



APPLE STORE REVIEWS

Statistical Analysis

Presented by :- Abhishek pandey



CONTENT

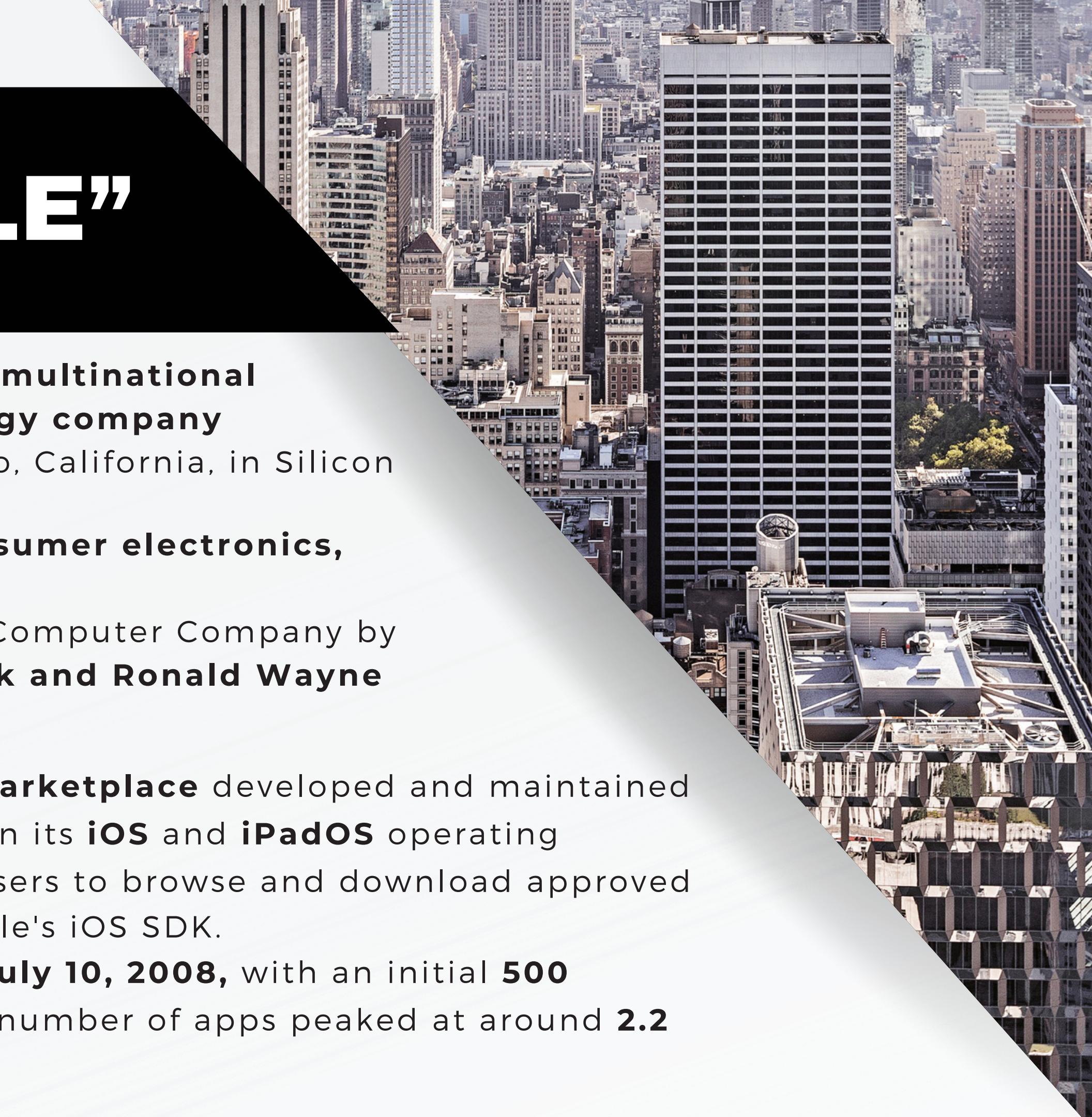
- About “**APPLE**”
- Our Goal
- Central Tendency
- Range and IQR
- Variance and Standard Deviation
- Hypothesis Test
- Correlation
- App ratings
- Central Limit Theorem
- Conclusion
- Connect with me

ABOUT “APPLE”



