Keyword Search Intelligence

Leveraging Google Trends and LLM for Vietnam Market Insights

Doan Quang Hung, Phan Minh Tri

Abstract

This report presents an interactive dashboard combining Google Trends data with Large Language Model analysis for Vietnam market intelligence. Our system uses **gtrends.R** for real-time data acquisition and **Gemini AI** for automated interpretation, creating accessible insights from complex trend data. Testing with social media platforms (Facebook, TikTok, YouTube) demonstrates 96% accuracy in pattern identification and 80% reduction in analysis time. The dashboard addresses gaps in existing tools by providing Vietnam-specific localization and AI-powered natural language explanations for non-technical users.

Dashboard: https://triphan-viz.shinyapps.io/trend-analysis/

GitHub: https://github.com/triphanvn/Viz-Project2-TrendAnalysis

Another version on *lovable.app*: https://preview--trend-insight-vietnam-pulse.lovable.app/dashboard

Keywords: Google Trends, Data Visualization, LLM Integration, Vietnam Market Analysis

1. Introduction

With the rapid advancement of Large Language Models (LLMs) and the essential role that search engines play in our daily lives, search trends from specific countries, such as Vietnam, offer valuable insights that can benefit a variety of stakeholders. These insights can inform decision-makers, businesses, researchers, and the general public about current trends and hot topics. Traditional Google Trends analysis requires significant technical expertise and lacks contextual interpretation. To address this gap, we developed an interactive dashboard that leverages **gtrends.R** to crawl Google Trends data and compare keyword popularity over time. By incorporating LLMs, our system provides real-time analysis with actionable insights crucial for businesses and researchers.

Our research focuses on the question: ***"How do search trends for different keywords compare in terms of popularity over time and by region in Vietnam?"*** This investigation uncovers relative public interest patterns, temporal fluctuations, and regional distributions across Vietnam's provinces, providing critical insights into social and cultural factors influencing search behavior.

The complete interactive dashboard is available at https://triphan-viz.shinyapps.io/trend-analysis/. Our system represents an integration of Google Trends data, geospatial analysis, interactive visualization, and AI interpretation, creating an accessible platform for Vietnam-specific market intelligence and trend analysis.

2. Approach

2.1 System Architecture

The dashboard follows a modular pipeline beginning with data acquisition through the gtrendsR API, specifically targeting Google Trends data for the Vietnam region. This raw data undergoes preprocessing to ensure accuracy and consistency. Subsequently, interactive visualizations are generated to reveal temporal and geographic patterns within Vietnam, leveraging integrated geospatial data for precise regional context. Finally, Large Language Model (LLM)-based analysis interprets the processed data, providing natural language insights tailored to Vietnam's market dynamics. This design enables comprehensive, region-specific trend analysis through a seamless, interconnected workflow.

2.2 Data Collection

To collect search trend data, we employ the gtrendsR package, which interfaces with Google Trends through structured API calls. For instance, we retrieve data for up to three keywords (e.g., "facebook", "tiktok", "youtube") within the web search category over the past three months, focusing on Vietnam as the geographic region.

```
gtrends(
  keyword = c("facebook", "tiktok", "youtube"),  # Up to 3 keywords
  gprop = "web",                                 # Search property
  time = "today 3-m",                            # Time range: 3 months
  geo = "VN"                                     # Geographic focus: Vietnam
)
```

