

UGANDA CHRISTIAN UNIVERSITY, MUKONO, UGANDA

**PROVENANCE-AWARE SENTIMENT
ANALYSIS AND MULTI-CRITERIA
INFLUENCER SELECTION: A
DECISION-SUPPORT FRAMEWORK
DEMONSTRATED FOR UGANDA
AIRLINES**

A Thesis Submitted in Partial Fulfillment of the Requirements for the
Award of the Degree of

Master of Science in Data Science and Analytics

of the

Faculty of Engineering

Department of Computing and Technology

By

Jordan Micheal Senyondo

Supervisor: Dr. Daphne Nyachaki Bitalo

April, 2025

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific references in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: _____

DATE: _____

APPROVAL

Title of the Research Thesis:

Provenance-Aware Sentiment Analysis and Multi-Criteria Influencer Selection: A Decision-Support Framework Demonstrated for Uganda Airlines

Statement of Approval:

This research thesis has been reviewed and approved for submission in partial fulfillment of the requirements for the award of a Master's degree. The undersigned confirms that the work meets the required academic standards for scholarly research and is approved for implementation.

Supervisor's Details

Name: DR. DAPHNE NYACHAKI BITALO

Title: SENIOR LECTURER

Department: COMPUTING

Signature: 

Date: 2/03/2026

Dedication

To my family, for love, patience, and the steady encouragement that carried me through this journey.

To my mentors in the **Department of Computing And Technology, Uganda Christian University**, for teaching me to think clearly and build carefully.

And to everyone who believes that technology can make communication more human.

Acknowledgments

This thesis was written in the **Department of Computing And Technology** at **Uganda Christian University**, in partial fulfillment of the requirements for the **Degree**. Sincere appreciation is extended to my supervisor, **Dr. Daphne N. BITALO**, for strict guidance, expertise, and sustained support from proposal development to completion.

Appreciation is also extended to faculty and staff in the Department for building an environment in which critical inquiry and effective practice reinforce one another. Gratitude is due to colleagues and friends who read drafts, questioned assumptions, and assisted with troubleshooting experiments; their time and honesty strengthened this thesis.

The communities behind open research and open-source tools are also acknowledged for making reproducible analysis feasible. Any errors or omissions remain the author's responsibility.

Finally, sincere thanks are due to family and loved ones for patience, prayers, and constant encouragement throughout this journey.

Abstract

Brands increasingly rely on creators to reach audiences, yet follower scale alone does not represent brand fit, audience relevance, engagement quality, or the emotional context of public discourse. This thesis develops and demonstrates a provenance-aware decision-support framework that combines descriptive sentiment context with evidence-aware, multi-criteria influencer ranking. Uganda Airlines provides the management scenario, while the empirical baseline is an observed legacy aviation-discourse dataset rather than a complete record of Uganda Airlines’ own social-media activity.

After tweet-ID deduplication, the observed baseline contains 14,485 unique tweets collected from 17 to 24 February 2015. Of these records, 62.70% are negative, 21.19% neutral, and 16.11% positive. A separate scenario layer contains 125,200 simulated records covering 2015–2026. The resulting 139,685-record union is therefore described as *simulation-augmented*; its longitudinal patterns are scenario results and not independently observed trends. The influencer campaign panel contains 683 observed and 60,000 simulated rows across 95 name–platform profiles representing 37 distinct influencer names. Sixty-four profiles meet the default eligibility threshold of at least five observed posts.

Influencer selection is operationalised through the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS). The ranking criteria are value alignment, audience fit, sentiment context, engagement quality, reach, and cost efficiency. The integrated top ten has no profiles in common with a reach-only top ten and two in common with a follower-count-only top ten, showing that the implemented ordering differs from scale-only selection. Criterion ablations identify engagement quality and value alignment as the most influential preferences in the current ordering. These are deterministic ranking results, not campaign-effect estimates. The simulated layer is reserved for stress testing rather than increasing empirical confidence.

The thesis contributes (i) a reusable, values-aware framework for linking social listening to influencer selection; (ii) a transparent method for separating observed evidence from simulation-based scenario analysis; and (iii) a working dashboard that exposes provenance, evidence thresholds, criterion contributions, and ranking sensitivity. The findings support managerial exploration and monitoring, but do not establish causal campaign uplift or externally validated real-world predictive accuracy.

Keywords: social listening; sentiment analysis; simulation augmentation; brand–creator alignment; influencer selection; TOPSIS; decision-support dashboard.

Table of Contents

Author's declaration	i
Dedication	iii
Acknowledgments	iv
Abstract	v
List of Tables	xi
List of Figures	xii
1 Introduction	1
1.1 Background to the Study	1
1.1.1 Global perspective	1
1.1.2 African perspective	2
1.1.3 East African perspective	3
1.1.4 Ugandan and Uganda Airlines perspective	4
1.1.5 Alignment with sustainable-development and national priorities	4
1.2 Problem Statement	5
1.3 Purpose of the Study	6
1.4 Rationale for the Study	6
1.5 Research Questions	6
1.6 Research Hypotheses	7
1.7 Objectives	8
1.8 Scope and Delimitations	8
1.8.1 Analytical Scope	8
1.8.2 Geographical Scope	9

1.8.3	Temporal Scope	9
1.8.4	Delimitations	9
1.9	Significance and Expected Contributions	9
1.9.1	Theoretical Significance	9
1.9.2	Methodological Significance	9
1.9.3	Practical Significance	10
1.9.4	Policy and Ethical Significance	10
1.10	Originality of the Study	10
1.11	Conceptual and Analytical Framework	11
1.12	Operational Definitions and Assumptions.	11
1.13	Ethical Considerations (Summary)	13
1.14	Structure of the Thesis	13
1.15	Mapping of Objectives, Research Questions, and Thesis Sections	14
2	Literature Review	15
2.1	Overview and Review Strategy	15
2.2	Influencer Credibility, Congruence, and Audience Fit	16
2.2.1	From scale to fit	16
2.2.2	Identification and engagement quality	16
2.3	Sentiment Analysis for Brand Discourse	17
2.3.1	Tasks and modelling paradigms	17
2.3.2	Evaluation, calibration, and the Brier score	18
2.4	Airline, Tourism, Brand Perception, and Engagement	18
2.5	Multi-Criteria Influencer Selection and Decision Support	19
2.6	African, East African, and Ugandan Context	20
2.6.1	African evidence and transfer limits	20
2.6.2	East African evidence	20
2.6.3	Ugandan policy and case positioning	20
2.7	Synthesis of the Literature	21
2.8	Research Gap and Conceptual Positioning	22
2.9	Implications for the Current Study	23
3	Research Design and Methodology	24
3.1	Research Design and Evidence Boundary	24
3.2	Case Setting, Units of Analysis, and Variables	24
3.3	Objective-to-Method Alignment	25

3.4	Data Sources, Sampling, and Immutability	26
3.4.1	Immutable source files and provenance	26
3.4.2	Canonical discourse cleaning	27
3.5	Methods by Specific Objective	27
3.5.1	Objective 1: To describe the sentiment distribution and supplied topic structure in the observed aviation-discourse baseline	27
3.5.2	Objective 2: To examine profile-level value-alignment, audience-fit, sentiment-context, engagement-quality, reach, and cost indicators in the observed campaign panel	28
3.5.3	Objective 3: To develop and demonstrate an evidence-aware TOP- SIS shortlist that incorporates sentiment context, brand-value compatibility, audience fit, engagement quality, reach, and cost efficiency, and to compare it with transparent ranking baselines	28
3.5.4	Objective 4: To examine source-mixture and criterion sensitivity and to present observed evidence and explicitly labelled scenario results separately in a management-oriented dashboard	28
3.6	Operational Analytical Pipeline	29
3.7	Simulation-Augmented Scenario Construction	31
3.8	Influencer Profiles, Eligibility, and TOPSIS	31
3.8.1	Evidence tiers and mode-specific profiles	31
3.8.2	TOPSIS decision rule	32
3.8.3	Ranking baselines and criterion ablations	32
3.9	Evaluation Protocols and Validation Boundaries	33
3.9.1	Brier-score justification and non-reporting decision	33
3.10	Dashboard and Reproducibility	34
3.11	Validity, Reliability, Ethics, and Limitations	34
4	Results and Findings	35
4.1	Overview	35
4.2	Data Provenance and Analysis Inputs	35
4.2.1	Canonical source counts	36
4.3	Results for Objective 1	37
4.3.1	Observed sentiment baseline	37
4.3.2	Simulated and combined scenario comparison	39
4.4	Results for Objective 2	39

4.5	Results for Objective 3	41
4.5.1	Comparison with transparent ranking baselines	42
4.6	Results for Objective 4	43
4.6.1	Criterion-ablation sensitivity	43
4.6.2	Source-mixture sensitivity	44
4.6.3	Monitoring and validation boundaries	44
4.6.4	Pending-reproduction holdout metrics	45
4.6.5	Dashboard-ready outputs	45
4.6.6	Direct answers to the research questions	47
5	Discussion	49
5.1	Purpose and Evidence Boundary	49
5.2	Interpretation by Objective and Research Question	50
5.3	Relationship to Prior Literature and Implications	53
5.4	Hypotheses Evaluation	54
5.5	Theoretical, Methodological, and Managerial Implications	55
5.6	Limitations and Threats to Validity	55
5.7	Future Work	56
5.8	Conclusion	57
6	Conclusion and Recommendations	58
6.1	Purpose, Scope, and Evidence Boundary	58
6.2	Summary of Findings by Objective	58
6.3	Conclusions	59
6.4	Contributions	60
6.5	Recommendations	61
6.5.1	Current dashboard use	61
6.5.2	Production-validation steps	61
6.5.3	Guidance for other brands	62
6.6	Limitations and Boundary Conditions	62
6.7	Future Research Directions	63
6.8	Closing Remarks	63
	References	64
	Data, Analysis, and Dashboard Access Register	68
	Artefact Register	68

[Dashboard Entry Point](#) 68

List of Tables

Table	Page
1.1 Alignment of objectives, research questions, methods, and results	14
2.1 Thematic synthesis and implications for the study	22
3.1 Full objective–research-question–method–result alignment	26
3.2 Operational pipeline from source records to dashboard outputs	30
4.1 Chapter 4 data and analysis artefacts	36
4.2 Generated provenance summary	37
4.3 Sentiment by provenance mode	38
4.4 Observed influencer evidence tiers	39
4.5 Observed default TOPSIS top ten	41
4.6 Integrated ranking compared with equal-weight and scale-only baselines . . .	42
4.7 One-criterion-at-a-time TOPSIS sensitivity	43
4.8 TOPSIS sensitivity under observed/simulated source-mixture scenarios . . .	44
4.9 Pending-reproduction simulation holdout metrics	45
4.10 Objective 4 dashboard signals derived from regenerated evidence	46
5.1 Bounded interpretation of the main findings	54
5.2 Assessment of the Chapter 1 ranking hypotheses	54
6.1 From current decision support to production validation	62
1 Access register for data and analysis artefacts	68

List of Figures

Figure	Page
1.1 Conceptual pipeline for the parameterised influencer marketing framework: brand inputs and data sources feed into data ingestion and preprocessing, sentiment and topic modelling, influencer–brand alignment scoring, and the decision and monitoring layer.	5
1.2 Conceptual decision-support model: observed indicators inform latent constructs and a composite selection score, which produces ranked recommendations with evidence, criterion, and scenario-monitoring outputs. The model does not estimate causal campaign outcomes.	12
4.1 Sentiment counts by provenance mode.	38
4.2 Observed profile counts by evidence tier.	40

Chapter 1

Introduction

1.1 Background to the Study

1.1.1 Global perspective

Influencer marketing has developed from celebrity-style endorsement into a creator ecosystem in which credibility, message value, product–creator congruence, audience identification, and content relevance shape how sponsored communication is received (Lou and Yuan 2019; Schouten et al. 2020; Sokolova and Kefi 2020). Experimental, survey, and large-scale review evidence shows that follower count can signal popularity, but does not by itself establish trust, brand fit, or purchase response (De Veirman et al. 2017; Weismueller et al. 2020; Belanche et al. 2021). Strategic influencer selection is consequently a multi-dimensional decision involving the source, message, audience, brand context, and disclosure environment rather than a search for the largest account (Hudders et al. 2021; Harrigan et al. 2021; Leung et al. 2022).

This shift also changes how marketing evidence should be interpreted. Credibility and congruence studies commonly examine attitudes, trust, engagement, or purchase intention under particular samples and platform conditions; their coefficients are not universal weights for every brand or market (Lou and Yuan 2019; Sokolova and Kefi 2020; Belanche et al. 2021). Reviews of influencer research similarly identify substantial variation in theories, measures, platforms, and outcomes, which makes transparent criteria and sensitivity analysis important when research findings are converted into a decision rule (Hudders et al. 2021; Tanwar et al. 2022; Joshi et al. 2025).

In parallel, sentiment analysis provides methods for organising public text into interpretable opinion signals, ranging from polarity classification to aspect-level analysis (Pang and Lee 2008; Liu 2012). Airline studies use social-media posts and brand content to examine passenger sentiment and engagement, but reliable interpretation requires traceable sampling, valid labels, suitable baselines, and attention to platform and domain language (Menon et al. 2019; Wu and Gao 2024). A descriptive sentiment distribution therefore answers a different question from classifier accuracy, probability calibration, causal campaign uplift, or production readiness.

Tourism and destination studies further show that social-media communication can shape engagement, electronic word of mouth, destination image, and travel planning, while the strength and direction of those relationships depend on the population and research design (Huerta-Álvarez et al. 2020; Sakshi et al. 2020). This evidence makes social-media analytics relevant to aviation and tourism communication, but it does not justify transferring an estimated effect or decision weight directly into a Uganda Airlines setting.

1.1.2 African perspective

African digital communication must be interpreted in a linguistically and institutionally diverse environment. The AfriSenti benchmark contains more than 110,000 native-speaker-annotated tweets across 14 African languages and documents the scarcity of high-quality labelled resources relative to the continent’s linguistic diversity (Muhammad et al. 2023). Its inclusion of Swahili and Kinyarwanda demonstrates progress toward locally grounded sentiment evaluation, while also showing that performance must be assessed per language and dataset rather than assumed from English-language benchmarks.

African tourism evidence also reinforces the importance of context. Research in an emerging destination setting links tourist- and organisation-generated social-media communication with brand equity and engagement, while South African work on tourism-oriented influencer video shows that source and content characteristics operate within a specific destination and audience environment (Huerta-Álvarez et al. 2020; Matiza and Slabbert 2024). These studies support the relevance of credibility, engagement, and destination communication, but they do not establish that the same indicators have identical meanings or effects across African aviation markets.

The broader influencer literature remains geographically concentrated outside Africa,

and many studies rely on cross-sectional intention measures rather than observed campaign outcomes (Hudders et al. 2021; Tanwar et al. 2022; Joshi et al. 2025). The African research need is therefore twofold: more locally observed evaluation and more transparent reporting of how public-data proxies are transformed into marketing recommendations. This thesis addresses the second need as a decision-support and evidence-governance problem; it does not claim to solve the first through simulated data.

1.1.3 East African perspective

East Africa introduces a more specific combination of multilingual communication, tourism dependence, mobile-first interaction, and limited task-specific language resources. Swahili is widely used across the region, yet Swahili NLP research has been constrained by data scarcity. SwahBERT was developed from a Swahili-specific corpus and outperformed multilingual BERT on most evaluated downstream tasks, illustrating why language-specific resources can matter (Martin et al. 2022). AfriSenti likewise includes native-speaker-annotated Swahili data, but neither resource validates an English-language airline classifier for Ugandan deployment (Muhammad et al. 2023).

