



School of Business and Management

Royal Holloway, University of London

**Assessing Consistency and Transparency in MotoGP Steward Decisions: An
Analysis of 2022 Rulings and Stakeholder Sentiment**

MN5821 CONSULTING PROJECT

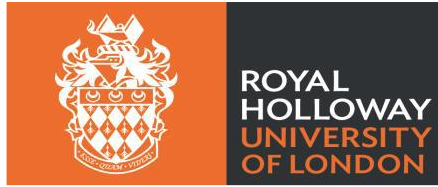
MSc Business Analytics

Student Name: Sushant Thapa Magar

College Number: 101064383

Supervisor: Dr: Timothy Ramjuan

Date: 28th August 2024



School of Business and Management

Royal Holloway, University of London

Declaration of Authenticity

This Consulting Project has been prepared on the basis of my own work, and where other published and unpublished source materials have been used, these have been acknowledged.

Word Count: 9,992

Number of contacts with supervisor: 35
including email and face to face meetings

Student Name: Sushant Thapa Magar

College Number: 101064383

Signature Sushant Thapa Magar

Date of Submission: 28-08-2024

Executive Summary

This consulting project critically examines the consistency, transparency, and stakeholder sentiment surrounding the MotoGP Stewards Panel's decisions during the 2022 season. Given the high stakes in professional racing, where decisions can significantly impact riders' careers and the sport's integrity, ensuring fair and consistent decision-making is crucial. MotoGP, Moto2, and Moto3 were the primary focus of this analysis, leveraging a mixed-methods approach that combines descriptive, diagnostic, and sentiment analysis. Instagram was chosen as the primary data source for sentiment analysis due to its high engagement levels among MotoGP stakeholders, providing a robust dataset for evaluating perceptions of fairness.

The project aimed to assess how consistent and transparent the Stewards Panel's decisions were and understand how these decisions were perceived by various stakeholders, particularly on social media. Key objectives included identifying contentious decisions, analyzing the consistency of sanctions across similar cases, and evaluating stakeholder sentiment. Ultimately, the project sought to provide actionable recommendations for enhancing decision-making processes within MotoGP.

The methodology employed in this project involved a mixed-methods approach, analyzing both quantitative and qualitative data. Descriptive analysis was conducted to identify key trends in rule enforcement, while diagnostic analysis explored the underlying causes of any inconsistencies observed. Sentiment analysis, supported by tools like BERT, was employed to gauge stakeholder's perception, with a focus on Instagram due to its significant role in MotoGP's digital engagement. Data from official MotoGP records were cross-referenced to ensure accuracy, and a manual verification process was used to enhance the reliability of sentiment classifications.

The analysis revealed that "Irresponsible Riding" was the most frequent infringement, accounting for over two-thirds of recorded infractions. This finding highlights persistent challenges in enforcing riding standards across all classes. "Long Lap Penalties" emerged as the most common sanction, preferred for their direct impact on race outcomes without being overly punitive. However, inconsistencies in the application of penalties across different circuits and rider classes raised concerns about fairness. However, inconsistencies in the application of penalties across different circuits and rider classes raised concerns about fairness.

The diagnostic phase identified specific factors contributing to these inconsistencies, such as differing interpretations of rules at various tracks and the varying severity of penalties for similar infractions. Sentiment analysis showed that negative reactions were most pronounced among riders, especially in cases involving severe penalties. This underscores the need for more standardized penalty guidelines and improved communication between the Stewards Panel and stakeholders to reduce perceptions of bias. Stakeholder trust, particularly among riders, appeared to be fragile, and the inconsistency in penalties only exacerbated this issue.

To address these issues, the project recommends standardizing penalty guidelines to ensure consistent treatment of similar infractions across all classes. This would involve developing clear, transparent rules that all stakeholders understand and agree upon. Enhancing rider education, particularly in Moto3, is also crucial.

Younger or less experienced riders in this class are more prone to risky behaviors, and targeted education programs could help reduce the incidence of infractions. Improving communication strategies is another key recommendation. By drawing on best practices from other equivalents MotoGP can increase transparency by clearly communicating the rationale behind decisions to riders, teams, and the public. Furthermore, the project suggests tailoring regulatory approaches to specific classes and circuits. Developing class- and circuit-specific regulations that address unique challenges would ensure that rules are appropriately enforced across different contexts. Finally, engaging media outlets more effectively is essential. Proactively working with media to manage public perception and ensure balanced reporting of decisions could help reduce misunderstandings and negative sentiments.

The findings of this project have significant implications for MotoGP's reputation and governance. By standardizing penalty guidelines and improving communication, the Stewards Panel can enhance stakeholder trust, reduce negative perceptions, and maintain the sport's integrity. These recommendations also have broader relevance for other sports facing similar challenges in decision-making and stakeholder management. The importance of transparency and consistency in maintaining fairness cannot be overstated, especially in a sport as competitive and scrutinized as MotoGP. Implementing these changes could position MotoGP as a leader in sports governance, demonstrating a commitment to fairness and integrity.

This project contributes to the field of sports governance by integrating sentiment analysis into the evaluation of steward decisions. It challenges the practical application of Stakeholder Theory in high-pressure environments like MotoGP, highlighting the complexities of balancing diverse stakeholder interests. The study also reinforces the importance of adaptive decision-making, as outlined in the Ecological Dynamics Approach. By offering practical steps for improving the fairness and transparency of MotoGP's stewarding processes, this project provides a valuable framework that can be applied to other sports.

While the project provides valuable insights, it is important to acknowledge certain limitations. The reliance on Instagram as the primary data source may introduce platform-specific biases, limiting the generalizability of the findings. Additionally, the focus on the 2022 season restricts the ability to draw conclusions about long-term trends. Future research should expand the analysis to include additional platforms and seasons, as well as qualitative methods such as interviews with key stakeholders, to provide a more comprehensive understanding of the Stewards Panel's decision-making processes.

In conclusion, this project successfully met its objectives by providing a detailed analysis of the 2022 MotoGP season's steward decisions. The findings and recommendations offer practical solutions for enhancing consistency, transparency, and fairness in MotoGP's decision-making. By implementing these recommendations, MotoGP can strengthen stakeholder trust and maintain its reputation as a competitive and fair sport. As the sport continues to evolve, it is crucial that its governance structures adapt accordingly, ensuring that all stakeholders feel confident in the integrity of the competition.

Acknowledgements

I would like to extend my deepest gratitude to all those who supported and guided me throughout the course of this project.

Firstly, I would like to express my sincere appreciation to my supervisor, Dr. Timothy Ramjuan, for his invaluable guidance, insightful feedback, and unwavering support throughout the project. His expertise and encouragement were instrumental in shaping the direction of my work and ensuring its successful completion.

I am equally grateful to my co-supervisor, Dr. Asmat Monaghan, whose detailed insights and thoughtful advice added significant depth to my analysis. Her mentorship played a critical role in refining my approach and broadening my understanding of the subject matter.

A special thanks goes to my project collaborators and clients, Dr. Mark Lycett and Dr. Jane Marriott, for their continuous support, constructive feedback, and for providing the necessary resources and datasets to bring this project to fruition. Their expertise and vision were pivotal in shaping the project's outcomes.

I would also like to extend my heartfelt thanks to my team member, Shruti Suryavanshi, for her dedication, collaboration, and hard work. Her contributions, especially in data collection and sentiment analysis, were invaluable, and her partnership throughout the project was truly appreciated.

Finally, I am grateful to everyone who provided support, feedback, and encouragement during this journey. This project would not have been possible without the collective efforts of all those involved.

Thank you all for your guidance, collaboration, and unwavering support.

Table of Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 1 |
| 1.1 | Background..... | 1 |
| 1.2 | Project Aim and Objectives..... | 2 |
| 1.3 | Research Approach | 3 |
| 1.4 | Report Outline | 4 |
| 2 | Trends and Theoretical Perspectives | 5 |
| 2.1 | Current Trends | 5 |
| 2.2 | Theoretical Frameworks..... | 6 |
| 2.3 | Critical Evaluation | 7 |
| 2.4 | Knowledge Gap and Insights Sought..... | 8 |
| 3 | Analytical Framework and Method | 9 |
| 3.1 | Research Methodology..... | 9 |
| 3.2 | Research Design | 10 |
| 3.3 | Ethical Considerations | 11 |
| 4 | The Data..... | 13 |
| 4.1 | Data Description | 13 |
| 4.2 | Source Justification | 14 |
| 4.3 | Data Preparation | 14 |
| 4.4 | Data Assumptions and Limitations | 16 |
| 5 | The Analysis | 17 |
| 5.1 | Analysis Progression..... | 17 |
| 5.2 | Descriptive Analytics | 17 |
| 5.3 | Diagnostic Analytics | 30 |
| 5.4 | Predictive Analytics | 36 |
| 5.5 | Prescriptive Analytics | 36 |
| 5.6 | Analytical Adjustments and Assumptions | 37 |
| 6 | Findings and Recommendations | 38 |
| 6.1 | Project Contribution..... | 39 |
| 6.2 | Recommendations..... | 39 |
| 6.3 | Project Limitations..... | 41 |
| 6.4 | Future Opportunities for Research..... | 41 |
| 7 | Conclusion..... | 43 |
| 8 | Bibliography..... | 45 |
| | Appendices | 47 |
| | Appendix A: Ethical Approval Form | 48 |
| | Appendix B: Project Development Diary | 55 |
| | Appendix C: Declaration of AI Tool Usage | 76 |
| | Appendix D: Change Approval Form..... | 77 |

1 Introduction

1.1 Background

Fair and transparent decision-making is crucial for MotoGP, a sport where race outcomes can significantly impact riders' careers and the championship's reputation. The MotoGP Stewards Panel's decisions shape not only race results but also perceptions of fairness among teams, riders, and fans. Inconsistent or biased decisions can erode trust and harm the sport's credibility, especially in such a competitive environment where margins between victory and defeat are often minimal. Research by (Petersen and Wichmann, 2021) highlights how implicit biases can influence decision-making in competitive environments, which may have serious consequences for fairness and integrity in sports like MotoGP.

Several incidents highlight the importance of consistent and transparent decision-making in MotoGP. A notable example is the 2015 Malaysian Grand Prix, where Valentino Rossi was penalized after an incident with Marc Márquez. Rossi's penalty, which forced him to start from the back of the grid in the final race, sparked widespread debate. Many viewed the decision as influenced by external pressures rather than fairness, affecting Rossi's championship chances and raising concerns about potential biases within the Stewards Panel. The controversy surrounding this penalty not only affected Rossi's title chances but also brought attention to potential biases within the Stewards Panel's decision-making process, as noted by BBC Sport.

These incidents underscore the need for transparency and consistency in MotoGP decision-making. Perceived unfairness erodes stakeholder trust and undermines the sport's integrity. Social media, particularly Instagram, amplifies public opinion, allowing riders, teams, and fans to express their views in real-time. This immediacy can shape the narrative around disciplinary decisions. As (Javani and Karimivand, n.d.) note, athletes often use social media to influence public opinion when they feel wronged. This report analyzes such cases to evaluate the current disciplinary framework and recommends improvements for greater transparency and fairness in Stewards Panel decisions.



Figure 1: Word cloud highlighting key terms from Rider and Team posts during the 2022 MotoGP season.

1.2 Project Aim and Objectives

The aim of this project is to critically assess the consistency and transparency of the MotoGP Stewards Panel's rulings during the 2022 season across MotoGP, Moto2, and Moto3 classes, focusing on stakeholder sentiment on Instagram. This approach enables a detailed examination of how these rulings are perceived by riders, teams, media, and fans, whose trust is crucial to the sport's integrity.

The project objectives include:

1. Identifying contentious decisions that sparked significant debate among stakeholders.
2. Classifying decisions by infringement type to analyze consistency across similar cases.
3. Evaluating stakeholder sentiment using Instagram to gauge perceptions of fairness and transparency.
4. Providing actionable recommendations to enhance the Stewards Panel's transparency, consistency, and trustworthiness.

By concentrating on the 2022 season and analyzing social media sentiment via instagram, this study aims to pinpoint areas for improvement and contribute to the ongoing discussion on better governance in MotoGP.

1.3 Research Approach

This project uses a mixed-methods approach, combining quantitative analysis and qualitative sentiment analysis to evaluate the consistency and transparency of the MotoGP Stewards Panel's rulings during the 2022 season. The quantitative analysis involves categorizing the Stewards' decisions by infringement types and sanctions, using official MotoGP records as the data source.

The qualitative aspect centers on sentiment analysis, with Instagram serving as the primary data source. Stakeholder sentiments, including those from riders, teams, media, and fans, are analyzed through third-party software that categorizes comments related to specific incidents. By comparing sentiments across various decisions, the project aims to assess the perceived fairness and effectiveness of the Stewards' rulings.

Key assumptions include Instagram's role as a primary platform for MotoGP community engagement, making it a representative source of sentiment. Additionally, it is assumed that the Stewards' criteria for disciplinary decisions remained consistent throughout the season, allowing for reliable comparisons. The accuracy of sentiment analysis tools, despite challenges like sarcasm or contextual language, is also assumed.

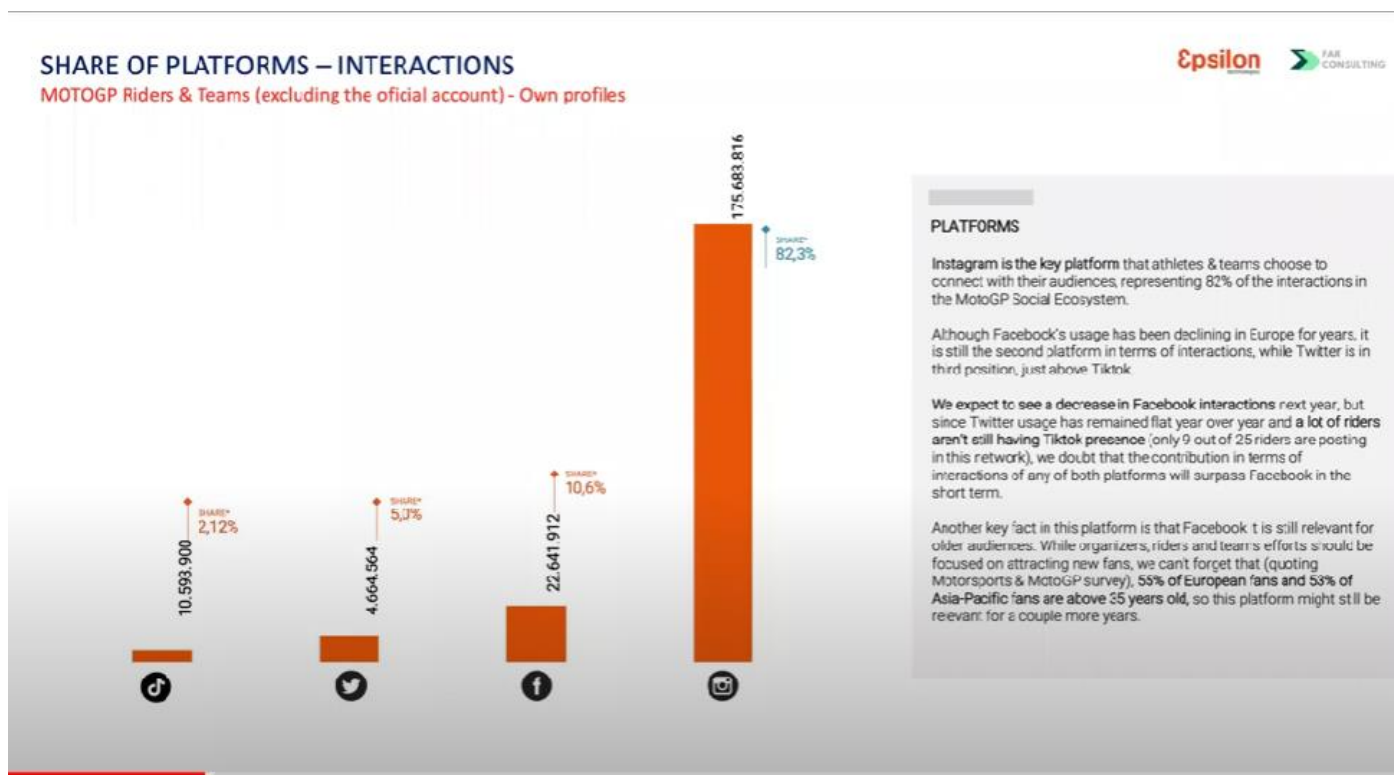


Figure 2: Distribution of social media interactions for MotoGP riders and teams, highlighting Instagram's dominance.

Understanding these assumptions is crucial for evaluating the analysis's potential limitations, ensuring a balanced examination of the Stewards' decision-making process.

1.4 Report Outline

The report provides a comprehensive analysis of the MotoGP Stewards Panel's decision-making process. It begins by exploring the context and significance of fair decision-making in MotoGP, using historical incidents to highlight the importance of consistency and transparency. Following this, the report discusses the theoretical framework and current trends in sports governance, establishing a foundation for understanding the broader implications of the Stewards' decisions.

The subsequent sections focus on the analytical framework and methodology used to evaluate the Stewards Panel's decisions, with a particular emphasis on key incidents from the 2022 season. This is followed by a sentiment analysis based on Instagram data, exploring how public discourse has responded to these decisions.

The report concludes with actionable recommendations to improve the transparency and consistency of the Stewards Panel's decisions. These suggestions aim to enhance stakeholder trust and ensure that MotoGP remains perceived as a fair and competitive sport. The structure guides readers from identifying issues to presenting evidence and practical solutions.

2 Trends and Theoretical Perspectives

This section explores the current trends and theoretical perspectives relevant to decision-making in MotoGP, particularly focusing on fairness and transparency in disciplinary actions. By analysing recent developments and key theories, this section sets the foundation for the subsequent analysis of the MotoGP Stewards Panel's decision-making process. The section is structured into four parts: current trends in the field, theoretical frameworks that inform our approach, a critical evaluation of these perspectives, and the identification of knowledge gaps that this project aims to address.

2.1 Current Trends

In recent years, MotoGP governance has faced increasing scrutiny, particularly regarding the Stewards Panel's consistency and transparency. High-profile incidents, like the 2022 Catalan Grand Prix involving Takaaki Nakagami, have sparked widespread debate. Nakagami's unpenalized braking, causing crashes, highlights the Stewards' challenge in maintaining perceived fairness.

A key factor driving this scrutiny is the rise of social media as a platform for stakeholder engagement. Platforms like Instagram now play a crucial role in shaping public perception, with fans, riders, and teams expressing opinions more visibly than ever. This shift mirrors the broader evolution of sports media, where social platforms like Facebook, Instagram, and TikTok have become primary means of fan engagement, surpassing traditional television (Forbes, 2022).

MotoGP's social media presence, with over 50 million followers—up 25% from the previous year—demonstrates this growing influence (MotoGP, 2023). Instagram posts, especially those addressing controversial decisions, regularly attract significant engagement, underscoring social media's role in shaping public discourse. This trend is reflective of the broader shift in sports media, where digital platforms drive fan engagement and revenue through sponsorships (Forbes, 2022).

The shift to digital engagement highlights the importance of understanding and managing public sentiment. Stewards' decisions are no longer confined to the racetrack but are scrutinized globally in real time. This amplifies the impact of each decision, making consistent and transparent rulings more critical (Javani and Karimivand, n.d.). Research shows that effective digital engagement significantly enhances fan interaction, further emphasizing the need for transparency in decision-making (Javani and Karimivand, n.d.).

These developments are crucial for this project, which analyzes Stewards' decision consistency through stakeholder sentiment on social media, providing insights into MotoGP governance.

2.2 Theoretical Frameworks

The decision-making processes within the MotoGP Stewards Panel can be understood through several theoretical frameworks, each offering a unique perspective on the dynamics at play. Stakeholder Theory, articulated by (Freeman and McVea, 2005), is foundational. It explains the relationships between the Stewards Panel and various stakeholders, including riders, teams, fans, sponsors, and regulatory bodies. This theory posits that organizations must consider all stakeholders' interests to maintain legitimacy and long-term success. (Donaldson and Preston, 1995) expanded on this by highlighting the descriptive, instrumental, and normative aspects of Stakeholder Theory, emphasizing its role in guiding decision-making and ethical obligations. (Mitchell et al., 1997) introduced stakeholder salience, prioritizing stakeholders based on power, legitimacy, and urgency, which helps the Stewards Panel navigate conflicting interests.

Sentiment Analysis and Opinion Mining are increasingly relevant in understanding how the Stewards Panel's decisions are perceived by the public and stakeholders. Techniques like BERT (Bidirectional Encoder Representations from Transformers), developed by (Devlin et al., 2018), allow the analysis of large volumes of social media data, capturing real-time sentiments. Studies like (Yu and Wang, 2015) analysis of Twitter activity during the 2014 FIFA World Cup show how fan emotions fluctuate in response to sports events. In MotoGP's highly publicized environment, sentiment analysis helps assess the broader impact of Stewards' decisions on stakeholders and manage reactions in real time.

The Ecological Dynamics Approach, discussed by (Balagué et al., 2008), offers an innovative perspective on sports decision-making. It integrates ecological psychology and dynamic systems theory, emphasizing the role of environmental, task, and personal constraints in shaping decisions. This approach is particularly relevant in MotoGP, where the Stewards Panel must adapt to constantly changing race conditions.

These theoretical frameworks collectively inform a comprehensive approach to analyzing the decision-making processes within the MotoGP Stewards Panel. They highlight the importance of considering multiple perspectives and the dynamic interplay of various factors in maintaining the integrity and transparency of the sport. Stakeholder Theory ensures that all interests are considered ethically.

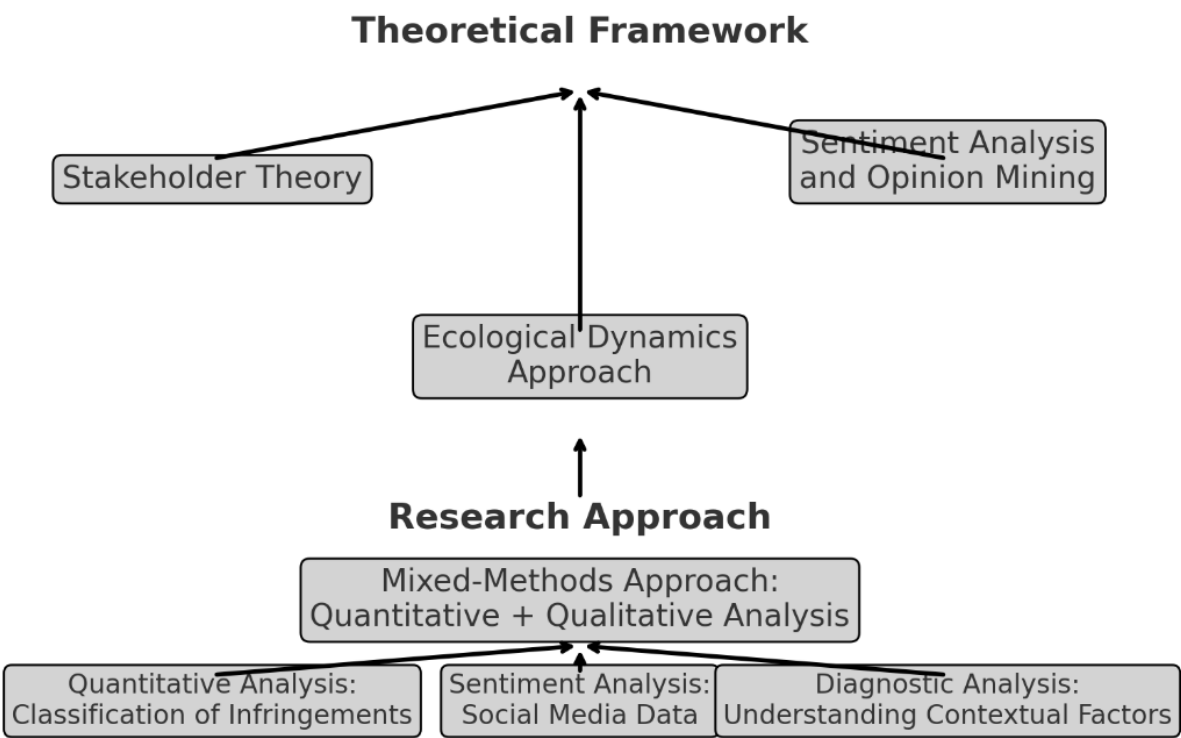


Figure 3: Theoretical Framework

2.3 Critical Evaluation

The critical evaluation of theoretical frameworks applied to the MotoGP Stewards Panel’s decision-making processes reveals both strengths and challenges. Stakeholder Theory is a fundamental framework for understanding the ethical obligations and responsibilities of the Stewards Panel. This theory emphasizes the need to consider the interests of all stakeholders, which is crucial in a sport as complex as MotoGP. Applying Stakeholder Theory in real-time, high-pressure environments like MotoGP can be challenging. The complexity of balancing the competing interests of riders, teams, fans, and sponsors sometimes overwhelms the decision-making process. This can lead to difficulties in making fair and transparent decisions. (Kujala et al., 2022) support this, highlighting that real-time stakeholder management often involves practical challenges, particularly in prioritizing conflicting demands.

Sentiment Analysis and Opinion Mining are increasingly used to gauge public and stakeholder sentiment. Tools like BERT for real-time social media analysis are valuable for understanding immediate reactions to the Stewards’ decisions. However, (Ashford et al., 2021) note that while sentiment analysis provides quick insights, it can sometimes misinterpret the nuanced and context-specific sentiments expressed by different

stakeholders. (Jiao et al., 2021) also highlight the limitations of sentiment analysis, particularly the risk of bias due to the selective nature of social media data, which can lead to misinterpretations.

The Ecological Dynamics Approach offers a model for decision-making by considering the non-linear interactions of personal, task, and environmental factors. This approach is particularly relevant for the fast-paced and unpredictable MotoGP environment. However, the abstract nature of this model poses challenges in consistent application across various scenarios. (Balagué et al., 2008) discuss these challenges, noting that while the approach provides valuable insights, its abstractness can make it difficult to apply in a structured manner.