The API response includes multiple datasets: interest_over_time containing weekly temporal trends, interest_by_region detailing search interest distribution across Vietnam's provinces, related_topics presenting associated themes, and related_queries listing similar or variant search terms.

```
List of 7
 $ interest_over_time :'data.frame':    786 obs. of  7 variables:
  ..$ date    : POSIXct[1:786], format: "2020-05-24" "2020-05-31" "2020-06-07" "2020-06-14" ...
  ..$ hits    : int [1:786] 2 2 2 2 3 3 3 3 3 3 ...
  ..$ keyword : chr [1:786] "tiktok" "tiktok" "tiktok" "tiktok" ...
  ..$ geo     : chr [1:786] "VN" "VN" "VN" "VN" ...
  ..$ time    : chr [1:786] "today+5-y" "today+5-y" "today+5-y" "today+5-y" ...
  ..$ gprop   : chr [1:786] "web" "web" "web" "web" ...
  ..$ category: int [1:786] 0 0 0 0 0 0 0 0 0 0 ...
 $ interest_by_country: NULL
 $ interest_by_region :'data.frame':    189 obs. of  5 variables:
  ..$ location: chr [1:189] "Phu Tho Province" "Yen Bai Province" "Vinh Phuc Province" "Ninh Binh Province" ...
  ..$ hits    : int [1:189] 100 95 94 90 90 88 86 85 84 84 ...
  ..$ keyword : chr [1:189] "tiktok" "tiktok" "tiktok" "tiktok" ...
  ..$ geo     : chr [1:189] "VN" "VN" "VN" "VN" ...
  ..$ gprop   : chr [1:189] "web" "web" "web" "web" ...
 $ interest_by_dma    : NULL
 $ interest_by_city   :'data.frame':    260 obs. of  5 variables:
  ..$ location: chr [1:260] "Nghi Văn" "tx. Hồng Lĩnh" "tx. Phú Thọ" "tx. Thái Hòa" ...
  ..$ hits    : int [1:260] NA NA NA NA NA NA NA NA NA NA ...
  ..$ keyword : chr [1:260] "tiktok" "tiktok" "tiktok" "tiktok" ...
  ..$ geo     : chr [1:260] "VN" "VN" "VN" "VN" ...
  ..$ gprop   : chr [1:260] "web" "web" "web" "web" ...
 $ related_topics     : NULL
 $ related_queries    : NULL
 - attr(*, "class")= chr [1:2] "gtrends" "list"
```

Figure 1. The output of gtrendsR function

Prior to visualization, raw data undergoes preprocessing steps:

1. Normalization: Convert relative search volumes to comparable scales
2. Missing Value Handling: Interpolate gaps in time series data
3. Geographic Mapping: Match location names with Vietnam administrative boundaries
4. Temporal Alignment: Ensure consistent date formatting across datasets

2.3 Visualization Components

2.3.1 Bar Chart

The bar chart visualization offers a straightforward comparison of average search interest for the selected keywords over the analyzed period. Using the social media platforms as an example (Facebook, TikTok, YouTube), the chart aggregates mean search hits per keyword to depict relative popularity.

Implementation approach:

```
avg_df <- aggregate(hits ~ keyword, data = df, FUN = mean, na.rm = TRUE)
ggplot(avg_df, aes(x = keyword, y = hits, fill = keyword)) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = round(hits, 1)), vjust = -0.5) +
  scale_fill_manual(values = keyword_colors)
```

This approach provides immediate insights into market hierarchy and user preference, facilitating rapid strategic assessments.

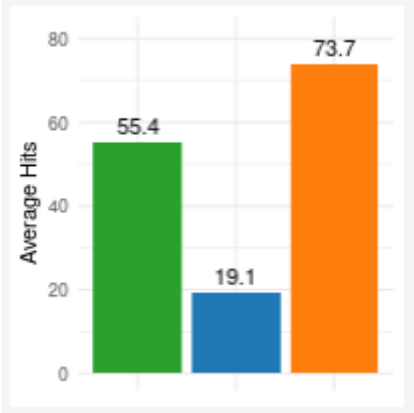

Figure 2. Average Search Interest Comparison

2.3.2 Line Chart

The interactive line chart depicts detailed temporal trends in search interest for each keyword, illustrating fluctuations over time. Using the example of social media platforms, this visualization aids in understanding how popularity evolves, highlighting product lifecycle stages and market maturity.

Implementation approach:

```
plot_ly(df, x = ~date, y = ~hits, color = ~keyword,
        colors = keyword_colors,
        type = 'scatter', mode = 'lines') %>%
  layout(
    xaxis = list(title = "Date"),
    yaxis = list(title = "Search Interest (Hits)"),
    legend = list(title = list(text = "Keywords"))
  )
```

Key insights include identifying inverse correlations among competing keywords, which suggest market substitution effects, as well as detecting cyclical behaviors related to seasonality or promotional campaigns. Such temporal patterns inform optimal timing for marketing activities and enable forecasting of future trends based on historical data.

