

Explanation of Visualizations

1. **Scatter Plot** – Visualizes relationships between two numerical variables.
2. **Histogram** – Shows the distribution of a single numerical variable.
3. **Boxplot** – Displays the spread and outliers in numerical data, grouped by a categorical variable.
4. **Bar Chart** – Compares the mean values of numerical data grouped by a category.
5. **Heatmap** – Shows correlations between numerical variables in a color-coded matrix.

```
import numpy as np
```

```
import pandas as pd
```

```
import matplotlib.pyplot as plt
```

```
import seaborn as sns
```

```
# Generate a sample dataset
```

```
np.random.seed(42)
```

```
data = pd.DataFrame({
```

```
    'A': np.random.randn(100), # Normally distributed data
```

```
    'B': np.random.rand(100) * 10, # Uniformly distributed data
```

```
    'C': np.random.randint(1, 100, 100), # Random integers
```

```
    'D': np.random.choice(['Category1', 'Category2', 'Category3'], 100) # Categorical data
```

```
})
```

```
# Scatter Plot
```

```
plt.figure(figsize=(6, 4))
```

```
sns.scatterplot(x=data['A'], y=data['B'])
```

```
plt.title('Scatter Plot')
```

```
plt.show()
```

```
# Histogram
```

```
plt.figure(figsize=(6, 4))
```

```
sns.histplot(data['A'], bins=20, kde=True) # kde=True adds a kernel density estimate
```

```
plt.title('Histogram of A')
```

```
plt.show()
```

```
# Box Plot
```

```
plt.figure(figsize=(6, 4))
```

```
sns.boxplot(x=data['D'], y=data['C']) # Box plot for categorical vs numerical
```

```
plt.title('Boxplot of C grouped by D')
```

```
plt.show()
```

```
# Bar Chart
```

```
plt.figure(figsize=(6, 4))
```

```
sns.barplot(x=data['D'], y=data['C'])
```

```
plt.title('Bar Chart of C grouped by D')
```

```
plt.show()
```

```
# Heatmap (Correlation Matrix)

plt.figure(figsize=(6, 4))

sns.heatmap(data[['A', 'B', 'C']].corr(), annot=True, cmap='coolwarm', fmt=".2f")

plt.title('Heatmap of Correlation Matrix')

plt.show()
```

instead of drawing plots btw 2 col is it possible to use all col and if yes then tell me the diff

Yes, you can visualize **all numerical columns** in your dataset instead of just two columns. The difference lies in how you structure the plots. Let's explore how each plot changes when applied to **all numerical columns** instead of just two.

1 Scatter Plot: Pairplot (All Numerical Columns)

Instead of a scatter plot between two specific columns, we use **Seaborn's pairplot()** to create scatter plots for all numerical column combinations.

Code:

```
sns.pairplot(data)
plt.show()
```

Difference:

- This will generate multiple scatter plots showing relationships between **every pair** of numerical columns.
 - It helps identify correlations and trends between multiple variables at once.
-

2 Histogram: Multiple Columns

Instead of a single histogram, we can plot multiple histograms for all numerical columns.

Code:

```
data.hist(figsize=(10, 6), bins=20, edgecolor='black')
plt.suptitle('Histograms of All Numerical Columns')
plt.show()
```

Difference:

- This generates **separate histograms** for each numerical column.
 - Helps compare the distributions of multiple numerical variables.
-

3 Boxplot: All Columns

To compare distributions across all numerical columns, we can plot **a single boxplot** with all numerical features.

Code:

```
plt.figure(figsize=(8, 5))
sns.boxplot(data=data)
plt.xticks(rotation=45) # Rotate labels if needed
plt.title('Boxplots of All Numerical Columns')
plt.show()
```

Difference:

- Instead of grouping by a categorical variable, this shows the distribution and outliers for **all numerical columns** side by side.
-

4 Bar Chart: Summary Statistics

For bar plots, we usually compare categorical vs. numerical data. If you only have numerical data, you can summarize them and plot mean values.

Code:

```
data.mean().plot(kind='bar', figsize=(6, 4))
plt.title('Mean Values of Numerical Columns')
plt.ylabel('Mean')
plt.show()
```

Difference:

- Instead of a categorical comparison, this bar chart shows **summary statistics** (mean, median, etc.) for all numerical columns.

5 📊 Heatmap: Correlation Matrix

The best way to visualize all numerical columns together is using a heatmap.

Code:

```
plt.figure(figsize=(6, 5))
sns.heatmap(data.corr(), annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Heatmap')
plt.show()
```

Difference:

- Instead of plotting one or two variables, the heatmap shows **correlations** between all numerical columns at once.
- Helps identify strong relationships between variables.

Summary of Differences

Plot Type	Two Columns	All Columns
Scatter Plot	Compares 2 variables	Pairplot compares all numerical columns
Histogram	Shows distribution of 1 column	Multiple histograms for each numerical column
Boxplot	Compares one variable across categories	Shows distributions of all numerical columns
Bar Chart	Compares numerical values across categories	Summarizes statistics like mean/median
Heatmap	Correlation between two columns	Correlation matrix for all numerical columns

Let me know if you need more clarification! ☐