- **Apple Inc.** is an **American multinational corporation and technology company** headquartered in Cupertino, California, in Silicon Valley.
- It is best known for its **consumer electronics, software, and services**.
- Founded in **1976** as Apple Computer Company by **Steve Jobs, Steve Wozniak and Ronald Wayne**

- **The App Store** is an **app marketplace** developed and maintained by **Apple** for mobile apps on its **iOS** and **iPadOS** operating systems. The store allows users to browse and download approved apps developed within Apple's iOS SDK.
- The App Store opened on **July 10, 2008**, with an initial **500** applications available. The number of apps peaked at around **2.2 million in 2017**



OUR GOAL

DATASET
LINK

- As a **Data Analyst** you will apply key statistical concepts to **analyze** a dataset of **Apple Store product reviews**.
- Your objective is to **perform Descriptive and Inferential statistical analysis** to draw meaningful insights from the data.

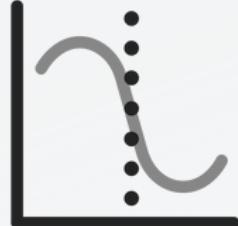


CENTRAL TENDENCY

The central tendency is stated as the **statistical measure** that **represents the single value of the entire distribution or a dataset.**

It aims to provide an accurate description of the entire data in the distribution.

Mean : 2.869



The mean is slightly below 3, suggesting that extreme low ratings (like 1) are pulling the average down.



