Czech University of Life Sciences Prague

Faculty of Economics and Management

Department of Statistics

Master's Thesis

Impact of Social Media Data on Consumer Behavior

Melika Amini Moghaddam

© 2025 CZU Prague

CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

Faculty of Economics and Management

DIPLOMA THESIS ASSIGNMENT

Bc. Melika Amini Moghaddam

Informatics

Thesis title

Impact of Social Media Data on Consumer Behavior

Objectives of thesis

This thesis aims to evaluate how social media content influences consumer purchase decisions and evaluate the effectiveness of social media analytics in predicting consumer behaviour. It aims to assess the impact of social media marketing on brand loyalty and discuss the ethical considerations in using social media data for consumer analysis. The thesis will investigate the influence of user-generated content on consumer trust and examine the efficacy of targeted social media advertising to ascertain how social media influencers affect how customers behave. Additionally, it will compare consumer engagement across different social media platforms, investigate the impact of social media trends on consumer behaviour, and evaluate the role of social media in crisis management and consumer perception.

Methodology

There are two components to the technique used in this thesis: theoretical and practical. The examination of secondary materials, such as published studies, academic publications, articles, and reliable web sites, will form the foundation of the theoretical portion.

The practical element will include both quantitative and qualitative research. For the quantitative study, an online survey will be created and delivered to a significant number of consumers.

Furthermore, the survey results will be evaluated to determine the links between social media interactions and customer behavior.

Interviews will be performed with a smaller, more targeted sample of participants for the qualitative study to acquire a better understanding of their experiences and perspectives of the effect of social media on their behavior.

The proposed extent of the thesis

60 – 80 pages

Keywords

Social media, consumer behavior, effect, statistical analysis, decision

---

Recommended information sources

Abbott, D., 2014. Applied Predictive Analytics : Principles and Techniques for the Professional Data Analyst. John Wiley & Sons, Incorporated. ISBN 9781118727935.

Bianchi, A., 2020. Driving Consumer Engagement in Social Media : Influencing Electronic Word of Mouth, Taylor & Francis Group, Oxford. Available from: ProQuest Ebook Central. [9 June 2024].

Godey, B., Manthiou, A., Pederzoli, D., Rokka, J., Aiello, G., Donvito, R. and Singh, R., 2016. Social media marketing efforts of luxury brands: Influence on brand equity and consumer behavior. Journal of business research, 69(12), pp.5833-5841.

Heinonen, K., 2011. Consumer activity in social media: Managerial approaches to consumers' social media behavior. Journal of consumer behaviour, 10(6), pp.356-364.

Stephen, A.T., 2016. The role of digital and social media marketing in consumer behavior. Current opinión in Psychology, 10, pp.17-21.

Vinerean, S., Cetina, I., Dumitrescu, L. and Tichindelean, M., 2013. The effects of social media marketing on online consumer behavior. International journal of business and management, 8(14), p.66.

Vlad, DE 2020, Concepts of Quality Connected to Social Media and Emotions, Springer Vieweg. in Springer Fachmedien Wiesbaden GmbH, Wiesbaden. Available from: ProQuest Ebook Central. [9 June 2024].

CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE
1906

Expected date of thesis defence

2024/25 SS – PEF

The Diploma Thesis Supervisor

doc. Ing. Tomáš Hlavsa, Ph.D.

Supervising department

Department of Statistics

Electronic approval: 11. 06. 2024

doc. Ing. Tomáš Hlavsa, Ph.D.

Head of department

Electronic approval: 04. 11. 2024

doc. Ing. Tomáš Šubrt, Ph.D.

Dean

Prague on 30. 03. 2025

Declaration

I declare that I have worked on my master's thesis titled " Impact of Social Media Data on Consumer Behavior" by myself and I have used only the sources mentioned at the end of the thesis. As the author of the master's thesis, I declare that the thesis does not break any copyrights.

In Prague, March 2025

Acknowledgement

I would like to thank Ing. Tomáš Hlavsa and all other friends and acquaintances for their advice and support during my work on this thesis.

Impact of Social Media Data on Consumer Behavior

Abstract

The thesis examines the influence of social media data on consumer behavior, emphasizing marketing insights and informatics applications. The advancement of digital platforms significantly impacts purchasing decisions, brand allegiance, and consumer trust via social media. The study investigates the effect of data-driven strategies—such as user-generated content, targeted advertising, and influencer marketing—on consumer preferences and engagement across various platforms.

Influencers, user-generated content, and personalized advertising all shape consumer views and trust. Furthermore, social media developments profoundly affect client relationships with firms in many contexts. The thesis use statistical modelling and analytics to assess behavioral patterns and digital interactions from an informatics perspective. The inquiry aims to examine the capacity of analytics to forecast purchasing trends and evaluate the ethical implications of utilizing consumer data, with the goal of uncovering the impact of social media on modern buying behavior. The aim is to clarify the complex relationship between social media use and consumer decision-making in the modern digital landscape.

The results show that user involvement, influencer content, and focused advertising as well as social media interaction all significantly affect customer trust and choices. Data patterns indicate substantial connections between social media presence and purchasing behavior, while qualitative research highlight the essential importance of trust and privacy in consumer perceptions.

.

Keywords: Social media, consumer behavior, user-generated content, influencer marketing, brand loyalty, predictive analytics, trust, privacy, targeted advertising.

Vliv dat ze sociálních médií na chování spotřebitelů

Abstrakt

Tato diplomová práce zkoumá vliv dat ze sociálních médií na chování spotřebitelů se zaměřením na marketingové poznatky a aplikace informatiky. Rozvoj digitálních platforem významně ovlivňuje nákupní rozhodování, loajalitu ke značce a důvěru spotřebitelů prostřednictvím sociálních médií. Studie se zabývá účinkem strategií založených na datech – jako je obsah vytvářený uživateli, cílená reklama a influencer marketing – na spotřebitelské preference a zapojení napříč různými platformami.

Influenceři, uživatelský obsah a personalizovaná reklama společně formují vnímání a důvěru spotřebitelů. Vývoj sociálních médií dále zásadně ovlivňuje vztahy zákazníků k firmám v různých kontextech. Diplomová práce využívá statistické modelování a analytické nástroje k posouzení vzorců chování a digitálních interakcí z pohledu informatiky. Cílem výzkumu je analyzovat schopnost analytiky předpovídat nákupní trendy a zhodnotit etické důsledky využívání spotřebitelských dat, s cílem odhalit dopad sociálních médií na současné nákupní chování. Záměrem je objasnit složité propojení mezi používáním sociálních médií a rozhodováním spotřebitelů v moderním digitálním prostředí.

Výsledky ukazují, že zapojení uživatelů, obsah od influencerů, cílená reklama i interakce na sociálních sítích významně ovlivňují důvěru a rozhodování zákazníků. Datové vzorce naznačují podstatné souvislosti mezi přítomností na sociálních médiích a nákupním chováním, zatímco kvalitativní výzkum zdůrazňuje klíčový význam důvěry a ochrany soukromí v rámci spotřebitelského vnímání.

.

Klíčová slova: Sociální média, chování spotřebitelů, obsah vytvářený uživateli, influencer marketing, loajalita ke značce, prediktivní analytika, důvěra, soukromí, cílená reklama.

Table of content

1. Introduction

The rapid evolution of social media has drastically transformed consumer-business interactions, affecting purchase choices, brand loyalty, and marketing methods. Initially, early platforms like Friendster and MySpace functioned as digital venues for social networking; but, as user participation increased, firms saw their potential as formidable marketing instruments. The advent of YouTube and Twitter in the mid-2000s proliferated user-generated content, allowing consumers to disseminate experiences globally. These interactions were important in creating public opinion, since online evaluations and testimonials profoundly impacted brand trust and consumer purchase behavior.

The rise of social media has become data-driven marketing methods essential for customer interaction. Platforms such as Instagram, established in 2010, transformed digital marketing by pioneering influencer marketing, whereby credible content providers directly influence customer trust and loyalty. Advancements in social media analytics have allowed companies to use user-generated data for predictive modeling, enhancing targeted marketing and maximizing engagement via tailored content distribution. This revolution signified a transition to algorithm-driven study of customer behavior, whereby machine learning models and AI-generated insights forecast and influence purchase choices.

In recent years, social commerce has become a predominant force, amalgamating product discovery, consumer evaluations, and effortless transactions inside social networks. Although these advances improve customer convenience, they also provoke issues over data privacy, ethical implications, and algorithmic bias, as companies increasingly depend on personal data for behavioral forecasting. The ethical ramifications of using customer data for marketing customization need rigorous examination, especially concerning openness, permission, and data security.

The study examines the convergence of social media data and consumer behavior from an informatics standpoint, emphasizing how corporations use content tactics, influencer partnerships, and targeted advertising to shape customer choices. The thesis seeks to examine the efficacy of social media analytics in forecasting purchase behavior, measure brand loyalty in digital environments, and investigate the ethical implications of data-driven marketing. This study will use quantitative and qualitative approaches to empirically evaluate the influence of social media on contemporary consumer behavior.

2. Objectives and Methodology

2.1 Objectives

Examining important aspects of digital engagement, data-driven marketing, and predictive analysis, this thesis seeks to investigate how social media affects consumer behavior. It looks at how social media material influences buying choices and assesses the accuracy of social media analytics in predicting customer behavior using behavioral monitoring and interaction patterns.

Furthermore, it will examine the influence of social media marketing on brand loyalty, exploring how companies use targeted advertising and interaction tactics to cultivate enduring customer connections. The research will examine the ethical ramifications of using consumer data in digital marketing, emphasizing privacy issues, algorithmic bias, and openness in data-informed decision-making.

The primary objective is to analyze the effect of user-generated content (UGC) on customer trust, evaluating the reliability and influence of peer reviews, influencer endorsements, and community interactions on purchasing choices. Focusing on how algorithm-driven ad placements and influencer endorsements affect customer behavior, this paper investigates the efficacy of targeted advertising on social media. Its goal is to assess user involvement on different platforms, hence determining platform-specific trends, behavioral differences, and the general effect of marketing tactics. The study also looks at how social media trends—such as viral content, changing themes, and digital environment shifts—affect consumer attitudes and buying choices.

Finally, the thesis will evaluate the function of social media in crisis management, analyzing its efficacy in influencing customer perception, controlling brand reputation, and alleviating the effects of negative incidents via swift digital reaction tactics.

2.2 Methodology

This thesis employs an approach that incorporates both theoretical and practical elements. The theoretical framework is established by an assessment of secondary materials, including published research, scholarly publications, papers, and credible online resources. This study offers a systematic comprehension of the connection between social media and consumer behavior by synthesizing known theoretical frameworks and actual data. The research utilizes consumer decision-making models to analyze how individuals process information and make purchasing decisions, engagement theories to investigate the impact of user interaction and content consumption, and social media marketing strategies encompassing influencer marketing, targeted advertising, and algorithmic content recommendations. Furthermore, ethical concerns like data privacy, algorithmic biases, and transparency in digital marketing are examined.

The practical component employs a mixed-methods approach, integrating quantitative and qualitative techniques to get profound insights. The quantitative component entails a systematic online survey administered to active social media users in Prague, examining critical factors such as the frequency of social media interaction, the impact of marketing and influencer content on purchasing choices, the importance of social media analytics in consumer behavior, and consumer trust in brands and digital advertisements. The gathered data were examined using statistical methodologies, including the Chi-Square test, logistic regression.

The Chi-Square test was used to determine if a significant link exists between categorical variables, such as the correlation between social media participation frequency and purchase choices. It is determined using the formula:

$$X^2 = \sum(\frac{(O-E)^2}{E} \quad (1)$$

where O denotes observed frequencies and E indicates predicted frequencies. The analysis presumes that the data are categorical, the observations are independent, and the predicted cell frequencies are sufficiently large—generally a minimum of five—to satisfy the test's criteria.

Logistic regression was used to evaluate the likelihood of a customer completing a purchase based on independent factors, including exposure to influencer material and engagement level. The likelihood of an event transpiring is assessed using the equation:

$$P(Y=1) = \frac{e^{(\beta 0+\beta 1X1+\beta 2X2+\cdots+\beta nXn)}}{1+e^{(\beta 0+\beta 1X1+\beta 2X2+\cdots+\beta nXn)}} \quad (2)$$

where $P(Y=1)$ denotes the likelihood of a purchase, X_i are predictor variables, and β_i are estimated coefficients. Additionally, regression analysis and predictive modeling were used to examine patterns and correlations within the data, yielding insights into customer behavior in the realm of social media marketing.

In binary logistic regression, model fit statistics are essential for assessing whether the addition of independent variables substantially enhances the model's predictive capability relative to a model devoid of predictors. SAS Studio offers many essential measures that facilitate the evaluation of overall model performance:

1. -2 Log Likelihood (-2LL)

The -2 Log Likelihood value quantifies the model's explanatory power regarding the observed outcomes. It originates from the likelihood function, which assesses the possibility of witnessing the stated data under the designated model.

- A reduced -2LL indicates a better model fit.
- A comparison is conducted between the intercept-only model, which lacks predictors, and the comprehensive model that includes predictors.
- A significant decrease in -2LL from the intercept-only model to the comprehensive model indicates that the predictors substantially enhance the model's efficacy.

2. Akaike Information Criterion (AIC)

The AIC evaluates model quality while imposing a penalty for complexity.

- It equilibrates goodness-of-fit and parsimony (model simplicity).
- Reduced AIC values signify a superior model, indicating enhanced fit without superfluous complexity.
- It is particularly advantageous for comparing various models.

3. Likelihood Ratio Chi-Square Test

This test contrasts the -2 Log Likelihood of the intercept-only model with that of the comprehensive model.

- It generates a Chi-Square statistic and a p-value.
- A statistically significant p-value (often $p < 0.05$) signifies that the model with predictors demonstrates a markedly superior fit to the data compared to the model devoid of predictors.
- This substantiates the overall relevance of the regression model.

The Wald Test assesses the statistical significance of various predictors inside a logistic regression model. It assesses if the coefficient of a predictor is considerably distinct from zero, indicating the variable's meaningful contribution to outcome prediction.

Formula:

$$W = (\frac{\beta_i}{SE(\beta_i)})^2 \tag{3}$$

Where:

- β_i is the estimated coefficient for predictor i
- $SE(\beta_i)$is the standard error of the coefficient

A p-value below 0.05 generally indicates that the predictor is statistically significant.

Odds Ratios (ORs) were used to assess the amount and direction of the impact of each variable in logistic regression. They are computed as follows:

$$OR = e^{\beta i} \tag{4}$$

- βi is the estimated regression coefficient

An odds ratio (OR) greater than 1 signifies heightened likelihood of the result with an increase in the predictor, while an OR less than 1 denotes diminished likelihood. This analysis facilitates comprehension of the influence magnitude of each variable on customer purchasing behavior.

The Confusion Matrix is an essential instrument for assessing the classification efficacy of the logistic regression model. The comparison involves the model's projected outcomes and the actual observed results, consisting of four primary components: True Positives (TP) denote accurately predicted purchases; False Positives (FP) signify predicted purchases that did not materialize; True Negatives (TN) represent accurately predicted non-purchases; and False Negatives (FN) pertain to purchases that occurred but were not anticipated by the model. From these values, critical performance measures like accuracy, sensitivity (true positive rate), and specificity (true negative rate) are computed to evaluate the model's overall dependability and predictive efficacy.

The Receiver Operating Characteristic (ROC) Curve is a graphical representation that demonstrates a model's capacity to differentiate between binary outcomes across different thresholds. It graphs the genuine positive rate (sensitivity) in relation to the false positive rate (1 - specificity). The Area Under the Curve (AUC) measures the model's overall capacity to differentiate across classes. Which AUC = 1.0 shows perfect model, AUC > 0.8: Strong model and AUC = 0.5: Indicates lack of discriminative ability (equivalent to random guessing).Elevated AUC values indicate superior model performance.