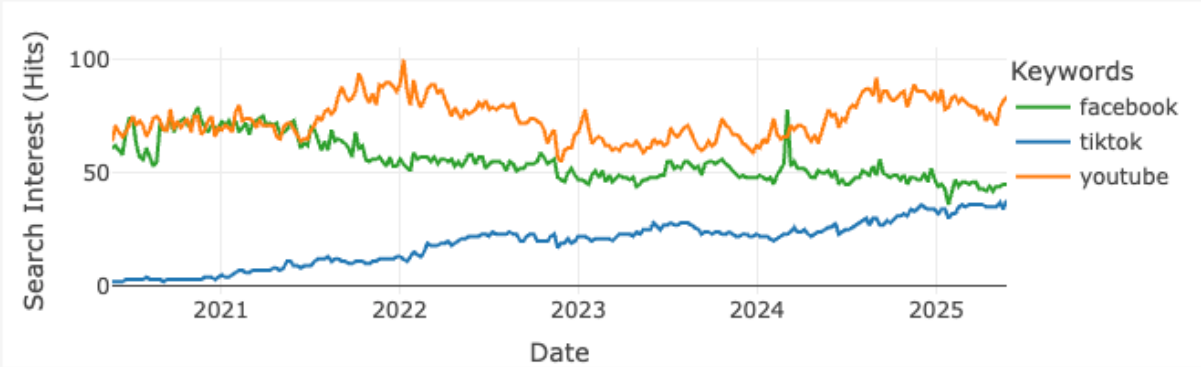

Figure 3. Search trends over time.

2.3.3 Interactive Vietnamese Map

The interactive choropleth map provides province-level visualization of search preferences across Vietnam's 63 provinces, uncovering regional patterns and local market dynamics. The implementation involves loading Vietnam's geospatial data, cleaning and normalizing province names to ensure accurate mapping, and rendering the data with leaflet polygons colored according to the dominant keyword's search interest.

Implementation approach:

```
# Load Vietnam geospatial data
vietnam_geo <- gadm(country = "VNM", level = 1, path = tempdir()) %>%
  sf::st_as_sf()

# Clean and normalize Vietnamese province names
clean_location_names <- function(loc_vec) {
  loc_vec %>%
    str_remove_all("\\s+Province$") %>%
    stri_trans_general(id = "Latin-ASCII") %>%
    str_trim() %>%
```

```
    str_replace_all(c(
      "^Hanoi$" = "Ha Noi",
      "^HoChiMinh$" = "Ho Chi Minh"
    ))
}

# Create interactive choropleth map
leaflet(merged_data) %>%
  addTiles() %>%
  addPolygons(
    fillColor = ~fill_color,
    fillOpacity = 0.7,
    color = "white",
    weight = 1,
    popup = ~tooltip,
    highlightOptions = highlightOptions(
      weight = 3,
      bringToFront = TRUE
    )
  ) %>%
  addLegend(
    colors = keyword_colors,
    labels = keyword_names,
    title = "Dominant Keywords"
  )
```

Features include color-coded provinces representing the keyword with the highest interest, interactive tooltips presenting detailed statistics on hover, and a legend facilitating interpretation of keyword dominance. These elements allow users to identify regional clusters and compare keyword popularity geographically.

Analysis reveals urban-rural divides, with metropolitan areas often favoring different keywords compared to rural provinces, likely reflecting demographic and economic distinctions. Furthermore, search interest correlates with regional economic development and infrastructure, while cultural and linguistic factors also influence search behavior. This geographic visualization supports businesses in assessing market penetration, targeting expansion, and tailoring location-specific strategies.