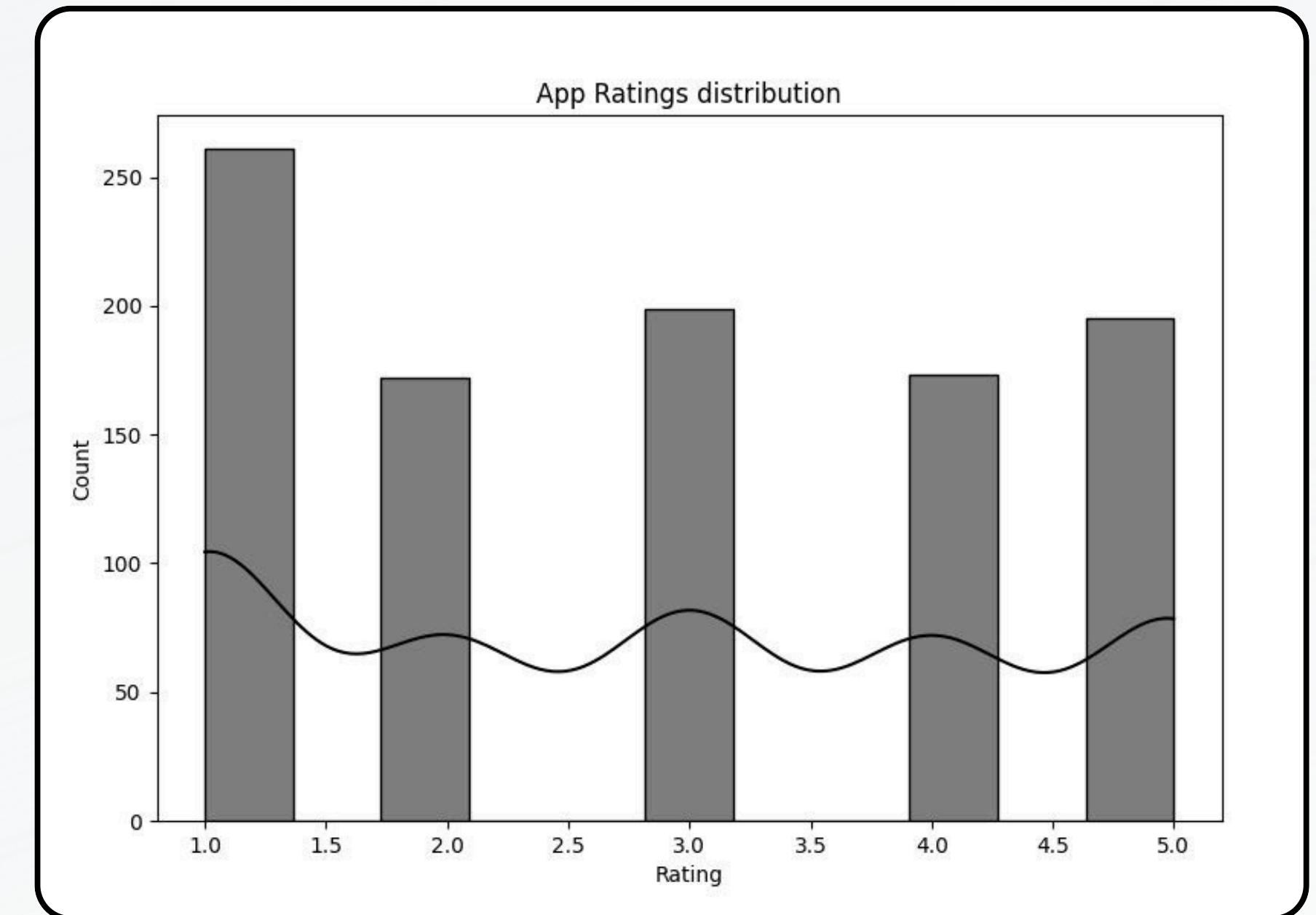
Median : 3

The median indicates that half of the ratings are below and half are above this value, making it a better representation of central tendency in this case.



Mode : 4

The mode being the most frequent rating indicates a significant number of users rated the app poorly.



RANGE AND IQR

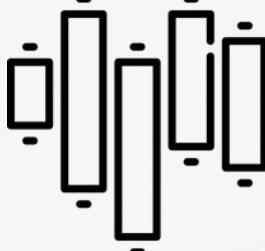
Range and interquartile range are two ways to measure the spread of values in a dataset:

- **Range**: Measures the difference between the minimum and maximum values in a dataset.
- **Interquartile Range (IQR)**: Measures the difference between the **first quartile (25th percentile)** and **third quartile (75th percentile)** in a dataset.



