



# *Apple Store Reviews*



-BY SHIVANI NAGAR

Statistics | Descriptive | Inferential



# Introduction :

This project analyzes a dataset of Apple Store product reviews to uncover meaningful insights through descriptive and inferential statistics.

As a Data Analyst, the objective is to explore user engagement and satisfaction by summarizing key metrics like ratings and likes, performing hypothesis testing, and leveraging statistical concepts such as the Central Limit Theorem and correlation analysis.

The findings aim to provide actionable insights, enabling stakeholders to optimize product offerings and enhance user experiences.

# Table of Contents :

- **Central Tendency Analysis**
  - Mean, Median, and Mode of App Ratings
  - Best Measure for Central Tendency
- **Spread of Data**
  - Range and Interquartile Range (IQR) of Purchase Amount
  - Interpretation of Data Spread
- **Variability Analysis**
  - Variance and Standard Deviation of Likes
  - Insights from Standard Deviation

- **Correlation Analysis**
  - Correlation Between Likes and Ratings
  - Nature of Correlation (Positive, Negative, or None)
- **Distribution Analysis**
  - Distribution of App Ratings
  - Skewness and User Satisfaction
- **Inferential Statistics**
  - Hypothesis Test: Instagram vs. WhatsApp Average Ratings
- **Central Limit Theorem**
  - Sampling Distribution of Ratings
  - Relationship to the Central Limit Theorem

