Public Sentiment Analysis of Distance-Based Fare Pricing in Rwanda

1. Introduction

In response to growing concerns over equity and sustainability in public transport, Rwanda recently transitioned from a flat-rate fare model to a distance-based fare pricing system. This policy change sparked a wave of public discussion across various platforms, including social media, online forums, and comment sections. The goal of this project is to design a data-driven solution that helps policymakers understand how citizens perceive and react to this new system. Through sentiment analysis and public engagement insights, Iaim to surface key trends, concerns, and misinformation.

2. Data Sources

To gain a holistic view of public sentiment, I aggregated data from multiple platforms:

- **Twitter/X screenshots**: Collected manually from posts dated around the implementation of the new fare system. Optical Character Recognition (OCR) was used to extract text from screenshots.
- **YouTube comments**: Extracted from a key YouTube video discussing the fare changes. Iutilized Google Cloud's YouTube Data API and Natural Language API for comment retrieval and sentiment tagging.
- **News reactions and user testimonials**: Additional qualitative feedback was captured from news article comment sections and public testimonials shared on local forums.

3. Data Preprocessing

- **Twitter/X Data**: OCR was applied using Tesseract to convert image text to machine-readable format. The resulting text was cleaned (removal of special characters, emojis, duplicates) and timestamped manually.
- **YouTube Comments**: Google Cloud NLP was used to assign sentiment scores ranging from -1 (negative) to +1 (positive). Comments were cleaned for analysis and categorized as positive, neutral, or negative based on score thresholds.

4. Sentiment Analysis & Visualization

Using Looker Studio, I visualized sentiment trends across the data sources. Key metrics include:

- Overall sentiment distribution (positive, neutral, negative)

- Sentiment trends over time (daily and weekly)
- Most frequently mentioned topics and keywords
- Comparative sentiment between platforms (Twitter vs YouTube)

Findings from YouTube:

- YouTube had a slightly more negative tone than Twitter.
- Frequent concerns were about fare fairness, accessibility for rural users, and unclear communication about implementation.

5. Key Findings

- **Mixed Reception**: While some users appreciate the fairness of paying based on distance, others expressed frustration over delays, poor communication, and technical limitations.
- **Implementation Concerns**: Many users were concerned about the need to tap cards multiple times, which may delay trips and cause crowding.
- **Misinformation and Misunderstanding**: Some users misunderstood how the system calculates fares, indicating a need for better public education.
- **Suggestions for Improvement**: Users proposed practical solutions such as adding more fare machines, improving the interface, and allowing smoother exits from buses.

6. Challenges Encountered

During the project, several challenges were faced:

- **Twitter Scraping Limitations**: Due to Twitter API access restrictions and limited scraping options, I could not programmatically retrieve real-time or historical tweets. Instead, I relied on screenshots and manual OCR.
- **OCR Inaccuracy**: Extracting text from Twitter/X screenshots using OCR sometimes resulted in errors or incomplete data, requiring manual correction and cleaning.
- **Time Constraints**: Given the breadth of data sources and the time needed for cleaning and formatting, only a subset of comments could be processed.
- **Platform Bias**: The volume and nature of feedback differs by platform, which may skew results toward the more expressive or reactive audiences on YouTube or Twitter.

7. Recommendations

Based on the analysis, I propose the following:

1. **Improve Public Communication**: Launch a campaign explaining how the distance-based fare system works and how users can use it efficiently.
2. **Enhance Infrastructure**: Consider increasing the number of card scanners and placing them at more strategic positions.

3. **Monitor Real-Time Feedback**: Implement a digital feedback channel (e.g., chatbot, survey link) for passengers to report issues quickly.
4. **Use Multi-Source Analysis**: Combine insights from social media, YouTube, and public feedback regularly to adjust implementation policies.

8. Conclusion

This project illustrates the power of public sentiment analysis in shaping inclusive transport policies. By integrating diverse data sources and tools like Google Cloud NLP and Looker Studio, I provided a multifaceted understanding of how Rwandans are responding to the distance-based fare policy. Ongoing sentiment monitoring and community engagement will be key to refining the system and ensuring its success.