The Hosmer-Lemeshow Test evaluates the adequacy of fit for a logistic regression model. It contrasts the anticipated and actual occurrence rates within groupings (often deciles) of ascending expected probability. An insignificant p-value ($p > 0.05$) indicates that the model adequately fits the data. A substantial p-value ($p < 0.05$) indicates a lack of fit, indicating that the model's projected probability diverges from actual results. The test assesses the congruence between the model's predictions and the observed data.

Spearman’s Rank Correlation Coefficient (ρ) quantifies the strength and direction of a monotonic association between two ordinal or ranked variables. It is particularly advantageous when the data fail to satisfy the normalcy assumptions necessary for Pearson correlation.

$$\rho = 1 - \frac{6\sum di^2}{n(n^2 - 1)} \tag{5}$$

- di : Difference in ranks for each pair of values
- n: Number of observations

The values of ρ vary from -1, indicating complete negative correlation, to +1, indicating perfect positive correlation, with 0 signifying no link

The qualitative component involves semi-structured interviews with chosen participants to investigate their personal experiences and attitudes toward social media marketing, their perceptions of brand trust and advertising efficacy, and their apprehensions regarding data privacy and personalized marketing. Thematic analysis was used to organize answers into principal themes within Python and Jupyter notebook, facilitating the discovery of prevailing trends in customer attitudes. The concluding phase entails synthesizing quantitative data with qualitative insights to guarantee a holistic comprehension of social media's impact on consumer decision-making.

To uphold ethical integrity, all participants provided informed permission, and data confidentiality was rigorously maintained. The study adheres to ethical standards for studies involving human participants, guaranteeing openness in data acquisition and processing. This research employs both quantitative and qualitative methodologies to provide a comprehensive and impartial analysis of social media's influence on consumer decision-making, using statistical and theme analyses to guarantee solid and dependable results.

3. Literature Review

Social media has profoundly altered customer behavior, affecting purchasing choices, brand loyalty, and trust. Predictive analytics refers to the use of data to uncover valuable patterns that can support future decision-making (Abbott, 2014). Its increasing integration into social media platforms enables companies to interpret user behavior and anticipate consumer preferences, leading to more strategic and data-informed marketing approaches. As organizations gather and organize large datasets, there is a natural shift toward leveraging this information to refine predictions, enhance decision-making, and improve overall efficiency. The process of discovering interesting and meaningful patterns in data is pointing to predictive analytics (Abbott, 2014).

User-generated content (UGC) and brand engagement are essential for cultivating consumer trust. Studies demonstrate that customers are more inclined to interact with firms that promptly address customer criticism on social media, hence enhancing brand loyalty. This engagement requires careful oversight, since social media crises may rapidly damage a brand's reputation if not handled properly.

Furthermore, social media trends and influencer marketing significantly affect client perceptions. Viral marketing campaigns and influencer endorsements often lead to increased product exposure and heightened purchase intent. Nonetheless, ethical issues emerge about the veracity of sponsored material and the possibility of deceptive marketing. Understanding these factors is essential for businesses seeking to improve their social media strategy while maintaining client trust in a more digital economy.

3.1 The Influence of Social Media Content on Consumer Purchase Decisions

Social media's rise has drastically changed consumer behavior, therefore altering how people interact with businesses, search for information, and decide what to buy. Traditionally, consumer decisions were largely influenced by one-way communication through conventional media channels such as television, radio, and newspapers (Kotler & Keller, 2016). Furthermore, very important for customer trust and decision-making were word-of-mouth referrals from intimate social circles.

However, the advent of digital platforms has introduced a bidirectional communication model, allowing consumers to interact with brands in real time, share experiences, and impact on one another's choices instantaneously (Stephen, 2016). Users of social media may give reviews, compare items, and ask questions of both peers and online influencers there, which has evolved into a center for interaction.

The growth of social media has converted it from a rudimentary entertainment platform into a significant influencer on customer trust, brand image, and purchase behavior. This transition has resulted in the development of novel marketing dynamics that use user participation and data-driven insights. Significant advancements encompass:

- **User-Generated Content (UGC):** Consumers create and share content about brands, shaping their perception and trust levels (Godey et al., 2016).
- **Influencer Marketing:** Influencers leverage their audience trust to promote products, significantly impacting purchasing behavior (Vinerean et al., 2013).
- **Targeted Social Media Advertising:** Brands utilize data-driven algorithms to deliver personalized advertisements based on user preferences and online behavior (Heinonen, 2011).
- **Consumer Insights through Data Analytics:** Social media provides valuable insights into consumer sentiment, trends, and behavior patterns, allowing businesses to optimize marketing strategies (Abbott, 2014).

The shift from passive consumption to active digital participation has transformed social media into a significant influence on consumer purchase behavior. Recent statistics indicate that 76.7% of the Czech population are active social media users, while 82.7% of internet users routinely connect with social platforms, underscoring the importance of social media as a principal source of consumer influence (DataReportal, 2024).

The progression of social media has profoundly influenced consumer behavior in Czechia over the past decade. Social media usage has risen significantly from 37.3% in 2012 to 76.7% in 2024, reflecting the increasing role of digital platforms in consumer engagement and business interactions (DataReportal, 2024). This surge in online activity has coincided with a substantial escalation in e-commerce expenditure, increasing from 50 billion CZK in 2012 to 250 billion CZK in 2024 (Statista, 2024).

The parallel growth of social media usage and online shopping underscores the role of digital platforms as key drivers of purchasing decisions, consumer trust, and brand engagement. Businesses in Czechia have leveraged social media marketing, influencer collaborations, and user-generated content to enhance customer interactions, reinforcing the transition towards digital commerce (GO-Globe, 2024).

The table depict the progression of social media's impact on consumer behavior in the Czech Republic from 2012 to 2024. The report underscores the consistent rise in social media use (%) and its association with the expansion of e-commerce expenditure (CZK billion).

Year	Social Media Usage (%)	E-commerce Spending (CZK Billion)
2012	37.3	50
2014	45	80
2016	55	110
2018	65	140
2020	72	196
2022	75	230
2024	76.7	250

Table 1: Digital Growth and Consumer Transformation in Czechia: 2012–2024-Source: DataReportal, 2024

In 2012, over 37.3% of the Czech population was actively engaging with social media platforms. Social media use increased significantly, reaching 45% in 2014 and 65% by 2018. By 2024, about 76.7% of the population will be active on social media (DataReportal, 2024). The swift embrace of digital platforms signifies a transformation in consumer behavior, as people increasingly depend on online interactions to identify items and make purchase choices.

Likewise, e-commerce expenditures in the Czech Republic have seen a similar rising trajectory. In 2012, internet shopping expenditures were 50 billion CZK, rising to 140 billion CZK by 2018 (Czech Journal of Economics, 2020). The most substantial rise occurred between 2020 and 2024, with e-commerce expenditure increasing from 196 billion CZK to an anticipated 250 billion CZK (Go-Globe, 2024). This rise corresponds with the proliferation of digital marketing tactics, such as social media advertising, influencer marketing, and customized promotions, which have enhanced customer participation and trust in online buying.

The graph illustrates the robust correlation between social media penetration and e-commerce growth. The concurrent rising trajectory of both metrics indicates that social media platforms significantly influence consumer buying behavior, as heightened engagement with digital content correlates with greater online expenditure. The data highlights the increasing importance of social media in the Czech market, as firms use digital platforms to impact customer choices and enhance e-commerce sales.

This transition underscores the significance of social media techniques in contemporary marketing, as firms that proficiently interact with customers online may enhance sales and foster customer loyalty (Eurostat, 2024).

3.1.1 The Shift from Passive Consumption to Active engagement

In the early 2000s, platforms like as Facebook, YouTube, and Twitter emerged, revolutionizing brand engagement with consumers (Kaplan & Haenlein, 2010). In addition to merely receiving marketing messages, these platforms enable customers to express ideas, provide feedback, and participate in brand discussions.

Subsequent platforms such as Instagram, Pinterest, and Snapchat introduced visual storytelling as social commerce evolved, enhancing consumer-brand engagement. Influencer marketing, which involves corporations collaborating with individuals possessing a strong online reputation to endorse their products, emerged from the heightened significance of peer influence and social proof in digital marketing (De Veirman, Cauberghe, & Hudders, 2017).

Influencer endorsements are more persuasive in shaping consumer choices than traditional advertising, as they typically appear authentic and pertinent. Research indicates that peer recommendations on social media significantly enhance purchase intent and brand trust in contrast to traditional advertisements (Cheung et al., 2017).

3.1.2 User-Generated Content (UGC) & the Role of Social proof

User-generated content (UGC) has gained significant importance in the digital economy regarding consumer engagement and purchasing behavior. User-generated content (UGC) encompasses any material, including blog posts, videos, reviews, social media updates, or testimonials, produced by consumers rather than corporations. Consumers today are not just passive recipients of marketing messages but active participants in brand conversations. Managerial approaches must evolve to accommodate changing social media dynamics where consumers dictate the narrative, influencing brand image through discussions, reviews, and online advocacy (Heinonen, 2011).

User-generated content (UGC) is perceived as authentic, credible, and pertinent, in contrast to traditional advertising, which is selectively curated by companies to convey promotional messages rather than genuine customer experiences. The proliferation of social media platforms and online review sites such as Instagram, TikTok, YouTube, Trustpilot, and Yelp has enabled consumers to articulate their opinions widely, therefore influencing the purchasing decisions of others (Cheung et al., 2017).

Social Proof Theory elucidates the growing impact of User-Generated Content (UGC) by suggesting that individuals, particularly in situations devoid of direct experience, rely on the behaviors and opinions of others when forming evaluations (Cialdini, 2008).

In a digital context, individuals are more predisposed to trust products and services that have received favorable reviews from other users rather than relying just on company advertising.

Research indicates that user-generated content fosters a sense of community and shared experience among consumers, resulting in superior engagement, trust, and conversion rates compared to traditional brand advertising (Erkan & Evans, 2016). Online reviews, social media testimonials, and influencer endorsements distinctly illustrate how consumer behavior is influenced. Study found that 78% of consumers trust UGC, such as online reviews and influencer endorsements, compared to only 42% who trust traditional advertising. This 36% trust gap highlights the growing preference for peer-generated content over brand messaging (Statista, 2024). Furthermore, another study shows that 74% of European customers consider UGC more trustworthy than branded content.

The success of UGC can largely be ascribed to its connection with Electronic Word-of-Mouth (eWOM) Theory, which elucidates how digital platforms enhance the visibility and impact of consumer recommendations compared to traditional word-of-mouth communication (Cheung & Thadani, 2012). The accessibility and rapid transmission of internet reviews and user testimonials further bolster social proof, enabling consumers to evaluate the quality and reliability of a product based on the experiences of numerous other users.

Moreover, particularly among younger consumers who are highly engaged on platforms like as TikTok and Instagram, the Fear of Missing Out (FOMO) is crucial for the persuasive impact of User-Generated Content. Trending products and viral social media content create a sense of urgency and prompt rapid purchasing decisions to avoid missing out on exclusive deals or sought-after experiences. Due of its influence, organizations have increasingly used user-generated content into their marketing strategies to enhance consumer trust and brand reputation. Numerous enterprises vigorously encourage users to generate and disseminate content through incentives such as discounts, gifts, or social media recognition. Influencer marketing, wherein firms collaborate with individuals possessing a strong online reputation and a loyal following, is one of the most effective applications of user-generated content (UGC).

While celebrities and macro-influencers have traditionally been used for sponsorships, firms are increasingly opting for micro-influencers—social media users with modest yet highly engaged audiences—perceived as more authentic and pertinent (De Veirman, Cauberghe, & Hudders, 2017). This alteration reflects customer preferences for authentic, experience-oriented content rather than refined corporate messaging.

User-generated content offers several advantages; nonetheless, its implementation presents certain ethical dilemmas and challenges. A significant issue is the rise of deceptive endorsements and fraudulent reviews as firms attempt to generate favorable user-generated content through sponsored testimonials or artificial engagement (Bianchi, A 2020). Research indicates that consumers quickly lose trust in organizations revealed to employ unethical marketing practices, potentially resulting in enduring reputational harm (Filieri et al., 2015).

Furthermore, regulatory bodies such as the European Commission and the U.S. Federal Trade Commission (FTC) have implemented regulations to ensure transparency in influencer marketing and require the accurate disclosure of sponsored content (Evans, Phua, Lim, & Jun, 2017).

The increasing significance of artificial intelligence (AI) and machine learning in user-generated content (UGC) management introduces an additional challenge. While artificial intelligence-driven systems can help detect fraudulent reviews, filter spam, and manage inappropriate material, concerns around algorithmic bias and the unintended suppression of legitimate consumer opinions arise. Excessive dependence on automated moderation methods may inadvertently suppress genuine yet negative evaluations, so undermining the integrity of user-generated content as a reliable data source. Achieving equilibrium between utilizing user-generated content for commercial purposes and ensuring ethical standards and transparency is a primary objective as organizations navigate these challenges.

The rise of user-generated content provides a peer-driven alternative to traditional advertising, therefore transforming consumer decision-making. As social media evolves, the methods by which users create, disseminate, and engage with user-generated content will also transform, therefore reinforcing its dominance in digital marketing and enhancing consumer trust. Future research should examine how emerging technologies—such as blockchain for verifying review authenticity, artificial intelligence (AI) for sentiment analysis, and augmented reality (AR) for immersive product experiences—will further transform the landscape of user-generated content (UGC) and its influence on consumer behavior (Rapid Innovation, 2024). Optimizing the potential of user-generated content as a reliable and transparent medium for brand interaction hinges on companies adhering to ethical marketing practices and safeguarding consumer trust.

3.2 The Effectiveness of Social Media Analytics in Predicting Consumer behavior

Growing dependency on social media for customer interaction has generated plenty of digital data. Social media analytics are today's tool for organizations seeking understanding of consumer preferences, industry trends, and behavior patterns. Social media analytics provides instantaneous insights into customer attitudes, interactions, and purchase habits unlike traditional market research dependent on polls and prior sales data (Abbott, 2014). Improving marketing strategies, optimizing involvement, and precisely predicting consumer behavior depend on these realizations.

Predictive modeling is a necessary ability of social media analytics. By evaluating sentiment, engagement indicators, and transactional data, advanced artificial intelligence systems help companies forecast changes in consumer demand and purchasing behavior. Monitoring social media conversation, keyword trends, and online activities can help companies to improve focused marketing plans and personalize customer experiences.

Evaluation of brand image, customer interaction, and the effectiveness of marketing efforts makes great use of social media data. By means of thorough information analysis, businesses may classify consumers based on behavioral traits, therefore enabling future purchase behavior and suitable modification of their marketing plans. Machine learning techniques provide unique content recommendations, automated trend recognition, and consumer sentiment research.

Moreover, social media behavioral analytics helps companies to spot early indicators of new market trends. Real-time engagement data used by platforms as Facebook, Instagram, and TikHub help businesses to spot popular products, therefore guiding their decisions on inventory control and new introductions.

Social media analytics raises questions about data security and algorithmic transparency even if it increases marketing precision. The great usage of consumer monitoring and behavioral profiling raises questions about the ethicality of data gathering and the degree to which algorithmic biases influence user decisions. Companies have to be able to effectively handle these issues to ensure that predictive analytics are used transparently and morally. Over 78% of global customers assert that corporations should confront AI bias, reflecting a significant need for openness and equity in algorithmic procedures. (Agility PR Solutions, 2024).

