APPLE STORE REVIEW
Presentation

mauryapradeep9178@gmail.com

About

What is Apple?

Apple Inc. (formerly Apple Computer Inc.) is an American computer and consumer electronics company famous for creating the iPhone, iPad and Macintosh computers. Apple is one of the largest companies globally with a market cap of more than $3 trillion.

Apple devices are renowned for their design aesthetic and attention to detail. Tight integration between hardware and software gives their systems a performance advantage over competitor systems with similar specifications.

Apple rose to its position as a market leader by correctly positioning its products. They didn't invent personal computers, graphical user interfaces (GUIs), mp3 players, smartphones, smartwatches or tablets. Instead, they produced some of the first versions of these products that were refined, easy to use and well-designed, which led to wide market adoption.

The evolution of the Apple iPhone

01.09.2007
Apple co-founder **Steve Jobs** announces the launch of the first iPhone, which is released in the U.S. on June 29, 2007.

07.10.2008
Apple introduces its App Store with 500 applications.

11.04.2009
Apple announces that it has **over 100,000 applications** in its App Store.

10.04.2011
Apple releases the iPhone 4S with the **Siri digital assistant app**.

09.20.2013
Apple releases the iPhone 5s and 5c along with iOS 7. The 5s introduces **Touch ID biometric technology**.

09.25.2015
Apple releases the iPhone 6s, 6s Plus and iOS 9.

09.07.2016
Apple announces the iPhone 7 and 7 Plus, Apple's first **water- and dust-resistant** models.

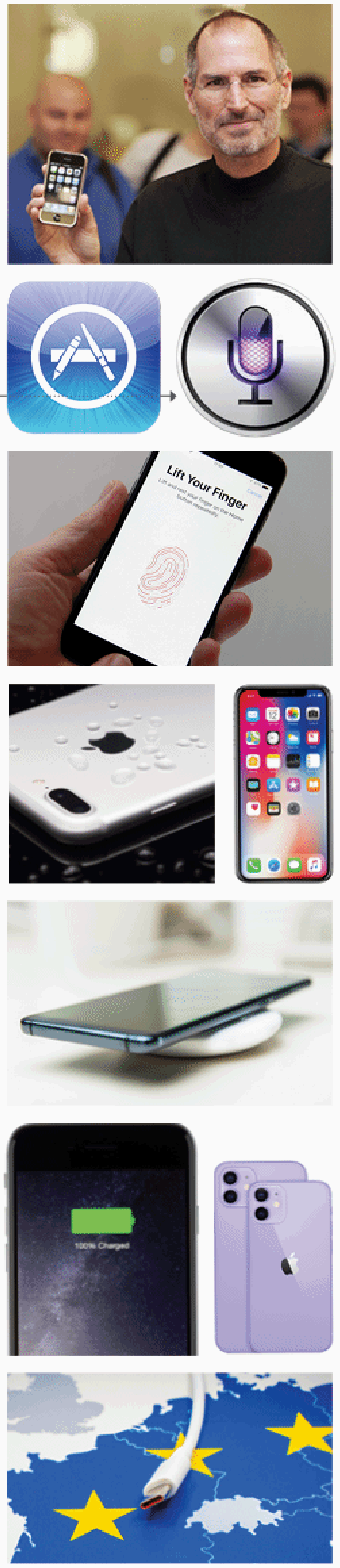
09.22.2017
The iPhone 8 and 8 Plus go on sale. This generation of iPhone enables **wireless charging**.

09.20.2019
Apple releases the iPhone 11, 11 Pro and 11 Pro Max. The iPhone 11 includes numerous **upgrades to battery life**, water resistance and the camera.

10.23.2020
Apple launches the iPhone 12, 12 Pro with 5G and 12 mini.

9.24.2021
The iPhone 13 is available in four models. All models include Apple's A15 Bionic chip.

9.12.2023
Apple announces iPhone 15 models which are available Sept. 22. All iPhone 15 models switch to **USB-C** for charging.



