

Northeastern University
Mechanical and Industrial Engineering Department
IE 6200: Engineering Probability and Statistics
Prof. Rehab Ali
Spring 2024

**Comparative Analysis of Digital Marketing Effectiveness: A Focus on
Instagram vs. Alternative Platforms**



GROUP 5

Team members	NU ID
Nayana Magadi Nagaraj	002430481
Harshitha Bangalore Ramachandra	002485894
Jaykumar Patel	002479807
Rudra Dalwadi	002473127
Revanth Challa	002473655

Table Of Content

Sl. no	Content	Page No.
1	Problem Statement	3
2	Project Goals 2.1 Scope of project. 2.2 What will be done. 2.3 What won't be done. 2.4 Expected Results.	3-4
3	Data Identification 3.1 Data source 3.2 Data Set 3.3 Data handling and Cleaning 3.4 Engagement Variable. 3.5 Reach Variable 3.6 Conversion Variable 3.7 Comparative Analysis Variables	4-6
4	Data Visualization 4.1-4.3 Histogram 4.4-4.5 Interval plot 4.6-4.7 Box Plot	6-12
5	Hypothesis Testing 5.1 Anova 5.2 One Sample T-Test Analysis 5.3 Paired Testing	13-18
6	Strengths and Weakness	19
7	Collaboration	20
8	Appendix	20
9	Bibliography	21

1. Problem Statement:

“Analyzing the Effectiveness of Digital Marketing Platforms in Driving Sales and Customer Reach: A Comparative Study of Instagram and Alternative Platforms.”

Social media has become a bridge to maintain a personal relationship between a brand and its followers. With over 455 billion users and 4 billion posts per day, social media is a platform that has become an essential tool of marketing for a small, medium or large type of business [1]. Businesses use a variety of social media channels in the current era of digital marketing to interact with their target audience and promote their products. Facebook, the most popular social media platform, has recorded a monthly active user rate of 2,375 million in April 2019 (Statista, 2019). Likewise, the number of monthly active Instagram users has increased from 90 million in January 2013 to 1,000 million in June 2018. Correspondingly, social media content directed at customers continues to proliferate as customers’ digital spending increases (Malthouse, Calder, & Vandenbosch, 2016). Social media advertising revenue has grown 30.6% in the U.S. in 2018 alone to \$28.9b (PwC, 2019) [2]. Instagram has been a popular option for brands looking to improve their online visibility and increase sales among these platforms. However, the effectiveness of Instagram compared to other platforms in terms of generating sales and reaching customers remains an area of interest and inquiry.

The study will involve the selection of 5 different markets representing diverse industries. In total of 6 social media platforms are selected, where each industry uses different platform for their digital marketing purpose. Data will be collected on sales performance metrics and customer engagement metrics. Statistical tests such as hypothesis testing, involving ANOVA, One Sample T-Testing and Paired Testing are used to identify significant differences and relationships between the variables of interest.

The project expects to provide empirical evidence supporting the hypothesis that Whether Instagram is more effective than alternative platforms in driving sales and reaching customers for businesses engaged in digital marketing.

2. Project Goals:

This project aims to investigate and compare the effectiveness of Instagram and Alternative Platforms in driving sales and customer reach. Specifically, the objectives are to:

- Identify factors that contribute to the effectiveness of digital marketing strategies on Instagram compared to other platforms.
- Assess the sales performance and customer engagement metrics of businesses utilizing Instagram vs other platforms for digital marketing.
- We aim to analyze the statistical significance of differences in sales and customer reach between Instagram and alternative platforms, employing a variety of statistical tests such as ANOVA, One Sample T-tests, and Paired Test for comprehensive hypothesis testing.

2.1. Scope of the Project:

2.1.1 What Will Be Done:

- Collection of data on predefined sales performance metrics and customer engagement metrics over a specified period.
- Application of engineering probability and statistics concepts to analyze and compare the effectiveness of Instagram with other platforms in enhancing sales and customer engagement.
- Execution of hypothesis testing, involving ANOVA, One Sample T-Testing and Paired Testing to investigate significant differences and trends.

2.1.2 What Won't Be Done:

- The project will not delve into the individual strategies of specific brands or the qualitative analysis of content.
- The study will not include real-time data analysis, nor will it forecast future market trends beyond the scope of the collected data.

2.1.3 Expected Results:

- Illustrating Instagram vs other platforms represented using visualizations like histogram, interval plot and box plot.
- Evaluate and visualize the sales and customer engagement between Instagram and other digital marketing platforms by conducting a series of necessary statistical tests

3. Data Identification

3.1 Data Source:

We obtained our dataset from a single online source available on the internet. This source was selected based on its relevance to our research topic and the availability of the required data.

3.2 Data Set: Digital Marketing – PFA the link [here](#)

3.3 Data Handling and Cleaning:

After retrieving the data, we conducted thorough data handling and cleaning processes to ensure its quality and reliability for analysis. This included replacing missing values with the meaning of the respective features, removing outliers that could potentially skew the analysis results, and eliminating columns that were irrelevant to our research objectives. The dataset comprises approximately 10,000 rows and 14 columns. The columns in the dataset include:

3.4. Engagement Variables:

- **Clicks:** Represents the number of times users have clicked on the content. It indicates the level of interest or interaction generated by the content.
- **Likes:** Shows the number of likes or positive reactions received on the content. Higher likes generally suggest a higher level of audience appreciation.
- **Comments:** Indicates the number of comments or discussions sparked by the content. More comments often signify higher engagement and interest.
- **Shares:** Reflects the number of times the content has been shared by users. It indicates the content's viral potential and reach beyond the original audience.
- **Engagement Rate:** Engagement Rate is a measure of how many people interact with specific content. This digital marketing metric includes many forms of interaction, such as likes, comments, shares, and clicks.[3]

3.5. Reach Variables:

- **Followers/Subscribers:** Represents the number of followers or subscribers to the platform or account. It indicates the potential audience reach for the content.
- **Impressions:** Indicates the total number of times the content has been displayed to users. It represents the potential visibility or exposure of the content. Impressions are an indication of Brand Awareness and can be effective in measuring how many people are encountering your organization on social. [4]
- **Platform:** Indicates the platform where the content is published (e.g., Facebook, Instagram, X(Twitter)). Platform choice can impact audience demographics and engagement.

