Apple Store Reviews

Statistical Analysis

PRESENTED BY :ANCHAL DAYAL

Problem Statment

This project aims to analyze Apple Store product reviews to gain insights into user ratings and engagement. By quantifying the central tendency of app ratings, assessing the spread of purchase amounts, evaluating correlations between likes and ratings, and comparing user perceptions of Instagram and WhatsApp through hypothesis testing, the analysis seeks to identify trends and areas for improvement. Ultimately, the goal is to provide actionable recommendations for app developers to enhance user satisfaction and engagement.

Rating Analysis

1. **Mean: 2.869**
   The mean is slightly below 3, suggesting that extreme low ratings (like 1) are pulling the average down.

2. **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.

3. **Mode: 1**
   The mode being the most frequent rating indicates a significant number of users rated the app poorly.

Conclusion: The median best represents the central tendency of the ratings, as it is not skewed by extreme values and reflects the "middle" user experience better than the mean.

Variance and Standard Deviation of Likes

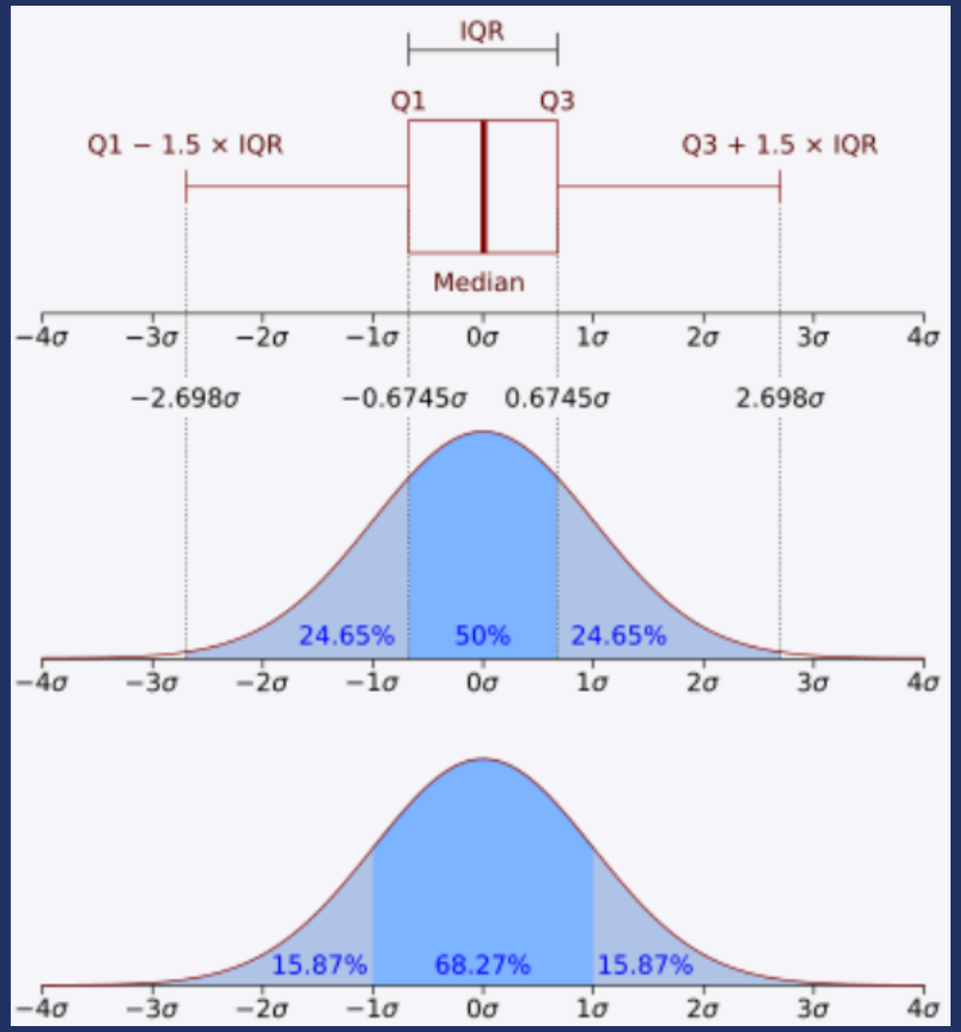
Variance
822.8547

Variance indicates the average squared deviation from the mean, reflecting the degree to which likes vary from the average.