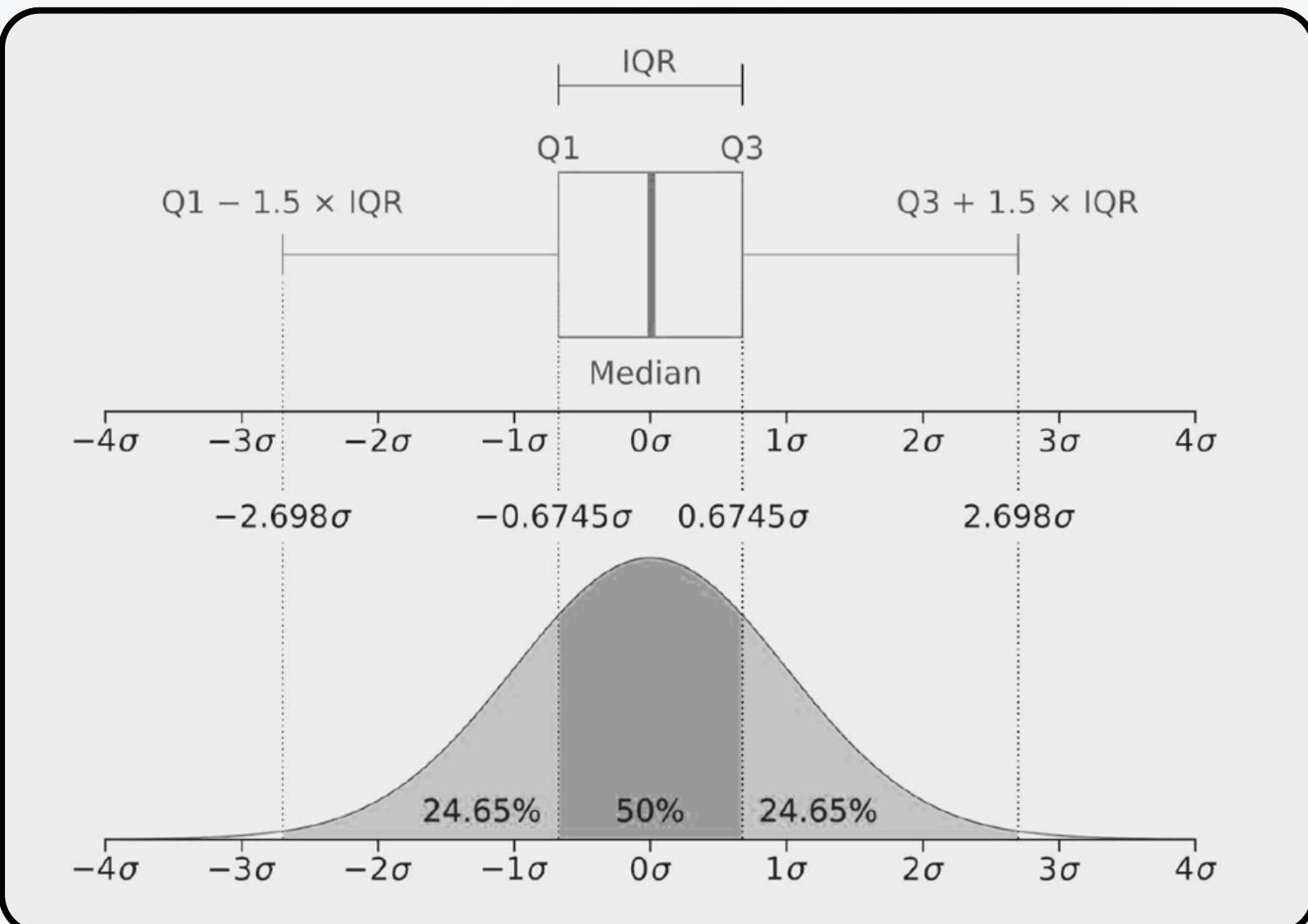
Range : 19.97

The range of 19.97 provides a comprehensive view of the variability in purchase amounts, showing the difference between the highest and lowest values.



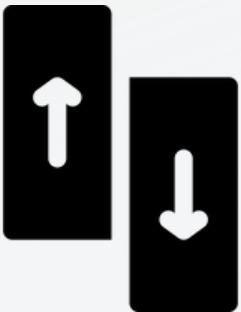
IQR : 10.1925

The IQR (10.1925), which measures the range of the middle 50% of the data, helps focus on typical purchase amounts without being influenced by outliers.