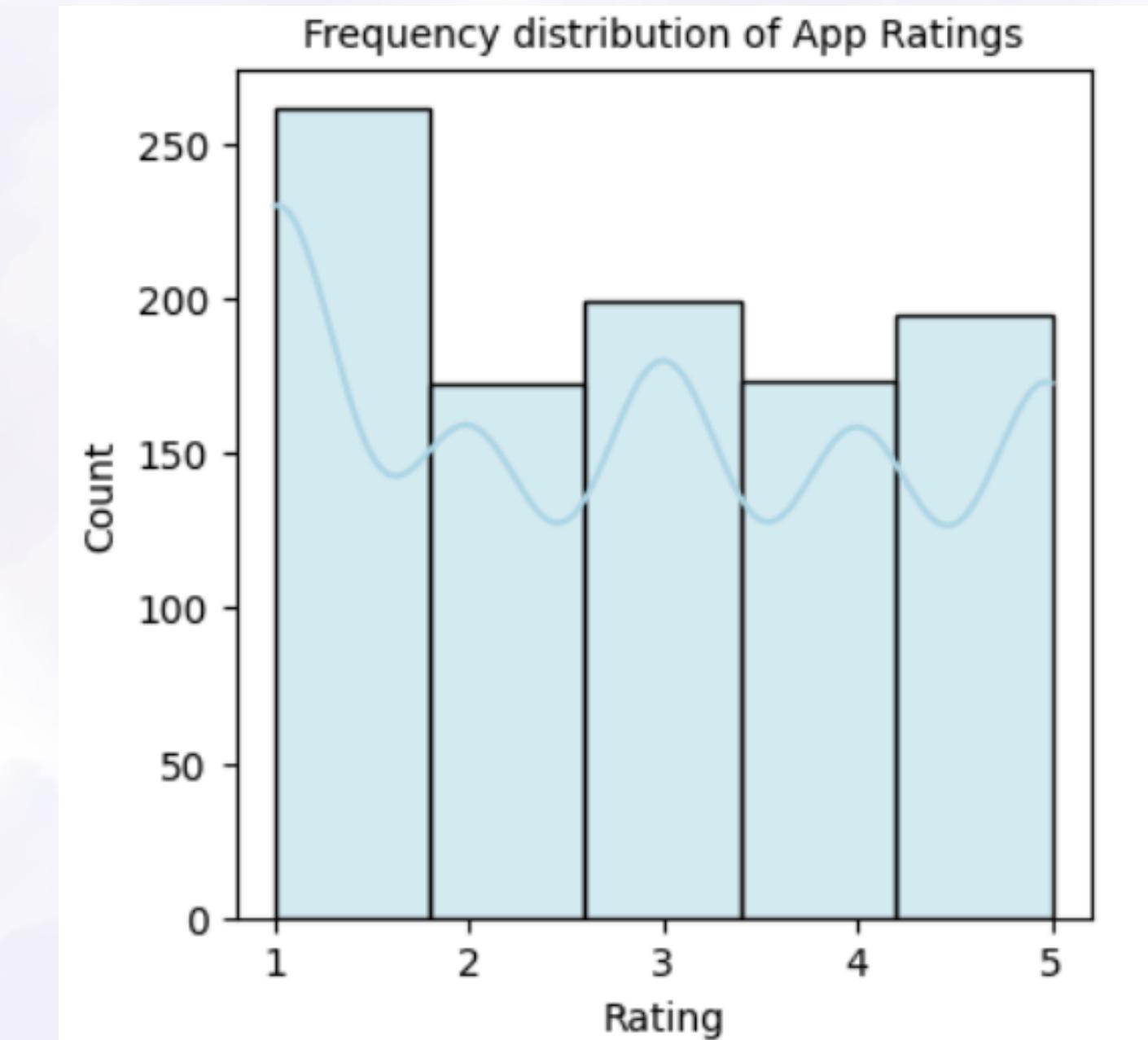
## Central Tendency Analysis

- Mean, Median, and Mode of App Ratings
- Best Measure for Central Tendency

- **Mean (2.869)** is close to the **Median (3.0)**, indicating a nearly symmetrical distribution.
- **Mode (1.0)** is significantly lower, suggesting a cluster of low ratings that skews the data.