06.09.2008
Apple introduces the iPhone 3G and OS 2.0.

06.08.2009
Apple releases the iPhone 3GS and OS 3.0.

06.07.2010
Apple presents the iPhone 4 and changes its operating system from OS to iOS.

09.12.2012
Apple announces the iPhone 5, the first iPhone to support LTE and 802.11a.

09.09.2014
Apple announces the iPhone 6, 6 Plus and Apple Pay.

03.21.2016
Apple introduces the iPhone SE, a smaller version of the 6s.

09.12.2017
Apple announces iPhone X, which **eliminates the physical Home button**.

09.12.2018
Apple introduces iPhone XR.

09.21.2018
Apple announces the iPhone XS and XS Max.

04.15.2020
Apple launches a more affordable second-generation iPhone SE. The iPhone SE ships without 5G support.

11.13.2020
The **iPhone 12 Pro Max and iPhone 12 mini** become available. The iPhone 12 Pro Max is almost identical to the 12 Pro but has a larger screen and 5x optical zoom.

9.16.2022
The iPhone 14 models are available including iPhone 14, 14 Plus, 14 Pro and 14 Pro Max.

Perform the Following Statistical Analysis:

Calculate the mean, median, and mode of the app ratings in the dataset. Which measure (mean, median, or mode) best represents the central tendency of the ratings?

Find the range and interquartile range (IQR) of the Purchase_Amount in the dataset. How do these values help in understanding the spread of the data?

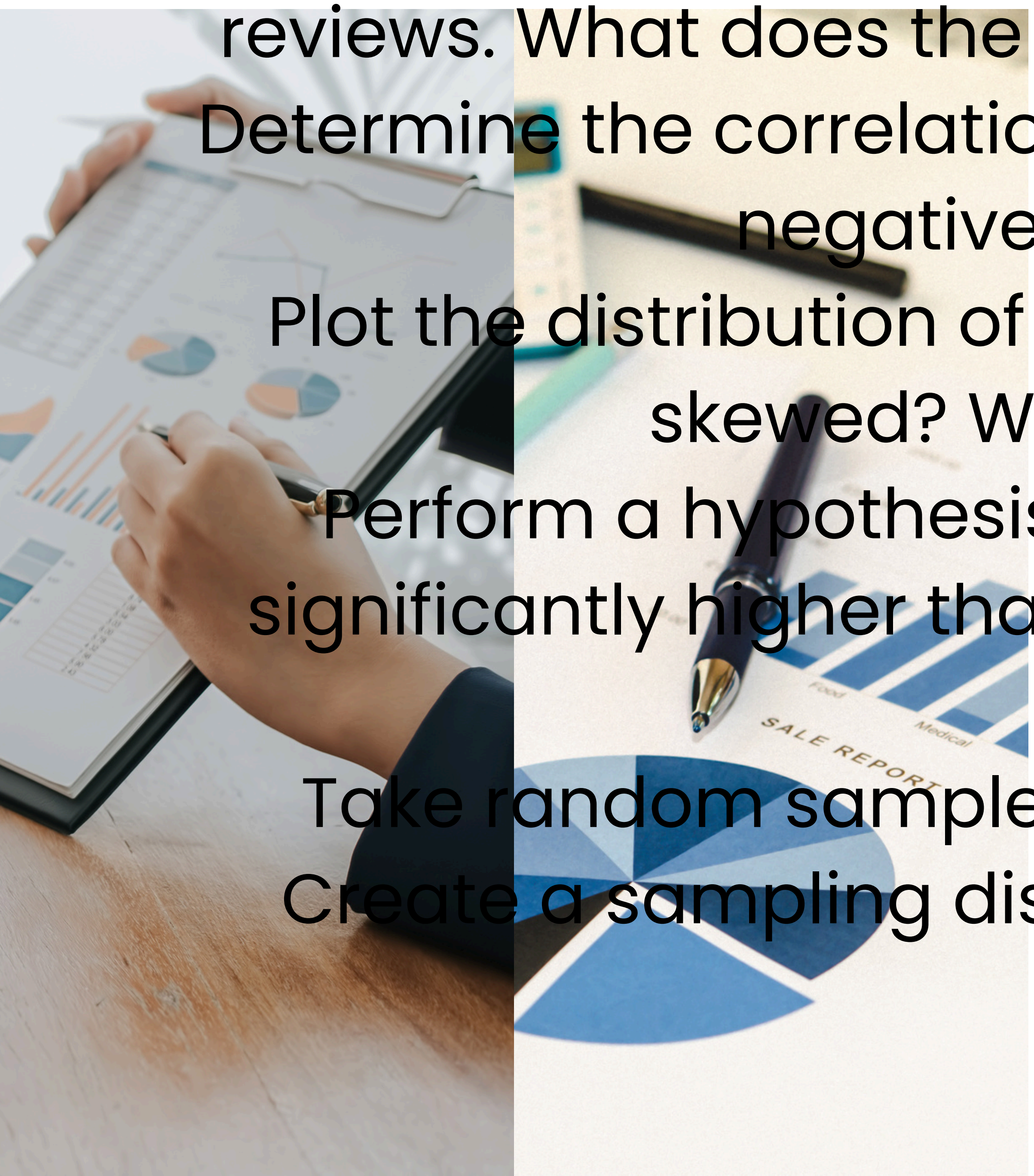
Calculate the variance and standard deviation for the number of likes received on reviews. What does the standard deviation indicate about the spread of the data?