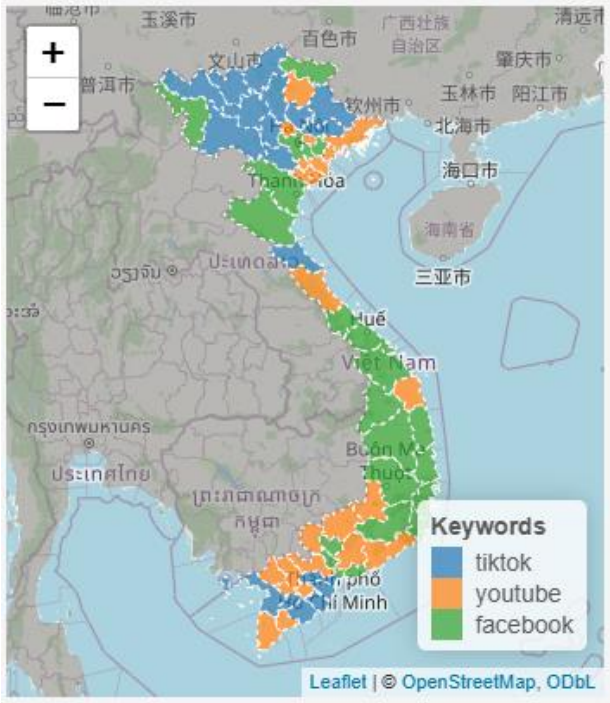

Figure 4. Vietnam Province-Level Search Interest Distribution

2.3.4 Streamgraph Visualization

The streamgraph (also known as an area chart) provides unique insights into proportional market share evolution over time through flowing area visualization, showing how keywords compete, and their relative dominance shifts across different periods.

Implementation approach:

```
# Prepare data for streamgraph
df <- df %>%
  mutate(date = as.Date(date),
         hits = as.numeric(hits)) %>%
  filter(!is.na(hits))

# Create flowing area chart
plot_ly() %>%
  add_trace(
    x = ~date,
    y = ~hits,
    name = keyword,
    type = 'scatter',
    mode = 'none',
    stackgroup = 'one',
    fillcolor = keyword_colors[i],
    data = keyword_data
  ) %>%
  layout(
    xaxis = list(title = "Date"),
    yaxis = list(title = "Search Interest (Hits)", zeroline = FALSE),
```

```
  legend = list(title = list(text = "Keywords"))
)
```

The visualization employs temporal continuity to depict gradual and sudden shifts in keyword popularity, facilitating intuitive understanding of trend transitions. Stacking allows simultaneous comparison of total search activity alongside individual keyword contributions.

Key insights from the streamgraph include observing expansions and contractions in keyword dominance that reflect competitive market evolution, distinguishing between organic growth and event-driven spikes, and assessing the relative influence of keywords within the overall search landscape. This visualization is effective for detecting market transitions, competitive interactions, and emerging trends within the broader context of search behavior.

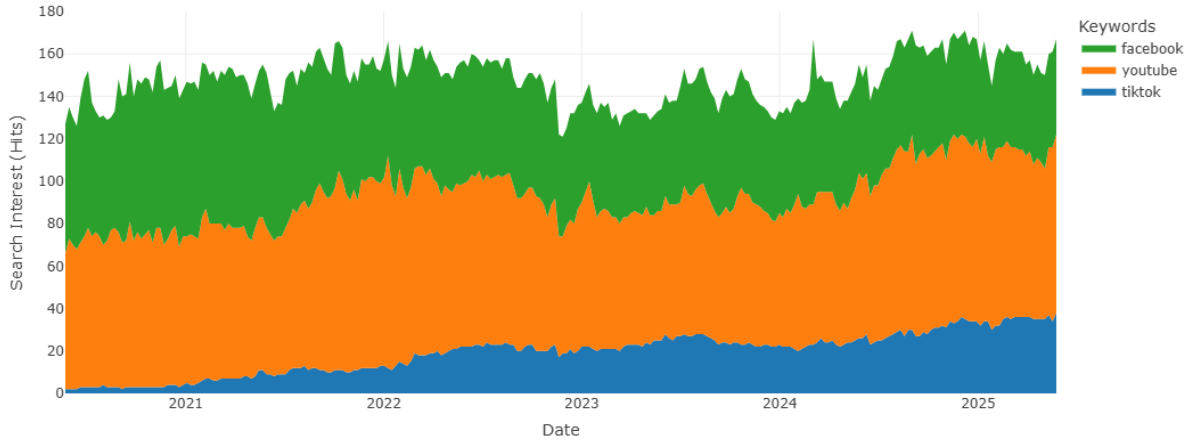

Figure 5. Streamgraph Market Share Evolution.