Empirical studies show unequivocally how social media exposure influences consumer buying behavior. Studies show that most consumers refer to social media before making a purchase, therefore stressing the influence of the platform on brand perception and buying intention. Data-driven insights drawn from social media interaction patterns significantly affect product adoption rates in sectors like fashion, technology, and beauty. Organizations that successfully use predictive social media analytics into their plan will have a competitive advantage as digital marketing develops. Modern consumer behavior research depends on social media analytics as a necessary instrument as it can evaluate client attitude, project future trends, and modify marketing plans.

3.2.1 Key Components and Applications of Social Media Analytics in Consumer Behavior prediction

Social media analytics has become an essential tool for businesses seeking to understand and predict consumer behavior. By leveraging data collection, sentiment analysis, predictive modeling, and machine learning algorithms, organizations can extract meaningful insights from vast amounts of digital interactions. These insights enable firms to refine marketing strategies, optimize customer engagement, and anticipate purchasing trends with greater precision. Data collection plays a crucial role in social media analytics, as it involves gathering information from various channels such as posts, comments, likes, shares, and user reviews. By compiling this data, organizations gain a holistic view of consumer behavior, enabling them to measure engagement, recognize influential users, and spot developing trends in the market. Sentiment analysis plays a crucial role in evaluating public perception of brands and products.

By analyzing text, images, and emojis, companies can determine whether consumer sentiment is positive, neutral, or negative. This capability is particularly useful in crisis management, enabling businesses to proactively address customer grievances and mitigate potential reputational risks. Additionally, understanding the emotional drivers behind consumer behavior allows organizations to craft more effective and personalized marketing messages.

A critical predictive capability within social media analytics is forecasting future consumer behavior based on historical data. Predictive analytics identifies recurring patterns, assisting firms in resource allocation and marketing strategy optimization. Platforms such as Amazon and Netflix effectively utilize predictive analytics to deliver personalized recommendations, thereby improving customer retention and conversion rates. Machine learning algorithms further enhance the predictive power of social media analytics by analyzing vast amounts of unstructured data.

These models continuously refine their predictions, identifying correlations between user engagement and purchasing behaviors. Businesses leverage these insights to create highly targeted advertising campaigns, ensuring that the right audience receives relevant content at the most opportune moments. Real-time analytics enable firms to adapt their marketing strategies dynamically, maintaining a competitive edge in rapidly evolving industries.

3.2.2 Challenges of Social Media analytics

Through analysis of vast volumes of internet data, social media analytics enables companies to grasp and forecast customer behavior. Tracking trends, consumer preferences, and product demand, companies use artificial intelligence and machine learning. Retailers, for instance, may predict which items will be hot, and service providers can evaluate client happiness by means of reviews and comments.

Targeted marketing—where companies employ artificial intelligence to classify consumers according on interests and behavior—is a main usage of social media analytics. This enables them to deliver tailored, more relevant adverts, hence enhancing engagement and revenues.

Social media analytics poses several issues, especially regarding data protection. As enterprises gather and scrutinize user data, ethical issues about permission and transparency emerge. A notable concern is algorithmic bias; if AI models are educated on prejudiced data, they may unjustly benefit certain groups, resulting in discriminatory consequences. Companies that use ethical and responsible AI processes may effectively address these difficulties, improve customer experience, anticipate market trends, and secure a competitive advantage. The future of social media analytics will hinge on the equilibrium between technological progress and user trust, security, and equity.

3.3 The Impact of Social Media Marketing on Brand Loyalty

The integration of social media marketing with AI-driven analytics has revolutionized brand loyalty strategies, enabling companies to personalize interactions, predict consumer behavior, and optimize engagement in real time. Machine learning models analyze user behavior, sentiment, and engagement patterns, allowing businesses to deliver hyper-personalized content that strengthens brand affinity (Abbott, 2014).

3.3.1 AI-Powered Engagement and Trust

By offering quick replies and customized product recommendations, AI-driven recommendation systems, chatbots, and sentiment analysis improve customer contacts. Faster problem resolution and proactive customer care help companies using AI-driven engagement strategies to retain customers at a 30% rise (Statista, 2024).

3.3.2 Cybersecurity and Consumer Trust in Brand Loyalty

Since data-driven marketing is becoming more and more important, brand loyalty depends much on cybersecurity. Customers are more willing to interact with companies showing excellent ethical artificial intelligence policies and data security. Studies reveal that 75% of customers will cease buying from a business after a data breach, therefore underscoring the direct connection between cybersecurity and consumer confidence (Cybersecurity Index, 2023).

Businesses using GDPR-compliant data processing, encryption, and AI-driven fraud detection have more confidence and loyalty (European Commission, 2023).

3.4 Consumer engagement across different social media platforms

Social media users promote engagement through likes, shares, comments, and group discussions. The platform's algorithm prioritizes large interactions, rendering brand-consumer engagements and user-generated content very influential. Businesses using Facebook for community engagement, targeted advertising, and the dissemination of extensive information. Businesses now engage just 5.2% of their followers each post on average (Hootsuite, 2023), indicating a significant decline in organic reach over the years, hence highlighting the necessity of sponsored events to improve exposure.

Instagram, boasting over 2 billion active users—90% of whom follow at least one corporate account—thrives on visual content, with Stories, reels, and influencer partnerships driving significant engagement. According to a Social Media Examiner (2023) research, influencer marketing on Instagram has an average engagement rate of 2.2%, far surpassing that of brand-generated content. Instagram Reels currently account for 30% of the time spent on the platform, significantly transforming interaction with short-form video content.

TikTok has rapidly emerged as the most captivating social media platform, with over 1.6 billion active users and an average daily usage time of 95 minutes per individual (Business of Apps, 2023). The site's algorithm prioritizes viral, short-form films, so allowing advertisers to organically engage extensive audiences. Research indicates that 67% of TikTok users assert that the platform incentivizes purchases even when they had no prior intention to buy, making it an essential medium for engagement-driven e-commerce. Hashtag challenges and user-generated content efforts are particularly successful in driving engagement.

Now rebranded as X, Twitter functions as a real-time discussion platform with more than 556 million active users (Statista, 2023). Brands on Twitter mostly participate in discourse through hashtags, polls, and interactive threads. Nonetheless, at just 0.05% (Rival IQ, 2023), the average engagement rate per post remains inferior to that of comparable networks. Despite this, 53% of Twitter users follow firms for updates, news, and customer service contacts, thereby offering an effective platform for real-time engagement and brand positioning.

LinkedIn, with over 900 million members, prioritizes industry-specific interaction through thought leadership content, corporate updates, and networking events. The typical post engagement rate on LinkedIn is 0.35% , indicating that interaction on this platform might be lower than on other networks. However, 80% of LinkedIn users engage with industry-specific content (Sprout Social, 2022), making this platform the preferred choice for business and professional networking.

3.5 The Impact of Social Media Trends on Consumer Behavior

Social media trends affect consumer behavior by means of exposure, social pressure, and FOMO—the fear of missing out. A trend gains urgency and attractiveness when it gathers momentum, therefore motivating individuals to look into and own popular items. According to research, 72% of consumers say their purchase decisions are influenced by social media trends (HubSpot, 2023). Platforms such as Twitter, Instagram, and TikTok act as accelerators helping viral goods and services to be quickly embraced.

User-generated content (UGC) is a highly powerful tool wherein consumers participate in trends by distributing their experiences, evaluations, or original work on a product. Product sales have surged significantly as a consequence of campaigns such #TikTokMadeMeBuyIt; TikTok indicates that 67% of users are driven to spend after popular content on the site. Especially in industries like fashion, cosmetics, and technology, Instagram Reels and YouTube Shorts help consumers find products.

Influencer-driven trends, in which social media personalities offer new products, lifestyle choices, or emerging society shifts to change consumer views, are absolutely essential. With 49% of consumers depending on influencer recommendations before making a purchase, influencer endorsements greatly impact buying intent. Particularly micro-influencers provide higher engagement rates and trustworthiness because of their supposed sincerity and close audience contacts.

Advancements in social media provide challenges, particularly about sustainability and ethical marketing, although providing opportunities for brand recognition. The ephemeral nature of viral trends can result in overconsumption and impulsive buying behavior, thereby causing problems related to waste and environmental impact. While fast fashion enterprises benefit from viral trends, they simultaneously endorse unsustainable spending habits and excessive manufacturing. Research highlights the necessity for organizations to promote mindful buying, since more than 60% of consumers express remorse about impulsive purchases influenced by social media trends (Statista, 2023).

The reliability and veracity of trending content present an additional challenge. Some trends emerge organically, while others are artificially created through sponsored events and influencer partnerships. Sponsored content is more significant to customers; therefore, transparency in marketing is essential to maintain trust. Regulations, particularly those from the Federal Trade Commission (FTC), require influencers to disclose sponsored affiliations to ensure ethical marketing practices. Companies that fail to maintain transparency risk losing customer confidence.

Contemporary consumer behavior is influenced by social media trends, which can serve as a valuable tool for firms to enhance connection and expand sales. Brands must, however, achieve equilibrium between trend-driven marketing and ethical responsibility to ensure their policies enhance customer confidence, sustainability, and transparency. Companies that adeptly adapt to social media trends will strengthen their connections with consumers and foster enduring brand loyalty as digital culture evolves.

3.6 The Role of Social Media in Crisis Management and Consumer perceptions

Crisis management encompasses the strategies and actions employed by organizations to identify, respond to, and mitigate hazards that threaten their operations, client relations, or reputation. The ability of social media to rapidly distribute information and promptly alter public opinion has become it an essential element of crisis management in the digital era. Social media reveals crises; therefore, while it provides organizations a platform for direct engagement with stakeholders, prompt, transparent, and astute measures are essential to maintain customer trust. Digital marketing strategies must focus on engagement-driven content as opposed to traditional push advertising. Studies suggest that interactive and visually stimulating posts lead to stronger consumer retention and brand recall (Stephen, 2016).

The detection and reaction to crises are enhanced by advancements in technology, particularly in artificial intelligence (AI) and cybersecurity. Artificial intelligence-driven sentiment analysis solutions allow firms to monitor consumer interactions, detect issues promptly, and formulate customized responses based on real-time data. Companies must also address cybersecurity challenges such as data breaches, misinformation campaigns, and hacking incidents to maintain brand integrity and customer trust

3.6.1 Crisis Management responses

In crisis management, social media both offers a tool and a danger. Negative news, false information, and consumer complaints can, on the one hand, become viral and aggravate public reaction. Conversely, companies may use social media deliberately to correct false information, answer customer questions, and show responsibility. According to studies, 76%

of customers want brands to react to crises on social media within 24 hours, and 63% feel businesses should take accountability and offer transparent information (Edelman Trust Barometer, 2023).

3.6.2 Social Media, Consumer Perception, and Cybersecurity Risks

A brand's response, together with the speed and authenticity of its message, shapes customer perception during crises. Research indicates that whereas 60% of customers would discontinue patronage of a business that mishandles a crisis, 47% exhibit increased loyalty towards companies that manage crises transparently (HubSpot, 2023). Social media enables businesses to craft narratives, although it simultaneously exposes them to cybersecurity threats that may exacerbate issues.

Data breaches and cyberattacks represent some of the most serious calamities for organizations, since they infringe upon consumer privacy and undermine trust. IBM Security (2023) reports that the average cost of a data breach is $4.45 million, resulting in 52% of customers losing trust in organizations that fail to protect their data.

Cybersecurity measures are essential for crisis prevention and response, as hackers often use vulnerabilities in social media to spread misinformation, impersonate businesses, or initiate phishing schemes.

Deepfake videos and automated bots, among other AI-generated misinformation, exacerbate crisis management challenges. Inaccurate narratives disseminate rapidly and influence public image before companies issue apologies. Companies must employ AI-driven content verification tools, cybersecurity systems, and rapid response teams to combat misinformation and maintain credibility through these technologies.

3.6.3 Best Practices for Crisis Management on Social Media

Crisis management is fundamentally reliant on social media, which enables organizations to respond swiftly, manage public perception, and maintain consumer confidence. A poorly managed crisis can inflict enduring damage on one's reputation. Organizations aiming to effectively navigate crises must adhere to a systematic plan that prioritizes digital security, transparency, and engagement.

A comprehensive crisis management strategy comprises the following essential components:

- **Real-time sentiment monitoring:** AI-driven analytics allow organizations to detect potential crises early by analyzing social media mentions, trends, and public sentiment before issues escalate.
- **Clear and proactive communication** –Companies are required to engage in clear, proactive communication to swiftly identify issues, provide accurate updates, and offer transparent solutions. The Edelman Trust Barometer 2023 indicates that 76% of people want companies to respond within 24 hours.
- **Direct consumer engagement-** Proactively addressing inquiries, correcting misinformation, and resolving complaints contribute to narrative construction and mitigate reputational damage (Sprinklr, 2023). Active consumer engagement Effective companies may reduce negative sentiment by as much as 35% within 48 hours (Sprout Social, 2023).
- **AI-powered automation and chatbots** –AI-driven automation and chatbots are automated systems that enable firms to efficiently address a substantial volume of consumer inquiries while maintaining a consistent and reassuring brand voice.
- **Mitigating misinformation and cybersecurity threats**: Digital crises such as data breaches and disinformation may undermine brand trust. Employing AI-driven content validation and cybersecurity protocols aids in mitigating misinformation and protecting reputational integrity.

An effectively implemented crisis management plan enables firms to mitigate damage, restore trust, and enhance brand resilience against digital threats. Organizations that integrate strategic communication, cybersecurity, and artificial intelligence into their crisis response plans will be better equipped to navigate the intricacies of modern crisis management (Hootsuite, 2023).

4. Practical Part

The practical component examines the impact of social media on consumer behavior via a mixed-methods approach that integrates quantitative and qualitative research. This research has three fundamental components: a systematic online questionnaire, statistical evaluation of the survey results, and comprehensive interviews. These approaches guarantee a thorough evaluation of consumer interaction with social media, the effects of diverse marketing tactics, and confidence in AI-driven commercials and influencer marketing.

The quantitative element comprises an online survey aimed at social media users in Prague. This poll assesses critical consumer behavior characteristics, including the frequency of social media use, engagement with brand commercials, confidence in influencer endorsements, and apprehensions about data privacy. The dataset has 161 replies, which will undergo statistical analysis to discern consumer behavior patterns and connections. Regression analysis and hypothesis testing will be performed using Python to investigate the correlation between social media exposure and purchase choices.

The qualitative component comprises comprehensive interviews with meticulously chosen people. These interviews will provide profound insights into consumer motives, confidence in digital content, and perspectives on AI-driven advertising. The data obtained from the interviews will undergo thematic analysis to augment and refine the quantitative results.

This study seeks to provide a comprehensive knowledge of how social media impacts consumer decision-making by the integration of statistical analysis, regression models, and qualitative theme assessment. The results will provide significant insights into the influence of social media marketing, customer trust, and AI-driven advertising on purchase behavior.

4.1 Quantitative Method

The quantitative study is based on an online survey administered to 161 social media users located in Prague. The poll aimed to elucidate critical aspects of consumer behavior shaped by social media, concentrating on marketing exposure, user trust, and purchase behavior.

The dataset includes both independent (explanatory) and dependent factors. The following is a summary of the variables used in the statistical analysis:

Based on the survey findings, the data renamed in a sensible manner to ensure readability in SAS Studio. Survey findings were stored and executed in SAS Studio, with the data shown using PROC PRINT.