In summary, while these frameworks provide essential guidance, each has its limitations. Stakeholder Theory, although ethically robust, can struggle in dynamic environments. Sentiment Analysis offers real-time insights but requires careful interpretation to avoid biases. The Ecological Dynamics Approach contributes to understanding adaptability, but its abstract nature can limit its practical application. By considering these strengths and limitations, the Stewards Panel can develop a more balanced and effective decision-making framework that meets the needs of all stakeholders in the fast-paced and ever-changing world of MotoGP racing.

2.4 Knowledge Gap and Insights Sought

Current research on decision-making in MotoGP lacks a comprehensive integration of real-time sentiment analysis into the decision-making processes of the Stewards Panel. This project will explore how sentiment data can be utilised to enhance the fairness and transparency of these decisions. By addressing this gap, the project aims to develop strategies that align decisions more closely with stakeholder expectations and improve overall governance.

3 Analytical Framework and Method

3.1 Research Methodology

Sentiment analysis was selected to capture stakeholder perceptions in the dynamic environment of MotoGP. Instagram, supported by literature (Alam et al., 2022), serves as an effective tool for real-time sentiment analysis, crucial in high-pressure contexts like MotoGP. Other research (Yu and Wang, 2015) further validate this approach, demonstrating its effectiveness in analyzing emotional responses in sports. Infringement classification adds depth to the analysis by evaluating the consistency and fairness of Stewards' decisions. This methodology, grounded in Stakeholder Theory, highlights the importance of balancing diverse interests (Donaldson and Preston, 1995; Freeman and McVea, 2005).

Sentiment analysis is particularly well-suited for this project's aim of assessing the consistency and transparency of the MotoGP Stewards Panel's rulings. This method enables real-time analysis of sentiments expressed by key stakeholders—riders, teams, media, and fans—on platforms like Instagram, where language can be highly contextual and nuanced. Alam et al. (2022) highlighted Instagram's effectiveness in analyzing public sentiment in critical, high-stakes scenarios, such as vaccine hesitancy. BERT's deep learning architecture excels in classifying sentiments by considering word context, enhancing accuracy in social media analysis (Devlin *et al.*, 2018). Recognizing that no automated tool is infallible, this project employs a hybrid approach, integrating BERT with manual cross-verification to ensure both scalability and precision. In addition, the classification of decisions into different infringement types and the evaluation of decision consistency across similar cases align with the project's broader objectives, ensuring that all aspects of the Stewards' decision-making process are comprehensively analyzed.

This project integrates both content analysis and sentiment analysis to effectively categorize and interpret vast amounts of social media data. While content analysis excels in systematically organizing and coding text, it alone is not scalable enough for the continuous data generated in contexts like MotoGP. Qualitative interviews offer deep insights but are impractical for capturing widespread online sentiments. To address these challenges, sentiment analysis with BERT enhances this project by adding scalability and precision, capturing the context and nuance that content analysis structures. The hybrid method of combining automated sentiment classification with manual cross-verification ensures both thoroughness and accuracy, surpassing what either method alone could achieve.

3.2 Research Design

The research process systematically captured and analyzed stakeholder sentiments and classified MotoGP Stewards Panel rulings during the 2022 season. Instagram posts and comments related to MotoGP, Moto2, and Moto3, focusing on specific hashtags, rider accounts, and team pages, were the primary data sources. Media outlet sentiments were analyzed through published articles due to limited social media engagement. The infringement dataset was obtained from Excel spreadsheets provided by collaborators, documenting the stewards' decisions throughout the season. ExportComment was chosen for automated data scraping, selected after considering ethical, legal, and effectiveness factors compared to other tools like TwSearch and Apify. The dataset was thoroughly cleaned to ensure integrity by removing irrelevant or duplicate entries. Non-English posts were translated using the deep_translator library, though the limitations of machine translation tools, as noted by (Patil and Davies, n.d.), were acknowledged. Sentiment classification was conducted using BERT, with manual cross-verification, and decisions were categorized to assess consistency across similar cases.

The research employed a combination of sentiment analysis and infringement classification to meet the project's objectives. Python was the primary programming language, utilizing libraries like Pandas for data manipulation, deep_translator for translation, and WordCloud for visualizing key sentiment terms. Tools like VADER and TextBlob were tested but ultimately not used due to their limitations. Excel was extensively used for organizing, cleaning, and analyzing data. Python Libraries supported the creation of visualizations, helping to present both sentiment and infringement classification findings clearly.

The research design was carefully aligned with the project's aim to assess stakeholders sentiment and classify decisions made by the MotoGP Stewards Panel. By combining sentiment analysis with classification of infringement types, the design provided a comprehensive examination of stakeholder perceptions and the consistency of rulings, addressing the research questions related to consistency, transparency, and public reaction. The iterative nature of the research allowed for adjustments based on initial findings, ensuring focused and relevant analysis.

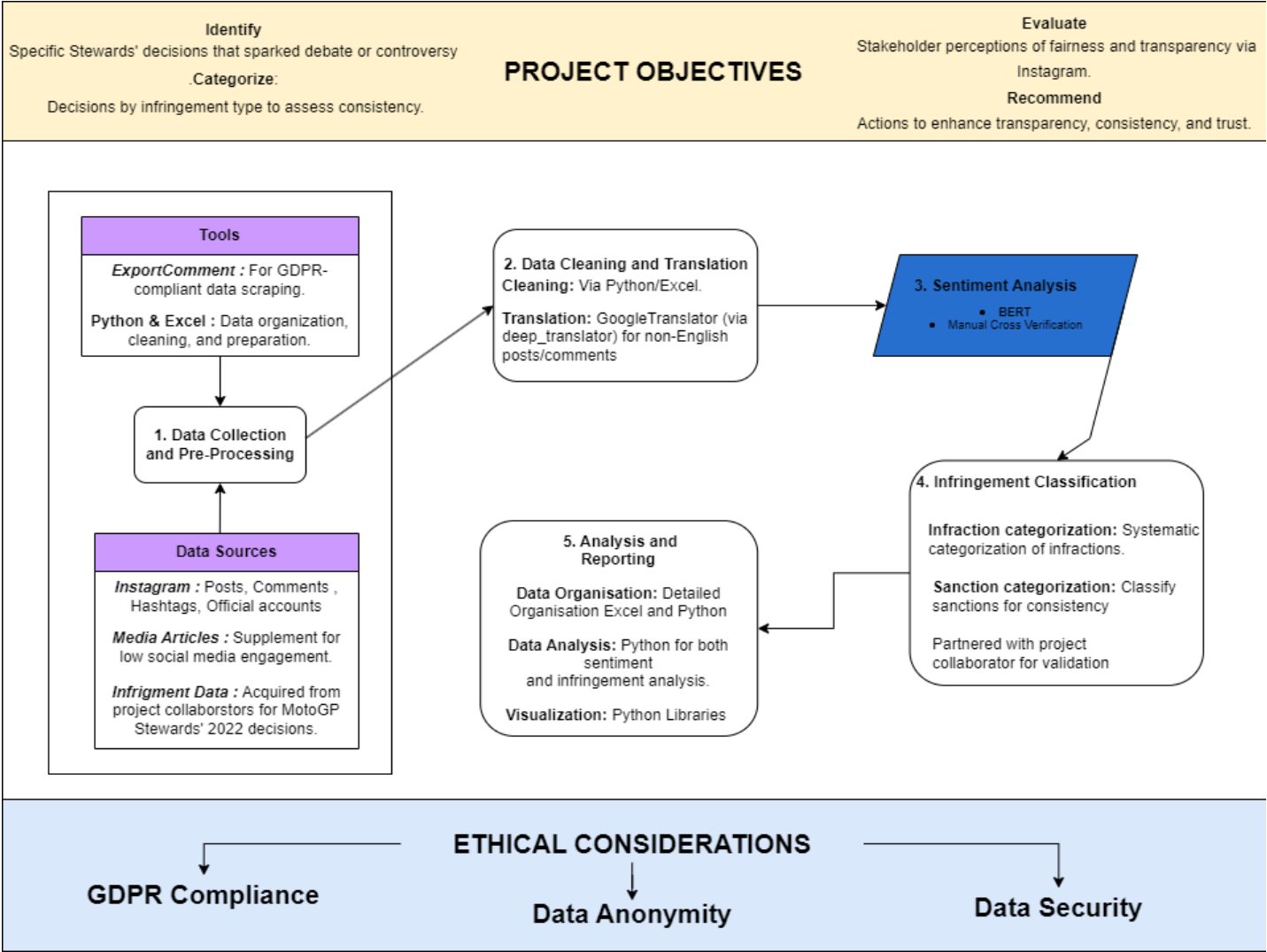
3.3 Ethical Considerations

This research carefully considered ethical issues related to data privacy and participant anonymity, particularly when using third-party tools like ExportComment for data scraping. The ethical framework guiding this research was rooted in principles of data ethics as macroethics (Floridi and Taddeo, 2016). This framework was integral in selecting data collection tools that complied with GDPR and ensuring that all research activities were aligned with ethical standards that prioritize participant privacy and data security. Communication with ExportComment confirmed their compliance with GDPR and their commitment to ethical data processing, focusing solely on publicly available content. All data collection activities adhered to GDPR guidelines, ensuring that no private information was accessed or stored. The ethical implications of using public data were discussed internally and externally to ensure compliance with Royal Holloway University of London's ethical standards. No covert operations or interactions with vulnerable populations were involved, minimizing the ethical risks associated with the research.

Participant privacy and anonymity were safeguarded by ensuring that all collected data was anonymized where possible, and sensitive information was excluded from the analysis. Our approach to data ethics was informed by the broader principles of data ethics as a macroethics, as emphasized by (Lundberg *et al.*, 2019). The research avoided identifying individuals, instead focusing on aggregated public sentiments expressed on social media platforms, thereby respecting user privacy and adhering to ethical guidelines for social media research.

Since the data utilized in this study was publicly accessible, formal informed consent was not required under GDPR guidelines. However, the research complied with domain-specific ethical guidelines by focusing on public sentiment analysis and avoiding any direct interaction with or identification of individuals. Ethical approval was obtained from the Royal Holloway University Ethics Committee, ensuring full compliance.

In line with the principles of ethical data management (Lundberg *et al.*, 2019), all collected data was anonymized and stored on password-protected systems, ensuring compliance with GDPR and institutional guidelines. This approach minimized risks to participant privacy and maintained the integrity of the data throughout the research process.



4 The Data

4.1 Data Description

The data used in this project is critical for evaluating the consistency and transparency of the MotoGP Stewards Panel's decisions during the 2022 season. The analysis draws on both received and created datasets, including the MotoGP Stewards Panel Decisions Dataset, Infringement Dataset, and various Sentiment Datasets.

MotoGP Stewards Panel Decisions Dataset (Received):

This dataset, provided by collaborators, documents stewards' decisions throughout the 2022 MotoGP season across MotoGP, Moto2, and Moto3 classes. It includes details such as the date, time, circuit, class, event stage, involved rider/team, infraction description, contravened articles, and penalties. Additional information covers hearing details, appeal rights, and contextual notes. This dataset is essential for analyzing contentious decisions and assessing ruling consistency.

Infringement Dataset (Created):

Developed by refining the MotoGP Stewards Panel Decisions Dataset, this dataset focuses on classifying decisions by infringement type, covering 56 decisions from 2022. It includes original variables and additional columns for Public, Media, Team, and Riders Sentiment. These enhancements support analyzing and visualizing stakeholder reactions, offering a comprehensive evaluation of decision perceptions and consistency.

Sentiment Datasets (Created):

The Sentiment Datasets were created from Instagram posts, comments, and blog articles related to the 2022 MotoGP season. The process began with collecting relevant URLs from Instagram and identifying articles that directly addressed the sentiments of riders, teams, and media outlets in relation to the project objectives. Approximately 20,000 public comments were collected, with 54 comments being identified and separated as originating from official accounts. These datasets are categorized into Riders Sentiment, Team Sentiment, Public Sentiment, and Media Sentiment. Media sentiment was derived from 87 blog articles from different media houses quoting riders and teams, which were then transferred to the respective sentiment datasets. Key variables include Infraction_ID, Stakeholder Type, Content_Type, Context, and Sentiment. These datasets are essential for analyzing stakeholder perceptions of stewards' decisions, offering detailed insights into the sentiment surrounding each infraction.

Other Datasets (Created):

Two additional datasets were utilized in this analysis. The Consolidated Accounts List was used primarily to cross-verify comments from official accounts, ensuring that these were correctly classified as stakeholder opinions rather than public sentiment. The Cleaned_Cross_Verification_Results dataset separated 54 official comments from the public domain, distinguishing between general public opinion and stakeholder perspectives for accurate sentiment analysis.

4.2 Source Justification

The data sources were selected based on their relevance and high engagement within the MotoGP community. Instagram emerged as a highly effective platform for capturing stakeholder sentiment, accounting for 82.3% of interactions among MotoGP riders and teams. The platform's ability to reflect real-time opinions was evident in the high engagement rates during the 2022 season, making it a reliable source for this analysis. Given the ethical challenges in data collection, the approach was guided by a data ethics framework emphasizing ethical responsibilities (Floridi and Taddeo, 2016). (Alam *et al.*, 2022) demonstrated that Instagram effectively captures public sentiment on issues like vaccine hesitancy, highlighting its capability for nuanced public reactions in contexts like MotoGP decisions. The Infringement/Decisions Dataset was recorded throughout the 2022 season, ensuring accuracy and integrity. Although other platforms like Twitter and TikTok were considered, Instagram's higher engagement led to its selection. Media sentiment was supplemented with insights from 87 blog articles due to insufficient Instagram content.

However, potential biases exist. Platform-specific bias arises from focusing on Instagram, possibly missing broader audience views. Demographic bias reflects the younger, tech-savvy Instagram user base, potentially underrepresenting older fans. Content type bias highlights how engaging posts might skew sentiment analysis. Acknowledging these biases helps contextualize findings and maintain transparency in the research.

4.3 Data Preparation

The data preparation process was crucial to ensuring the accuracy and reliability of this project's analysis.

Steps Taken to Prepare the Data for Analysis:

The process began with verifying and organizing the Stewards panel data, focusing on the 2022 MotoGP, Moto2, and Moto3 classes. The dataset was cleaned to retain only relevant data and refined to emphasize infraction and sanction classifications. Sentiment columns for Public, Media, Team, and Riders were integrated, creating a comprehensive master dataset for analysis.

For the Sentiment Datasets, significant data cleaning and transformation were performed. Public sentiment data from Instagram was cross-verified against a consolidated accounts list to exclude comments from official accounts, ensuring only genuine public opinions were analyzed. A Python script facilitated this process, culminating in a clean dataset ready for sentiment analysis. Also a layered approach was exercised for accurate sentiment classification of public data on the assumption the public comments on the post endorse the sentiment of the post and a complex rubric spanning all possible combinations of the sentiment of post and that of comment was fed to the python for nuanced sentiment classification. However, public comments with relevant keywords such as penalty, outrageous, etc. we filtered and processed for sentiment for accuracy.

Data Transformation or Feature Engineering Performed:

The dataset was enhanced by refining infraction classifications, categorizing penalties and standardizing terminology. A unique Infraction_ID was created for each decision, linking sentiments from all stakeholders to specific infractions. Sentiment data was organized into distinct datasets—Riders, Team, Media, and Public Sentiment—with key variables like Infraction_ID, Stakeholder Type, and Sentiment.

Data Flow, Integration and Feature Engineering

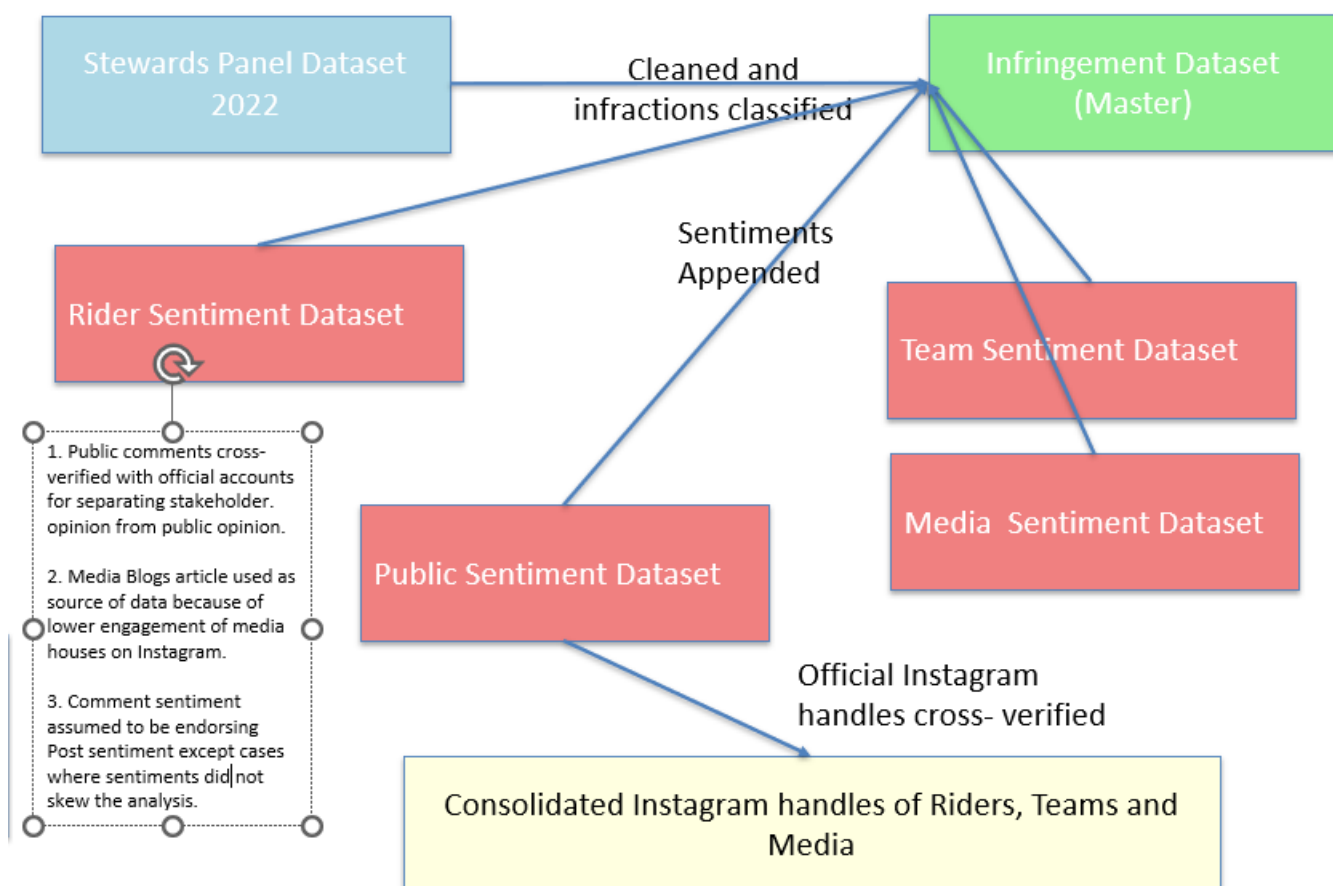


Figure 5: Data Integration and Management

Handling Missing or Inconsistent Data:

Missing data issues were resolved through communication with collaborators, addressing gaps like missing Race Class details. Blog articles were analyzed to supplement Media Sentiment, compensating for limitations in Instagram data extraction.

Challenges Encountered and How They Were Addressed:

Challenges included- data collection methods and GDPR compliance. Python-based web scraping faced limitations, so third-party tools like ExportComments.com were used, with ethical considerations carefully managed. A cross-verification process filtered official comments from public sentiment data, and media content was refined to distinguish original sentiment from stakeholder quotes. These measures ensured robust, accurate datasets aligned with the project's objectives while adhering to ethical standards.

4.4 Data Assumptions and Limitations

During the data preparation process, several critical assumptions were made. It was assumed that the Infringement Dataset provided by the clients was both complete and accurate, as the analysis relied heavily on this dataset to evaluate the consistency of the MotoGP Stewards Panel's decisions. Additionally, the decision to focus exclusively on the 2022 season across MotoGP, Moto2, and Moto3 was based on the assumption that this would provide a comprehensive and relevant view of the disciplinary actions taken by the Stewards Panel. Instagram was selected as the primary platform for sentiment analysis due to its high engagement levels among MotoGP fans and stakeholders, under the assumption that it adequately represents the broader MotoGP audience. In the sentiment data, instances labeled as "No Response" were assumed to indicate neutrality rather than a lack of sentiment, reflecting a neutral stance rather than an absence of opinion.

However, these assumptions also brought inherent limitations. Focusing solely on Instagram may have introduced platform-specific biases, potentially skewing the sentiment analysis toward the demographics of Instagram users. Additionally, machine translation of non-English posts may have led to misinterpretations due to the nuances of language. To mitigate these issues, a manual review of key data and translations was conducted, and supplementary data sources, such as blog articles, were incorporated to provide a broader and more balanced perspective. These steps were essential in maintaining the integrity of the analysis while acknowledging and addressing the limitations of the data.

5 The Analysis

5.1 Analysis Progression

The analysis aimed to assess the consistency, sentiment, and transparency of the MotoGP Stewards Panel's decisions during the 2022 season through a structured, multi-phase approach. Descriptive analysis was the starting point, identifying key trends and statistics, such as common infringement types and penalty distributions across classes. This phase revealed crucial patterns, including frequent track limit violations and time penalties, setting the stage for deeper investigation.

Building on these findings, diagnostic analysis explored the causes behind the patterns. Cross-tabulation and correlation analysis examined relationships between infringement types and sanctions, revealing inconsistencies in penalties for similar infractions across different contexts. For instance, differing penalties for track limit violations at various circuits suggested potential biases in decision-making. Sentiment analysis enriched the diagnostic phase by uncovering stakeholder reactions, identifying areas for improvement in stewarding fairness and consistency.

The analytical sequence progressed logically from broad trend identification to specific problem areas. Descriptive analysis informed the diagnostic phase, allowing targeted analysis of key issues. Sentiment analysis followed to provide context for stakeholder perceptions, directly aligning with evaluating steward decisions.

While predictive and prescriptive modeling were beyond the project's scope, the diagnostic findings led to practical recommendations. Biases and data limitations were acknowledged throughout, ensuring the reliability and robustness of the conclusions within the given constraints.

5.2 Descriptive Analytics

The data exploration began with refining the MotoGP Stewards Panel Decisions Dataset from the 2022 season. This involved cleaning and organizing variables like infringement types, sanctions, and metadata, resolving inconsistencies through cross-referencing and consultations. Sentiment analysis variables—Public, Rider, Team, and Media Sentiment—were then integrated, creating a comprehensive master dataset. This dataset, including the refined Infringement Dataset, was systematically explored for consistency and accuracy, forming a solid foundation for the descriptive analysis of trends and patterns in stewards' rulings. A data dictionary was also developed to guide further analysis and planning.

Key Metrics and Statistics

The 2022 MotoGP season data revealed significant insights into rider behaviour and rule enforcement. Among these, "Irresponsible Riding" was the most frequent infringement, accounting for over two-thirds of all recorded infractions. This indicates persistent challenges in ensuring safe riding practices in high-stakes racing. The most commonly imposed sanctions were "Long Lap Penalties" (19 instances) and "Grid Penalties" (10 instances), reflecting a preference for penalties that directly influence race outcomes without being overly punitive.