Standard Deviation
28.6854

The standard deviation (28.6854) represents the spread in the same units as the original data, highlighting significant variability in user engagement with reviews.

Conclusion: A high standard deviation suggests diverse reactions among users, indicating that while some reviews received few likes, others received many.



Spread of Purchase Amount

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.

Conclusion: The range offers an overview of data variability, while the IQR gives insight into the spread of typical purchase amounts, aiding in identifying anomalies.

Correlation between Likes and Rating

1

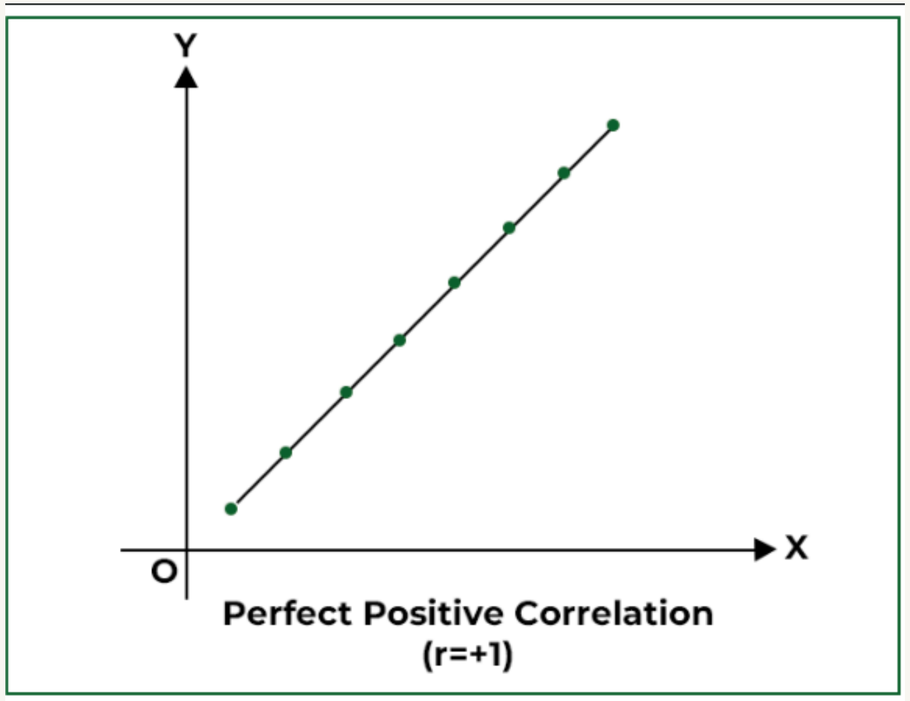
Correlation Coefficient
0.8425

2

Analysis
A correlation coefficient of 0.8425 indicates a strong positive correlation between likes and ratings, suggesting that higher likes tend to accompany higher ratings.

3

Conclusion
This strong correlation implies that user satisfaction and engagement (likes) are closely related.