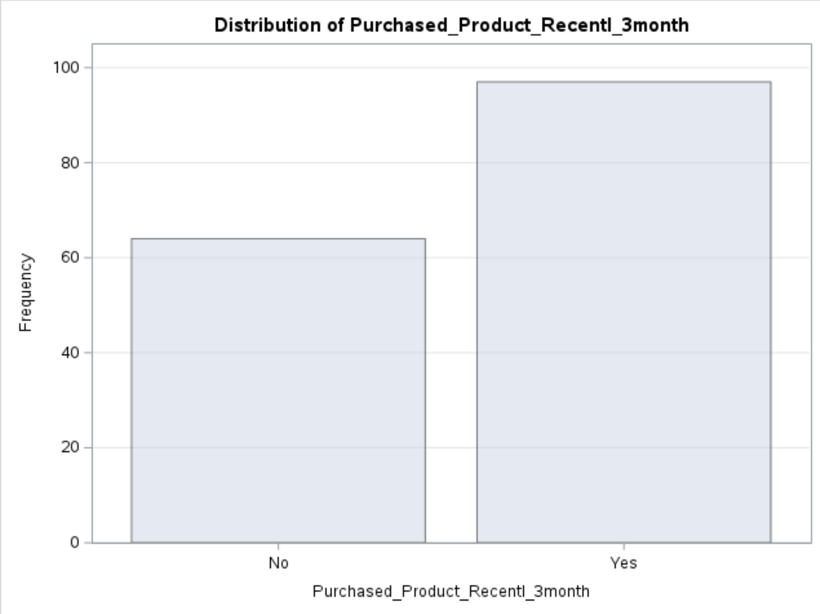
Comfort_with_Targeted_Ads	Binary	Yes/No – whether the user feels comfortable with targeted ads
Followed_Brand_Ad	Binary	Yes/No – followed a brand after an ad
SocialMedia_Influence_Shopping	Ordinal	Scale (less likely(1)-more likely(5)) – degree to which social media content influences shopping
Trust_Building_Marketing	Ordinal	Scale (Likert) (less likely (1)-more likely(5)) – trust in social media marketing
Ad_Personalization_Perception	Ordinal	Scale – perception of ad personalization (positive/neutral/negative)
Ad_Promotion_Engagement	Ordinal	Rarely/Sometimes/Often – frequency of engagement with ads
Age_Group	Categorical	Age ranges – e.g., 18–24, 25–34, etc.
Boycotted_Brand_SocialMedia	Binary	Yes/No – if they boycotted a brand based on SM exposure
Brand_Content_Engagement	Categorical	Like, Share, Comment, etc. – type of brand interaction
Brand_Content_Interaction	Ordinal	Frequency – Rarely, Sometimes, Often
Crisis_Response_Impact	Binary	Yes/No – whether crisis handling impacts trust
Daily_Social_Media_Usage	Ordinal	<1hr, 1–3hrs, >3hrs – amount of daily social media usage
Employment_Status	Categorical	Student, Employed, Unemployed, etc.
Gender	Categorical	Male, Female, Other
Highest_Education	Categorical	College, Bachelor's, Master's, etc.
Importance_of_Reviews	Ordinal	1(low)–5 (high)Likert scale – importance of reviews

Influential_Content_Type	Categorical	e.g., Videos, Reviews, Testimonials, etc.
Influential_Platform_Purchase	Categorical	TikTok, Instagram, Facebook, etc.
Likelihood_Purchase_UserContent	Ordinal	1–5 Likert scale – likelihood of purchase based on UGC
Loyalty_Brand_Engagement	Binary	Yes/No – brand loyalty through engagement
Most_Used_Platform	Categorical	TikTok, Facebook, Instagram, etc.
Privacy_Concerns_Trust	Binary	Yes/No – if privacy concerns impact trust
Privacy_Settings_Ad_Tracking	Binary	Yes/No – whether privacy settings were changed due to ad tracking
Purchased_Influencer_Recommendation	Binary	Yes/No – purchase due to influencer recommendation
Purchased_Product_Recentl_3months	Binary	Yes/No – purchase in the last 3 months
Relevance_of_Recommendations	Ordinal	1–5 Likert – how relevant user finds recommendations
SM_algorithm_influence	Ordinal	1–5 Likert – perceived influence of algorithms
Trends_Influence_Purchase	Binary	Yes/No – trends' influence on purchase decisions

Table 2: Survey questions and categorization-Source: own Survey results 2025

Descriptive analyses were performed on all major categorical and binary variables to summarize the sample's characteristics and provide the groundwork for later inferential tests. These included factors such as social media impact, recent purchasing behavior, brand involvement, advertising interactions, and platform use.

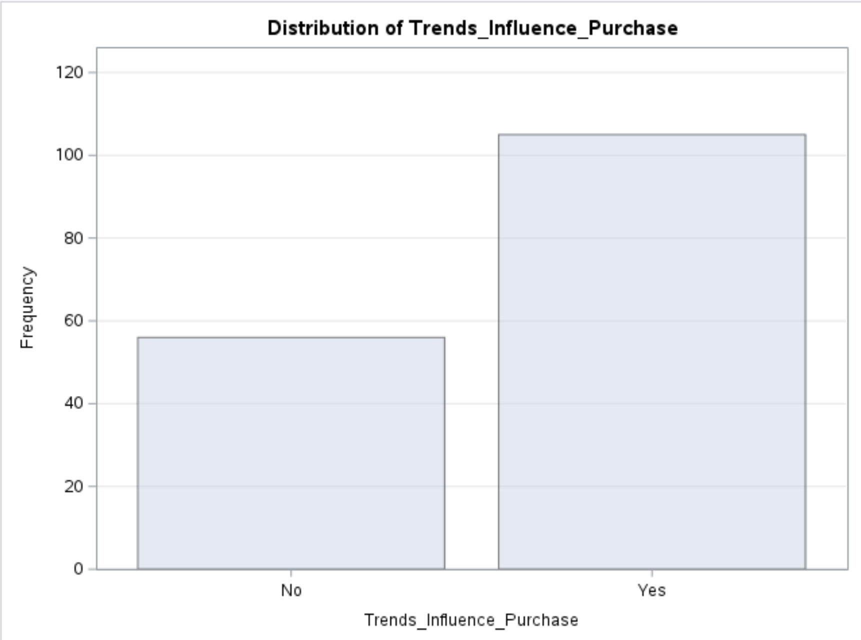
Purchased_Product_Recentl_3month	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	64	39.75	64	39.75
Yes	97	60.25	161	100.00



Picture 1: Distribution of purchase within last 3 month- Source:SAS studio

A majority (60.2%) reported purchasing a product in the last 3 months due to social media influence

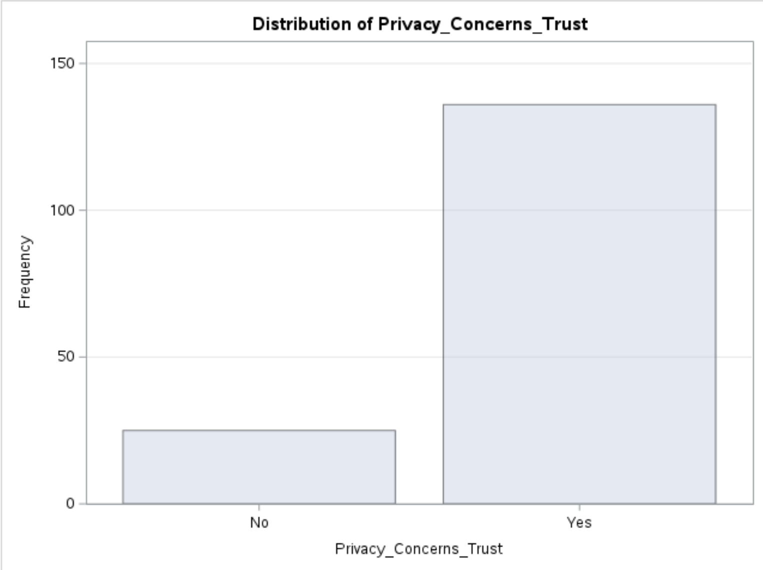
Trends_Influence_Purchase	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	56	34.78	56	34.78
Yes	105	65.22	161	100.00



Picture 2: Distribution of trend influence purchase- Source:SAS studio

As per Picture 2, Trend Impact shows that 65.2% were influenced by social media trends in their purchasing choices.

Privacy_Concerns_Trust	Frequency	Percent	Cumulative Frequency	Cumulative Percent
No	25	15.53	25	15.53
Yes	136	84.47	161	100.00



Picture 3: privacy concern among participants of survey- Source: SAS studio

Privacy Concerns indicates a significant 84.5% expressed concerns about privacy and trust online

4.1.1 Evaluation of social media content on consumer purchase decisions

All essential assumptions necessary for the Chi-square test were thoroughly satisfied in this study. Initially, both variables employed—SM_Influence_Grouped (classified as Low, Medium, and High) and Purchased_Product_Recently_3month (Yes/No)—are unequivocally categorical, satisfying the first criterion.

The survey methodology guaranteed the independence of each observation, since each respondent provided just a single answer to the dataset. The assumption about predicted frequency was fulfilled, since all anticipated cell counts exceeded 5, as seen in the SAS output table. This verifies that the test retains its statistical validity. The sample size of 161 exceeds the minimal threshold necessary for a Chi-square test. Ultimately, the two variables included in the test were mutually exclusive and exhaustive, indicating that each participant was assigned to precisely one group for each variable, without any overlap or omission. Consequently, the findings of the Chi-square test may be regarded as both valid and trustworthy.

Following to that, hypothesis set as per below:

- H_0: There is no significant relationship between the influence of social media on shopping decisions and whether a consumer has made a recent purchase.
- H_1:There is a significant relationship between the influence of social media on shopping decisions and whether a consumer has made a recent purchase.

The SAS code generates a new grouped variable named SM_Influence_Grouped. The original Likert-scale variable SocialMedia_Influence_Shopping has been categorized into three groups: Low (values 1 and 2), Medium (value 3), and High (values 4 and 5). The used name literals ('column-name'n) to accurately reference the original variable that had a leading space. This change guarantees proper category sizes for Chi-square analysis and facilitates the evaluation of social media impact levels.

```
proc print data=work.import;
run;

data grouped;
    set work.import;

    length SM_Influence_Grouped $6;

    /* Reference the column with leading space using a name literal */
    if ' SocialMedia_Influence_Shopping'n in (1, 2) then SM_Influence_Grouped = 'Low';
    else if ' SocialMedia_Influence_Shopping'n = 3 then SM_Influence_Grouped = 'Medium';
    else if ' SocialMedia_Influence_Shopping'n in (4, 5) then SM_Influence_Grouped = 'High';
run;

proc print data=work.grouped;
run;
```

Picture 4: Grouping Variable SocialMedia_Influence_Shopping -Source: SAS studio

A Chi-square test of independence was conducted to examine the relationship between perceived levels of social media influence (Low, Medium, High) and recent purchasing behavior (whether the respondent purchased a product in the last three months). The result was statistically significant, $\chi^2(2, N = 161)$ = p45, $p < .0001$, indicating a meaningful association between the two variables. The Likelihood Ratio Chi-square test ($\chi^2 = 41.2657$, $p < .0001$) further confirmed this significance. Additionally, the Mantel-Haenszel Chi-square value of 17.8068 ($p < .0001$) suggested a strong linear trend — as the level of perceived social media influence increased, so did the likelihood of making a recent purchase. All assumptions for the Chi-square test were fully met, including sufficient sample size, independence of observations, categorical data, and expected frequencies greater than five in all cells.

These findings confirm that social media content has a statistically significant impact on consumer purchase decisions. The rejection of the null hypothesis (H_0) constitutes the result of the test, leading to the acceptance of the alternative hypothesis (H_1).

Statistics for Table of SM_Influence_Grouped by Purchased_Product_Recentl_3month

Statistic	DF	Value	Prob
Chi-Square	2	37.2445	<.0001
Likelihood Ratio Chi-Square	2	41.2657	<.0001
Mantel-Haenszel Chi-Square	1	17.8068	<.0001
Phi Coefficient		0.4810	
Contingency Coefficient		0.4334	
Cramer's V		0.4810	

Sample Size = 161

Table 3:P_value results- Source: SAS studio

4.1.2 Social media analytics effect on predicting consumer behavior

The study used Binary Logistic Regression to evaluate the efficacy of social media analytics in forecasting consumer purchase behavior. Our objective was to identify the elements associated with social media engagement and customization that substantially affect a consumer's propensity to purchase a product influenced by social media content. Given that the dependent variable (Purchased_Product_Recently_3months) is binary (Yes/No), we choose Binary Logistic Regression as the suitable statistical method. Logistic regression estimates the likelihood of an outcome (buying) based on many predictor factors.

This approach was chosen from SAS Studio under:

- Tasks and Utilities → Predictive Regression Models → Binary Logistic Regression.

Binary Logistic Regression was chosen due to the following reasons:

1. It estimates the probability of an event occurring (purchase).
2. It quantifies the impact of each independent variable on the dependent variable using odds ratios.

In the logistic regression study, we concentrated on a subset of binary predictor variables that are most relevant to social media analytics and consumer behavior forecasting. The decision was grounded upon theoretical significance and practical use in marketing analytics.

The dependent variable in this research was the respondent's purchase of a product influenced by social media material during the preceding three months. This variable functioned as the result we sought to anticipate.

We picked six binary explanatory factors as independent classification predictors as per table 5. Logistic Regression Model Setup as per below:

Dependent Variable (Response):

Purchased_Product_Recentl_3month

- Binary outcome indicating whether the respondent made a purchase influenced by social media within the last three months.
- Coded as: *Yes* = 1 (event of interest), *No* = 0.

Explanatory Variables:

Comfort_with_Targeted_Ads

Whether the respondent is comfortable with social media platforms using data for targeted advertising.

Privacy_Settings_Ad_Tracking

- Indicates if the user has changed privacy settings to restrict ad tracking.

Loyalty_Brand_Engagement

- Reflects consumer loyalty based on social media engagement with brands.

Purchased_Influencer_Recommendat

- Whether the respondent has purchased a product based on influencer recommendations.

Trends_Influence_Purchase

- Whether the respondent is influenced by social media trends in their purchasing decisions.

Followed_Brand_Ad

- Indicates if the respondent followed a brand after seeing a social media advertisement.

The logistic regression model calculated the likelihood of a purchase decision, with "Yes" denoting a successful purchase and "No" signifying no purchase. The evaluation of model performance used many statistical metrics, including the Akaike Information

Criterion (AIC), log-likelihood values, and the concordance %. The importance of each predictor was evaluated using the Wald Test and the Likelihood Ratio Test.

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	218.381	191.184
SC	221.463	212.754
-2 Log L	216.381	177.184

Table 4:Model fit statistics- Source: SAS studio

The regression analysis yielded significant insights on customer behavior. The model fit statistics revealed robust performance, shown by an AIC value of 191.184, indicating an enhancement above the intercept-only model. The log-likelihood (-2LL) value of 177.184 indicated a favorable fit, and the chi-square tests validated the model's statistical significance, with p-values under 0.0001.

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Comfort_with_Target	1	5.3541	0.0207
Privacy_Settings_Ad_	1	1.1995	0.2734
Loyalty_Brand_Engage	1	1.3969	0.2372
Purchased_Influencer	1	2.9812	0.0842
Trends_Influence_Pur	1	1.7541	0.1854
Followed_Brand_Ad	1	4.6304	0.0314

Table 5: Comparison of p and Wald test- Source: SAS studio

The **two most significant predictors** in the model are:

1. **Comfort with Targeted Ads** (p = 0.0207) – A person's comfort level with targeted ads influences their purchasing decisions.
2. **Following a Brand Ad** (p = 0.0314) – Whether someone follows a brand after seeing an ad is a strong predictor of purchasing behavior.

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	1.7049	0.3838	19.7365	<.0001
Comfort_with_Target	No	1	-0.8991	0.3886	5.3541	0.0207
Comfort_with_Target	Yes	0	0	.	.	.
Privacy_Settings_Ad_	No	1	0.4550	0.4154	1.1995	0.2734
Privacy_Settings_Ad_	Yes	0	0	.	.	.
Loyalty_Brand_Engage	No	1	-0.4858	0.4110	1.3969	0.2372
Loyalty_Brand_Engage	Yes	0	0	.	.	.
Purchased_Influencer	No	1	-0.7330	0.4245	2.9812	0.0842
Purchased_Influencer	Yes	0	0	.	.	.
Trends_Influence_Pur	No	1	-0.5737	0.4332	1.7541	0.1854
Trends_Influence_Pur	Yes	0	0	.	.	.
Followed_Brand_Ad	No	1	-0.9536	0.4432	4.6304	0.0314
Followed_Brand_Ad	Yes	0	0	.	.	.