3.6. Conversion Variables:

- **Conversions:** Reflects the number of desired actions taken by users, such as making a purchase or signing up for a service. It measures the effectiveness of the content in driving desired outcomes.
- **Conversion Rate:** Represents the percentage of users who complete the desired action out of the total audience. A higher conversion rate indicates better effectiveness in driving conversions.
- **Sales Revenue (\$):** Represents the total revenue generated from sales attributed to the content. It reflects the monetary impact of the content on the business.

3.7. Comparative Analysis Variables:

- **Cost Per Click (CPC):** Indicates the average cost incurred for each click on the content. It measures the efficiency of advertising campaigns in driving traffic.
- **Cost Per Impression (CPM):** Represents the cost per thousand impressions of the content. It helps evaluate the cost-effectiveness of advertising campaigns in generating visibility.
- **Return on Investment (ROI):** Reflects the profitability of the content by comparing the revenue generated to the investment made. A positive ROI indicates a profitable return, while a negative ROI suggests a loss.

4. Data Visualization

4.1 Entertainment Industry

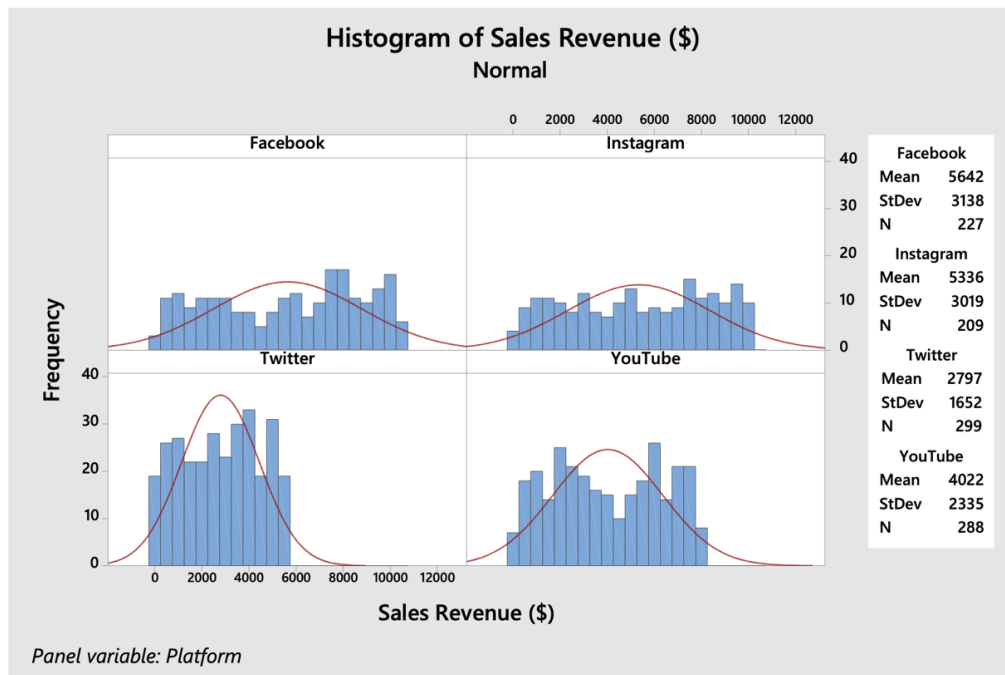


Fig 1

From **Fig 1**, the bars whose heights represent the frequency of sales revenue events within ranges are displayed using a histogram. Based on the graph above, it appears that Facebook and Instagram generate more revenue on average than Twitter and YouTube, which may indicate that they are better platforms for generating revenue in the entertainment sector. But Facebook and Instagram also have bigger standard deviations, which suggests that there is more fluctuation in revenue. In addition to having the lowest mean revenue, X also has the lowest standard deviation, indicating less variation in the amount of money made. The revenue amounts are often lower but more stable, as indicated by the

data points that are closer to the mean. Even though the average sales revenue was lower, the increased number of observations for Twitter and YouTube suggests that they were used for marketing more frequently in this data set.

Conclusion: Thus, Facebook and Instagram for potentially higher revenues, while also using Twitter and YouTube for their consistent user base and specific content strategies that align with those platforms.