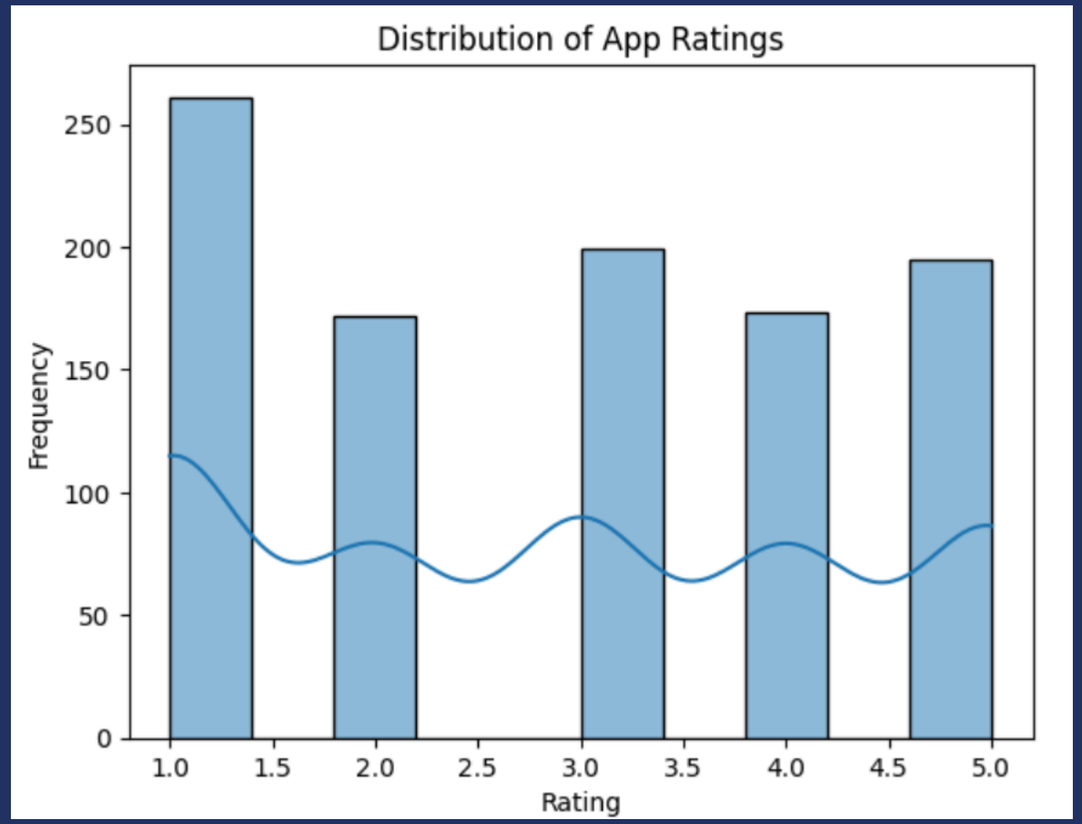


Distribution of App Ratings

1. **Skewness**
(assumed based on visual analysis)

2. **Analysis**
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.

3. **Conclusion**
A positively skewed distribution may suggest overall dissatisfaction, while a negatively skewed distribution would imply user satisfaction.