### Conclusion:

The **Median (3.0)** is the best measure of central tendency as it is unaffected by outliers and better represents the typical rating.



## Spread of Data

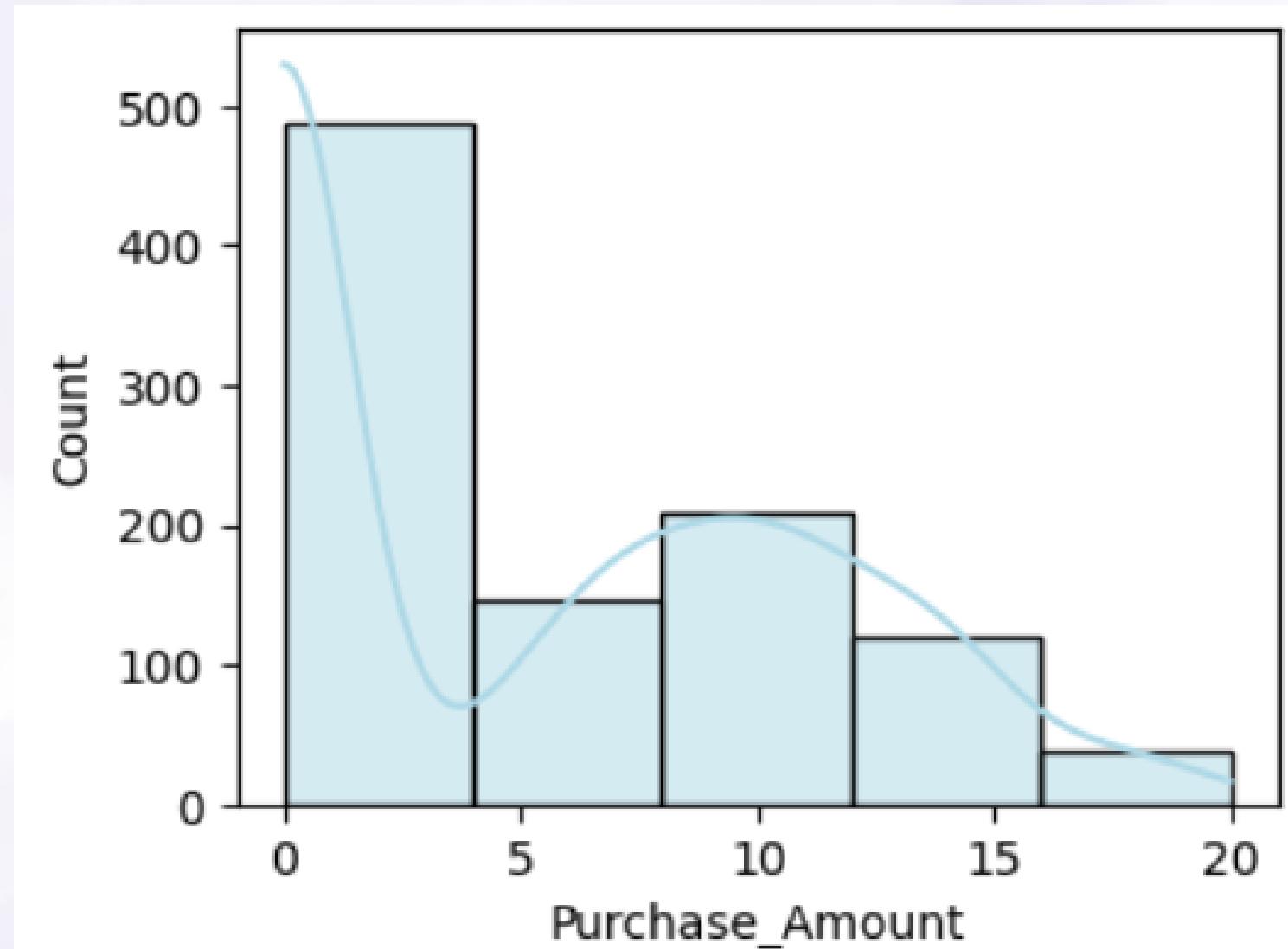
- Range and Interquartile Range (IQR) of Purchase Amount
- Interpretation of Data Spread