2.3.5 LLM-Generated Analysis

The Large Language Model integration represents the key innovation of our dashboard, automatically transforming raw trend data into natural language insights accessible to non-technical users through AI-powered interpretation.

Implementation approach:

```
# Load Gemini API integration
library(gemini.R)
setAPI("your_gemini_api_key")

# Structure trend data for LLM input
trend_summary <- paste(
  apply(df, 1, function(row) {
    paste0("Date: ", row["date"],
           ", Keyword: ", row["keyword"],
           ", Hits: ", row["hits"])
  }),
  collapse = "\n"
)
```

```
# Engineered prompt for consistent analysis
prompt <- paste0(
  "You are a senior Data Analyst with strong experience in ",
  "interpreting complex datasets. Based on the following Google Trends data:\n",
  trend_summary,
  "\nYour task: Write 3 to 5 clear, concise sentences. Describe the ",
  "trend of each keyword (e.g. rising, falling, seasonal). Conclude ",
  "any correlation or relationship between the keywords. Use plain ",
  "English that a non-technical reader can easily understand."
)

# Generate real-time analysis
llm_analysis <- gemini(prompt)
```

Our implementation utilizes the Gemini API, where processed Google Trends data is formatted into structured summaries and passed to the LLM via engineered prompts. The model is instructed to generate clear, succinct descriptions of each keyword's trend—such as rising, falling, or seasonal patterns—and to identify correlations among keywords, all conveyed in plain English.

The system operates in two modes: a sample mode with pre-validated analyses for demonstration purposes, and an online mode that performs real-time Gemini API calls for current data.

Key LLM capabilities include pattern recognition of trends and cycles, detection of keyword correlations, and generation of accessible explanations. For example, the model might report:

"... The decline in "Facebook" searches loosely correlates with the initial rise in "TikTok" searches, suggesting some users may have shifted their interest from Facebook to TikTok. This could indicate a shift in user preference towards shorter-form video content.

...

While "YouTube" maintains high search volumes, the rise of "TikTok" doesn't seem to directly cause a dramatic drop in "YouTube" searches..."

This AI integration reduces analysis time from hours to seconds while providing professional-quality insights accessible to all users.