Hypothesis Test for Instagram vs. WhatsApp Ratings

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: μInstagram <= μ Whatsapp

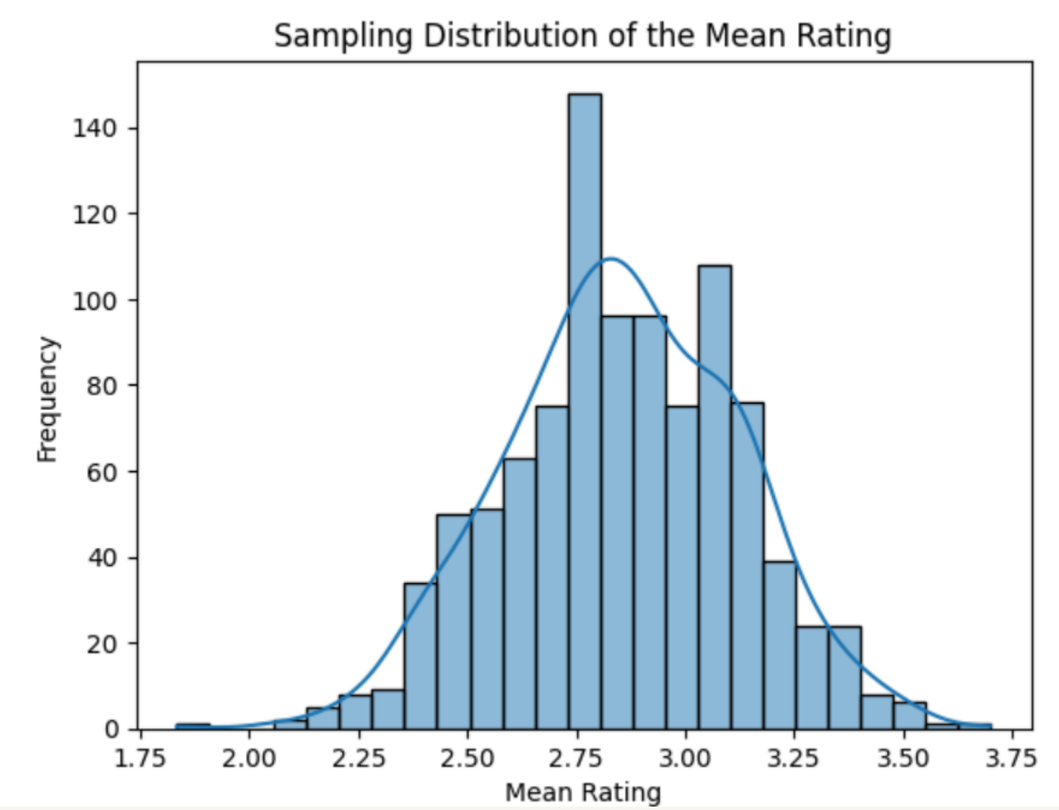
Alternative Hypothesis (H_1):
The average rating for Instagram is greater than the average rating for WhatsApp.
H_1: μInstagram > μ Whatsapp

T-statistic: -0.7967 **P-value: 0.7868**

Interpretation:
A high p-value indicates we fail to reject the null hypothesis.

Conclusion:
There is no significant difference between average ratings for Instagram and WhatsApp, suggesting user perceptions are similar.

Sampling Distribution and Central Limit Theorem



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

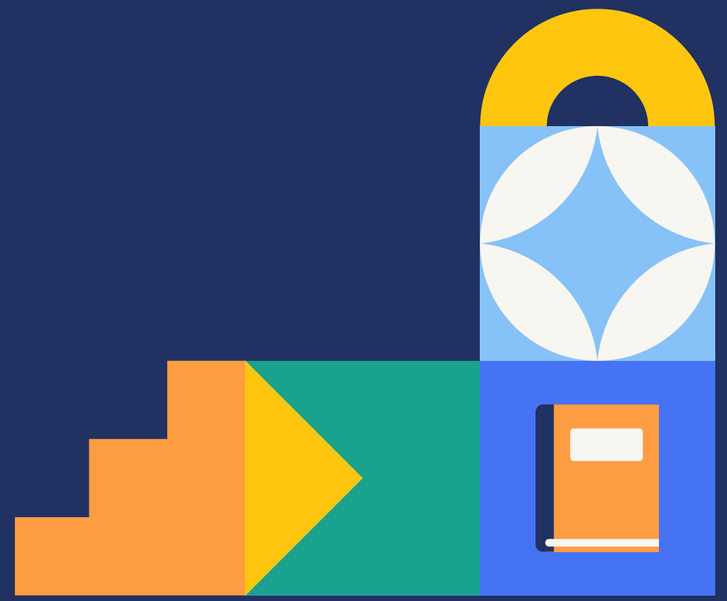
Analysis:

- **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 ≥ 30).**
- **In this case, the histogram shows a clear normal shape, validating the application of the CLT.**

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

Conclusion: Analysis of Apple Store Reviews

- 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.



Get In Touch

LinkedIn | Github | Email