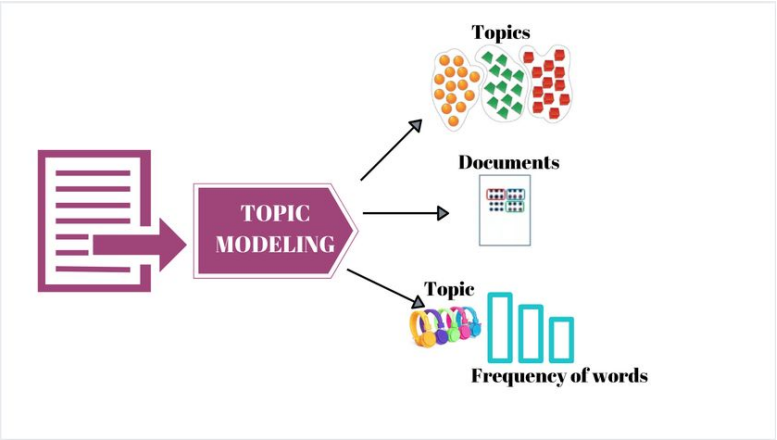


01 August 2023

Exploring Bangla News Trends: A Comprehensive Topic Modeling Analysis on The Daily Star Bangla News Portal

https://colab.research.google.com/drive/1DJN8ja2G3BqyL5LLai2E_Gw9Jd5y2pDo?usp=sharing

Kazi Shadman Sakib
4th Year Undergraduate Student

Department of Computer Science & Engineering
University of Dhaka

Quick Summary

Conducted an in-depth analysis of the Daily Star Bangla news portal, aiming to gain insights into the most frequent topics discussed over a span of seven days. Leveraging the **Latent Dirichlet Allocation (LDA)** technique, performed topic modeling on three distinct datasets: **daily news**, **peak user hour news (9 PM - 11 PM)**, and **off-peak user hour news (4 AM - 6 AM)**. Additionally, merged the data from all seven days to perform topic modeling on a weekly basis.

Tools and Technologies :

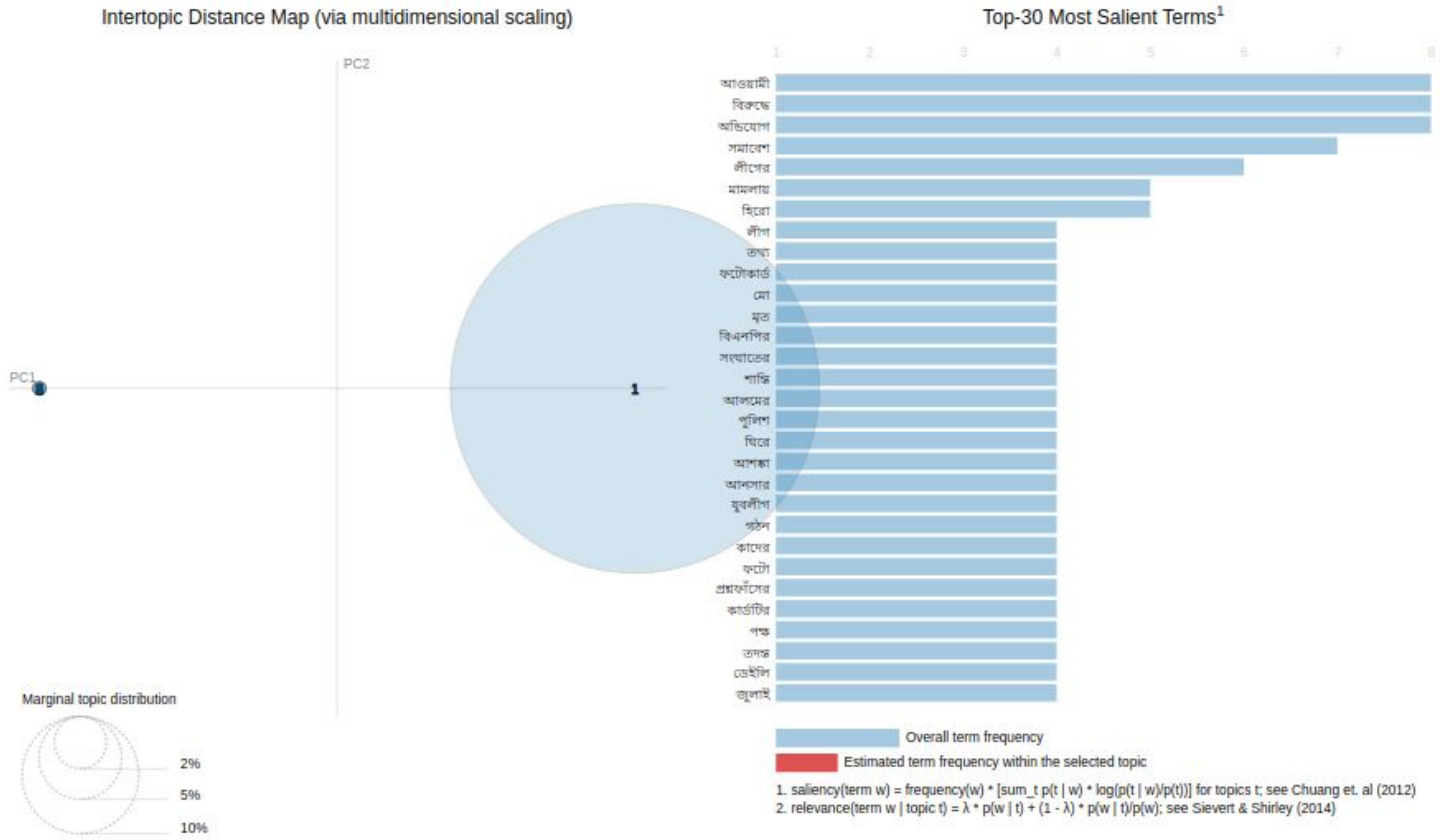
- Python
- Requests
- BeautifulSoup
- Web Scraping
- Gensim
- NLTK (Natural Language Toolkit)
- BNLP (Bengali Natural Language Processing Toolkit)
- LDA (Latent Dirichlet Allocation)
- PyLDAvis
- Matplotlib
- Googletrans
- Google Colab

Daily News

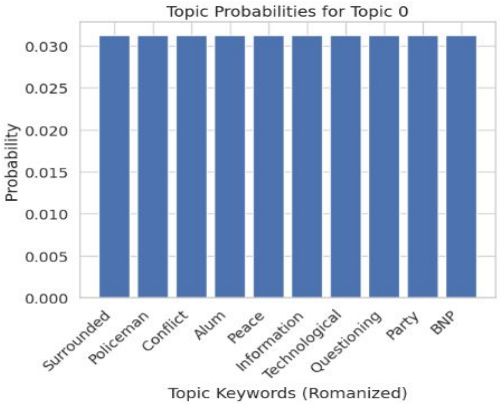
Kazi Shadman Sakib
4th Year Undergraduate Student

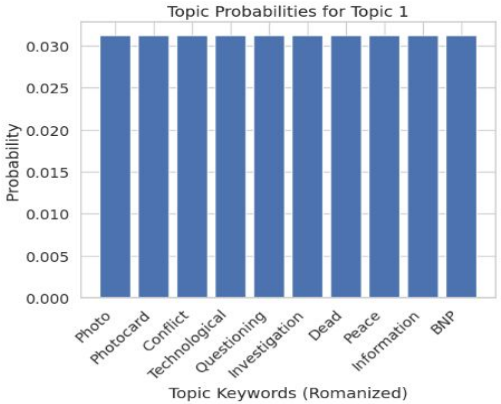
Department of Computer Science & Engineering
University of Dhaka

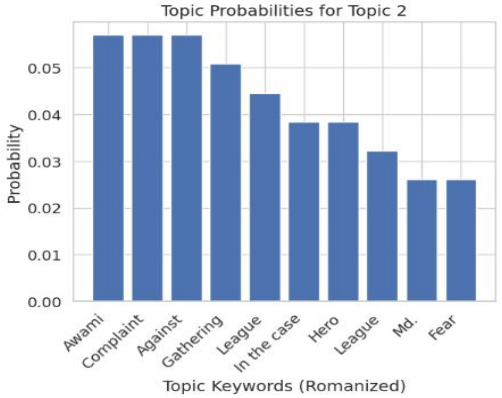
What was the trending topic on July 24, 2023?

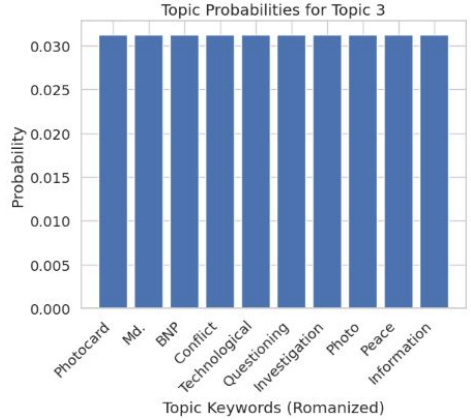


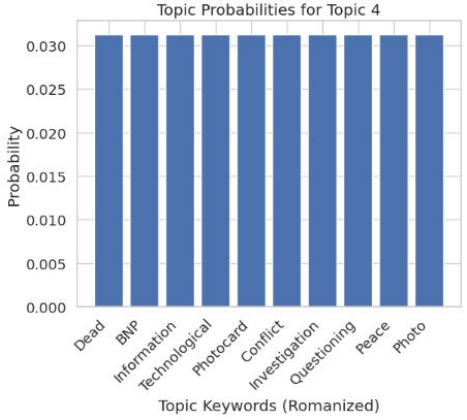
What was the trending topic on July 24, 2023? (Continued)







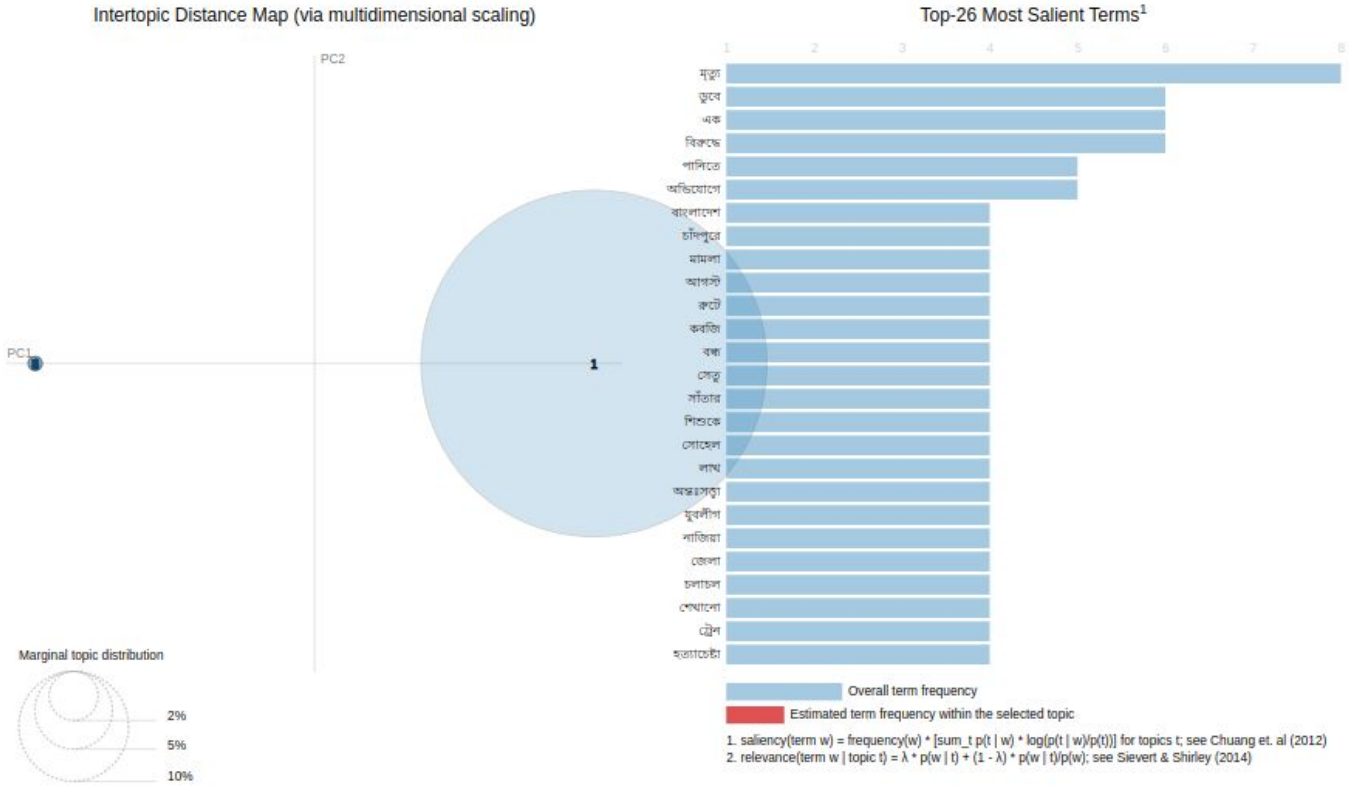




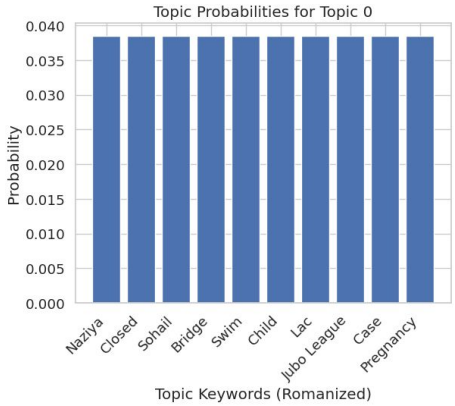
Most trending topic of the day is Topic: 2 with probabilities :

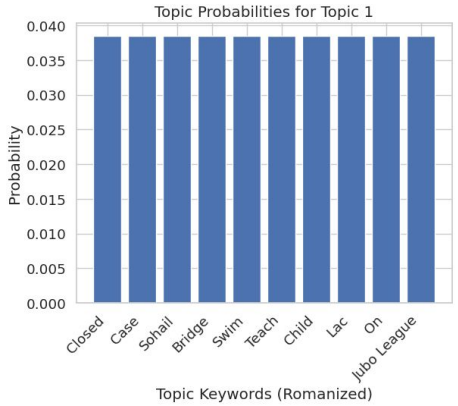
0.057*"আওয়ামী" + 0.057*"অভিযোগ" + 0.057*"বিরুদ্ধে" + 0.051*"সমাবেশ" + 0.045*"লীগের" + 0.038*"মামলায়" + 0.038*"হিরো" + 0.032*"লীগ" + 0.026*"মো" + 0.026*"আশঙ্কা"

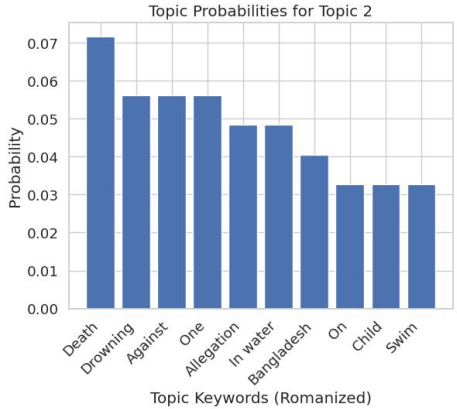
What was the trending topic on July 25, 2023?