Regional tourism research provides a second strand of evidence. A study of 278 travellers sampled at major Tanzanian airports found that affective and behavioural customer engagement were positively related to different forms of tourism electronic word of mouth (Kanje et al. 2020). This supports treating engagement as multidimensional and context-dependent rather than equating raw interaction volume with influence. It also strengthens the connection between aviation gateways, tourism communication, and social-media engagement without implying that an influencer shortlist causes tourist arrivals or commercial returns.

Taken together, the East African literature supports locally validated language resources, contextual engagement measures, and cautious transfer of international influencer models. It also exposes the absence of a traceable Uganda Airlines campaign dataset containing verified audience attributes, calibrated sentiment probabilities, and downstream campaign outcomes. The present study therefore uses explicit provenance, evidence thresholds, and sensitivity analysis to demonstrate a bounded framework rather than claiming regional predictive validity.

1.1.4 Ugandan and Uganda Airlines perspective

Uganda Vision 2040 identifies tourism, air transport, strategic marketing, research, and ICT infrastructure as components of national transformation (National Planning Authority 2013). Within that policy environment, social-media communication is relevant to destination image, service reputation, and travel planning, but international and regional findings remain conditional on their samples, platforms, and outcome measures (Kanje et al. 2020; Huerta-Álvarez et al. 2020; Sakshi et al. 2020). Uganda Airlines is therefore used as a management scenario, not as the verified source of the complete observed discourse corpus and not as an organisation whose existing marketing performance has been audited by this study.

For a Ugandan aviation decision maker, the practical problem is how to review sentiment context, brand-value compatibility, audience-fit proxies, engagement quality, reach, and cost without reducing the decision to follower count. This study responds by developing a parameterised framework that a brand can instantiate by supplying two inputs:

1. a brand-values vector (for example, safety, reliability, affordability, sustainability);
and
2. an industry taxonomy (for example, aviation versus cosmetics).

The pipeline then (a) summarises the available sentiment and topic fields and (b) ranks eligible influencer profiles using documented decision criteria. The empirical component uses an observed legacy aviation-discourse baseline and an observed campaign panel, while a separately labelled simulated layer supports stress testing. This design responds to the local evidence constraint by making provenance and limitations visible; it does not substitute simulation for Ugandan campaign observation. Figure 1.1 summarises the end-to-end process from brand inputs to decision-support outputs.

1.1.5 Alignment with sustainable-development and national priorities

The study is aligned most directly with Sustainable Development Goal (SDG) 9 through its emphasis on transparent digital innovation and an auditable decision-support artefact

(United Nations Department of Economic and Social Affairs 2026b). It has a secondary connection to SDG target 8.9 because aviation and destination communication can support tourism-oriented economic activity (United Nations Department of Economic and Social Affairs 2026a). At national level, the framework is consistent with Uganda Vision 2040’s emphasis on tourism development, strategic marketing, ICT capability, and improvement of air-transport connectivity (National Planning Authority 2013). These are areas of potential contribution: the present thesis does not measure national development impact, employment creation, tourism growth, or transport performance.

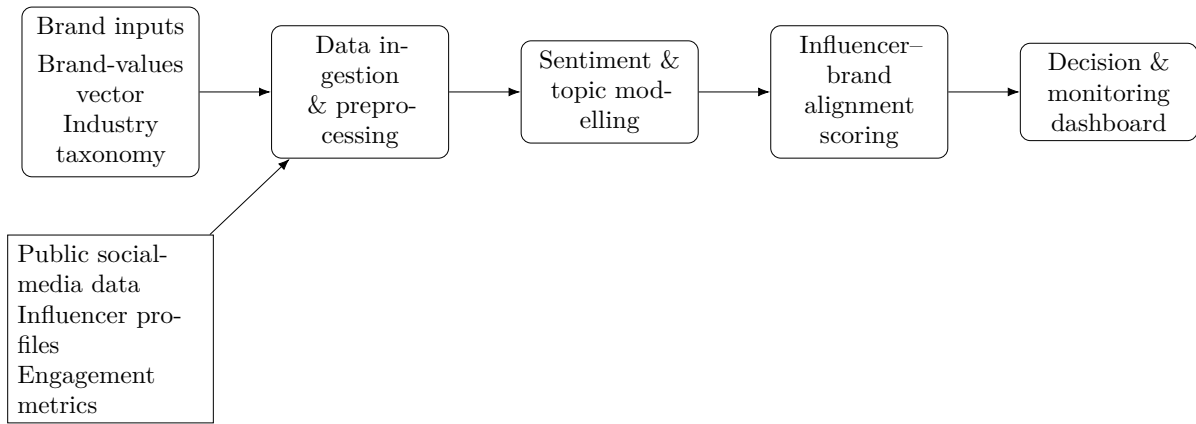


Figure 1.1: Conceptual pipeline for the parameterised influencer marketing framework: brand inputs and data sources feed into data ingestion and preprocessing, sentiment and topic modelling, influencer–brand alignment scoring, and the decision and monitoring layer.

1.2 Problem Statement

Influencer selection based only on follower counts or generic engagement rates obscures the different considerations involved in brand–creator fit. Research identifies credibility, congruence, content relevance, audience characteristics, and engagement quality as distinct influences on endorsement response (De Veirman et al. 2017; Schouten et al. 2020; Belanche et al. 2021; Harrigan et al. 2021). However, those findings do not by themselves provide a transparent, context-sensitive procedure for combining sentiment information with influencer-selection criteria, particularly where observed data are sparse and scenario data are also used.

The specific problem addressed by this thesis is therefore a decision and evidence-governance problem: how can public-discourse summaries and profile-level indicators

be converted into an auditable influencer shortlist without concealing data provenance, proxy limitations, user-selected preferences, or uncertainty? The problem is relevant to a reputation-sensitive aviation management scenario, but the available evidence cannot determine whether Uganda Airlines' existing influencer practice is ineffective or whether the proposed framework improves campaign outcomes. The study consequently focuses on descriptive evidence, reproducible ranking, sensitivity, and transparent presentation rather than causal uplift.

1.3 Purpose of the Study

The purpose of the study is to design, implement, and evaluate a parameterised framework that summarises available aviation-discourse sentiment, organises influencer-selection indicators, and shows how a multi-criteria shortlist responds to alternative assumptions. Uganda Airlines is used as the management scenario. Observed evidence establishes the descriptive baseline, while explicitly labelled simulation-augmented analysis supports stress testing rather than observed longitudinal inference.

1.4 Rationale for the Study

The rationale follows a direct chain from problem to intervention. First, scale-only selection omits constructs that the literature treats as relevant, including congruence, audience fit, and engagement quality. Second, sentiment studies and influencer-selection studies are often presented as separate analytical tasks, while evidence provenance and decision preferences are not always visible in the resulting recommendation. Third, an aviation management scenario makes these limitations important because public commentary is heterogeneous and commercial outcomes cannot be inferred from engagement proxies alone. The intervention is therefore an auditable pipeline that connects descriptive sentiment context, profile indicators, TOPSIS ranking, evidence thresholds, and scenario-labelled sensitivity outputs.

1.5 Research Questions

To align the analysis with the specific objectives in Section 1.7, the study is guided by four research questions, each corresponding to one objective.

RQ1 (linked to O1). What sentiment distribution and supplied topic structure characterise the observed aviation-discourse baseline?

RQ2 (linked to O2). What do the available observed campaign records show about profile-level value alignment, audience fit, sentiment context, engagement quality, reach, and cost indicators?

RQ3 (linked to O3). How can the six criteria be integrated in an evidence-aware TOPSIS shortlist, and how does that shortlist differ from equal-weight and scale-only baselines?

RQ4 (linked to O4). How sensitive is the ranking to source-mixture and criterion assumptions, and how can observed evidence and scenario results be distinguished in a management-oriented dashboard?

1.6 Research Hypotheses

The following analytical hypotheses guide the quantitative comparisons. They concern associations and ranking behaviour; they do not assert causal campaign uplift.

H1 (Integrated versus scale-only ranking). The integrated TOPSIS specification produces a materially different ordering from reach-only and follower-count-only rankings, assessed through top-ten overlap and Spearman rank correlation.

H2 (Criterion sensitivity). Removing a decision criterion can change the ordering produced by the integrated specification; the size of that change is assessed descriptively through top-ten overlap, leader stability, and Spearman rank correlation.

No hypothesis of causal campaign uplift or improved sentiment-classification accuracy is tested. A comparison between generic and domain-adapted sentiment classifiers is retained as future research because the current study lacks the traceable labelled observed evaluation set needed for that test.

1.7 Objectives

Main Objective

To develop a transparent decision-support framework that integrates provenance-aware sentiment summaries, profile-level decision indicators, and evidence-aware influencer ranking, demonstrated through a Uganda Airlines management scenario.

Specific Objectives

O1. To describe the sentiment distribution and supplied topic structure in the observed aviation-discourse baseline.

O2. To examine profile-level value-alignment, audience-fit, sentiment-context, engagement-quality, reach, and cost indicators in the observed campaign panel.

O3. To develop and demonstrate an evidence-aware TOPSIS shortlist that incorporates sentiment context, brand-value compatibility, audience fit, engagement quality, reach, and cost efficiency, and to compare it with transparent ranking baselines.

O4. To examine source-mixture and criterion sensitivity and to present observed evidence and explicitly labelled scenario results separately in a management-oriented dashboard.

1.8 Scope and Delimitations

1.8.1 Analytical Scope

The analysis uses public aviation-discourse text and influencer campaign records. The observed discourse baseline is drawn from Twitter, while the simulated scenario layer adds Twitter and Facebook records. Influencer profiles span Facebook, Instagram, Twitter, and YouTube. Text is the primary modality, supplemented by engagement, reach, cost, profile, and provenance metadata. No claim is made that every record concerns Uganda Airlines directly.

1.8.2 Geographical Scope

Uganda Airlines is the management case and the dashboard defaults to a Uganda-focused campaign scenario. The observed legacy corpus is a broader aviation-discourse baseline with user-supplied location fields; it is not a probability sample of Ugandan public opinion. The framework is designed to be reusable when the brand parameters and industry taxonomy are replaced.

1.8.3 Temporal Scope

The observed discourse baseline covers 17–24 February 2015 after tweet-ID deduplication. Observed campaign records cover 2023–2024. Simulated discourse and campaign records extend across 2015–2026 to support explicitly labelled longitudinal scenarios. Consequently, 2015–2026 trends are scenario results rather than observed evidence covering the entire period.

1.8.4 Delimitations

Advertisements that are paid for without giving credit to the creator are not included. Deep multimodal features such as audio and video are outside the analysis. The observed sources are convenience data rather than representative samples, and simulated records depend on their generation assumptions. The study therefore focuses on descriptive associations, ranking sensitivity, and decision support rather than causal inference (Leung et al. [2022](#); Harrigan et al. [2021](#)).

1.9 Significance and Expected Contributions

1.9.1 Theoretical Significance

The study connects influencer-congruence concepts with sentiment context and multi-criteria decision support (Schouten et al. [2020](#); Belanche et al. [2021](#)). It does not establish a new causal theory or estimate how these constructs affect campaign outcomes.

1.9.2 Methodological Significance

The methodological value lies in keeping observed and simulated evidence separate, applying observed-evidence eligibility thresholds, documenting TOPSIS criteria and

weights, and reporting ranking sensitivity. This makes the transformation from source records to shortlist more auditable.

1.9.3 Practical Significance

The dashboard provides marketing teams with a transparent decision tool for comparing creators, platforms, evidence levels, criterion contributions, and sensitivity scenarios. It supports campaign planning but does not promise a particular sentiment effect or replace managerial judgement (Morgan 2012; Grandhi et al. 2021).

1.9.4 Policy and Ethical Significance

The framework foregrounds disclosure, provenance, aggregated reporting, and the distinction between public fields and verified audience attributes. These controls support responsible interpretation but do not by themselves establish full legal or platform-policy compliance (Jin and Phua 2014; Kaplan and Haenlein 2010).

1.10 Originality of the Study

The originality of this study does not lie in proposing a new sentiment-classification algorithm or proving that influencer selection improves campaign outcomes. Its contribution is the integration of sentiment context, brand-value alignment, audience-fit indicators, engagement quality, reach, cost, and multi-criteria ranking within one provenance-aware decision-support framework. The framework separates observed evidence from simulated scenarios, prevents simulated volume from increasing a profile’s evidence tier, exposes the criteria and preferences that determine a shortlist, and presents assumption sensitivity in a reviewable dashboard.

This is principally a design, integration, and evidence-governance contribution. In the Uganda Airlines management scenario, it demonstrates how limited public-style evidence can be organised without representing proxy associations, simulated records, or deterministic rankings as causal campaign validation. The practical value is therefore transparency and reproducibility in shortlisting, while outcome effectiveness remains a separate question for future observed campaign research.

1.11 Conceptual and Analytical Framework

The theoretical framework connects the constructs on three levels:

Public discourse layer. The sentiment polarity and intensity and the topic of the aspects of services (e.g., safety, punctuality, and service quality) are mapped to category discourse.

Influencer layer. Creator characteristics include topical affinity, past audience sentiment, engagement-quality metrics, audience-segment fit and brand-value similarity (such as the cosine of creator content embeddings and brand-values keywords).

Decision layer. TOPSIS combines value alignment, audience fit, sentiment context, engagement quality, reach, and cost efficiency after placing all criteria in a documented direction and normalising user weights. Scores support platform-specific recommendations and scenario testing; they are not estimates of causal sentiment lift.

These three layers have been mapped as in Figure 1.2 into a measurement model that disaggregates observed indicators, latent constructs, the composite selection score and downstream outcomes.

1.12 Operational Definitions and Assumptions.

Influencer

An artist whose content is above-baseline engaging in a niche and has domestic authority (Harrigan et al. [2021](#)).

Brand-value Alignment

Semantic proximity between creator content and brand values, operationalised via keyword and integrating similarity with a check of manual codes (Belanche et al. [2021](#)).

Sentiment

Positive, neutral, or negative valence. In the present analysis these are supplied source labels that are summarised descriptively; the thesis does not claim to have reproduced the classifier that generated them (Pang and Lee [2008](#); Liu [2012](#)).

