

# EDA-Histogram Analysis

---

Histogram analysis is a crucial step in Exploratory Data Analysis (EDA) that helps to understand the distribution of a continuous variable. Here's a summary:

## What is a Histogram?

A histogram is a graphical representation of the distribution of a numerical variable, showing the number of data points that fall into each interval or bin.

## Purpose of Histogram Analysis in EDA:

- 1. Understand Distribution Shape:** Histograms help to identify if the data follows a normal distribution, skewed distribution, bimodal distribution, or any other type.
- 2. Identify Outliers and Anomalies:** By examining the histogram, you can spot unusual values that don't fit the overall pattern.
- 3. Determine Data Range:** Histograms provide an idea of the minimum, maximum, and typical value ranges in the dataset.
- 4. Visualize Distribution Parameters:** You can estimate parameters like mean, median, quartiles, and interquartile range (IQR) from the histogram.

## Example:

Suppose we have a dataset containing exam scores for a group of students. We want to use a histogram to understand the distribution of these scores.

## Dataset:

Student ID	Exam Score
1	85
2	90
3	78
...	...
1000	92

## Histogram:

Here's a simple histogram with the exam scores, using a bin width of 10 points:

Exam Score

```
-----  
(60-70) : 20  
(70-80) : 50  
(80-90) : 100  
(90-100): 30
```

### Interpretation:

From this histogram, we can infer:

1. The scores are skewed to the right, indicating a higher proportion of students scoring above 80.
2. There are no outliers or anomalies in the data; all values fall within expected ranges.
3. Most students scored between 80-90, suggesting that there is some central tendency in the data.

### Code Example:

Using Python with Pandas and Matplotlib libraries:

```
import pandas as pd  
import matplotlib.pyplot as plt  
  
# Load dataset  
df = pd.read_csv('exam_scores.csv')  
  
# Plot histogram  
plt.hist(df['Exam Score'], bins=10, edgecolor='black')  
plt.title('Histogram of Exam Scores')  
plt.xlabel('Score Range')  
plt.ylabel('Frequency')  
plt.show()
```

By analyzing the histogram, you can gain insights into the distribution of your data and make informed decisions for further analysis or modeling.