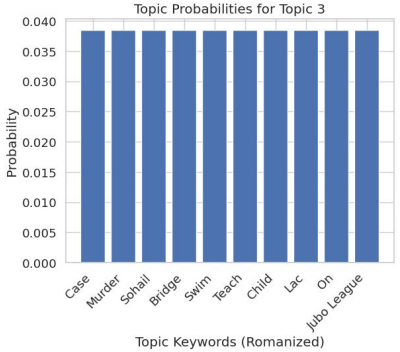


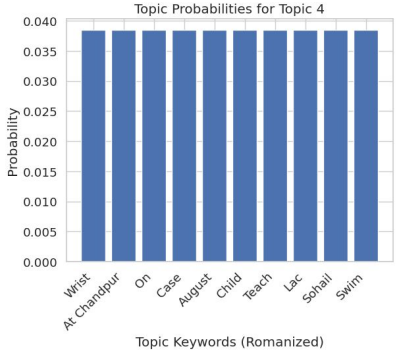
What was the trending topic on July 25, 2023? (Continued)







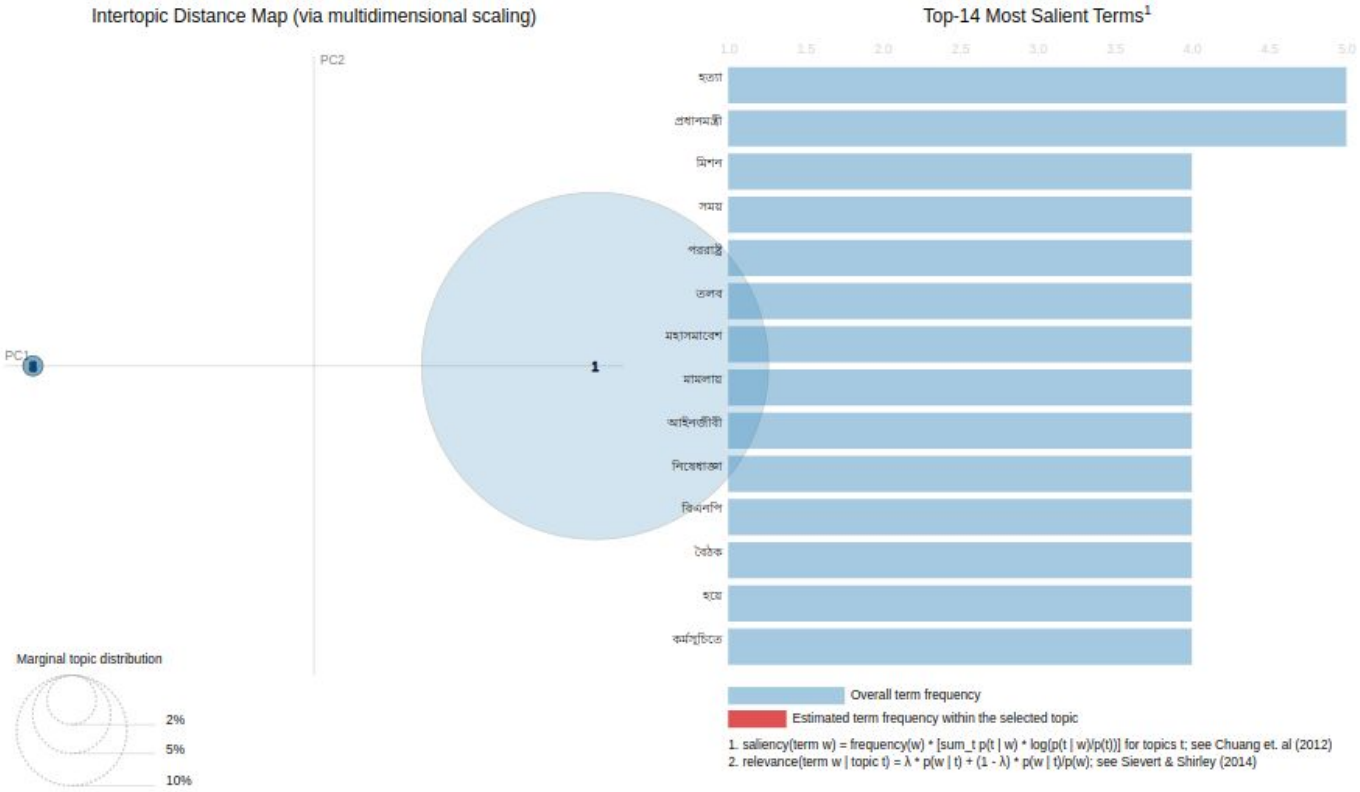




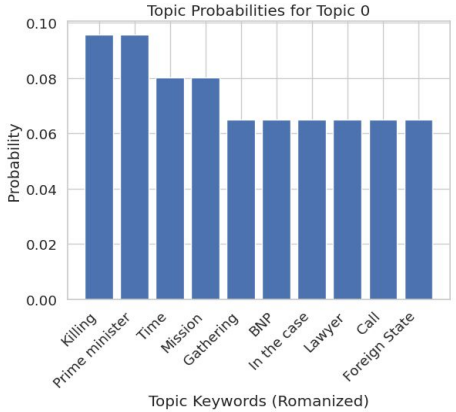
Most trending topic of the day is
Topic: 2 with probabilities :

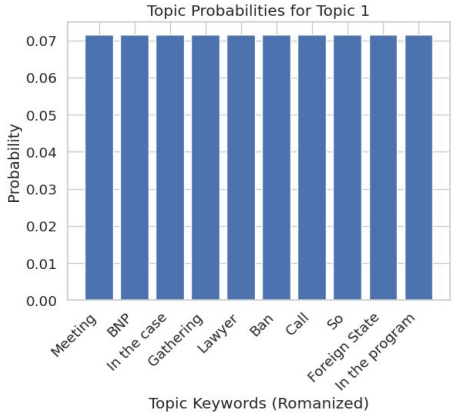
0.072*"মৃত্যু" + 0.056*"ডুবে" + 0.056*"
বিরুদ্ধে" + 0.056*"এক" + 0.048*"
অভিযোগে" + 0.048*"পানিতে" + 0.041*"
বাংলাদেশ" + 0.033*"রুটে" + 0.033*"
শিশুকে" + 0.033*"সাঁতার"

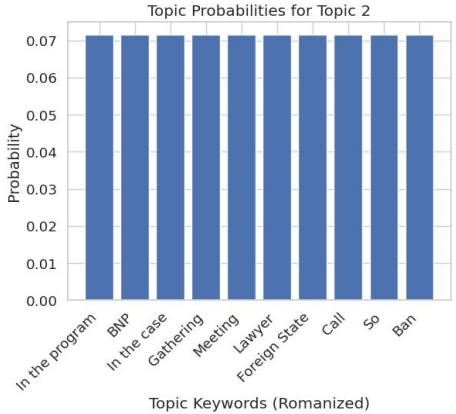
What was the trending topic on July 26, 2023?

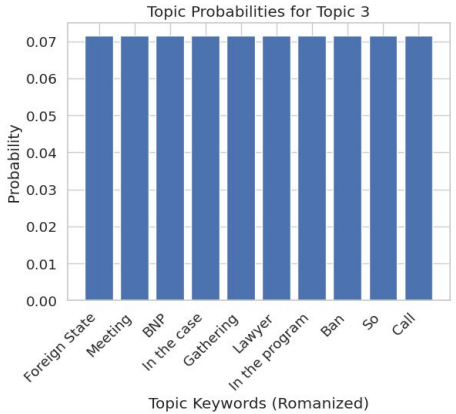


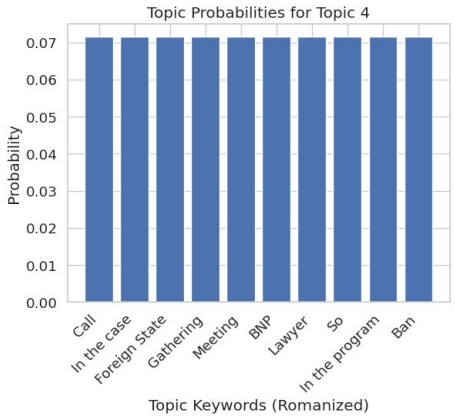
What was the trending topic on July 26, 2023? (Continued)







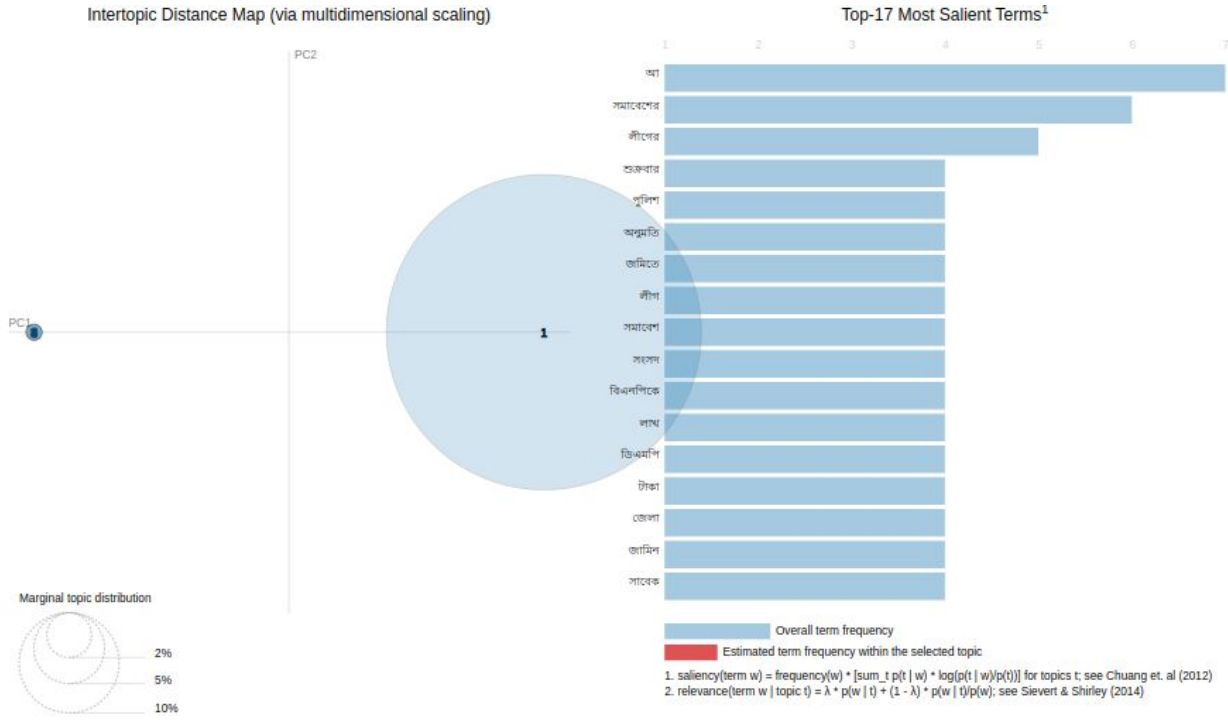




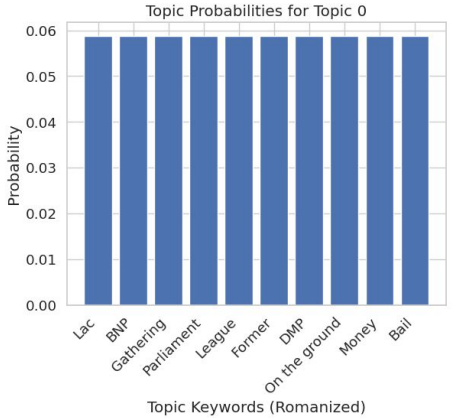
Most trending topic of the day is
Topic: 0 with probabilities :

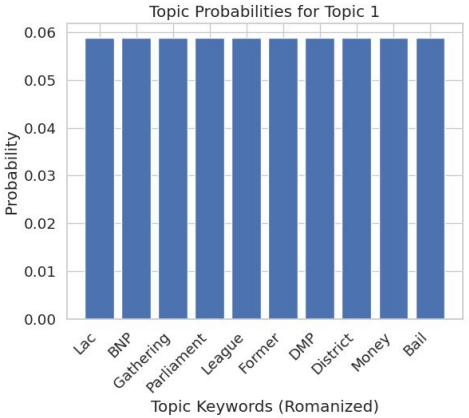
```
0.096*"হত্যা" + 0.096*"প্রধানমন্ত্রী" +
0.080*"সময়" + 0.080*"মিশন" + 0.065*"
মহাসমাবেশ" + 0.065*"বিএনপি" + 0.065*"
মামলায়" + 0.065*"আইনজীবী" + 0.065*"
তলব" + 0.065*"পররাষ্ট্র"
```

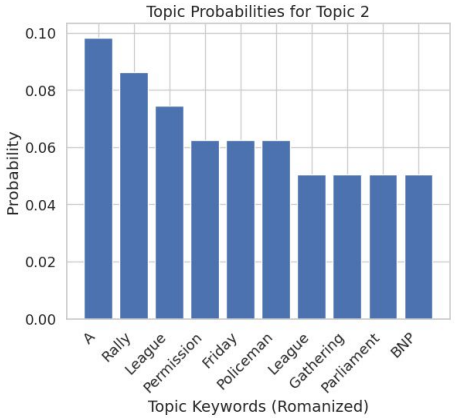
What was the trending topic on July 27, 2023?