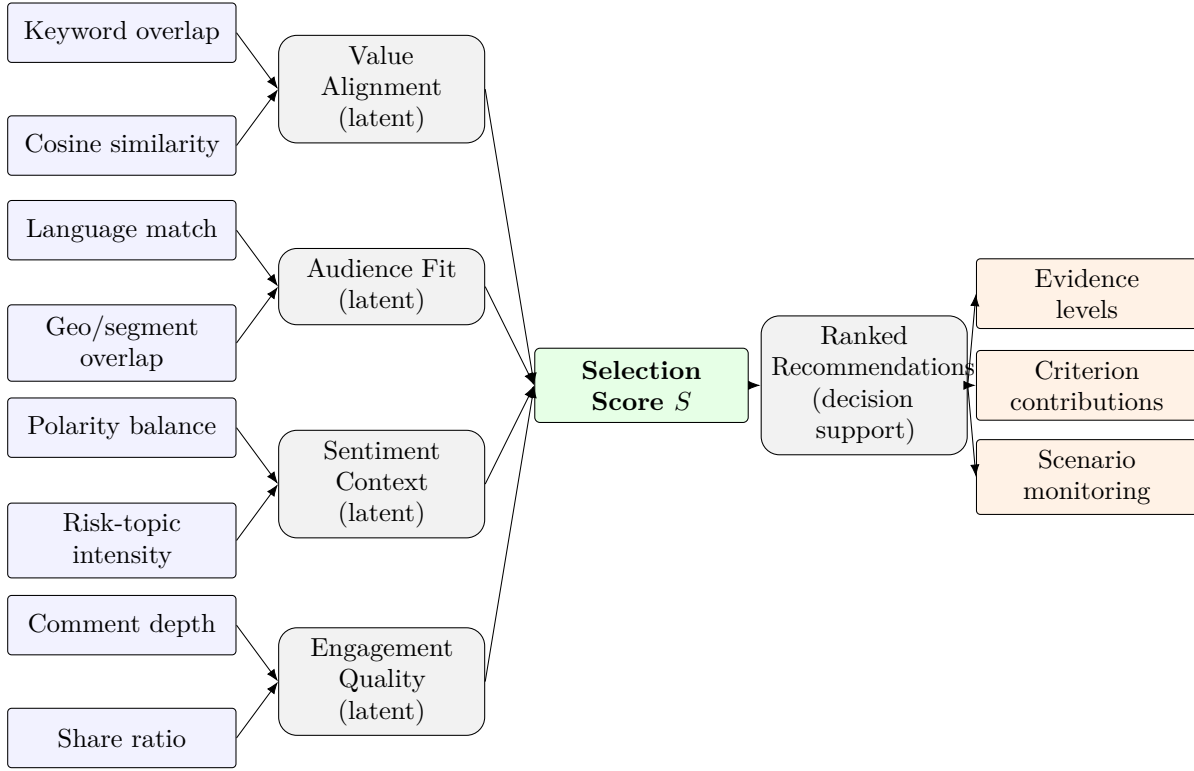


Figure 1.2: Conceptual decision-support model: observed indicators inform latent constructs and a composite selection score, which produces ranked recommendations with evidence, criterion, and scenario-monitoring outputs. The model does not estimate causal campaign outcomes.

Engagement Quality

A normalised measure of significant interactions (depth of comments and share ratio, along with emotion in responses) instead of bare vanity values like likes.

Observed Dataset

Records declared as `source=real` after validation. When tweet IDs repeat, the rule retains the first complete row per tweet ID. Observed results are the primary empirical evidence in the thesis.

Simulated Dataset

Records declared as `source=synthetic` or `source=synthetic_cross_platform`. These records are generated for scenario analysis and stress testing and are not described as

collected social-media observations.

Simulation-Augmented Dataset

The union of the observed and simulated datasets. Any statistic calculated partly or wholly from this union is reported as a scenario result rather than independent empirical validation. Simulation augmentation supports stress testing rather than independent empirical confirmation.

Assumptions

The observed sources are assumed to provide a useful directional aviation-discourse baseline, not a statistically representative estimate of all customers or Ugandan residents. Public-field location and audience variables are treated as imperfect proxies. Simulation is assumed to be useful for controlled stress testing only to the extent that its documented generation rules preserve relevant source distributions. Bot and spam filtering may reduce noise but cannot eliminate systematic bias.

1.13 Ethical Considerations (Summary)

The research uses supplied public-style records and presents results in aggregate. Working data use anonymised or hashed identifiers where available, and the analysis does not infer private audience attributes. Platform terms, data-protection obligations, sponsorship disclosure, and data minimisation remain requirements for any future live collection or production deployment (Jin and Phua [2014](#); Kaplan and Haenlein [2010](#)).

1.14 Structure of the Thesis

The remainder of the thesis is organised as follows:

- **Chapter Two** reviews literature on influencer marketing, sentiment analysis, and aviation branding, structured to lead from global perspectives to regional insights and then to the Ugandan context.
- **Chapter Three** presents the research design, case setting, data sources, processing pipeline, models, and evaluation procedures.

- **Chapter Four** reports results that are explicitly organised around the four research objectives and corresponding research questions.
- **Chapter Five** discusses the findings in relation to the literature, evaluates the hypotheses, and outlines theoretical and managerial implications as well as limitations and areas for further work.
- **Chapter Six** concludes the thesis and provides recommendations and an implementation roadmap for Uganda Airlines and other brands.

1.15 Mapping of Objectives, Research Questions, and Thesis Sections

Table 1.1 provides an explicit chain from each objective and research question to the data, method, and result that answer it. This matrix is intentionally organised by evidence rather than by promises about downstream campaign effects.

Table 1.1: Alignment of objectives, research questions, methods, and results

Obj.	Question and evidence	Method	Result and permitted conclusion
O1	RQ1; 14 485 observed discourse posts.	Tweet-ID deduplication and descriptive sentiment/topic summaries.	Section 4.3: distribution and supplied topic field for the retained 2015 baseline; no classifier-accuracy or later-trend claim.
O2	RQ2; 683 observed campaign rows across 95 profiles.	Profile aggregation, evidence tiers, and descriptive correlations.	Section 4.4: proxy patterns and evidence coverage; no audience-membership or causal-effect claim.
O3	RQ3; 64 eligible observed profiles.	TOPSIS, equal-weight comparison, scale-only baselines, and criterion ablations.	Section 4.5: reproducible relative shortlist and ranking sensitivity; no probability-of-success claim.
O4	RQ4; observed, simulated, and combined analytical modes.	Source-mixture sensitivity, provenance-labelled stress tests, and dashboard implementation.	Section 4.6: transparent scenario and dashboard outputs; no observed longitudinal or production-performance claim.

Chapter 2

Literature Review

2.1 Overview and Review Strategy

This chapter synthesises literature that directly supports the study’s four objectives. The review is organised around five themes: influencer congruence and credibility; sentiment analysis for brand discourse; airline, tourism, and destination engagement; multi-criteria decision support; and the African, East African, and Ugandan application context. Rather than listing studies in sequence, each section compares what the literature establishes, the methods used, the limitations that remain, and the implication for the present study.

Priority is given to peer-reviewed research that operationalises influencer fit, sentiment classification, airline or tourism social-media analysis, or decision modelling. Global reviews and experiments establish the principal constructs; African and East African studies are then used to examine contextual transfer, language resources, engagement, and tourism communication. Foundational methodological sources define established concepts or evaluation procedures. Official United Nations and Government of Uganda sources are used only for policy alignment; they are not treated as empirical proof of campaign effectiveness. This separation is important because the study’s observed corpus is a limited aviation-discourse baseline, not a probability sample of Ugandan opinion or a complete archive of Uganda Airlines communication.

2.2 Influencer Credibility, Congruence, and Audience Fit

2.2.1 From scale to fit

Influencer marketing differs from conventional celebrity endorsement partly because creators can develop sustained identification and perceived proximity with their audiences. However, the commercial value of that relationship is conditional. De Veirman et al. show that follower count affects perceived popularity but interacts with product divergence and brand attitude (De Veirman et al. 2017). Schouten et al. find that identification, credibility, and product–endorser fit help explain endorsement response (Schouten et al. 2020). Belanche et al. extend this reasoning by treating congruence among influencer, product, and consumer as a central mechanism (Belanche et al. 2021). Message value and source credibility also contribute to trust in branded influencer content, while credibility and parasocial interaction can shape purchase intentions under specified platform and sample conditions (Lou and Yuan 2019; Sokolova and Kefi 2020).

The shared implication is that reach is not interchangeable with relevance. Advertising disclosure and the source’s perceived expertise, trustworthiness, and attractiveness further affect endorsement evaluation (Weismueller et al. 2020). Nevertheless, these studies mostly estimate attitudes or intentions in specified experimental or survey settings. Reviews of the field show variation in source, message, audience, platform, method, and outcome definitions (Hudders et al. 2021; Tanwar et al. 2022; Joshi et al. 2025). They do not establish that a particular weighted formula will transfer unchanged to an East African airline campaign. Accordingly, this thesis operationalises value alignment and audience fit as transparent decision inputs rather than validated causal drivers.

2.2.2 Identification and engagement quality

Influencer-identification research increasingly combines network, content, and engagement information instead of using a single popularity threshold (Harrigan et al. 2021). Marketing-effectiveness research also distinguishes exposure from the quality of interaction and the organisational capability required to translate analytics into performance (Leung et al. 2022; Morgan 2012). Tourism-oriented influencer research in South Africa similarly treats the influencer and the content as situated persuasive cues rather than assuming that audience size alone determines travel response (Matiza and Slabbert 2024).

This supports including engagement quality alongside reach, while cautioning against treating likes, comments, shares, or a composite score as a direct measure of persuasion, conversion, or return on investment.

The literature therefore supports a multi-criteria shortlist, but it leaves the weights as a decision preference requiring disclosure and sensitivity analysis. That limitation motivates Objective 3 and the comparison between the integrated ranking, equal-weight TOPSIS, and scale-only baselines.

2.3 Sentiment Analysis for Brand Discourse

2.3.1 Tasks and modelling paradigms

Sentiment analysis commonly assigns text to positive, negative, or neutral classes, while aspect-based analysis links sentiment to specific entities or service attributes (Pang and Lee 2008; Liu 2012). Lexicon-based methods are interpretable and inexpensive but may struggle with domain vocabulary, negation, sarcasm, and code-switching. Supervised models can learn contextual boundaries from labelled examples, but their validity depends on label quality and similarity between training and deployment data. Hybrid and transfer-learning approaches attempt to combine domain knowledge with statistical modelling.

African-language research demonstrates why this transfer problem is material. AfriSenti provides native-speaker annotations for more than 110,000 tweets in 14 African languages, including Swahili and Kinyarwanda, and reports language-specific benchmark experiments (Muhammad et al. 2023). SwahBERT uses a Swahili-specific corpus and outperforms multilingual BERT on most of its evaluated downstream tasks, supporting the broader principle that locally appropriate language resources can outperform generic multilingual transfer under some conditions (Martin et al. 2022). Neither study validates the supplied English-language labels used in this thesis; instead, they strengthen the case for future Ugandan and East African language-specific evaluation.

Domain adaptation is consequently a plausible design direction, not an automatic improvement. A claim that an aviation-adapted model is better than a generic model requires the same traceable observed test set, an identified baseline, a fixed split, and uncertainty-aware comparison. The present study does not possess that evaluation design; it therefore reports supplied sentiment labels descriptively and retains domain-adaptation

evaluation as future work.

2.3.2 Evaluation, calibration, and the Brier score

Accuracy alone can conceal poor minority-class performance. Macro-averaged F1 and Matthews correlation coefficient are useful complements when sentiment classes are imbalanced (Matthews 1975). When a model produces probabilities that will inform decisions, calibration is a separate concern: a score of 0.8 should correspond to an event occurring approximately 80% of the time among comparable predictions (Guo et al. 2017).

For a binary outcome, the Brier score is the mean squared distance between predicted probability and observed outcome,

$$\text{BS} = \frac{1}{N} \sum_{i=1}^N (p_i - y_i)^2,$$

where p_i is the predicted probability and $y_i \in \{0, 1\}$ is the observed label (Brier 1950). Lower values are better and zero is perfect. Its value must be interpreted relative to a transparent reference forecast, such as the observed outcome prevalence, because a raw score has no universal adequacy threshold. For multiclass sentiment, the squared error is summed across class probabilities and the scoring convention must be stated. The present evidence register does not contain reproducible record-level holdout probabilities and labels, so this thesis does not report a Brier result. It identifies Brier score as an appropriate future calibration measure once those inputs are available.

2.4 Airline, Tourism, Brand Perception, and Engagement

Airline social-media research demonstrates the analytical value of customer posts while also illustrating the difference between descriptive monitoring and outcome validation. Menon et al. compare design and content factors associated with airline brand-post engagement on Facebook and Twitter, showing that engagement patterns depend on platform and message characteristics (Menon et al. 2019). Wu and Gao process airline-related tweets using machine-learning and lexical analysis and examine abrupt changes in sentiment over time (Wu and Gao 2024). Elemonie discusses social media in the formation

of airline image, reinforcing the reputation-management relevance of digital discourse (Elemonie 2020).

Tourism research broadens this evidence base. In Tanzania, affective and behavioural customer engagement were associated with different forms of electronic word of mouth among 278 travellers sampled at major airports (Kanje et al. 2020). Research in an emerging destination context connects tourist- and organisation-generated social-media communication with brand equity and brand engagement, while social-media travel-planning research treats actual platform use as a product of several behavioural and informational factors (Huerta-Álvarez et al. 2020; Sakshi et al. 2020). These findings make engagement, destination image, and electronic word of mouth relevant to aviation-linked marketing decisions.

Together, the airline and tourism studies support the use of social text for service, reputation, and destination analysis, but they do not justify treating every aviation-related post as brand-specific evidence. Nor do they demonstrate that a creator recommended by a multi-criteria score will cause higher bookings or improved sentiment. This distinction is particularly important here because the observed discourse source covers a short legacy airline window and is not wholly Uganda Airlines-specific.

2.5 Multi-Criteria Influencer Selection and Decision Support

Influencer shortlisting contains competing benefit and cost criteria. Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) provides a transparent way to normalise criteria, apply declared preferences, identify ideal-best and ideal-worst profiles, and rank each alternative by relative closeness to those ideals (Hwang and Yoon 1981). Its strength is auditability: criteria, direction, normalisation, and weights can be documented. Its limitation is preference dependence. A stable computation is not necessarily a valid commercial decision if its input measures or weights do not represent the intended construct.

For that reason, decision support should include at least three safeguards. First, evidence eligibility should depend on observed records rather than simulated volume. Second, changes in weights, criteria, and source mixture should be reported as sensitivity analyses. Third, a TOPSIS coefficient should be described as a relative ranking value, not a

probability of campaign success. These principles connect the literature to Objectives 2–4.

2.6 African, East African, and Ugandan Context

2.6.1 African evidence and transfer limits

The global literature provides concepts and methods, but direct transfer to African settings is limited by language variation, data access, platform coverage, and the scarcity of openly documented labelled corpora. AfriSenti addresses part of the language-resource gap through native-speaker annotation across 14 languages, yet its language-specific benchmarks also caution against assuming uniform model performance across the continent (Muhammad et al. 2023). African tourism-influencer research shows the relevance of creator cues and destination communication, but remains tied to its national, platform, and outcome context (Matiza and Slabbert 2024). The appropriate conclusion is therefore a context-specific validation gap, not a claim that African organisations lack social-media marketing activity.

2.6.2 East African evidence

East African evidence is directly relevant in two areas. First, Swahili-specific modelling shows that local corpora and language models can improve downstream NLP performance relative to generic multilingual models in the evaluated tasks (Martin et al. 2022). Second, Tanzanian tourism evidence shows that customer engagement is multidimensional and relates differently to seeking, giving, and passing electronic word of mouth (Kanje et al. 2020). These studies justify attention to language, context, and engagement quality, but they do not provide a validated weighting scheme for influencer selection or a Uganda Airlines campaign counterfactual.

2.6.3 Ugandan policy and case positioning

Uganda Vision 2040 places tourism, air-transport connectivity, strategic marketing, research, and ICT infrastructure within the country’s development direction (National Planning Authority 2013). SDG 9 provides a broader policy rationale for responsible digital innovation, while SDG target 8.9 links tourism policy with employment, culture, and local products (United Nations Department of Economic and Social Affairs 2026b; United Nations Department of Economic and Social Affairs 2026a). These policies

establish relevance for a transparent aviation decision-support artefact. They do not validate the framework’s predictions or establish national-development impact.

Uganda Airlines is therefore positioned as a management scenario rather than a claim of proprietary organisational diagnosis. The practical research question is whether observed and scenario data can be separated while sentiment context and influencer criteria are brought into one inspectable process. Any production or impact claim would require Uganda Airlines-specific observed data, documented campaign decisions, and independently measured outcomes.

2.7 Synthesis of the Literature

Table [2.1](#) compares the principal evidence streams and shows how their limitations shape the current design.

