SVKM'S NMIMS Deemed to be UNIVERSITY

MUKESH PATEL SCHOOL OF TECHNOLOGY MANAGEMENT & ENGINEERING

Technical Internship Program 2024

Project Proposal

Social Media Data Analyst Intern

Name: Darshita Anyawada
Roll No: N225
MBA Tech Computer Engineering
SAP ID: 70472100184
Date: 01/06/2024

About the Company:

YouVah stands as India's pioneering internship platform exclusively designed for teenagers. Its mission is to empower high school students across the nation by providing them with early career exposure through internship experiences tailored to their interests and aspirations.

With a unique approach, YouVah enables teenagers to engage in real-world work that aligns with corporate expectations, allowing them to gain valuable insights and skills essential for their future careers. All internships offered by YouVah are co-created with leading companies, ensuring that students receive hands-on training in relevant fields.

At YouVah, the significance of mentorship in guiding students through their internship journey is deeply understood. The platform provides access to a diverse guild of industry mentors who offer expert advice and domain guidance, ensuring that every mentern (mentor-intern) receives personalized support and encouragement.

Beyond internships, YouVah fosters a vibrant community where teenagers can connect with like-minded peers who share their interests. Students can build meaningful friendships and collaborate on projects through the platform, creating a supportive environment for growth and exploration.

Trusted and backed by prestigious institutions such as IIM Ahmedabad and IIM Bangalore, YouVah is committed to impacting teenagers' lives across India. Its vision is to create a one-stop solution for students seeking to explore unconventional career paths and shape their futures from an early age.

Abstract:

This report outlines the progress and methodology of a project undertaken during my tenure as a Social Media Data Analyst Intern at YouVah, focusing on data scraping and analysis to optimize content strategy on Instagram. The project aimed to extract valuable insights by scraping Instagram data using tools like Selenium, Requests, and Instaloader, followed by in-depth analysis utilizing Pandas, Matplotlib, and Seaborn.

The primary objective was to decipher viewer engagement patterns, assess the effectiveness of different content types, and identify optimal posting times. Through meticulous data collection and analysis, key findings emerged, notably that informative reels outperform trendy ones regarding viewer engagement. Additionally, the study revealed specific hashtags that significantly amplify viewership.

These insights offer actionable recommendations for refining YouVah's content strategy on Instagram, enabling the platform to tailor its approach to maximize engagement and reach. By leveraging data-driven insights, YouVah can enhance its presence on Instagram and foster deeper connections with its audience, ultimately advancing its social media objectives.

Introduction:

In the rapidly evolving social media landscape, understanding content engagement dynamics is crucial for optimizing reach and impact. YouVah seeks to enhance its Instagram content strategy through data-driven insights. This report aims to dissect YouVah's Instagram performance metrics, analyzing various factors influencing viewer engagement. We specifically focus on the performance of trendy vs. informative reels, the impact of hashtags, and optimal posting times. These insights will help YouVah tailor its content to maximize engagement and bolster its Instagram presence.

Problem Statement:

The primary challenge in this project was to scrape and analyze Instagram data to uncover patterns and insights that can inform YouVah's content strategy. This involved:

1. Data Extraction: Securely scraping Instagram data, managing API limitations, and handling dynamic web content.

2. Data Cleaning: Converting scraped data into a usable format, handling missing values, and ensuring data integrity.

3. Data Analysis: Identifying trends in viewer engagement, analyzing the effectiveness of content types, and determining optimal posting schedules.

4. Visualization and Reporting: Creating intuitive visualizations and compiling the findings into actionable recommendations.

Proposed Solution:

To address these challenges, the project leveraged a combination of web scraping and data analysis tools:

1. Data Collection:
   - Selenium: Automated browser interactions to scrape dynamic Instagram content.
   - Requests and Instaloader: Extracted data such as likes, views, captions, and hashtags.
   - XPath and Regular Expressions: Located specific elements on web pages and extracted relevant data points.

2. Data Cleaning:
   - Pandas: Processed and cleaned the data, handling missing values and converting data types.
   - Regular Expressions: Extracted hashtags from captions for further analysis.

3. Data Analysis:
   - Pandas, Matplotlib, Seaborn: Conducted exploratory data analysis (EDA), visualizing relationships between metrics and deriving insights.
   - Engagement Rate Calculation: Calculated engagement rates based on likes and views to gauge audience interaction.

4. Visualization and Reporting:
   - Tableau: Created interactive visualizations for deeper insights.
   - Documentation: Summarized findings and provided recommendations for optimizing YouVah's Instagram content strategy.

Implementation:

1. Data Collection

Automated Data Collection:

- Selenium WebDriver: Automated the extraction of metrics such as likes, views, captions, and hashtags from Instagram posts.
- XPath and Regular Expressions: Located specific elements and extracted relevant data points.

  Metrics Extracted: Likes, views, captions, hashtags.

Manual Data Collection:
Supplemented automated extraction with manually collected data for comprehensive analysis.

2. Data Cleaning

Data Import and Preprocessing:

- Imported data into a Pandas DataFrame.
- Handled missing values and converted numeric data for analysis.
- Mapped categorical variables to numerical values for quantitative analysis.
- Extracted hashtags using regular expressions.

Data Integrity Checks:
Ensured data consistency and validated transformations for accuracy.

3. Data Analysis

- Exploratory Data Analysis (EDA):
  - Scatter Plot: Visualized the relationship between likes and views.
  - Bar Plots: Compared average views for trendy vs. informative content and their combinations.
  - TimeSeries Analysis: Identified optimal posting times by analyzing average views by hour and day.

- Hashtag Analysis:
  Identified topperforming and frequently used hashtags.

- Engagement Rate Calculation:
  Calculated the percentage of likes relative to total views.

4. Additional Analysis

Tableau Visualization: Leveraged Tableau for interactive exploratory analysis.

5. Output

Documentation and Presentation: Compiled the cleaned dataset, visualizations, and insights into a comprehensive report with actionable recommendations.

Conclusion:

The analysis provided valuable insights into YouVah's Instagram content performance:

1. Content Strategy Refinement: Prioritize informative reels over trendy ones to enhance engagement.

2. Hashtag Utilization: Use top forming hashtags strategically to increase reach.

3. Optimal Posting Times: Align posting schedules with peak engagement periods to maximize visibility.

4. Continuous Monitoring: Regularly monitor performance metrics and adapt strategies based on emerging trends.

By leveraging these data-driven insights, YouVah can optimize its Instagram presence, drive engagement, and achieve its social media objectives.

References:

1. Selenium: https://www.selenium.dev/
2. Requests: https://docs.pythonrequests.org/en/latest/
3. Instaloader: https://instaloader.github.io/
4. Pandas: https://pandas.pydata.org/
5. Matplotlib: https://matplotlib.org/
6. Seaborn: https://seaborn.pydata.org/
7. Tableau: https://www.tableau.com/
8. Instagram API Documentation: https://developers.facebook.com/docs/instagramapi/
9. YouVah: https://www.youvah.in/