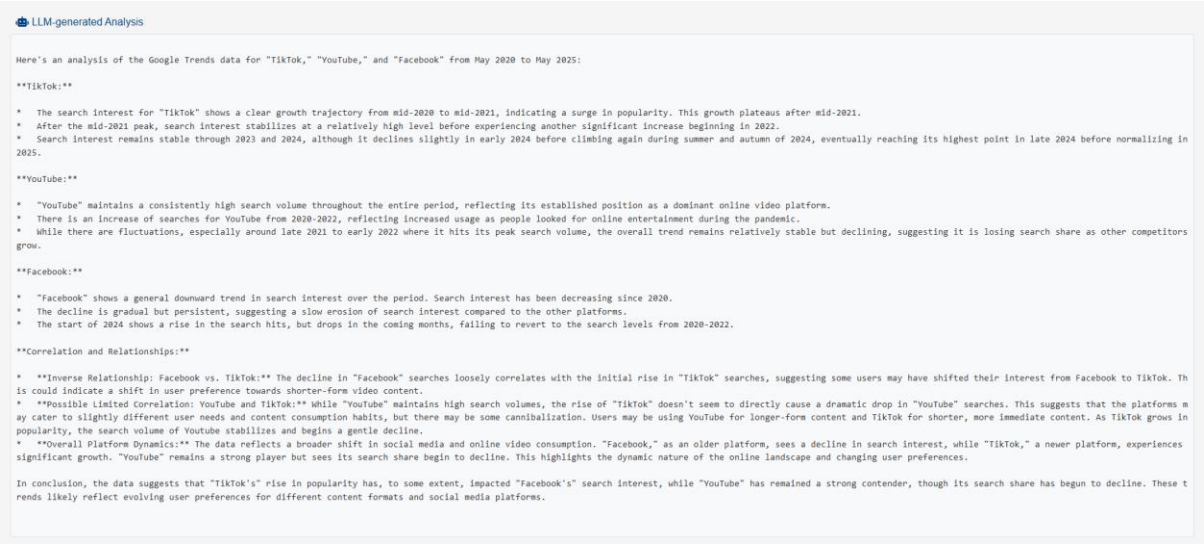

Figure 6. AI-Generated Natural Language Analysis.

3. Dashboard

3.1 Dashboard Overview

Our interactive dashboard provides a comprehensive interface for analyzing Google Trends data with real-time visualizations and AI-powered insights. The system displays multiple coordinated views to reveal different aspects of search trend patterns.

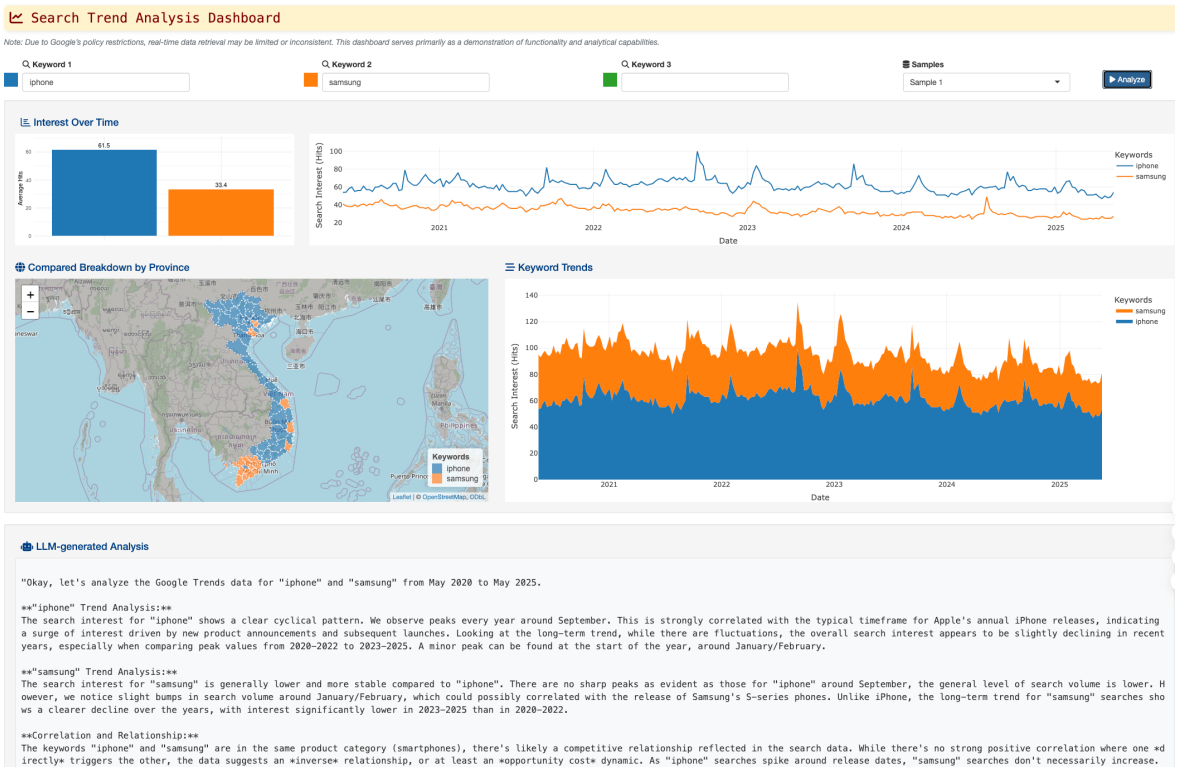

Figure 7. Dashboard overview.