Table 2.1: Thematic synthesis and implications for the study

Theme	What the literature supports	Unresolved limitation	Response in this study
Influencer fit	Global experiments and reviews associate credibility, congruence, message value, and identification with attitudes, trust, or intentions.	Effects and measures are context-dependent; estimated relationships are not automatically portable weights.	Declare criteria and integrated, equal-weight, and scale-only rankings.
Influencer identification	Content, network, and engagement indicators are complementary.	Public indicators proxy audience composition and commercial response.	Report evidence counts; avoid audience-membership or ROI claims.
Sentiment analysis	Global methods organise opinion text; AfriSenti and SwahBERT show the value of African and Swahili-specific resources.	Accuracy and calibration require traceable local labels, splits, and probabilities, and baselines.	Treat supplied labels descriptively; reserve quality claims for reproducible evaluation.
Airline and tourism evidence	Airline studies show platform and sentiment patterns; African and East African tourism studies connect engagement, eWOM, and destination communication.	Results do not identify causal campaign effects or transfer automatically across brands and countries.	Use a bounded baseline; state that it is not wholly Uganda Airlines-specific.
Decision support	TOPSIS makes benefit/cost preferences explicit.	Results depend on measures, weights, and eligible alternatives.	Add evidence thresholds, baselines, and sensitivity tests.
Regional application	African language resources and East African tourism evidence support contextualisation; Ugandan policy values tourism, air transport, research, and ICT.	Regional relevance is not Uganda Airlines-specific model or campaign validation.	Use a management scenario, expose provenance, and bound all impact claims.

2.8 Research Gap and Conceptual Positioning

Four gaps emerge from the synthesis. First, global influencer studies establish that credibility, congruence, and message value matter, but their measures and estimated relationships do not supply a universally valid weighting scheme for an East African aviation setting. Second, African and Swahili NLP research has expanded language resources, yet no source reviewed here validates the supplied labels or an airline-specific sentiment model for Uganda Airlines. Third, airline and East African tourism studies analyse engagement, electronic word of mouth, destination image, or service perceptions,

while the strategic selection of creators remains a distinct decision problem. Fourth, no reviewed study combines observed-evidence thresholds, sentiment context, six disclosed ranking criteria, scale-only baselines, criterion ablations, source-mixture sensitivity, and provenance-aware dashboard reporting for a Ugandan airline management scenario.

The study addresses an *integration and evidence-governance gap*. It combines supplied sentiment context, brand-value alignment, audience-fit proxies, engagement quality, reach, cost, and TOPSIS within a pipeline that keeps provenance, evidence thresholds, and decision preferences visible. Its originality is not a new sentiment algorithm and its results do not prove marketing improvement. The intended contribution is a reproducible decision-support framework demonstrated in a Uganda Airlines management scenario, accompanied by explicit baseline comparisons and sensitivity limits.

2.9 Implications for the Current Study

The synthesis produces four design commitments. Objective 1 reports only what the observed discourse labels support. Objective 2 treats alignment and audience fields as profile-level proxies with visible evidence counts. Objective 3 uses TOPSIS but compares the integrated specification with equal-weight and scale-only alternatives. Objective 4 separates observed summaries from simulated stress tests and presents sensitivity rather than predictive or causal validation. Chapter 3 translates these commitments into the operational pipeline, and Chapter 4 reports results under the same evidence boundary.

Chapter 3

Research Design and Methodology

3.1 Research Design and Evidence Boundary

This study uses a quantitative, provenance-aware research design to demonstrate a values-aware framework for sentiment summarisation, influencer profiling, and multi-criteria selection. The empirical baseline is an observed legacy airline-discourse dataset; the extended scenario layer is explicitly simulated. The case is therefore a demonstration of a reusable analytical framework for an aviation-service setting, not a report of a live Uganda Airlines social-media collection. The framework combines descriptive text analytics, profile-level summary measures, and Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) ranking (Pang and Lee 2008; Liu 2012; Hwang and Yoon 1981).

Three terms are used consistently throughout the study. *Observed* denotes rows declared `source=real` after validation and tweet-ID deduplication. *Simulated* denotes rows declared `source=synthetic` or `source=synthetic_cross_platform`. *Simulation-augmented* denotes their combination for scenario analysis. Observed evidence is the basis for empirical descriptive claims; any result containing simulated rows is a scenario result and is not independent empirical validation.

3.2 Case Setting, Units of Analysis, and Variables

The observed discourse source is a legacy airline-Twitter dataset, identified in the source file by `source_reference=legacy_dataset_airline_tweets`. Its observed window is 17–24 February 2015. It supplies an aviation-discourse baseline rather than contempo-

aneous brand-specific monitoring. The campaign panel supplies post-level influencer records for the profile and ranking exercises; its observed records span 1 January 2023–14 May 2024. Simulation extends both datasets for controlled scenario analysis, but the combined date range (2015–2026) must not be interpreted as a fully observed longitudinal period.

The units of analysis are (i) individual discourse posts, (ii) influencer-platform profiles and their posts, and (iii) aggregate summaries by provenance mode. Key variables are sentiment class, topic, platform and location fields for discourse; and value alignment, audience fit, sentiment context, engagement quality, reach, and cost where available for profiles. The latter six criteria form the TOPSIS input vector. Date, platform, topic, location, and sentiment fields remain available in the simulation layer to support labelled scenario distributions; their presence does not turn simulated records into observed platform activity.

3.3 Objective-to-Method Alignment

Table 3.1 writes out each approved objective and linked research question in full, then identifies the method and the corresponding results section. The wording is identical to Chapter 1 so that the reader does not have to infer what O1–O4 represent. The table distinguishes observed descriptive evidence from simulation-augmented scenario work and does not imply causal campaign effects.

Table 3.1: Full objective–research-question–method–result alignment

Specific objective	Linked research question	Method, output, and results location
Objective 1: To describe the sentiment distribution and supplied topic structure in the observed aviation-discourse baseline.	RQ1: What sentiment distribution and supplied topic structure characterise the observed aviation-discourse baseline?	Tweet-ID deduplication and provenance-separated descriptive sentiment and supplied-topic summaries. Output: observed distribution and topic-field scope. Results: Section 4.3.
Objective 2: To examine profile-level value-alignment, audience-fit, sentiment-context, engagement-quality, reach, and cost indicators in the observed campaign panel.	RQ2: What do the available observed campaign records show about profile-level value alignment, audience fit, sentiment context, engagement quality, reach, and cost indicators?	Profile aggregation, evidence tiers, platform summaries, and descriptive Pearson correlations. Output: observed proxy profiles with evidence counts. Results: Section 4.4.
Objective 3: To develop and demonstrate an evidence-aware TOPSIS shortlist that incorporates sentiment context, brand-value compatibility, audience fit, engagement quality, reach, and cost efficiency, and to compare it with transparent ranking baselines.	RQ3: How can the six criteria be integrated in an evidence-aware TOPSIS shortlist, and how does that shortlist differ from equal-weight and scale-only baselines?	TOPSIS, equal-weight, reach-only, follower-only, and criterion-ablation comparisons (Hwang and Yoon 1981). Output: shortlist, top-ten overlap, and Spearman agreement. Results: Section 4.5.
Objective 4: To examine source-mixture and criterion sensitivity and to present observed evidence and explicitly labelled scenario results separately in a management-oriented dashboard.	RQ4: How sensitive is the ranking to source-mixture and criterion assumptions, and how can observed evidence and scenario results be distinguished in a management-oriented dashboard?	Source-mixture and criterion sensitivity, provenance-separated scenario summaries, and dashboard presentation. Output: robustness and dashboard signals with explicit boundaries. Results: Section 4.6.

3.4 Data Sources, Sampling, and Immutability

3.4.1 Immutable source files and provenance

The analysis reads, but does not alter, two immutable raw files: `chapters/chapter04/data/tweets_expanded.csv` and `chapters/chapter04/data/campaign_expanded.csv`. Derived facts and Chapter 4 tables are regenerated from these sources; no cleaning operation overwrites a source CSV. The discourse file contains 139,840 raw rows: 14,640 labelled `real`, 120,000 labelled `synthetic`, and 5,200 labelled `synthetic_cross_platform`. The

campaign file contains 60,683 rows: 683 **real** and 60,000 **synthetic**. These source values are provenance labels, not inferred labels.

The source-specific sample sizes reported after processing are given in the generated provenance summary. The observed campaign panel has 683 rows across 95 name-platform profiles representing 37 distinct influencer names. The simulation-augmented campaign panel has 60,683 rows. It is used for scenario exploration only; observed campaign rows remain the basis for default evidence-aware ranking.

3.4.2 Canonical discourse cleaning

The canonical discourse dataset retains the first complete occurrence of each **tweet_id**, preserving input order. Completeness is determined from the required **tweet_id**, **text**, and **source** fields: a later duplicate replaces an earlier occurrence only when it is more complete. Rows without a tweet ID are retained because no ID-based duplicate decision can be made. This rule removes 155 duplicate tweet-ID occurrences and produces 139,685 canonical rows: 14,485 observed and 125,200 simulated. The observed rows are Twitter records from the February 2015 window; the simulated rows may include Twitter and Facebook scenario records.

No undocumented near-duplicate, bot, language-confidence, geolocation, or spam filter is claimed in this thesis. Text, topic, location, and engagement fields are analysed as supplied in the immutable sources. This makes the transformation boundary auditable and prevents unsupported assertions about live collection or additional cleaning.

3.5 Methods by Specific Objective

3.5.1 Objective 1: To describe the sentiment distribution and supplied topic structure in the observed aviation-discourse baseline

Objective 1 uses the canonical observed discourse records produced after tweet-ID deduplication. Sentiment class and the supplied topic field are counted and converted to within-mode percentages. No new sentiment labels or latent topics are inferred. The method therefore answers RQ1 descriptively and produces the inputs reported in Section 4.3.

3.5.2 Objective 2: To examine profile-level value-alignment, audience-fit, sentiment-context, engagement-quality, reach, and cost indicators in the observed campaign panel

Objective 2 groups observed campaign rows by influencer name and platform, retains the number of observed posts for each profile, and summarises the six supplied decision indicators. Evidence tiers are assigned from observed counts, and Pearson correlations are used only to describe pairwise associations among supplied proxy measures. The method answers RQ2 without treating the proxies as verified audience membership, persuasion, conversion, or return on investment; results appear in Section 4.4.

3.5.3 Objective 3: To develop and demonstrate an evidence-aware TOPSIS shortlist that incorporates sentiment context, brand-value compatibility, audience fit, engagement quality, reach, and cost efficiency, and to compare it with transparent ranking baselines

Objective 3 applies TOPSIS to the six criteria for the profiles meeting the observed-evidence threshold. The resulting ordering is compared with equal-weight TOPSIS, reach-only ranking, follower-count-only ranking, and one-criterion-at-a-time ablations. Top-ten overlap, leader identity, and Spearman rank correlation describe how the decision rule differs from its comparators. These methods answer RQ3 as a ranking-behaviour question, not as a campaign-effectiveness test; results appear in Section 4.5.

3.5.4 Objective 4: To examine source-mixture and criterion sensitivity and to present observed evidence and explicitly labelled scenario results separately in a management-oriented dashboard

Objective 4 recomputes the ranking under 60/40, 70/30, and 80/20 observed/simulated source mixtures and presents observed, simulated, and combined outputs as distinct analytical modes. Dashboard signals include provenance counts, evidence-tier exclusions, shortlist sensitivity, and the publication status of unreproduced holdout metrics. The

method answers RQ4 by testing declared assumptions and preserving provenance, not by validating causal or predictive performance; results appear in Section 4.6.

3.6 Operational Analytical Pipeline

The analytical pipeline is implemented as a sequence of bounded transformations rather than a single predictive model. Table [3.2](#) records the input, operation, output, and principal validation check at each stage.

Table 3.2: Operational pipeline from source records to dashboard outputs

Step	Input	Operation and output	Validation or boundary
1	Immutable discourse and campaign CSVs	Parse required fields and normalise declared source values to observed or simulated provenance.	Unknown provenance is rejected rather than silently treated as observed.
2	Discourse rows	Retain the first complete occurrence of each tweet ID; produce the canonical discourse table.	Report raw rows, duplicates removed, canonical rows, and source counts.
3	Canonical discourse rows	Summarise supplied sentiment and topic fields separately by analytical mode.	Descriptive summary only; no new label or classifier-quality claim.
4	Campaign rows	Group by influencer name and platform; calculate source-specific profile averages.	Preserve observed and simulated post counts and date windows.
5	Observed profile aggregates	Assign evidence tiers and restrict default observed/combined ranking to profiles with at least five observed posts.	Simulated volume cannot increase eligibility or confidence.
6	Six profile criteria	Vector-normalise, apply declared weights, and calculate TOPSIS closeness coefficients.	Treat the coefficient as a relative ranking value, not success probability.
7	Eligible profiles	Recompute rankings under equal weights, scale-only baselines, criterion omissions, and 60/40–80/20 source mixtures.	Report rank agreement and overlap as deterministic sensitivity, not causal validation.
8	Generated facts and ranking outputs	Present observed, simulated, and combined views in the dashboard and Chapter 4 tables.	Observed mode is the default; scenario provenance remains visible.

The implementation reads the two source files without overwriting them. The evidence builder regenerates source counts, descriptive summaries, and evidence tiers, while the dashboard analytics modules regenerate profile aggregates and TOPSIS rankings. The separate ranking-baseline script produces the comparator and ablation tables used

in Chapter 4. This division allows each reported result to be traced to an explicit transformation rather than to an undocumented manual calculation.

3.7 Simulation-Augmented Scenario Construction

Simulation was necessary because the observed discourse baseline covers only eight days in 2015 and the observed campaign panel is limited to 683 rows. It enables stress testing of the dashboard and ranking workflow over a broader, explicitly labelled scenario layer without presenting that layer as newly observed social-media behaviour. Every generated-style row carries its original `source` value, which is normalised to the *simulated* provenance field for analysis; `real` is normalised to *observed*. Unknown provenance values are not silently treated as observed.

The scenario layer retains the supplied distributions represented by its sentiment, temporal, platform, topic, and location fields, while keeping each row’s provenance visible. In particular, its sentiment shares are reported separately from the observed shares in the generated sentiment-by-mode table; similarity of those summaries is a calibration property of the scenario data, not a validation result. The simulation-augmented union is therefore suitable for controlled comparisons and stress tests, but it is not an independently observed 2015–2026 dataset.

3.8 Influencer Profiles, Eligibility, and TOPSIS

3.8.1 Evidence tiers and mode-specific profiles

Profiles are defined by influencer name and platform. Observed, simulated, and combined modes calculate source-specific profile aggregates separately. A profile with fewer than five observed posts is *insufficient* (1–4) and is excluded from default observed and combined rankings. Eligible profiles are *exploratory* (5–9 observed posts), *moderate* (10–19), or *strong* (20 or more). The generated evidence-tier table reports the resulting tier counts. Simulated post volume does not increase a profile’s evidence tier or confidence.

For a combined scenario, each source-specific profile metric is combined before TOPSIS vector normalisation, using the declared source-mixture assumption:

$$m_i^{(0.70)} = 0.70m_{i,\text{observed}} + 0.30m_{i,\text{simulated}}.$$

TOPSIS then vector-normalises the combined metrics across eligible profiles before applying criterion weights. The source mixture is not an estimated population proportion. Sensitivity analysis recomputes combined rankings at 60/40, 70/30, and 80/20 observed/simulated weights and records top-ten overlap and rank movement. This assesses a ranking’s dependence on the declared mixture; it does not validate the ranking against observed campaign outcomes.