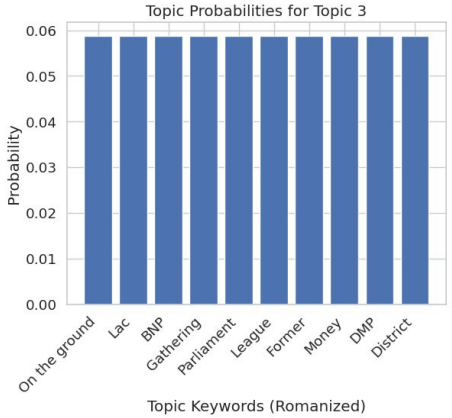


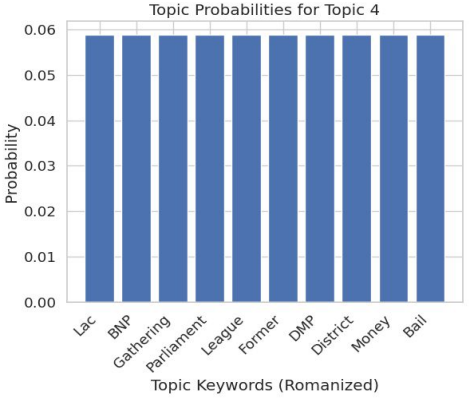
What was the trending topic on July 27, 2023? (Continued)







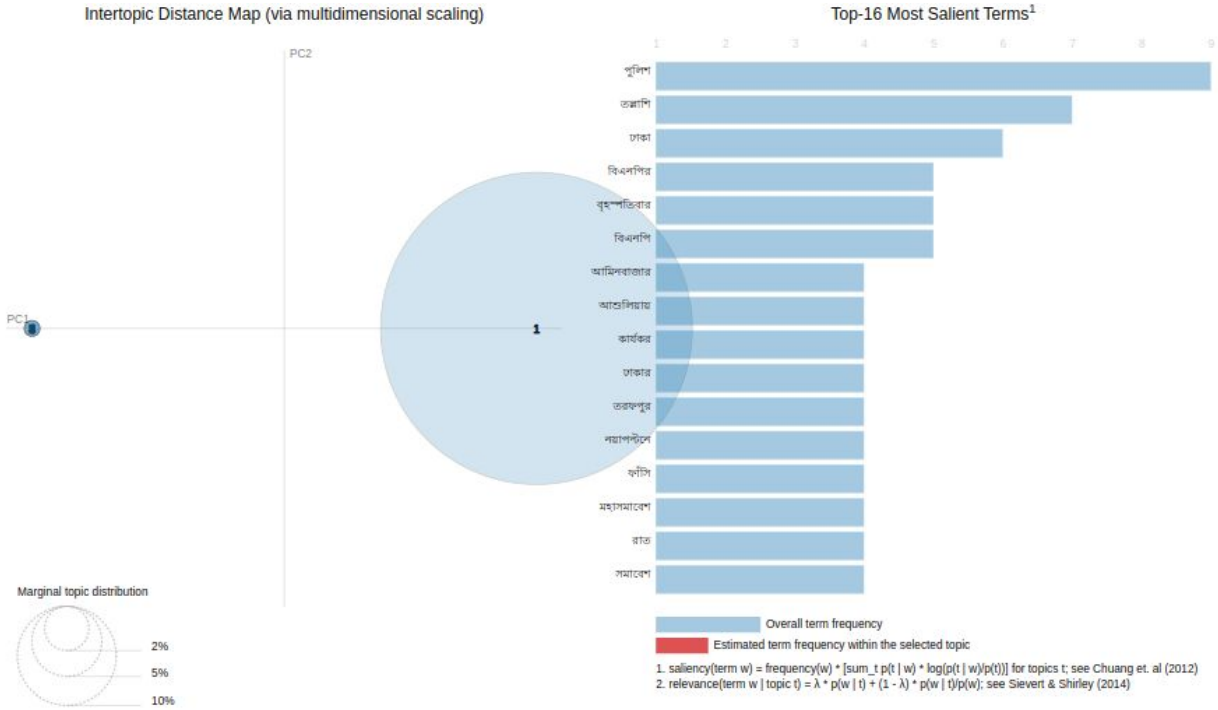




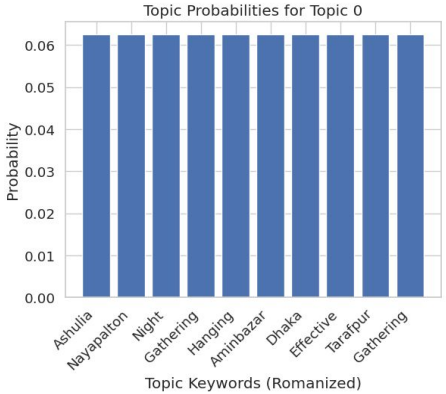
Most trending topic of the day is Topic: 2 with probabilities :

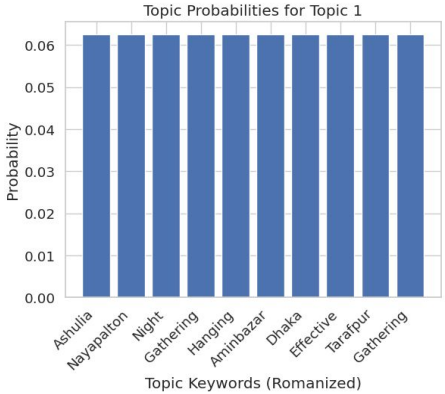
```
0.098*"আ" + 0.086*"সমাবেশের" +
0.074*"লীগের" + 0.062*"অনুমতি" +
0.062*"শুক্রবার" + 0.062*"পুলিশ" +
0.050*"লীগ" + 0.050*"সমাবেশ" +
0.050*"সংসদ" + 0.050*"বিএনপিকে"
```

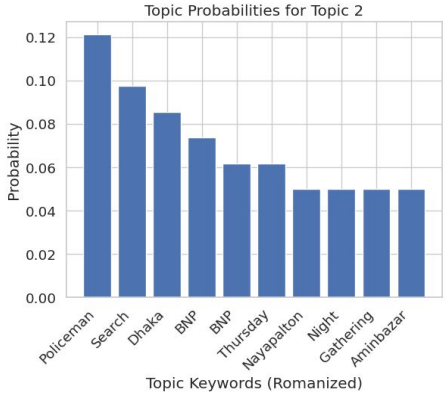
What was the trending topic on July 28, 2023?

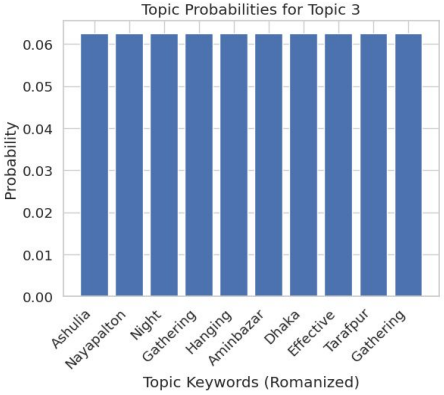


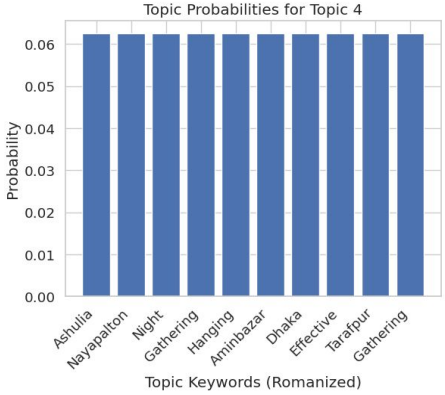
What was the trending topic on July 28, 2023? (Continued)







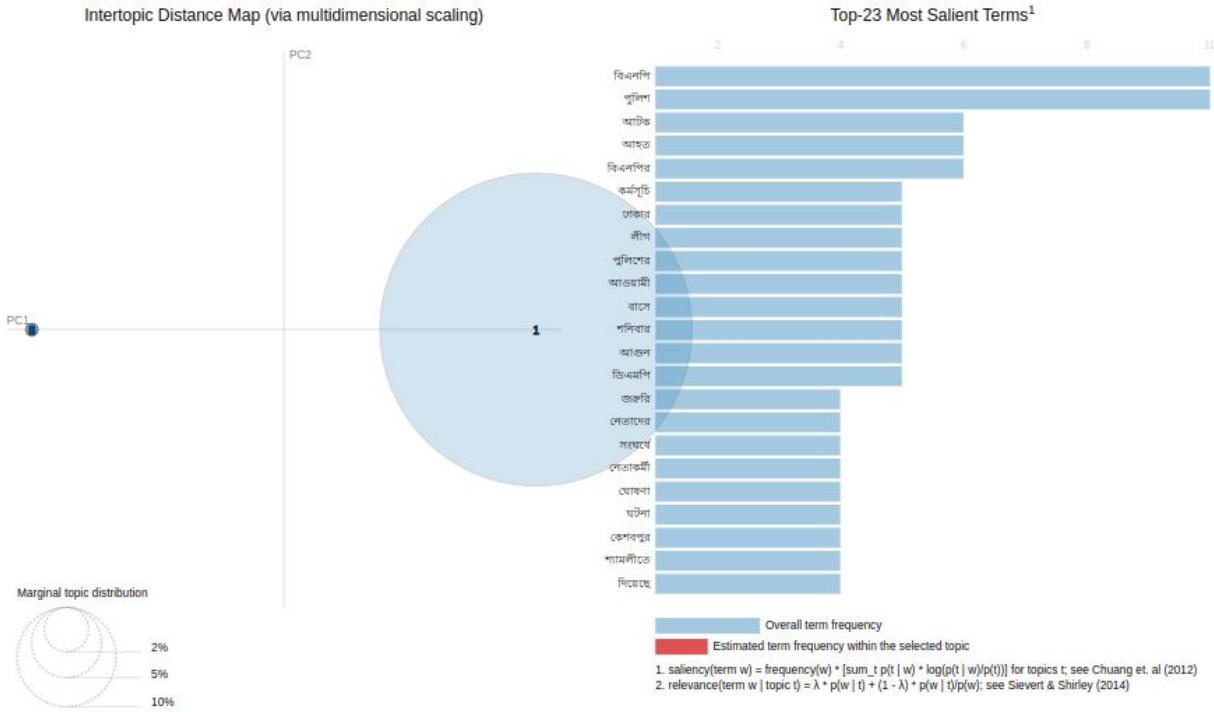




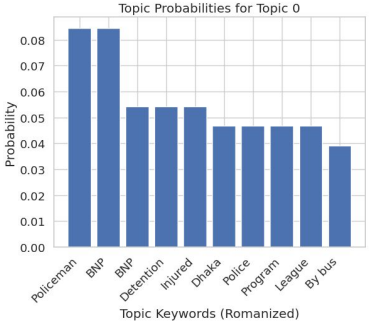
Most trending topic of the day is
Topic: 2 with probabilities :

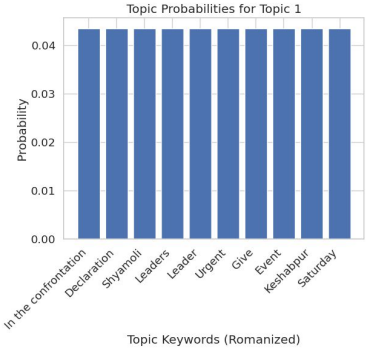
```
0.121*"পুলিশ" + 0.097*"তল্লাশি" + 0.086*"
ঢাকা" + 0.074*"বিএনপির" + 0.062*"বিএনপি"
+ 0.062*"বৃহস্পতিবার" + 0.050*"নয়াপল্টনে" +
0.050*"রাত" + 0.050*"মহাসমাবেশ" +
0.050*"আমিনবাজার"
```

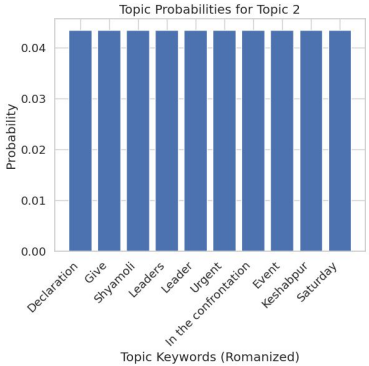
What was the trending topic on July 29, 2023?

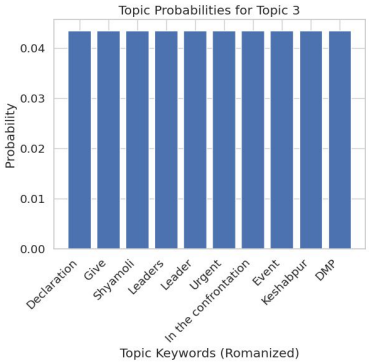


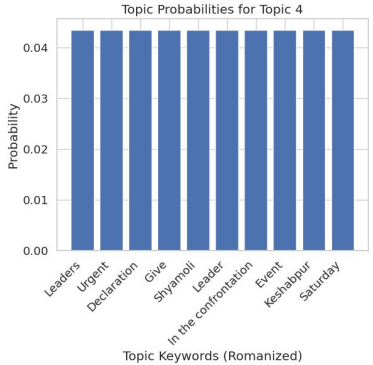
What was the trending topic on July 29, 2023? (Continued)







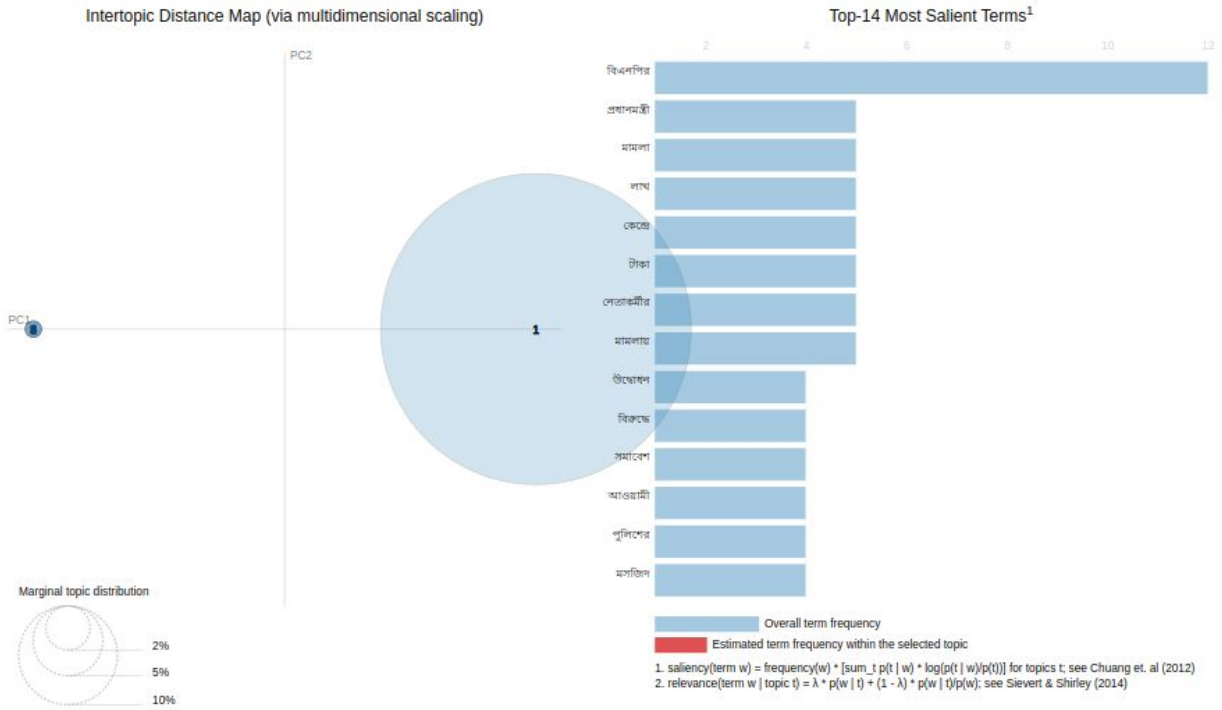




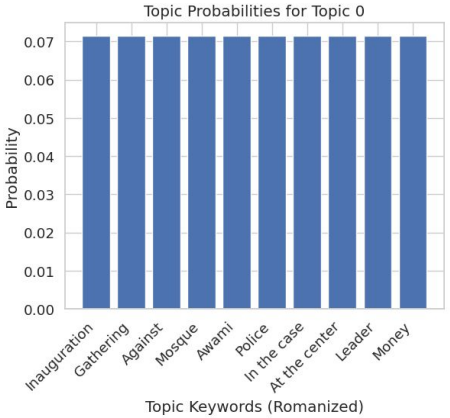
Most trending topic of the day is
Topic: 0 with probabilities :

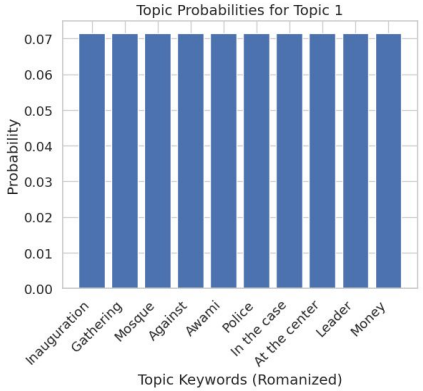
```
0.084*"পুলিশ" + 0.084*"বিএনপি" + 0.054*"
বিএনপির" + 0.054*"আটক" + 0.054*"আহত"
+ 0.047*"ঢাকার" + 0.047*"পুলিশের" +
0.047*"কর্মসূচি" + 0.047*"লীগ" + 0.039*"
বাসে"
```

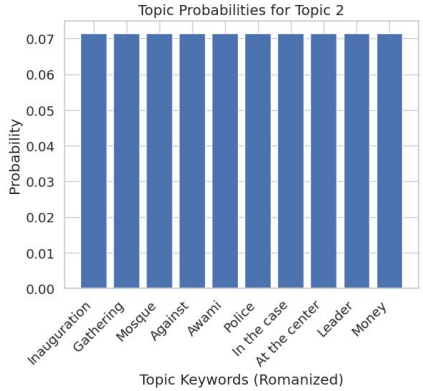
What was the trending topic on July 30, 2023?