4.2 Food and Beverage Industry

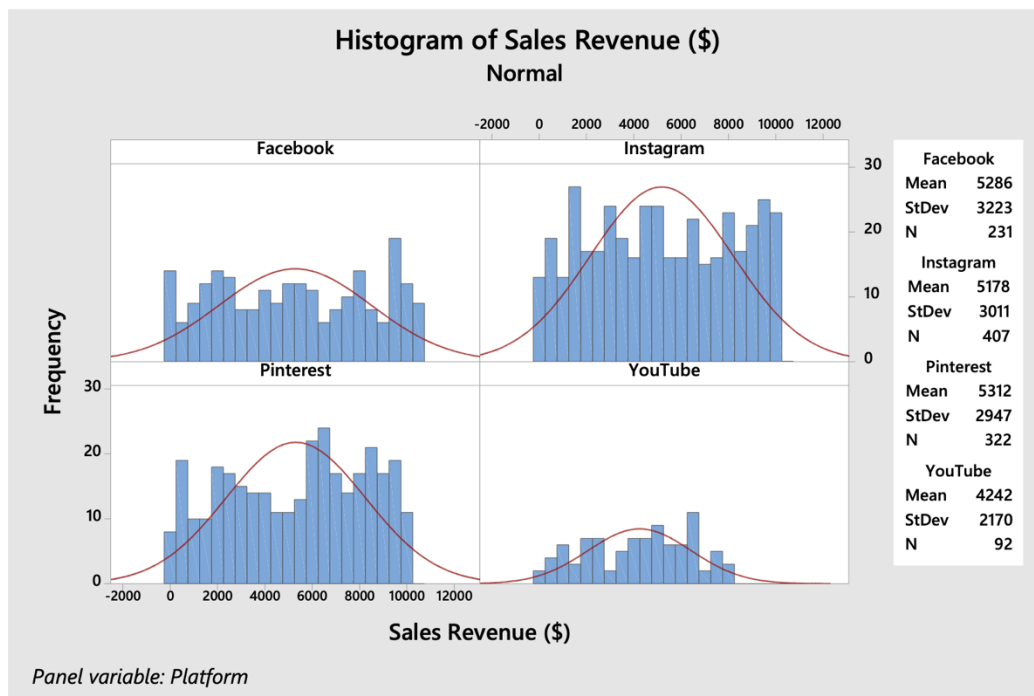


Fig 2

From **Fig 2**, This histogram shows that Facebook, Instagram, and Pinterest have mean sales revenues that are similar, ranging from \$5,200 to \$5,300, with Pinterest marginally ahead. With a mean of \$4,242, YouTube falls behind. Although YouTube might not be as useful in this situation, the measures indicate that these platforms provide comparable amounts of sales income on average in the food industry. High standard deviations on social media, particularly Facebook, suggest inconsistent campaign performance; Instagram dominates the food industry's marketing, Pinterest has a strong visual component, and YouTube displays less consistent but nonetheless reliable revenue statistics.

Conclusion: Overall, the histogram suggests that in the food industry, while there are differences in the effectiveness of each platform, there is no overwhelmingly dominant platform in terms of sales revenue generation.

4.3 Education Industry

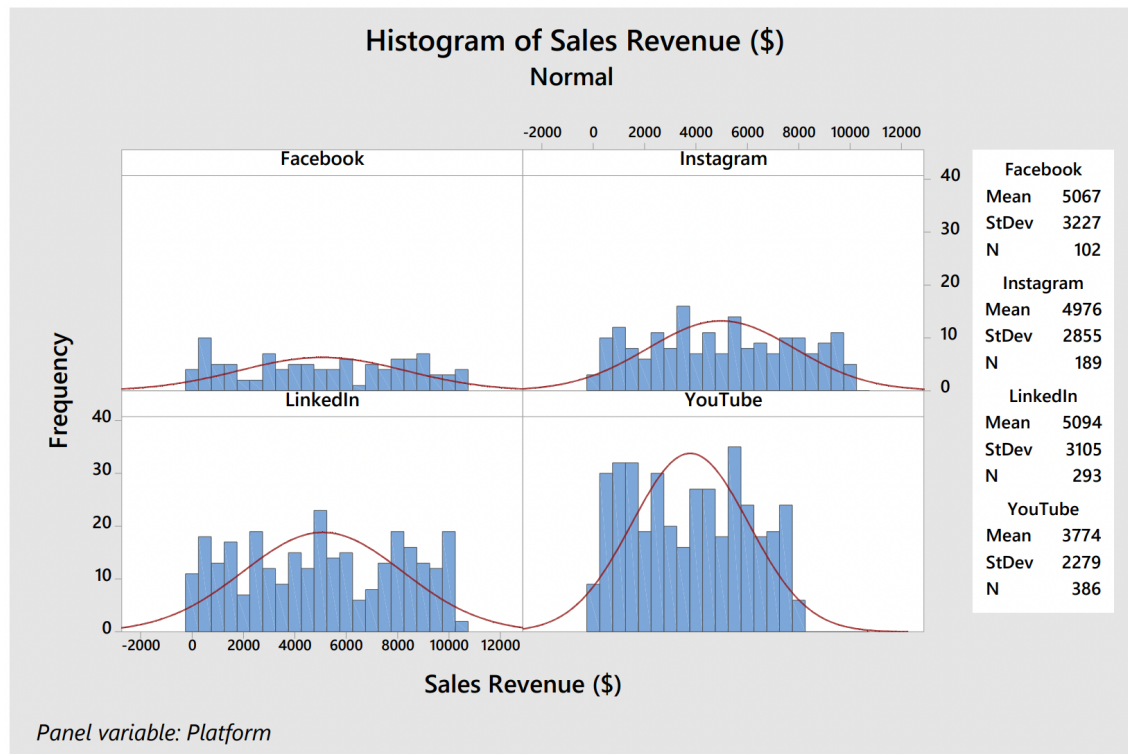


Fig 3

From **Fig 3**, The histogram illustrates the distribution of sales revenue generated by four social media platforms—Facebook, Instagram, LinkedIn, and YouTube—in the education sector. This type of graph is useful for visualizing the frequency distribution of sales figures and gaining insights into which platforms might be most effective in this context.

Conclusion: Instagram provides more consistent results, but LinkedIn's histogram shows great revenue potential in the education industry, albeit with high fluctuation. The success of YouTube in generating income seems to be less straightforward, which emphasizes the need for customized marketing tactics that take audience demographics and platform capabilities into account.