The Range of the `Purchase_Amount` is **19.97**, which is the difference between the maximum (**19.97**) and minimum (**0.0**) purchase amounts.

The **IQR** is **10.19**, calculated as the difference between the third quartile (**Q3=10.19**) and the first quartile (**Q1=0.0**).

- The data has a large spread, with some high-value purchases pushing the maximum to \$19.97. However, the low Q1 shows a cluster of free or minimal-value purchases.
- The relatively high IQR (\$10.19) compared to the Range (\$19.97) indicates that the variability in the middle 50% of data is significant.

This implies diverse spending behaviors among users.



## Variability Analysis

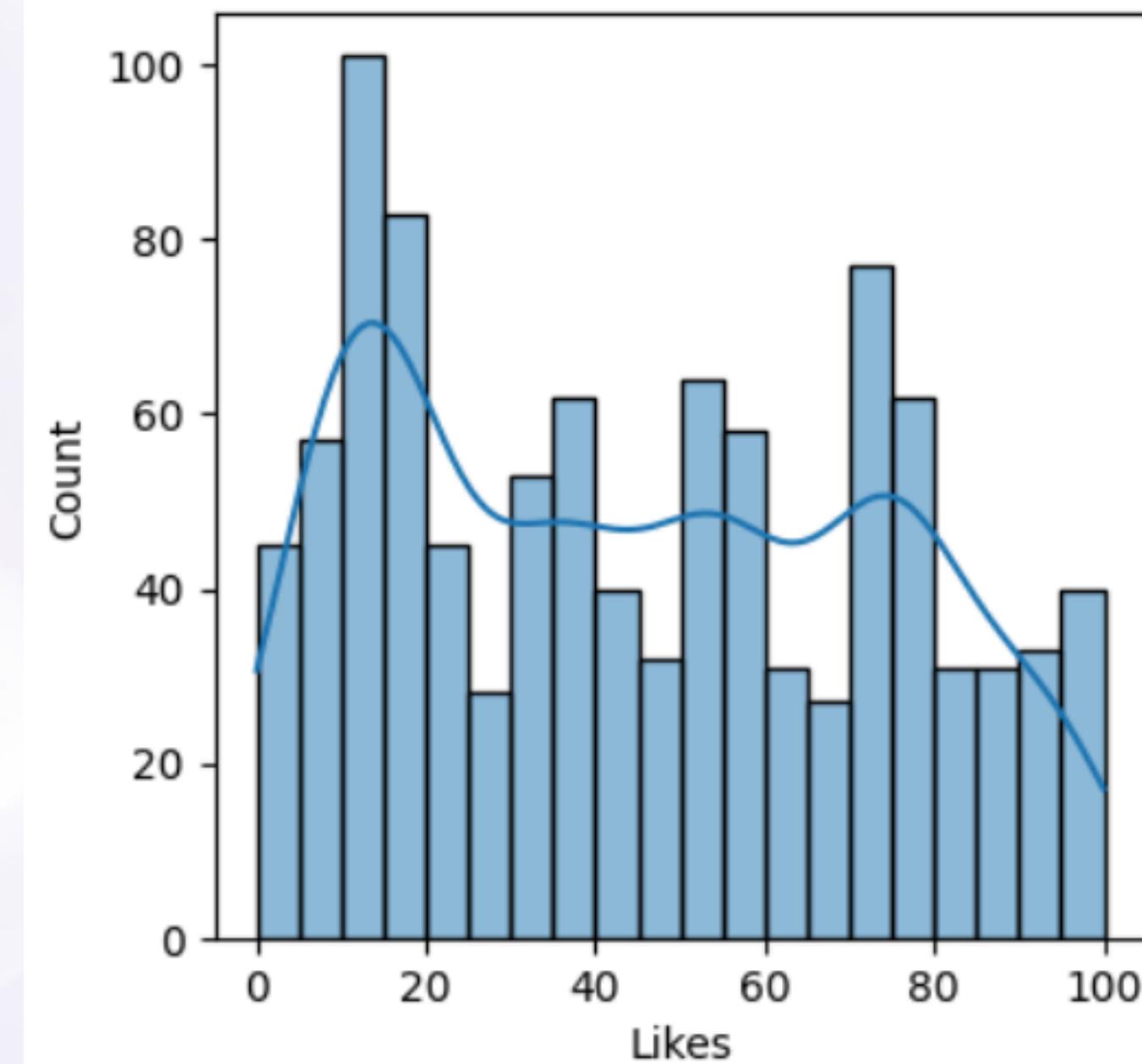
- Variance and Standard Deviation of Likes
- Insights from Standard Deviation