3.8.2 TOPSIS decision rule

For profile i , the TOPSIS criteria vector is

$$\mathbf{x}_i = (\text{ValueAlign}_i, \text{AudienceFit}_i, \text{SentimentContext}_i, \text{EQ}_i, \text{Reach}_i, \text{Cost}_i),$$

where higher values are preferred for the first five criteria and lower cost is preferred for the final criterion. Each criterion is vector-normalised, $r_{ij} = x_{ij} / \sqrt{\sum_i x_{ij}^2}$, and weighted with non-negative decision weights that are normalised to sum to one. The weighted matrix is $v_{ij} = w_j r_{ij}$. TOPSIS then defines ideal-best and ideal-worst criterion vectors according to their benefit or cost direction, calculates distances d_i^+ and d_i^- , and ranks by the closeness coefficient

$$\text{CC}_i = \frac{d_i^-}{d_i^+ + d_i^-} \in [0, 1].$$

The score is a transparent multi-criteria decision-support quantity. It is not a measured probability of campaign success, and any missing cost is visibly handled as an imputed ranking input rather than evidence of a real price.

3.8.3 Ranking baselines and criterion ablations

The integrated default ranking is compared with three non-causal reference specifications. Equal-weight TOPSIS assigns the same preference to all six criteria. The reach-only and follower-count-only baselines sort the same 64 eligible observed profiles using the named scale indicator. These baselines answer whether the integrated ordering is materially different from a simple scale rule; they do not estimate what would have happened in a real campaign.

Criterion sensitivity is assessed by setting one default TOPSIS weight to zero at a time and renormalising the remaining weights. Each comparator is evaluated against the integrated default using top-ten overlap, leader identity, and Spearman rank correlation across all eligible profiles. This is an ablation of the decision specification, not an ablation of a trained predictive model.

3.9 Evaluation Protocols and Validation Boundaries

The study separates five kinds of evaluation rather than treating them as interchangeable. First, sentiment percentages, topic counts, platform counts, and profile summaries are *descriptive* results. They characterise the selected provenance mode and do not by themselves establish sentiment-classification accuracy. Second, a labelled sentiment-validation claim may be made only for a traceable labelled set with a documented labelling procedure, split, and source composition. No untraceable gold-standard annotation, inter-annotator agreement, cross-validation result, calibration result, or ablation result is claimed here.

Third, the existing resonance figures in the evidence register are candidate *simulation holdout performance*: accuracy 0.8882, precision 0.8819, recall 0.8658, F1 0.8737, and ROC-AUC 0.9277. The evidence register marks them pending reproduction because the generating script, labels, split, and source composition have not yet been reproduced. Accordingly, these metrics cannot establish external real-world accuracy or observed campaign effectiveness. Fourth, TOPSIS sensitivity is a deterministic robustness check under the stated 60/40, 70/30, and 80/20 source-mixture assumptions; it is not predictive validation. Fifth, real-world effectiveness—such as causal uplift, conversion, reach, or reputation change caused by a selected influencer—is unmeasured in the present data and requires a separately designed, observed campaign evaluation.

3.9.1 Brier-score justification and non-reporting decision

If record-level predicted probabilities and observed binary labels are available, calibration should be evaluated separately from discrimination. The Brier score is

$$\text{BS} = \frac{1}{N} \sum_{i=1}^N (p_i - y_i)^2,$$

where p_i is the predicted probability and y_i is the observed outcome (Brier 1950). It is appropriate when a probability will inform a decision because it penalises confident incorrect predictions as well as less accurate probability assignments. Lower is better, zero is perfect, and interpretation should include a prevalence-based reference forecast and, where feasible, a calibration plot (Guo et al. 2017).

No Brier value is reported for the present holdout metrics. The available evidence register contains only aggregate candidate figures and does not preserve the record-level

probabilities, labels, split, generating code, or source composition required to reproduce a Brier score. Reporting one would therefore create an appearance of calibration evidence that the study cannot verify. Brier evaluation remains part of the production-validation protocol once those artefacts are available.

3.10 Dashboard and Reproducibility

The dashboard prototype reads the same source files and exposes observed, simulated, and combined modes with provenance counts and methodology assumptions. Observed is the default analytical mode. A normal collection failure is returned as unavailable or blocked; it does not trigger an unlabelled simulation. If a demonstration simulation is explicitly enabled, generated comments and their aggregates are labelled simulated.

The evidence builder records input paths, source-value counts, canonical counts, date windows, sentiment summaries, evidence tiers, and validation labels in `defence/evidence/defence-facts.json`. It writes the Chapter 4 CSV tables from those facts. This provenance trail enables a reader to reproduce the reported descriptive and scenario summaries without changing the raw files.

3.11 Validity, Reliability, Ethics, and Limitations

Construct validity is supported by transparent definitions of the analytical modes, the stated TOPSIS criteria, and the documented eligibility threshold. Reliability is limited to reproducibility of the declared transformations and deterministic ranking inputs; it is not evidence that simulated records reproduce real social-media behaviour. External validity is constrained by the legacy airline baseline, the short observed discourse window, public-data proxies, and the simulation design. The framework can be adapted to another brand by supplying documented values and taxonomy, but that adaptation requires fresh observed data and validation.

The analysis uses the supplied public-style records in aggregate form and reports provenance rather than treating simulation as observation. It avoids claims about private audience attributes, causal influence, or live platform access. The principal limitations are source specificity, residual sentiment ambiguity, limited observed campaign evidence per profile, proxy-based audience and alignment measures, and the absence of traceable observed outcome labels for real-world effectiveness.

Chapter 4

Results and Findings

4.1 Overview

This chapter reports the thesis results using the evidence boundary defined in Chapter 3. Observed evidence is reported first. Simulated and simulation-augmented outputs are presented only as scenario results and are not treated as independently observed social-media behaviour or real-world campaign-effect evidence.

The chapter follows the four research objectives using the wording established in Chapter 1. Section 4.2 first establishes the data provenance and source counts. Section 4.3 answers Objective 1 through observed sentiment and supplied-topic summaries. Section 4.4 answers Objective 2 through profile evidence, decision indicators, and descriptive associations. Section 4.5 answers Objective 3 through the TOPSIS shortlist and transparent baseline comparisons. Section 4.6 answers Objective 4 through criterion and source-mixture sensitivity, dashboard interpretation, and validation boundaries.

4.2 Data Provenance and Analysis Inputs

Table 4.1 lists the artefacts used to reproduce the chapter. The immutable source CSVs are read-only inputs; the generated JSON and CSV tables are derived outputs used for the counts, percentages, rankings, and figures in this chapter.

Table 4.1: Chapter 4 data and analysis artefacts

Artefact	Local reference	Used for
Discourse source	chapters/chapter04/data/ tweets_expanded.csv	Objective 1 sentiment summaries and Objective 4 monitoring inputs
Campaign source	chapters/chapter04/data/ campaign_expanded.csv	Objective 2 profile evidence and Objective 3 ranking inputs
Evidence register	defence/evidence/defence- facts.json	Canonical counts, provenance windows, evidence tiers, and validation labels
Generated Chapter 4 tables	chapters/chapter04/tables/ table4-1...table4-10	Reproducible descriptive, ranking, baseline, and ablation outputs
Dashboard code and tests	company_interactive_ dashboard/	Mode-aware analytics, TOPSIS, recommender, and dashboard output checks

The artefact structure separates source evidence from presentation. The two CSV files are the immutable inputs; the evidence register and generated tables record the transformations used for reporting; and the dashboard consumes the same analytical outputs. This separation matters because a value displayed in the interface can be traced to a generated artefact instead of being treated as an undocumented dashboard calculation.

4.2.1 Canonical source counts

The discourse source contains 139 840 raw rows. Deduplication removes 155 tweet-ID duplicates and leaves 139 685 canonical rows. Of these canonical rows, 14 485 are observed and 125 200 are simulated. The observed discourse window is 17–24 February 2015. The 2015–2026 window belongs to the simulated and simulation-augmented scenario layer, not to observed data.

The campaign source contains 60 683 rows: 683 observed and 60 000 simulated. The

observed campaign records span 1 January 2023 to 14 May 2024. Across the campaign source there are 95 influencer-platform profiles and 37 distinct influencer names.

Table 4.2: Generated provenance summary

Dataset	Mode	Records	Interpretation
Campaign	Combined	60 683	Simulation-augmented scenario
Campaign	Observed	683	Observed campaign rows
Campaign	Simulated	60 000	Simulated campaign rows
Discourse	Combined	139 685	Canonical simulation-augmented records
Discourse	Observed	14 485	Observed records after tweet-ID deduplication
Discourse	Simulated	125 200	Simulated records after tweet-ID deduplication

The provenance summary shows that the combined datasets are numerically dominated by simulation: observed records account for approximately 10.37% of the canonical discourse dataset and 1.13% of the campaign dataset. A combined count is therefore not stronger observed evidence. For descriptive claims, the observed rows remain the relevant denominator; the larger combined totals are used only to exercise the scenario and dashboard pathways.

4.3 Results for Objective 1: To describe the sentiment distribution and supplied topic structure in the observed aviation-discourse baseline

4.3.1 Observed sentiment baseline

The observed discourse baseline contains 14 485 canonical posts, all from the February 2015 legacy airline-Twitter source. Its sentiment distribution is negative-heavy: 9082 negative posts (62.6993%), 3069 neutral posts (21.1874%), and 2334 positive posts (16.1132%). This is a descriptive result for the observed source, not an accuracy result for a sentiment classifier.

Table 4.3: Sentiment by provenance mode

Mode	Negative		Neutral		Positive	
	Count	%	Count	%	Count	%
Observed	9082	62.6993	3069	21.1874	2334	16.1132
Simulated scenario	78 706	62.8642	26 826	21.4265	19 668	15.7093
Combined scenario	87 788	62.8471	29 895	21.4017	22 002	15.7512

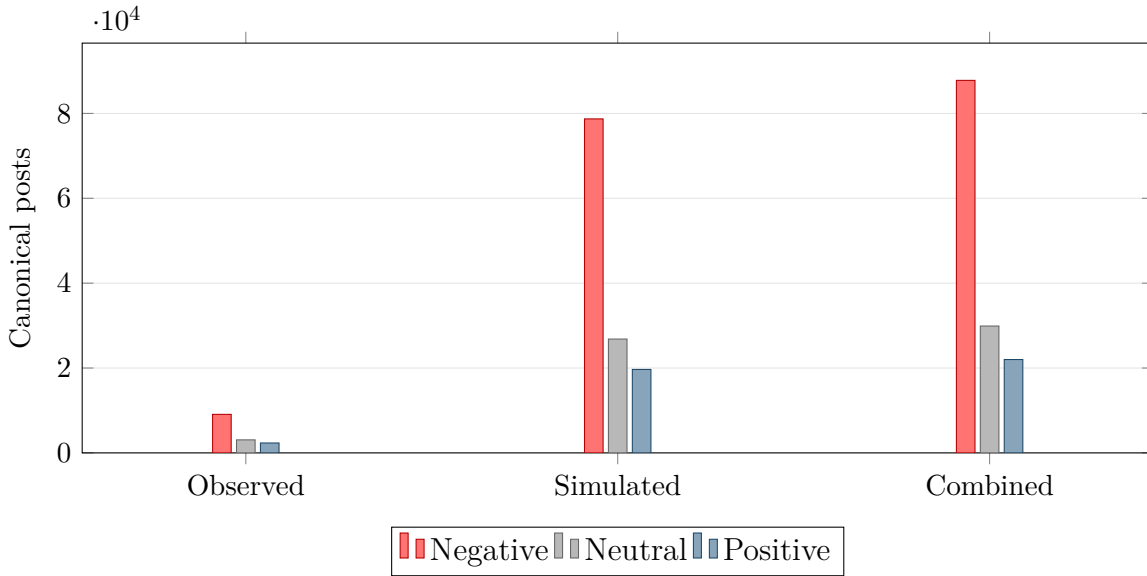


Figure 4.1: Sentiment counts by provenance mode.

Figure 4.1 shows that negative labels are the largest category in every analytical mode. The much larger simulated and combined bars reflect row volume, not stronger evidence. The comparable percentage distributions in Table 4.3 arise from how the scenario layer was calibrated; they do not provide independent confirmation of sentiment after the observed February 2015 window.

The observed topic field is also bounded: all 14 485 observed discourse records carry the supplied `legacy_general` topic label. Chapter 4 therefore does not claim observed aspect discovery, geographic inference, or message-propagation effects from this baseline.

4.3.2 Simulated and combined scenario comparison

The simulated discourse layer contains 125 200 canonical records. It has 78 706 negative, 26 826 neutral, and 19 668 positive records. The simulation-augmented combined scenario contains 139 685 canonical records, with 87 788 negative, 29 895 neutral, and 22 002 positive records.

These scenario summaries are useful for dashboard stress testing because they preserve labelled source separation and expose how outputs look when observed and simulated rows are viewed together. They do not convert the 2015–2026 scenario range into a fully observed longitudinal record, and they do not validate future sentiment behaviour.

4.4 Results for Objective 2: To examine profile-level value-alignment, audience-fit, sentiment-context, engagement-quality, reach, and cost indicators in the observed campaign panel

Objective 2 requires profile-level evidence before any ranking claim is made. The observed campaign panel supplies 683 observed rows. All 95 influencer-platform profiles have at least one observed post, but only 64 profiles meet the default TOPSIS eligibility rule of at least five observed posts. Those 64 eligible profiles contain 590 observed posts. The remaining 31 profiles have 93 observed posts in total and are excluded from the default observed and combined rankings.

Table 4.4: Observed influencer evidence tiers

Tier	Observed posts/profile	Profiles	Observed posts	Default status
Insufficient	1–4	31	93	Excluded
Exploratory	5–9	48	345	Eligible
Moderate	10–19	13	165	Eligible
Strong	20+	3	80	Eligible

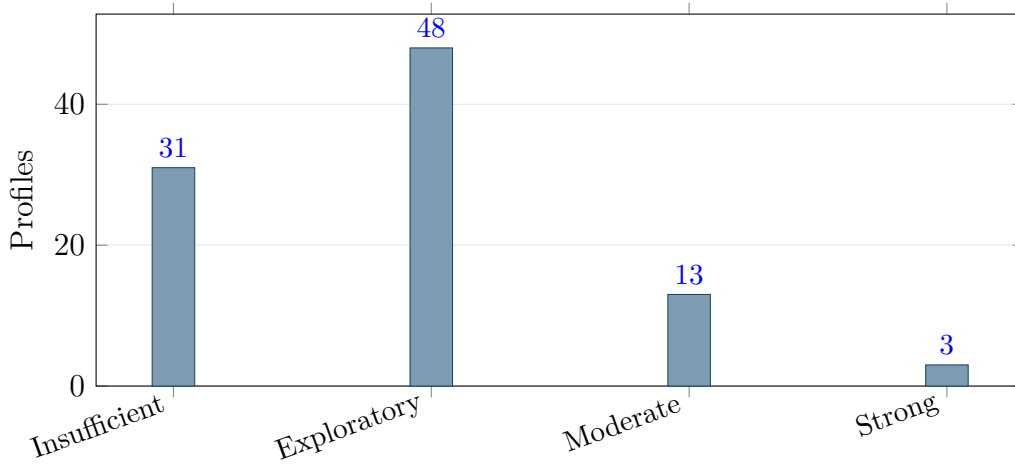


Figure 4.2: Observed profile counts by evidence tier.

The evidence filter retains 64 of the 95 profiles (67.37%) and 590 of the 683 observed campaign rows (86.38%) for the default ranking. Most eligible profiles are still in the exploratory tier: 48 profiles have only five to nine observed posts. The filter therefore prevents the sparsest profiles from being ranked by default, but it does not make the retained profiles representative or prove their campaign effectiveness.