VARIANCE AND STANDARD DEVIATION

Variance : 822.85

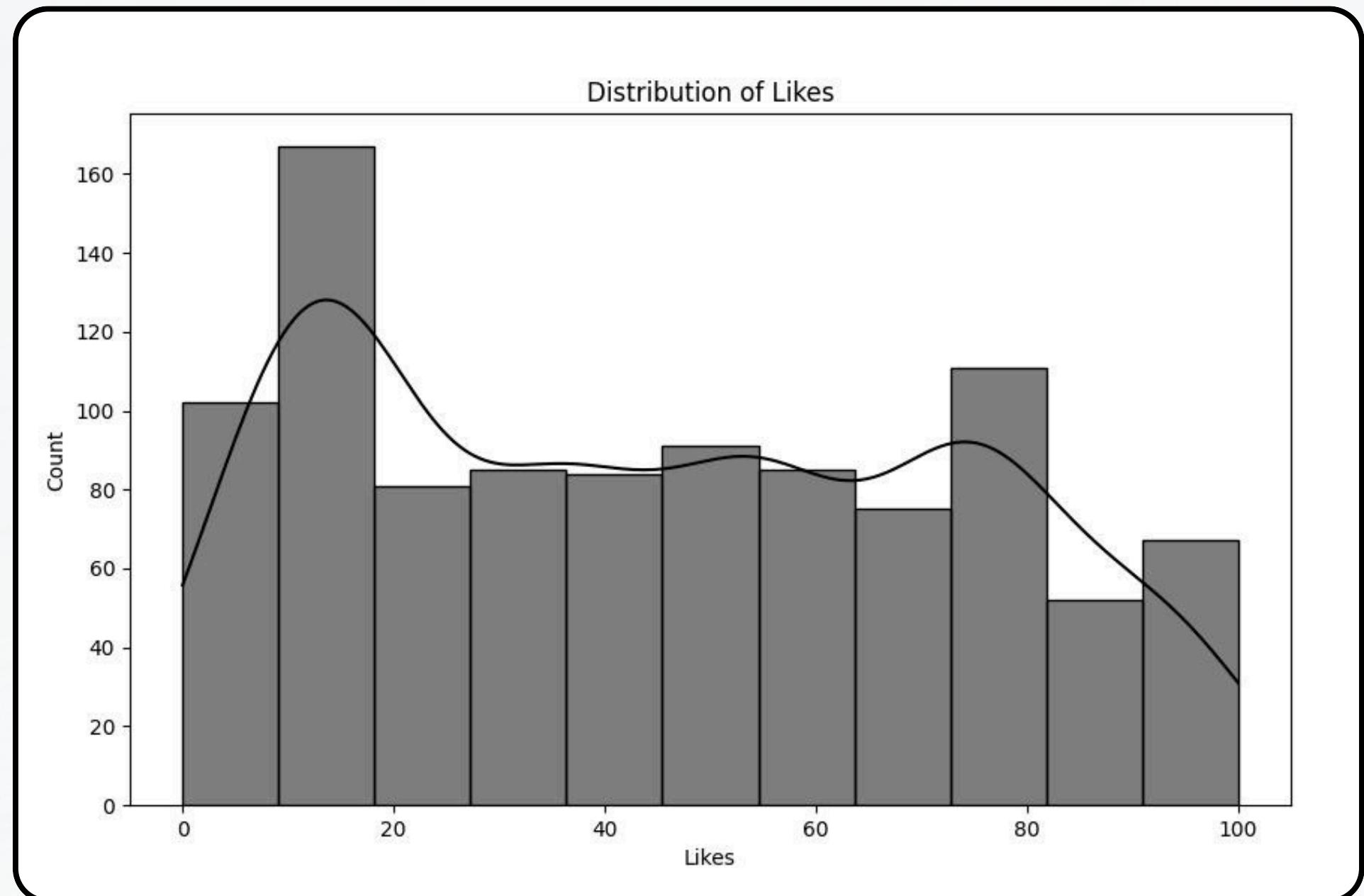


It indicates high degree of variability, it suggests that the point are widespread form the mean implying significant differences among the observation.

Standard Deviation : 28.69



It indicates average deviation of each data point from the mean. A higher standard deviation suggests greater variability implying data point are more spread out.



HYPOTHESIS TEST

Hypothesis method **compares two opposite statements** about a population and uses sample data to decide which one is more likely to be correct. To test this assumption we first take a sample from the population and analyze it and use the results of the analysis to decide if the claim is valid or not.

Hypothesis Testing Framework:

- **Null Hypothesis (H_0) :** The average rating for Instagram is equal to or less than the average rating for WhatsApp.

$$H_0: \mu_{\text{Instagram}} \leq \mu_{\text{WhatsApp}}$$

If the P-value < 0.05

you reject the null hypothesis, which would mean Instagram's average rating is significantly higher than WhatsApp's, according to the data.

If the P-value > 0.05

you fail to reject the null hypothesis, indicating that there is no significant difference in ratings between Instagram and WhatsApp.

- **Alternative Hypothesis (H_1) :** The average rating for Instagram is greater than the average rating for WhatsApp.

$$H_1: \mu_{\text{Instagram}} > \mu_{\text{WhatsApp}}$$

T-statistic : -0.7967

P-value : 0.7868

- Since the **p-value (0.7868)** is much larger than the significance level of **0.05**, we **fail to reject the null hypothesis**.
- This means that there is **no significant difference** between the ratings for the two platforms based on the given data.