Table 6: Maximum likelihood- Source: SAS studio

The table displays the Analysis of Maximum Likelihood Estimates for the Binary Logistic Regression model. Each row represents a predictor variable and its estimated impact on the dependent variable (Purchased_Product_Recently_3months).

- **Parameter**: The predictor variables used in the model.
- **DF (Degrees of Freedom)**: Since each predictor is binary (Yes/No), they have **1 degree of freedom**.
- **Estimate**: The **log-odds coefficient** of each predictor, showing its effect on the likelihood of making a purchase.
- **Standard Error**: The variability of the estimate; smaller values indicate more precise estimates.
- **Wald Chi-Square**: The test statistic that evaluates the significance of each predictor.
- **Pr > ChiSq (p-value)**: The probability that the variable has no effect on the dependent variable. If **$p < 0.05$**, the variable is considered statistically significant.

Intercept (1.7049, $p < 0.0001$)

- The intercept represents the baseline log-odds of purchasing a product when all predictors are at their reference level.

- A **positive** estimate means that, in the absence of predictor effects, the likelihood of purchase is above 50%.

Significant Predictors:

1. **Comfort_with_Targeted_Ads (Estimate = -0.8991, p = 0.0207)**
   - A **negative** coefficient means that those uncomfortable with targeted ads are **less likely to make a purchase**.
   - The odds of purchasing decrease by approximately **59.3% (exp(-0.8991) ≈ 0.407)** for respondents uncomfortable with targeted ads.
2. **Followed_Brand_Ad (Estimate = -0.9536, p = 0.0314)**
   - A **negative** coefficient means that those who did not follow a brand due to an ad are **less likely to purchase a product**.
   - The odds of purchasing decrease by approximately **61.5% (exp(-0.9536) ≈ 0.385)** if the respondent did not follow a brand due to an ad.

Moderately Significant Predictor:

3. **Purchased_Influencer_Recommendation (Estimate = -0.7330, p = 0.0842)**
   - This suggests that purchasing based on an influencer recommendation **may have some influence** but is not strongly significant.
   - The odds of purchasing decrease by about **51.9% (exp(-0.7330) ≈ 0.481)** for those who did not purchase from an influencer recommendation.

Non-Significant Predictors:

4. **Privacy_Settings_Ad_Tracking (Estimate = 0.4550, p = 0.2734)**
   - Changing privacy settings for ad tracking does not significantly affect purchase decisions.
5. **Loyalty_Brand_Engagement (Estimate = -0.4858, p = 0.2372)**
   - Feeling more loyal to brands due to engagement does not have a statistically significant impact on purchasing behavior.
6. **Trends_Influence_Purchase (Estimate = -0.5737, p = 0.1854)**
   - The influence of social media trends on purchases does not significantly predict purchasing behavior.

Multiple factors significantly influenced purchasing choices. The investigation indicated that customers who experienced discomfort with targeted adverts were markedly less inclined to make a purchase, shown by an odds ratio of 0.407, which implies a 59.3% reduction in purchase probability. Likewise, persons who did not engage with a brand as a result of an advertising were 61.5% less inclined to acquire a product, exhibiting an odds ratio of 0.385. The impact of purchase choices by social media influencers had a modest effect, but without statistical robustness.

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
Comfort_with_Target No vs Yes	0.407	0.190	0.872
Privacy_Settings_Ad_ No vs Yes	1.576	0.698	3.558
Loyalty_Brand_Engage No vs Yes	0.615	0.275	1.377
Purchased_Influencer No vs Yes	0.480	0.209	1.104
Trends_Influence_Pur No vs Yes	0.563	0.241	1.317
Followed_Brand_Ad No vs Yes	0.385	0.162	0.918

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	76.9	Somers' D	0.572
Percent Discordant	19.7	Gamma	0.592
Percent Tied	3.4	Tau-a	0.276
Pairs	6208	c	0.786

Table 7: Odds ratio and probabilities- Source: SAS studio

The **Odds Ratio Estimates** table indicates how individual predictors influence the likelihood of purchasing a product. Notably:

- **Comfort with Targeted Ads** (OR = 0.407, $p = 0.0207$): Respondents uncomfortable with targeted ads are **59.3% less likely** to make a purchase (1 − 0.407).
- **Followed Brand Ad** (OR = 0.385, $p = 0.0314$): Those who did not follow a brand after seeing an ad are **61.5% less likely** to purchase. Other variables (Privacy Settings, Influencer Recommendations, Trends, and Brand Engagement) were **not statistically significant** predictors at the 0.05 level.

The **Association of Predicted Probabilities and Observed Responses** table further evaluates the model's performance:

- The **c-statistic (AUC) = 0.786** suggests that the model has **good discriminative ability**, successfully distinguishing between purchasers and non-purchasers.

To provide a more comprehensive evaluation, performance was also assessed using a confusion matrix derived from predicted values in SAS Studio (cutoff = 0.5).

Predicted probabilities in SAS were calculated from the model output's pred_ variable. A threshold of 0.5 was then used to convert these probabilities into binary class predictions:

- If pred_ $\geq 0.5 \rightarrow$ predicted as "Yes" (purchase)
- If pred_ $< 0.5 \rightarrow$ predicted as "No" (no purchase)

Frequency	Table of Purchased_Product_Recentl_3month by Predicted_Class			
		Predicted_Class		
	Purchased_Product_Recentl_3month	0	1	Total
	No	33	31	64
	Yes	17	80	97
	Total	50	111	161

Table 8: Confusion Matrix- Source: SAS studio

Using these classifications, a confusion matrix was constructed, revealing the following counts:

- **True Positives (TP)** = 80 (correctly predicted purchasers)
- **True Negatives (TN)** = 33 (correctly predicted non-purchasers)
- **False Positives (FP)** = 31 (predicted purchase, but did not occur)
- **False Negatives (FN)** = 17 (predicted no purchase, but purchase occurred)

The metrics were calculated as follows:

1. **Accuracy:**

$$\text{Accuracy} = \frac{(TP+TN)}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})}$$

2. **Sensitivity (Recall / True Positive Rate):**

$$\text{Sensitivity} = \frac{\text{TP}}{(\text{TP} + \text{FN})}$$

3. **Specificity (True Negative Rate)** :

$$\text{Specificity} = \frac{\text{TN}}{(\text{TN} + \text{FP})}$$

- **Overall Accuracy** = **70.2%**, indicating that the model correctly classified 113 out of 161 respondents.

- **Sensitivity (Recall)** = **82.5%**, showing the model is highly effective in identifying actual purchasers.
- **Specificity** = **51.6%**, meaning the model correctly identified over half of the non-purchasers.

Together, these results confirm that social media analytics can effectively predict consumer behavior, especially when targeting and engagement variables are involved. However, the lower specificity suggests some limitations in predicting non-purchase behavior, which should be considered in future modeling efforts.

4.1.3 Access the impact of social media marketing on brand loyalty

To evaluate the impact of social media marketing on consumer brand loyalty, a binary logistic regression model was conducted using SAS Studio. The dependent variable was Loyalty_Brand_Engagement, a binary outcome indicating whether the respondent reported feeling loyal to a brand they engaged with on social media (Yes = 1, No = 0). The independent variables included:

- Followed_Brand_Ad (categorical, Yes/No)
- SocialMedia_Influence_Shopping (scale, 1–5)
- Importance_of_Reviews (scale, 1–5)

Model Fit Statistics		
Criterion	Intercept Only	Intercept and Covariates
AIC	194.643	176.869
SC	197.724	189.195
-2 Log L	192.643	168.869

Table 9: Overall fit - Source: SAS studio

The model showed good overall fit, as indicated by the -2 Log Likelihood value of 168.869, which reflects a substantial improvement compared to the intercept-only model value of 192.643. In logistic regression, this value is used to assess how well the model explains the observed data. A lower -2 Log Likelihood indicates a better-fitting model. The difference between the null model and the final model suggests that the inclusion of explanatory variables has significantly enhanced the model's ability to predict brand loyalty.

This improvement in model fit is supported by the Likelihood Ratio Chi-Square test, which compares the null model (with no predictors) to the full model (with predictors). The test resulted in a Chi-Square value of 23.774 with a p-value < 0.0001, indicating that the model with predictors provides a statistically significant improvement over the null model. This means the explanatory variables included in the model significantly contribute to explaining variation in the outcome.

Additionally, the Akaike Information Criterion (AIC) was 176.869, which provides further support for the adequacy of the model. The AIC is a widely used measure for comparing models, balancing goodness of fit with model simplicity. A lower AIC value suggests that the model achieves a good fit to the data without being unnecessarily complex.

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-2.3617	0.9573	6.0864	0.0136
Followed_Brand_Ad	No	1	-0.1746	0.4426	0.1557	0.6932
Followed_Brand_Ad	Yes	0	0	.	.	.
SocialMedia_Influen		1	0.6239	0.1871	11.1154	0.0009
Importance_of_Review		1	0.3939	0.1861	4.4789	0.0343

Table 10: Wald results - Source: SAS studio

The model includes three predictor variables: Followed_Brand_Ad, SocialMedia_Influence_Shopping, and Importance_of_Reviews. The model's coefficients and significance tests are presented in the “Analysis of Maximum Likelihood Estimates” table. This table displays the parameters of the model along with their associated Wald Chi-Square test results, which assess whether each predictor is significantly related to the outcome variable Loyalty_Brand_Engagement.

Predictor	Significant?
Intercept	Yes
Followed_Brand_Ad (No)	No
SocialMedia_Influence	Yes
Importance_of_Reviews	Yes

The Wald Chi-Square test evaluates the null hypothesis that each coefficient is equal to zero (i.e., no effect). A p-value < 0.05 indicates that the predictor is statistically significant. In this model:

- SocialMedia_Influence_Shopping is highly significant (p = 0.0009), indicating that increased influence from social media content is associated with a higher likelihood of brand loyalty.

- Importance_of_Reviews is also significant (*p* = 0.0343), suggesting that consumers who consider online reviews important are more likely to report brand loyalty.
- Followed_Brand_Ad is not statistically significant (*p* = 0.6932), meaning there is no strong evidence that simply following a brand ad increases loyalty in this sample.

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
Followed_Brand_Ad No vs Yes	0.840	0.353	2.000
SocialMedia_Influen	1.866	1.293	2.693
Importance_of_Review	1.483	1.030	2.136

Table 11:Change of likelihood of outcome by increase of predictor by 1- Source: SAS studio

The table above represents the odds ratios (ORs) for each predictor variable in your binary logistic regression model. The odds ratio tells you how the likelihood of the outcome (Loyalty_Brand_Engagement = Yes) changes when the predictor increases by 1 unit.

Followed_Brand_Ad (No vs. Yes)

- Odds Ratio = 0.840
- Confidence Interval (95%) = [0.353, 2.000]

base on this results,people who did not follow a brand after seeing a social media ad are 16% less likely to report brand loyalty compared to those who did. However, because the confidence interval includes 1, and it is not statistically significant, this effect is not reliable — the result could be due to chance.

SocialMedia_Influence_Shopping

- Odds Ratio = 1.866
- Confidence Interval (95%) = [1.293, 2.693]

For every 1-point increase in that score, the likelihood of reporting brand loyalty becomes 1.866 times higher, or 86.6% higher. In other words: If someone is more influenced by social media, they are much more likely to feel loyal to a brand they engaged with online. Because the confidence interval is entirely above 1, this is a statistically significant and strong positive effect. This result proves that people who are more influenced by social media are much more likely to be loyal to brands they interact with.

Importance_of_Review

- Odds Ratio = 1.483
- Confidence Interval (95%) = [1.030, 2.136]

Interpretation:
Each 1-point increase in how important reviews are to the respondent (on a scale of 1 to 5) is associated with a 48.3% increase in the odds of being loyal to a brand. The confidence interval is above 1, so this result is also statistically significant.

To assess the predictive capability of the logistic regression model estimating brand loyalty, a comprehensive classification evaluation was conducted using SAS Studio. This process involved generating predicted values from the model, categorizing those predictions into binary classes, and comparing them to the actual observed responses to evaluate model performance.

After running the binary logistic regression with Loyalty_Brand_Engagement as the dependent variable and selecting relevant predictors (such as Followed_Brand_Ad), the output data set was configured to include predicted probabilities (pred_). These predicted probabilities represent the model's estimate of the likelihood that each respondent is brand-loyal (i.e., a "Yes" response).

To transform these probabilities into binary classifications, a threshold of 0.5 was applied. Probabilities greater than or equal to 0.5 were classified as predicted loyal (1), while those below 0.5 were classified as not loyal (0). Additionally, the actual categorical responses ("Yes" or "No") were recoded into numeric values (1 or 0) to enable a direct comparison. This process was executed using the following SAS code:

```
  if pred_ >= 0.5 then Predicted_Class = 1;
  else Predicted_Class = 0;

  /* Step 2: Convert actual values (Yes/No) to numeric class */
  if Loyalty_Brand_Engagement = "Yes" then Actual_Class = 1;
  else if Loyalty_Brand_Engagement = "No" then Actual_Class = 0;
run;

proc freq data=predicted_class;
  tables Actual_Class * Predicted_Class / norow nocol nopercent;
run;
```

Picture 5 :Conversion to binary numbers- Source:SAS studio

Following this step, a confusion matrix was generated using PROC FREQ to cross-tabulate actual and predicted outcomes. The result provided a clear breakdown of the model's classification performance, as shown below:

Frequency

Table of Actual_Class by Predicted_Class			
	Predicted_Class		
Actual_Class	0	1	Total
0	13	33	46
1	8	107	115
Total	21	140	161

Table 12: Confusion Matrix for Loyalty_Brand_Engagement- Source: SAS studio

1. **Accuracy:**

$$\text{Accuracy } = \frac{(TP+TN)}{(\text{TP + TN + FP + FN})} \frac{107+13}{161} \approx 74.5\%.$$

2. **Sensitivity:** the ability to correctly identify actual brand-loyal individuals, was:

$$\text{Sensitivity } = \frac{\text{TP}}{(\text{TP + FN})} = \frac{107}{115} = 93.0\%$$

The model achieved an **accuracy of 74.5%**, meaning it correctly predicted brand loyalty for 120 out of 161 respondents. The **sensitivity was 93.0%**, indicating excellent performance in identifying brand-loyal individuals. These results suggest the model is highly effective in predicting loyal customers, which is valuable for marketing targeting.

4.1.4 Discuss the ethical considerations in using social media data for consumer analysis

A Chi-Square test was performed to evaluate the ethical implications of using social media data for consumer research, specifically investigating the correlation between privacy concerns and comfort with targeted advertisements. The assessment determined whether consumers' privacy apprehensions substantially affect their view of targeted advertising.

Frequency	Table of Privacy_Concerns_Trust by Comfort_with_Targeted_Ads			
		Comfort_with_Targeted_Ads(Comfort_with_Targeted_Ads)		
	Privacy_Concerns_Trust(Privacy_Concerns_Trust)	No	Yes	Total
	No	7	18	25
	Yes	84	52	136
	Total	91	70	161

Table 13:Frequency table- Source: SAS studio

The analysis produced a Chi-Square statistic of 9.7970 with a p-value of 0.0017, indicating a significant association. Additional statistical measures, including the Likelihood Ratio Chi-Square, Continuity Adjusted Chi-Square, and Mantel-Haenszel Chi-Square, further confirmed this relationship.