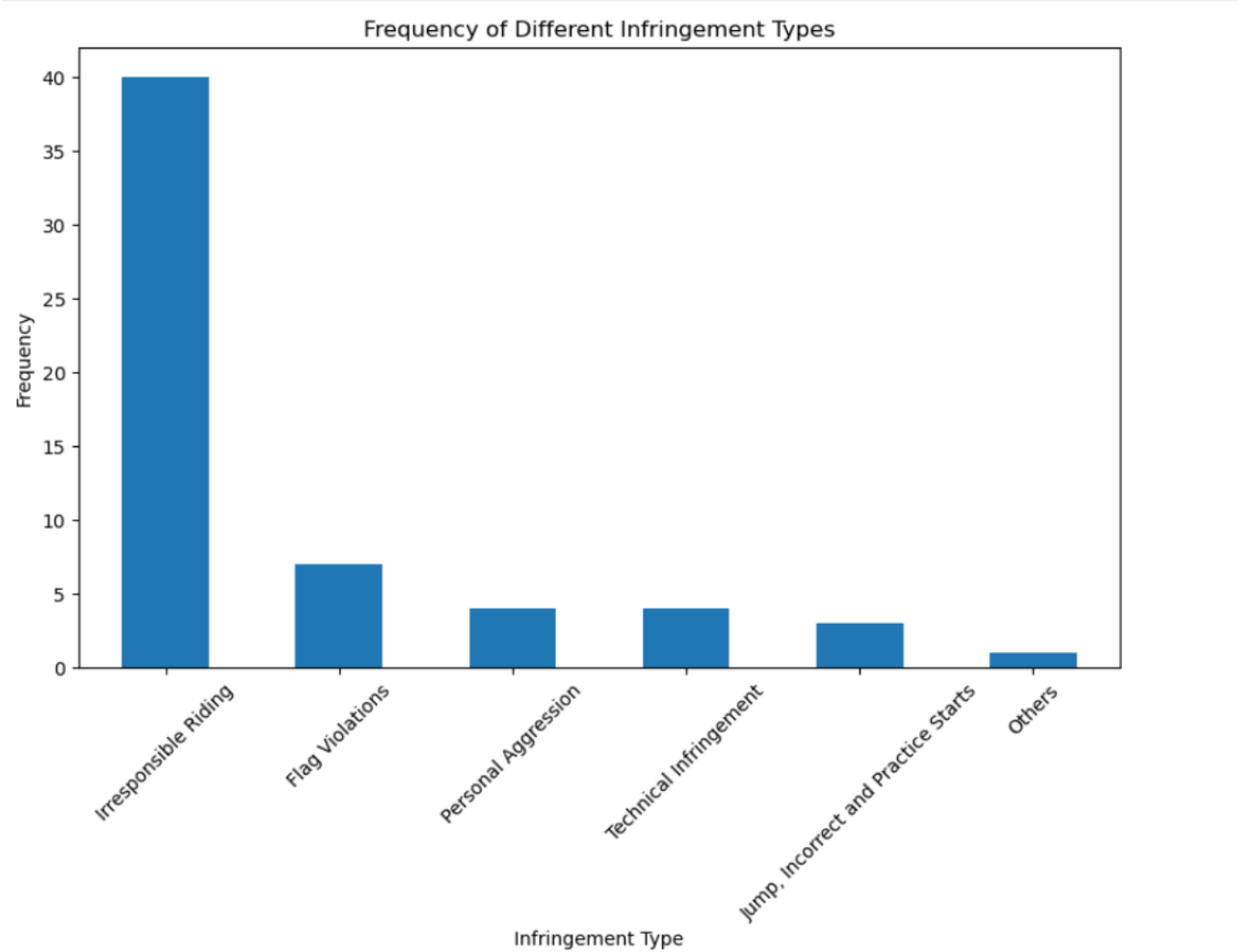


Figure 6: Infringement Frequency

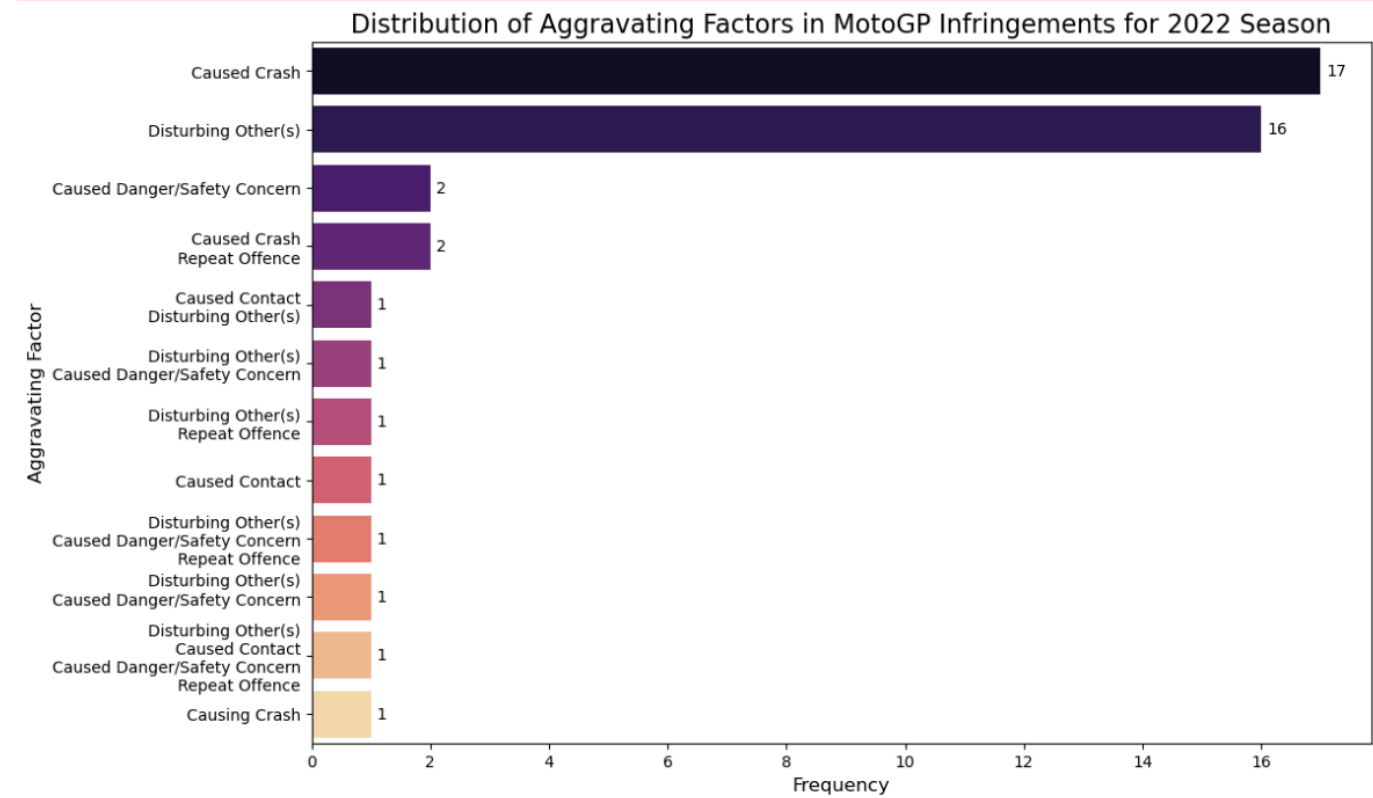


Figure 7: Frequency of Sanction Classifications (Including Multiple Sanctions per Decision)

Sub-classifications of infringements provide further granularity. "Riding slowly-On the racing line" was the most common behaviour leading to penalties, occurring 19 times. This behaviour poses significant risks as it disrupts the race flow and increases collision likelihood. "Overly aggressive" behaviour followed, with 10 instances. Other sub-classifications such as "Riding slowly-Departing Pit Lane," "Rejoining track limit violations unsafely," "Riding Irresponsibly," and "Overly ambitious" were less common, indicating these behaviours are either less frequent or less often penalized.

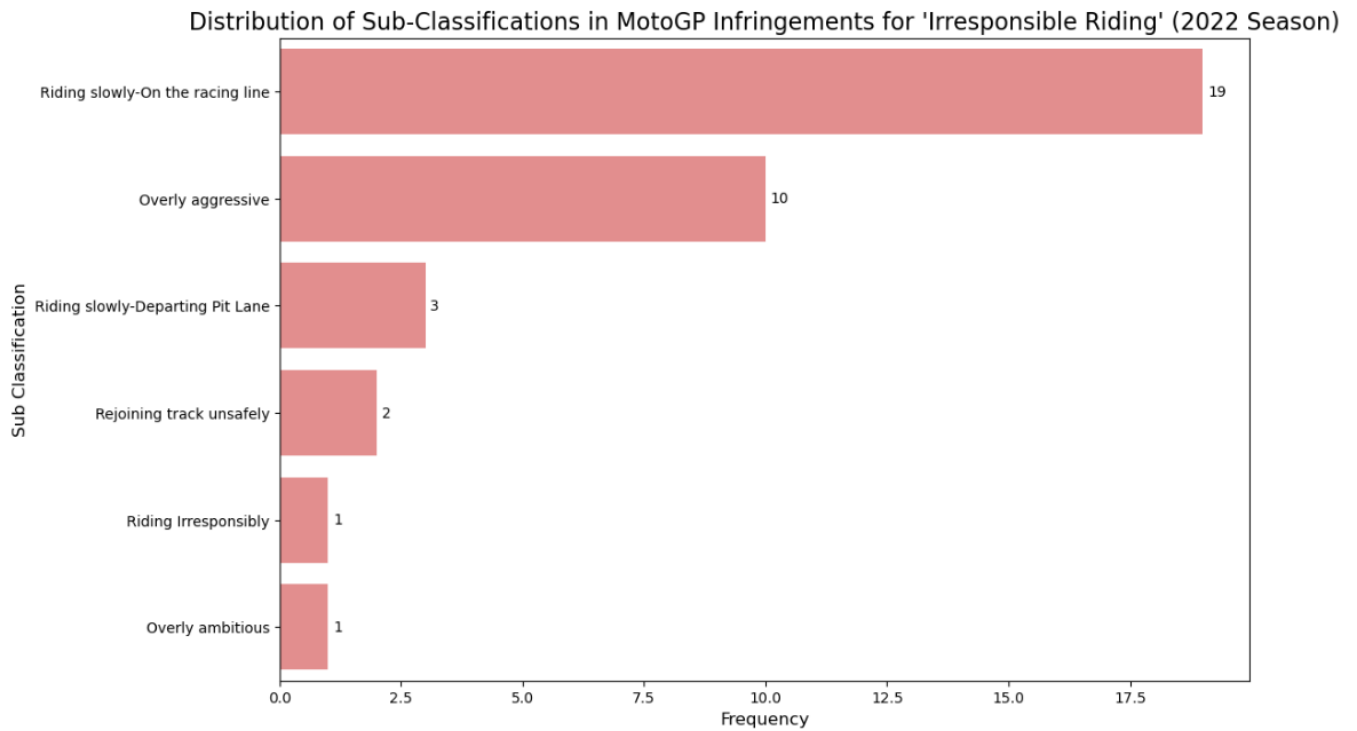


Figure 8: Irresponsible Riding Sub classification

Aggravating Factors

Aggravating factors, such as "Caused Crash" and "Disturbing Other(s)," were critical in determining the severity of sanctions, appearing 17 and 16 times, respectively. These factors often justify harsher penalties due to their direct impact on rider safety and race integrity. Lesser, but still important, factors included "Caused Danger/Safety Concern" and "Repeat Offence." Identifying these factors is crucial for understanding the broader context of rule violations and ensuring consistent application of sanctions.

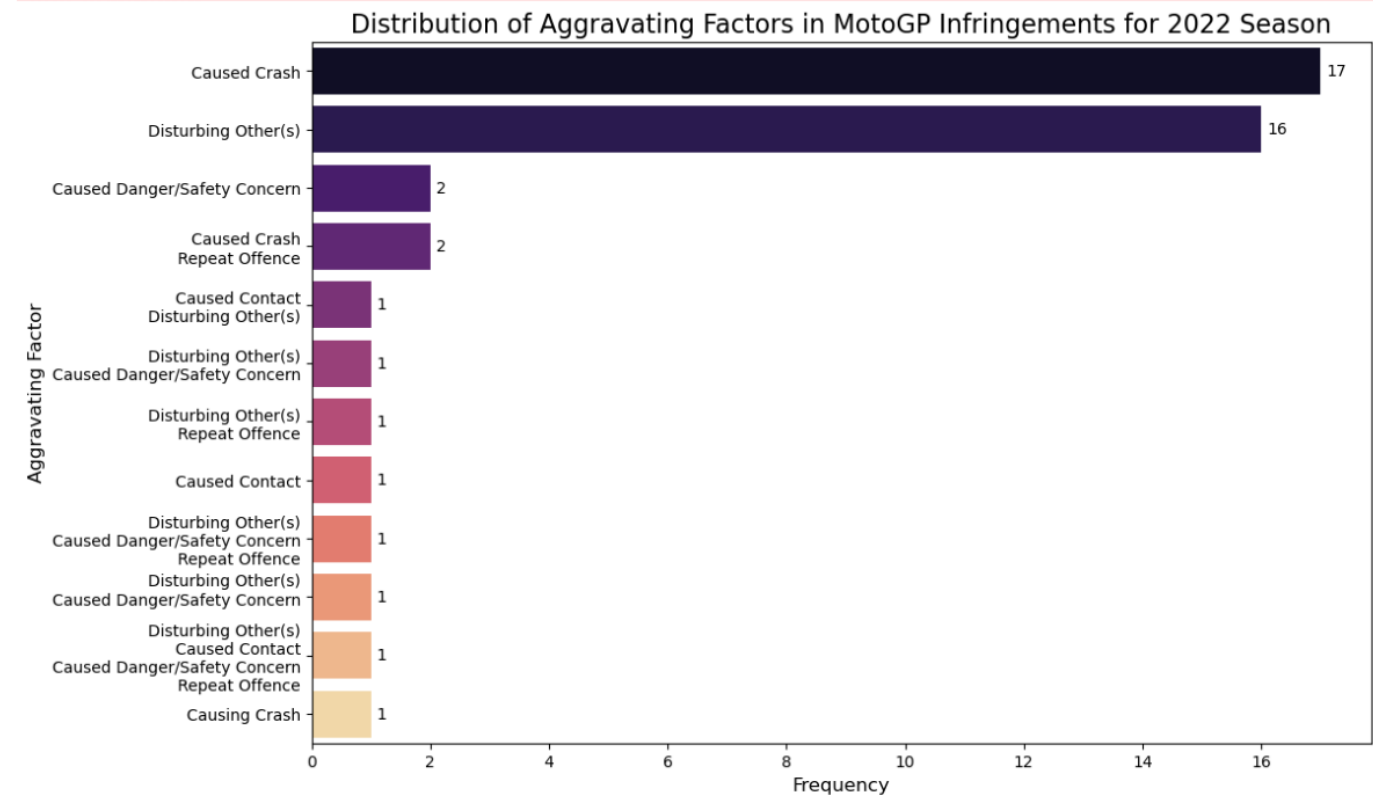


Figure 9: "Distribution of Aggravating Factors by Infringements (Including Multiple Factors per Infraction)"

Class-Specific Trends

In Moto3, "Irresponsible Riding" was particularly prevalent, accounting for 75% of all infringements, suggesting younger or less experienced riders may be more prone to risky behaviour. In contrast, Flag Violations were most common in Moto2, where strategic rule infractions may be more prevalent. Meanwhile, Technical Infringements were more frequent in Moto3, possibly reflecting the challenges faced by newer teams and riders in managing technical aspects. These trends underscore the need for tailored regulatory approaches for each class, such as stricter enforcement in Moto3 and enhanced technical support.

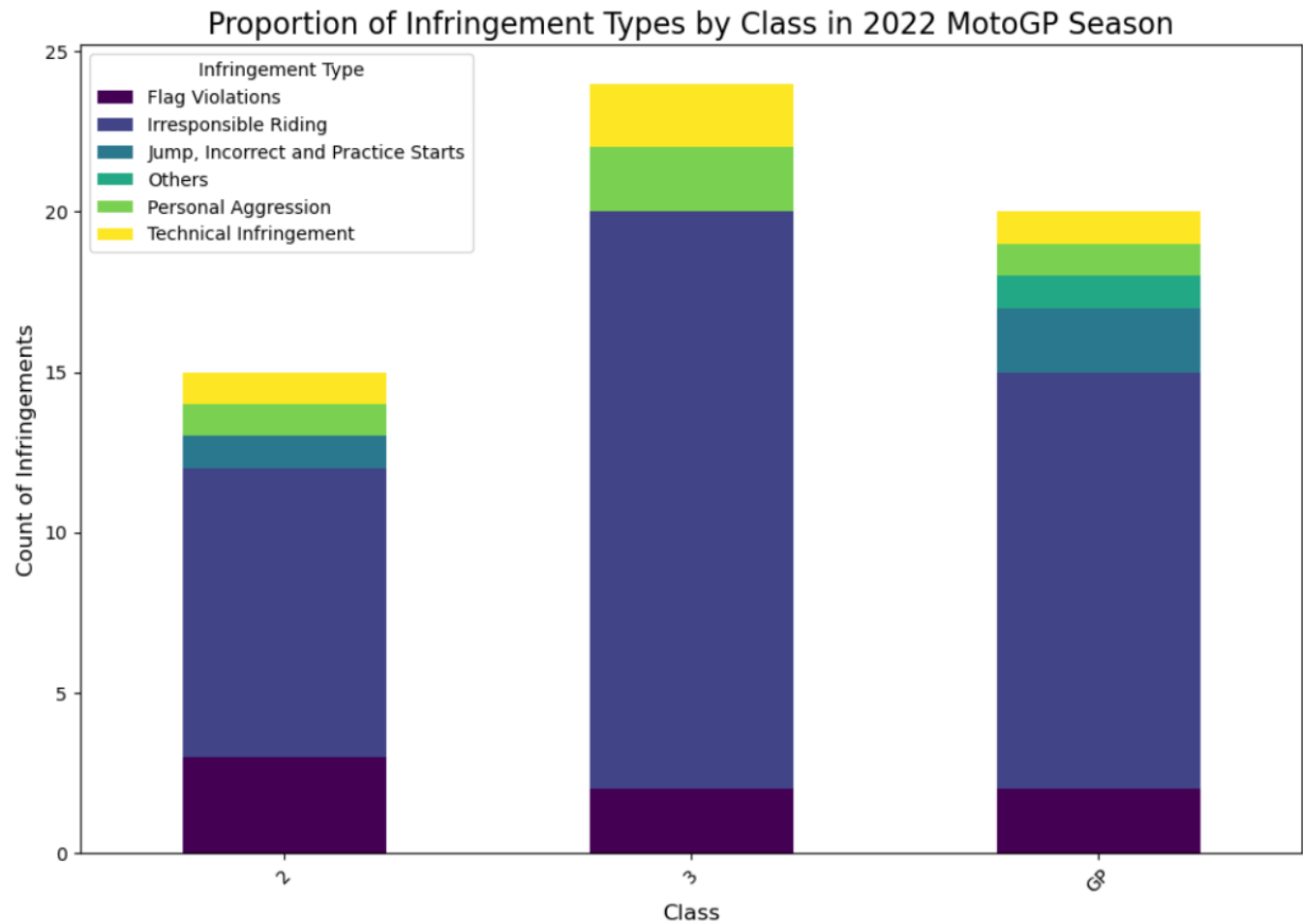


Figure 10: Proportion of Infringement by Class

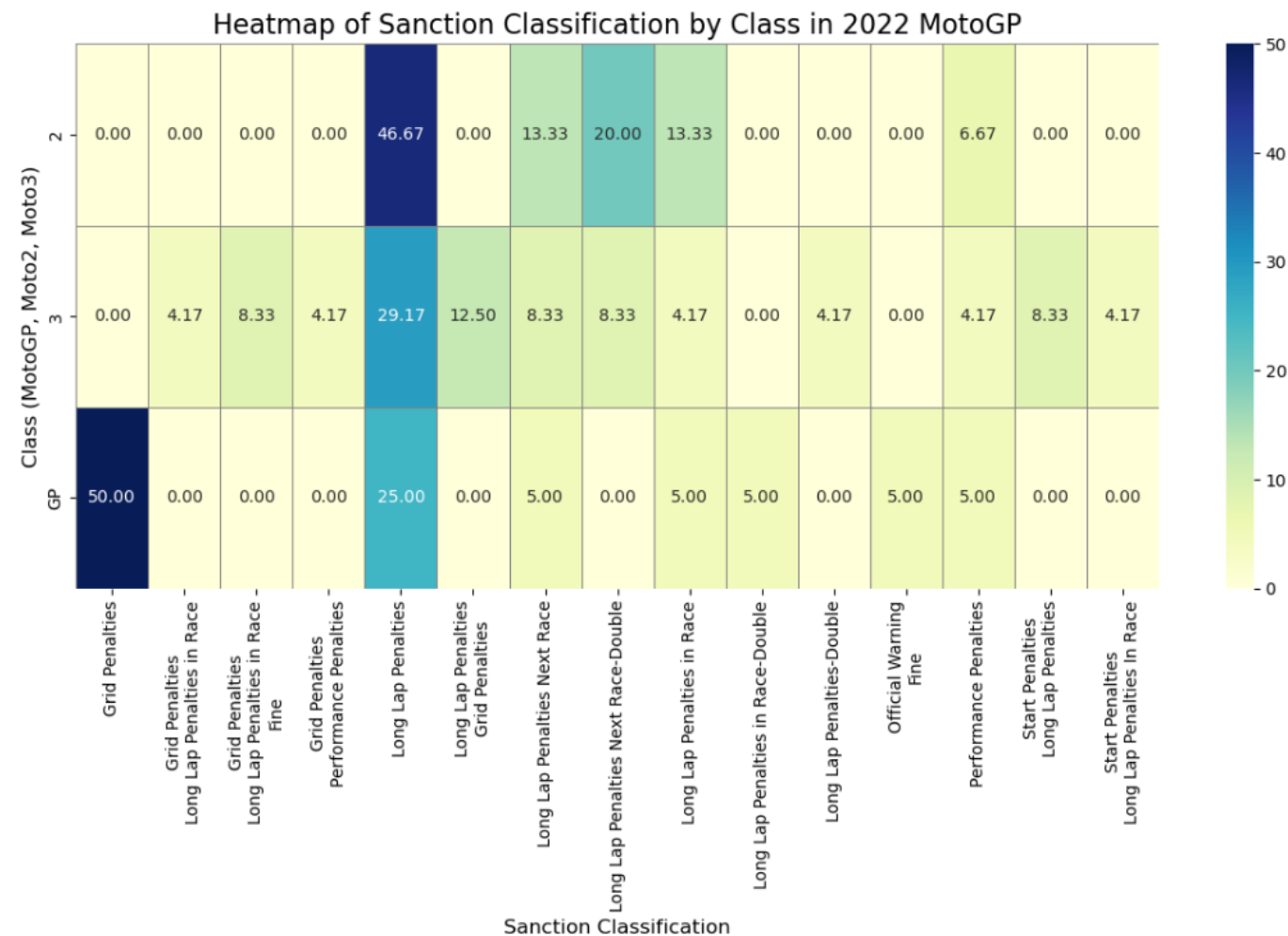


Figure 11: Sanction Classification by Class Heatmap'

Circuit-Specific and Temporal Trends

The "Assen" circuit recorded the highest number of unique sanctions (8), followed by "Valencia" (6) and "Jerez" (5). These findings suggest that certain circuits may pose greater challenges to riders, increasing the likelihood of infractions. Understanding circuit-specific trends is essential for race organizers, informing decisions on track modifications or race management strategies to reduce violations. In terms of race stages, the "Race" stage saw the most unique sanctions (13), which is expected due to the high-pressure environment. FP3 (12 sanctions) and qualifying sessions Q2 (10) and Q1 (7) also recorded high sanction counts, highlighting the need for close monitoring during these critical stages.

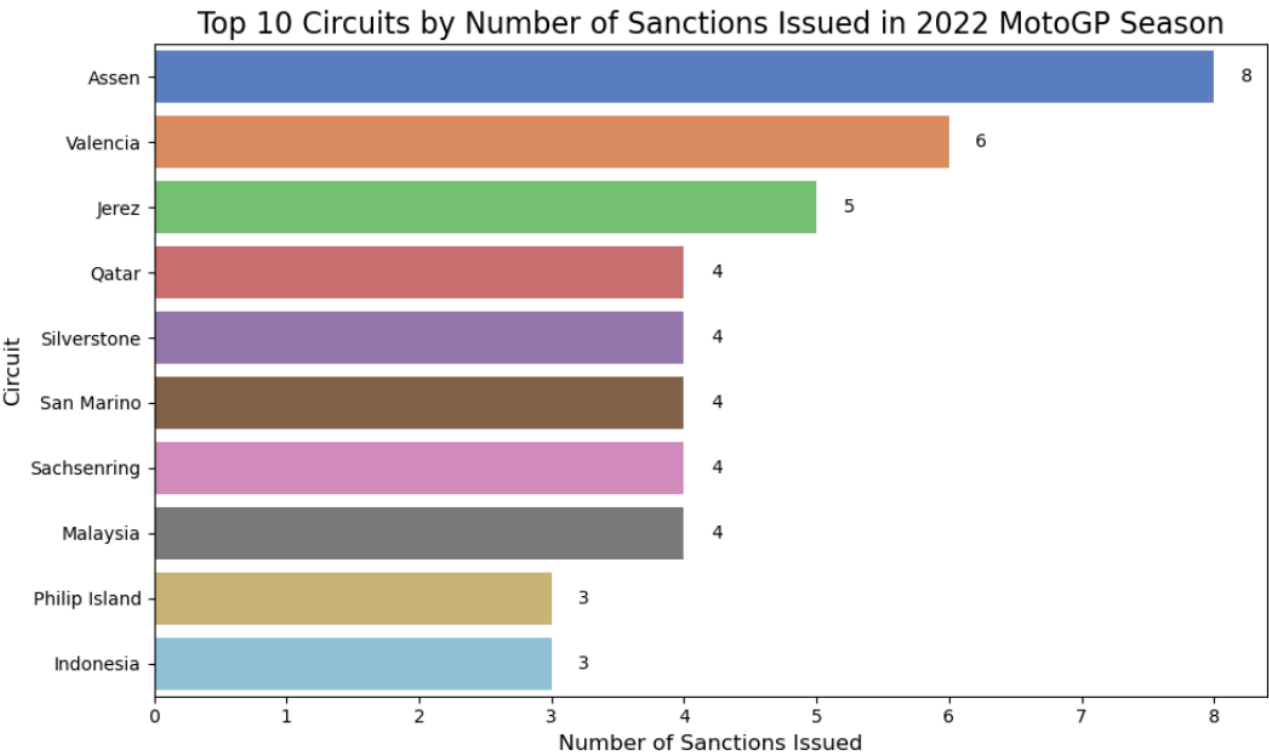


Figure 12: Circuits with Most Sanctions Issued 2022

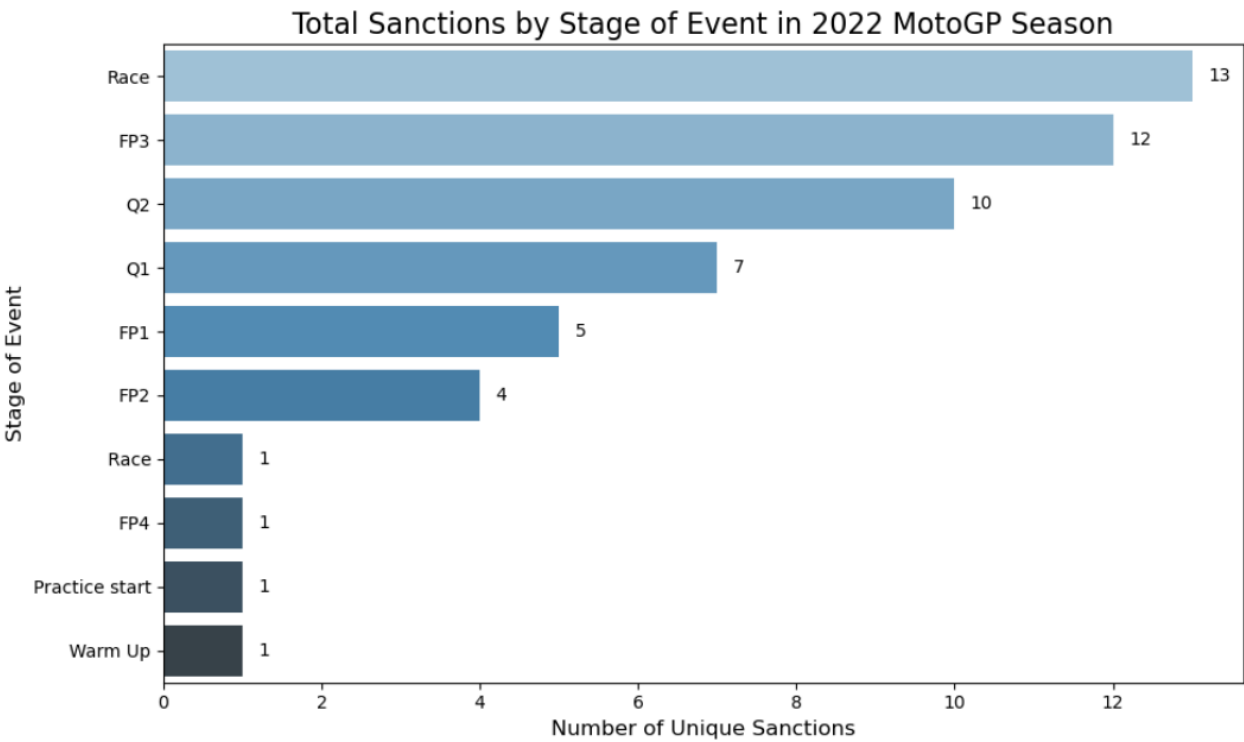


Figure 13: Stages with Most Sanctions Issued

Judicial Challenges

"Irresponsible Riding" was frequently associated with hearings, particularly "Riding Slowly on the Racing Line," which led to 9 hearings out of 19 cases. This suggests that such decisions are often contested, possibly due to their subjective nature or significant impact on race outcomes. Technical Infringements were the most contested, with 3 out of 4 cases leading to appeals, reflecting the complexity of technical regulations and the potential for interpretational differences.