The tier names describe observed sample size only. In particular, the *strong* tier means 20 or more observed posts for a profile; it is not a claim that the profile has proven campaign impact. Audience fit, value alignment, sentiment context, engagement quality, reach, and cost are therefore treated as decision-support inputs rather than causal measures.

Across the 683 observed campaign rows, the strongest association with the supplied high-resonance proxy is value alignment ($r = 0.7802$), followed by sentiment context ($r = 0.5592$); audience fit has a weaker association ($r = 0.1300$). Reach and engagement quality are negatively associated ($r = -0.6433$). For Objective 2, the practical implication is that the supplied indicators do not move together uniformly: scale is not a substitute for interaction quality, and the available audience-fit proxy contributes little linear association with the supplied resonance proxy. These are descriptive Pearson correlations within the observed panel. They do not establish causal influence, and the high-resonance field remains a supplied proxy rather than an externally validated campaign outcome.

4.5 Results for Objective 3: To develop and demonstrate an evidence-aware TOPSIS shortlist that incorporates sentiment context, brand-value compatibility, audience fit, engagement quality, reach, and cost efficiency, and to compare it with transparent ranking baselines

The observed default TOPSIS ranking was regenerated through the dashboard analytics and TOPSIS modules. The export used only observed profile metrics from profiles with at least five observed posts. It did not use simulated post volume to make a profile eligible.

Table 4.5: Observed default TOPSIS top ten

Rank	Influencer	Platform	Category	Observed posts	TOPSIS score
1	Bryan Morel Publications	Twitter	Lifestyle	6	0.6967
2	Fat Cat Podcast UG	Facebook	Travel	8	0.6685
3	Matifa Kajoke	Instagram	Lifestyle	10	0.6564
4	Harland MaYanja	Twitter	Lifestyle	7	0.6260
5	Matifa Kajoke	Facebook	Food	9	0.5809
6	Allan Toniks	Facebook	Food	6	0.5771
7	Shane Singer	Twitter	Travel	5	0.5749
8	Fatboy Gagamel	Twitter	Lifestyle	6	0.5748
9	Mary Luswata	Twitter	Lifestyle	8	0.5632
10	Spice Diana	Instagram	Entertainment	21	0.5490

Table 4.5 answers Objective 3 as an observed decision-support ranking. The top score is 0.6967 for Bryan Morel Publications on Twitter. Because the score is a TOPSIS closeness coefficient, it should be read as a relative multi-criteria ranking value, not as a probability of campaign success.

4.5.1 Comparison with transparent ranking baselines

Table 4.6 compares the integrated default with equal-weight TOPSIS and two scale-only rules. All specifications use the same 64 observed-eligible profiles. Spearman’s ρ measures agreement across the complete ordering, while top-ten overlap gives a directly interpretable shortlist comparison.

Table 4.6: Integrated ranking compared with equal-weight and scale-only baselines

Specification	Top-ten overlap	Spearman ρ	Leader
Integrated default TOPSIS	10	1.0000	Bryan Morel Publications (Twitter)
Equal-weight TOPSIS	7	0.8402	Fat Cat Podcast UG (Facebook)
Reach-only	0	−0.4058	Natukunda Jane (Instagram)
Follower-count-only	2	−0.0109	Solome Katende (Facebook)

The integrated ordering is similar, but not identical, to equal-weight TOPSIS: seven profiles are shared in the top ten and the overall rank correlation is 0.8402. In contrast, the reach-only top ten has no profile in common with the integrated top ten and its rank correlation is negative. The follower-only baseline shares two top-ten profiles and has near-zero overall rank correlation. These results support H1 as a statement about ranking behaviour: the integrated decision rule is materially different from selecting on scale alone. They do not show that the integrated shortlist produces better campaign outcomes.

4.6 Results for Objective 4: To examine source-mixture and criterion sensitivity and to present observed evidence and explicitly labelled scenario results separately in a management-oriented dashboard

Objective 4 examines how the ranking changes when declared assumptions are altered and how those changes are communicated without mixing observed evidence with simulation. Criterion-ablation and source-mixture results are reported first, followed by the dashboard signals and their validation boundaries.

4.6.1 Criterion-ablation sensitivity

Table 4.7 removes one criterion at a time from the integrated TOPSIS weights. This shows which declared preferences most affect the current ordering; it does not estimate a criterion’s causal contribution to engagement or resonance.

Table 4.7: One-criterion-at-a-time TOPSIS sensitivity

Omitted criterion	Top-ten overlap	Spearman ρ	Leader
Value alignment	8	0.7734	Bryan Morel Publications (Twitter)
Audience fit	10	0.9957	Bryan Morel Publications (Twitter)
Sentiment context	10	0.9863	Bryan Morel Publications (Twitter)
Engagement quality	1	0.4632	Natukunda Jane (Instagram)
Reach	10	0.9959	Bryan Morel Publications (Twitter)
Cost	10	0.9809	Bryan Morel Publications (Twitter)

The largest change occurs when engagement quality is omitted: only one default top-ten profile remains and the leader changes. Removing value alignment also produces a noticeable, though smaller, change. Audience fit and sentiment context are implemented

in the ranking, but their individual removal leaves the current top ten unchanged and produces rank correlations above 0.98. H2 is therefore supported only in a qualified form: criterion effects are heterogeneous, and the available profiles do not show a large standalone ranking effect for audience fit or sentiment context under the default specification.

4.6.2 Source-mixture sensitivity

Table 4.8 reports the 60/40, 70/30, and 80/20 observed/simulated source-mixture sensitivity scenarios generated by the dashboard TOPSIS code. These rows are scenario-labelled because they combine observed and simulated profile metrics before ranking.

Table 4.8: TOPSIS sensitivity under observed/simulated source-mixture scenarios

Scenario	Top-10 overlap vs. observed	Top-10 overlap vs. 70/30	Shared-profile rank changes	Leader score
60/40	0.90	1.00	2	0.7730
70/30	0.90	1.00	0	0.7574
80/20	1.00	0.90	2	0.7410

Across the three combined scenarios, 9 profiles remain common to all top-ten lists and the observed default leader, Bryan Morel Publications on Twitter, remains ranked first. The leader’s score falls from 0.7730 under the 60/40 mixture to 0.7410 under the 80/20 mixture, showing that numerical scores change even when the first-ranked profile does not. The sensitivity result is therefore a deterministic check on the declared source-mixture assumption. It is not predictive validation and does not show observed campaign effectiveness.

4.6.3 Monitoring and validation boundaries

Objective 4 concerns dashboard-ready monitoring outputs and robustness boundaries. The observed baseline supports current descriptive panels for 14 485 discourse posts and 683 campaign rows. The simulation-augmented layer supports labelled stress testing over 139 685 canonical discourse records and 60 683 campaign rows, but it must remain visibly separate from observed evidence.

4.6.4 Pending-reproduction holdout metrics

The resonance metrics in Table 4.9 are retained only as pending-reproduction simulation holdout performance. The evidence register marks them as non-publishable for real-world accuracy because the generating script, labels, split, and source composition have not been reproduced.

Table 4.9: Pending-reproduction simulation holdout metrics

Metric	Value
Accuracy	0.8882
Precision	0.8819
Recall	0.8658
F1 score	0.8737
ROC-AUC	0.9277

These figures cannot be described as predictive readiness, internal-coherence evidence, or observed real-world campaign accuracy. They are retained as a transparent limitation and as a candidate metric set requiring reproduction before publication.

4.6.5 Dashboard-ready outputs

Table 4.10 summarises the dashboard outputs that follow directly from the regenerated evidence and ranking artefacts.

Table 4.10: Objective 4 dashboard signals derived from regenerated evidence

Signal	Reported value	Interpretation boundary
Observed analytics panel	14 485 discourse posts; 683 campaign rows	Descriptive observed baseline only
Evidence filter	64 eligible profiles; 31 insufficient profiles excluded	Simulated volume does not raise evidence tier or confidence
Scenario comparison	125 200 simulated discourse rows; 60 000 simulated campaign rows	Stress-test layer, not independent empirical validation
Ranking sensitivity	60/40, 70/30, and 80/20 source-mixture checks; 9 common top-ten profiles	Deterministic scenario sensitivity, not causal or predictive validation
Holdout metric card	Accuracy 0.8882, F1 0.8737, ROC-AUC 0.9277	Pending reproduction and non-publishable as real-world evidence

The observed analytics panel establishes the empirical base available to the dashboard: 14 485 canonical discourse posts and 683 campaign rows. Managers can use these panels to inspect the supplied sentiment distribution and campaign-profile indicators, but the counts represent the selected source files rather than live Uganda Airlines monitoring or a representative sample of Ugandan travellers.

The evidence filter shows the practical effect of the eligibility rule. Only 64 of 95 profiles enter the default ranking, while 31 profiles are withheld because they have fewer than five observed posts. This makes sparse evidence visible and prevents simulated volume from creating an appearance of observed support. It is a minimum-data safeguard, not proof that every eligible profile is reliable or commercially effective.

The scenario comparison exposes 125 200 simulated discourse rows and 60 000 simulated campaign rows as a separate stress-test layer. Their purpose is to test interface behaviour and source-mixture assumptions over a larger scenario. Because the records are generated rather than independently observed, agreement between scenario and observed summaries cannot be treated as external validation.

The ranking-sensitivity signal summarises the 60/40, 70/30, and 80/20 source-mixture

tests. Nine profiles remain in every top ten and the leader remains stable, indicating limited shortlist movement within the tested range. The changed scores and two shared-profile rank movements show that the result is not invariant. A manager should therefore review both the shortlist and its assumptions rather than treating rank one as an unconditional recommendation.

Finally, the holdout metric card preserves accuracy, F1, and ROC-AUC values for audit purposes but marks them pending reproduction. Without the generating code, record-level predictions, labels, split, and source composition, the dashboard cannot use them as evidence of calibrated or real-world performance. Taken together, the five signals answer Objective 4 by showing what is observed, what is excluded, what is simulated, how the ranking responds to assumptions, and which validation claims remain unavailable.

4.6.6 Direct answers to the research questions

Research Question 1: What sentiment distribution and supplied topic structure characterise the observed aviation-discourse baseline?

Answer. The observed baseline contains 62.6993% negative, 21.1874% neutral, and 16.1132% positive posts. All observed records carry the supplied `legacy_general` topic label. The study therefore establishes a negative-heavy distribution for the selected February 2015 aviation source, but it does not establish observed aspect-level topic discovery, live sentiment, or later longitudinal change.

Research Question 2: What do the available observed campaign records show about profile-level value alignment, audience fit, sentiment context, engagement quality, reach, and cost indicators?

Answer. The 683 observed campaign rows provide supplied proxy indicators for 95 influencer-platform profiles. Value alignment has the strongest association with the supplied high-resonance proxy ($r = 0.7802$), sentiment context has a moderate association ($r = 0.5592$), audience fit has a weak association ($r = 0.1300$), and reach is negatively associated with engagement quality ($r = -0.6433$). Only 64 profiles meet the default five-post threshold; 31 are excluded. These associations describe the panel and do not verify audience membership or commercial effect.

Research Question 3: How can the six criteria be integrated in an evidence-

aware TOPSIS shortlist, and how does that shortlist differ from equal-weight and scale-only baselines?

Answer. After vector normalisation and application of the declared weights and criterion directions, TOPSIS produces an observed shortlist headed by Bryan Morel Publications, Fat Cat Podcast UG, and Matifa Kajoke. The integrated top ten shares seven profiles with equal-weight TOPSIS, none with reach-only ranking, and two with follower-count-only ranking. The result demonstrates that the implemented multi-criteria rule differs materially from scale-only sorting; it does not demonstrate superior campaign outcomes.

Research Question 4: How sensitive is the ranking to source-mixture and criterion assumptions, and how can observed evidence and scenario results be distinguished in a management-oriented dashboard?

Answer. Criterion ablation produces heterogeneous changes: omitting engagement quality changes the leader and leaves only one default top-ten profile, while omitting value alignment produces a smaller but noticeable change. Across the 60/40, 70/30, and 80/20 source mixtures, nine profiles remain common to all top tens and the leader remains stable, although scores and some positions move. The dashboard distinguishes these scenario results through mode labels, provenance counts, evidence filters, and explicit publication-status warnings for unreproduced metrics. The sensitivity evidence is deterministic and does not validate causal or predictive performance.

Chapter 5

Discussion

5.1 Purpose and Evidence Boundary

This chapter interprets the results in Chapter 4 against the objectives and research questions stated in Chapter 1. The central distinction is between the observed baseline and the simulated or simulation-augmented scenario layer. Empirical interpretation begins with the observed records: 14 485 canonical discourse posts from the February 2015 legacy airline-Twitter source and 683 observed campaign records from January 2023 to May 2024. The simulated discourse and campaign records are retained for labelled stress tests and source-mixture scenarios; they are not additional observed social-media activity or independent empirical validation.

This distinction also governs the claims made below. Similar observed and simulated sentiment shares indicate that the simulation was calibrated to resemble the supplied sentiment distribution. They do not validate the simulation, demonstrate future sentiment behaviour, or establish a longitudinal trend. Likewise, the TOPSIS results are decision-support rankings under documented criteria and assumptions, rather than estimates of causal campaign uplift.

5.2 Interpretation by Objective and Research Question

Objective 1: To describe the sentiment distribution and supplied topic structure in the observed aviation-discourse baseline

Research Question 1: What sentiment distribution and supplied topic structure characterise the observed aviation-discourse baseline?

For Objective 1 and RQ1, the observed baseline is negative-heavy: 9082 of 14 485 canonical posts (62.6993%) are negative, 3069 (21.1874%) are neutral, and 2334 (16.1132%) are positive. This is a descriptive account of the observed source, not a classifier-accuracy result. It supports the limited conclusion that the retained 2015 discourse source contains more negative than neutral or positive labels.

The thematic result is deliberately narrow. The observed records carry the supplied `legacy_general` topic label, so the study cannot claim observed aspect discovery, geographic inference, propagation effects, or incident-specific narratives. Aviation terminology and domain adaptation remain useful design concepts for a future labelled analysis, but the available evidence does not establish that a domain-adapted model improved classification over an off-the-shelf alternative.

The simulated layer has sentiment shares close to the observed baseline (62.8642% negative, 21.4265% neutral, and 15.7093% positive). Those similarities are expected under simulation calibration. They make the scenario layer useful for checking provenance-aware dashboard displays and stress-testing source mixtures, but they are not confirmation of the observed distribution beyond the underlying source or of sentiment in later years.

Objective 2: To examine profile-level value-alignment, audience-fit, sentiment-context, engagement-quality, reach, and cost indicators in the observed campaign panel

Research Question 2: What do the available observed campaign records show about profile-level value alignment, audience fit, sentiment context, engagement quality, reach, and cost indicators?

For Objective 2 and RQ2, the campaign panel provides profile-level inputs for alignment, audience fit, sentiment context, engagement quality, reach, and cost. These inputs should be read as supplied public-data proxies rather than direct measures of audience membership, persuasion, conversion, or return on investment. In the observed campaign records, the supplied high-resonance probability proxy is strongly associated with value alignment (Pearson $r = 0.78024905$) and moderately associated with sentiment context ($r = 0.5591605$); its association with audience fit is weak ($r = 0.12995486$). Reach is negatively associated with engagement quality ($r = -0.64331418$). These are descriptive associations within the observed campaign panel. Because the resonance quantity is itself a proxy and the data are not an experimental campaign design, none of the correlations identifies a causal effect or validates a commercial outcome.

The observed evidence is also uneven across profiles. All 95 influencer-platform profiles have at least one observed campaign post, but only 64 meet the default threshold of five observed posts. The evidence tiers therefore communicate observed sample size, not demonstrated effectiveness. They appropriately limit the default ranking to profiles for which at least a small amount of observed evidence exists and keep 31 lower-evidence profiles out of the default shortlist.