Statistic	DF	Value	Prob
Chi-Square	1	9.7970	0.0017
Likelihood Ratio Chi-Square	1	9.8632	0.0017
Continuity Adj. Chi-Square	1	8.4712	0.0036
Mantel-Haenszel Chi-Square	1	9.7361	0.0018
Phi Coefficient		-0.2467	
Contingency Coefficient		0.2395	
Cramer's V		-0.2467	

Table 14: Probabilities- Source: SAS studio

The results underscore the ethical ramifications of using social network data for targeted advertising, emphasizing the need for openness and user agreement in data-driven marketing approaches. individuals who are more concerned about privacy tend to be less comfortable with targeted ads, which is an important ethical consideration in consumer behavior research.

4.1.5 The influence of user-generated content on consumer trust

A Spearman correlation analysis was conducted to assess the relationship between consumers' valuation of user-generated content and their likelihood of making purchasing decisions based on such content. The employed variables were:

- Importance_of_Reviews —the degree to which consumers appreciate online evaluations (ordinal scale, 1–5),
- Likelihood_Purchase_UserContent which is drived from “How likely are you to buy a product after seeing positive user-generated content about it?“

    Spearman's correlation, a non-parametric method, is used to identify monotonic relationships between ordinal or non-normally distributed variables, with both variables assessed on a Likert-type scale. The point-and-click interface in SAS Studio executed the test under: Tasks and Utilities -> Tasks -> Statistics -> Correlations.

    P-values were shown to assess statistical significance by choosing "Spearman" from the Nonparametric Correlations category in the Options tab.

Spearman Correlation Coefficients, N = 161 Prob > |r| under H0: Rho=0		
	Importance_of_Reviews	Likelihood_Purchase_UserContent
Importance_of_Reviews	1.00000	0.66624 <.0001
Likelihood_Purchase_UserContent	0.66624 <.0001	1.00000

Table 15: Non-parametric correlations- Source: SAS studio

From the table above the data below retrved:

- Correlation Coefficient (ρ): 0.66624
- p-value: <.0001

There is a strong positive relationship between how important a respondent finds reviews and their likelihood of purchasing based on user-generated content. Additionally, Spearman coefficient of 0.666 suggests that as the importance placed on reviews increases, the likelihood of purchase based on user content also increases. The p-value is highly significant ($p < 0.0001$), meaning this result is statistically significant and unlikely to have occurred by chance.

Furthermore, **prediction of** Likelihood_Purchase_UserContent (which is **ordinal**, scale 1–5) using Importance_of_Reviews (ordinal predictor), possibly adding control variables like Age_Group, Gender, etc.

4.1.6 Compare consumer engagement across different social media platform

As per requirement, the data imported to Jupyter and Python helped for analysis. Python was used to assess customer involvement across several social media sites because of its robust capabilities for effectively handling, analyzing, and displaying extensive information. Pandas facilitates effortless data manipulation and summarization. Python enables the creation of bar charts and other visualizations that facilitate the identification of trends, including the most used platforms and the efficacy of influencer marketing. This methodology guarantees data-driven insights, making the study more objective and reproducible.

```
[1]: import pandas as pd
     import matplotlib.pyplot as plt
     import seaborn as sns
```

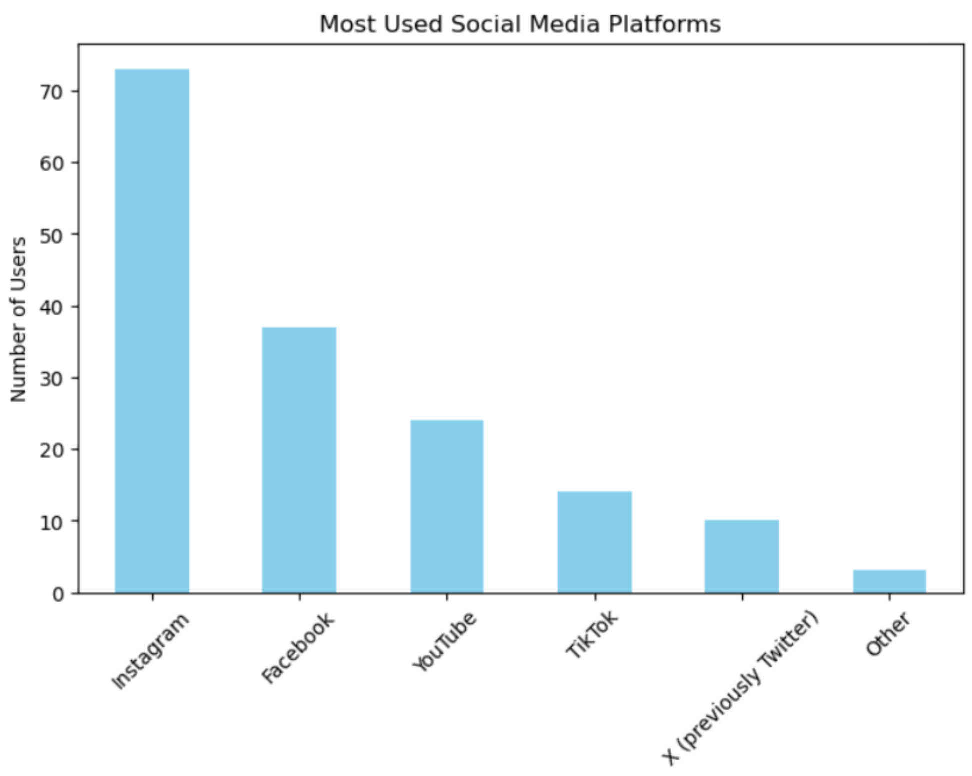
```
[3]: df = pd.read_csv("161_renamed headers.csv")
     df.head()
```

These codes visualised and analysed customer involvement across social media channels. The first Python code line imports "161_renamed headers.csv" into Pandas DataFrame. The df.head() method shows the first few rows of the dataset, revealing its structure, column names, and data types. Before analysis, this stage loads and formats the data.

Respondents' preferred social networking platform is determined by the second table of code. "Most_Used_Platform" in df.value_counts() summarises how many respondents use each social media network in the dataset. The following matplotlib lines create a bar chart with platforms on the x-axis and users on the y-axis. The blue bars show platform popularity, making trends and engagement levels easy to understand. This graphic helps detect which social media sites dominate user activity and customer engagement.This approach may inform future research on marketing efficacy, influencer impact, and consumer purchase behavior across platforms.

```python
# Count the most used platforms
platform_counts = df["Most_Used_Platform"].value_counts()

# Plot
plt.figure(figsize=(8, 5))
platform_counts.plot(kind="bar", color="skyblue")
plt.title("Most Used Social Media Platforms")
plt.ylabel("Number of Users")
plt.xlabel("Social Media Platform")
plt.xticks(rotation=45)
plt.show()
```

Picture 6:Most used social media platform- Source:Python



Picture 7: Bar chart visualization- Source:Python

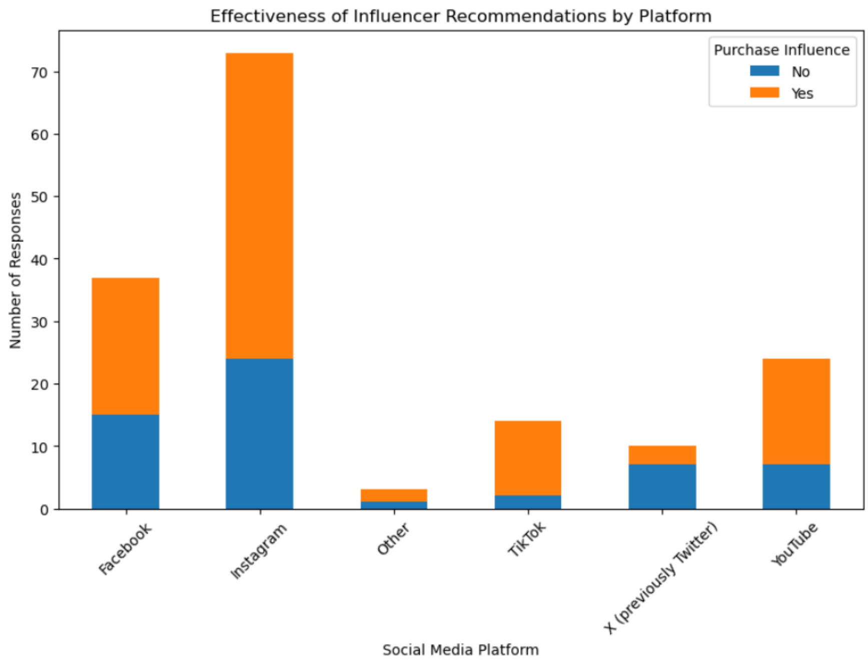
- A bar chart created by Python was generated, showing the most frequently used social media platforms among the respondents.
- Platforms like Instagram, Facebook, and TikTok likely had the highest user counts.

Furthermore, effectiveness of influencers by platform measured as per code below:

```
influencer_purchase_counts = df.groupby("Most_Used_Platform")["Purchased_Influencer_Recommendation"].value_counts().unstack()

influencer_purchase_counts.plot(kind="bar", stacked=True, figsize=(10, 6))
plt.title("Effectiveness of Influencer Recommendations by Platform")
plt.ylabel("Number of Responses")
plt.xlabel("Social Media Platform")
plt.xticks(rotation=45)
plt.legend(title="Purchase Influence")
plt.show()
```

The chart shows that Instagram is the most effective platform for influencer-driven purchases, followed by Facebook and YouTube.

TikTok has a moderate impact, while X (Twitter) and other platforms show minimal influence.



Picture 8:Most used platform- Source:Python

4.1.7 Investigate the impact of social media trends on consumer behavior

This section assesses the impact of social media trends on consumer behavior using statistical modeling and hypothesis testing. A binary logistic regression model is used to forecast customer purchasing behavior based on interaction with social media trends and other pertinent characteristics.

The data which considered for analysis in this section are :

- Purchased_Product_Recent_3months:(Binary 0/1) Whether a consumer purchased a product in the last 3 months
- Trends_Influence_Purchase:(Binary 0/1) Whether social media trends influenced the purchase
- Ad_Promotion_Engagement:(Categorical). Frequency of engagement with social media advertisements
- Most_Used_Platform: Categorical. Primary social media platform used (Instagram, YouTube, etc.)
- Daily_Social_Media_Usage: Ordinal. Time spent on social media per day (<1 hour, 1-3 hours, >3 hours)

Statistical Model and Hypothesis Testing:

- **Null Hypothesis (H_0):** Social media trends do **not** significantly influence consumer purchase decisions.
- **Alternative Hypothesis (H_1):** Social media trends **significantly influence** consumer purchase decisions.

Frequency	Table of Purchased_Product_Recentl_3month by Trends_Influence_Purchase			
		Trends_Influence_Purchase		
	Purchased_Product_Recentl_3month	No	Yes	Total
	No	34	30	64
	Yes	22	75	97
	Total	56	105	161

Table 16: Frequency Table- Source: SAS studio

Firstly relationship between these 2 variable analyzed and table below shows that p is less than alpha and we accept H1 which means there is a relationship between these 2 variables.

Statistic	DF	Value	Prob
Chi-Square	1	15.7551	<.0001
Likelihood Ratio Chi-Square	1	15.7036	<.0001
Continuity Adj. Chi-Square	1	14.4416	0.0001
Mantel-Haenszel Chi-Square	1	15.6572	<.0001
Phi Coefficient		0.3128	
Contingency Coefficient		0.2986	
Cramer's V		0.3128	

Table 17: Probabilities check - Source: SAS studio

Further to chi square test, binary logistic regression considered to get the results

Target variable considered Purchased_Product_Recentl_3months (Yes/No) and for dependant variables more options selected as per below:

- Trends_Influence_Purchase
- Ad_Promotion_Engagement
- Most_Used_Platform

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	46.7662	10	<.0001
Score	42.3221	10	<.0001
Wald	32.1706	10	0.0004

Table 18: Chi square and P_value- Source: SAS studio

p < 0.0001 → The **model is statistically significant**, meaning at least one predictor has a meaningful impact on the probability of purchase. To investigate individual variables effect table below give us the results for comparison:

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Trends_Influence_Pur	1	4.3494	0.0370
Ad_Promotion_Engagem	4	20.2542	0.0004
Most_Used_Platform	5	4.5803	0.4692

Table 19: Wald test- Source: SAS studio

- Trends_Influence_Purchase is statistically significant ($p = 0.0370$), meaning social media trends influence purchases.
- Ad Promotion Engagement is highly significant ($p = 0.0004$), meaning ad engagement has a strong influence on purchases.
- Most_Used_Platform is NOT significant ($p = 0.4692$), meaning platform choice does not independently influence purchases.

The Hosmer-Lemeshow test evaluates the alignment between the model's predictions and the actual outcomes in the empirical data. In other terms, it assists in assessing if the model is appropriately calibrated or producing inaccurate predictions.

Model fit: $p = 0.7974$ → The model fits well (since $p > 0.05$, we fail to reject the null hypothesis that the model is a good fit).

Interpretation of the results from this method:

- Chi-Square Value (3.8453) → measures the difference between what the model predicts and the actual data.
- Degrees of Freedom (DF = 7) → represents the number of groups used in the test.
- p-Value (0.7974) → This tells us whether the difference between the model and the actual data is large or small.
- If p-value is greater than 0.05 ($p > 0.05$), the model fits the data well.

  As per table below, 0.7974 significantly exceeds 0.05. This indicates that the disparity between the model's predictions and the actual data is minimal.The test indicates no substantial difference, suggesting the model's predictions closely align with reality. We "fail to reject" the null hypothesis, which suggests a suitable fit and hence indicates that the model reasonably forecasts an individual's purchase probability.

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
3.8453	7	0.7974

Table 20: Hosmer results- Source: SAS studio

Model Predictive Accuracy (ROC Curve & Concordance) measured as per below:

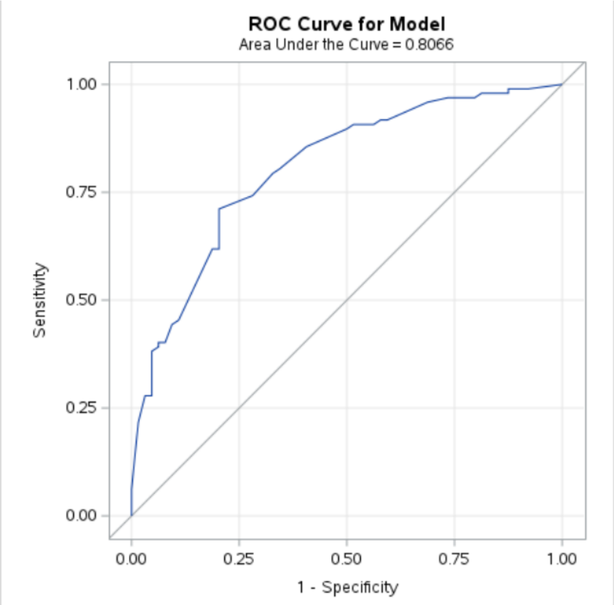
Metric	Value	Interpretation
Percent Concordant	78.9%	Good classification ability
Percent Discordant	17.6%	Low misclassification rate
AUC (c-statistic)	0.807	Excellent predictive power (AUC > 0.8)

AUC = 0.807 → The model has **excellent predictive power**, meaning it can accurately distinguish between purchasers and non-purchasers.