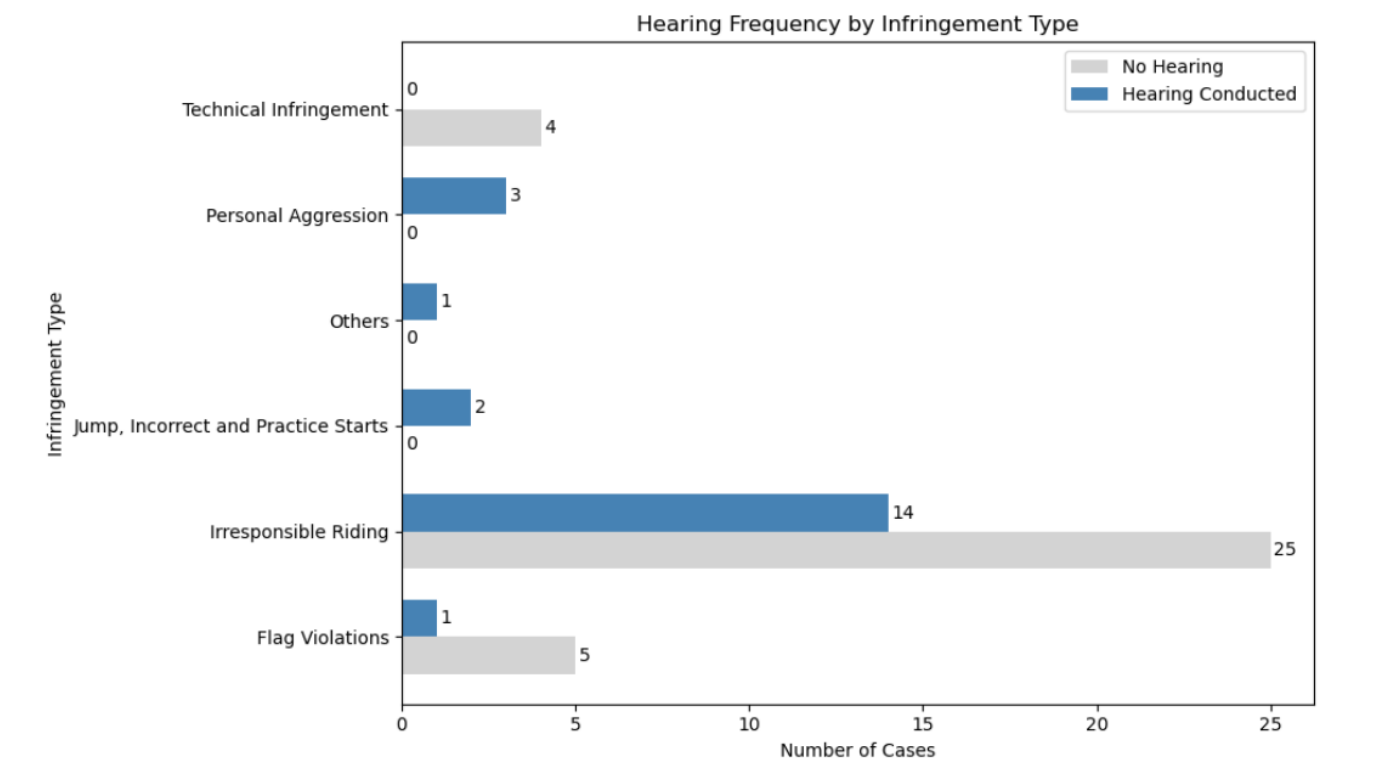


Figure 14: Hearing Frequency by Infringement Type

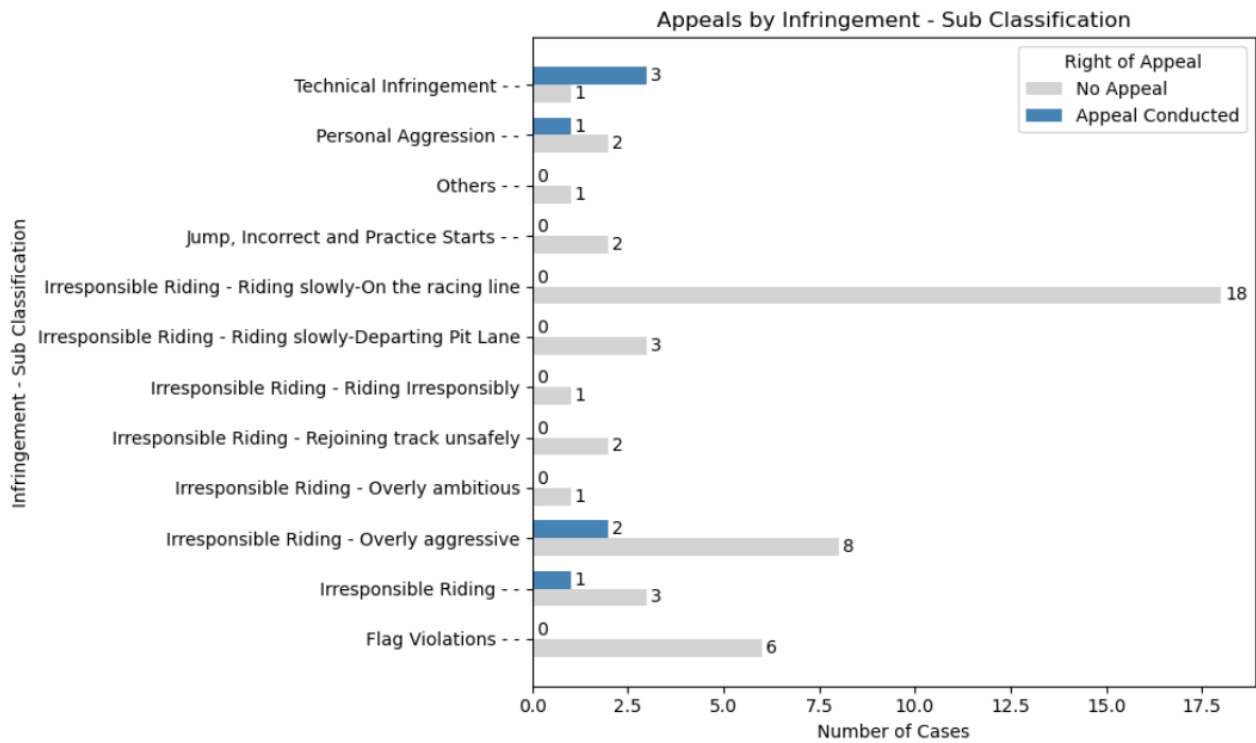


Figure 15: Appeal by Infringement/Sub Classification

Steward Panel Perceptions

The panel group "Somolinos, Spencer, De Fabritis" issued the highest number of sanctions (18) and received predominantly neutral public sentiment (16 out of 18). In contrast, the "Somolinos, Spencer, Bohnhorst" group had a higher proportion of negative sentiments (5 out of 12), indicating their decisions may have been perceived as inconsistent or overly harsh. Understanding public sentiment towards steward panels is vital for assessing the perceived fairness of race regulation.

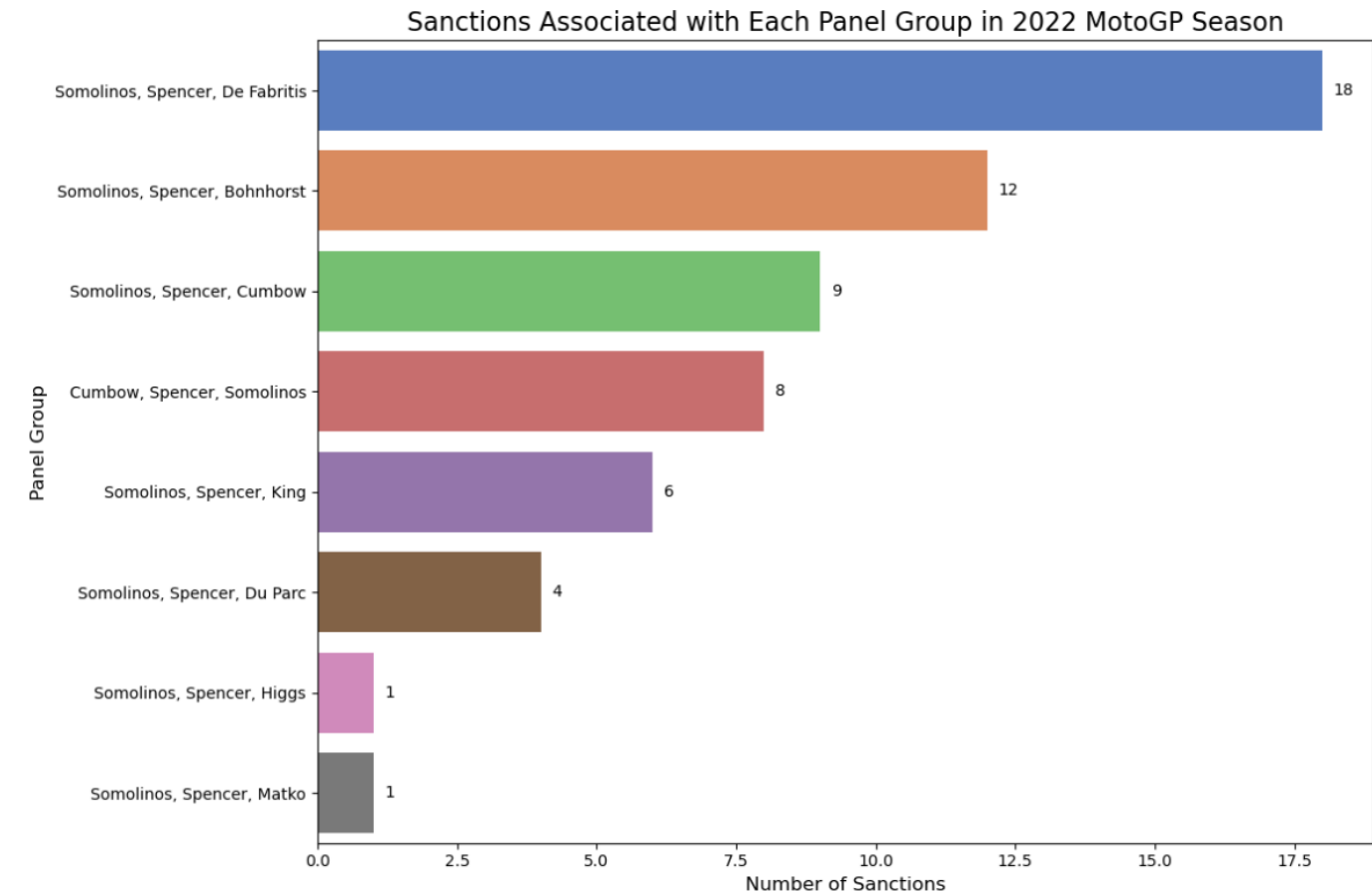


Figure 16: Sanctions Issued by Panel Group

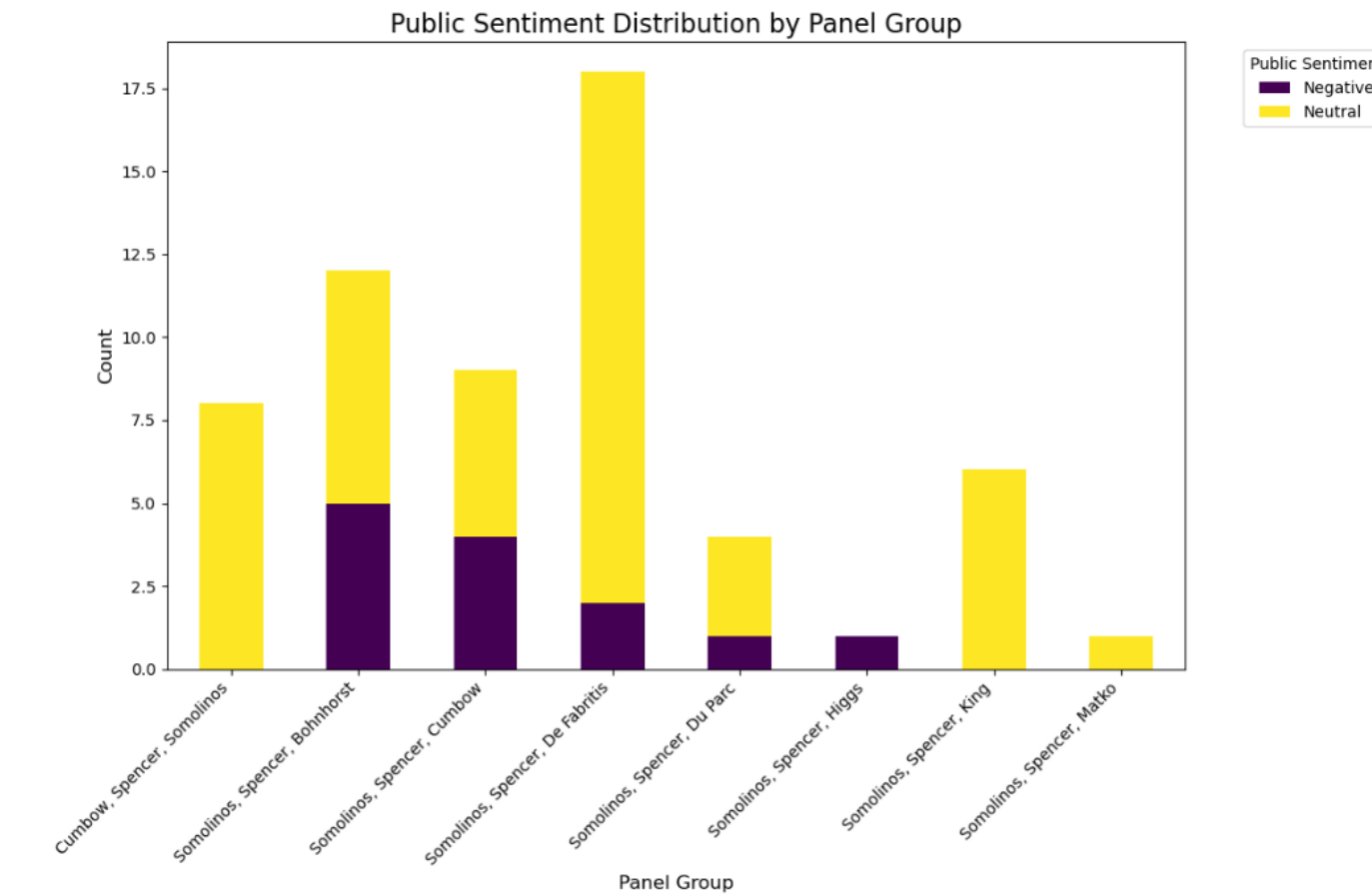


Figure 17: Public Sentiment by Panel Group

Stakeholder Sentiment Quantification

Sentiment analysis revealed that Leopard Racing and Sic58 Squadra Corse were the only teams associated with negative sentiments, each recording one negative sentiment. Mooney VR46 Racing Team had the most neutral sentiments (6), indicating frequent mentions without significant negative impact. Media Articles were the primary medium for expressing negative sentiments, with 7 riders, including Jack Miller (5 mentions) and Fabio Quartararo (4), prominently featured. The role of media in shaping public perception is evident, as riders frequently mentioned in negative contexts may face challenges in managing their public image.

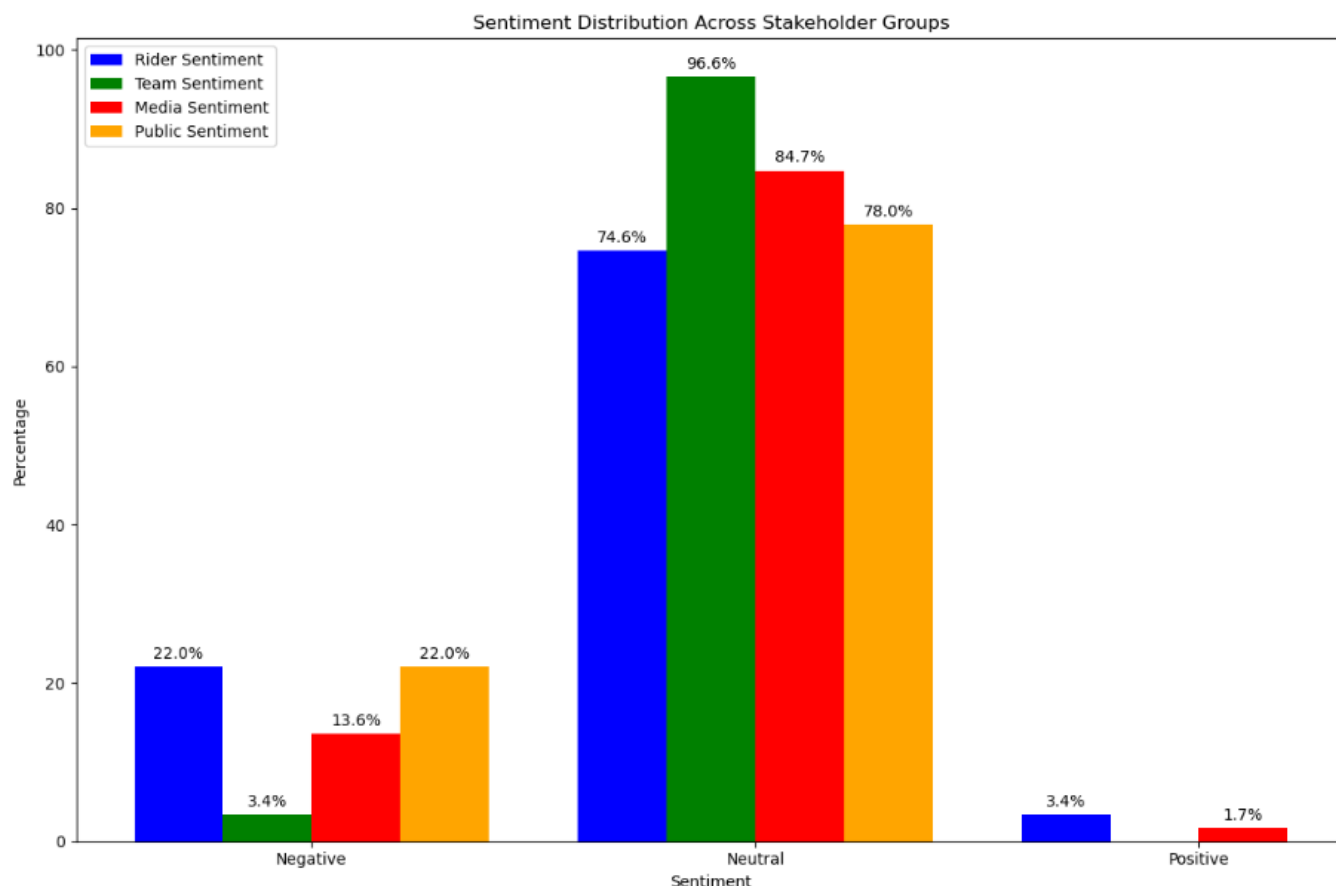


Figure 18: Sentiment of Stakeholders

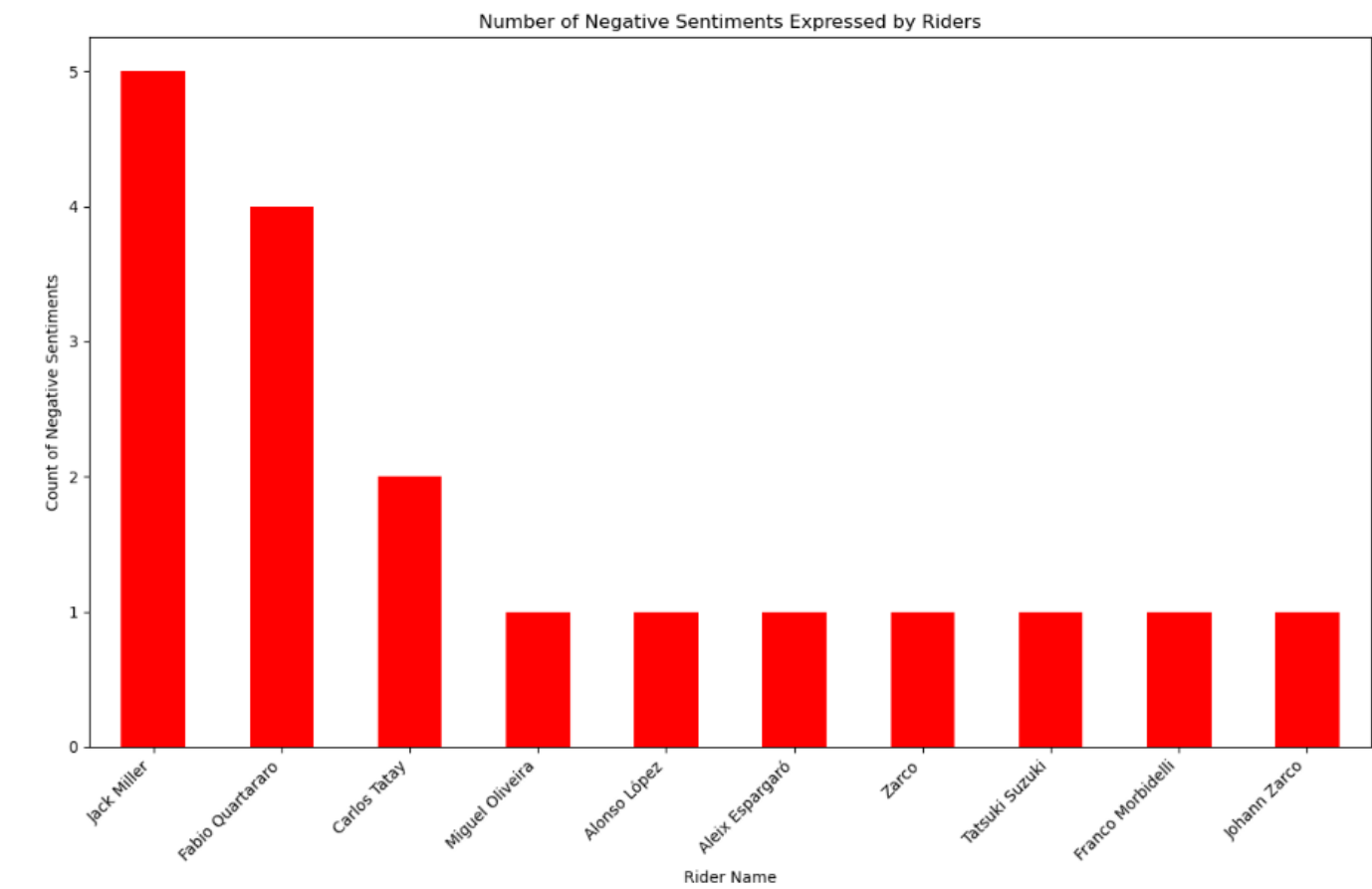


Figure 19: Negative Sentiments by Rider

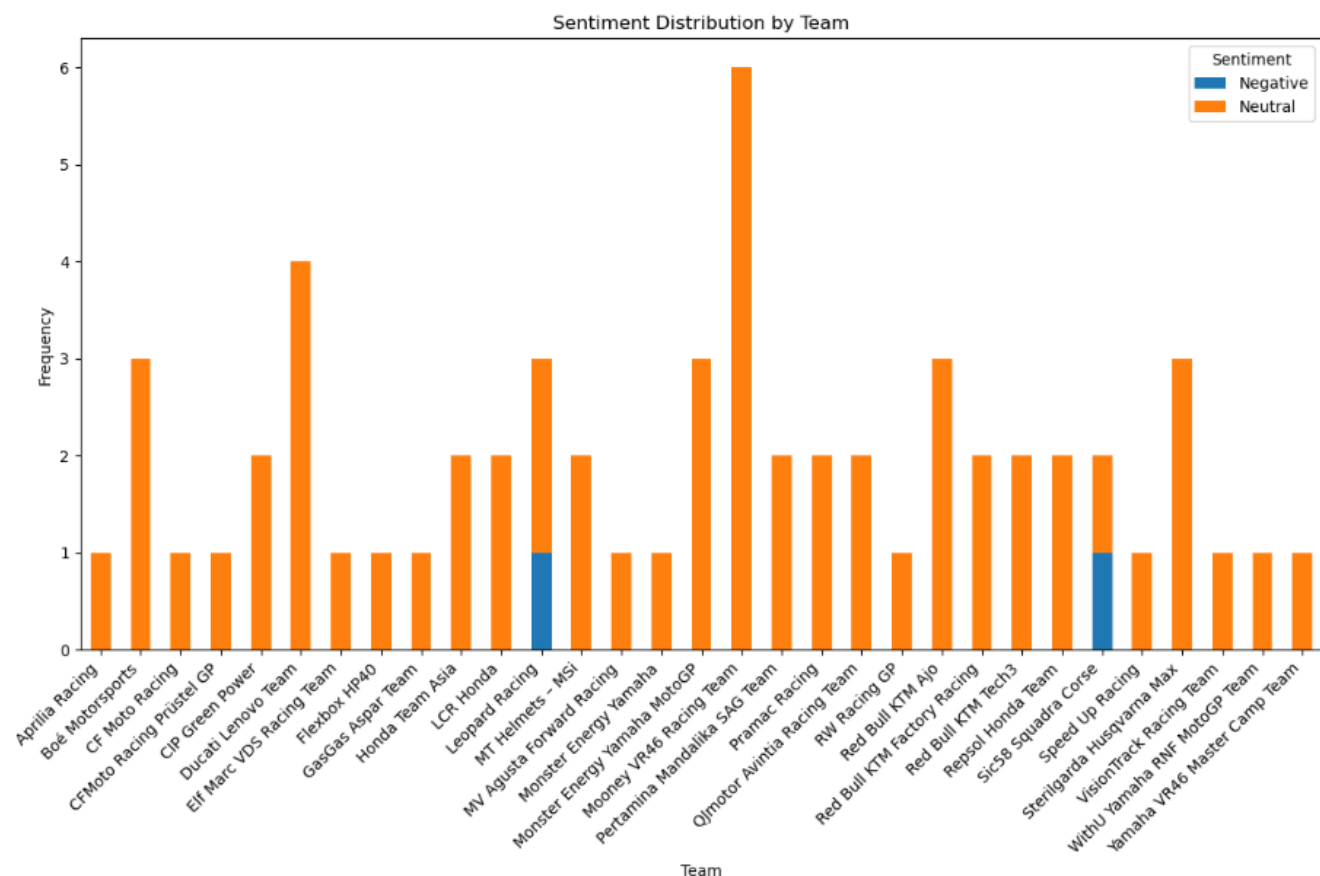


Figure 20: Sentiment Distribution by Team

Rider-Specific Trends

Morbidelli (#21) and Vietti (#13) had the most sanctions, with 4 each, followed by Miller (#43) and Sasaki (#71) with 3. These findings suggest that certain riders may be more prone to infractions, reflecting more aggressive riding styles or a tendency to push the boundaries of race regulations. Understanding these trends is crucial for race officials in making decisions about targeted interventions to prevent future infractions.