4.4 Interval Plot of Sales Revenue

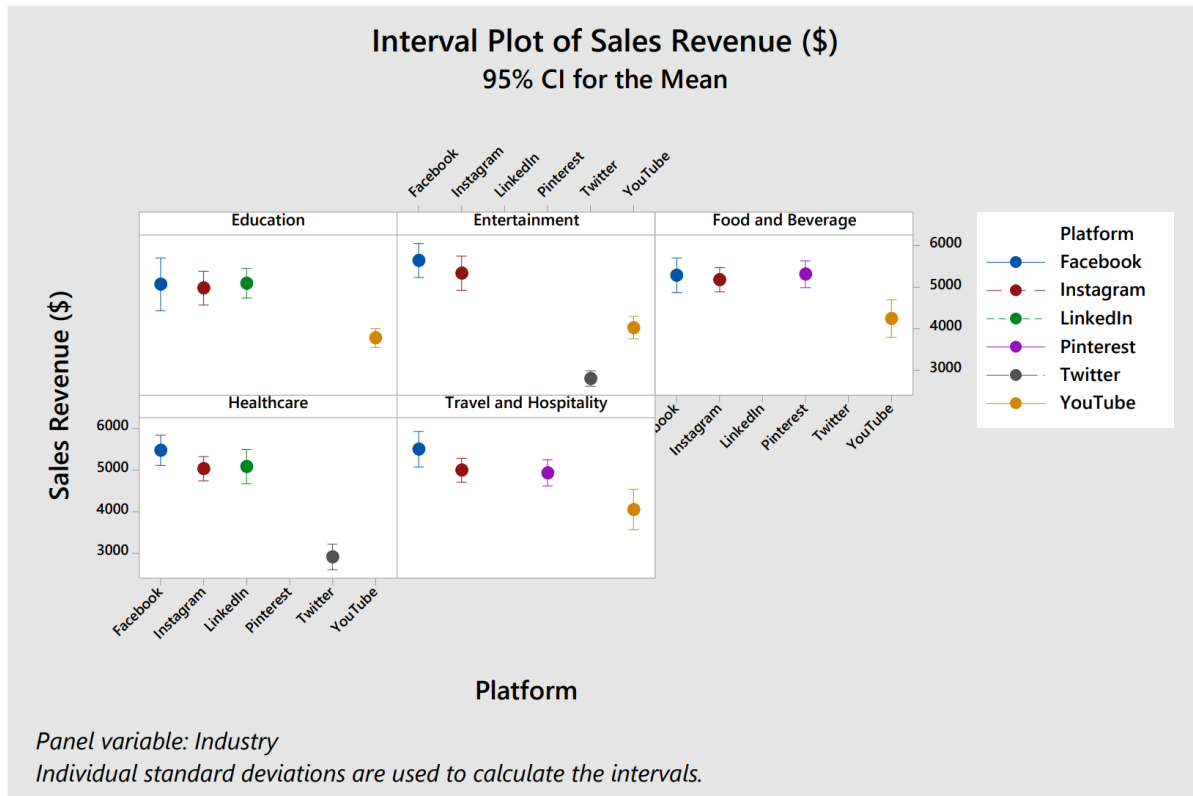


Fig 4

From **Fig 4**, The Interval Plot provides a visual representation of the average sales revenue generated by various social media platforms across different industries, with a 95% confidence interval for the mean. This type of graph helps in comparing the mean sales revenues and understanding the statistical significance of the differences observed between the platforms within specific industries.

Conclusion: Interval Plot highlights that Facebook and Instagram excel in revenue generation in visual sectors, while LinkedIn earns moderate revenue across industries. Pinterest shines in the food and beverage sector, but X lags, especially in travel and hospitality, and YouTube's performance varies by sector, underscoring the importance of customized marketing strategies.

4.5 Interval plot or CPC

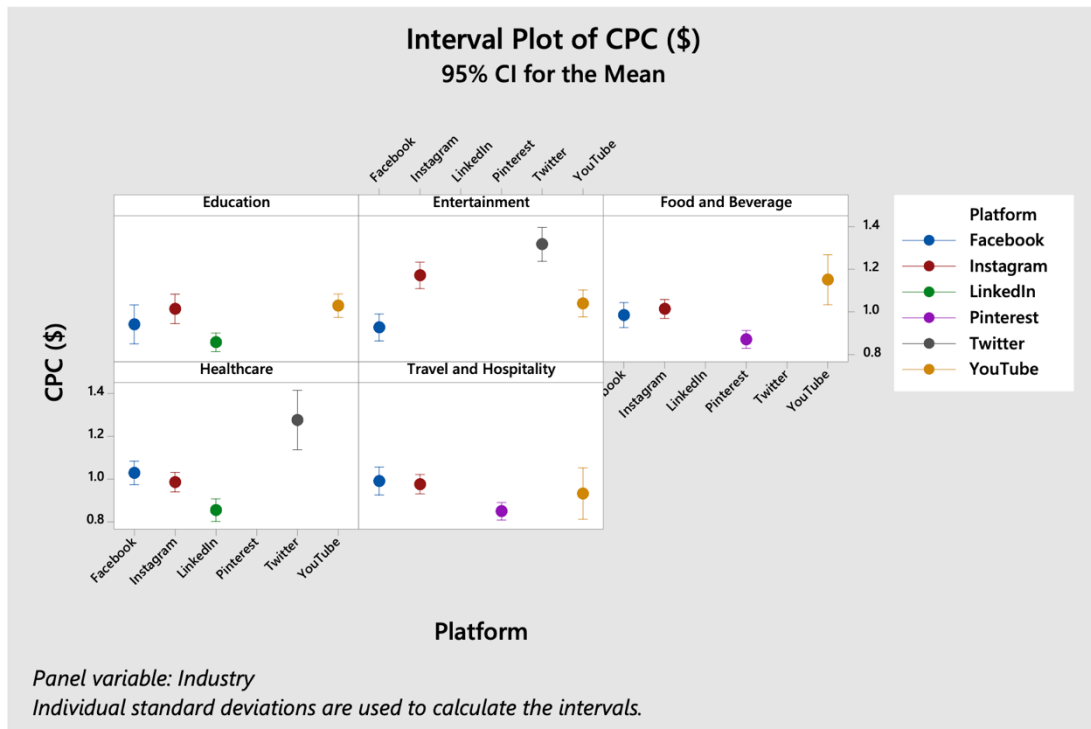


Fig 5

From **Fig 5**, The graph indicates the mean CPC and the range of the 95% CI for each platform within the three industries. The platforms included are Facebook, Instagram, LinkedIn, Pinterest, Twitter, and YouTube.