Objective 3: To develop and demonstrate an evidence-aware TOPSIS shortlist that incorporates sentiment context, brand-value compatibility, audience fit, engagement quality, reach, and cost efficiency, and to compare it with transparent ranking baselines

Research Question 3: How can the six criteria be integrated in an evidence-aware TOPSIS shortlist, and how does that shortlist differ from equal-weight and scale-only baselines?

For Objective 3 and RQ3, the contribution is a reproducible application of TOPSIS to the observed eligible profiles. The observed top three are Bryan Morel Publications (Twitter), Fat Cat Podcast UG (Facebook), and Matifa Kajoke (Instagram). Their positions arise from the stated multi-criteria inputs and the selected decision rule; a TOPSIS closeness coefficient is a relative ranking value, not a probability of campaign success.

The baseline comparisons establish that the integrated ordering is not a disguised scale ranking. It shares no top-ten profiles with the reach-only list and two with the follower-count-only list; the corresponding complete-order Spearman correlations are -0.4058 and -0.0109 . Equal-weight TOPSIS is closer to the default preference structure, sharing seven top-ten profiles with $\rho = 0.8402$, but it selects a different leader. This answers RQ3 by demonstrating how the six-criterion ordering differs from transparent alternatives. It is evidence about the implemented decision rule, not about downstream campaign effectiveness.

Objective 4: To examine source-mixture and criterion sensitivity and to present observed evidence and explicitly labelled scenario results separately in a management-oriented dashboard

Research Question 4: How sensitive is the ranking to source-mixture and criterion assumptions, and how can observed evidence and scenario results be distinguished in a management-oriented dashboard?

For Objective 4 and RQ4, the current contribution is a provenance-aware dashboard output design. It can present the observed baseline separately from simulated and combined scenarios, expose evidence tiers and source counts, and display TOPSIS rankings with their sensitivity checks. These are useful decision-support and audit features.

The criterion ablations show the largest changes when engagement quality or value alignment is omitted. Audience fit and sentiment context remain conceptually justified inputs, but their individual omission changes the current ordering much less. The combined scenarios test a declared observed/simulated source mixture rather than a model of the population. Across the 60/40, 70/30, and 80/20 checks, nine profiles are common to all top-ten lists and the leader remains Bryan Morel Publications, although the leader score and some shared-profile positions change. These findings show that sensitivity is heterogeneous and assumption-bound; they do not validate the ranking against campaign outcomes or prove that the leader will remain best in deployment.

The observed evidence does not support a claim of observed longitudinal evolution, live operation, near-real-time monitoring, or campaign-performance tracking over time. The 2015–2026 range belongs to the simulated and simulation-augmented scenario layer. It may be used to demonstrate labelled longitudinal stress-test views, but not to infer

observed trends or to validate ongoing monitoring performance.

5.3 Relationship to Prior Literature and Implications

The observed results are consistent with global literature that treats credibility, congruence, audience relevance, and engagement quality as considerations beyond raw scale (Lou and Yuan 2019; Schouten et al. 2020; Belanche et al. 2021; Harrigan et al. 2021). Reviews caution that source, message, audience, platform, and outcome measures vary across studies (Hudders et al. 2021; Joshi et al. 2025). The present evidence is therefore not sufficient to conclude that fit causes superior campaign performance. The association between the supplied resonance proxy and value alignment is informative for candidate screening, while the weak association with the audience-fit proxy cautions against treating a single public indicator as decisive.

The regional literature reinforces that caution. AfriSenti demonstrates both progress and variation in sentiment resources across African languages, while SwahBERT shows the value of language-specific resources for evaluated Swahili tasks (Muhammad et al. 2023; Martin et al. 2022). Tanzania tourism evidence also treats engagement as multidimensional rather than reducible to raw volume (Kanje et al. 2020). Aviation-specific vocabulary and local language resources are therefore plausible future improvements, but this thesis does not provide a traceable labelled observed baseline comparison from which such an improvement claim can be made. The unreproduced resonance holdout figures reported in Chapter 4 are consequently not evidence of real-world predictive accuracy, calibrated thresholds, or operational readiness.

Table 5.1: Bounded interpretation of the main findings

Finding	Interpretation	Boundary for use
Observed sentiment is 62.6993% negative	Describes the retained February 2015 observed discourse baseline.	Not classifier validation, aspect discovery, or evidence of later sentiment patterns.
Value-alignment/proxy association is $r = 0.78024905$	Supports inspecting value alignment as one decision input.	The outcome is a supplied proxy; the association is non-causal.
Integrated and scale-only rankings differ	Demonstrates that the implemented multi-criteria ordering is not equivalent to sorting by reach or followers.	Not proof that the integrated ranking produces better campaign outcomes.
Similar simulated sentiment shares	Reflects simulation calibration and enables stress testing.	Not empirical validation or observed longitudinal evidence.

5.4 Hypotheses Evaluation

Table 5.2: Assessment of the Chapter 1 ranking hypotheses

Hypothesis	Evidence and rationale	Assessment
H1: Integrated versus scale-only ranking	The integrated top ten overlaps with 0 reach-only profiles and 2 follower-only profiles. Complete-order correlations are -0.4058 and -0.0109 , respectively.	Supported for ranking behaviour.
H2: Criterion sensitivity	Removing engagement quality changes the leader and leaves one default top-ten profile; omitting value alignment leaves eight. Omitting the other criteria leaves the default top ten intact, although lower positions change.	Partially supported; effects are heterogeneous.

These assessments concern deterministic ranking behaviour. They do not test campaign uplift or causal effects.

H1 is supported within its bounded ranking interpretation because the integrated ordering differs materially from both scale-only comparators. H2 receives qualified support:

the model is sensitive to criterion removal, but the magnitude is concentrated in engagement quality and value alignment rather than being uniform across criteria. A domain-adaptation hypothesis is not included in the evaluated hypotheses because it requires a reproducible labelled observed comparison with a generic sentiment baseline.

5.5 Theoretical, Methodological, and Managerial Implications

The conceptual contribution is a parameterised framework that joins brand-value alignment, sentiment context, audience-fit proxies, engagement-quality measures, and multi-criteria decision analysis. It makes the relevant constructs explicit and portable, while leaving their causal relationships open for future testing. This is a design and integration contribution rather than a confirmed theory test.

Methodologically, the study contributes a provenance-aware analytical pattern: observed records, simulated records, and simulation-augmented scenarios are separated; eligibility depends on observed evidence; and TOPSIS results are accompanied by scale-only baselines, equal-weight comparison, criterion ablations, and source-mixture checks. The framework preserves domain adaptation as an extensible modelling concept and TOPSIS as an auditable selection mechanism, without representing either as a verified performance gain.

For managers, the dashboard contribution is a structured way to inspect observed baseline summaries, evidence tiers, criterion inputs, and scenario-labelled ranking changes. It should support judgement and documentation, not automate creator selection. Its present outputs are not a substitute for a validated production monitoring system or for campaign experimentation.

5.6 Limitations and Threats to Validity

Provenance and coverage. The observed discourse baseline is restricted to a short 2015 legacy Twitter source, while observed campaign evidence is restricted to 683 records from the available platform panel. The simulated records extend the scenario layer but do not extend observed coverage. Platform access, source composition, and public-data availability restrict generalisation.

Synthetic calibration dependence. The close observed and simulated sentiment shares depend on how the simulation was calibrated. Such agreement is not independent confirmation and cannot establish trends, stability, or external validity beyond the observed source.

Proxy and construct validity. Value alignment, audience fit, sentiment context, engagement quality, and high resonance are operational proxies. In particular, language and location indicators cannot establish audience membership, while the resonance probability is not a directly observed business outcome. The correlation results are non-causal and may be affected by omitted variables and source-specific measurement choices.

Model evaluation. The reported resonance holdout metrics have not been reproduced with a traceable generating script, labels, split, and source composition. They cannot substantiate real-world accuracy, calibration, threshold reliability, or model improvement. There is likewise no traceable labelled observed comparison for the domain-adaptation hypothesis.

Decision validity. TOPSIS depends on chosen criteria and weights. Eligibility rules improve transparency but do not eliminate sparse-profile uncertainty. The source-mixture sensitivity analysis is deterministic and limited to its stated scenarios; it is not an outcome-based validation.

5.7 Future Work

Future research should first collect and preserve provenance for real campaign exposures, creator choices, audience responses, and independently measured outcomes. A preregistered or otherwise traceable evaluation can then compare the framework with non-contextual and single-factor selection rules on engagement, sentiment, conversion, or return-on-investment outcomes, and estimate causal effects where an appropriate design is feasible. Those outcome hypotheses would be distinct from the ranking-behaviour hypotheses evaluated in this thesis.

A second priority is a labelled observed aviation corpus with a documented generic baseline, domain-adapted variant, train/test split, and error analysis. That work would permit a direct assessment of the proposed domain-adaptation improvement and of any

calibration or threshold claims. Additional work may extend coverage to more platforms and modalities, while retaining source restrictions and privacy safeguards explicitly.

5.8 Conclusion

The Uganda Airlines case demonstrates a provenance-aware framework for organising observed discourse and campaign inputs into transparent decision-support outputs. The observed baseline establishes a negative-heavy 2015 sentiment distribution and supplies limited profile-level campaign evidence; the simulated and combined layers provide labelled stress-test scenarios. TOPSIS, domain-adaptation concepts, and the dashboard remain useful components of the framework, but their practical performance, causal effects, and production readiness require future validation with traceable real campaign evidence.

Chapter 6

Conclusion and Recommendations

6.1 Purpose, Scope, and Evidence Boundary

This thesis developed a parameterised framework for data-informed influencer selection and monitoring using Uganda Airlines as an aviation case. The framework combines a brand-values vector, an industry taxonomy, sentiment context, public-data alignment and audience-fit proxies, engagement-quality inputs, and TOPSIS ranking. Its evidence is deliberately bounded: 14 485 canonical observed discourse posts form the empirical baseline, while 125 200 simulated discourse records and 60 000 simulated campaign records support labelled scenario analysis. The 139 685-record discourse union is simulation-augmented, not fully observed longitudinal social-media evidence.

6.2 Summary of Findings by Objective

- **Objective 1: To describe the sentiment distribution and supplied topic structure in the observed aviation-discourse baseline.** In response to RQ1, the observed February 2015 baseline is negative-heavy: 62.6993% negative, 21.1874% neutral, and 16.1132% positive. The supplied observed topic field is `legacy_general`; therefore, the study does not claim observed aspect discovery or later observed sentiment trends.
- **Objective 2: To examine profile-level value-alignment, audience-fit, sentiment-context, engagement-quality, reach, and cost indicators in the observed campaign panel.** In response to RQ2, the panel contains 683 rows

and 95 influencer-platform profiles. Value alignment and sentiment context are associated with the supplied high-resonance probability proxy ($r = 0.78024905$ and $r = 0.5591605$, respectively), audience fit has a weak association ($r = 0.12995486$), and reach is negatively associated with engagement quality ($r = -0.64331418$). These are non-causal proxy associations, not observed return-on-investment effects.

- **Objective 3: To develop and demonstrate an evidence-aware TOPSIS shortlist that incorporates sentiment context, brand-value compatibility, audience fit, engagement quality, reach, and cost efficiency, and to compare it with transparent ranking baselines.** In response to RQ3, TOPSIS produces a reproducible observed ranking for the 64 profiles with at least five observed posts. Its top ten shares seven profiles with equal-weight TOPSIS, none with reach-only ranking, and two with follower-count-only ranking. The implemented rule therefore differs from scale-only sorting, but the comparison does not validate campaign effectiveness.
- **Objective 4: To examine source-mixture and criterion sensitivity and to present observed evidence and explicitly labelled scenario results separately in a management-oriented dashboard.** In response to RQ4, criterion ablations show the largest changes when engagement quality or value alignment is omitted, while nine profiles remain common across the tested source-mixture top tens and the leader remains stable. The dashboard presents observed summaries, simulated and combined stress-test scenarios, evidence tiers, ranking inputs, and sensitivity outputs with explicit provenance. The available evidence does not establish live operation, observed longitudinal trends, or predictive monitoring performance.

Similar observed and simulated sentiment shares should be understood as a result of simulation calibration. They support stress testing of scenario displays, not validation of the simulated data or confirmation of sentiment behaviour after the observed period.

6.3 Conclusions

The study's principal conclusion is that a transparent, values-aware decision-support framework can be specified and demonstrated with limited observed evidence. It makes the criteria for influencer shortlisting explicit, distinguishes observed evidence from

simulated scenarios, and documents where judgement enters the ranking. This is valuable for reproducibility and governance, even though it does not prove that the framework improves campaign outcomes.

The second conclusion is methodological. Provenance must remain visible throughout social-listening and influencer analytics. Separating observed, simulated, and simulation-augmented results prevents scenario outputs from being mistaken for empirical confirmation. The observed eligibility rule and TOPSIS source-mixture checks add useful controls, but they do not remove dependence on proxy measurements or decision preferences.

The third conclusion is prospective. Domain adaptation, a brand-values vector, an industry taxonomy, and a dashboard interface are reusable design components. The current evidence does not verify that domain adaptation improves accuracy, that calibrated thresholds are reliable, or that the dashboard can operate as a production monitoring system. Those claims require reproducible labelled evaluation and real campaign validation.

6.4 Contributions

Conceptual contribution

The thesis articulates a parameterised framework that links influencer-congruence concepts with sentiment context and multi-criteria selection (Belanche et al. 2021; Schouten et al. 2020). It clarifies the constructs that a future study can test—brand-value alignment, audience fit, engagement quality, and discourse sentiment—without claiming a new causal theory or that downstream effects have been established.

Methodological contribution

The methodological contribution is an auditable evidence structure: provenance-specific summaries, separation of simulation-augmented scenarios from observed results, observed-evidence eligibility tiers, scale-only ranking baselines, criterion ablations, and source-mixture sensitivity checks (Hwang and Yoon 1981). Domain adaptation is retained as a documented modelling concept that can be evaluated when a labelled observed corpus and baseline comparison are available. The unreproduced holdout metrics are preserved as a limitation rather than treated as validation evidence, and no Brier score is reported

without record-level probabilities and labels.

Practical contribution

The dashboard prototype can organise observed baseline statistics, profile evidence counts, criterion-level decision inputs, TOPSIS rankings, and labelled source-mixture scenarios in one reviewable interface. It supports managerial review and transparent documentation of assumptions. It is not evidence of a deployed live or near-real-time service, and its outputs should be reviewed alongside campaign context and human judgement.

6.5 Recommendations

6.5.1 Current dashboard use

The current dashboard can be used as a bounded decision-support artefact. Users should begin in observed mode, inspect source counts and evidence tiers, restrict default recommendations to profiles with at least five observed posts, and record the criteria and weights used for a TOPSIS shortlist. Simulated and combined views should be used only as clearly labelled stress-test scenarios. The 60/40, 70/30, and 80/20 source-mixture outputs can document sensitivity to that declared assumption; they should not be presented as forecast accuracy or campaign validation.

6.5.2 Production-validation steps

Before operational deployment, Uganda Airlines should collect traceable real campaign data: creator selection decisions, exposure dates, paid and organic reach, audience responses, costs, and independently measured outcomes. A pre-specified evaluation should compare the framework with non-contextual and single-factor baselines, rather than infer improvement from proxy associations. Where feasible, experimental or quasi-experimental designs should measure incremental effects.

The organisation should also create a labelled observed aviation corpus and document the model, generic baseline, split, source composition, and error analysis before relying on sentiment-model performance, calibration, or alert thresholds. Production work should include platform-compliant collection, provenance logging, privacy and advertising-disclosure controls, and periodic checks for source or model drift. Thresholds

and weighting choices should be empirically validated for the intended setting rather than adopted as fixed presets.