5.3 Diagnostic Analytics

Diagnostic analytics uncovers patterns in MotoGP Stewards Panel decisions during the 2022 season. This analysis examined contentious decisions, classified by infringement type, and evaluated consistency in penalties. It explored stakeholder sentiments—riders, teams, and media—and how factors like rider nationality and penalty type influenced perceptions. Descriptive analytics identified sanction distribution, while correlation analysis revealed patterns such as the frequent use of Long Lap Penalties for irresponsible riding. Root cause analysis delved into the reasons for varied penalties, especially for severe infractions like personal aggression. Sentiment analysis highlighted negative reactions to severe sanctions, and cross-tabulation showed fewer appeals after hearings. The analysis considered variables like race type, timing, and external factors, offering insights into how Stewards’ decisions were shaped and perceived.

Results of Diagnostic Analysis:

Analysis of contentious cases

Contentious cases, like 2022-CGP-004, saw strong negative reactions from both the public and riders due to severe penalties and aggravating factors like "Caused Danger." However, in cases like 2022-CGP-007, there was a divergence in sentiment, with riders accepting the penalty but the public viewing it as overly harsh. In less contentious cases, such as 2022-CGP-0018, neutral sentiments prevailed, indicating that clear communication and context play a key role in acceptance of severe penalties.

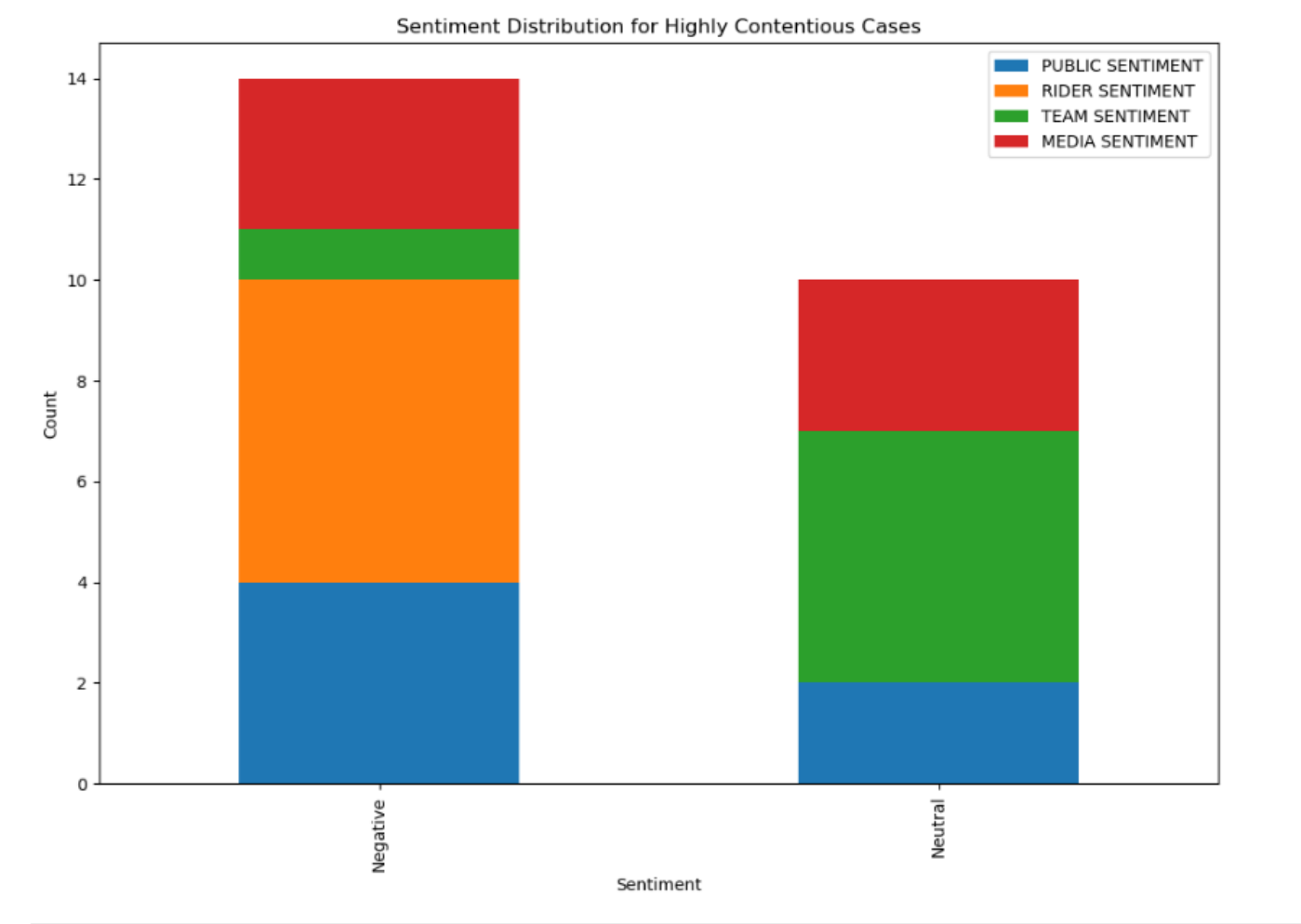


Figure 21: Sentiment Distribution by Highly Contentious Cases

Analysis of Sanction Distribution Across Infringement Types

Irresponsible riding was the most penalized offense, leading to 40 incidents, often resulting in Long Lap Penalties. For flag violations, 5 out of 7 cases also led to Long Lap Penalties, while grid penalties were common for jump starts, applied in 2 out of 3 cases. Personal aggression and technical infringements received compounded penalties and performance penalties, respectively. This pattern shows a clear link between specific infractions and the types of penalties imposed, highlighting how different violations are consistently met with distinct consequences in racing.

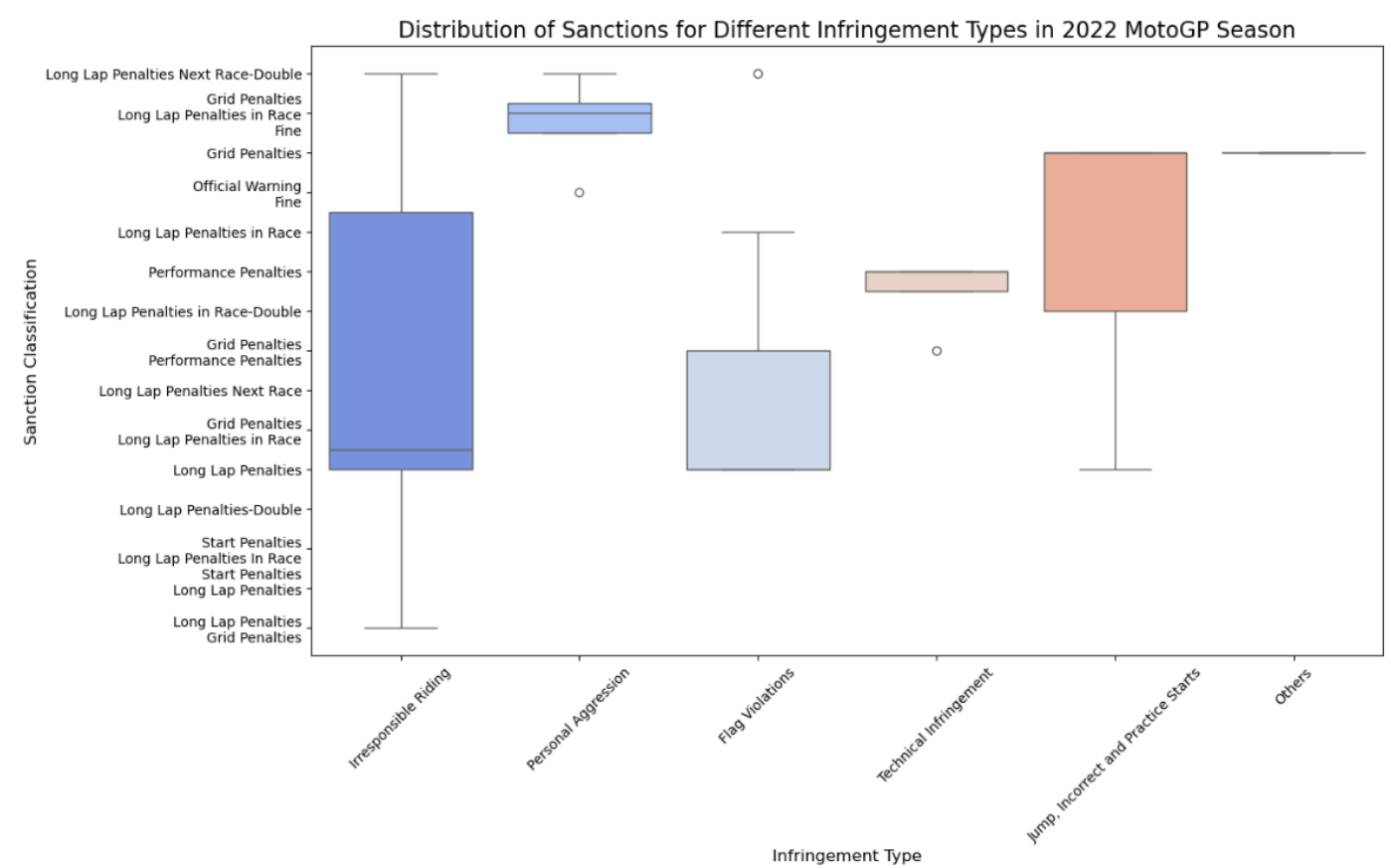


Figure 22: Sanction Distribution By Infringement Type

Correlation Analysis Between Hearings and Appeals

In 30 cases where no hearing was held, no appeal was made, indicating general acceptance of decisions without formal proceedings. However, in 4 cases without a hearing, appeals were filed, suggesting that the lack of a hearing may sometimes lead to dissatisfaction. When hearings were conducted, 18 cases did not lead to an appeal, reflecting the effectiveness of hearings in resolving disputes. However, 3 cases resulted in appeals despite a hearing, showing that formal proceedings can still leave some stakeholders dissatisfied, indicating room for improvement in the decision-making process.

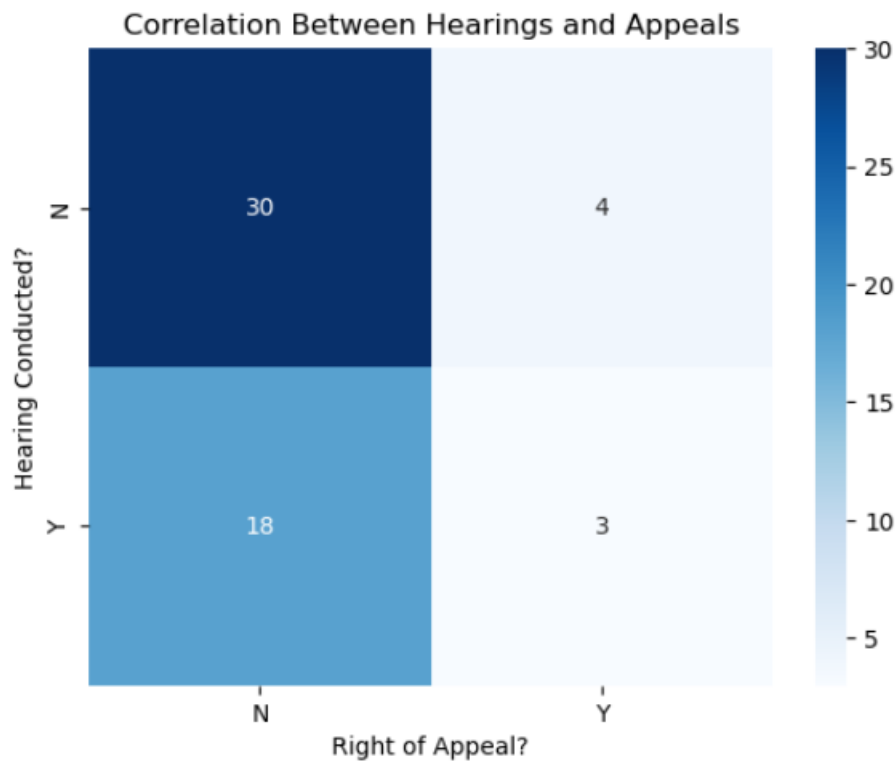


Figure 23: Correlation between Hearing and Appeal

Sentiment Analysis of Infractions

Neutral sentiment was dominant across stakeholders, with 96.61% neutrality among teams, 84.75% among media, and 74.58% among riders. This suggests a general acceptance or indifference towards the Stewards' decisions. However, negative sentiment was most prominent among riders (22.03%), highlighting some dissatisfaction with the rulings, while positive sentiment remained very low across all groups: 3.39% for riders, 1.69% for media, and none for teams. This distribution points to potential areas where the decision-making process may need refinement, particularly in addressing rider concerns.

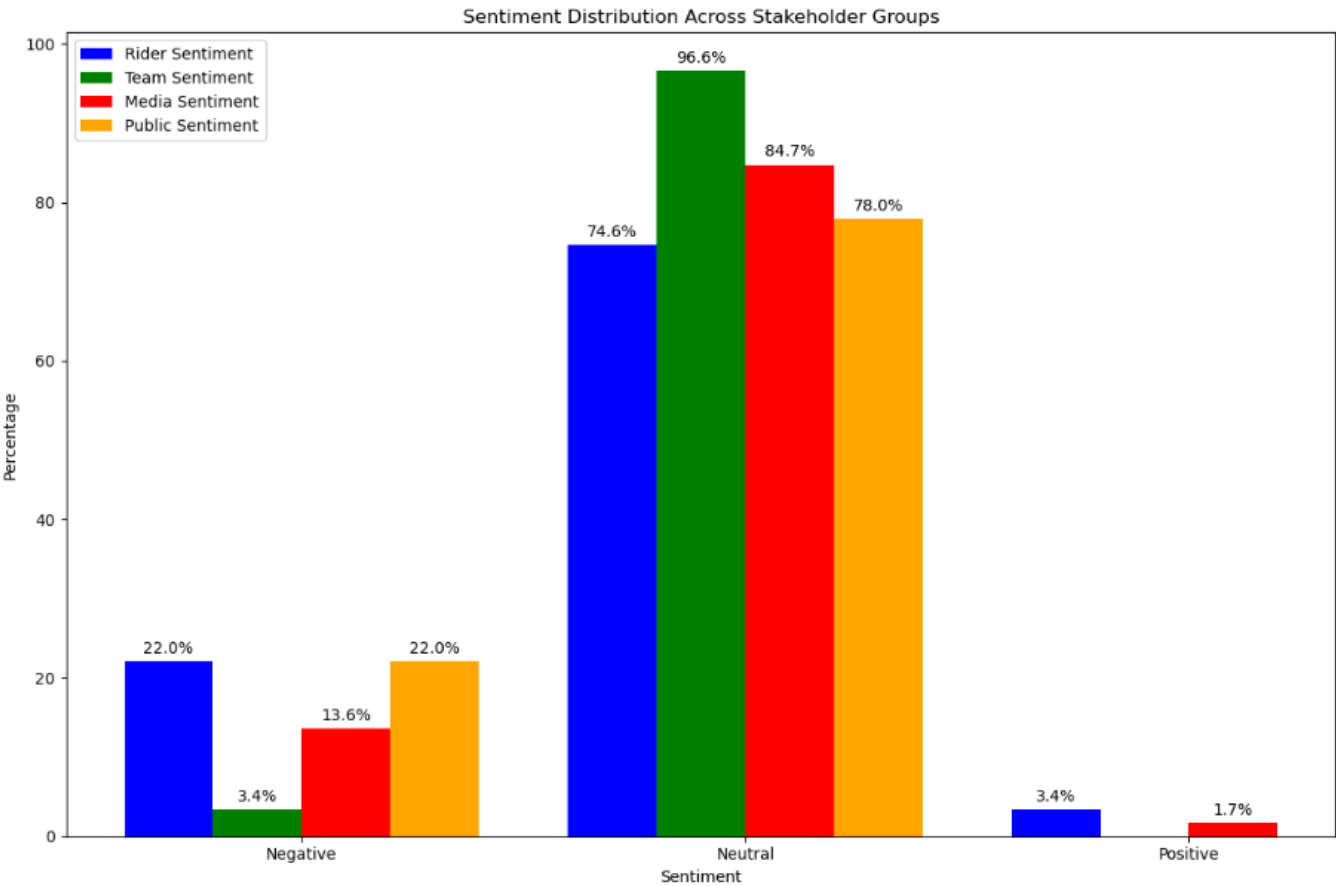


Figure 24: Sentiment Distribution Across Stakeholders

Correlation Analysis Between Sanction Severity and Stakeholder Sentiments

More severe sanctions led to a higher proportion of negative sentiment, rising from 25.81% for mild sanctions to 39.29% for severe ones. Despite this, neutral sentiment remained dominant for both mild (86.29%) and severe (80.36%) sanctions, indicating broad acceptance of penalties as a routine part of the sport. Positive sentiment was low but slightly higher for severe sanctions, increasing from 0.81% to 1.79%.

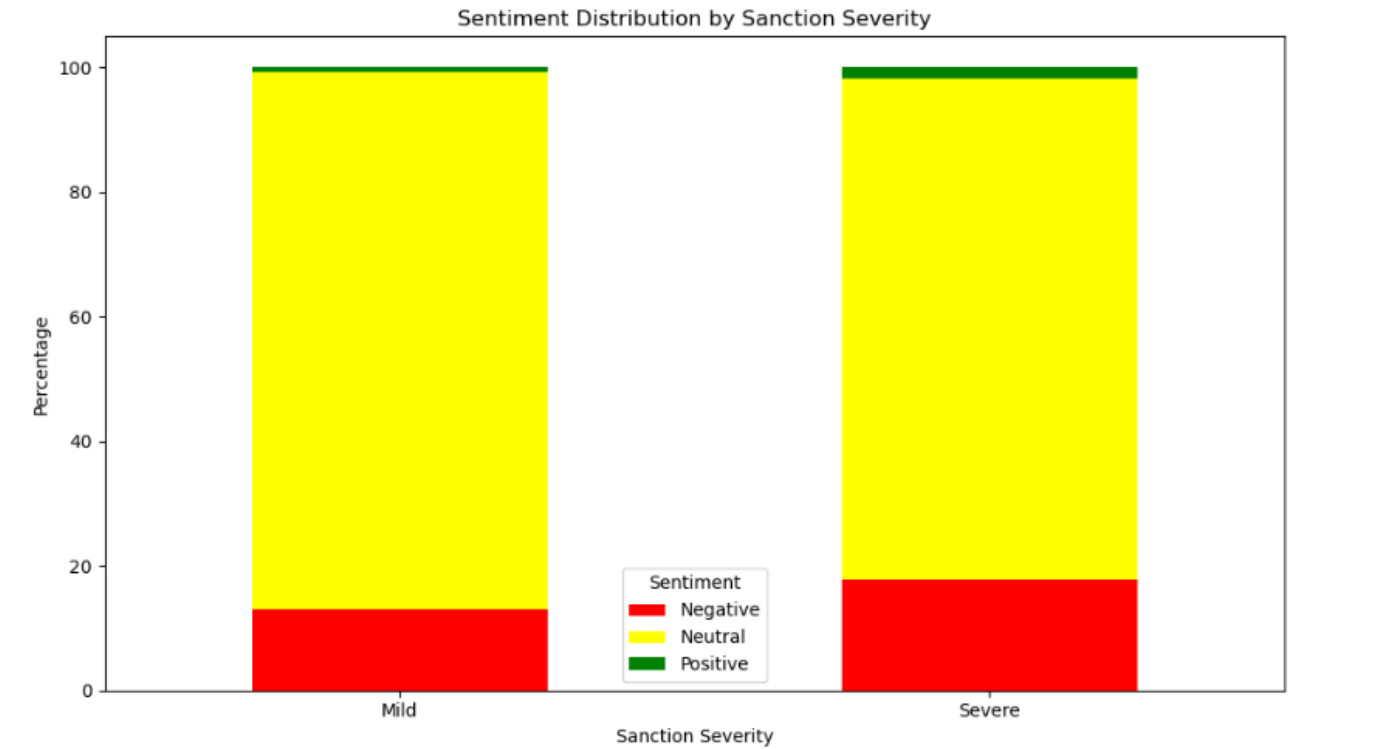


Figure 25: Overall Sanction Severity

Analysis of Divergence in Media Sentiments Regarding Infractions

Media sentiment varied significantly by source. Websites like "[www.the-race.com](#)" and "[www.paddock-gp.com](#)" consistently showed negative. There were also 17 neutral mentions recorded across various outlets, including "[www.gpone.com](#)".

The analysis identified statistically significant correlations, particularly between "Irresponsible Riding" and "Long Lap Penalties," indicating consistent application of rules. However, while strong correlations were found, establishing causality proved more complex. For example, the link between hearings and reduced appeals suggests a potential causal relationship, but further qualitative analysis would be needed to confirm this. The analysis also acknowledges limitations, such as the scope of the dataset being restricted to the 2022 season and potential biases, including unreported infractions or inconsistencies in documentation, which could impact the validity of the findings. These limitations should be considered when interpreting the results.

5.4 Predictive Analytics

While comprehensive predictive analytics was not implemented in this project, several advanced techniques could enhance MotoGP's decision-making process. With a more extensive historical dataset, predictive analytics could be invaluable. For example, machine learning models, such as Random Forests, could identify patterns in past decisions and predict likely outcomes in future steward rulings (Breiman, 2001). Regression analysis could also forecast the likelihood of specific infringements based on past trends, providing stewards with foresight into potential rule violations. Predictive models are increasingly applied in sports, with artificial data and simulation frameworks demonstrating their effectiveness in forecasting and decision-making (Garnica-Caparrós et al., 2022).

In practical terms, predictive analytics could help stewards anticipate the outcomes of decisions by analyzing historical data on how different penalties were received by riders, teams, and the public. For instance, if data shows that certain penalties tend to provoke negative reactions, stewards could adjust their approach, perhaps issuing warnings instead of penalties for minor infractions, to maintain fairness and avoid backlash. Predictive models could also aid in race planning by identifying tracks or situations where infringements are more likely, enabling proactive measures to ensure smoother race operations.

However, the effectiveness of predictive models depends on various factors. The dynamic nature of MotoGP, with constantly changing rules and conditions, may limit the long-term reliability of predictions. Additionally, challenges such as obtaining consistent historical data across seasons need to be addressed. Inaccuracies in data could lead to unreliable predictions, potentially undermining the effectiveness of these models. Moreover, over-reliance on predictions might overlook real-time nuances, which are crucial in a fast-paced sport like MotoGP.