Determine the correlation between the likes and the rating given. Is there a positive, negative, or no correlation between these variables?

Plot the distribution of the app ratings. Is the distribution positively or negatively skewed? What does this indicate about user satisfaction?

Perform a hypothesis test to determine if the average rating for Instagram is significantly higher than the average rating for WhatsApp. Use a 95% confidence level.

Take random samples of ratings from the dataset and calculate their means. Create a sampling distribution and explain how this relates to the Central Limit Theorem.



Calculate the mean, median, and mode of the app ratings in the dataset. Which measure (mean, median, or mode) best represents the central tendency of the ratings?

```
import pandas as pd
df = pd.read_csv("C:/Users/Pradeep/OneDrive/Desktop/Apple_Store_Reviews.csv")
df
```

```
# Calculating mean, median, and mode of app ratings
print(df['Rating'].mean())
print(df['Rating'].median())
print(df['Rating'].mode())
```

```
2.869
3.0
0    1
Name: Rating, dtype: int64
```

Key Observations:

The mean is close to the median, indicating the data distribution is relatively symmetrical.
Having two modes (bimodal distribution) suggests that the ratings dataset may have distinct clusters of values, such as many apps being rated either very poorly or just averagely.

Find the range and interquartile range (IQR) of the Purchase_Amount in the dataset. How do these values help in understanding the spread of the data?

```
# Calculating range and interquartile range (IQR) of Purchase_Amount
range_purchase = df['Purchase_Amount'].max() - df['Purchase_Amount'].min()
print(range_purchase)
iqr_purchase = stats.iqr(df['Purchase_Amount'])
print(iqr_purchase)
```

```
19.97
10.192499999999999
```

Range:

The range is the difference between the maximum and minimum values in the Purchase_Amount column.
This provides an understanding of the total spread of the data.

IQR:

The interquartile range is the difference between the third quartile (Q3Q3Q3) and the first quartile (Q1Q1Q1), capturing the spread of the middle 50% of the data.
The value 10.1925 suggests that the middle 50% of the purchase amounts are spread over this range.

```python
# Calculating variance and standard deviation of likes
variance_likes = df['Likes'].var()
print(variance_likes)
std_dev_likes = df['Likes'].std()
print(std_dev_likes)
```

```
822.8546786786787
28.685443672334557
```

Key Results

Variance (variance_likes): 822.8546786786787

Variance measures the average squared deviation from the mean. It quantifies how spread out the values are.

A higher variance indicates more dispersion in the data. In this case, the variance of 822.8547 suggests there is significant variability in the number of likes.