### Standard Deviation (28.69) :

The standard deviation (28.69) shows that while there is moderate variability in the number of likes received, the spread is not excessively wide. This suggests that most reviews perform similarly in terms of likes, with occasional outliers receiving significantly more or fewer likes.

### Variance (822.85) :

The variance quantifies the spread of the data in squared units. The relatively high value of 822.85 reflects the presence of variability in the number of likes but is harder to interpret intuitively compared to the standard deviation.



## Correlation Analysis

- Correlation Between Likes and Ratings
- Nature of Correlation (Positive, Negative, or None)

**Value of Correlation (0.8425) :**

Correlation values range from -1 to 1:

+1 : Perfect positive correlation.

0 : No correlation.

-1 : Perfect negative correlation.

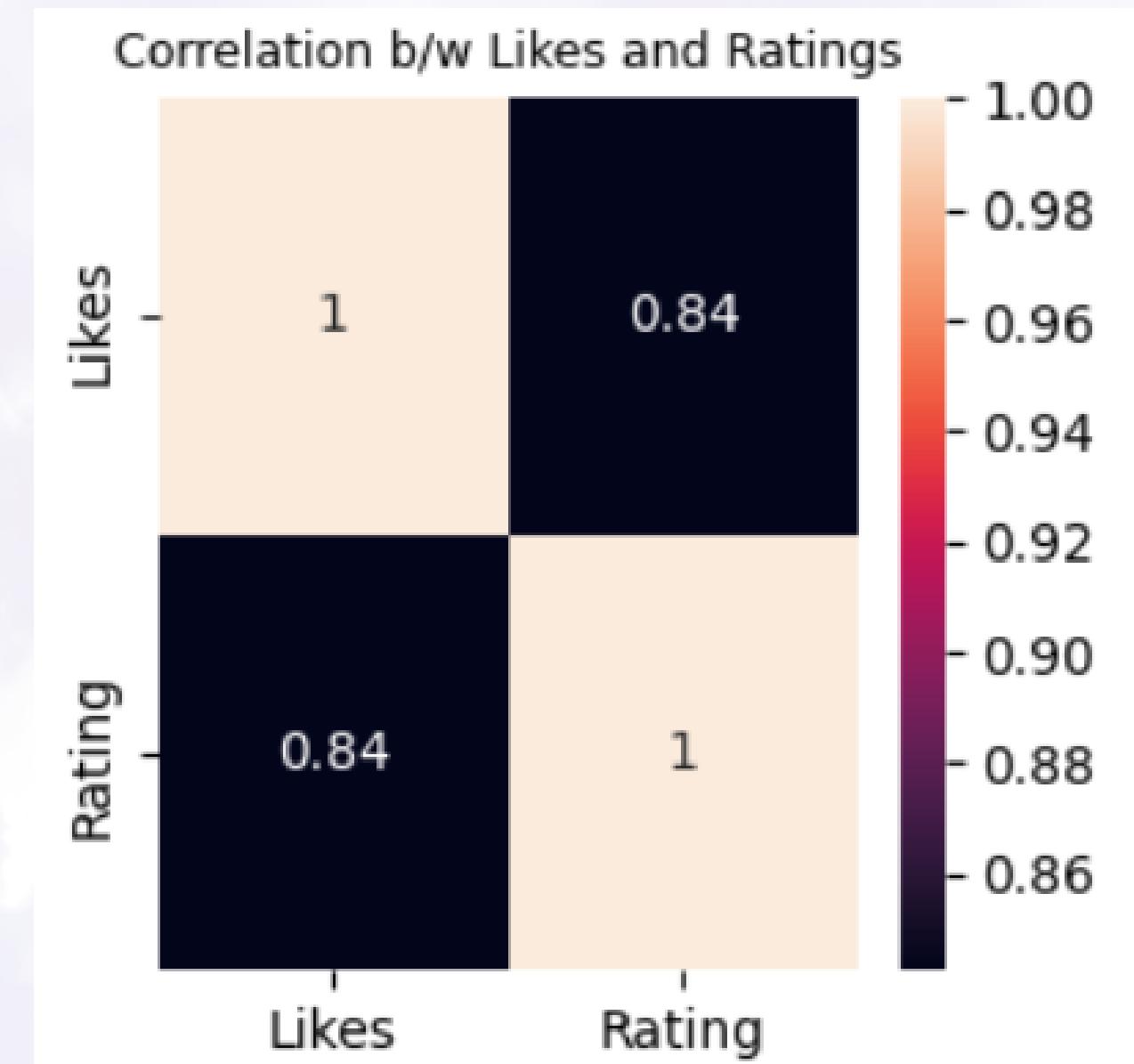
A correlation of **0.8425** is close to +1, suggesting a strong positive relationship.

**What It Means :**

As the rating of a review increases, the number of likes it receives also tends to increase.

High ratings are strongly associated with high engagement (likes),

indicating that users are more likely to like reviews they perceive as positive or useful.