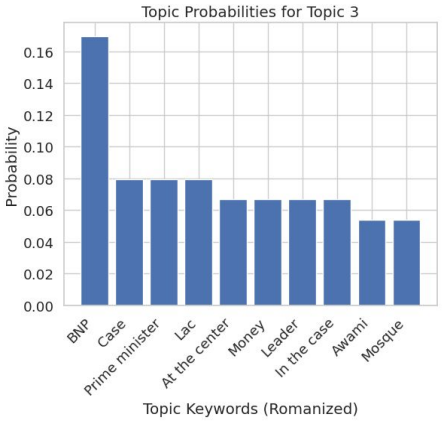


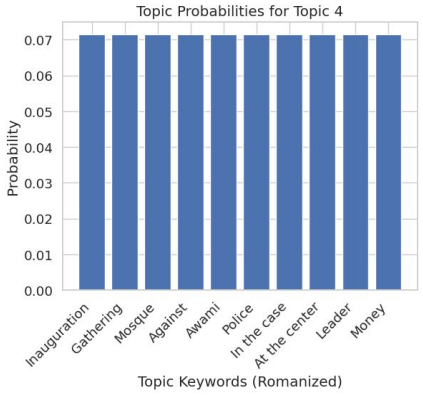
What was the trending topic on July 30, 2023? (Continued)







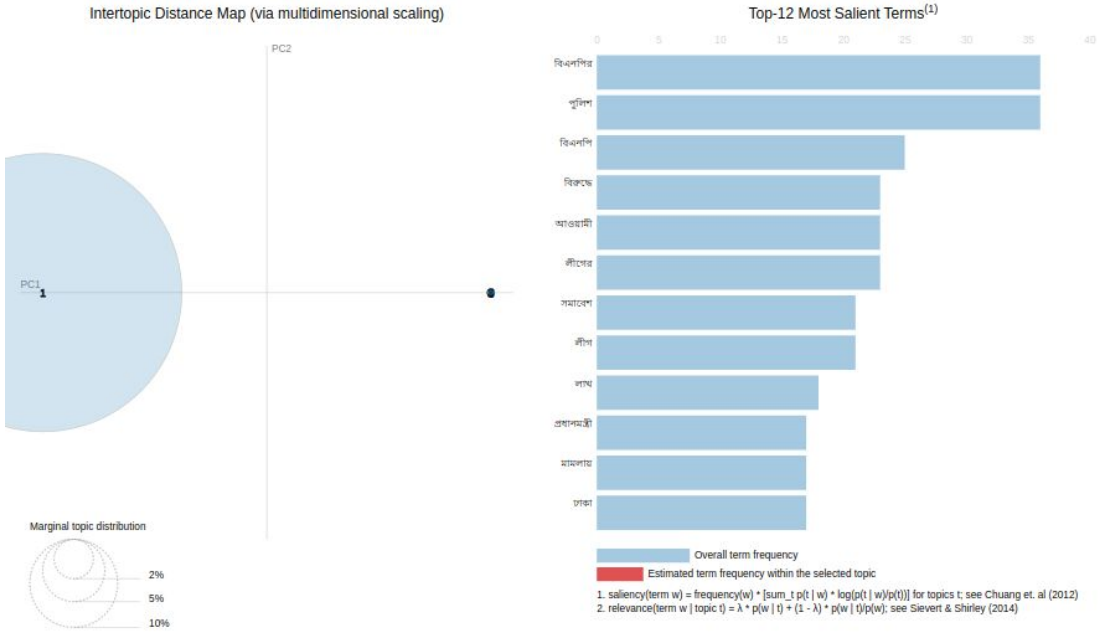




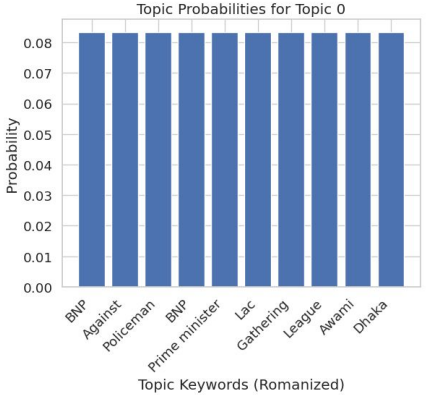
Most trending topic of the day is
Topic: 3 with probabilities :

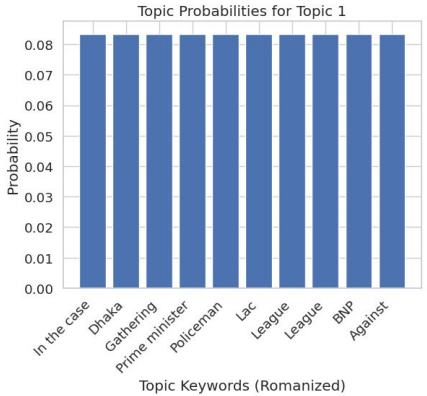
```
0.170*"বিএনপির" + 0.080*"মামলা" +
0.080*"প্রধানমন্ত্রী" + 0.080*"লাখ" + 0.067*"
কেন্দ্রে" + 0.067*"টাকা" + 0.067*"নেতাকর্মীর"
+ 0.067*"মামলায়" + 0.054*"আওয়ামী" +
0.054*"মসজিদ"
```

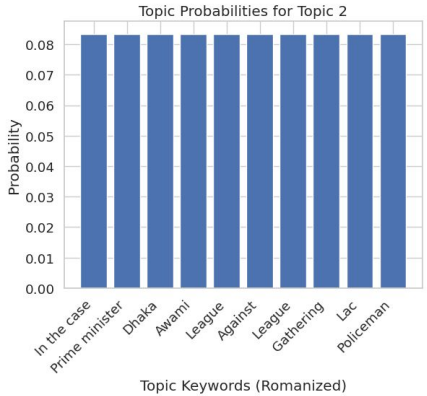
What is the trending topic of this week?

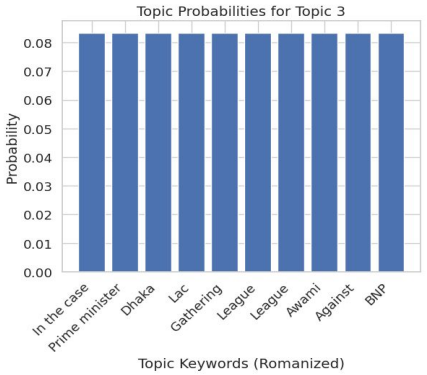


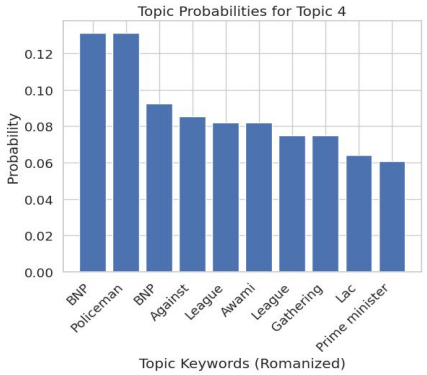
What is the trending topic of this week? (continued)











Most trending topic of the day is
Topic: 4 with probabilities :

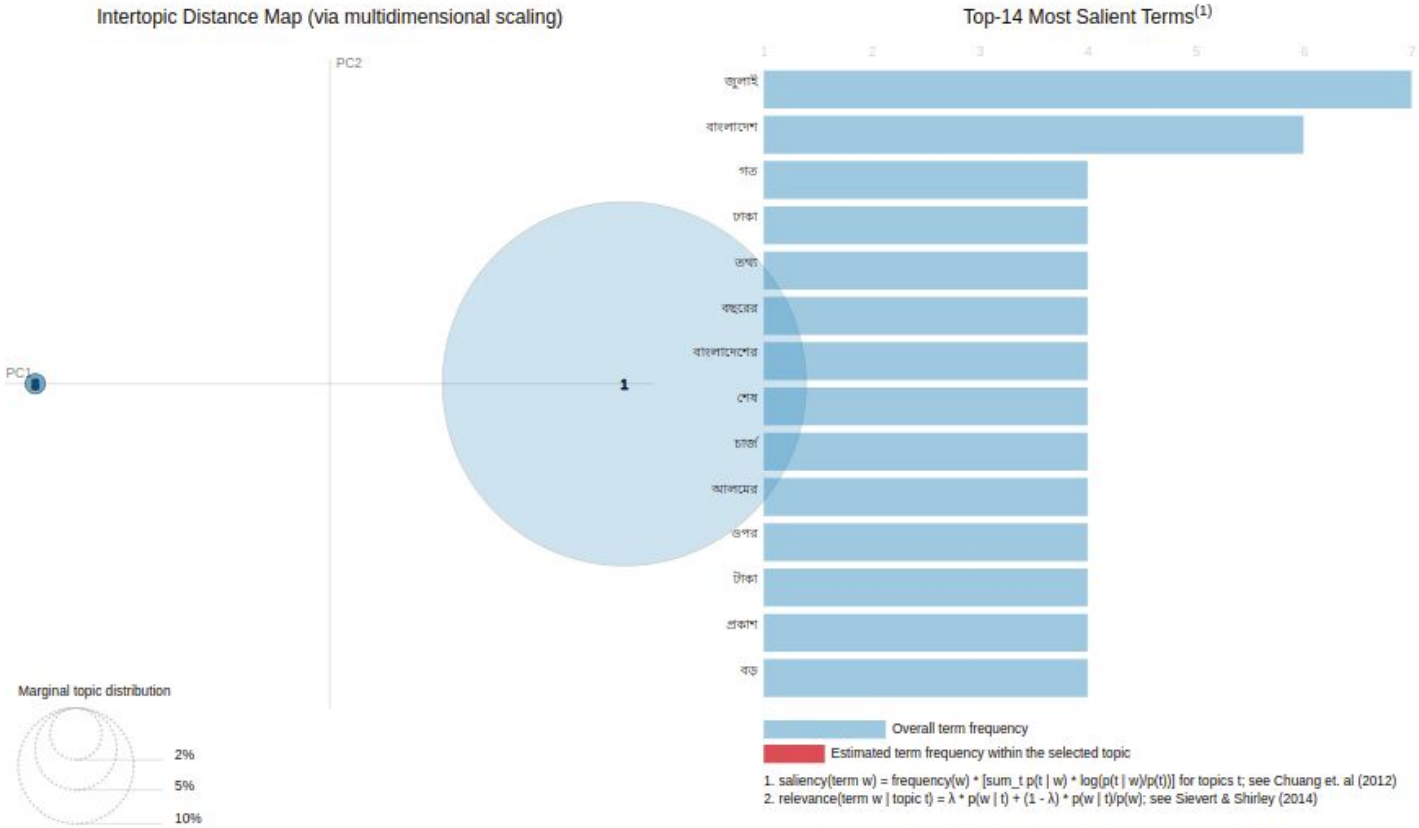
0.131*"বিএনপির" + 0.131*"পুলিশ" +
0.092*"বিএনপি" + 0.085*"বিরুদ্ধে" +
0.082*"লীগের" + 0.082*"আওয়ামী" +
0.075*"লীগ" + 0.075*"সমাবেশ" + 0.064*"
লাখ" + 0.061*"প্রধানমন্ত্রী"

Peak User Hour News

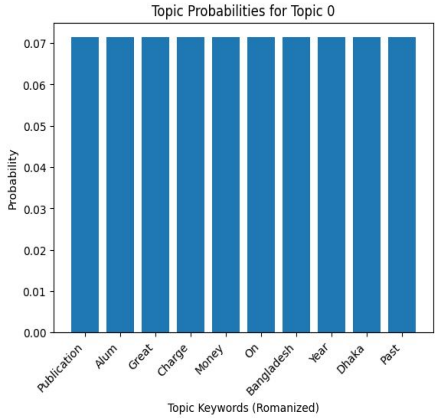
Kazi Shadman Sakib
4th Year Undergraduate Student

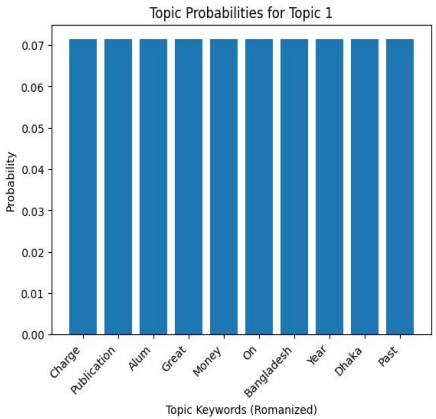
Department of Computer Science & Engineering
University of Dhaka

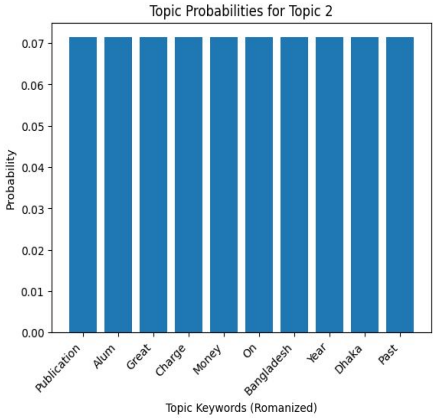
What was the most popular topic during the peak user hour on July 24, 2023?

