Web Scraping & Data Analysis Report: StoryMirror Authors

# 1. Project Requirements

The objective of this project was to:  
- Extract author and story data from the StoryMirror platform.  
- Collect detailed metrics including:  
 - Author name, bio, total claps.  
 - Top 3 stories per author with title, claps, snippet, and URL.  
- Save the data into a structured Excel file.  
- Analyze and visualize data to understand content performance.

# 2. Solutions & Tools Used

## Web Scraping Solution

- Browser Automation Tool: Selenium  
- Dynamic Content Handling: Scroll & wait logic implemented.  
- Output: Pandas DataFrame exported to Excel.

## Data Analysis & Visualization

- Libraries: pandas, matplotlib, seaborn  
- Visuals Generated:  
 - Bar plots (total claps, top stories, average claps)  
 - Histogram (snippet length)  
 - Scatter plot (snippet length vs claps)  
 - Horizontal bars  
 - Word cloud (optional)

# 3. Code Breakdown

## A. Web Scraping Code

Used Selenium to launch a browser using ChromeDriverManager.  
Scroll handling and element extraction (name, bio, claps, and top 3 stories) implemented.  
Data is stored in a dictionary and saved using pandas to Excel.

## B. Visualization Code

Used seaborn and matplotlib for various charts.  
Data was cleaned and prepared using pandas.  
Charts generated include bar plots, histograms, scatter plots, and optional word cloud.

# 4. Libraries & Their Functions

|  |  |
| --- | --- |
| Library | Function |
| selenium | Browser automation to interact with StoryMirror dynamically |
| pandas | Data handling, transformation, and export to Excel |
| matplotlib | Low-level plotting framework |
| seaborn | High-level statistical plotting on top of matplotlib |
| webdriver\_manager | Automatically manages the correct ChromeDriver version |
| wordcloud (optional) | Visualization of frequently used words in story titles |
|  |  |

# 5. Insights from Charts

🔥 Total Claps by Author: Indicates overall popularity. Some authors had significantly higher engagement.  
📖 Top 10 Stories by Claps: Spotlights individual stories gaining the most traction.  
📈 Average Claps per Story: Highlights content consistency across authors.  
✂️ Snippet Length Distribution: Most snippets range between 100–300 characters.  
💬 Snippet Length vs Claps: Shows correlation between snippet length and engagement.  
🏆 Most Clapped Story per Author: Identifies best-performing content from each author.  
📚 Story Claps (All Stories): Bar graph shows overall engagement spectrum.

# 6. Conclusion

This project successfully scraped structured data from a dynamic content website using Selenium and transformed it into insightful visualizations using Pandas and Seaborn. These insights help understand content strategies, engagement patterns, and optimize future publishing decisions on platforms like StoryMirror.

# 7. Visual Charts