## Distribution Analysis

- Distribution of App Ratings
- Skewness and User Satisfaction

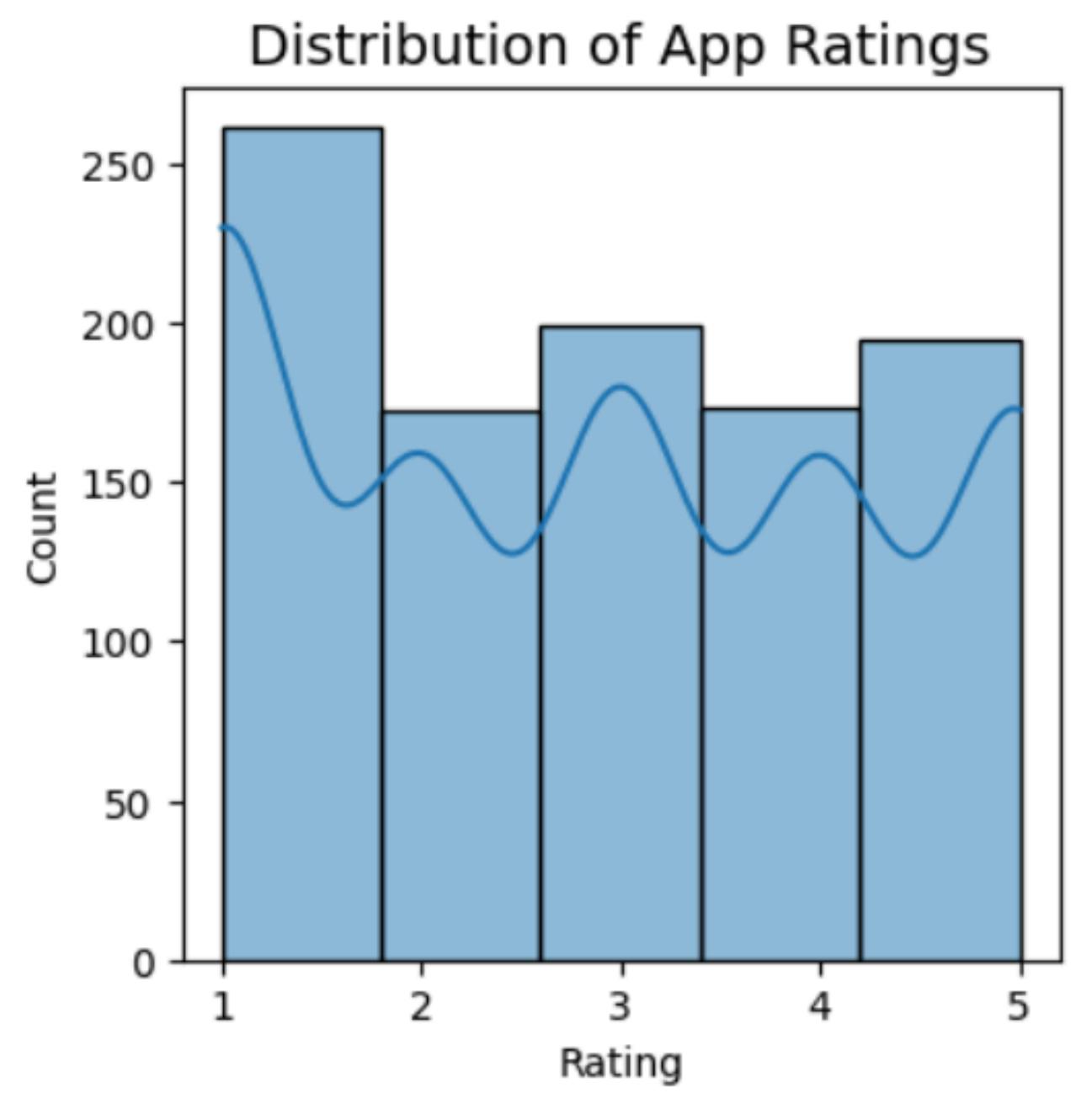
A skewness value close to 0 indicates that the distribution is approximately symmetrical.

**Skewness Value:** The skewness of the app ratings is **0.1018**, which is **close to zero**. This indicates an approximately symmetrical distribution with a slight positive skew.

The histogram confirms this, as most ratings are clustered near the mean with a small rightward tail.

**Conclusion:** The **nearly symmetrical distribution** suggests balanced user satisfaction.

This indicates that the app performs moderately well in meeting user expectations.



# Inferential Statistics

## 1. Hypothesis Test: Instagram vs. WhatsApp Average Ratings

Define the Hypotheses

Null Hypothesis ( $H_0$ ):

- Average rating for Instagram is not significantly higher than the average rating for WhatsApp.

Alternative Hypothesis ( $H_a$ ):

- Average rating for Instagram is significantly higher than the average rating for WhatsApp.