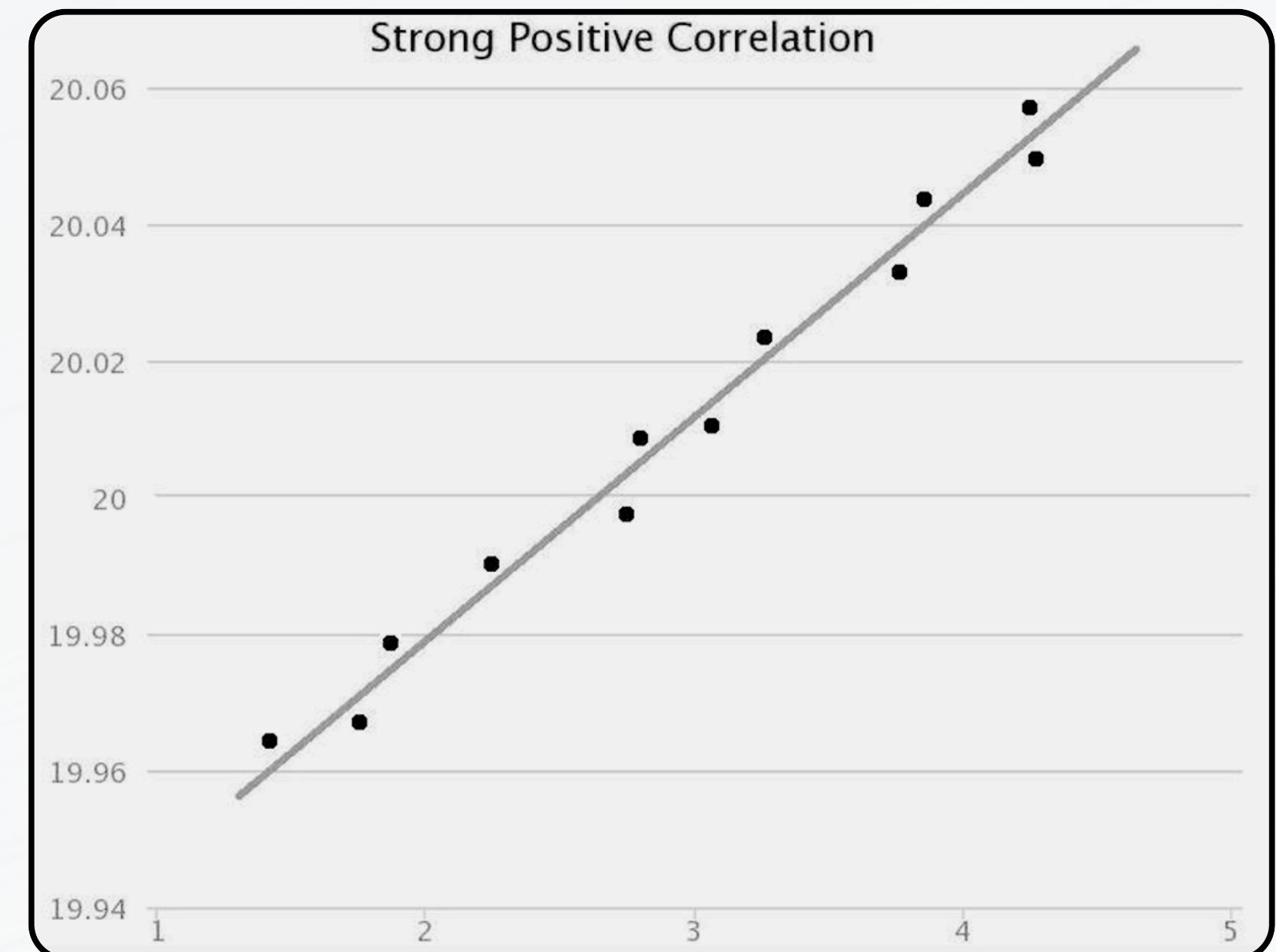
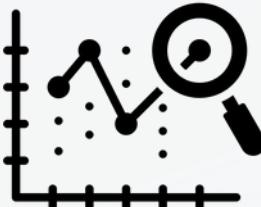
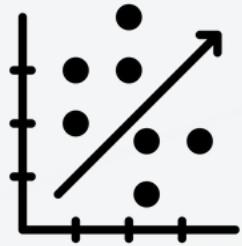
CORRELATION

Correlation is a statistical measure that describes the extent to which two or more variables are related to one another.

It indicates how changes in one variable are associated with changes in another variable.

Correlation : 0.84254

- A value above **0.8** generally represents a **strong correlation**, meaning there is a clear and predictable trend between the two variables.
- This positive correlation could mean that products with higher ratings receive more likes, perhaps because users who are satisfied with the product are more likely to engage further by liking it.
- It also suggests that **higher-rated products are likely perceived more favorably**, which can lead to greater interaction, like and share behavior on the platform.