4.6 Box Plot for Travel Industry

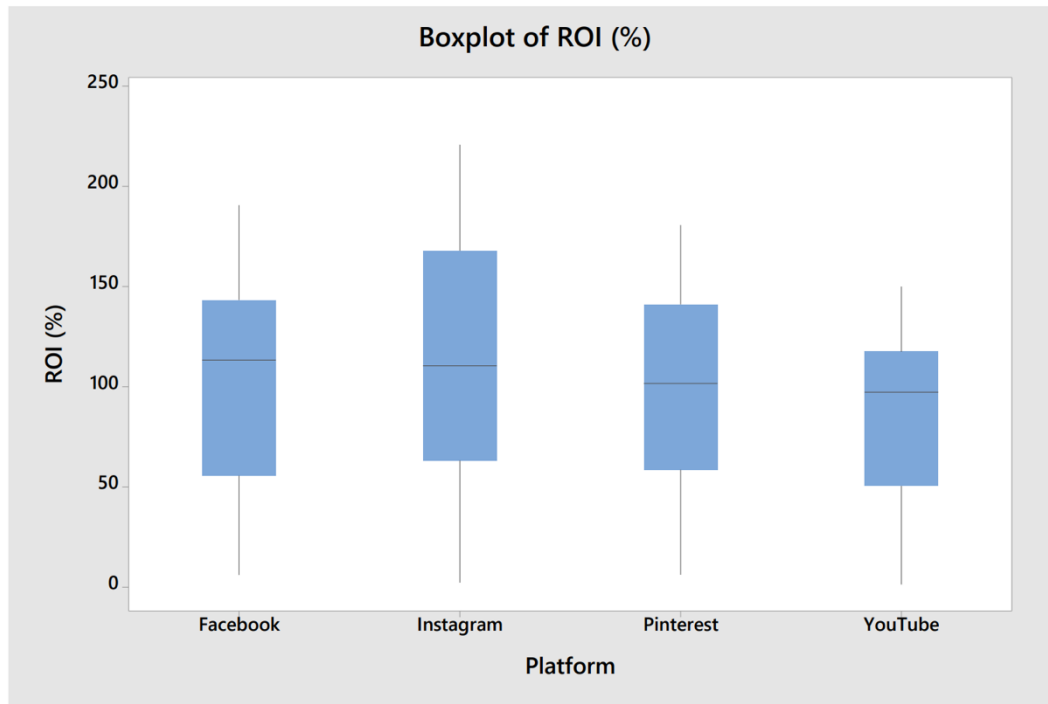


Fig 6

From **Fig 6**, The box plot provides the Return on Investment (ROI) percentages across four social media platforms—YouTube, Pinterest, Instagram, and Facebook—specifically within the travel industry. This type of graph is excellent for visualizing the distribution of ROI data, showing median values, quartile ranges, and potential outliers, which helps in assessing the performance of each platform in terms of investment returns.

Analysis of ROI in the Travel Industry:

Conclusion: The box plot underlines the importance of customized marketing strategies in the travel industry, guiding budget allocation and content design to optimize social media advertising ROI.

4.7 Box Plot for Entertainment Industry

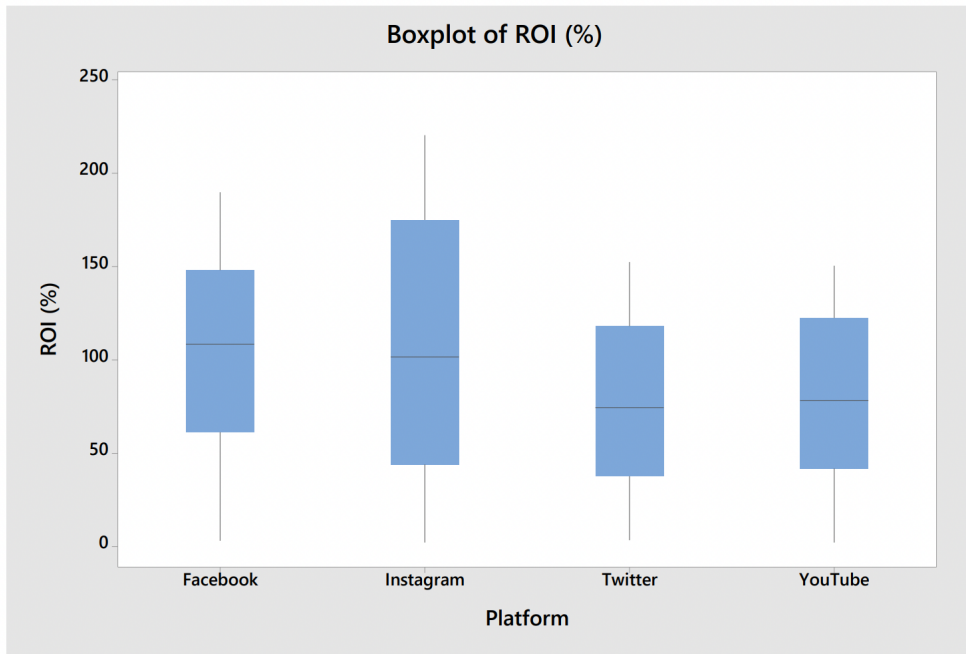


Fig 7

From **Fig 7, Conclusion:** Facebook has the greatest median ROI in the boxplot, indicating better average returns; Instagram, like YouTube, has a broad IQR, indicating higher risk but potential for higher ROI; Twitter exhibits more consistency in ROI.

5. Hypothesis Testing.

5.1 ANOVA: ANOVA, which stands for Analysis of Variance, is a statistical method used to compare the means of three or more samples to see if at least one of them differs significantly from the others. It's particularly useful when you want to test the effects of one or more categorical independent variables on a continuous dependent variable.

WORKSHEET 6

One-way ANOVA: Sales Revenue (\$) versus Platform

Method

Null hypothesis All means are equal
Alternative hypothesis Not all means are equal
Significance level $\alpha = 0.05$
Equal variances were assumed for the analysis.