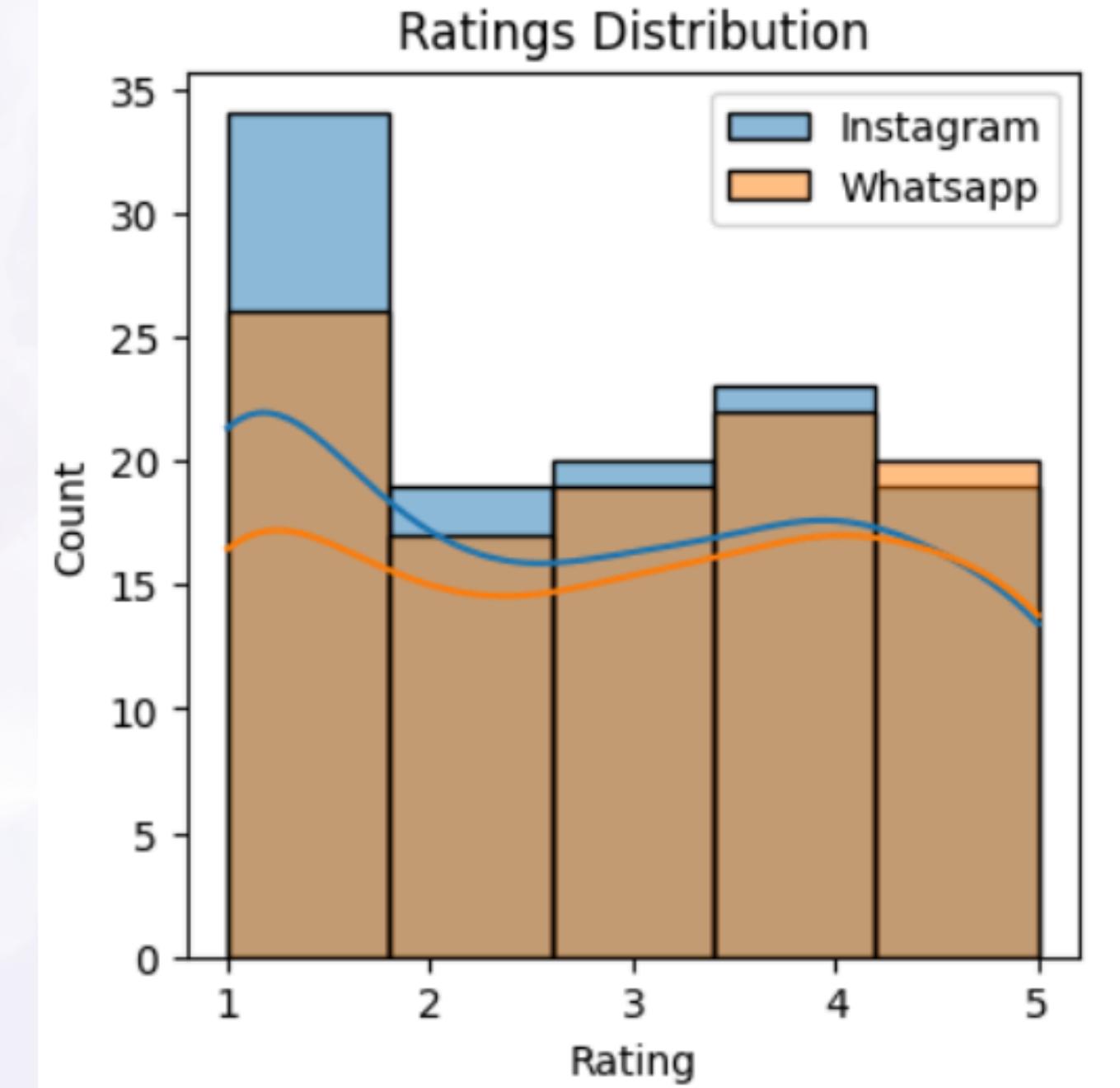
T-Test Results

- T-Statistic: -0.797
- P-Value: 0.787

Interpretation:

Since  $p > 0.05$ , we fail to reject the null hypothesis ( $H_0$ ).

This means there is insufficient evidence to conclude that Instagram's average rating is significantly higher than WhatsApp's average rating at the 95% confidence level.



## Central Limit Theorem

- Sampling Distribution of Ratings
- Relationship to the Central Limit Theorem

### Results :

1. Population Mean: 2.869

2. Sampling Mean: 2.8782

- The sampling mean is very close to the population mean, as expected under the CLT.