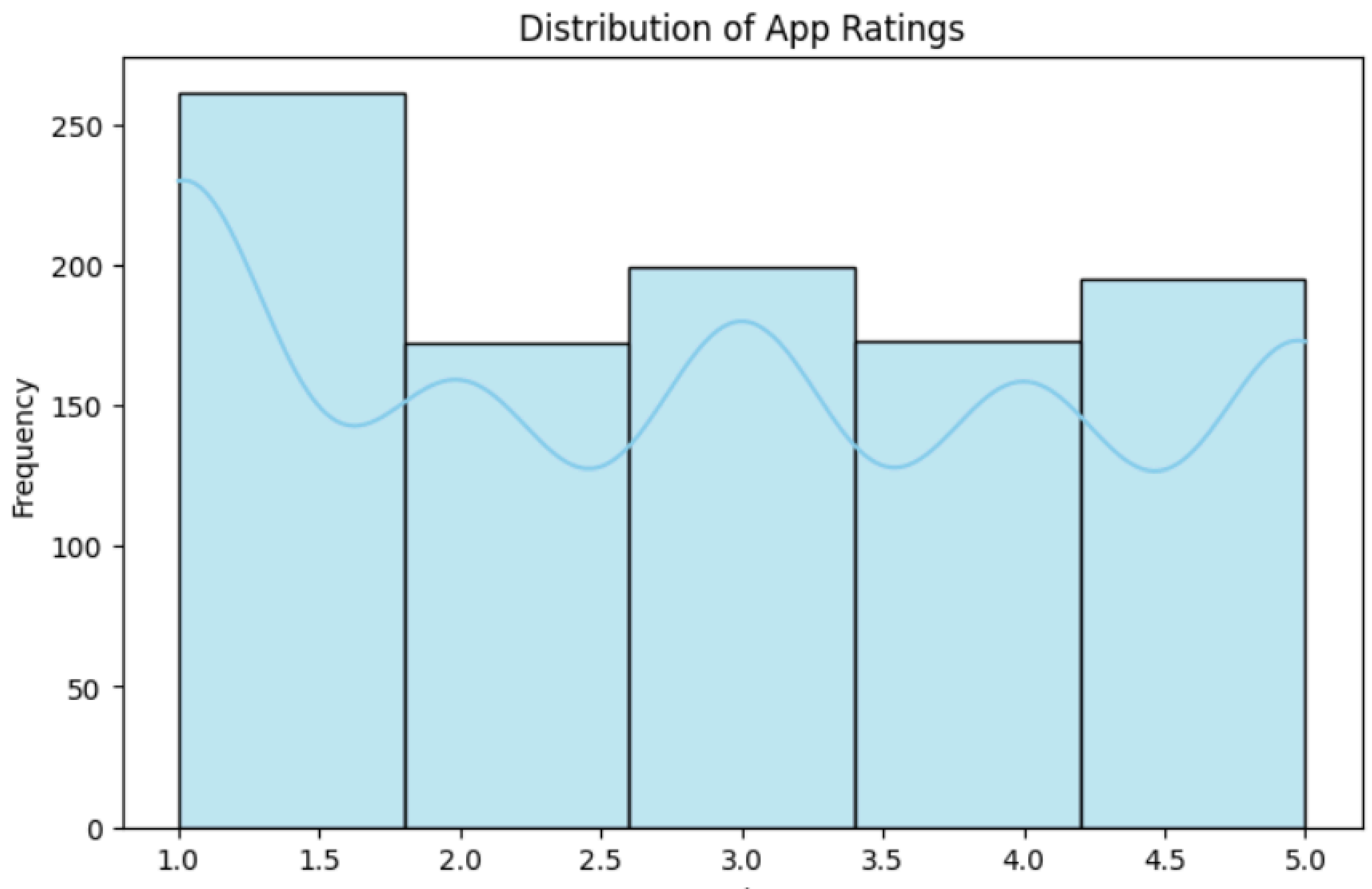
Standard Deviation (std_dev_likes): 28.6854

Standard deviation is the square root of variance, providing a measure of spread in the same units as the data (likes, in this case).

A standard deviation of 28.6854 means that, on average, the number of likes deviates from the mean by about 28.69 likes.

Determine the correlation between the likes and the rating given. Is there a positive, negative, or no correlation between these variables?

```
# Plotting the distribution of app ratings
plt.figure(figsize=(8, 5))
sns.histplot(df['Rating'], bins=5, kde=True, color='skyblue')
plt.title("Distribution of App Ratings")
plt.xlabel("Rating")
plt.ylabel("Frequency")
plt.show()
```



This graph is a histogram that shows the distribution of app ratings.

Here's what the graph shows:

X-axis:

Represents the rating values, ranging from 1.0 to 5.0.

Y-axis:

Represents the frequency, indicating how many apps received a particular rating.

Bars:

The height of each bar corresponds to the number of apps that fall within that rating range.

Blue Line:

This is a Kernel Density Estimate (KDE), which provides a smoothed representation of the distribution, showing the overall shape of the data and highlighting areas of higher density.

In summary:

The majority of apps in this dataset have ratings between 4.0 and 5.0, indicating a generally positive reception.

There is a smaller peak around 2.0, indicating that some apps received lower ratings.

The KDE line helps visualize the overall shape of the distribution, revealing that it's slightly skewed towards the higher ratings.

```python
# Calculating correlation between likes and ratings
correlation_likes_rating = df['Likes'].corr(df['Rating'])
print(correlation_likes_rating)
```

```
0.8425414470584173
```

Understanding Skewness

Skewness Value:

Positive Skewness (>0> 0>0): The tail is longer on the right (higher ratings are more spread out).