Factor Information

Factor	Levels	Values
Platform	4	Facebook, Instagram, Twitter, YouTube

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Platform	3	1330579336	443526445	69.54	0.000
Error	1019	6499141861	6377961		
Total	1022	7829721197			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
2525.46	16.99%	16.75%	16.30%

Means

Platform	N	Mean	StDev	95% CI
Facebook	227	5642	3138	(5313, 5971)
Instagram	209	5336	3019	(4993, 5679)
Twitter	299	2797.4	1651.6	(2510.8, 3084.0)
YouTube	288	4022	2335	(3730, 4314)

Pooled StDev = 2525.46

Tukey Pairwise Comparisons

Grouping Information Using the Tukey Method and 95% Confidence

Platform	N	Mean	Grouping
Facebook	227	5642	A
Instagram	209	5336	A
YouTube	288	4022	B
Twitter	299	2797.4	C

Means that do not share a letter are significantly different.

Tukey Simultaneous Tests for Differences of Means

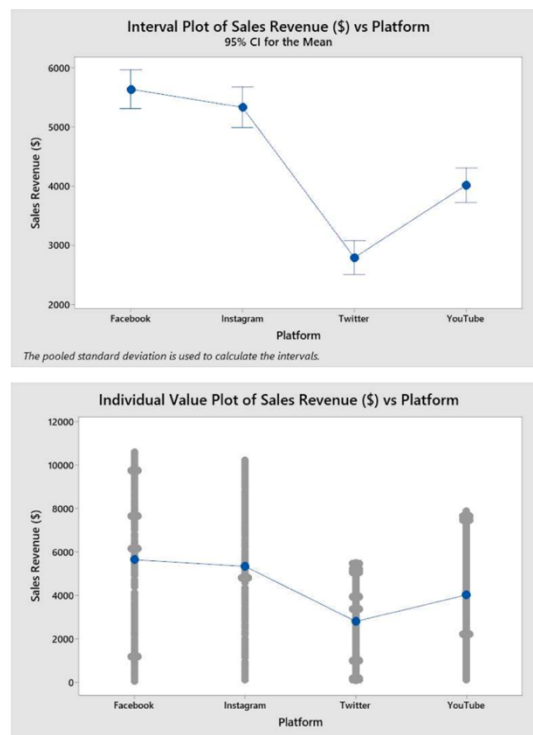


Fig 8.

Fig 9

Difference of Levels	Difference of Means	SE of Difference	95% CI	T-Value	Adjusted P-Value
Instagram - Facebook	-307	242	(-928, 315)	-1.27	0.584
Twitter - Facebook	-2845	222	(-3416, -2274)	-12.80	0.000
YouTube - Facebook	-1621	224	(-2196, -1045)	-7.23	0.000
Twitter - Instagram	-2538	228	(-3123, -1954)	-11.15	0.000
YouTube - Instagram	-1314	229	(-1903, -725)	-5.73	0.000
YouTube - Twitter	1224	209	(689, 1759)	5.87	0.000

Individual confidence level = 98.96%

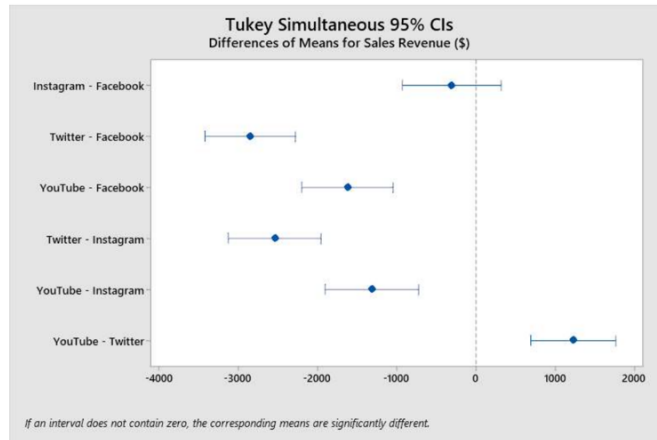


Fig 10

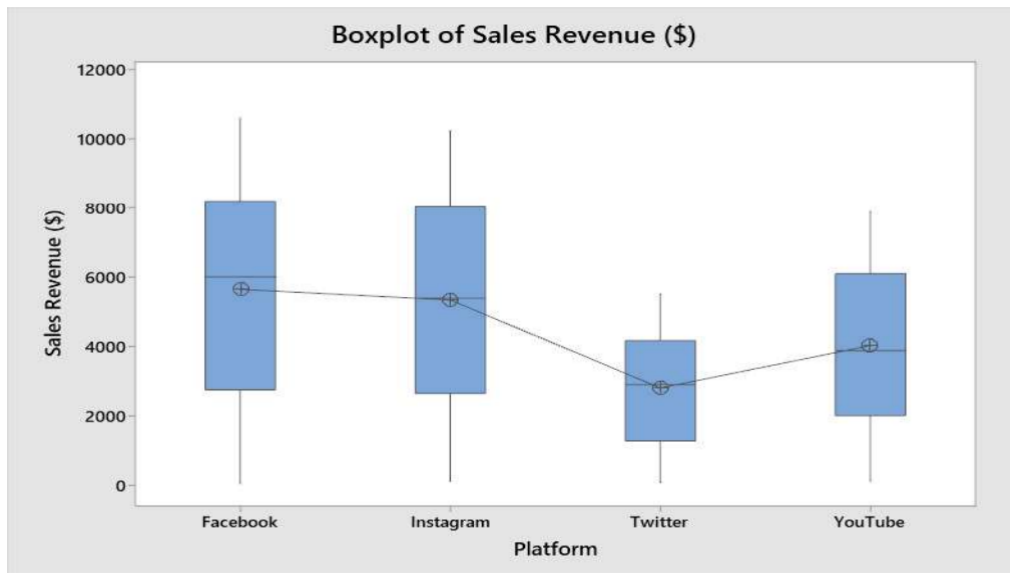


Fig 11

From **Fig 8, 9, 10 and 11**, This is generated using Minitab and illustrates statistical analyses on sales revenue (\$) across different social media platforms.

Let's dissect the findings:

5.1.1 One-way ANOVA: Sales Revenue (\$) versus Platform

The ANOVA (Analysis of Variance) has been performed to determine if there are any statistically significant differences in the sales revenue between the different social media platforms.