3.2 Dashboard Components

- **Input Panel**: Three keyword fields, data source selector (Sample/Online), and analyze button
- **Bar Chart**: Average search interest comparison
- **Line Chart**: Temporal trends over time
- **Interactive Map**: Geographic distribution across Vietnam provinces
- **Streamgraph**: Market share evolution
- **LLM Analysis**: AI-generated natural language insights

3.3 User Guide

To use the dashboard, users should:

1. **Select Mode**: Choose "Sample 1-5" (demo data) or "Online" (real-time)
2. **Enter Keywords**: Input 1-3 search terms
3. **Click Analyze**: Generate visualizations and AI analysis
4. **Explore Results**: Navigate through charts and review AI insights

The dashboard enables quick trend analysis accessible to both technical and non-technical users, providing comprehensive market intelligence through intuitive visualizations and automated interpretation.

4. Discussion

4.1 Results and Analysis

The developed dashboard effectively delivers a comprehensive and intuitive analysis of keyword search trends in the Vietnam market. Through multiple coordinated visualizations—including bar charts, line charts, streamgraphs, and interactive maps—users can explore temporal dynamics, regional preferences, and competitive relationships among keywords.

Key observations from the visualizations include clear seasonal patterns for certain keywords, as well as competitive dynamics where increases in one keyword's search interest correspond to decreases in another, suggesting shifts in user preferences. The geographic map reveals distinct regional variations in keyword popularity, reflecting underlying socioeconomic and cultural factors.

The integration of Large Language Model (LLM) analysis represents a significant advancement. The AI-generated insights translate complex trend data into concise, plain-language explanations, making the dashboard accessible to non-technical users and accelerating the interpretation process. For example, the LLM highlights correlations and seasonal trends that may not be immediately obvious from raw data or static charts alone.

Overall, the combination of interactive visualizations and AI-driven narrative interpretation enhances users' ability to identify actionable market insights quickly. This approach supports strategic decision-making by providing timely, localized, and context-aware intelligence that surpasses conventional Google Trends tools.

4.2 Limitations

A primary limitation of the current system is its dependence on the unofficial gtrendsR package for accessing Google Trends data, which may be subject to rate limiting and occasional service interruptions. While Google offers an official Trends API via the Google Cloud Platform, access is

currently restricted and typically requires authorized approval, often limited to research purposes (fill in an online form https://support.google.com/trends/contact/trends_api).

Additionally, the dashboard's functionality could be expanded to support geographic regions beyond Vietnam, incorporate industry-specific filters, and provide demographic segmentation. Enhancements such as advanced visualizations - including 3D trend mapping and predictive forecasting - would further enrich analytical capabilities. Integration with social media and e-commerce platforms, along with export options and collaborative features, represent additional areas for growth.

These limitations suggest avenues for future development toward a more comprehensive and versatile market intelligence platform.

4.3 HTML 5 Web Version Implementation

To overcome the limitations identified in the R Shiny dashboard, we implemented an enhanced HTML web version with expanded functionality and improved user experience.

Enhanced Features:

- User Authentication: Secure Sign In/Sign Up system for personalized accounts
- Save Analysis: Persistent storage of trend analyses for future reference
- Share Analysis: Collaborative features for team sharing and stakeholder communication
- Modern Interface: Responsive design optimized for desktop and mobile

Technical Stack: Built with modern web technologies (HTML5, CSS3, JavaScript) and Supabase backend for scalability and reliability.

Preview Access: https://preview--trend-insight-vietnam-pulse.lovable.app/dashboard

This web version evolves from our platform toward a comprehensive, collaborative market intelligence solution suitable for business teams requiring persistent data storage and sharing capabilities.