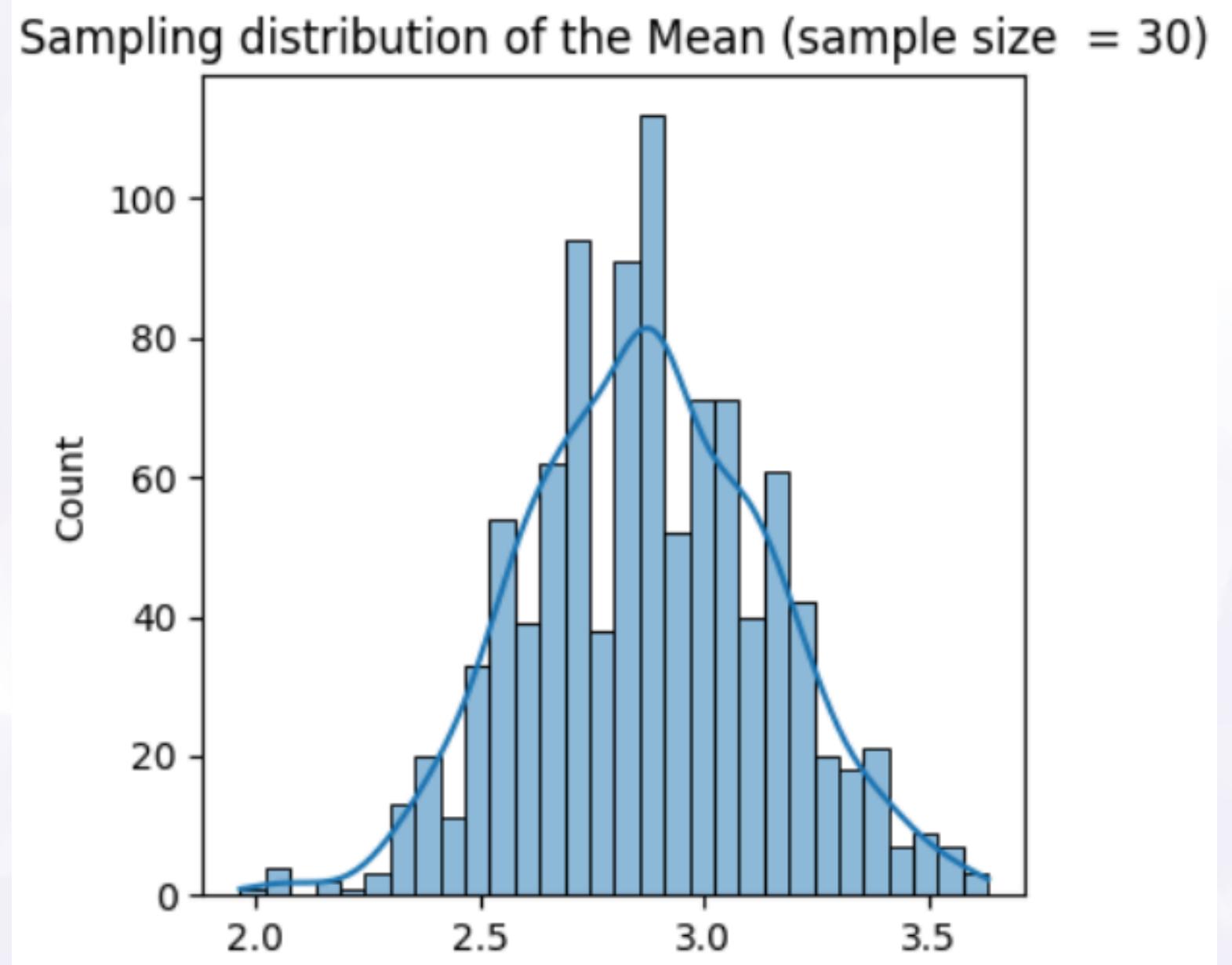
3. Standard Error: 0.268

- Indicates the variability of the sample means is much smaller than the variability in the population.

### Conclusion :

The analysis confirms the Central Limit Theorem: the sampling distribution of the mean is normal with a mean close to the population mean (2.869), and the standard error (0.268) shows reduced variability in sample means.

This enables reliable statistical inference.



# *Resources for Further Exploration !!*

I've attached the following resources for your reference:

1. Dataset Link: The Apple Store Reviews dataset used for this analysis.
2. Python Analysis File (.ipynb): A complete notebook containing the analysis, visualizations, and all the libraries used.



# Let's Connect !!

Thank you for reviewing the analysis of the Apple Store Reviews dataset.  
I'm passionate about turning data into actionable insights and would love to connect!

LinkedIn: [Shivani Nagar](#)

Email : [ShivaniEmail](#)

[Github here](#)

Let's collaborate to uncover more stories hidden in the data!