Looking forward, integrating predictive analytics into MotoGP's decision-making process would require several key steps. A comprehensive dataset spanning multiple seasons is essential, focusing on factors influencing steward decisions and stakeholder reactions. By refining and updating these models, stakeholders could better anticipate decision outcomes and their potential impacts, ultimately contributing to more consistent and fair decision-making across the sport.

5.5 Prescriptive Analytics

While prescriptive analytics was not conducted in this project, it holds significant potential for improving decision-making in MotoGP. Data-driven approaches could help stewards make more balanced decisions by recommending penalties that are fair and minimize negative reactions. This would ensure consistent treatment of similar cases, reducing perceptions of bias or unfairness.

In practical terms, prescriptive analytics could assist stewards by evaluating penalty options and predicting their impacts. Analyzing past cases could provide recommendations on appropriate penalties under various

circumstances, aiding more informed decisions. Simulation techniques could also test different decision-making strategies, considering factors like race conditions or rider behavior, to identify the most effective approaches. These tools could offer data-backed recommendations for handling complex cases, potentially reducing contentious decisions and improving overall fairness. However, successful implementation would depend on the accuracy of predictive models and overcoming challenges like organizational buy-in and integrating new decision-making frameworks.

Looking ahead, implementing prescriptive analytics would require expanding the project's scope to include predictive models. This would involve developing a comprehensive dataset and setting clear objectives for optimizing decision-making processes. Engaging stakeholders through workshops or pilot programs would be crucial to ensure that recommendations are practical and aligned with MotoGP's goals. By incorporating prescriptive analytics, MotoGP could enhance the fairness and consistency of its decision-making processes.

The analysis of the 2022 MotoGP season revealed key patterns in rule enforcement and stakeholder reactions, highlighting areas for improvement in steward decisions. Although predictive and prescriptive analytics were not fully implemented, their potential to enhance decision-making was acknowledged. Integrating advanced models with continuous data collection and stakeholder engagement could further improve fairness and consistency, offering practical steps to enhance race regulation and maintain the sport's integrity.

5.6 Analytical Adjustments and Assumptions

Throughout the analysis, critical adjustments were made to enhance accuracy and relevance, driven by the need for deeper insights into MotoGP Stewards' decisions. Initially, broad data categorizations were refined to capture nuances within infractions, enabling a more detailed understanding of penalty patterns. Recognizing the significance of stakeholder perceptions, sentiment analysis was integrated, revealing the multifaceted impact of sanctions on riders, teams, media, and the public. This led to the incorporation of predictive and prescriptive analytics, shifting the focus from merely analysing past data to anticipating future reactions and guiding strategic decision-making. Key assumptions, such as uniform stakeholder perceptions and consistent interpretations of sanctions, simplified the analysis but also introduced limitations, potentially overlooking nuanced reactions and discrepancies. The assumption that past patterns predict future outcomes underscored the predictive models, though unique future scenarios might challenge their accuracy. These adjustments, while necessary for navigating complex data, also highlight the importance of continuous refinement. Despite inherent limitations, the analysis successfully offers robust insights into the consistency, fairness, and stakeholder perceptions of MotoGP sanctions. This refined approach equips stakeholders with actionable strategies for improving decision-making processes, ensuring that future sanctions are perceived as fair and justified, ultimately enhancing the integrity of the sport. The balance between detailed data exploration and forward-looking analytics positions this analysis as a vital tool for driving positive outcomes in MotoGP's regulatory framework.

6 Findings and Recommendations

The analysis of the 2022 MotoGP season reveals several key insights into the nature of infringements, the consistency of sanctions, and stakeholder sentiment. One of the most significant findings is the dominance of "Irresponsible Riding" as the most frequent infringement across all classes. This suggests either insufficient enforcement of riding standards or a need for enhanced rider training to promote safer practices on the track. The frequent imposition of "Long Lap Penalties" reflects the stewards' preference for penalties that directly impact race outcomes. However, the variation in penalty application raises concerns about consistency. While safety and fairness remain paramount, the inconsistent treatment of similar infractions indicates that the decision-making process may require further standardization to ensure uniform enforcement across all cases.

Aggravating factors such as "Caused Crash" and "Disturbing Other(s)" justify the severe penalties often associated with these infractions, given their direct impact on rider safety. However, less frequent factors like "Caused Danger/Safety Concern" and "Repeat Offence" also significantly influence sanction severity. This broad consideration of factors is logical but can lead to subjective weighting, which may result in inconsistent penalties and stakeholder dissatisfaction, particularly if the penalties are perceived as disproportionate.

Class-specific trends also emerged, with a notable high incidence of "Irresponsible Riding" in Moto3. This suggests that younger or less experienced riders may be more prone to risky behaviours, highlighting the need for targeted interventions, such as more comprehensive training programs or stricter enforcement of riding standards within this class. Additionally, variations in Flag Violations and Technical Infringements across different classes emphasize the necessity of tailored regulatory approaches to ensure fair and consistent enforcement across all competition levels. Addressing these class-specific trends is crucial to maintaining safety and fairness, as a one-size-fits-all approach may not be effective.

Circuit-specific and temporal trends further reveal potential areas for intervention. Circuits like Assen and Valencia emerged as hotspots for rule violations, suggesting that track design or competitive conditions may pose greater challenges to riders. Moreover, the high number of sanctions during the "Race" stage and FP3 highlights these periods as critical for rule enforcement. Riders tend to push the limits more aggressively during these times, indicating that better-timed interventions could reduce the likelihood of severe penalties and improve overall safety.

Judicial challenges and perceptions of the steward panels also significantly influence how decisions are received. The high frequency of hearings, particularly for "Irresponsible Riding" and Technical Infringements, suggests that these issues are often contested, reflecting their perceived seriousness. Furthermore, differences in public sentiment across steward panels indicate that some panels may be viewed as more controversial or less fair in their decision-making. This perception could undermine the credibility

of the stewarding process and calls for clearer guidelines and more transparent communication to reduce the perception of bias.

Stakeholder sentiment, particularly from riders, shows higher levels of negative sentiment compared to teams and media, who generally display neutral reactions. This suggests that riders are more likely to feel that decisions are unfair, especially regarding the perceived severity of penalties. Additionally, the divergence in media sentiment across different outlets highlights the significant influence of media in shaping public perceptions. Outlets like "www.the-race.com" tend to focus more on the negative aspects of incidents, while others maintain a neutral stance. This variation underscores the importance of managing media narratives to ensure fair and accurate reporting of decisions.

The findings from the MotoGP analysis align with and extend the theoretical frameworks discussed earlier. Stakeholder Theory emphasizes balancing the interests of diverse groups, and the results demonstrate the challenges the Stewards Panel faces in satisfying all stakeholders, especially when decisions are perceived as inconsistent. The application of Sentiment Analysis, as validated by studies like (Yu and Wang, 2015), proved effective in capturing real-time public and rider sentiment, revealing significant dissatisfaction among riders with certain penalties. The Ecological Dynamics Approach further contextualizes the dynamic and complex nature of decision-making in MotoGP, where adapting to rapidly changing race conditions is crucial. These frameworks collectively validate the complexity and the need for nuanced decision-making in this high-stakes environment.

6.1 Project Contribution

This project contributes to the existing knowledge by highlighting the interplay between fairness, stakeholder satisfaction, and real-time decision-making in a dynamic sporting environment. The findings support Stakeholder Theory by demonstrating the difficulty of balancing conflicting interests in MotoGP. The use of Sentiment Analysis in this context also contributes to its growing application in sports management, emphasizing its value in understanding stakeholder reactions. Moreover, the integration of the Ecological Dynamics Approach reinforces the importance of adaptive decision-making in sports, offering insights that could be applicable to other high-pressure environments.

6.2 Recommendations

To address the issues identified, MotoGP should prioritize the standardization of penalty guidelines to ensure consistent treatment of similar infractions across all classes. Drawing on principles from the International Olympic Committee's sanctioning guidelines, MotoGP can create a framework that emphasizes fairness and proportionality in penalties. By developing these guidelines with input from key stakeholders—such as stewards, teams, and riders—in upcoming seasons, MotoGP can reduce contested penalties significantly. This approach not only enhances fairness but also strengthens the credibility of the decision-making process. Potential challenges may include resistance from stakeholders who feel that the

new guidelines might not adequately address specific situations. To mitigate this, MotoGP should engage stakeholders early in the process and provide clear communication on the benefits of standardization.

Additionally, it is crucial to enhance rider education and training, particularly for younger riders in Moto3. Implementing mandatory training programs that focus on reducing risky behaviours can help lower the incidence of "Irresponsible Riding" in the upcoming season. These programs should be developed in collaboration with experienced riders and training schools, ensuring that they are both practical and effective. Rolling out these initiatives during the off-season will prepare riders for safer competition in the next season. A potential challenge might be reluctance from riders or teams due to the additional time commitment required. To overcome this, MotoGP could offer incentives or integrate the training with existing team activities to ensure participation.

Improving communication and transparency in the decision-making process is another essential step. By implementing a clear communication strategy, MotoGP can provide stakeholders with detailed explanations of penalties and decision-making processes. This strategy should be informed by best practices from other sports organizations, aiming to increase stakeholder understanding and acceptance of decisions. The challenge of ensuring consistent and timely communication can be addressed by designating a specific communication team responsible for this task, ensuring regular updates and feedback loops.

Furthermore, regulatory approaches should be tailored to the specific needs of different classes and circuits. By developing class- and circuit-specific regulations that address unique challenges, MotoGP can ensure that infractions are handled appropriately and fairly. These regulations should be introduced before the next season and periodically reviewed to assess their effectiveness.

Engaging with media outlets is also vital in shaping public perception. Establishing regular communication with key media outlets will help ensure accurate and balanced reporting of decisions. This includes providing detailed explanations of controversial decisions and facilitating direct communication between stewards and the media. The aim is to monitor and improve the sentiment of media coverage, ensuring that public perceptions are fair and informed. Challenges might include biased reporting from certain media outlets. To mitigate this, MotoGP could offer exclusive interviews or access to information in exchange for balanced coverage.

Finally, conducting post-season reviews of the most contentious decisions can build trust and ensure continuous improvement. By incorporating structured post-event analysis, as recommended in the IOC guidelines, MotoGP can refine its guidelines and processes for future seasons. These reviews should involve all relevant stakeholders and be conducted within three months of the season's end, ensuring that feedback is timely and actionable.

In conclusion, the findings of this analysis highlight the need for more consistent enforcement, improved communication, and tailored approaches to different challenges within MotoGP. By addressing these areas, MotoGP can enhance fairness, safety, and credibility in its sanctioning process.

6.3 Project Limitations

The project faced several constraints that impacted both the scope and the depth of the analysis. One of the primary limitations was data availability and coverage. Initially, the project aimed to analyse sentiment across multiple platforms (Instagram, Twitter, TikTok, Facebook) and over a broader timeframe (2021-2024). However, due to time constraints, technical challenges with web scraping, and limitations in data collection tools, the analysis was narrowed to focus primarily on Instagram and the 2022 season. This limitation restricted the ability to perform a more comprehensive trend analysis or time series analysis across multiple seasons, potentially affecting the generalizability of the findings.

Resource constraints also played a significant role. The technical difficulties in using automated tools like BERT for nuanced sentiment analysis, combined with the need for manual cross-verification, limited the scalability of the analysis. The project had to balance between accuracy and the time required for manual verification, which may have introduced some bias in the data interpretation.

Bias in data collection was another potential issue, particularly related to the platform-specific nature of the analysis. By focusing solely on Instagram, the study might not fully capture the sentiment of all MotoGP stakeholders, especially those who are more active on other platforms. Additionally, the inherent limitations of automated translation tools might have led to misinterpretations of non-English sentiments, affecting the accuracy of sentiment classification.

The generalizability of findings is also impacted by the narrowed focus of the study. While the analysis provides valuable insights into the 2022 season, the findings may not be fully applicable to other seasons or platforms. The exclusion of broader timeframes and additional platforms limits the ability to draw conclusions about long-term trends or the overall effectiveness of the MotoGP Stewards Panel's decisions across different contexts.

6.4 Future Opportunities for Research

Given the limitations of this project, there are several opportunities for future research that could extend and enhance the findings. Expanding the analysis to multiple platforms and seasons would provide a more comprehensive understanding of stakeholder sentiment over time. Incorporating data from platforms like Twitter and TikTok, alongside Instagram, would allow for a more diverse and holistic view of stakeholder reactions and might reveal platform-specific trends in public sentiment.

Time series analysis or predictive modelling could be valuable in future research to assess how sentiments evolve over time and potentially predict stakeholder reactions to future decisions. This could be particularly

useful for the MotoGP Stewards Panel to anticipate public and stakeholder responses to their rulings and make adjustments accordingly.

Another area for future exploration is the inclusion of qualitative methods such as interviews or focus groups with key stakeholders. While sentiment analysis provides quantitative insights, qualitative data could offer deeper understanding of the reasons behind specific sentiments, especially in contentious cases.

Finally, exploring the impact of specific interventions by the Stewards Panel—such as changes in penalty guidelines or communication strategies—could provide actionable insights into improving the fairness and transparency of their decision-making processes. Future research could also investigate the role of media coverage in shaping public perceptions and how the Stewards Panel might engage more effectively with the media to manage narratives around their decisions.

7 Conclusion

This consulting project aimed to critically assess the consistency and transparency of the MotoGP Stewards Panel's decisions during the 2022 season, focusing on stakeholder sentiment as expressed on Instagram. Through a structured, multi-phase analysis, including descriptive, diagnostic, and predictive analytics, the report has explored key patterns in rule enforcement and stakeholder reactions, providing actionable recommendations to enhance fairness and trust in MotoGP's decision-making processes.

The primary objective of the project was to identify contentious decisions, classify decisions by infringement type, and evaluate stakeholder sentiment to gauge perceptions of fairness and transparency. The findings revealed significant insights, such as the dominance of "Irresponsible Riding" as the most frequent infringement and the prevalence of "Long Lap Penalties" as a common sanction. These patterns indicate ongoing challenges in enforcing riding standards and highlight areas where the decision-making process could be more consistent.

The diagnostic analysis further uncovered inconsistencies in penalties for similar infractions, particularly across different circuits and rider classes. This suggests that the Stewards Panel's decision-making may sometimes be influenced by external factors, such as race conditions or rider nationality, leading to perceptions of bias. Sentiment analysis supported these findings, showing that negative reactions were more common among riders, particularly in cases involving severe penalties.

Reflecting on the analytical process, the mixed-methods approach proved effective in capturing the complexity of the MotoGP Stewards Panel's decisions. The combination of sentiment analysis with traditional statistical methods allowed for a nuanced understanding of how decisions are perceived by various stakeholders. However, the reliance on Instagram as the primary data source, while justified by its high engagement levels, introduced some limitations. Future research could expand the analysis to include other platforms and timeframes to provide a more comprehensive view of stakeholder sentiment.

The project's contributions to the field of sports governance are significant. By integrating sentiment analysis into the evaluation of steward decisions, this study provides a novel approach that could be applied to other sports facing similar challenges in maintaining fairness and transparency. The recommendations for standardizing penalty guidelines, enhancing rider education, and improving communication processes are not only practical but also grounded in the findings of the analysis, offering concrete steps that MotoGP could implement to improve its decision-making framework.

This project also challenges some existing knowledge in the field. While Stakeholder Theory emphasizes the need to balance the interests of various groups, the findings of this study suggest that achieving this balance is more complex in practice, particularly in high-stakes environments like MotoGP. The divergence in stakeholder sentiments, particularly between riders and other groups, underscores the difficulty of making

decisions that satisfy all parties. This highlights the importance of transparency and clear communication in mitigating perceptions of bias and unfairness.

The practical implications of this work are far-reaching. By addressing the inconsistencies in decision-making and improving the transparency of the Stewards Panel, MotoGP can enhance stakeholder trust and maintain its reputation as a fair and competitive sport. The recommendations provided in this report, if implemented, could lead to more consistent enforcement of rules, improved safety on the track, and greater stakeholder satisfaction.

Looking ahead, there are several potential future directions for research. Expanding the analysis to include a broader range of data sources and timeframes would provide a more comprehensive understanding of long-term trends in steward decisions. Additionally, exploring the impact of specific interventions, such as changes in penalty guidelines or communication strategies, could offer valuable insights into how the Stewards Panel can continue to evolve and improve.

In conclusion, this consulting project has successfully met its objectives by providing a detailed analysis of the MotoGP Stewards Panel's decisions during the 2022 season. The findings and recommendations presented here offer practical solutions to enhance the fairness, consistency, and transparency of decision-making in MotoGP. As the sport continues to grow and evolve, it is essential that its governance structures adapt accordingly, ensuring that all stakeholders feel confident in the integrity of the competition. This project contributes to that goal, providing a foundation for ongoing improvements in MotoGP's stewarding processes and offering valuable insights for the broader field of sports governance.

8 Bibliography

- Alam, M.T., Sohail, S.S., Ubaid, S., Shakil, Ali, Z., Hijji, M., Saudagar, A.K.J., et al. (2022), “It’s Your Turn, Are You Ready to Get Vaccinated? Towards an Exploration of Vaccine Hesitancy Using Sentiment Analysis of Instagram Posts”, *Mathematics*, MDPI, Vol. 10 No. 22, doi: 10.3390/math10224165.
- Ashford, M., Abraham, A. and Poolton, J. (2021), “Understanding a player’s decision-making process in team sports: A systematic review of empirical evidence”, *Sports*, MDPI AG, 1 May, doi: 10.3390/SPORTS9050065.
- Balagué, N., Hristovski, R., Vazquez, P. and Phd In Biomedical, N.B. (2008), *ECOLOGICAL DYNAMICS APPROACH TO DECISION MAKING IN SPORT. TRAINING ISSUES*, Vol. 4.
- BBC Sport (2015) ‘Moto GP: Valentino Rossi penalised for Marc Marquez clash’, *BBC Sport*. Available at: <https://www.bbc.co.uk/sport/motorsport/34630634> (Accessed: 12 August 2024).
- Breiman, L. (2001), *Random Forests*, Vol. 45.
- Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. (2018), “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”.
- Donaldson, T. and Preston, L.E. (1995), *The Stakeholder Theory of the Corporation: Concepts, Evidence, and Implications*, Source: *The Academy of Management Review*, Vol. 20.
- Floridi, L. and Taddeo, M. (2016), “What is data ethics?”, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Royal Society of London, 28 December, doi: 10.1098/rsta.2016.0360.
- Freeman, R.E.E. and McVea, J. (2005), “A Stakeholder Approach to Strategic Management”, *SSRN Electronic Journal*, Elsevier BV, doi: 10.2139/ssrn.263511.
- Garnica-Caparrós, M., Memmert, D. and Wunderlich, F. (2022), “Artificial data in sports forecasting: a simulation framework for analysing predictive models in sports”, *Information Systems and E-Business Management*, Springer Science and Business Media Deutschland GmbH, Vol. 20 No. 3, pp. 551–580, doi: 10.1007/s10257-022-00560-9.
- International Olympic Committee (IOC), 2018. Guidelines for sports organisations on the sanctioning of competition manipulation. [PDF] Available at: <https://stillmed.olympic.org/media/Document%20Library/OlympicOrg/IOC/What-We-Do/Protecting-Clean-Athletes/Competition-manipulation/Sanctioning-Guidelines-EN.pdf>
- Javani, V. and Karimivand, H. (n.d.). “Social Media and Sports Stakeholders Challenges and Opportunities”, doi: 10.22103/JNSSM.2022.19154.1068.
- Jiao, F., Guo, T., Li, J., Li, P., Cao, R. and Shen, Y. (2021), “Research on the Emotion Regulation Effect of Home Sports Based on Social Media Data Analysis and Online Experimental Investigation”, *Proceedings - 2021 International Conference on Information Technology and Contemporary Sports, TCS 2021*, Institute of Electrical and Electronics Engineers Inc., pp. 357–362, doi: 10.1109/TCS52929.2021.00079.

- Kujala, J., Sachs, S., Leinonen, H., Heikkinen, A. and Laude, D. (2022), “Stakeholder Engagement: Past, Present, and Future”, *Business and Society*, SAGE Publications Ltd, Vol. 61 No. 5, pp. 1136–1196, doi: 10.1177/00076503211066595.
- Lundberg, I., Narayanan, A., Levy, K. and Salganik, M.J. (2019), “Privacy, Ethics, and Data Access: A Case Study of the Fragile Families Challenge”, *Socius*, SAGE Publications Inc., Vol. 5, doi: 10.1177/2378023118813023.
- Mitchell, R.K., Agle, B.R. and Wood, D.J. (1997), Toward a Theory of Stakeholder Identification and Salience: Defining the Principle of Who and What Really Counts, *Source: The Academy of Management Review*, Vol. 22.
- Patil, S. and Davies, P. (n.d.). “Use of Google Translate in medical communication”, doi: 10.2307/26519051.
- Petersen, T.S. and Wichmann, S.S. (2021), “Fairness, implicit bias testing and sports refereeing: An argument for why professional sports organisations ought to promote fairness in sport through testing referees for implicit biases”, *Journal of the Philosophy of Sport*, Routledge, Vol. 48 No. 1, pp. 97–110, doi: 10.1080/00948705.2020.1866994.
- Weimer, J. (2022) ‘The inside scoop on how sports have been affected by the latest social media trends’, *Forbes*. Available at: <https://www.forbes.com/sites/jacksonweimer/2022/09/09/the-inside-scoop-on-how-sports-have-been-affected-by-the-latest-social-media-trends/> (Accessed: 13 August 2024).
- Yu, Y. and Wang, X. (2015), “World Cup 2014 in the Twitter World: A big data analysis of sentiments in U.S. sports fans’ tweets”, *Computers in Human Behavior*, Elsevier Ltd, Vol. 48, pp. 392–400, doi: 10.1016/j.chb.2015.01.075.