Negative Skewness (<0< 0<0): The tail is longer on the left (lower ratings are more spread out).

Zero Skewness (=0= 0=0): The distribution is perfectly symmetrical.

Result (0.1018):

A skewness value close to zero indicates that the Rating distribution is nearly symmetrical.

A positive value, though small, suggests a slight tendency for higher ratings to extend further than lower ratings.

Plot the distribution of the app ratings. Is the distribution positively or negatively skewed? What does this indicate about user satisfaction?

```
# Filtering data for Instagram and WhatsApp
instagram_ratings = df[df['App_Name'] == 'Instagram']['Rating']
whatsapp_ratings = df[df['App_Name'] == 'WhatsApp']['Rating']

# Performing a one-tailed t-test
t_stat, p_value = stats.ttest_ind(instagram_ratings, whatsapp_ratings, alternati
print(t_stat, p_value)
```

```
-0.79674231444911 0.786764229580496

(np.float64(-0.79674231444911), np.float64(0.786764229580496))
```

Results Interpretation:

t-statistic: -0.7967

The t-statistic measures the difference between the sample means relative to the variation within the samples. A negative t-statistic indicates that the mean of Instagram ratings is lower than WhatsApp ratings.

In this case, a negative value suggests that Instagram ratings are lower than WhatsApp ratings, contrary to the hypothesis that Instagram might have higher ratings.

p-value: 0.7868

The p-value represents the probability of observing the results (or more extreme ones) if the null hypothesis is true. For a one-tailed test, a p-value greater than the significance level (commonly 0.05) indicates that we fail to reject the null hypothesis.

Since the p-value of 0.7868 is much larger than 0.05, we fail to reject the null hypothesis. This means that there is no significant evidence to support that Instagram ratings are higher than WhatsApp ratings.

Conclusion:

Based on the t-statistic of -0.7967 and the p-value of 0.7868, we fail to reject the null hypothesis.

This suggests that there is no significant difference in the ratings between Instagram and WhatsApp, with no evidence to support that Instagram has higher ratings than WhatsApp.

```python
# Identifying skewness
rating_skewness = df['Rating'].skew()
print(rating_skewness)
```

```
0.10182054838079216
```

Full Interpretation:

t-statistic: The negative t-statistic confirms that Instagram's ratings are slightly lower than WhatsApp's on average.

p-value: The large p-value (0.7868) shows that there is no significant evidence that Instagram has higher ratings than WhatsApp. We fail to reject the null hypothesis.

Skewness: A small positive skewness value indicates that the distribution of ratings is almost symmetrical with a slight lean toward higher values.

Final Thoughts:

