APPLE STORE REVIEW

Descriptive and inferential statistical analysis

A presentation by Jagruti Jadhav

GET INSPIRED

Without data, you’re just another person with an opinion

W. Edwards Deming

Objective of the Analysis

- Understand the central tendency and spread of app ratings.
- Analyze user engagement through likes and reviews.
- Determine relationships between variables (e.g., likes and ratings).
- Conduct hypothesis testing to compare app ratings (Instagram vs WhatsApp).
- Explore sampling distribution and its relation to the Central Limit Theorem.

KEY METRICS CALCULATED

Mean Rating	2.869	Average user rating
Median Rating	3	Middle rating score
Mode Rating	1	Most frequent rating (lowest score)
Variance (Likes)	822.85	High spread in user engagement
Standard Deviation	28.69	High variability in likes
Correlation (Likes vs Ratings)	0.842	Strong positive correlation

Central Tendency of App Ratings

- The mean rating (2.869) is slightly below the median (3.0), indicating a slight skew towards lower ratings.
- The mode (1) shows that the most common rating is the lowest possible score, suggesting a notable group of dissatisfied users.
- The median seems to be the best representation of central tendency because it is less influenced by extreme low ratings

2.869

MEAN

3

MEDIAN

1

MODE

Range and Interquartile Range (IQR) of Purchase Amount

19.97

The range shows the difference between the highest and lowest values of Purchase_Amount. A range of 19.97 indicates a wide spread in the data, meaning there are customers making both very low and very high purchases

10.19

The IQR measures the spread of the middle 50% of the data. An IQR of 10.19 suggests that most purchase amounts are closely clustered around the middle, even though there are some extreme values (outliers) causing a larger overall range.

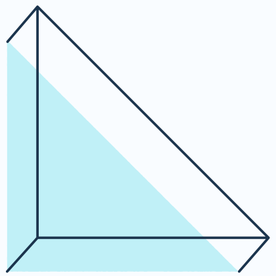
Variance and Standard Deviation for Likes

Variance

The variance measures how much the number of likes on reviews varies from the average. A variance of 822.85 indicates considerable fluctuation in likes across different reviews.

Standard Deviation

The standard deviation provides a more interpretable measure of spread. A standard deviation of 28.69 suggests that the number of likes on reviews deviates significantly from the mean, indicating that some reviews receive far more likes than others, while others receive fewer likes.