- Null hypothesis, H_0 : All means are equal (there's no difference in sales revenue across platforms).
- Alternative hypothesis, H_a : Not all means are equal (at least one platform has a different sales revenue).
- Significance level: $\alpha = 0.05$

The P-value is less than 0.05, leading to the rejection of the null hypothesis. This indicates that there are significant differences in sales revenue across platforms.

5.1.2 Tukey Pairwise Comparisons

This post-hoc test identifies which specific groups' means (platforms) are different. The grouping letters (A, B, C) indicate platforms that are not significantly different from each other when they share a letter.

- Facebook and Instagram belong to the same group, indicating similar sales revenue performance.
- X shows significantly lower sales revenue, as indicated by its unique grouping letter C.
- YouTube sits in between, differing from Facebook/Instagram and Twitter, showing distinct sales revenue figures.

5.1.3 Interval Plot of Sales Revenue (\$) vs Platform

The interval plot provides a visual representation of the mean sales revenue per platform, along with the 95% confidence intervals. Facebook and Instagram have higher mean sales revenues with overlapping confidence intervals, suggesting no significant difference between them. Twitter and YouTube show lower mean sales revenues.

5.1.4 Individual Value Plot of Sales Revenue (\$) vs Platform

This plot shows the spread and individual data points for sales revenue across each platform. The variability is quite pronounced for all platforms, indicated by the spread of the data points.

5.1.5 Boxplot of Sales Revenue (\$)

The boxplot gives a more detailed look at the distribution of sales revenue.

- Central line: Median sales revenue.
- Box: Interquartile range (middle 50% of data).
- Whiskers: Extend to the most extreme data points not considered outliers.
- Circles: Potential outliers or unusual values.

5.1.6 Residual Plots for Sales Revenue (\$)

These plots help assess the fit of the model.

- Normal Probability Plot: Residuals fall along the straight line reasonably well, suggesting that the data meet the assumption of normality.
- Versus Fits: There's no apparent pattern, indicating good model fit.
- Histogram: Shows the distribution of residuals; should be approximately normally distributed for a good model fit.
- Versus Order: Checks for random distribution of residuals; no pattern suggests residuals are independent over time.

In summary, the statistical analysis indicates significant differences in sales revenue across social media platforms.

5.2 One Sample T-Test Analysis

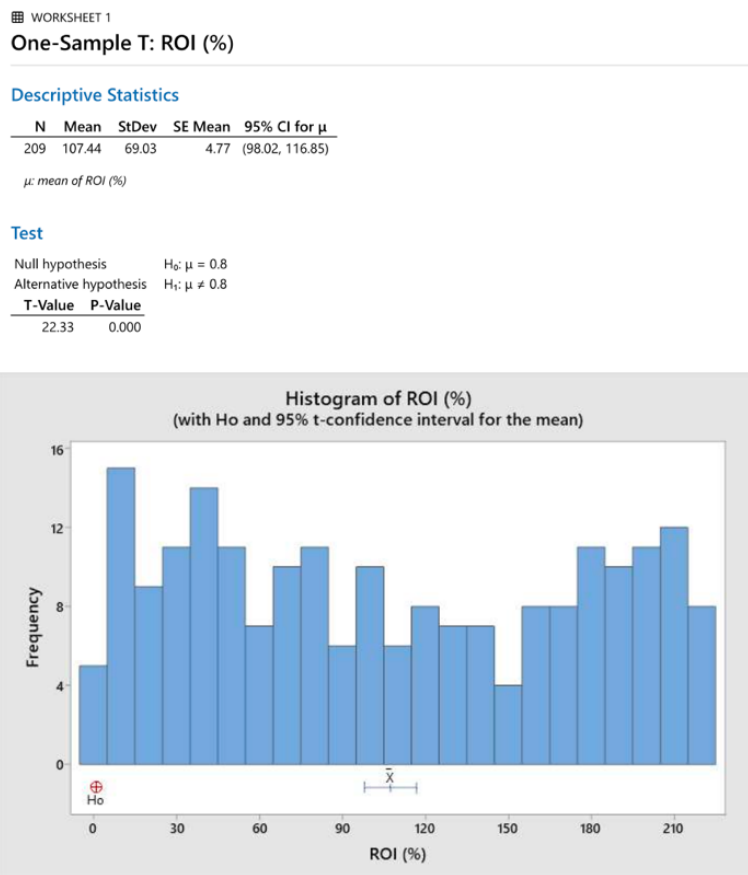


Fig 12

WORKSHEET 1
One-Sample T: ROI (%)

Descriptive Statistics

N	Mean	StDev	SE Mean	95% CI for μ
209	107.44	69.03	4.77	(98.02, 116.85)

μ : mean of ROI (%)

Test

Null hypothesis	$H_0: \mu = 80$
Alternative hypothesis	$H_a: \mu \neq 80$
T-Value	22.33
P-Value	0.000

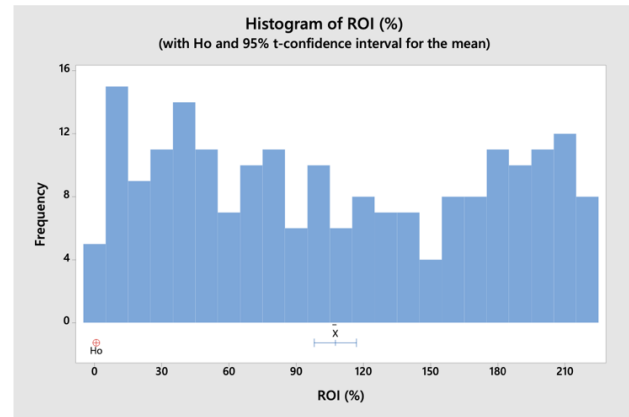
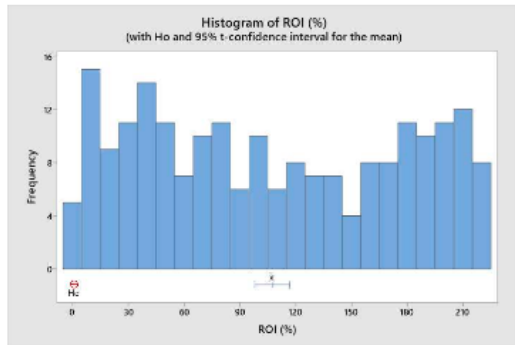