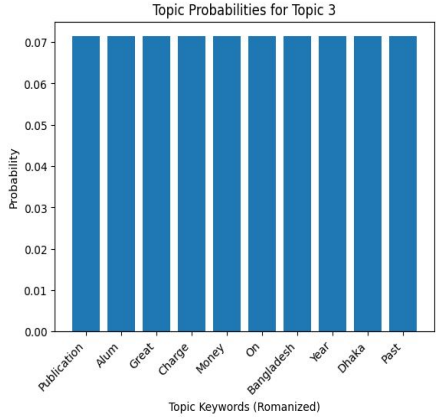


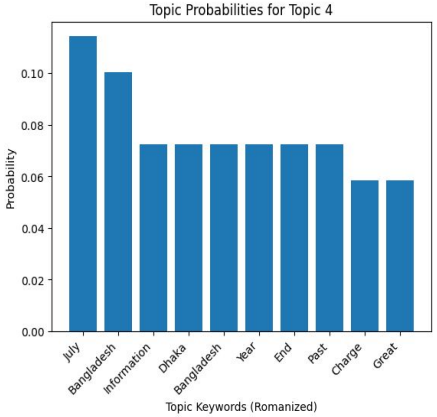
What was the most popular topic during the peak user hour on July 24, 2023? (Continued)







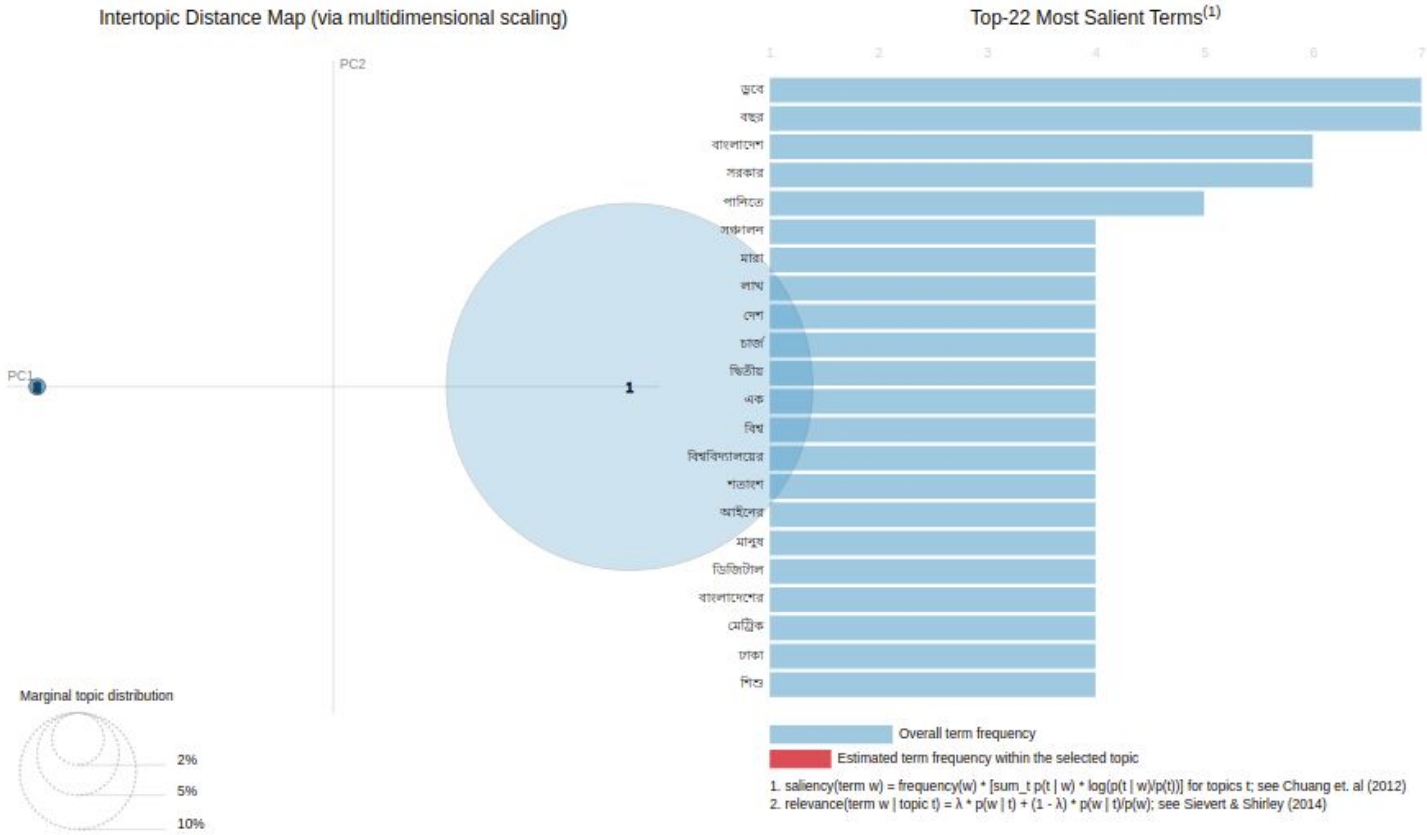




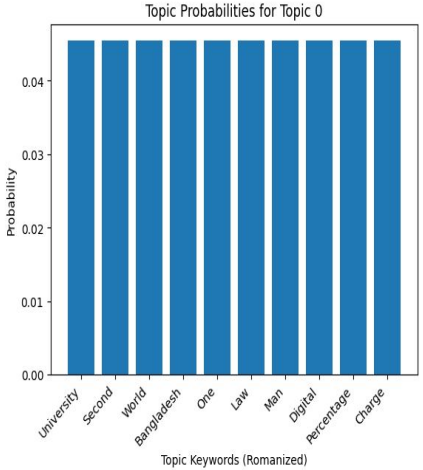
Most trending topic of the day is Topic: 4 with probabilities :

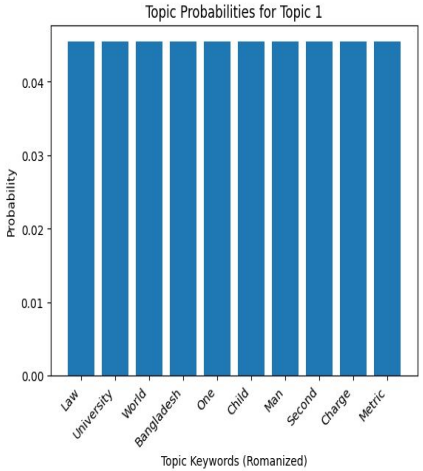
```
0.114*"জুলাই" + 0.100*"বাংলাদেশ" +
0.072*"তথ্য" + 0.072*"ঢাকা" + 0.072*"
বাংলাদেশের" + 0.072*"বছরের" + 0.072*"
শেষ" + 0.072*"গত" + 0.058*"চার্জ" +
0.058*"বড়"
```

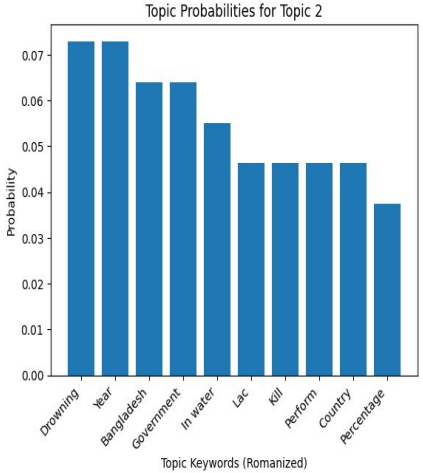
What was the most popular topic during the peak user hour on July 25, 2023?

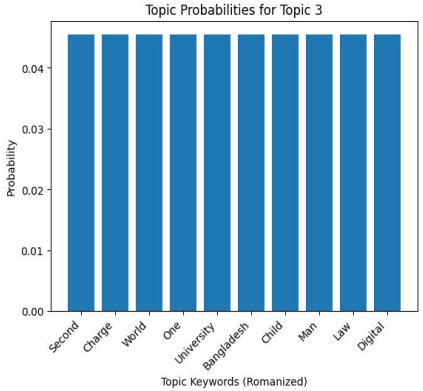


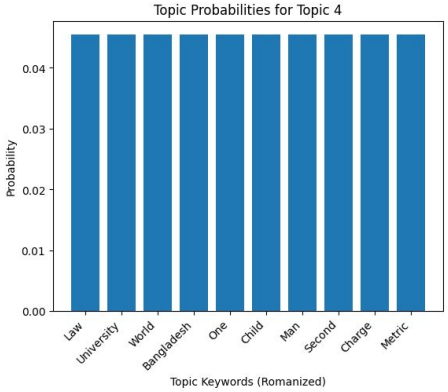
What was the most popular topic during the peak user hour on July 25, 2023? (Continued)







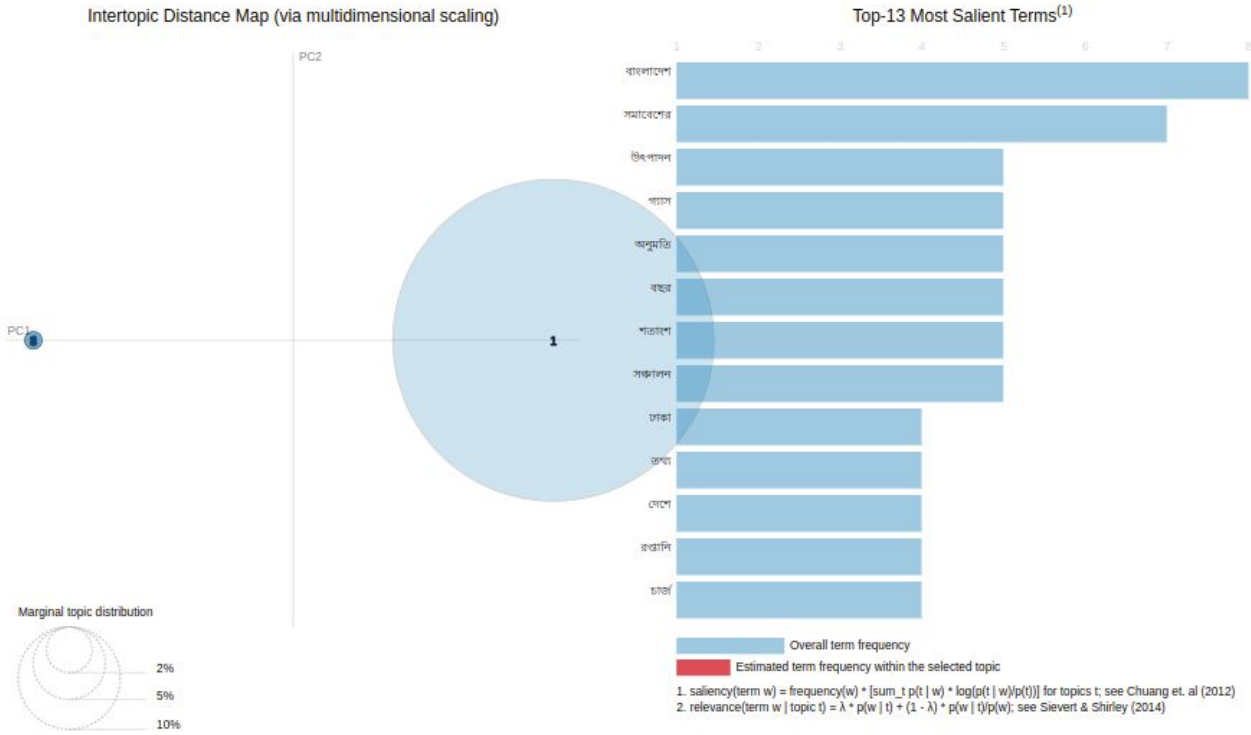




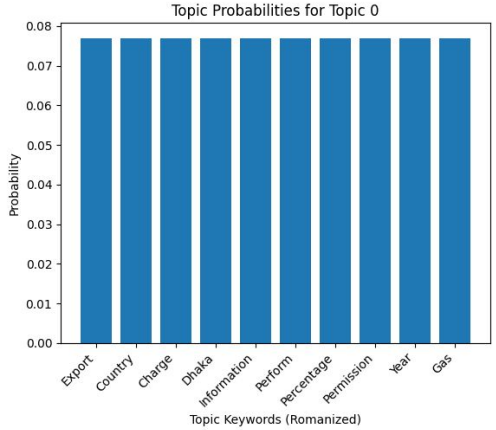
Most trending topic of the day is Topic: 2 with probabilities :

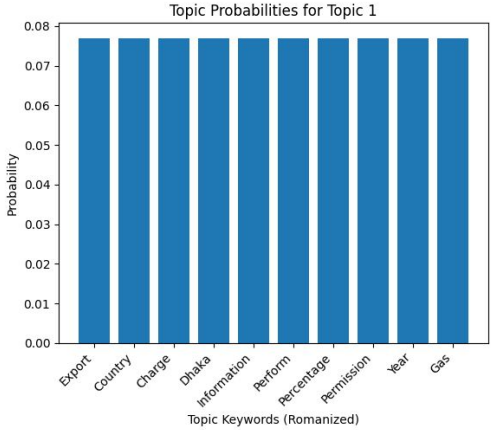
```
0.073*"ডুবে" + 0.073*"বছর" + 0.064*"বাংলাদেশ" + 0.064*"সরকার" + 0.055*"পানিতে" + 0.046*"লাখ" + 0.046*"মারা" + 0.046*"সঞ্চালন" + 0.046*"দেশ" + 0.037*"শতাংশ"
```

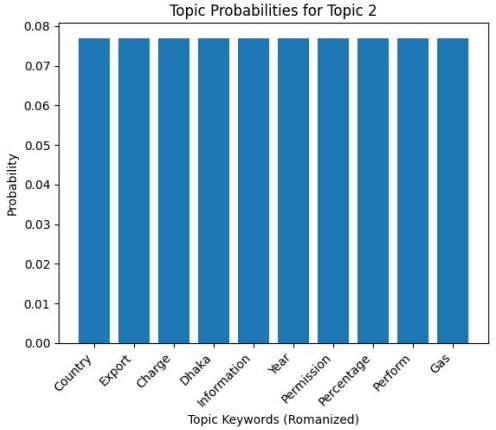
What was the most popular topic during the peak user hour on July 26, 2023?

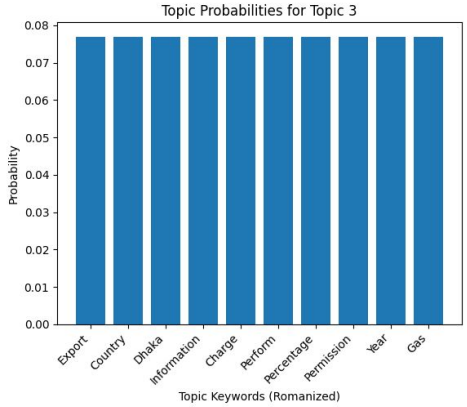


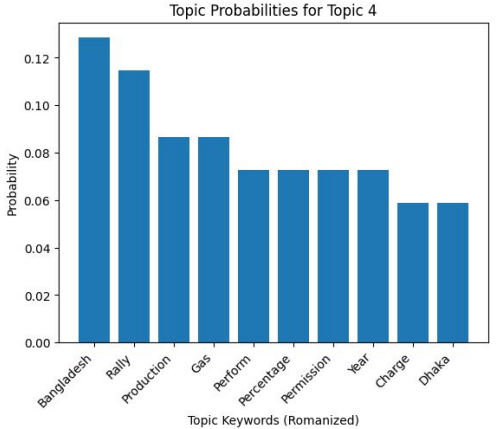
What was the most popular topic during the peak user hour on July 26, 2023? (Continued)