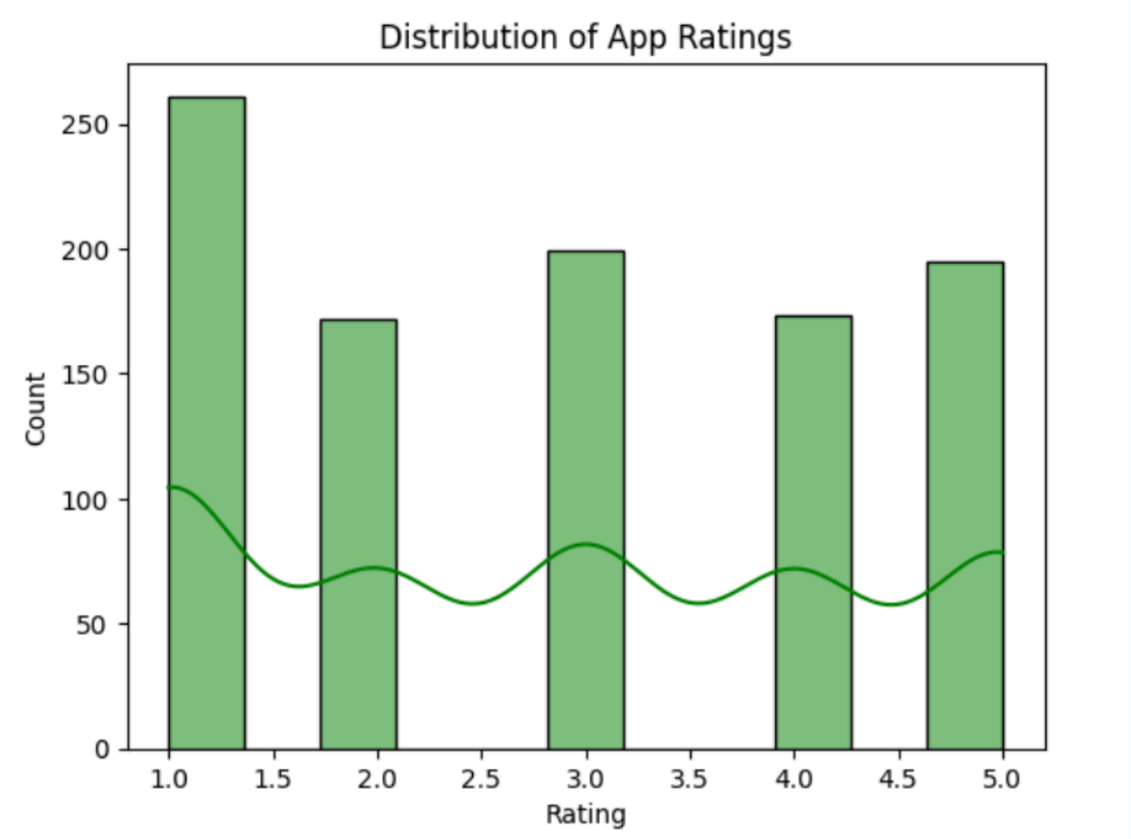
Correlation Between Likes and Ratings

0.842

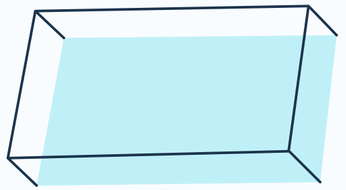
- The strong positive correlation (0.842) indicates that higher ratings tend to receive more likes on reviews.
- Users are more likely to engage with positive reviews, which can be a valuable insight for app developers aiming to increase user engagement.

Plot the Distribution of App Ratings

- The distribution is positively skewed, with more users giving lower ratings.
- Most users rated the apps at 1.0, indicating dissatisfaction.
- Only a smaller portion of users gave higher ratings.



Hypothesis Test: Instagram vs WhatsApp Average Ratings

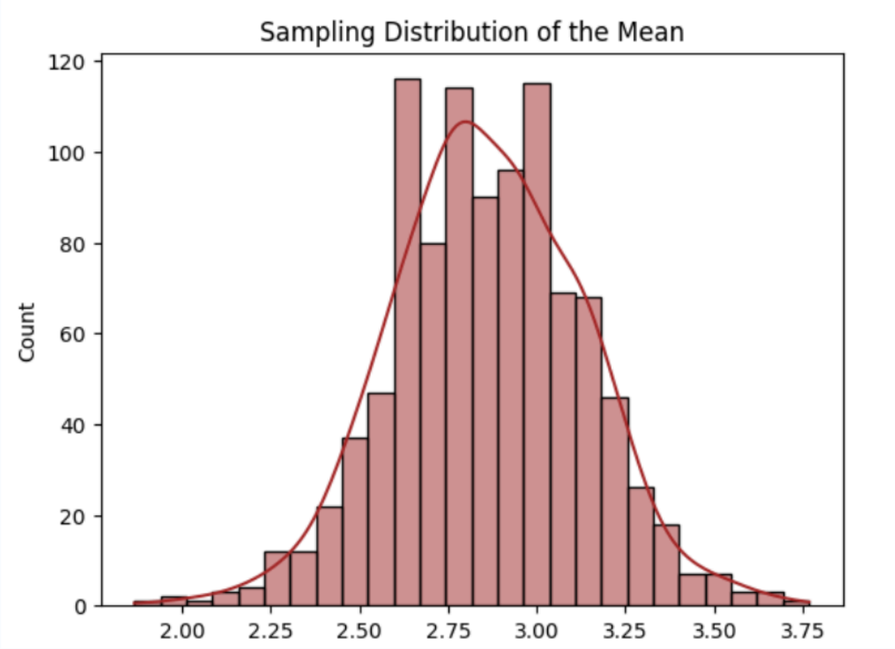


Null Hypothesis (H_0): The average rating for Instagram is not significantly higher than for WhatsApp.

Alternative Hypothesis (H_1): The average rating for Instagram is significantly higher than for WhatsApp.

- **Since the p-value (0.426) is greater than 0.05, we fail to reject the null hypothesis.**
- **This means there is no significant difference between the average rating of Instagram and WhatsApp at the 95% confidence level**

Sampling Distribution and Central Limit Theorem



The Central Limit Theorem explains why the sampling distribution of sample means tends to be normal, regardless of the population distribution. This property is crucial for inferential statistics because it allows us to use the normal distribution to make predictions and calculate probabilities about sample means.

What we Learn ?

KEY SUMMARY

- User ratings are skewed towards lower scores, indicating potential dissatisfaction with the apps.
- The number of likes on reviews varies significantly, with a strong positive correlation between likes and ratings.
- There is no significant difference between Instagram and WhatsApp ratings based on the hypothesis test.
- The Central Limit Theorem holds, as the sampling distribution of the mean forms a normal curve even with skewed data.

THANK YOU

Github LinkedIN