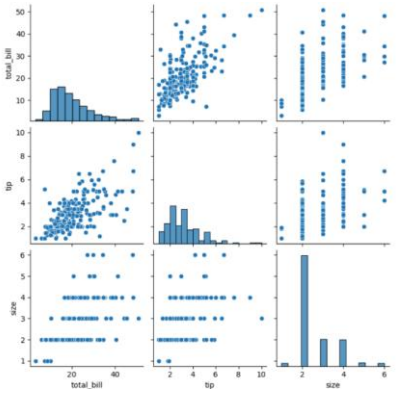


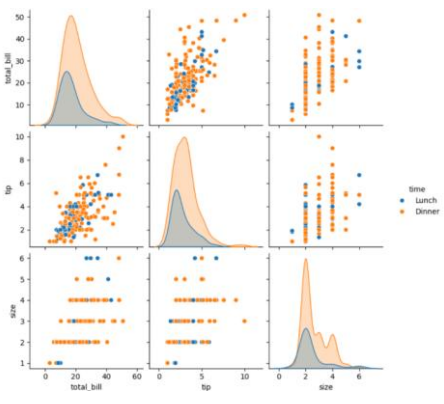


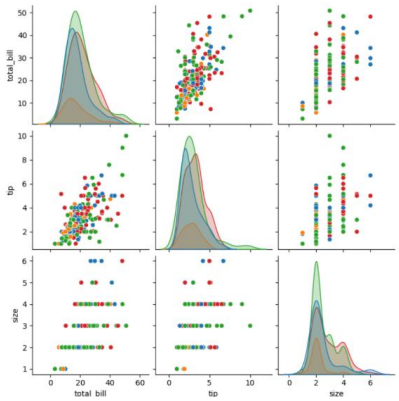


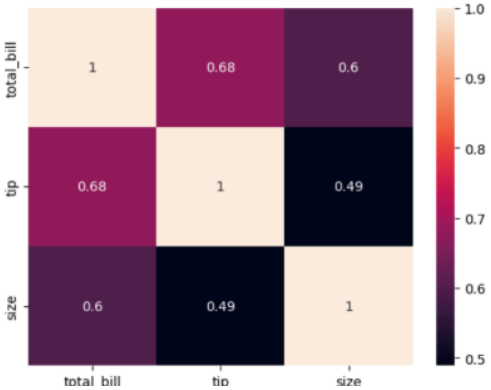


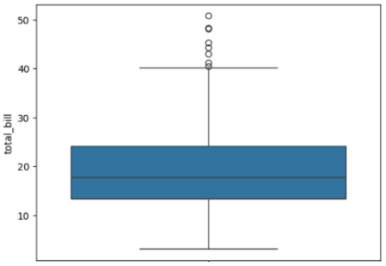


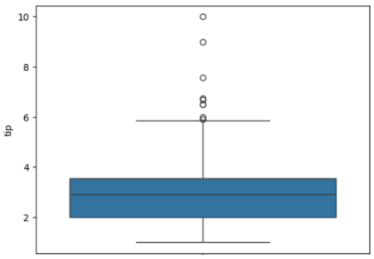


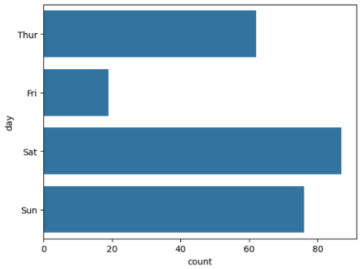


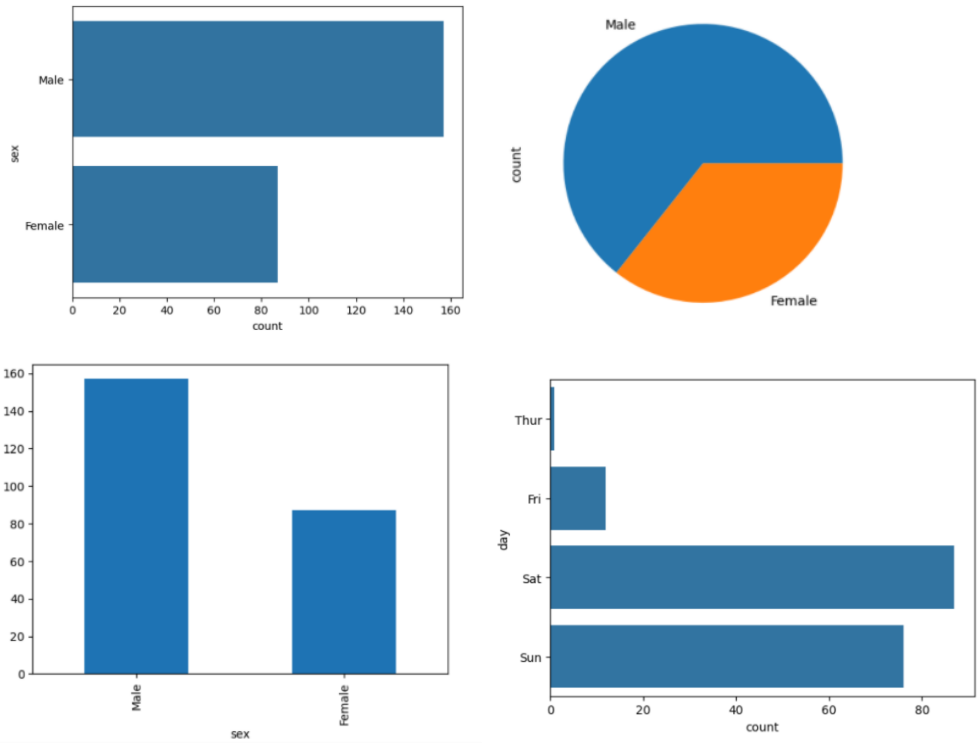












Result:

The experiment successfully executed various EDA techniques, demonstrating both **Quantitative Analysis** (using displot, jointplot, boxplot, and heatmap for numerical features) and **Qualitative Analysis** (using countplot and value_counts for categorical features). The analysis highlighted key findings in the tips dataset, such as the positive correlation between total_bill and tip and the busiest days/times, thereby demonstrating the core importance of EDA in summarizing and understanding data structure.