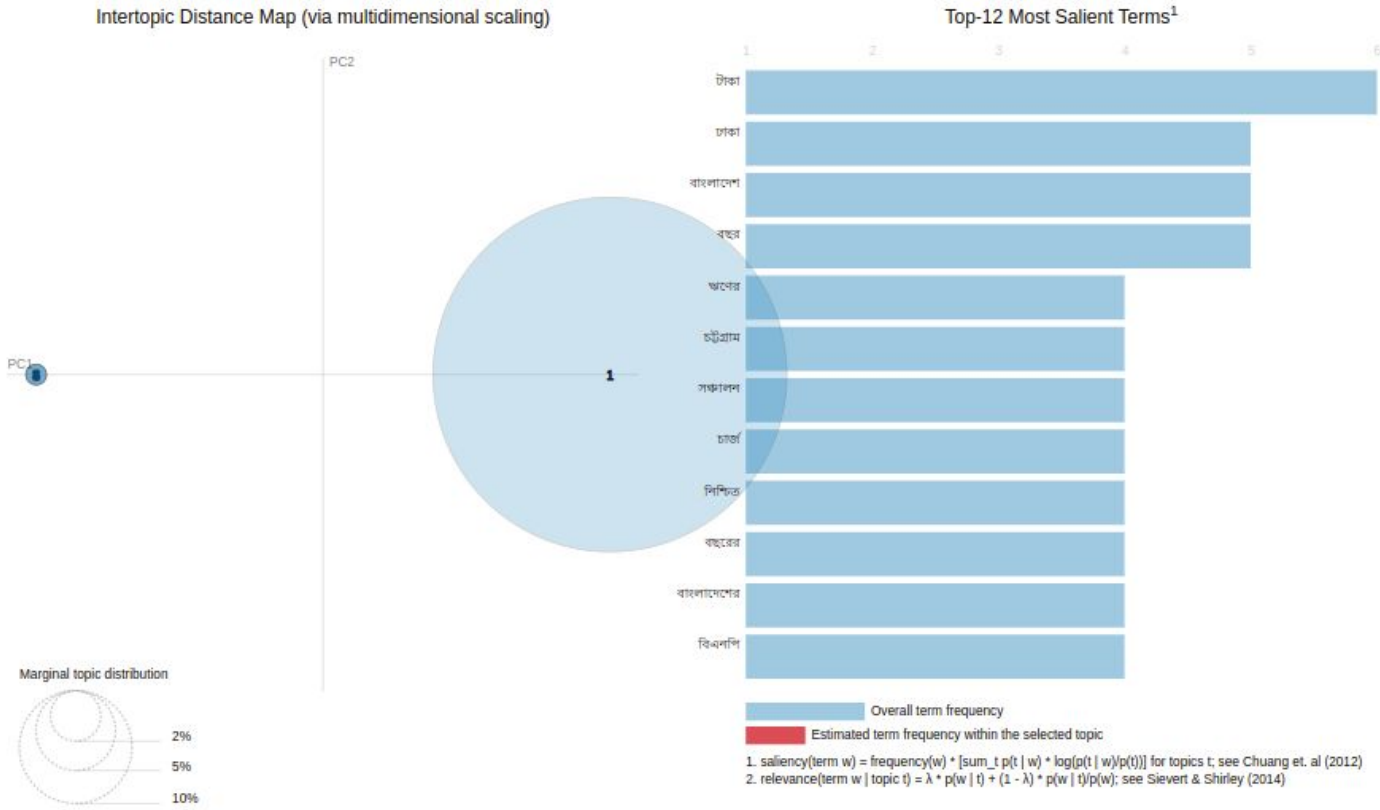




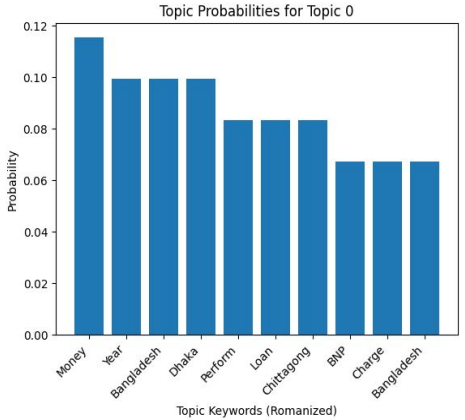
Most trending topic of the day is Topic: 4 with probabilities :

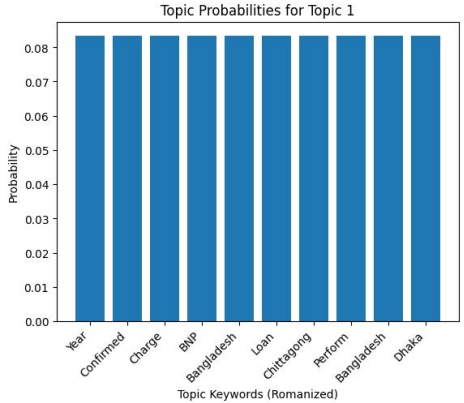
```
0.128*"বাংলাদেশ" + 0.115*"সমাবেশের" +
0.087*"উৎপাদন" + 0.087*"গ্যাস" +
0.073*"সঞ্চালন" + 0.073*"শতাংশ" +
0.073*"অনুমতি" + 0.073*"বছর" +
0.059*"চার্জ" + 0.059*"ঢাকা"
```

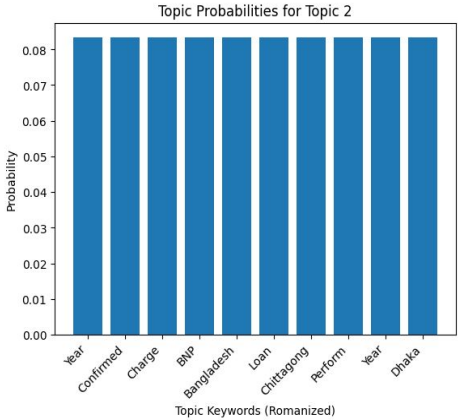
What was the most popular topic during the peak user hour on July 27, 2023?

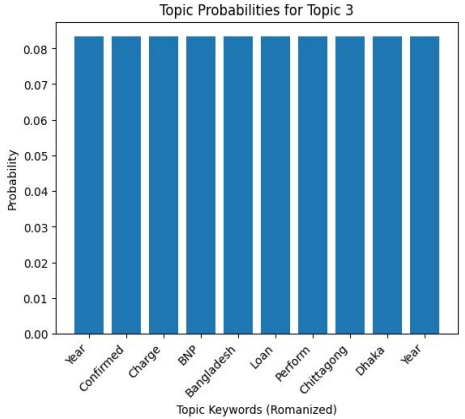


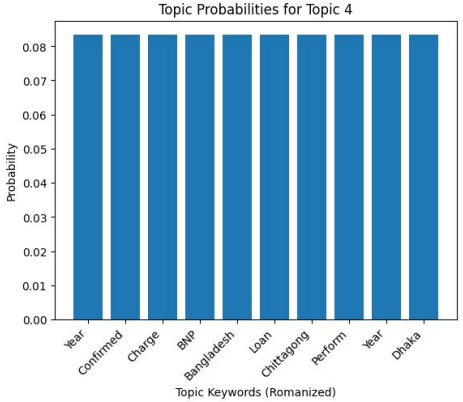
What was the most popular topic during the peak user hour on July 27, 2023? (Continued)







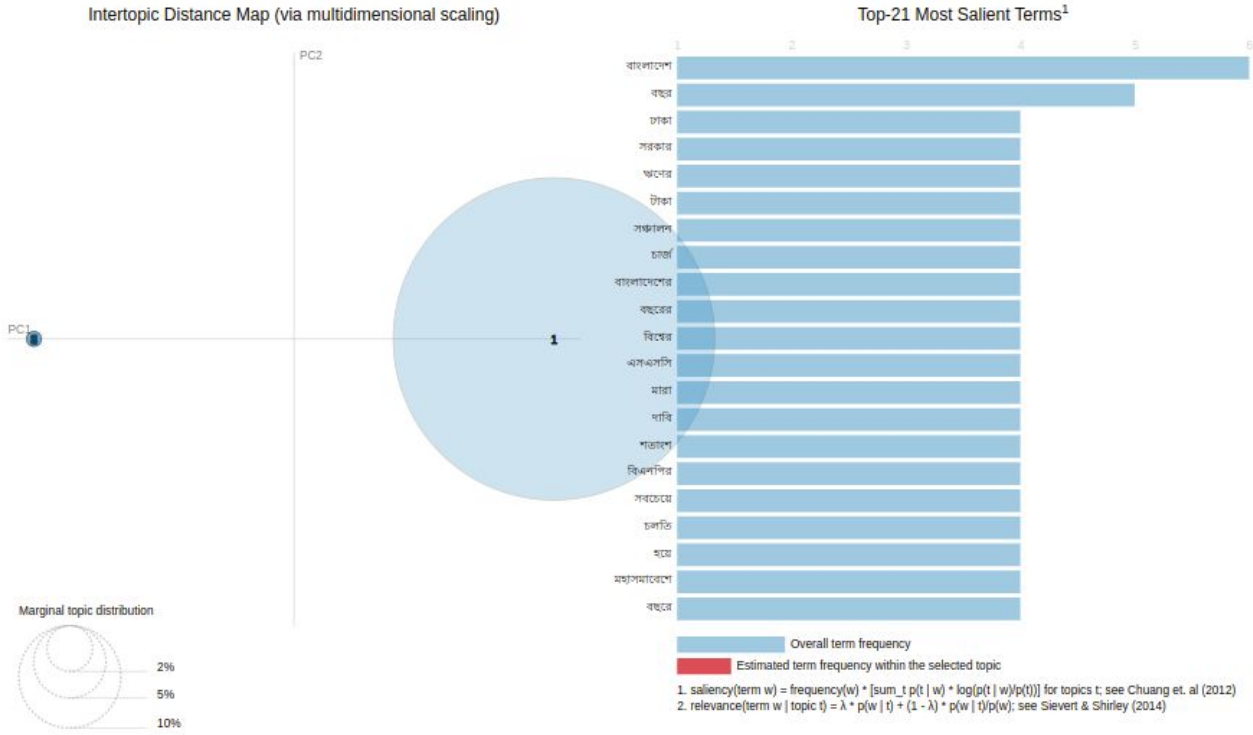




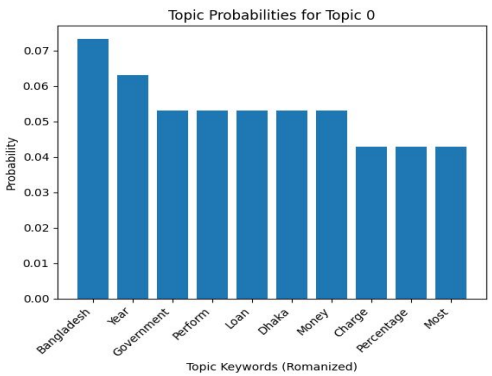
Most trending topic of the day is Topic: 0 with probabilities :

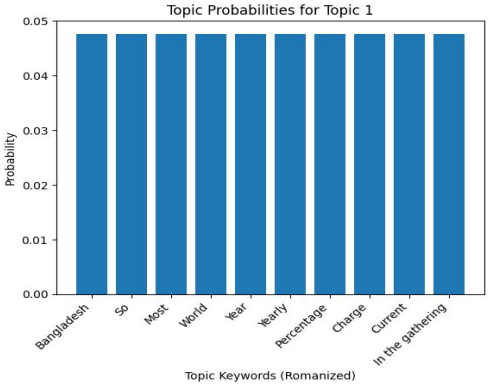
```
0.115*"টাকা" + 0.099*"বছর" + 0.099*"
বাংলাদেশ" + 0.099*"ঢাকা" + 0.083*"
সঞ্চালন" + 0.083*"ঋণের" + 0.083*"চট্টগ্রাম"
+ 0.067*"বিএনপি" + 0.067*"চার্জ" +
0.067*"বাংলাদেশের"
```

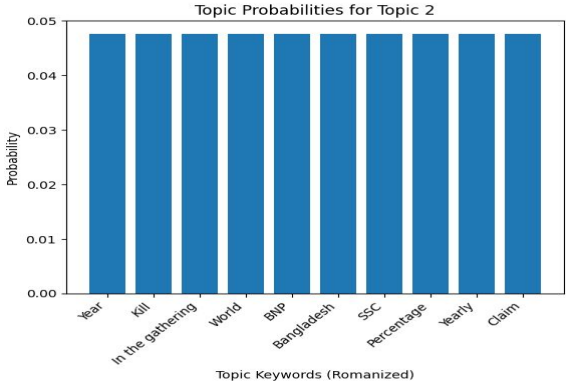
What was the most popular topic during the peak user hour on July 28, 2023?

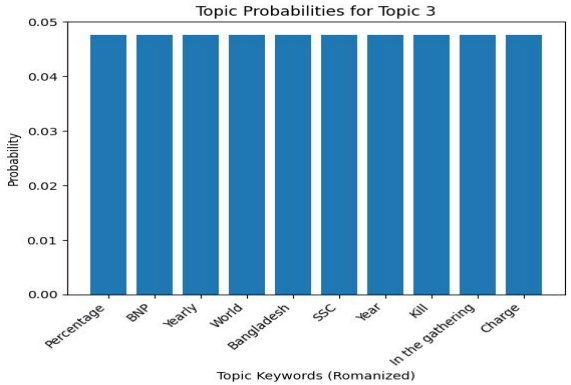


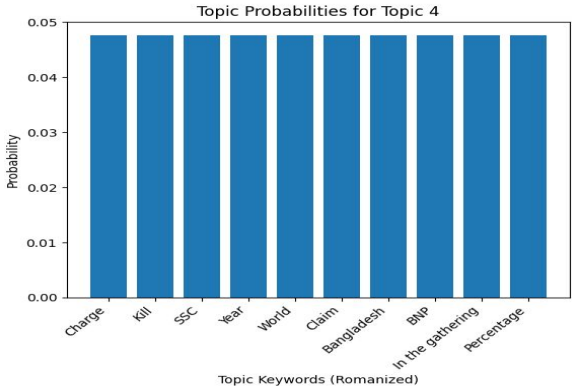
What was the most popular topic during the peak user hour on July 28, 2023? (Continued)







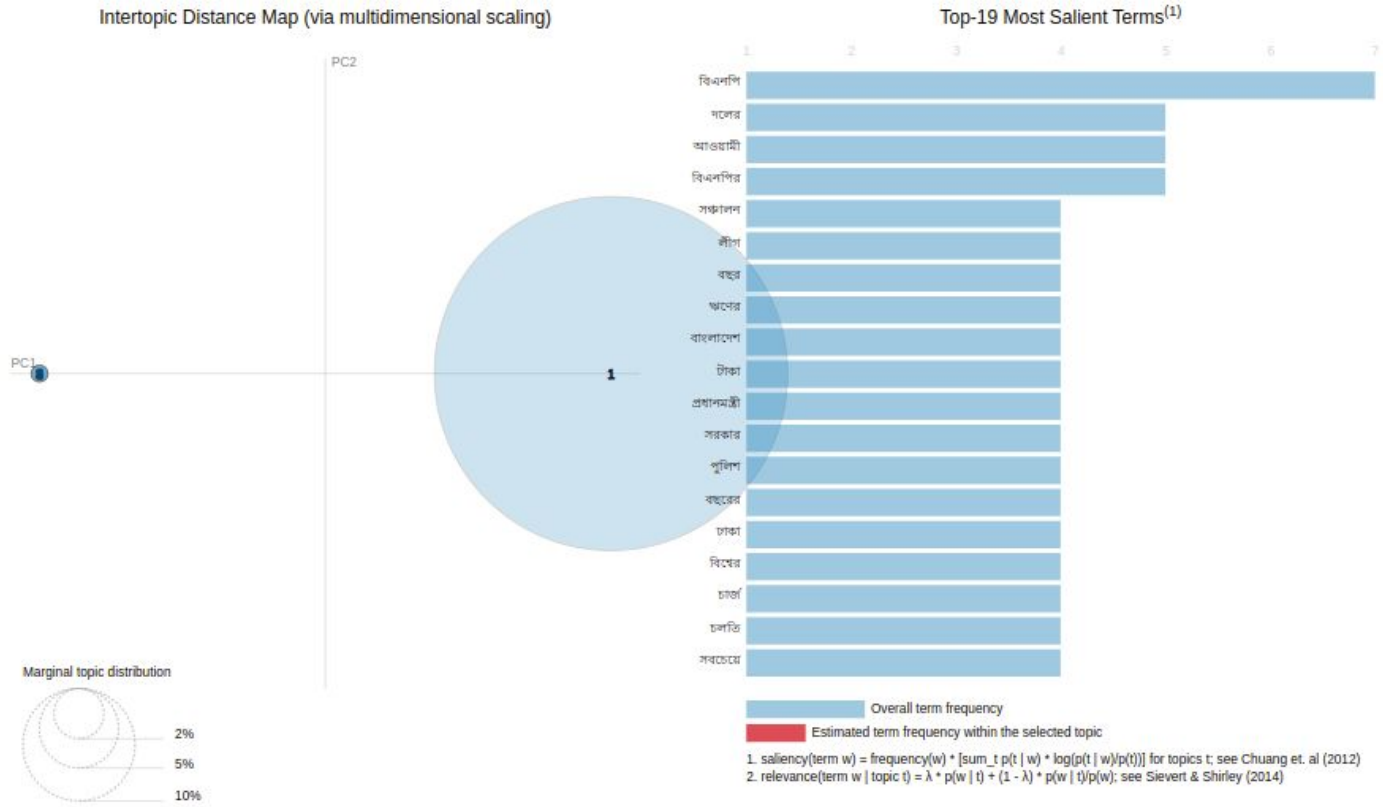




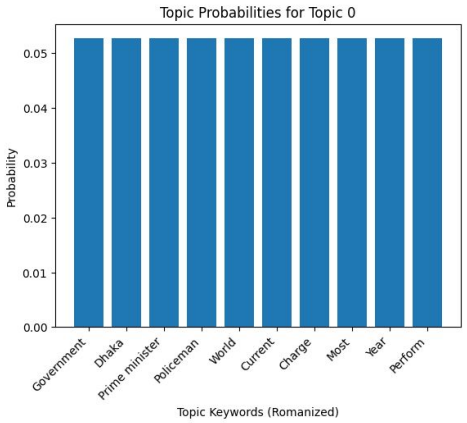
Most trending topic of the day is Topic: 0 with probabilities :