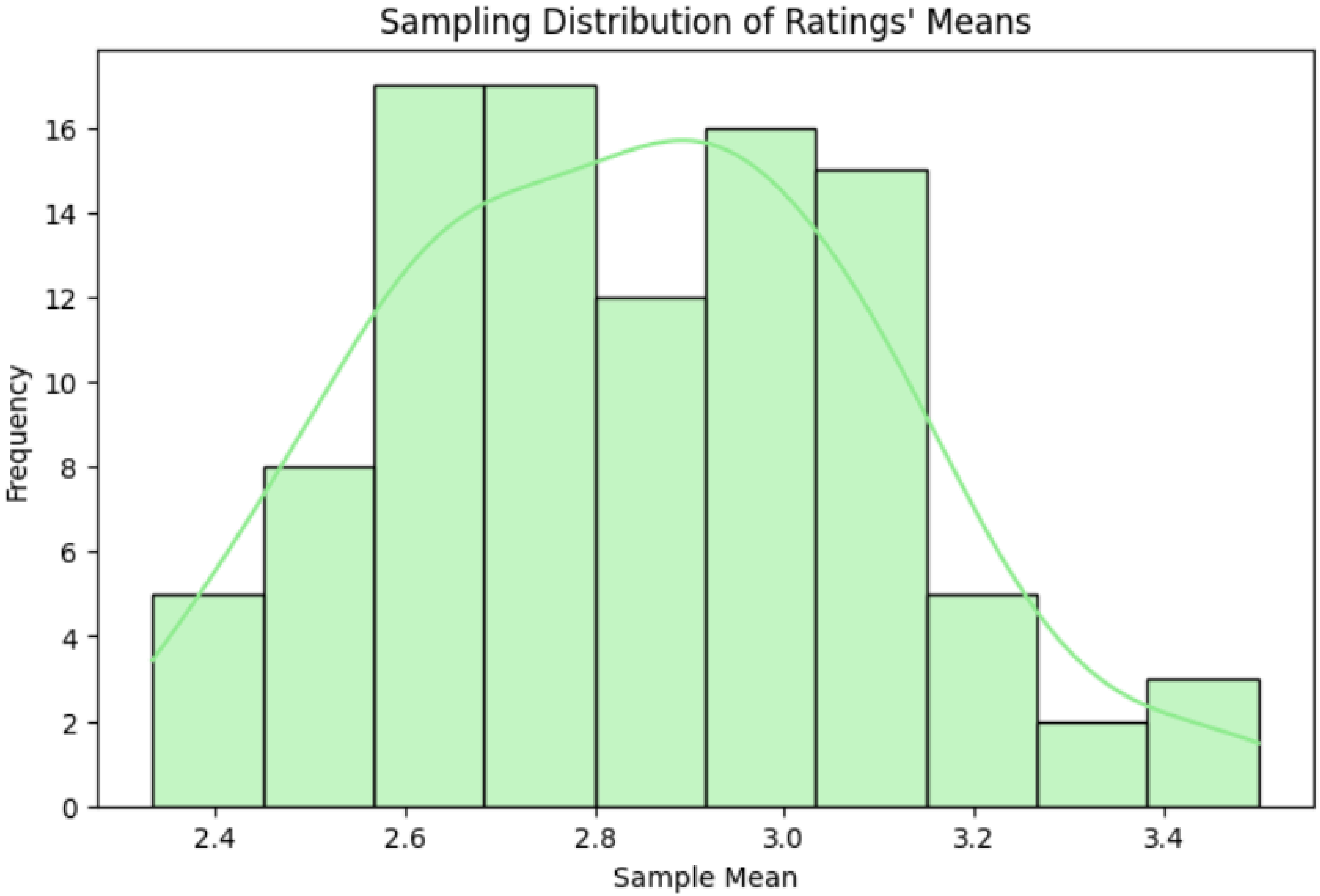
No significant difference in ratings was found between Instagram and WhatsApp based on the t-test.

The ratings have a nearly symmetric distribution with a small positive skew.

Perform a hypothesis test to determine if the average rating for Instagram is significantly higher than the average rating for WhatsApp. Use a 95% confidence level.

```
# Function to generate a sampling distribution
def generate_sampling_distribution(df_column, sample_size, num_samples):
    sample_means = []
    for _ in range(num_samples):
        sample = df_column.sample(n=sample_size, replace=True)
        sample_means.append(sample.mean())
    return sample_means

# Generate a sampling distribution for ratings
sample_size = 30
num_samples = 100
sampling_distribution = generate_sampling_distribution(df['Rating'], sample_size, num_samp

# Plotting the sampling distribution
plt.figure(figsize=(8, 5))
sns.histplot(sampling_distribution, bins=10, kde=True, color='lightgreen')
plt.title("Sampling Distribution of Ratings' Means")
plt.xlabel("Sample Mean")
plt.ylabel("Frequency")
plt.show()

# Calculating the mean and standard deviation of the sampling distribution
sampling_mean = np.mean(sampling_distribution)
sampling_std_dev = np.std(sampling_distribution)

sampling_mean, sampling_std_dev
```

Take random samples of ratings from the dataset and calculate their means. Create a sampling distribution and explain how this relates to the Central Limit Theorem.



The graph in the picture is a histogram, which represents the sampling distribution of ratings' means.

Insights:

Distribution Shape:

The distribution appears to be roughly bell-shaped, indicating a normal distribution. This is expected according to the central limit theorem, which states that the distribution of sample means approaches a normal distribution as the sample size increases.

Central Tendency:

The peak of the distribution is around 2.8, suggesting that the mean of the ratings is approximately 2.8.

Spread:

The distribution ranges from around 2.4 to 3.5, indicating the variability in the sample means. This spread is measured by the standard deviation, which can be estimated from the histogram.

Sample Size:

The number of samples used to generate this histogram can be estimated by adding the frequencies of all bars.

INGOUDE COMPANY

Home Product Service Contact

THANK YOU

Thank you for your attention! A well-prepared business report is a powerful tool for success.

www.reallygreatsite.com