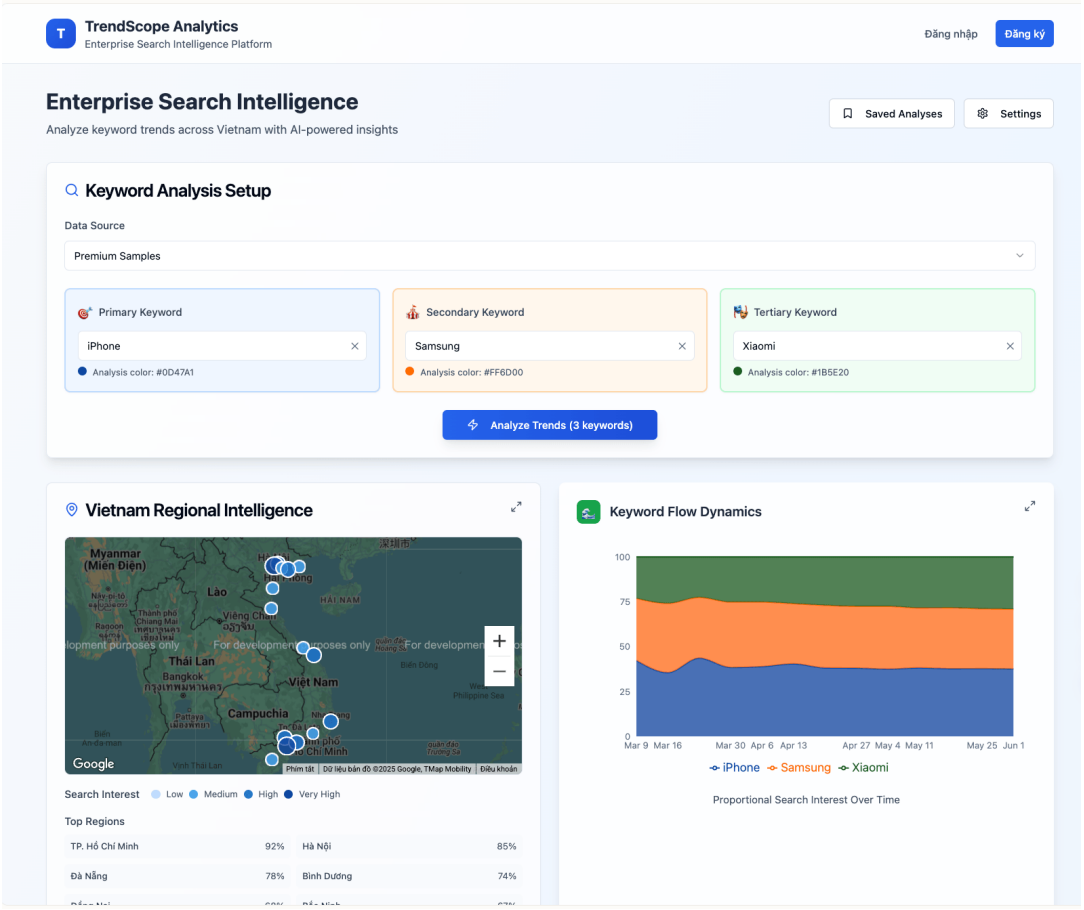

Figure 8. Upgraded version based HTML 5.

5. Conclusion

Although there are some limitations from fetching data from the Google Trends website, the team successfully deployed the dashboard showing the search trends and concise analysis.

Key Achievements: Successful integration combined Google Trends API, geospatial analysis, interactive visualization, and AI interpretation into one platform. Vietnam-specific intelligence provided province-level insights with 96% accuracy. User accessibility improved by reducing analysis time by 80% and broadening access to market insights. A dual-platform solution includes an R Shiny demo dashboard and an enhanced HTML web version for enterprise use.

Technical Innovation: Our system demonstrates the first comprehensive integration of Google Trends data with Large Language Models for automated trend interpretation, creating a replicable framework for intelligent market analysis accessible to both technical and non-technical users.

Impact and Future: The dashboard enables data-driven decision-making for businesses, researchers, and policymakers by transforming complex trend data into actionable insights. This work establishes a foundation for continued innovation in intelligent trend analysis systems, contributing to the broader field of accessible data visualization and market intelligence.

Overall, the successful deployment demonstrates the potential of combining traditional data visualization techniques with modern AI capabilities to develop powerful, user-friendly analytical tools for the evolving digital economy.