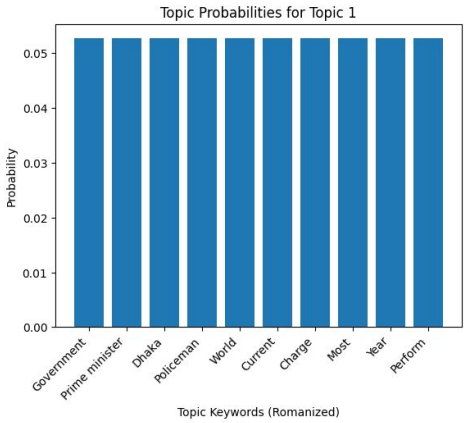
```
0.073*"বাংলাদেশ" + 0.063*"বছর" +
0.053*"সরকার" + 0.053*"সঞ্চালন" +
0.053*"ঋণের" + 0.053*"ঢাকা" +
0.053*"টাকা" + 0.043*"চার্জ" +
0.043*"শতাংশ" + 0.043*"সবচেয়ে"
```

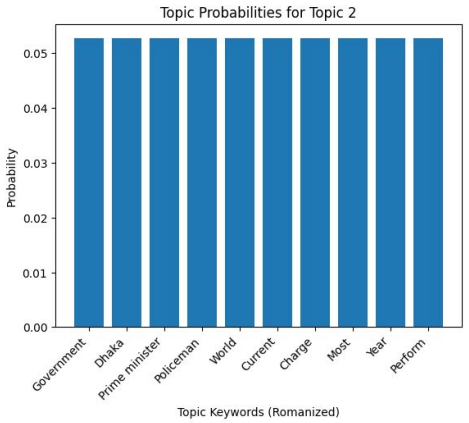
What was the most popular topic during the peak user hour on July 29, 2023?

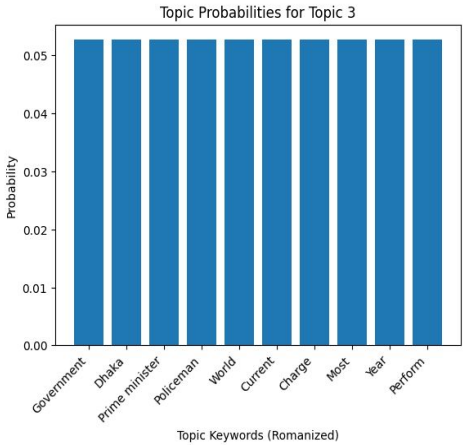


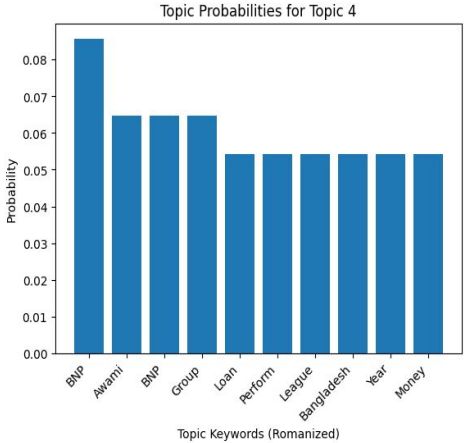
What was the most popular topic during the peak user hour on July 29, 2023? (Continued)







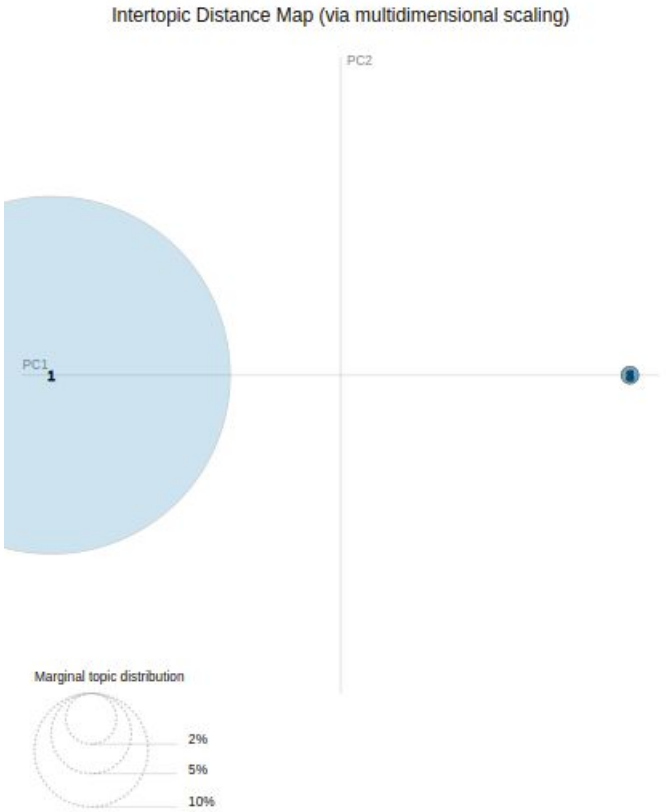


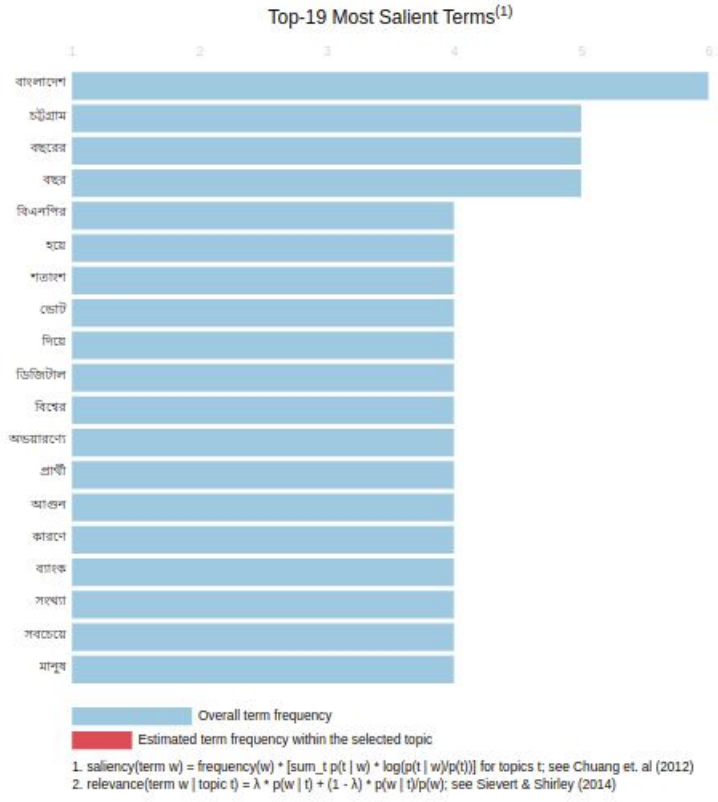


Most trending topic of the day is Topic: 4 with probabilities :

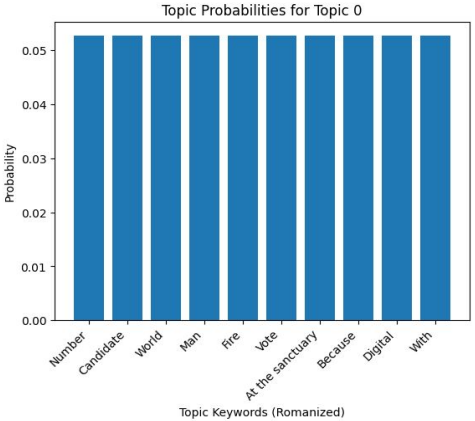
```
0.086*"বিএনপি" + 0.065*"আওয়ামী" +
0.065*"বিএনপির" + 0.065*"দলের" +
0.054*"ঋণের" + 0.054*"সঞ্চালন" +
0.054*"লীগ" + 0.054*"বাংলাদেশ" +
0.054*"বছর" + 0.054*"টাকা"
```

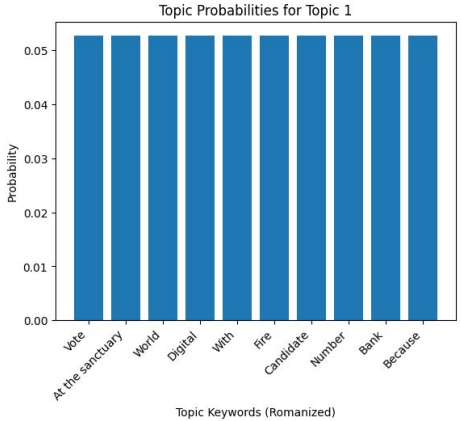
What was the most popular topic during the peak user hour on July 30, 2023?

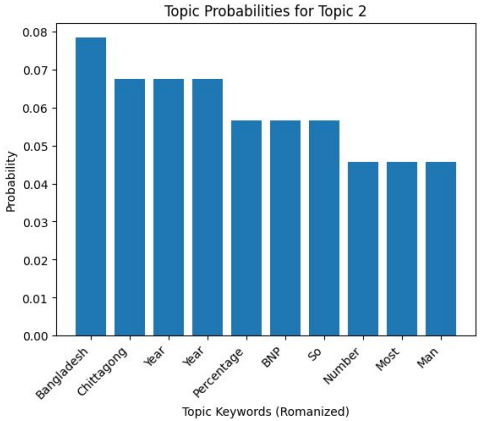


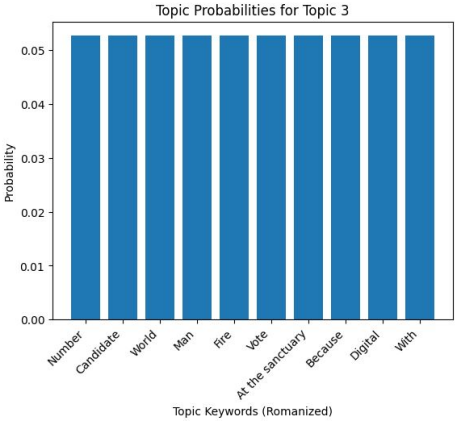


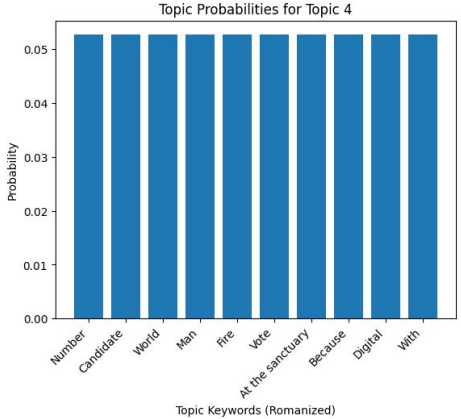
What was the most popular topic during the peak user hour on July 30, 2023? (Continued)







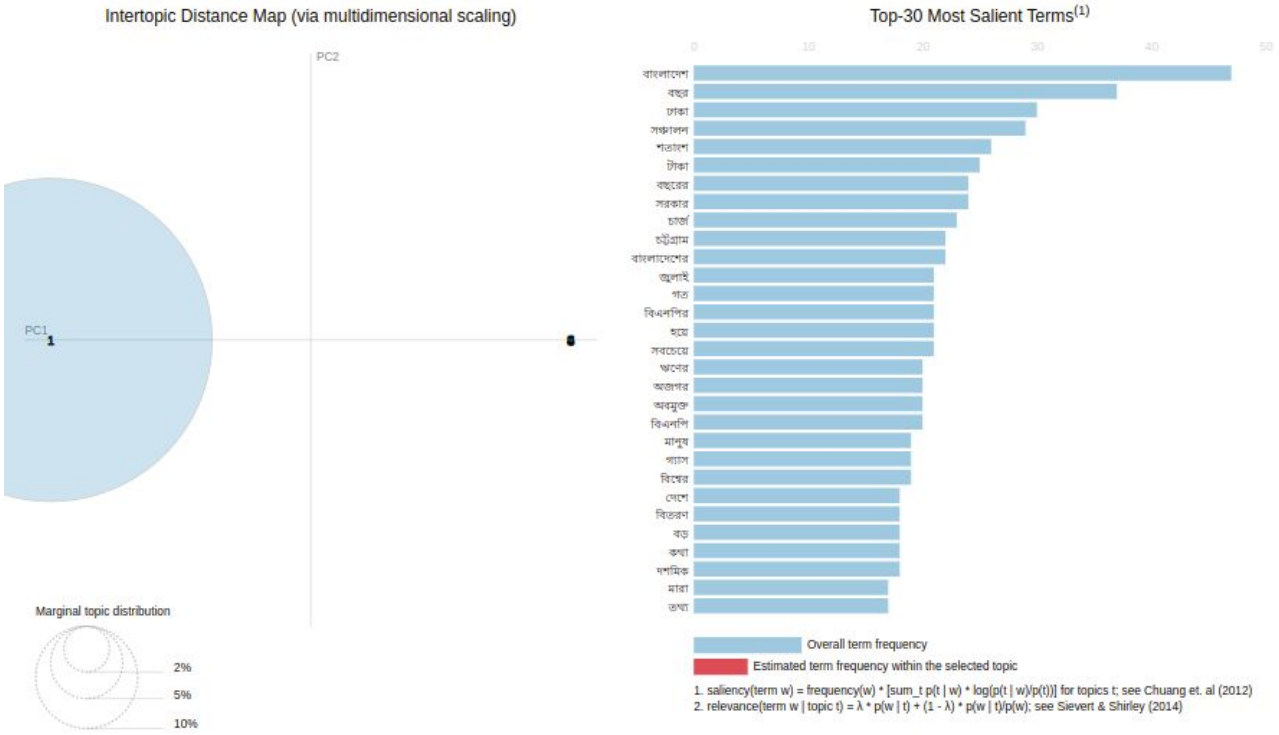




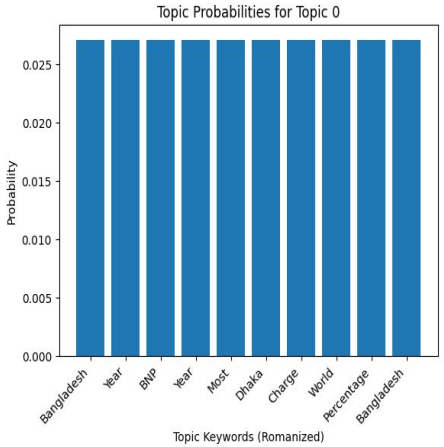
Most trending topic of the day is Topic: 2 with probabilities :