Appendices

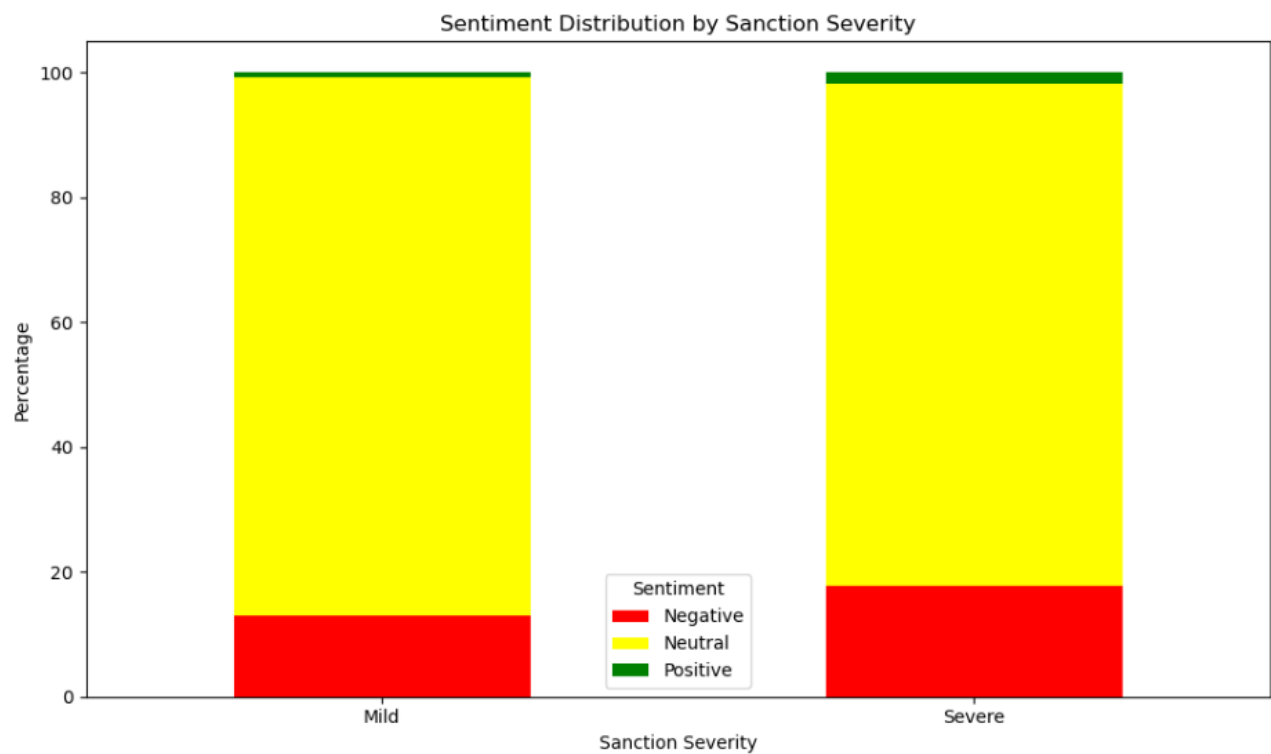


Figure 26: Sentiment Distribution by Sanction Severity

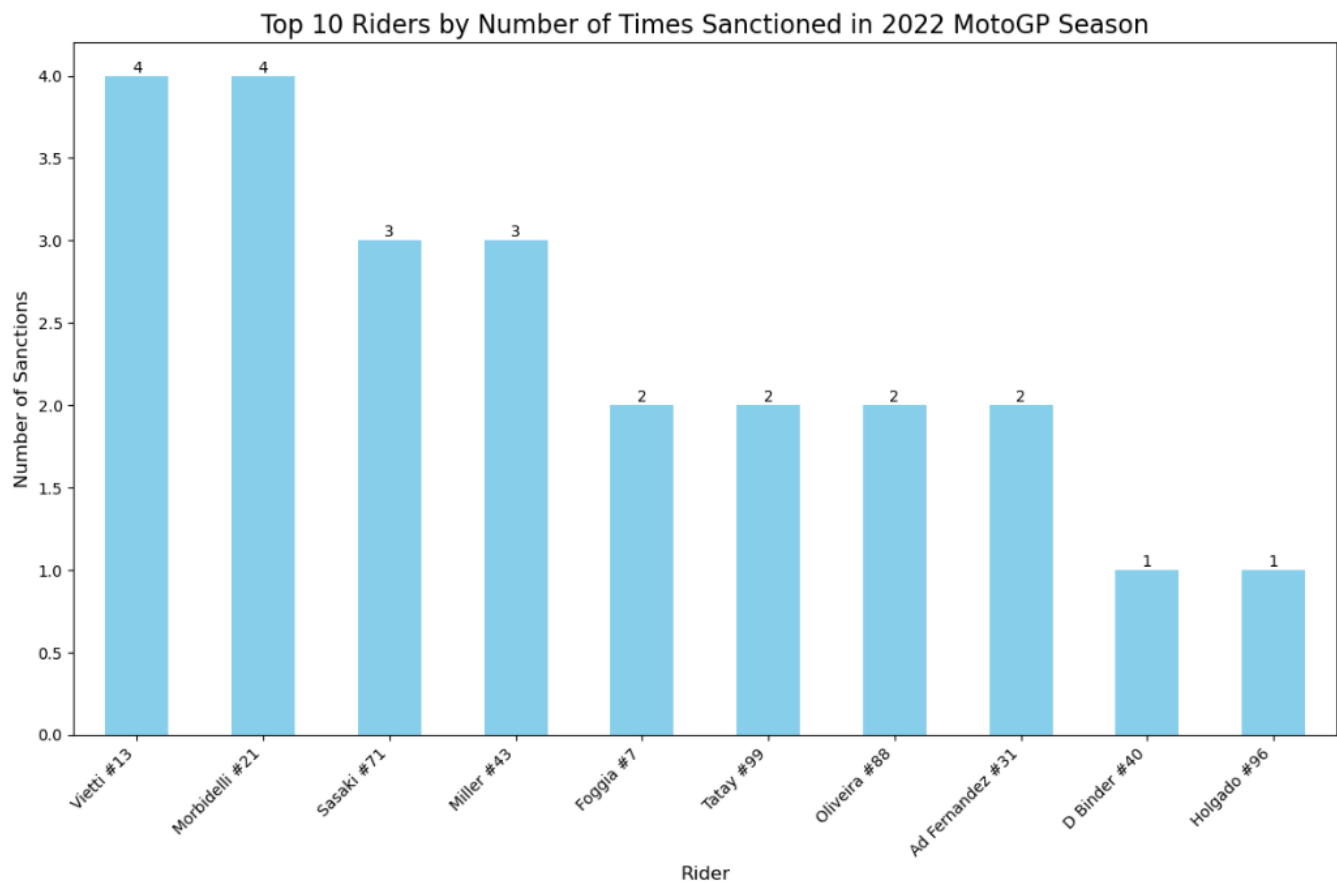


Figure 27: Riders with Most Penalty

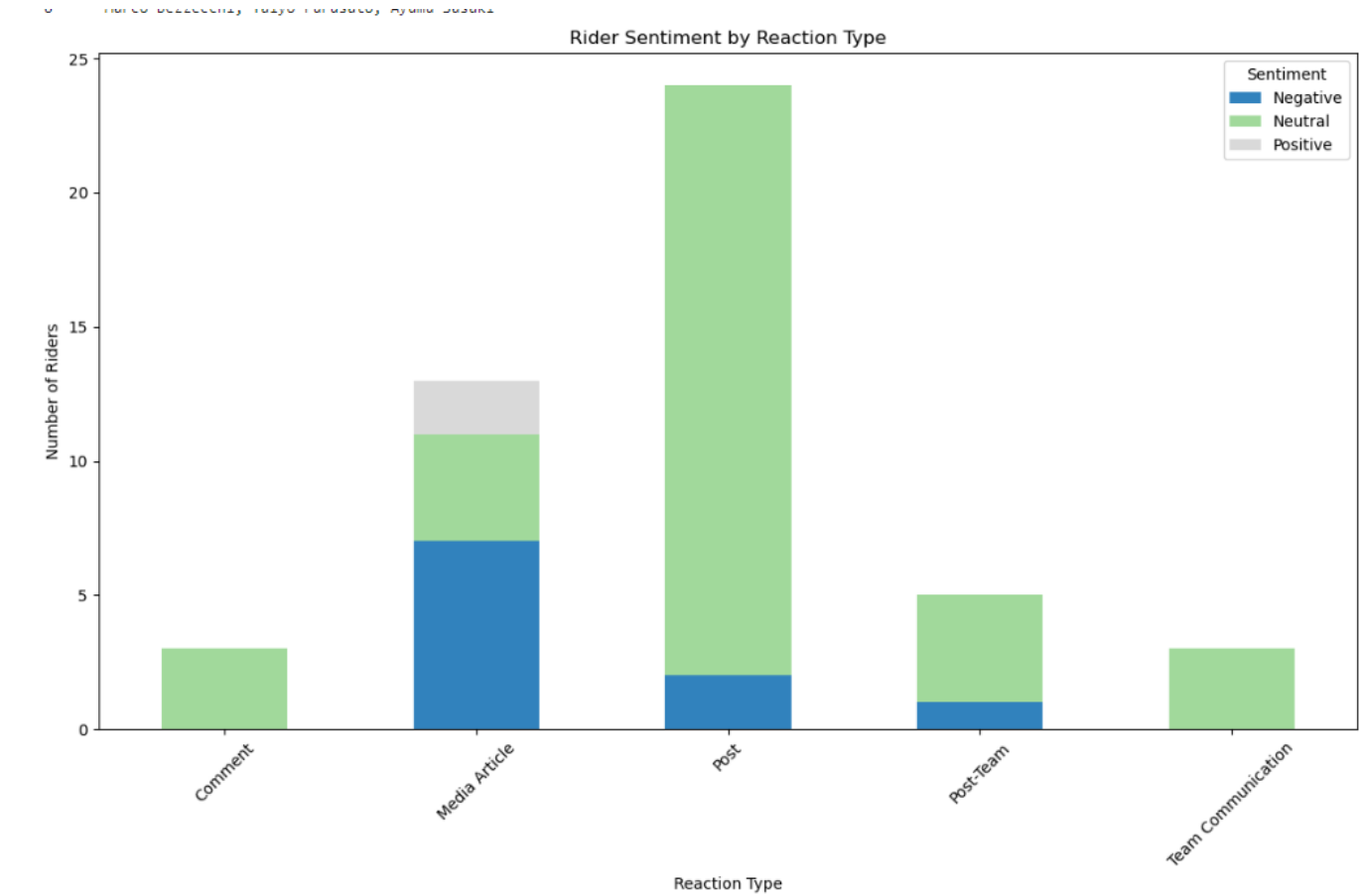


Figure 28: Rider Sentiment By Reaction Type

Infringement

Categorization Framework

| Infringement Types | Sub Classification | Aggravating Factors |
|-------------------------------------|--|--|
| Irresponsible Riding | 1. Overly aggressive 2. Riding slowly on the racing line 3. Rejoining track unsafely 4. Riding irresponsibly 5. Overly ambitious | 1. Caused Contact 2. Caused Crash 3. Disturbing Other(s) 4. Caused Danger/Safety Concern 5. Repeat Offence |
| Flag Violations | - | 1. Caused Crash |
| Technical Infringement | - | - |
| Jump, Incorrect and Practice Starts | - | - |
| Track Limits Violations | - | - |
| Failure to Comply with Penalties | - | - |
| Others | - | - |

Sanction Classification Framework

| Broad Sanction Classification | Examples from Sanction Column |
|-------------------------------|---|
| Long Lap Penalties | - Double long lap penalty for next event |
| | - Long lap penalty for race |
| | - Long lap penalty for Silverstone race |
| | - Long lap penalty next appearance |
| | - Long lap penalty for Portimao race |
| Grid Penalties | - Pit lane start + long lap penalty for race |
| | - Start race from pit lane, 5-second delay after green light |
| | - Start from back of grid for race |
| | - Long lap penalty + Start from the back of the grid for race |
| | - 3 grid position penalty for race |
| | - Start from the back of the grid for race |
| | - Three grid position penalty for race |
| Fine | - Fine €1,000 (paired with other sanctions) |
| Performance Penalties | - Q2 results disqualified |
| | - Race 1 results disqualified |
| | - Cancellation of last run of FP2 where offence committed |
| | - Cancellation of Q2 lap where pressures didn't comply |
| Official Warning | - Official warning |

| DATA DICTIONARY MASTER DATA SET(Infringement Data) | |
|--|---|
| Column Name | Description |
| Infraction ID | Unique identifier for each infraction. |
| DATE | Date when the infraction was recorded. |
| TIME | Time when the infraction was recorded. |
| PANEL MEMBERS | Names of the stewards panel members involved in the decision-making. |
| CIRCUIT | Name of the circuit where the infraction occurred. |
| CLASS | Class of the event (e.g., MotoGP, Moto2, Moto3). |
| STAGE | Stage of the event (e.g., Practice, Qualifying, Race). |
| RIDER | Rider's name and race number involved in the infraction. |
| TEAM | Name of the rider's team. |
| DETAILS OF INFRACTION | Detailed description of the infraction committed by the rider. |
| Infringement Type | Broad category of the infraction (e.g., Irresponsible Riding, Technical Infringement). |
| Sub Classification | More specific categorization within the infringement type (e.g., Overly aggressive, Riding slowly). |
| Aggravating factor | Any factors that worsened the severity of the infraction. |

| | |
|--------------------------------|---|
| ARTICLES CONTRAVENED | Specific articles of the MotoGP regulations that were violated by the infraction. |
| SANCTION | Penalty imposed for the infraction (e.g., Long lap penalty, Grid penalty, Fine). |
| Sanction Classification | Categorization of the sanction (e.g., Grid Penalties, Long Lap Penalties, Performance Penalties). |
| HEARING? | Indicates if a hearing was conducted (Y/N). |
| RIGHT OF APPEAL? | Indicates if there was a right of appeal against the decision (Y/N). |
| NOTES | Additional notes regarding the decision or the infraction. |
| PUBLIC SENTIMENT | Sentiment score or classification of public reactions to the infraction and decision (e.g., Positive, Neutral, Negative). |
| RIDER SENTIMENT | Sentiment score or classification of the rider's reaction to the infraction and decision (e.g., Positive, Neutral, Negative). |
| TEAM SENTIMENT | Sentiment score or classification of the team's reaction to the infraction and decision (e.g., Positive, Neutral, Negative). |
| MEDIA SENTIMENT | Sentiment score or classification of the media's reaction to the infraction and decision (e.g., Positive, Neutral, Negative). |

| Rationale for Layered Approach for Public Sentiment (due to limitations of BERT/ VADER in understanding context without fine tuning) | | | |
|---|---------------------------------|--------------------------------|--|
| Post Sentiment | Public Comment Sentiment | Final Comment Sentiment | Rationale |
| Positive | Positive | Positive | The post expresses a positive sentiment around the infraction, and the comment agrees, reinforcing the positive view of the decision. |
| Positive | Negative | Negative | The post is positive about the infraction, but the comment disagrees, introducing a negative perspective. The final sentiment reflects the disagreement with the positive stance of the post, hence negative. |
| Positive | Neutral | Positive | The post is positive regarding the infraction, and the comment is neutral, neither supporting nor opposing the sentiment. Since the comment does not counter the positive sentiment, the overall sentiment remains positive. |
| Negative | Positive | Negative | The post is negative about the infraction, expressing dissatisfaction. The comment, while positive, is likely responding to the negative sentiment and offering support, but it does not negate the negativity of the post. Therefore, the overall sentiment remains negative. |
| Negative | Negative | Positive | The post is negative about the infraction, but the comment disagrees with the post's stance by being negative toward it. This implies that the commenter might actually support the infraction, leading to a positive overall sentiment. |
| Negative | Neutral | Negative | The post is negative about the infraction, and the comment remains neutral, offering no counter-sentiment. The overall sentiment remains negative. |

| | | | |
|--------------|----------|--------------|--|
| Neutral | Positive | Neutral | The post is neutral, likely presenting facts or observations without strong sentiment. The comment, although positive, supports the neutrality of the post rather than introducing a new emotional direction. The positive comment does not strongly influence the overall tone, so the final sentiment remains neutral. |
| Neutral | Negative | Negative | The post is neutral, presenting information or observations without strong emotion. However, the comment introduces a negative sentiment, likely critiquing or expressing dissatisfaction with the content or context of the post. Since the negative comment shifts the overall tone away from neutrality, the final sentiment becomes negative. |
| Neutral | Neutral | Neutral | The post is neutral, providing information or a statement without any strong emotional tone. The comment also remains neutral, neither supporting nor opposing the content of the post. Since both the post and the comment are neutral, the overall sentiment remains neutral, as there is no shift towards positivity or negativity. The neutrality of both the post and comment maintains a balanced, impartial tone. |
| No Sentiment | Positive | No Sentiment | The post does not relate to any infraction, penalty, or decision, and thus is categorized as having "No Sentiment." Even though the comment expresses a positive sentiment, it is not relevant to the analysis of infractions or penalties, so the overall sentiment remains categorized as "No Sentiment." |
| No Sentiment | Negative | No Sentiment | The post does not address any infraction, penalty, or decision, and is categorized as having "No Sentiment." The comment, although neutral, does not introduce any relevant perspective on infractions or penalties. Therefore, the overall sentiment is maintained as "No Sentiment." |
| No Sentiment | Neutral | No Sentiment | Since the post itself does not mention any infraction, penalty, or decision and is categorized as having "No Sentiment," the negative sentiment of the comment is also irrelevant in this context. Thus, the overall sentiment remains as "No Sentiment." |

Methodological Choice/ Assumption

| Assumption/Methodological Choice | Rationale | Project Stage |
|---------------------------------------|---|---------------|
| The significance of social media data | Given MotoGP's strong social media presence, it was assumed that Instagram and other platforms would provide relevant sentiment data. This shaped the decision to focus heavily on these platforms for data collection. | Foundational |
| Timeline-based data collection | Collecting data within specific time frames (e.g., immediate reactions within 48 hours, short-term up to a week) was essential to capture the evolution of sentiment and to reflect real-world reactions. | Foundational |
| Prioritizing Instagram | Instagram was selected as the primary platform due to its high engagement metrics and content relevance, aligning with the project's objectives. | Foundational |

| | | |
|--|---|-----------------------|
| Treating 'No Response' as Neutral | Non-responses were treated as neutral sentiments, reflecting the assumption that silence typically indicates indifference or strategic neutrality in this context. | Analytical Adjustment |
| Language translation is necessary | Since MotoGP has a global fanbase, it was assumed that comments in various languages would be relevant, necessitating the use of tools like Google Translate for consistency. | Foundational |
| Cross-verification of official accounts | To ensure that sentiment analysis focused on public opinion, a cross-verification process was implemented to filter out comments from official MotoGP accounts, riders, and teams. | Analytical Adjustment |
| Categorization of infractions and sanctions | The project classified infractions into specific categories to facilitate the analysis, especially in differentiating similar cases for comparative purposes. | Analytical Adjustment |
| Stakeholder engagement reflects public sentiment | The sentiment of key stakeholders (riders, teams, media) was assumed to provide an accurate reflection of broader public sentiment. | Analytical Adjustment |
| Sentiment Analysis Tools (Manual vs. Automated) | TextBlob and VADER were tested but deemed inadequate for capturing nuanced sentiments. BERT, though powerful, required domain-specific tuning. Hence, a manual approach was chosen for its contextual accuracy. | Analytical Adjustment |
| Need for ethical compliance (GDPR) | Compliance with GDPR and ethical standards was a primary concern, leading to decisions to use GDPR-compliant tools and avoid certain data collection methods (e.g., web scraping without permission). | Foundational |
| The need for post-project adjustments | It was assumed that after the project deadline, additional adjustments or full deliverables may be necessary, based on the pilot results. | Analytical Adjustment |

Appendix A: Ethical Approval Form

ROYAL HOLLOWAY, UNIVERSITY OF LONDON
ETHICAL APPROVAL FORM

For staff and student Consulting Projects and research projects involving data collection from research participants (observations, interviews, questionnaires, group discussions, recordings, video etc.).

This form should be discussed and completed jointly by both student and supervisor (and in the case of staff, with their immediate line manager) with each keeping a signed copy of the form.

If the proposed work involves human participants and is judged by the supervisor/line manager potentially to give rise to ethical problems, ethical approval must be sought in advance. The supervisor will recommend whether the completed/signed form and any supporting material should be considered only by the Department’s internal approval procedures or be referred to the College Ethics Committee.

| <u>To be completed by the applicant</u> | Please select | |
|--|--------------------------|-------------------------------------|
| | YES | NO |
| 1. Will the study be <i>covert</i> in any way? | <input type="checkbox"/> | <input checked="" type="checkbox"/> |
| 2. Will resulting data be used for purposes outside this study? | <input type="checkbox"/> | <input checked="" type="checkbox"/> |
| 3. Are you working with a vulnerable population? | <input type="checkbox"/> | <input checked="" type="checkbox"/> |
| 4. Is it possible that your study will cause distress or harm to participants? | <input type="checkbox"/> | <input checked="" type="checkbox"/> |

If the answer to any of the above questions is “YES” please supply relevant supporting materials and explanations.

The working title of my Consulting Project/project is:

Evaluating MotoGP Stewards' Decisions: Social Media Sentiment Analysis of Riders, Teams, Fans, and Media

I am fully aware that the research carried out for my Consulting Project requires that I take due care of ethical issues.

I will ensure that consent is obtained from all participants which, saving exceptional cases, will be in writing.

These issues have been discussed with my supervisor.

| Student Name | Student Signature |
|---------------------------|-------------------|
| Sushant Thapa Magar | Sushant |
| <hr/> | |
| College Number: 101064383 | |
| <hr/> | |



To be completed by the supervisor (students) or line manager (staff)

Issues of ethics, copyright and data protection have been considered where necessary as indicated in the attached material and appropriate measures have been recommended. All necessary materials have been seen and the Ethics Committee’s *Notes for Guidance* have been consulted.

Please tick **ONLY ONE** box:



To be completed by the supervisor (students) or line manager (staff)

Issues of ethics, copyright and data protection have been considered where necessary as indicated in the attached material and appropriate measures have been recommended. All necessary materials have been seen and the Ethics Committee’s *Notes for Guidance* have been consulted.

| | |
|--|-------------------------------------|
| <u>Please tick ONLY ONE box:</u> | |
| 1. No referral necessary | <input checked="" type="checkbox"/> |
| 2. Form to be referred to departmental ethical approval procedures | <input type="checkbox"/> |
| 3. An application must be made to College Ethics Committee | <input type="checkbox"/> |

(For 2. and 3. please append supporting documents as required e.g. research project proposal (questionnaires, consent forms).

| | |
|------------------------|-----------------------------|
| Supervisor Name | Supervisor Signature |
| Dr T Ramjaun | <i>T Ramjaun</i> |
| ----- | ----- |

9 Appendix B: Project Development Diary

My MN5821 Development Diary

| Date | Code | My Diary Entries |
|------------|--------|---|
| 10-06-2024 | C | <ul style="list-style-type: none"> ❖ Held the project kick-off meeting to officially start the project. ❖ Discussed the objectives, key timelines, and milestones, with an emphasis on analyzing public opinion on MotoGP stewards' decisions. ❖ During the key discussion points, the project objectives were established, focusing on evaluating sentiment on social media concerning decisions made by MotoGP stewards. We formulated an initial project timeline that considered potential data collection challenges and other unexpected issues. Additionally, we established a routine meeting schedule and agreed on communication methods within the team to ensure consistent and effective collaboration. |
| 10-06-2024 | D | <ul style="list-style-type: none"> ❖ Due to time constraints, we decided against using APIs and instead focused on learning web scraping techniques for platforms like Instagram, TikTok, X, and Facebook. ❖ We decided to explore web scraping techniques for social media posts, utilizing tools such as Python, BeautifulSoup, Selenium, Scrapy, and Chrome WebDriver. |
| 12-06-2024 | C D | <ul style="list-style-type: none"> ❖ My colleague focused on delving into web scraping and its associated challenges, leveraging AI tools for code samples and YouTube tutorials to grasp the process in preparation for receiving data from clients. ❖ Collaboratively, we began studying the ethical and legal considerations, potential bans on social media accounts due to web scraping, and the costs involved. Additionally, I synthesized the project brief to create a comprehensive plan and workflow for the overall project, and started developing a checklist for the same. ❖ Developed a basic checklist of what can be done to be adjusted further with the receipt of Stewards Decisions Dataset. |
| 13/06/2024 | D | <ul style="list-style-type: none"> ❖ Received dataset relating to Infringement with the following suggestions/discussions: <ol style="list-style-type: none"> 1) Focusing on only the MotoGP class (due to the time constraint) 2) 21-23 dataset- Missing entries (to be reviewed later) 3) 24 dataset-updated information till the Mugello race. Whether or not to include the Assen race? (to be decided, more likely to include). Also, some entries marked with red question marks require more information. 4) No Further Action Dataset- Only in 2024 since it is new. Only include in the Dashboard (as per the client). More likely to form part of the whole analysis (makes sense). No action by the Stewards can often prompt as much outpouring of views on social media as steward action can. |
| 14/06/2024 | C | <ul style="list-style-type: none"> ❖ Joined my colleague in learning about web scraping and started hands-on testing with Python. Viewed YouTube videos for additional guidance and insights. The research into web scraping techniques and their application continued throughout the week. |

| Date | Code | My Diary Entries |
|------------|------------|---|
| 15/06/2024 | M | <p>❖ Email to Supervisor: Provided an update on the timeline and next steps, covering the following points:</p> <ol style="list-style-type: none"> 1) Objectives: Overview of project goals. 2) Timeline Summary: A brief recap of the project schedule. 3) Key Milestones: Important project milestones. 4) Platforms: Information on the tools to be used, including Tableau/Power BI for data visualization and Python for data analysis. |
| 29/06/2024 | D M | <p>❖ Observations and plans:</p> <ol style="list-style-type: none"> 1) Verify and organize Stewards panel data from 2021 to mid-2023. 2) Remove irrelevant data not related to MotoGP from the datasets and send back for review. 3) Address missing race data due to access changes during 2023. 4) Assessment to confirm the use of updated data (Assen), which is more likely. 5) Ensure consistency with previous data formats. 6) Assess red question marks related to specific regulations and infringements. 7) Incorporate 'No Further Action' data into the dashboard for comprehensive analysis. 8) Analyze social media reactions and comments on incidents where no action was taken. <p>❖ Emailed clients and supervisors with the update in the project.</p> |
| 1/07/2024 | C | <ul style="list-style-type: none"> ❖ Basic data set cleaning. ❖ Deciding on the stakeholder (breaking down the email also) with the rationale behind it. ❖ Basic research on how to tackle the language barrier issue (Reactions on social media can be in languages other than English). ❖ All the difficulties and limitations note it down for further discussion (possibly in Excel/word) ❖ How to web scrape around a particular infringement in general? ❖ Is it possible to web scraping as a whole (by defining conditions- hastags, words, phrases, time frame, stakeholders name and so on) than having to go through each infringement manually? ❖ How to tackle the time of the reaction issue (For stakeholders can reach to a decision in a weeks or a month) <p>❖ We discussed the limitations encountered during our web scraping efforts, including multiple errors and bugs that arose. Additionally, we noted that Twitter had banned web scraping as of mid-2023 to improve the quality of its app, which has impacted our ability to collect data from that platform.</p> |
| 1/07/2024 | D | <p>❖ Cleaned the original dataset to remove Race Classes- 2 and 3.</p> |

| Date | Code | My Diary Entries |
|------------|------|--|
| 1/07/2024 | C | <ul style="list-style-type: none"> ❖ How significant are Hearing and Right of appeal for the overall analysis? Are these good variables for the dashboard purpose? How do we deal with appeals rejected in terms of overall analysis? ❖ How do you classify the infringement types? Are they pre-defined since we have already decided if there can be a request for appeal or not? Can we get a supplementary dataset that includes the pre-defined rules/types for the same? ❖ What does the note column in general mean? ❖ How can the date be relevant for the penalty imposed? ❖ Why is offense type inconsistently defined? ❖ What do the red question marks mean here? <p>How can the columns/variables be used for the overall analysis?</p> |
| 1/07/2024 | D | <ul style="list-style-type: none"> ❖ Mail to be sent to client regarding the questions in the data set by 4th July, 2024. |
| 2/07/2024 | D | <ul style="list-style-type: none"> ❖ Use the riders' Wikipedia links to compile comprehensive lists of riders and teams for each season from 2018 to 2024. ❖ Plan to manually collect articles and comments from key MotoGP blogs and websites, as well as social media accounts of these media houses, to understand professional opinions on stewards' decisions and rider performances (Other Stakeholders). |
| 2/07/2024 | C | <ul style="list-style-type: none"> ❖ Reflect on the feasibility and effectiveness of manually collecting data from Wikipedia, blogs, and social media accounts. Consider the initial observations and challenges in integrating these different data sources. ❖ Can it be done by using automated Python Scripts? ❖ Check the feasibility of web scraping and API for all the data collection. This is the most difficult and important technical part of the analysis. |
| 2/07/2024 | M | <ul style="list-style-type: none"> ❖ Have and make a bird's eye view of why and how to proceed with data collection keeping all the stakeholders and channels in mind. |
| 5/07/2024 | M | <ul style="list-style-type: none"> ❖ Emailed the signed Ethics form to supervisor. |
| 8/07/2024 | C | <ul style="list-style-type: none"> ❖ We reviewed the issues with web scraping, including platform restrictions and technical difficulties, and recognized the need for additional guidance. Due to time constraints and the complexity of the problems, we decided to reach out to our supervisors for assistance and clarification. Drafted an email to supervisors outlining our current challenges and seeking their input on alternative approaches and dataset clarifications. |
| 9/07/2024 | M | <ul style="list-style-type: none"> ❖ Received the signed Ethics form to supervisor. And uploaded the same. |
| 10/07/2024 | C | <ul style="list-style-type: none"> ❖ Sent an email to supervisors addressing the difficulties encountered with web scraping from Instagram, Twitter, Facebook, and TikTok, including technical barriers, platform restrictions, and time constraints. Detailed the challenges of dynamic content, anti-scraping measures, and legal risks, and discussed potential alternative approaches for data collection. Additionally, included requests for clarifications on the provided datasets to ensure alignment with project goals. |

