

Test Task: Viral Content Analysis

GitHub Repository Link

[Maidanchuklrina/Virus_Content_Analysis](https://github.com/Maidanchuklrina/Virus_Content_Analysis)

Introduction

This report summarizes the results of an analysis of viral content in social media (Instagram, YouTube, Facebook, TikTok). The research focuses on identifying key factors influencing post popularity: likes, comments, views, shares, hashtags, and posting time.

Data Overview

The dataset contains information on social media posts, including:

- Post text
- Number of likes, comments, and shares
- Platform (Instagram, Facebook, TikTok, YouTube)
- Posting time
- Used hashtags

Analysis Methods

1 Data Preprocessing:

- Removing incorrect records
- Handling missing values
- Standardizing platform names

2 Viral Content Analysis:

- Identifying the top 5% viral posts
- Comparing activity across different platforms
- Analyzing the distribution of likes, comments, and views

3 Hashtag Analysis:

- Identifying the most frequently used hashtags
- Evaluating the connect of hashtags on engagement
- Correlation analysis of hashtag popularity

4 Posting Time Analysis:


- Identifying the optimal posting time
- Visualizing user engagement trends

Key Findings


- **YouTube and Instagram have the highest proportion of viral content.** This suggests the effectiveness of their recommendation algorithms.
- **Optimal posting time:** The highest user engagement occurs in the **evening hours (21:00-22:00) and afternoon (12:00-13:00)**.
- **Hashtags reflect common sentiments of people.** The most popular ones are thematic and emotionally charged hashtags.
- **Religious content has high engagement.** It often includes key terms that encourage interaction.

Some Visualizations


Distribution of Likes and Viral Posts by Platforms

 **Conclusion:** Viral posts exhibit a long tail with high engagement levels.




Impact of Posting Time on Virality

 **Conclusion:** Evening and early morning hours have the highest viral potential.

Top 20 Popular Hashtags

 **Conclusion:** Hashtags with religious and emotional context show high engagement.


Future Analysis Directions

- 1  **Categorizing posts by topic.** Using clustering or classification techniques to identify the most viral themes. Analyzing the relationship between topic and engagement metrics (likes, comments, shares).
- 2  **Impact of multimedia content.** Examining the dependency between the length of the post and engagement levels. Evaluating which types of content perform best on different platforms.
- 3  **Virality dynamics over time.** Analyzing how viral posts gain popularity over a certain period. Identifying trends and long-term viral effects.

Additional Materials

- `Virus_Content_Analysis.ipynb` – Jupyter Notebook with analysis code
- `README.md` – Project documentation
- `Test Task Report Tremementum.pdf` – PDF report with key findings

 **Author:** Maidanchuk Iryna

 **Date:** 25.01.2025