Table 6.1: From current decision support to production validation

Area	Current capability or next action	Evidence boundary
Observed review	Inspect baseline sentiment, campaign profiles, evidence tiers, and observed TOPSIS shortlist.	Descriptive public-data evidence; not campaign-outcome validation.
Scenario review	Compare labelled 60/40, 70/30, and 80/20 source-mixture outputs.	Deterministic stress test; not a forecast.
Campaign validation	Collect real exposure, cost, response, and outcome data; compare pre-specified baselines.	Required before performance or causal claims.
Model validation	Build a traceable labelled observed evaluation set and benchmark a generic model.	Required before accuracy, calibration, or threshold claims.

6.5.3 Guidance for other brands

Other brands can reuse the framework by defining their values and industry vocabulary, while retaining the same provenance discipline. They should treat value alignment, audience fit, and engagement measures as context-specific operationalisations, validate them against their own real outcomes, and avoid assuming that a simulated scenario transfers across platforms, markets, or time periods.

6.6 Limitations and Boundary Conditions

The conclusions are limited by provenance, coverage, and measurement. The observed discourse evidence comes from a restricted legacy Twitter period, and the campaign evidence comes from a limited public-data panel. Platform and source restrictions may omit relevant audiences and activity. The simulation layer is calibration-dependent: similarity to observed sentiment shares does not independently verify it. Alignment, audience, engagement, and high-resonance quantities are proxies whose validity may differ by platform and market; their observed associations are non-causal.

The available holdout metrics have not been reproduced with traceable labels, split, generating code, and source composition, so they do not demonstrate real-world model quality or reliable thresholds. No traceable labelled observed comparison evaluates a domain-adapted model against an off-the-shelf baseline. Finally, the observed evidence is too limited to establish longitudinal trends, production monitoring performance, or campaign uplift.

6.7 Future Research Directions

Future work should prioritise real campaign validation, labelled observed sentiment evaluation, and broader platform coverage before extending the system. Once those foundations exist, researchers may test causal designs for influencer selection, improve proxy measures with consented first-party analytics, examine multimodal content, and conduct fairness and privacy audits. Each extension should continue to report its provenance, source restrictions, and validation dataset explicitly.

6.8 Closing Remarks

The thesis provides a transparent foundation for discussing influencer selection with Uganda Airlines data: an observed baseline, a clearly labelled simulation-augmented stress-test layer, and an auditable TOPSIS decision process. Its strongest contribution is this disciplined separation of evidence and assumptions. Claims about model improvement, threshold reliability, causal effects, observed trends, and production operation remain future validation tasks rather than findings of the present study.

References

- Belanche, Daniel, Luis V. Casaló, Carlos Flavián, and Sergio Ibáñez-Sánchez (2021). “Understanding influencer marketing: The role of congruence between influencers, products and consumers”. In: *Journal of Business Research* 132, pp. 186–195. DOI: [10.1016/j.jbusres.2021.03.067](https://doi.org/10.1016/j.jbusres.2021.03.067) (cit. on pp. 1, 5, 9, 11, 16, 53, 60).
- Brier, Glenn W. (1950). “Verification of Forecasts Expressed in Terms of Probability”. In: *Monthly Weather Review* 78.1, pp. 1–3. DOI: [10.1175/1520-0493\(1950\)078<0001:VOFEIT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2) (cit. on pp. 18, 33).
- De Veirman, Marijke, Veroline Cauberghe, and Liselot Hudders (2017). “Marketing through Instagram influencers: the impact of number of followers and product divergence on brand attitude”. In: *International Journal of Advertising* 36.5, pp. 798–828. DOI: [10.1080/02650487.2017.1348035](https://doi.org/10.1080/02650487.2017.1348035) (cit. on pp. 1, 5, 16).
- Elemonie, Anastasia (2020). “The role of social media in shaping airline’s image”. In: *Jurnal Komunikasi Profesional* 4.2, pp. 105–118. DOI: [10.25139/jkp.v4i2.2965](https://doi.org/10.25139/jkp.v4i2.2965) (cit. on p. 19).
- Grandhi, Bala, Neeraj Patwa, and Kashif Saleem (2021). “Data-driven marketing for growth and profitability”. In: *EuroMed Journal of Business* 16.4, pp. 381–398. DOI: [10.1108/EMJB-09-2018-0054](https://doi.org/10.1108/EMJB-09-2018-0054) (cit. on p. 10).
- Guo, Chuan, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger (2017). “On Calibration of Modern Neural Networks”. In: *Proceedings of the 34th International Conference on Machine Learning (ICML)*, pp. 1321–1330 (cit. on pp. 18, 33).
- Harrigan, Paul, Trish M. Daly, Kristof Coussement, J. A. Lee, Geoffrey N. Soutar, and U. Evers (2021). “Identifying influencers on social media”. In: *International Journal of Information Management* 56, p. 102246. DOI: [10.1016/j.ijinfomgt.2020.102246](https://doi.org/10.1016/j.ijinfomgt.2020.102246) (cit. on pp. 1, 5, 9, 11, 16, 53).
- Hudders, Liselot, Steffi De Jans, and Marijke De Veirman (2021). “The Commercialization of Social Media Stars: A Literature Review and Conceptual Framework on the

- Strategic Use of Social Media Influencers”. In: *International Journal of Advertising* 40.3, pp. 327–375. DOI: [10.1080/02650487.2020.1836925](https://doi.org/10.1080/02650487.2020.1836925) (cit. on pp. 1, 3, 16, 53).
- Huerta-Álvarez, Rocío, Jesús J. Cambra-Fierro, and Maria Fuentes-Blasco (2020). “The Interplay between Social Media Communication, Brand Equity and Brand Engagement in Tourist Destinations: An Analysis in an Emerging Economy”. In: *Journal of Destination Marketing & Management* 16, p. 100413. DOI: [10.1016/j.jdmm.2020.100413](https://doi.org/10.1016/j.jdmm.2020.100413) (cit. on pp. 2, 4, 19).
- Hwang, Ching-Lai and Kwangsun Yoon (1981). *Multiple Attribute Decision Making: Methods and Applications*. Berlin, Heidelberg: Springer. DOI: [10.1007/978-3-642-48318-9](https://doi.org/10.1007/978-3-642-48318-9) (cit. on pp. 19, 24, 26, 60).
- Jin, S.-A. A. and J. Phua (2014). “Following celebrities’ tweets about brands: Source credibility, buying intention, and social identification”. In: *Journal of Advertising* 43.2, pp. 181–195. DOI: [10.1080/00913367.2013.827606](https://doi.org/10.1080/00913367.2013.827606) (cit. on pp. 10, 13).
- Joshi, Yatish, Weng Marc Lim, Khyati Jagani, and Satish Kumar (2025). “Social media influencer marketing: foundations, trends, and ways forward”. In: *Electronic Commerce Research* 25.2, pp. 1199–1253. DOI: [10.1007/s10660-023-09719-z](https://doi.org/10.1007/s10660-023-09719-z) (cit. on pp. 1, 3, 16, 53).
- Kanje, Patrokil, Goodluck Charles, Elly Tumsifu, Lena Mossberg, and Tommy Andersson (2020). “Customer Engagement and eWOM in Tourism”. In: *Journal of Hospitality and Tourism Insights* 3.3, pp. 273–289. DOI: [10.1108/JHTI-04-2019-0074](https://doi.org/10.1108/JHTI-04-2019-0074) (cit. on pp. 3, 4, 19, 20, 53).
- Kaplan, Andreas M. and Michael Haenlein (2010). “Users of the world, unite! The challenges and opportunities of Social Media”. In: *Business Horizons* 53.1, pp. 59–68. DOI: [10.1016/j.bushor.2009.09.003](https://doi.org/10.1016/j.bushor.2009.09.003) (cit. on pp. 10, 13).
- Leung, Felix F., Fangfang Gu, Yan Li, J. Z. Zhang, and Robert W. Palmatier (2022). “Influencer Marketing Effectiveness”. In: *Journal of Marketing* 86.6, pp. 93–115. DOI: [10.1177/00222429221102889](https://doi.org/10.1177/00222429221102889) (cit. on pp. 1, 9, 16).
- Liu, Bing (2012). *Sentiment Analysis and Opinion Mining*. Vol. 5. Synthesis Lectures on Human Language Technologies 1. Morgan & Claypool. DOI: [10.2200/S00416ED1V01Y201204HLT016](https://doi.org/10.2200/S00416ED1V01Y201204HLT016) (cit. on pp. 2, 11, 17, 24).
- Lou, Chen and Shupeí Yuan (2019). “Influencer Marketing: How Message Value and Credibility Affect Consumer Trust of Branded Content on Social Media”. In: *Journal of Interactive Advertising* 19.1, pp. 58–73. DOI: [10.1080/15252019.2018.1533501](https://doi.org/10.1080/15252019.2018.1533501) (cit. on pp. 1, 16, 53).

- Martin, Gati, Medard Edmund Mswahili, Young-Seob Jeong, and Jiyoung Woo (2022). “SwahBERT: Language Model of Swahili”. In: *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Seattle: Association for Computational Linguistics, pp. 303–313. DOI: [10.18653/v1/2022.naacl-main.23](https://doi.org/10.18653/v1/2022.naacl-main.23) (cit. on pp. 3, 17, 20, 53).
- Matiza, Tafadzwa and Elmarie Slabbert (2024). “The Celebrity Influencer: Delineating the Effect of Tourism-Oriented Short Video Marketing on International Tourist Conation”. In: *Journal of Global Marketing* 37.4, pp. 331–349. DOI: [10.1080/08911762.2024.2377741](https://doi.org/10.1080/08911762.2024.2377741) (cit. on pp. 2, 16, 20).
- Matthews, Brian W. (1975). “Comparison of the Predicted and Observed Secondary Structure of T4 Phage Lysozyme”. In: *Biochimica et Biophysica Acta (BBA) - Protein Structure* 405.2, pp. 442–451. DOI: [10.1016/0005-2795\(75\)90109-9](https://doi.org/10.1016/0005-2795(75)90109-9) (cit. on p. 18).
- Menon, Vishnu et al. (2019). “How to Grow Brand Post Engagement on Facebook and Twitter for Airlines? An Empirical Investigation of Design and Content Factors”. In: *Journal of Air Transport Management* 79, p. 101678. DOI: [10.1016/j.jairtraman.2019.05.002](https://doi.org/10.1016/j.jairtraman.2019.05.002) (cit. on pp. 2, 18).
- Morgan, Neil A. (2012). “Marketing and business performance”. In: *Journal of the Academy of Marketing Science* 40.1, pp. 102–119. DOI: [10.1007/s11747-011-0279-9](https://doi.org/10.1007/s11747-011-0279-9) (cit. on pp. 10, 16).
- Muhammad, Shamsuddeen Hassan et al. (2023). “AfriSenti: A Twitter Sentiment Analysis Benchmark for African Languages”. In: *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*. Singapore: Association for Computational Linguistics, pp. 13968–13981. DOI: [10.18653/v1/2023.emnlp-main.862](https://doi.org/10.18653/v1/2023.emnlp-main.862) (cit. on pp. 2, 3, 17, 20, 53).
- National Planning Authority (2013). *Uganda Vision 2040*. Tech. rep. Kampala: Government of Uganda. URL: <https://www.npa.go.ug/wp-content/uploads/2023/03/VISION-2040.pdf> (visited on 09/16/2026) (cit. on pp. 4, 5, 20).
- Pang, Bo and Lillian Lee (2008). “Opinion mining and sentiment analysis”. In: *Foundations and Trends in Information Retrieval* 2.1–2, pp. 1–135. DOI: [10.1561/15000000011](https://doi.org/10.1561/15000000011) (cit. on pp. 2, 11, 17, 24).
- Sakshi, Urvashi Tandon, Myriam Ertz, and Harbhajan Bansal (2020). “Social Vacation: Proposition of a Model to Understand Tourists’ Usage of Social Media for Travel Planning”. In: *Technology in Society* 63, p. 101438. DOI: [10.1016/j.techsoc.2020.101438](https://doi.org/10.1016/j.techsoc.2020.101438) (cit. on pp. 2, 4, 19).

- Schouten, Alexander P., Loes Janssen, and Marjolijn Verspaget (2020). “Celebrity vs. influencer endorsements in advertising: the role of identification, credibility, and product-endorser fit”. In: *International Journal of Advertising* 39.2, pp. 258–281. DOI: [10.1080/02650487.2019.1634898](https://doi.org/10.1080/02650487.2019.1634898) (cit. on pp. 1, 5, 9, 16, 53, 60).
- Sokolova, Karina and Hajer Kefi (2020). “Instagram and YouTube Bloggers Promote It, Why Should I Buy? How Credibility and Parasocial Interaction Influence Purchase Intentions”. In: *Journal of Retailing and Consumer Services* 53, p. 101742. DOI: [10.1016/j.jretconser.2019.01.011](https://doi.org/10.1016/j.jretconser.2019.01.011) (cit. on pp. 1, 16).
- Tanwar, Abhinav S., Harsh Chaudhry, and Manoj K. Srivastava (2022). “Trends in Influencer Marketing: A Review and Bibliometric Analysis”. In: *Journal of Interactive Advertising* 22.1, pp. 1–27. DOI: [10.1080/15252019.2021.2007822](https://doi.org/10.1080/15252019.2021.2007822) (cit. on pp. 1, 3, 16).
- United Nations Department of Economic and Social Affairs (2026a). *Goal 8: Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all*. URL: <https://sdgs.un.org/goals/goal8> (visited on 09/16/2026) (cit. on pp. 5, 20).
- (2026b). *Goal 9: Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation*. URL: <https://sdgs.un.org/goals/goal9> (visited on 09/16/2026) (cit. on pp. 5, 20).
- Weismueller, Jason, Paul Harrigan, Shasha Wang, and Geoffrey N. Soutar (2020). “Influencer Endorsements: How Advertising Disclosure and Source Credibility Affect Consumer Purchase Intention on Social Media”. In: *Australasian Marketing Journal* 28.4, pp. 160–170. DOI: [10.1016/j.ausmj.2020.03.002](https://doi.org/10.1016/j.ausmj.2020.03.002) (cit. on pp. 1, 16).
- Wu, Shengyang and Yi Gao (2024). “Machine Learning Approach to Analyze the Sentiment of Airline Passengers’ Tweets”. In: *Transportation Research Record* 2678.2. DOI: [10.1177/03611981231172948](https://doi.org/10.1177/03611981231172948) (cit. on pp. 2, 18).

Data, Analysis, and Dashboard Access Register

This appendix documents concrete access references for the datasets, generated analysis artefacts, and deployed dashboard prototype referred to in Chapters 3, 4, and 6.

Artefact Register

Table 1: Access register for data and analysis artefacts

Artefact	Reference	Notes
Discourse analysis dataset	<code>tweets_cleaned.csv</code>	Input used for sentiment, trend, and retweet analyses.
Influencer analysis datasets	<code>influencers_cleaned.csv</code> ; <code>top_influencers_overall.csv</code>	Inputs used for influencer profiling, MCDA/TOPSIS ranking, and sensitivity checks.
Generated figure outputs	<code>chapters/chapter04/figures/</code>	Exported PNG outputs used in Chapter 4 figures.
Dashboard prototype	Railway dashboard URL	Deployed dashboard page that consolidates Objective 4 monitoring signals.

Dashboard Entry Point

Deployed dashboard URL:

<https://influencersentimentanalysisdashborad.up.railway.app/>

Clickable web link:

[Open deployed dashboard](#)