| Date | Code | My Diary Entries |
|------------|------|---|
| 12/07/2024 | C | <ul style="list-style-type: none"> ❖ Received an updated Excel penalties sheet from the client for the Assen and Sachsenring races. Noted a correction on line 93 where Rins was correctly identified as belonging to Yamaha Factory Racing. No updates needed for the 'no further action' issues. |
| 15/07/2024 | C | <ul style="list-style-type: none"> ❖ Meeting with Supervisors: <ul style="list-style-type: none"> • Discussed rescoping the project to address ongoing challenges. • Agreed to maintain the immovable project deadline of August 28th while using it as a pilot phase for learning. • Plan to submit full deliverables to the client after August 28th, with project rescoping details to be finalized with the client. ❖ Web Scraping Challenges: <ul style="list-style-type: none"> • Decided to explore third-party web scraping tools, such as Apify. • Identified need for client funding to cover costs associated with these tools due to difficulties with Python. • Challenges include technical limitations, frequent changes in web page structures, and platform restrictions affecting consistency and reliability. |
| 17/07/2024 | C | <ul style="list-style-type: none"> ❖ Email Sent to The Clients: <ul style="list-style-type: none"> • Clarifications Requested: Sought information on standardized infraction types, missing Race Class details, and the status of entries marked with red question marks in the datasets. • Web-Scraping Challenges: Explained issues with scraping data from Instagram, Twitter, Facebook, and TikTok due to restrictions and time constraints; proposed using third-party tools like Apify and Export Comment. • Project Scope Adjustments: Suggested dividing the project into a pilot phase due by August 28, with full deliverables to follow; emphasized the need to adjust the scope based on pilot results. ❖ Client Response: Suggesting a discussion with supervisors to determine the next steps. Highlighted the potential risks of scraping, including ethical concerns and possible legal issues, and proposed exploring third-party tools if ethically approved or rescoping the project to use only publicly available datasets. |
| 18/07/2024 | C | <ul style="list-style-type: none"> ❖ Received clarification on the dataset questions and reflected on those. |

| Date | Code | My Diary Entries |
|------------|------|--|
| 18/07/2024 | C | <ul style="list-style-type: none"> ❖ Supervisors Support: <ul style="list-style-type: none"> • TwSearch Tool: Provided access to 100 tweets using the term '#MotoGP Steward' and information on paid options for more extensive searches. • GDPR and Copyright Considerations: Shared insights on ethical and legal considerations for using social media data, including UK GDPR guidelines, copyright, and sui generis database rights. Emphasized the need for careful handling and potential legal consultation. ❖ Collaboration: Reviewed Dr. Asmat's recommendations and data sources, including potential ethical issues and legal considerations. Discussed internally and with team members the implications of using third-party tools like TwSearch, Apify, Export Comments and adhering to GDPR and copyright regulations. ❖ Reflection: The feedback highlights the importance of navigating legal and ethical aspects of data collection. We need to ensure compliance with GDPR and consider the use of third-party tools for data collection. This requires detailed planning and possible legal consultation to align with ethical standards. We plan to incorporate these insights into our project scope and discuss with supervisors how to proceed effectively. ❖ Comparison of Social Media Platforms: Decided to prioritize Instagram based on its engagement metrics and explore tools like Export Comments as alternatives. Discussed limitations of various tools and attached sample data. |
| 23/07/2024 | D | <ul style="list-style-type: none"> ❖ Web-Scraping Tools: Apify is generally GDPR-compliant but requires careful user adherence to regulations. ExportComments.com raises significant GDPR compliance concerns, particularly regarding data subject rights and lawful processing. ❖ Scope Change: Consider focusing on Instagram due to its high engagement. Evaluate its feasibility for our needs and whether it complements existing panel decision data. ❖ Alternative Sources: Review the PDF with alternative fan opinion sources provided by Dr. Asmat for potential use in data collection. ❖ Ongoing discussion with Shruti about Instagram focus and tool compliance. Reviewed additional data sources. Decided to contact Apify and ExportComments.com to inquire about their GDPR compliance policies. ❖ Facing compliance challenges with web-scraping tools. Apify seems safer but needs careful use. ExportComments.com presents major risks. Finalize data collection strategy with an emphasis on ethical and legal considerations. |
| 24/07/2024 | C | <ul style="list-style-type: none"> ❖ Emailed the client with extra clarification regarding the dataset. |

| Date | Code | My Diary Entries |
|------------|------|---|
| 25/07/2024 | D | <ul style="list-style-type: none"> ❖ Decided that we will meet on Monday to discuss Apify's GDPR compliance, compare it with Export Comments, define the scope of social media platforms and stakeholders, review deadlines, and address dataset questions. ❖ We have confirmed Apify's GDPR compliance and contacted both Apify and Export Comments about their agreements with social media platforms. We have also applied for Reddit API keys and are exploring additional blogs and websites for stakeholder data. Confirmed that export comments also complies to GDPR. ❖ Received clarifications from the client about the dataset and the clarity is now established. The initial exaggeration of issues stemmed from confusion over multiple steward notes for a single penalty. This has been resolved, leaving only the original red marks as missing information, now significantly reduced. |
| 29/07/2024 | D | <p>Summary of Meeting with The Clients:</p> <ul style="list-style-type: none"> ❖ Export Comments Subscription: Decided to use Export Comments for data collection with a flexible monthly subscription. Awaiting supervisors' confirmation on ethical compliance and financial implications. ❖ Social Media Focus: Focus will be solely on Instagram for the project. The analysis will cover the 2022 season (March to November 2022) to align with the deadline. ❖ Language Handling: Comments will be in English. If non-English comments are encountered, Python will be used for translation. A flag will be maintained to indicate translated comments. ❖ Media Stakeholders: Concentrate on Instagram for media stakeholder sentiment. Blogs will not be included to meet the project deadline. ❖ Dataset Focus: Analysis will be on the 2022 season data, with the deliverable due on August 28th. ❖ Stewards Decisions Data: Appeal data will be excluded from the main analysis but included in the dashboard. Red question marks in the dataset will be addressed by Sushant. ❖ Infraction Classification: Different infractions will be classified using language models and submitted for approval. <p>Next Steps:</p> <ul style="list-style-type: none"> ❖ Proceed with the above decisions. Adjustments will be made if deadlines change. Awaiting feedback and any further clarifications. |

| Date | Code | My Diary Entries |
|------------|------|---|
| 30/07/2024 | D | <p>❖ We decided to wait for approval to fund and use Export Comments. In the meantime, we have divided tasks as follows:</p> <p>1. Task Allocation:</p> <ul style="list-style-type: none"> • Shruti: Responsible for collecting URLs for data scraping based on her best judgment. • Sushant: Will develop a checklist for categorizing collected URLs to ensure comprehensive coverage of all channels and stakeholders. Will also work on categorizing the infractions and sending for approval. <p>2. Checklist Development:</p> <ul style="list-style-type: none"> • Create a checklist to categorize URLs by: <ul style="list-style-type: none"> ▪ Social Media Platform ▪ Media Stakeholder Type ▪ Relevance to Analysis Objectives • Ensure the checklist allows for tracking and verifying that all intended channels and stakeholders are covered. |
| 01/08/2024 | M | <p>❖ Started planning on the reporting part and how to use the word count available for each section. Completed the intro part of the report and went through journal articles supporting the context and significance of the project as a whole.</p> |
| 03/08/2024 | D | <p>❖ Sent email to clients for last missing details in 2022 season and also categorisation of infractions to be approved by the clients.</p> |
| 06/08/2024 | C | <p>❖ Worked on the infraction classification as advised by the client.</p> |
| 07/08/2024 | D | <p>❖ Received update from client regarding funding of third party tool for exporting comments. Started planning the project in that manner.</p> |
| 07/08/2024 | D | <p>❖ Confirmed with clients via email whether the analysis should focus solely on MotoGP, as previously discussed during the scope redefinition, or if it should encompass Moto2, Moto3, and MotoGP collectively for the year 2022 on Instagram.</p> |
| 08/08/2024 | D | <p>❖ Created separate sheets for Riders and Teams accounts to accurately cross-verify comments exported from relevant Instagram posts. This organization ensures precise attribution of reactions and improves the accuracy of the sentiment analysis. By organizing the accounts this way, I was able to systematically match comments with their respective stakeholders, ensuring that each reaction was accurately attributed to the correct rider, team, or media outlet.</p> <p>Worked on Classification of infraction.</p> |

| Date | Code | My Diary Entries |
|------------|------|---|
| 09/08/2024 | D | ❖ Created separate sheets for Media accounts as above. Also vaguely determined which were useable based on followers, popularity as well as content relevant and communicated to shruti regarding the same. |
| 10/08/2024 | C | <ul style="list-style-type: none"> Started the Literature Review and how this part of the report should look like to maintain a birds eye view. |
| 11/08/2024 | C | <ul style="list-style-type: none"> Started brainstorming on how we can mobilize the media as stakeholders while working on the infractions. I focused on analyzing media sentiment related to penalties by reviewing various blogs, news outlets, and Instagram accounts. During this process, I noticed that media content often featured rider opinions or interviews, which required a more nuanced approach to accurately capture the media's sentiment. To ensure comprehensive coverage, I selected the most relevant media channels from a list suggested by Dr. Jane (clients) and carefully determined the frequency of posts, making sure all penalties and related opinions were adequately represented. |
| 12/08/2024 | D | <ul style="list-style-type: none"> ❖ Communicated refined dataset of infraction classification to the clients with proper details. Also received some feed back and started working on that: • Distinguished between "Long Lap in Race" and "Long Lap Next Race" penalties for clearer categorization. • Grouped start penalties under the "Grid Penalties" heading as per the feedback. • Ensured consistent use of British English, specifically changing "offense" to "offence." • Clarified that there are no sub-classifications for any infringement type other than "Irresponsible Riding." |
| 13/08/2024 | D | <ul style="list-style-type: none"> ❖ Final refinement on the infraction classification to align with client requirement: • Track Limits Violation: Added "Track Limits" as a classification in the system, ensuring its presence even if not recorded in the 2022 season data. • Personal Aggression: Introduced a new classification for "Personal Aggression" to capture incidents involving fights between riders or aggression towards marshals. • Aggravating Factors: Applied relevant aggravating factors across all infractions for consistency and comprehensiveness. • Performance Penalty: Replaced the terms "Time Penalty" and "Disqualification" with "Performance Penalty" to better encapsulate various penalties within the sanction classification. • Final Dataset Submission: Attached the final dataset version for review, with all adjustments incorporated as per the feedback. |

| Date | Code | My Diary Entries |
|------------|------|---|
| 14/08/2024 | C | <ul style="list-style-type: none"> ❖ Worked on finding proper literature supporting the case and the differences thereon. Also started to brainstorm on the analysis part and planned the following: <ul style="list-style-type: none"> • Reorganized Sentiment Data: Grouped the sentiment data into three primary categories: Riders and Teams Sentiment Dataset, Media Sentiment Dataset, and Public Sentiment Dataset, ensuring a structured approach for analysis. • Dataset Structure Implementation: Designed each dataset to exclude profile IDs, focusing on key fields such as Infraction_ID, Entity_Name, Content_Type, Content_Text, and Date, allowing for streamlined analysis. • Integration Strategy: Established a method to link sentiments back to specific infractions using a unique Infraction_ID, enabling cross-referencing across different datasets (Riders and Teams, Media, Public). • Sentiment Aggregation Approach: Developed a workflow to start with individual sentiment analysis for each stakeholder, then aggregate these into an overall sentiment for each infraction using rules like majority rule or weighted approach. • Data Maintenance: Maintained separate tabs in Excel for each dataset and ensured consistent column headers across all datasets for easy integration and analysis. • Visualization and Reporting: Planned for the creation of a dashboard or report to visualize overall sentiments and provide drill-down capabilities into detailed sentiments from individual stakeholders. |
| 15/08/2024 | C | <ul style="list-style-type: none"> • Started to work on excel sheet for Rider and Team Sentiment Analysis. Organized the sentiment data by adding an "Infraction ID" column to uniquely identify each infraction, ensuring all reactions could be traced back to specific incidents. To categorize the reactions, I created three new columns: "Stakeholder Type," "Stakeholder Name," and "Reaction Type." This structure was implemented to facilitate clearer analysis of the data, enabling easy identification of who is reacting (riders, teams, etc.) and how (posts, comments). The rationale behind this was to streamline the process of linking stakeholder sentiments directly to infractions, allowing for more efficient and accurate analysis in the subsequent stages of the project. |
| 16/08/2024 | D | <ul style="list-style-type: none"> ❖ Completed the Riders and Team Sentiment Dataset. But more information may come from cross verification with the public sentiment data, if riders have made any comments thereon, and from media data if they have quoted riders thereon. And the following challenges were faced and resolved: <ul style="list-style-type: none"> • Vague Comments: Used Python for sentiment analysis but cross-verified manually to ensure accuracy, as these comments are mostly neutral. Automated tools assist, but human judgment captures nuances. • No Responses: Treated "No Response" cases as neutral since silence likely indicates neutrality in contexts where responses are expected. • Team Quotes: Considered team quotes of riders as endorsements, reflecting the team's stance and maintaining consistent sentiment classification. |

| Date | Code | My Diary Entries |
|------------|------|---|
| 16/08/2024 | C | <ul style="list-style-type: none">❖ Followed up on Media accounts with Shruti and had a discussion. Started working on Sentiment Analysis for Public Comments which should start with cross verification of comments from the riders.❖ One significant challenge was dealing with the varying nature of media sources. Instagram accounts and other social media platforms were initially considered, but due to limitations in extracting relevant narrative content from these platforms, the focus shifted to using URLs from more traditional media sources.❖ The process required manually extracting relevant comments and narratives from the articles linked in the URLs provided. These comments were then planned to be analyzed for sentiment to gauge the general media and public perception of the penalties. |

| | | |
|------------|---|---|
| 16/08/2024 | D | <ul style="list-style-type: none"> ❖ Objective: Today, the goal was to conduct a cross-verification between the public sentiment dataset (exported comments) and the consolidated accounts list of riders, teams, and media. After cross-verification, the aim was to clean the original dataset by removing comments that were identified as coming from official accounts. ❖ Tasks and Process: <ul style="list-style-type: none"> ❖ Cross-Verification Implementation: <ul style="list-style-type: none"> • Loaded the "Sentiment Data-Public.xlsx" file, which contains multiple sheets corresponding to different infractions. • Loaded the "Consolidated accounts list.xlsx" file, which includes the Instagram handles of riders, teams, and media accounts. • Developed a Python script to iterate through each sheet in the sentiment data file and cross-check the Username column against the Instagram Handle column from the consolidated accounts list. • Added a new column, "Type," to the cross-verified results to classify whether the comment was made by a rider, team, or media account. ❖ Results of Cross-Verification: <ul style="list-style-type: none"> • Successfully identified and extracted all relevant comments made by riders, teams, and media across all sheets in the sentiment data file. • Ensured that the results included a "Sheet Name" column to track where each comment originated and a "Type" column to indicate the account type. ❖ Cleaning the Original Dataset: <ul style="list-style-type: none"> • Created a new Python script to iterate through the original sentiment dataset and remove any comments that matched the cross-verified list. • Carefully handled cases where certain sheets in the original dataset did not contain any Username data or where vague or incomplete data was present. • The cleaned dataset was saved to a new Excel file for further analysis. |
| | C | <ul style="list-style-type: none"> ❖ Verification of Cleaning Process: <ul style="list-style-type: none"> • Performed a verification step by checking if any of the Username entries from the cross-verified list were still present in the cleaned dataset. • The verification confirmed that no such entries remained, indicating that the cleaning process was successful. ❖ Challenges Encountered: <ul style="list-style-type: none"> • Handling Hyperlinked Data: Initially encountered issues with the 'Instagram Handle' column because the data was stored as hyperlinks. This required parsing and adjusting the code to correctly read and process these values. • SettingWithCopyWarning: Faced multiple warnings related to chained assignments, which were resolved by using .loc for explicit indexing. • Invalid Index Handling: Managed challenges related to mapping the 'Type' column from non-unique indexes by ensuring uniqueness before proceeding with the mapping. ❖ Outcome: <ul style="list-style-type: none"> • Successfully completed the cross-verification and cleaning process. The cleaned dataset is now ready for the next phase of sentiment analysis, having removed all comments identified as originating from official accounts. |

| Date | Code | My Diary Entries |
|------------|---------------------------------|--|
| | | <ul style="list-style-type: none"> Ensured that the process is fully documented and reproducible for future reference. |
| 17/08/2024 | <p>D</p> <p>C</p> | <ul style="list-style-type: none"> ❖ Finalised the URLs of Blog/ New of media and sports outlet relevant to reactions around the infractions as well as penalties imposed by Steward Panel Decision. ❖ Tranformed multiple sheets from an Excel file containing media narratives into a single sheet, enabling sentiment analysis using Pandas Library in Python. The focus was on assessing media sentiment around infractions or penalties imposed by MotoGP stewards. Part of the task involved identifying the media source from URLs and converting them into recognizable media outlet names.This step was crucial to standardize the data for further analysis. ❖ Researched on how TextBlob, VADER and BERT can be used for sentiment analysis. Did trial and error for each to see if the sentiment generated really represented the nuanced setiment around the decisions from the relevant stake holders |

| | | |
|------------|---|--|
| 17/08/2024 | D | <div><div><div>❖ Contextual Limitations:</div><div>TextBlob and VADER are indeed lightweight tools designed for general sentiment analysis, but they often struggle with the nuanced and complex contexts found in detailed articles, especially those involving sports narratives like MotoGP. They primarily operate on a word-level sentiment analysis, which means they can't effectively capture the subtleties and context of a statement. This leads to oversimplified or inaccurate sentiment classifications, particularly when dealing with mixed sentiments or specific situations like penalties and reactions.</div><div>BERT, while more sophisticated, can better understand context through its deep learning capabilities. However, it requires extensive fine-tuning on domain-specific data to be truly effective. Without this fine-tuning, BERT might not capture the specific sentiment nuances related to MotoGP penalties and rider reactions accurately.</div></div><div><div>❖ Inability to Focus on Quoted Reactions:</div><div>Sentiment analysis in this project needed to be highly specific, focusing only on quoted reactions from riders and teams regarding penalties. TextBlob and VADER analyze sentiment across entire texts without distinguishing between general narrative and direct quotes, leading to potential misinterpretations. Even BERT, although capable of handling context better, does not inherently distinguish between narrative and quotes unless explicitly programmed to do so. This would require additional processing and customization, increasing the complexity and time required for the analysis.</div></div><div><div>❖ Complex Implementation and Resource Requirements:</div><div>BERT is a powerful tool but implementing it for nuanced sentiment analysis is resource-intensive. It requires significant computational power and expertise to fine-tune and deploy effectively, particularly for domain-specific tasks like analyzing MotoGP articles. The complexity of setting up such a system, combined with the need for large amounts of labeled data, made it impractical for this project, where quick, accurate sentiment analysis was needed without extensive overhead.</div></div><div><div>❖ Why Manual Analysis?</div><div>❖ Contextual Accuracy: Manual analysis allowed us to fully understand and analyze the entire article's context, accurately distinguishing between media, rider, and team sentiments.</div><div>❖ Focus on Quoted Reactions: The manual approach enabled us to isolate and analyze only the quoted statements from riders and teams, ensuring a true reflection of their sentiments.</div><div>❖ Simplified Process: This method reduced the need for complex programming and computational resources while providing more accurate and nuanced results.</div></div><div><div>Importance of Quoted Reactions:</div><div>❖ Quoted reactions from riders and teams were essential in our analysis because they represent the most accurate and direct expression of their sentiments and opinions. Unlike general narrative or third-party interpretations, quotes provide unfiltered insights into how the individuals involved genuinely felt about the stewards' decisions. This focus ensures that our sentiment analysis is based on the actual words of the riders and teams, avoiding any potential bias or misrepresentation that might occur if we relied on paraphrased or interpreted content.</div></div></div> |
|------------|---|--|

| Date | Code | My Diary Entries |
|------------|------|--|
| 18/8/2024 | D | Final Approach: <ul style="list-style-type: none"> ❖ We moved ahead with manual analysis of articles, combined with manual cross-checking for each entry, ensuring that sentiment was accurately tagged as positive, neutral, or negative. This approach allowed us to fully understand and capture the context, providing the most reliable and nuanced sentiment analysis possible. Task completed. |
| 18/08/2024 | D | <ul style="list-style-type: none"> ❖ Finalized the analysis of contentious cases using diagnostic analytics. Focused on the distribution of sanctions and correlation analysis between hearings and appeals. Noted patterns in stakeholder sentiment, particularly the relationship between sanction severity and negative responses. Drafted initial findings based on these insights. ❖ Finalizing the analysis and drafting findings based on stakeholder sentiment patterns. |
| 18/08/2024 | C | <ul style="list-style-type: none"> ❖ Ensured accurate interpretation of sentiment data and managed potential biases due to selective social media data. ❖ Dealt with the accuracy of sentiment interpretation and bias in data. |
| 18/08/2024 | D | <ul style="list-style-type: none"> ❖ Began writing the discussion section and align findings with theoretical frameworks ❖ Planned to align findings with theoretical frameworks. |
| 19/08/2024 | D | <ul style="list-style-type: none"> ❖ Integrated additional literature into the critical evaluation section. Finalized descriptive analytics, focusing on class-specific trends and the influence of circuit-specific factors on sanction distribution. Worked on creating visualizations to better illustrate key findings. ❖ Integrated literature and finalized analytics based on class and circuit trends. |
| 19/08/2024 | C | <ul style="list-style-type: none"> ❖ Balanced the depth of analysis with the word count constraints ❖ Managed analysis depth and word count. |
| 19/08/2024 | D | <ul style="list-style-type: none"> ❖ Planned to complete the discussion and start drafting the recommendations based on findings. ❖ Planned to draft recommendations. |
| 20/08/2024 | D | <ul style="list-style-type: none"> ❖ Conducted a thorough review of the literature to enhance the theoretical framework. Incorporated additional insights from studies on stakeholder theory and sentiment analysis to support findings. Finalized the diagnostic analytics phase and began drafting the conclusion. ❖ Enhanced the theoretical framework with literature. |
| 20/08/2024 | C | <ul style="list-style-type: none"> ❖ Ensured the robustness of the theoretical framework in light of the new data. |

| Date | Code | My Diary Entries |
|------------|------|---|
| 21/08/2024 | D | <div><div>❖</div>Finalized the recommendations section, ensuring they were specific, measurable, achievable, relevant, and time-bound. Incorporated feedback from collaborators on the practicality of the recommendations. Completed the final review of the findings and discussion sections.</div> <div><div>❖</div>Finalized recommendations and incorporated feedback.</div> |

Code for the Development Diary

10 Development Diary Reference Notes

The Development Diary is an essential element of this project, functioning as a personal chronicle of your journey. It is designed to capture pivotal moments, critical decisions, fundamental assumptions, and any challenges encountered along the way. This diary is your space to meticulously record every assumption you make, every decision and choice along with their justifications and reasoning, every obstacle faced, and how you navigated or circumvented it. Regular updates are crucial for tracking the development of your project and your analytical thought process. This consistent documentation provides a valuable narrative for learning and development throughout the project. This diary will not only illustrate your progress but also act as a crucial reference when compiling your final report.

Below is a table outlining the types of entries you might make, with clear explanations and relevant examples to guide you in chronicling your project experience.

| Code | Definition | Examples |
|------|--|--|
| A | Assumption: A belief or statement taken to be true with limited evidence. | <ul style="list-style-type: none">- Assuming a certain customer demographic will respond positively to a new product based on market trends.- Presuming that historical sales data can predict future trends. |
| D | Decision: A conclusion or resolution reached after consideration. | <ul style="list-style-type: none">- Opting to use a linear regression model based on the data's characteristics.- Choosing to focus on a particular dataset after evaluating its relevance and completeness. |
| C | Challenge: Any obstacle or issue encountered during the project. | <ul style="list-style-type: none">- Encountering missing data and deciding to use imputation techniques.- Dealing with unexpected results in data analysis and revisiting the methodology to ensure accuracy. |
| M | Miscellaneous: Any other observations or experiences not covered above. | <ul style="list-style-type: none">- Noting a particularly useful resource or piece of advice.- Reflecting on a personal learning moment or a change in perspective during the project. |

Remember that the more detailed and consistent your entries are, the more valuable the Development Diary will be as a resource for understanding and improving your project approach. Your diary entries are likely to be simple notes written in the first person. These entries are only for your own use and should not be repeated verbatim when writing your report.

Appendix C: Declaration of AI Tool Usage

This declaration must be completed if you have used any AI tool in the completion of this assignment. This includes Quillbot, Grammarly, Paperpal, Elicit, Scispace, ChatGPT, Claude, Perplexity or any other tools of this kind.

There are many ethical and acceptable uses of AI tools that have been explained to you during lectures and workshops. If you have any doubt about your use of AI, then please contact your tutor to seek advice.

There is, however, **one significant 'unethical' use** that can result in an academic misconduct investigation. You must not use AI-generated text, whether this is a single phrase, sentence, or paragraph, presented as your own work in your submission. This also includes allowing any AI tool (e.g. Grammarly) to rephrase of your work, because it will no longer be your work.

| AI Tool | Subscription (YES/NO) | How and for What purpose was this tool used? |
|----------|--------------------------|---|
| CHAT GPT | NO | Guidance in the Analysis |
| | | Summarizing |
| | | Python coding |
| | | Asking all the whys and the rationale thereof |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

Appendix D: Supplements