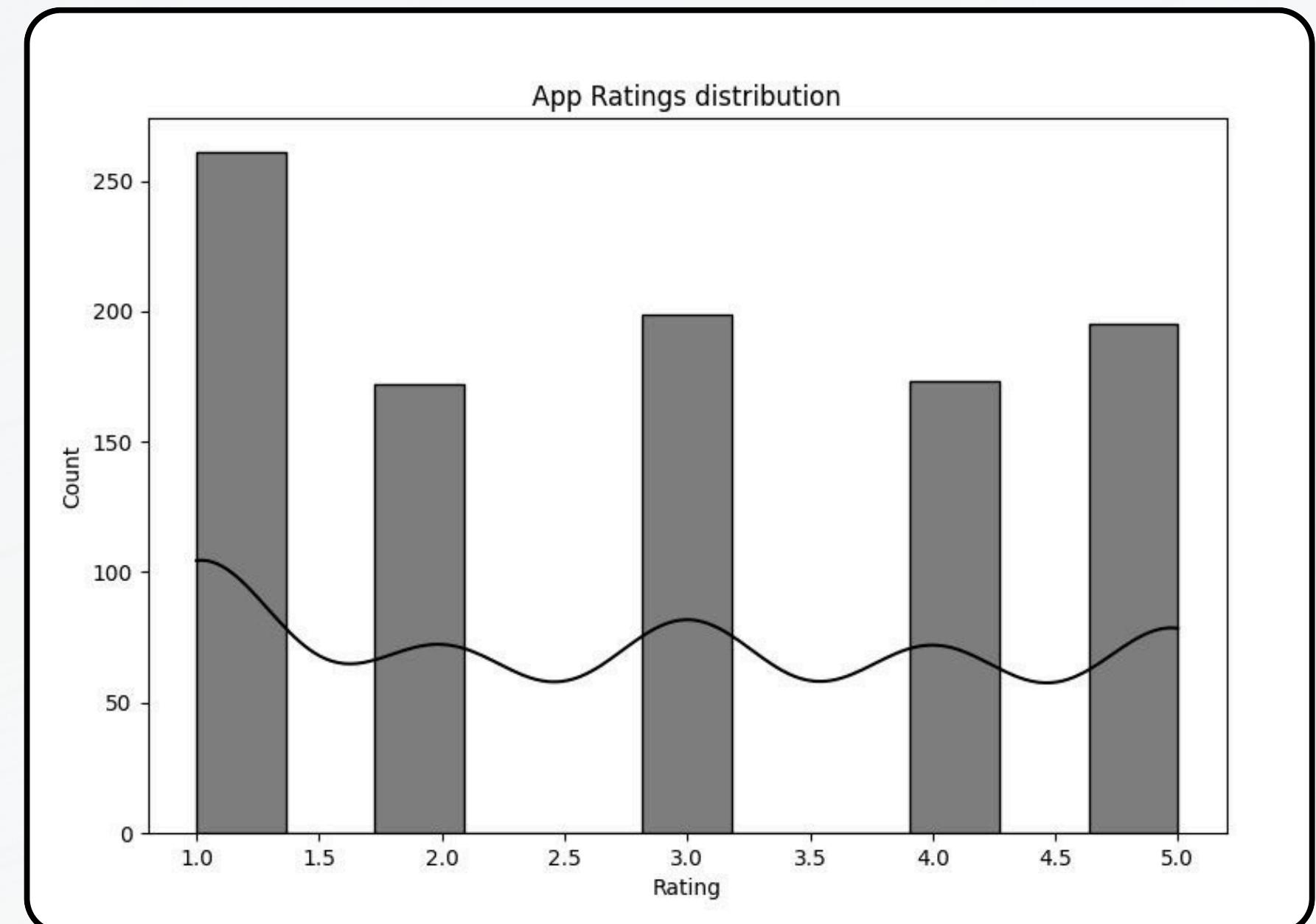
APP RATINGS

App rating helps both other users and app developers. For users, it improves the decision-making process of trying an app.

For developers, it provides insights that can be crucial for the app's success.

If the distribution is positively skewed, this indicates that a majority of users rated the app lower, with a few high ratings pulling the average up. Conversely, a negatively skewed distribution would suggest most users rated the app higher.

- The distribution appears to be negatively skewed as the bars are higher towards the lower rating and comparatively lower towards the higher rating.
- Left-skewed distribution indicates a **high number of dissatisfied users**.
- Many users are giving low rating suggesting significant **issue with the app**.
- Fewer user are giving high ratings (**5 star**) indicating **positive experience** amongst but **for few users only**.