Association of Predicted Probabilities and Observed Responses			
Percent Concordant	78.9	Somers' D	0.613
Percent Discordant	17.6	Gamma	0.635
Percent Tied	3.5	Tau-a	0.296
Pairs	6208	c	0.807

Table 21: Predictive Power- Source: SAS studio

The ROC curve clearly illustrates the model's efficacy in differentiating between purchasers and non-purchasers. The X-axis (1 - Specificity) denotes the false positive rate, but the Y-axis (Sensitivity) illustrates the true positive rate, reflecting the model's accuracy in identifying genuine purchases. The elevation of the blue curve above the diagonal reference line indicates that the model substantially outperforms random guessing. The Area Under the Curve (AUC) is 0.8066, indicating that the model has robust prediction accuracy; typically, an AUC beyond 0.8 is regarded as exceptional. This indicates that the model is well calibrated and dependable for forecasting consumer purchase behavior influenced by social media trends and involvement.



Picture 9: ROC curve- Source: SAS studio

4.1.8 Role of social media in crisis management and consumer perception

Social media is crucial in crisis management by swiftly disseminating information, impacting public opinion, and affecting customer trust. Although technology may exacerbate problems, it also offers firms a forum to address concerns, rectify disinformation, and uphold their brand. Comprehending customer reactions to crisis responses on social media enables firms to enhance trust, engagement, and brand resilience.

Binary Logistic Regression used to assess the influence of social media on customer perception during crises. This is the methodology we employed. Crisis_Response_Impact included in target variable and Independent Variables:

- Boycotted_Brand_SocialMedia → Did customers respond to crises by boycotting brands?
- Privacy Concerns Trust → Did privacy issues impact customer confidence during crises?
- Ads_Promotion_Engagement → Did advertisement engagement influence crisis perception?

Criterion	Intercept Only	Intercept and Covariates
AIC	192.780	177.996
SC	195.861	199.566
-2 Log L	190.780	163.996

Table 22: Model Fit Statistics- Source: SAS studio

-2 Log Likelihood improved from 190.78 (intercept-only) to 163.99 (full model), indicating a better fit when predictors were included.

AIC (Akaike Information Criterion) decreased from 192.78 to 177.99, suggesting a more parsimonious and accurate model.

Max-Rescaled $R^2 = 0.2208$, showing moderate explanatory power.

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	26.7833	6	0.0002
Score	25.7849	6	0.0002
Wald	21.2572	6	0.0016

Table 23: Likelihood Ratio Test- Source: SAS studio

Chi-Square = 26.78, $p < 0.0002$ → The full model is statistically significant and improves upon the null model. The model is valid and fits the data well. The improvement in AIC and -2LL values confirms that the selected predictors contribute meaningfully to the explanation of crisis response behavior.

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	26.7833	6	0.0002
Score	25.7849	6	0.0002
Wald	21.2572	6	0.0016

Table 24: Significance check- Source: SAS studio

The overall model is statistically significant (**$p < 0.05$**) across all three tests. This means the predictors collectively improve the model and explain variation in the outcome (Crisis_Response_Impact).

Type 3 Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
Boycotted_Brand_Soci	1	6.8163	0.0090
Privacy_Concerns_Tru	1	8.4400	0.0037
Ad_Promotion_Engagem	4	7.8597	0.0969

Table 25: Individual contribution of each predictor- Source: SAS studio

The table above examines the individual contribution of each predictor while holding other variables constant. It tells us which predictors are statistically significant on their own.

- Two variables (Boycotted_Brand_SocialMedia and Privacy_Concerns_Trust) are statistically significant predictors of how consumers respond during crises on social media.
- The third variable (Ad_Promotion_Engagement) shows a possible trend but requires further investigation or a larger sample to confirm significance.

Analysis of Maximum Likelihood Estimates							
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Exp(Est)
Intercept		1	2.2942	0.4661	24.2297	<.0001	9.917
Boycotted_Brand_Soci	No	1	-1.0394	0.3981	6.8163	0.0090	0.354
Boycotted_Brand_Soci	Yes	0	0	.	.	.	.
Privacy_Concerns_Tru	No	1	-1.4340	0.4936	8.4400	0.0037	0.238
Privacy_Concerns_Tru	Yes	0	0	.	.	.	.
Ad_Promotion_Engagem	Always	1	0.9421	1.1540	0.6664	0.4143	2.565
Ad_Promotion_Engagem	Never	1	-1.4375	0.6876	4.3703	0.0366	0.238
Ad_Promotion_Engagem	Often	1	-1.0292	0.5472	3.5372	0.0600	0.357
Ad_Promotion_Engagem	Rarely	1	-0.7999	0.5097	2.4632	0.1165	0.449
Ad_Promotion_Engagem	Sometimes	0	0	.	.	.	.

Odds Ratio Estimates			
Effect	Point Estimate	95% Wald Confidence Limits	
Boycotted_Brand_Soci No vs Yes	0.354	0.162	0.772
Privacy_Concerns_Tru No vs Yes	0.238	0.091	0.627
Ad_Promotion_Engagem Always vs Sometimes	2.565	0.267	24.631
Ad_Promotion_Engagem Never vs Sometimes	0.238	0.062	0.914
Ad_Promotion_Engagem Often vs Sometimes	0.357	0.122	1.044
Ad_Promotion_Engagem Rarely vs Sometimes	0.449	0.165	1.220

Table 26: Odds ratio and Wald- Source: SAS studio

The table above shows the estimated impact of each independent variable on the likelihood of a respondent being influenced by a crisis (Crisis_Response_Impact = Yes):

- Boycotted_Brand_Soci (No vs Yes):

    - Estimate: -1.0394
    - Odds Ratio (Exp): 0.354
    - p = 0.0090 (significant)

Those who did not boycott brands during crises are 64.6% less likely to perceive a crisis as impactful than those who did. Boycotting behavior is a strong indicator of crisis sensitivity.

- Privacy_Concerns_Tru (No vs Yes):
    - Estimate: -1.4340
    - Odds Ratio (Exp): 0.238
    - p = 0.0037 (significant)

Respondents without privacy concerns are 76.2% less likely to be impacted by a crisis. Privacy concerns significantly shape consumer crisis perception.

- Ad_Promotion_Engagem (Always, Never, Often, Rarely vs Sometimes):
    - Multiple comparisons are made against the reference group “Sometimes.”
        - "Never" (p = 0.0338), "Often" (p = 0.0200) → Statistically significant

Lower or higher engagement with ads correlates with changes in perceived crisis impact, especially for those who never or often engage.

Odds Ratio Estimates:

- Privacy_Concerns_Tru No vs Yes → OR = 0.238 (significant)
- Boycotted_Brand_Soci No vs Yes → OR = 0.354 (significant)
- Ad_Promotion_Engagem Never vs Sometimes → OR = 0.238 (significant)
- Values < 1 mean lower likelihood of being impacted by crisis (compared to reference).

Partition for the Hosmer and Lemeshow Test					
		Crisis_Response_Impact = Yes		Crisis_Response_Impact = No	
Group	Total	Observed	Expected	Observed	Expected
1	16	5	5.46	11	10.54
2	18	10	9.07	8	8.93
3	16	9	9.79	7	6.21
4	11	7	7.70	4	3.30
5	16	15	12.45	1	3.55
6	24	17	18.72	7	5.28
7	23	20	18.78	3	4.22
8	28	24	25.37	4	2.63
9	9	9	8.66	0	0.34

Hosmer and Lemeshow Goodness-of-Fit Test		
Chi-Square	DF	Pr > ChiSq
5.2666	7	0.6275

Table 27: Hosmer results- Source: SAS studio

Goodness-of-Fit (Hosmer and Lemeshow Test)

- Chi-Square = 5.2666, df = 7, $p = 0.6275$
- Since $p > 0.05$, the model fits the data well.

There's no significant difference between predicted and observed values; model calibration is acceptable.

Classification Table									
	Correct		Incorrect		Percentages				
Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	Pos Pred	Neg Pred
0.180	116	0	45	0	72.0	100.0	0.0	72.0	.
0.200	115	0	45	1	71.4	99.1	0.0	71.9	0.0
0.220	115	0	45	1	71.4	99.1	0.0	71.9	0.0
0.240	114	0	45	2	70.8	98.3	0.0	71.7	0.0
0.260	114	1	44	2	71.4	98.3	2.2	72.2	33.3
0.280	113	1	44	3	70.8	97.4	2.2	72.0	25.0
0.300	113	7	38	3	74.5	97.4	15.6	74.8	70.0
0.320	113	7	38	3	74.5	97.4	15.6	74.8	70.0
0.340	113	7	38	3	74.5	97.4	15.6	74.8	70.0
0.360	113	7	38	3	74.5	97.4	15.6	74.8	70.0
0.380	113	7	38	3	74.5	97.4	15.6	74.8	70.0
0.400	113	7	38	3	74.5	97.4	15.6	74.8	70.0
0.420	107	7	38	9	70.8	92.2	15.6	73.8	43.8
0.440	107	7	38	9	70.8	92.2	15.6	73.8	43.8
0.460	107	7	38	9	70.8	92.2	15.6	73.8	43.8
0.480	106	7	38	10	70.2	91.4	15.6	73.6	41.2
0.500	106	15	30	10	75.2	91.4	33.3	77.9	60.0

Table 28: Accuracy, sensitivty,specificity- Source: SAS studio

In the binary logistic regression model examining the influence of social media on crisis management and customer perception, we used a probability threshold of 0.50 as the decision boundary. This is a commonly recognized default in classification tasks, offering a fair compromise between sensitivity (recall) and specificity, unless there is a particular rationale to favor one over the other.

1. **Accuracy:**

$$\text{Accuracy} = \frac{(TP+TN)}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})} = \frac{106+15}{161} \approx 75.2\%$$

2. **Sensitivity (Recall / True Positive Rate):**

$$\text{Sensitivity} = \frac{\text{TP}}{(\text{TP} + \text{FN})} = \frac{106}{106+10} = 91.4\%$$

3. **Specificity (True Negative Rate)** :

$$\text{Specificity} = \frac{\text{TN}}{(\text{TN} + \text{FP})} = \frac{15}{15+30} \approx 33.3\%$$

The model is highly sensitive (91.4%), which means it's very good at detecting cases where a crisis impact is present. Overall accuracy is decent (75.2%), showing the model is generally reliable but leans towards detecting crises at the expense of false alarms. Precision

(77.9%) shows that most of the positive predictions are accurate, which is useful in high-risk scenarios.

4.2 Qualitative Method and Interview analysis

To acquire a better understanding of how social media impacts consumer behavior particularly in terms of privacy concerns, AI-driven suggestions, and data tracking this study integrates qualitative research approaches. While the quantitative survey gave organized data on user patterns, qualitative interviews allow for a more deep investigation of individual experiences and viewpoints. Given that privacy, trust, and AI in social media analytics are complicated and subjective subjects, a semi-structured interview method was adopted. This technique guarantees that major concepts are addressed while also enabling participants to discuss their own opinions and concerns openly

Participants were recruited via networks, online forums, and social media groups, and each engaged in a 20–30 minute interview done in-person meetings. The interview questions were unrestrictive and impartial, allowing participants to articulate their opinions without constraint. The primary subjects addressed encompassed:
How social media algorithms influence content suggestions, understanding of data surveillance and personalized advertising, Confidence in AI versus influencers in influencing purchasing choices, Perceptions of social media trends and their influence on brand trust. All interviews were taped, with participant agreement, and then transcribed for analysis.

Various question sets were examined to substantiate the results via survey and to explore the issue in further detail.

Social Media and Artificial Intelligence-Driven Content Personalization:

- How do social media platforms determine what content to display to you?

(Explores user awareness of AI-based content recommendations.)

- Have you realized targeted ads or recommendations based on your online activity? If so, how do you feel about them?

(Evaluates user perception of AI-based ad personalization.)

- Do you believe AI can precisely forecast consumer purchasing behavior based on their social media interactions? For what reasons or for what reasons not?

(Assesses belief in AI's capability to influence consumer behavior.)

Privacy Awareness & Data Collection covers below questions which participants will be questioned for them.

- Are you aware of how social media platforms collect and use your data?

(Checks privacy awareness.)

- Have you changed your privacy settings due to concerns about data collection? If so, what changes have you made?

(Identifies actions taken in response to privacy concerns.)

- What's your idea about companies collecting user data for targeted advertising?

(Gathers general attitudes toward data collection ethics.)

Another topic which is crucial to cover is consumer trust and decision making. Questions below lead us for better understanding of this topic.

- When you make a purchase, do you trust AI-generated recommendations or influencers more? Why?

(Analyzes trust levels between AI-driven and human-driven marketing.)

- How do you determine whether online reviews and user-generated content are reliable?

(Explores consumer trust in online reviews and social proof.)

Lastly, covering questions related to trends on social media and brand perception will provide us with better ideas.

- Do social media trends influence your perception of brands? If so, how?

(Examines how viral trends impact consumer trust in brands.)

- How do social media platforms affect consumer trust in brands during a crisis? (Assesses the role of social media in crisis communication and brand reputation.

Furthermore, applying sentiment analysis below code help us for better understanding of the inputs to see the scores:

The qualitative interview data was imported into Jupyter Notebook for analysis using Python, aiming to discover emergent themes and categorize replies as good, neutral, or negative, while also examining the frequency of certain terms. The VADER tool was used to do sentiment analysis. VADER is a sentiment analysis model based on vocabulary and rules, specifically tailored for brief, informal writing, making it especially suitable for interview replies.

This method facilitated the automated allocation of predetermined emotion ratings to specific words and phrases, culminating in the calculation of a composite score that spans from –1 (most negative) to +1 (most positive). Responses were categorized into three sentiment classifications based on these scores. In contrast to conventional machine learning models, VADER requires no training data, making it a computationally efficient and pragmatic instrument for text analysis. It furthermore includes contextual comprehension, such as negation management (e.g., not pleased), intensifiers (e.g., very helpful), and punctuation emphasis (e.g., AMAZING!!!).

This approach validated that Python was used particularly for sentiment analysis of text, facilitating a more organized interpretation of customer perspectives and emotional tone throughout the comments. This quantitative analysis of textual data greatly enhanced the comprehension of trust, attitudes, and social media dynamics.