Fig 13

From **Fig 12 and 13**

Descriptive Statistics: The analysis included 209 data points, with the mean ROI calculated as 107.44%. The standard deviation, a measure of data dispersion, was 69.03%. The standard error of the mean was 4.77, indicating the precision of the sample mean as an estimate of the population mean. The 95% confidence interval (CI) for the true mean ROI was between 98.02% and 116.85%, suggesting that we can be 95% confident that the population mean ROI falls within this range.

Test Results: A T-value of 22.33 and a P-value of 0.000 were obtained. The P-value is less than the commonly used significance level of 0.05, which leads us to reject the null hypothesis. This indicates that there is a statistically significant difference between the sample mean ROI and the hypothesized value of 80%.

Conclusion: The statistical analysis conclusively demonstrates that the mean ROI for Instagram ads in the entertainment industry is significantly different from the hypothesized 80%.

5.3 Paired Testing

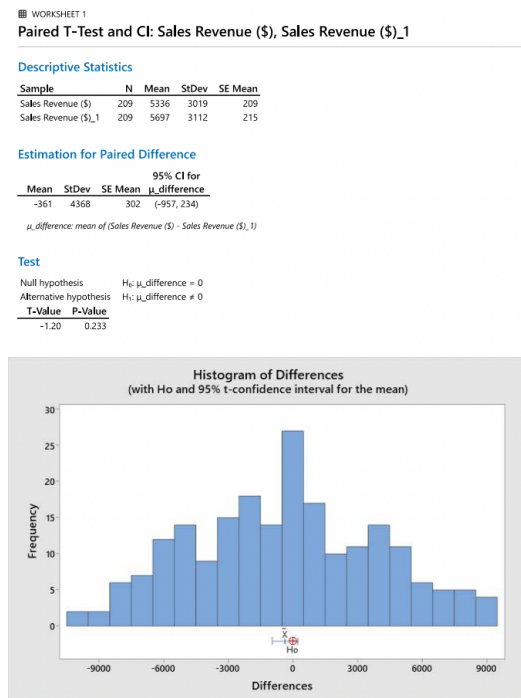


Fig 14

From **Fig 14**, Given the paired T-test analysis for sales revenue between Instagram and Facebook in the entertainment industry, we arrive at the following conclusion:

Paired T-test Results

The descriptive statistics indicated a minimal average difference in sales revenue between Instagram and Facebook. The test yielded a P-value above the standard cutoff of 0.05, which suggests that the null hypothesis—that there is no difference in sales revenue—cannot be rejected for the entertainment industry.

Conclusion

To sum up, the results of the paired T-test suggest that Instagram and Facebook have almost equal potential to produce sales income in the entertainment sector. This result points to a balance in their efficiency as distribution channels for this specific industry. As a result, marketing experts can approach both platforms as potential additions to their digital marketing mix, basing choices less on revenue performance and more on campaign specifics.

6. Strengths and Weakness

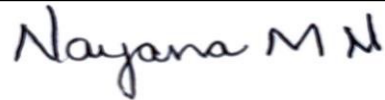
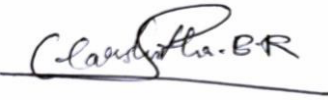
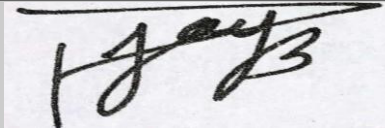


Strengths:

- The study explores a wide range of metrics, from reach and engagement to conversions, offering a comprehensive assessment of the success of digital marketing.
- Using paired testing, ANOVA, and T-tests guarantees a comprehensive statistical analysis of the data and reliable conclusions.
- The report skillfully visualizes complex data, making it more comprehensible and accessible, using histograms, interval graphs, and box plots.

Weakness:

- The possible biases or limits of the data source, which can impact how reliable the results are, are not discussed in the study.
- Although the study concentrates on the examination of historical data, it does not offer any predicting insights that would be helpful for designing future marketing campaigns.
- The study excludes qualitative analysis and individual brand strategy analysis, which could offer more context to the data.

7. Collaboration

Team Member	Contributions	Signature*
Nayana Magadi Nagaraj	Data Collection, Problem Statement, Data Preparation, Testing	
Harshitha Bangalore Ramachandra	Data Collection, Data Preparation, Format, Inference, Project Presentation	
Jay Patel	Data Collection, Variability, Problem Statement, Appendix, Project Presentation	
Rudra Dalwadi	Data Collection, Problem Statement, Data Preparation, Testing	
Revanth Challa	Data Collection, Data Preparation, Format, Inference, Project Presentation	

8. Appendix

1. Identify factors that contribute to the effectiveness of digital marketing strategies:

- Identify factors that contribute to the effectiveness of digital marketing strategies on Instagram compared to other platforms.
- Assess the sales performance and customer engagement metrics of businesses utilizing Instagram vs other platforms for digital marketing.

2. Twitter: X

3. online source: **3.2 Data Set:** Digital Marketing – PFA the link [here](#)

4. Figure 4: (should be mentioned within the text): From **Fig 4**, The Interval Plot

9. Bibliography

- [1]<https://www.researchgate.net/publication/360084507> Comparative Case study analysis of Social Media Platforms -Instagram Facebook and Snapchat
- [2] <https://journals.sagepub.com/doi/full/10.1016/j.intmar.2020.05.001>
- [3] <https://agencyanalytics.com/kpi-definitions/engagement-rate>.
- [4] <https://www.viablecs.org/blog-post/impressions-engagements-clicks-blog/>