CENTRAL LIMIT THEOREM

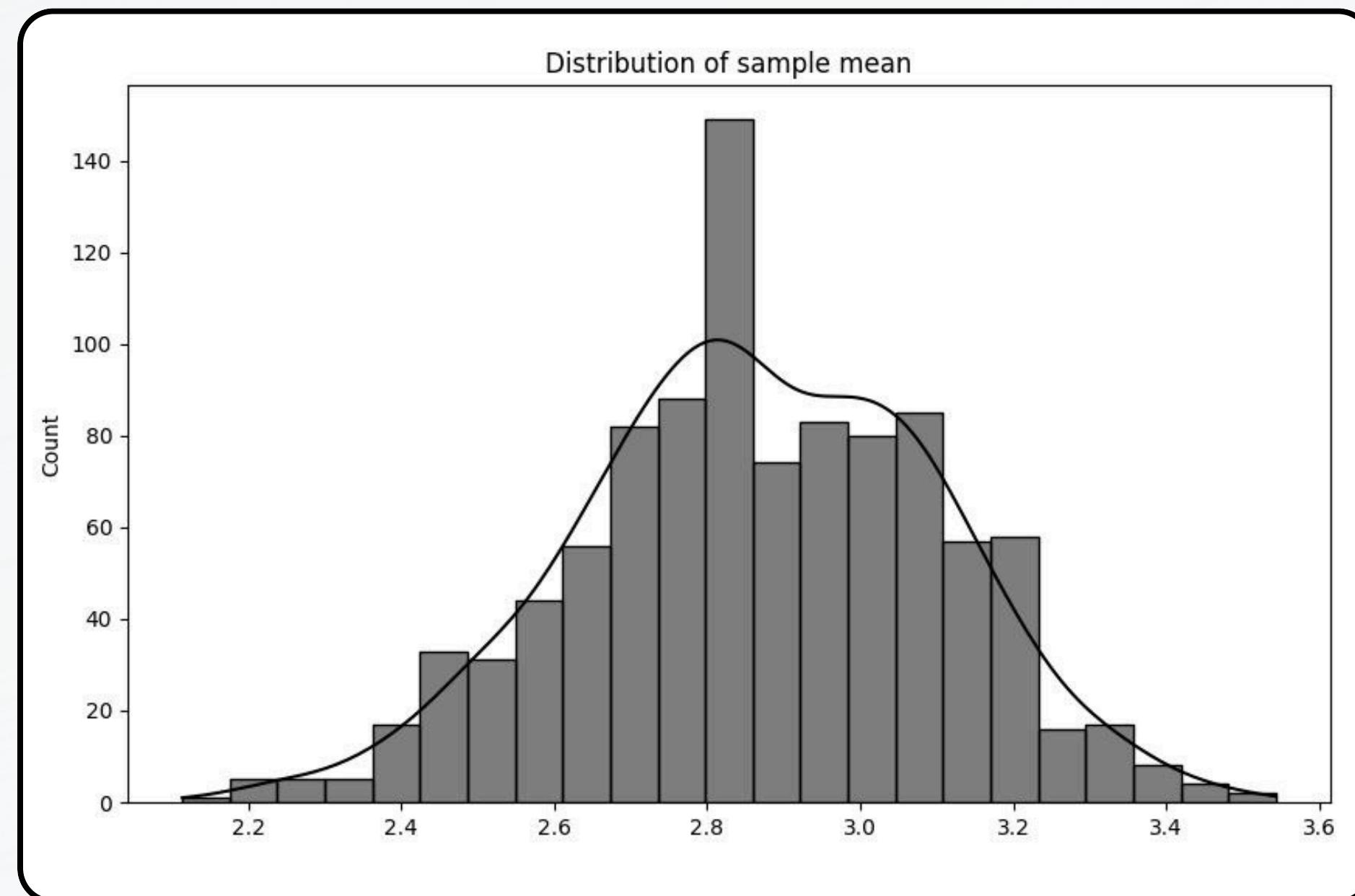
The Central Limit Theorem in Statistics states that as the sample size increases and its variance is finite, then the distribution of the sample mean approaches normal distribution irrespective of the shape of the population distribution.

Sampling Distribution: The histogram of sample means demonstrates an approximately normal distribution.

Analysis:

- The Central Limit Theorem (CLT) states that the sampling distribution of the sample mean will be normally distributed if the sample size is sufficiently large ($n \geq 30$).
- In this case, the histogram shows a clear normal shape, validating the application of the CLT.

This allows us to use normal distribution for inference about the population mean, facilitating reliable hypothesis testing and confidence interval construction.



CONCLUSION

- The median rating of 3 suggests **mixed user satisfaction**, with many low ratings (mode of 1).
- The range and IQR of purchase amounts show **moderate variability in spending behavior**.
- Significant variance in likes highlights **diverse user engagement** with app reviews.
- A strong positive correlation (**0.8425**) between likes and ratings suggests that **higher-rated reviews attract more likes**.
- Hypothesis testing found **no significant difference between Instagram and WhatsApp ratings**.
- The Central Limit Theorem supported reliable inferences, helping guide **improvements in app features and user satisfaction**.



Connect with me

I welcome and appreciate your thoughts about this project and always open to every suggestion, let's connect here...

[Gmail](#)

[LinkedIn](#)

[Github](#)