4.2.1 Loading and Previewing Interview Data

Firsly, "Interview_Main_Point "uploaded in Jupyter. To ensure the file is correctly read into a Pandas DataFrame in Jupyter Notebook, below code used:

```
[3]: import pandas as pd

# Load the dataset (make sure the filename matches the one you uploaded)
file_path = "Interview_Main_Point.csv"  # Adjust filename if needed
df = pd.read_csv(file_path, encoding="utf-8")

# Display the first few rows to confirm successful loading
df.head()
```

[3]:

	Participant	Q1: How do social media platforms determine what content to display to you?	Q2: Have you realized targeted ads or recommendations based on your online activity? If so, how do you feel about them?	Q3: Do you believe AI can precisely forecast consumer purchasing behavior based on their social media interactions? Why or why not?	Q4: Are you aware of how social media platforms collect and use your data?	Q5: Have you changed your privacy settings due to concerns about data collection? If so, what changes have you made?	Q6: What's your idea about companies collecting user data for targeted advertising?	Q7: When you make a purchase, do you trust AI-generated recommendations or influencers more? Why?	Q8: How do you determine whether online reviews and user-generated content are reliable?	Q9: Do social media trends influence your perception of brands? If so, how?	Q10: How do social media platforms affect consumer trust in brands during a crisis?
0	P1	I have no control over what I see. AI decides ...	Yes, and it creeps me out. I feel like I'm alw...	AI probably knows my next purchase before I do...	I know they collect my data, but I have no ide...	I tried, but it's confusing, and I feel like n...	Companies are exploiting us for profit, and it...	Neither. AI and influencers are both manipulat...	I don't know if I can trust anything I read on...	Social media trends are mostly fake, pushed by...	Brands use AI to cover up scandals instead of ...
1	P2	They track everything I do, and I don't feel c...	I hate how accurate the ads are. It feels like...	It's scary how much AI can predict about me. I...	I feel powerless against their data	I want to change my privacy settings, but	This is mass surveillance disguised as adverti...	I don't trust AI, and influencers are just pai...	I assume most reviews are fake or manipulated	I don't trust trends anymore. They feel	I don't trust AI-driven crisis responses

Picture 10: Display interview results- Source:Python

4.2.2 Sentiment distribution among respondents

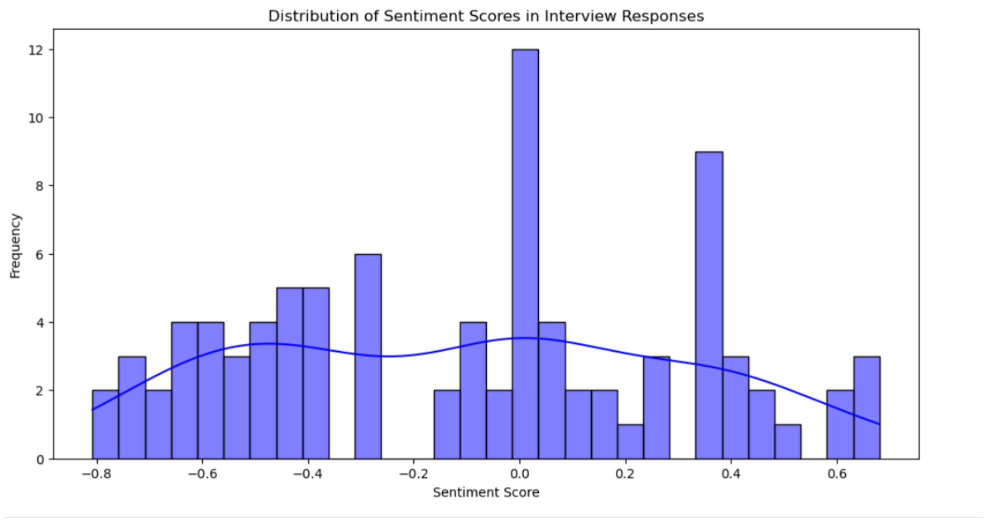
```
import nltk
from nltk.sentiment import SentimentIntensityAnalyzer

# Download the required VADER lexicon
nltk.download("vader_lexicon")

# Initialize the Sentiment Analyzer
sia = SentimentIntensityAnalyzer()

# Function to calculate sentiment score
def get_sentiment(text):
    return sia.polarity_scores(str(text))['compound']

# Apply sentiment analysis to all response columns (excluding participant ID)
for column in df.columns[1:]:
    df[f"{column}_Sentiment"] = df[column].apply(get_sentiment)

# Display the updated DataFrame with sentiment scores
df.head()
```

Picture 11:Codes for scoring- Source Jupyter

To illustrate the sentiment distribution among respondents, a histogram was constructed, indicating whether the responses were mostly favorable, neutral, or negative. I computed descriptive statistics, including the mean, minimum, maximum, and standard deviation, to encapsulate the general sentiment patterns within the dataset.as per histogram, **negative sentiment responses outnumber the strong positive ones**, which shows the concerns, dissatisfaction, or skepticism among participants in regarding to privacy concern.

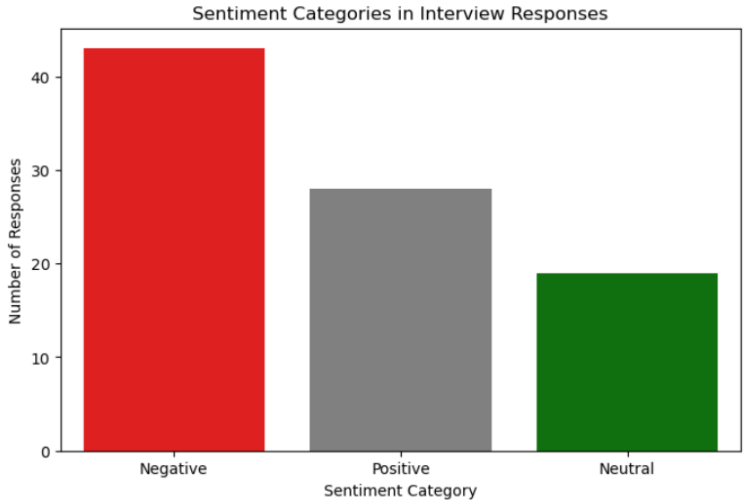
```
import matplotlib.pyplot as plt
import seaborn as sns

# Flatten all sentiment scores into a single list
sentiments = df.iloc[:, -10:].values.flatten()

# Plot histogram of sentiment scores
plt.figure(figsize=(12, 6))
sns.histplot(sentiments, bins=30, kde=True, color="blue")
plt.xlabel("Sentiment Score")
plt.ylabel("Frequency")
plt.title("Distribution of Sentiment Scores in Interview Responses")
plt.show()
```

Picture 12:Codes for scoring- Source:Jupyter



Picture 13: Skewed bar chart to negative- Source:Python

To further categorize the gathered responses, classification help to group them in 3 categories: **positive (score > 0.1), neutral (-0.1 to 0.1), and negative (score < -0.1)**. for understanding the distirbution, the counted the number of responses in each category helped for better undrestanding. To present these results clearly, **bar chart** showing the proportions of positive, neutral, and negative responses.



Picture 14: Categorization of interview results- Source:Python

The bar chart indicates that negative emotion prevails in the interview replies, with over 40 responses categorized as unfavorable. This indicates that a considerable proportion of participants voiced worries, unhappiness, or criticism about the issues addressed. The emphasis of the interviews on privacy and trust problems suggests that the pronounced negative attitude likely reflects respondents' perceptions of significant privacy threats, data abuse, or trust deficiencies in social media and consumer behavior.

5. Results and Discussion

This study investigated the influence of social media marketing on customer behavior and brand perception using a synthesis of theoretical concepts and unique survey-based research. The theoretical research indicated a notable increase in social media use and digital commerce in the Czech Republic over the last decade. In conjunction with this digital shift, influencer marketing, user-generated content, and data-driven customization have become pivotal in brand-consumer engagement. Consumers often exhibit more confidence in peer-shared material compared to conventional advertising, with digital word-of-mouth becoming a significant influence on attitudes and choices. The research's practical component included analyzing survey data to identify trends in customer behavior. The investigation indicated that those who voiced concern with tailored advertising or refrained from interacting with brand marketing on social media were much less inclined to make purchases affected by social content. The results indicate that trust and perceived relevance substantially influence customer responses to marketing initiatives on digital media.

The findings indicated that customers impacted by social media material and who prioritize online reviews tend to establish higher brand loyalty. This underscores the significance of perceived authenticity and peer validation in establishing enduring customer connections. Brands that effectively use these social components seem more adept at cultivating loyalty among their audience.

Ethical issues were obviously evident from the results. A significant correlation was shown between privacy concerns and unease over targeted advertising, indicating that data collection and use directly influence consumer trust. Moreover, customers that place significant importance on peer reviews are more inclined to engage with user-generated material, underscoring the need of trust and openness in the contemporary digital landscape. The study indicated that, in crisis management, customers who possess ethical awareness or privacy concerns tend to react negatively to brand disputes on social media. Boycotting activity and trust issues profoundly affected people' perceptions of brand activities amid crises. The models were proficient in identifying customers who responded vigorously to crises, but they were less adept at recognizing those who remained unaffected. The results together underscore the intricate aspects of social media impact. Genuine content, ethical procedures, and emotionally astute marketing significantly influence customer reactions. These observations are especially pertinent to information technology, as they highlight the need of developing ethical and responsive systems—such as AI-driven suggestions and privacy-aware analytics—that bolster trust and relevance in digital marketing.

6. Conclusion

The thesis examines the changing dynamics of social media marketing and its effects on consumer behavior, highlighting the roles of digital content, peer influence, and ethical factors in shaping purchase choices and brand image. The amalgamation of theoretical and practical viewpoints substantiated that consumer trust is progressively linked to authenticity, transparency, and relevance in digital communication. As organizations shift from conventional advertising to customized and influencer-centric techniques, customers are increasingly scrutinizing issues related to privacy and data ethics.The research highlights the dual function of developing technology in contemporary marketing. AI-driven analytics, targeted content distribution, and automated interaction tools like chatbots improve marketing efficiency and consumer experience. Conversely, these technologies provoke substantial ethical concerns over algorithmic bias, spying, and data security. The results indicated that privacy issues directly affect consumer comfort with targeted advertising and crisis response behavior, illustrating that confidence in digital networks is not only a technical matter but a strategic one.

The use of statistical models, data processing using Python, and machine learning principles to examine customer behavior illustrates that IT serves not just as an analytical tool, but as a fundamental catalyst for marketing change. As predictive algorithms become more sophisticated, the need for ethical AI practices and enhanced cybersecurity frameworks also increases. Contemporary consumers possess heightened awareness about the collection and use of personal data; neglecting to address these apprehensions may compromise the efficacy of even the most sophisticated digital marketing. This research concludes that the future of successful social media marketing resides at the convergence of data-driven innovation and ethical IT practices. Businesses must employ AI and platform analytics to comprehend and engage people while ensuring the deployment of these technologies is characterized by fairness, transparency, and stringent data security. Establishing customer confidence in the digital era will increasingly rely on the effective collaboration of IT and marketing experts to develop systems that are both intelligent and responsible.

7. References

- Abbott, D. (2014) *Applied Predictive Analytics: Principles and Techniques for the Professional Data Analyst*. Newark: John Wiley & Sons, Incorporated. Available from: ProQuest Ebook Central, pp. 5, 10, 12, 27, 36.
- Agility PR Solutions (2024) 'Consumers are growing wary of AI bias and will reward businesses that eliminate it'. Available at: https://www.agilitypr.com/pr-news/public-relations/consumers-are-growing-wary-of-ai-bias-and-will-reward-businesses-that-eliminate-it/.
- Backlinko. (2025) *X (Twitter) Statistics: How Many People Use X? (2025)*. Available at: https://backlinko.com/twitter-users.
- Bianchi, A. (2020) *Driving Consumer Engagement in Social Media: Influencing Electronic Word of Mouth*. Oxford: Taylor & Francis Group. ISBN 9780367422128.
- Cheung, C.M. and Thadani, D.R. (2012) 'The impact of electronic word-of-mouth communication: A literature analysis and integrative model', *Decision Support Systems*, 54(1), pp. 461–470.
- Cialdini, R.B. (2008) *Influence: Science and Practice*. 5th ed. Boston: Allyn & Bacon. Chapter 4, pp. 98–140. ISBN 9780205609994.
- Czech Journal of Economics and Finance. (2020) *E-commerce Expenditure Trends in the Czech Republic*. Available at: https://www.cjournal.cz/files/440.pdf.
- DataReportal. (2024) *Digital 2024: Czechia – Social Media Statistics*. Available at: https://datareportal.com/reports/digital-2024-czechia.
- De Veirman, M., Cauberghe, V. and Hudders, L. (2017) 'Marketing through Instagram influencers: The impact of number of followers and product divergence on brand attitude', *International Journal of Advertising*, 36(5), pp. 798–828.
- El País. (2025) *Bluesky contra Threads: así está la batalla por heredar el espacio de X*. Available at: https://elpais.com/tecnologia/2025-01-01/bluesky-contra-threads-asi-esta-la-batalla-por-heredar-el-espacio-de-x.html.
- Erkan, I. and Evans, C. (2016) 'The influence of eWOM in social media on consumers' purchase intentions: An extended approach to information adoption', *Computers in Human Behavior*, 61, pp. 47–55.
- European Parliament. (2020) *The impact of the General Data Protection Regulation (GDPR) on artificial intelligence*. Available at:

https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf.

- Evans, N.J., Phua, J., Lim, J. and Jun, H. (2017) 'Disclosing Instagram Influencer Advertising: The Effects of Disclosure Language on Advertising Recognition, Attitudes, and Behavioral Intent', *Journal of Interactive Advertising*, 17(2), pp. 138–149.
- Filieri, R., Hofacker, C.F. and Alguezaui, S. (2015) 'What makes online reviews helpful? A diagnosticity-adoption framework to explain informational and normative influences in eWOM', *Journal of Business Research*, 68(6), pp. 1261–1270.
- Godey, B., Pederzoli, D., Aiello, G., Donvito, R., Wiedmann, K-P. and Hennigs, N. (2016) 'Social media marketing efforts of luxury brands: Influence on brand equity and consumer behavior', *Journal of Business Research*, 69(12), pp. 5833–5841. DOI: 10.1016/j.jbusres.2016.04.181, p. 53.
- GO-Globe. (2024) *E-commerce Growth in Czech Republic: Market Trends and Projections*. Available at: https://www.go-globe.com/e-commerce-in-czech-republic-statistics.
- Heinonen, K. (2011) 'Consumer activity in social media: Managerial approaches to consumers’ social media behavior', *Journal of Consumer Behaviour*, 10(6), pp. 356–364. DOI: 10.1002/cb.376.
- Hong, T.-P., Serrano-Estrada, L., Saxena, A. and Biswas, A. (eds) (2022) *Deep Learning for Social Media Data Analytics*. Cham: Springer International Publishing. DOI: 10.1007/978-3-031-10869-3.
- Hootsuite. (2025) *2025 Facebook statistics every marketer needs [Original data]*. Available at: https://blog.hootsuite.com/facebook-statistics/.
- Hootsuite. (2025) *Social media benchmarks: Q1 2025 data + tips*. Available at: https://blog.hootsuite.com/social-media-benchmarks/.
- International Telecommunication Union (ITU). (2021) *Global Cybersecurity Index 2020*. Available at: https://www.itu.int/en/ITU-D/Cybersecurity/Documents/GCIv5/2401416_1b_Global-Cybersecurity-Index-E.pdf.
- Kaplan, A.M. and Haenlein, M. (2010) 'Users of the world, unite! The challenges and opportunities of Social Media', *Business Horizons*, 53(1), pp. 59–68. DOI: 10.1016/j.bushor.2009.09.003.

- Kotler, P. and Keller, K.L. (2016) *Marketing Management*. 15th ed. Harlow: Pearson Education. ISBN 9780133856460.
- Statista. (2023) *U.S. X reach by age group 2023*. Available at: https://www.statista.com/statistics/265647/share-of-us-internet-users-who-use-twitter-by-age-group/.
- Statista. (2024) *E-commerce share of total turnover in Czechia (2012–2024)*. Available at: https://www.statista.com/statistics/1016832/enterprises-e-commerce-share-total-turnover-czechia.
- Statista. (2024) *Global audience of X/Twitter 2024*. Available at: https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/.
- Statista. (2024) *Social media – statistics & facts*. Available at: https://www.statista.com/topics/1164/social-networks/.
- Statista. (2024) *X/Twitter: number of worldwide users 2019–2024*. Available at: https://www.statista.com/statistics/303681/twitter-users-worldwide/.
- Statista. (2025) *Biggest social media platforms by users 2025*. Available at: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/.
- Stephen, A.T. (2016) 'The role of digital and social media marketing in consumer behavior', *Current Opinion in Psychology*, 10, pp. 17–21. Available at: https://doi.org/10.1016/j.copsyc.2015.10.016.
- The Social Shepherd. (2025) *21 Essential Twitter (X) Statistics You Need to Know in 2025*. Available at: https://thesocialshepherd.com/blog/twitter-statistics.
- Eurostat. (2024) *E-commerce statistics*.Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=E-commerce_statistics.
- Vinerean, S., Cetina, I., Dumitrescu, L. and Tichindelean, M. (2013) 'The effects of social media marketing on online consumer behavior', *International Journal of Business and Management,* 8(14), pp. 66–77.
- Vlad, D.E. (2020) *Concepts of Quality Connected to Social Media and Emotions*. Wiesbaden: Springer Vieweg. ISBN 9783658288679.

8. List of pictures and tables

8.1 List of pictures

8.2 List of tables