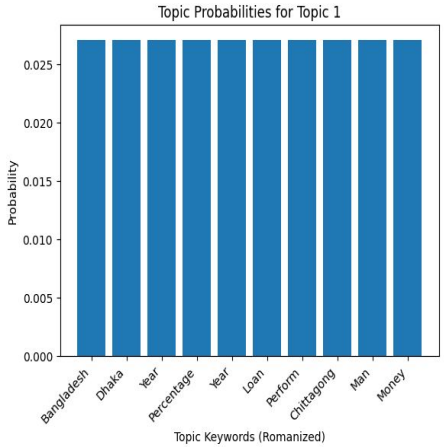
```
0.078*"বাংলাদেশ" + 0.068*"চট্টগ্রাম" +
0.068*"বছর" + 0.068*"বছরের" +
0.057*"শতাংশ" + 0.057*"বিএনপির" +
0.057*"হয়ে" + 0.046*"সংখ্যা" + 0.046*"
সবচেয়ে" + 0.046*"মানুষ"
```

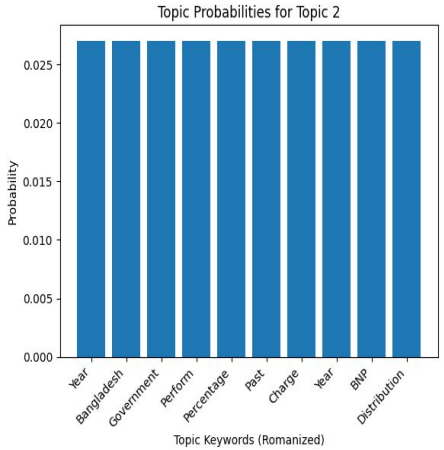
What is the most popular topic during the peak user hour of this week?

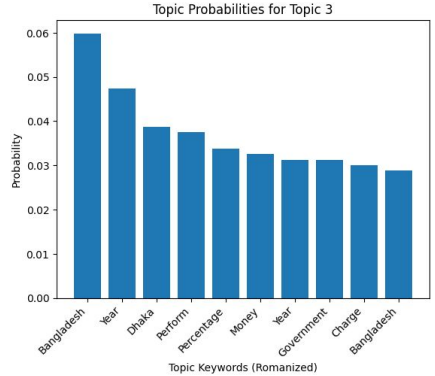


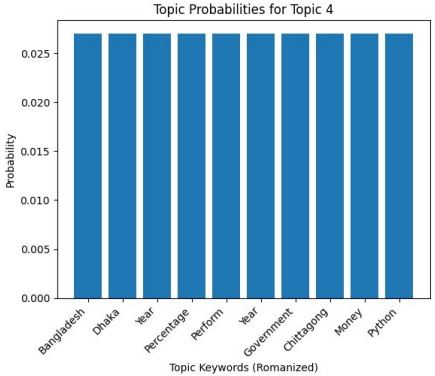
What is the most popular topic during the peak user hour of this week? (continued)











Most trending topic of the day is Topic: 3 with probabilities :

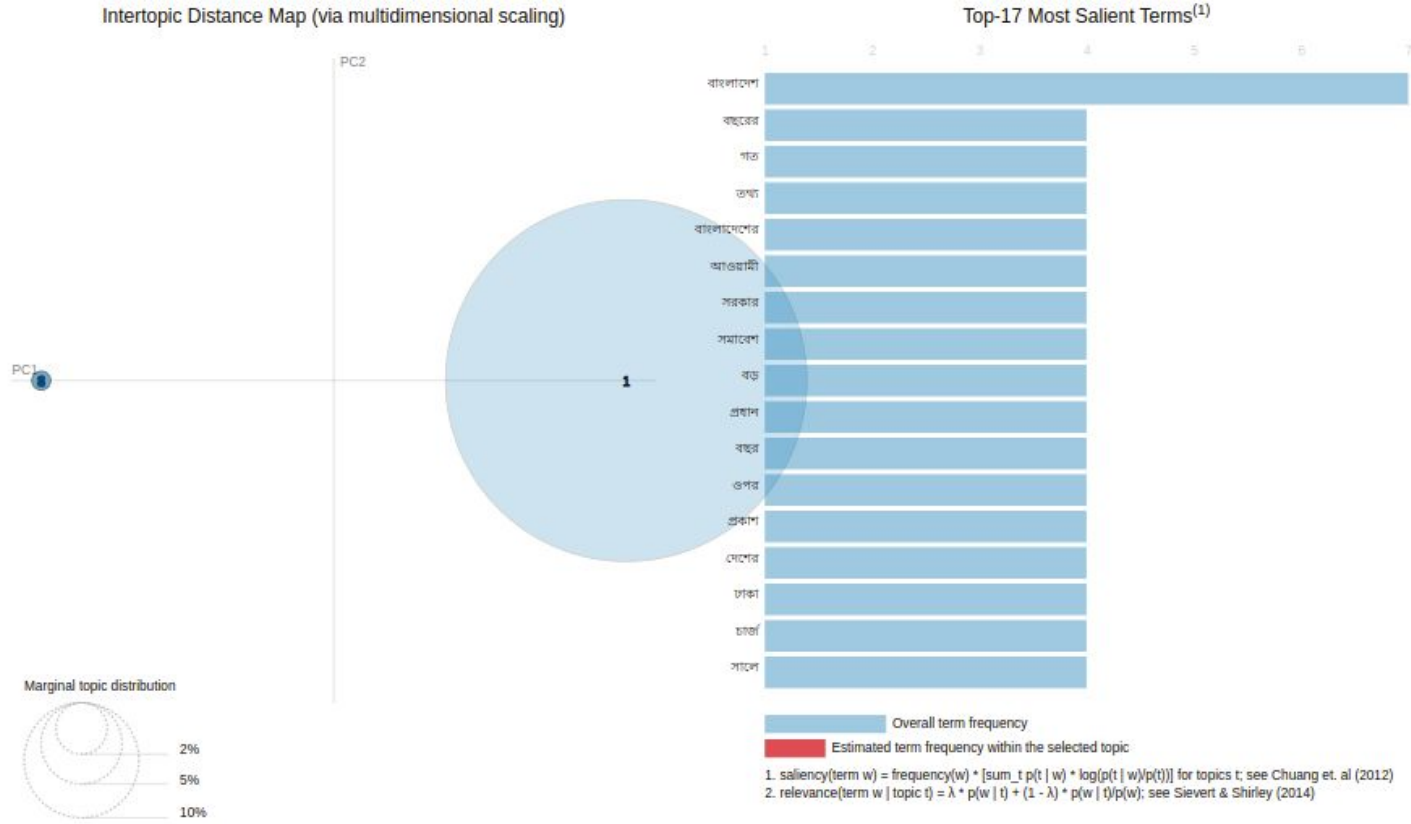
```
0.060*"বাংলাদেশ" + 0.047*"বছর" +
0.039*"ঢাকা" + 0.037*"সঞ্চালন" +
0.034*"শতাংশ" + 0.033*"টাকা" +
0.031*"বছরের" + 0.031*"সরকার" +
0.030*"চার্জ" + 0.029*"বাংলাদেশের"
```

Off Peak User Hour News

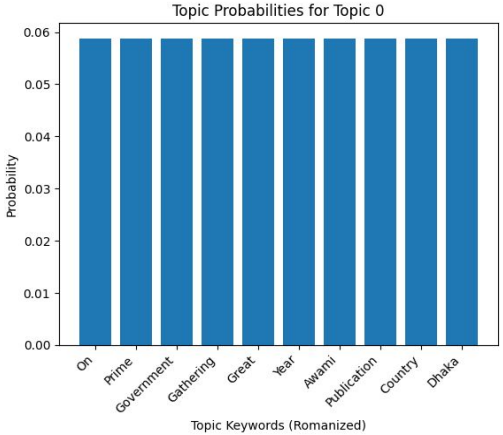
Kazi Shadman Sakib
4th Year Undergraduate Student

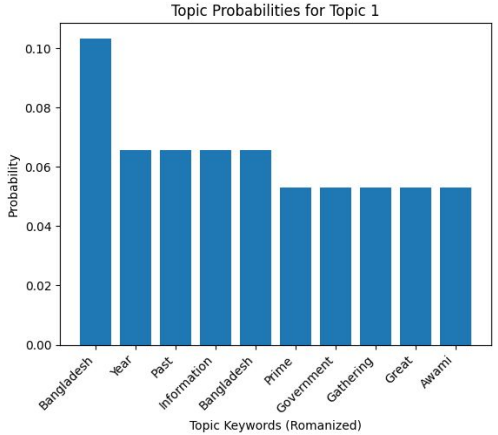
Department of Computer Science & Engineering
University of Dhaka

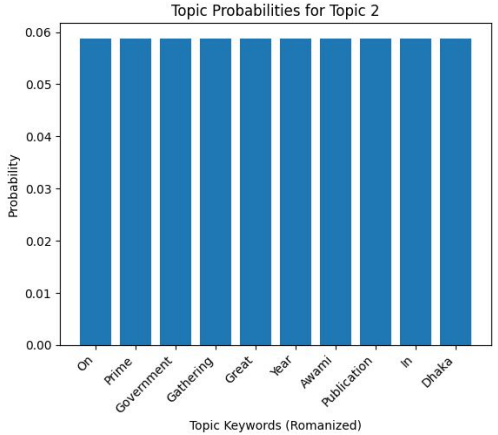
What was the most popular topic during the off peak user hour on July 24, 2023?

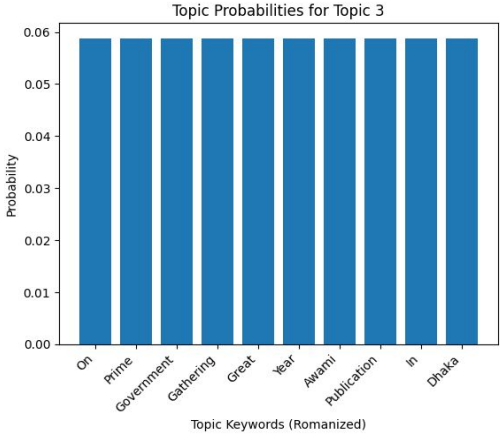


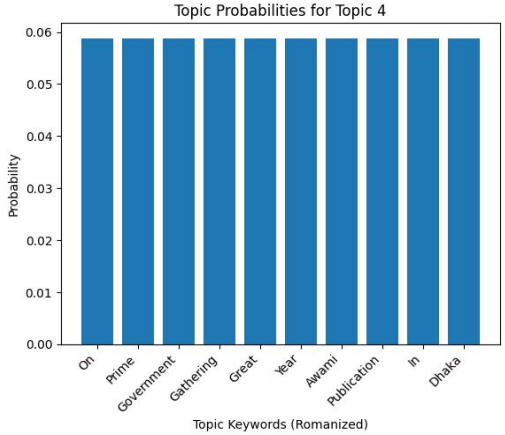
What was the most popular topic during the off peak user hour on July 24, 2023? (Continued)







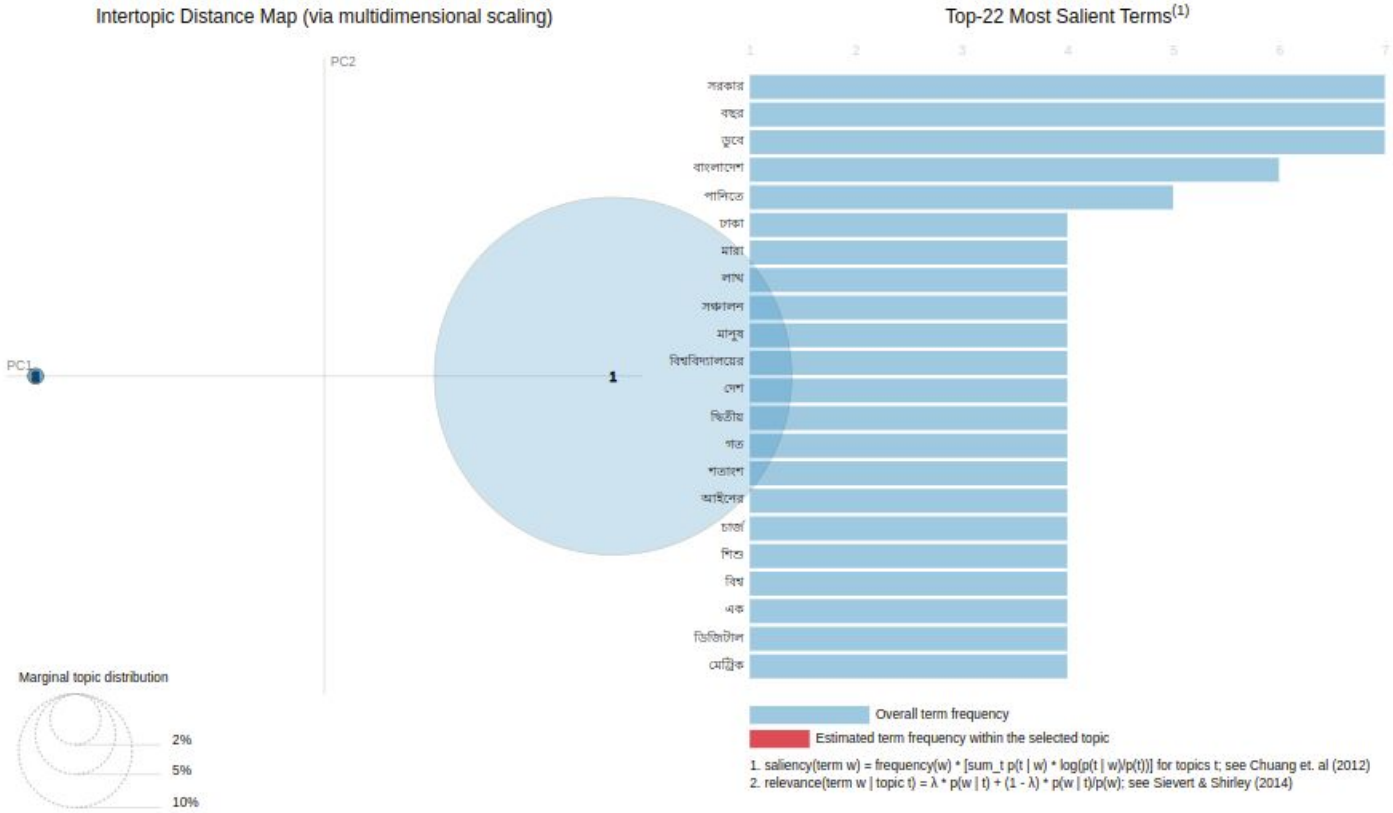




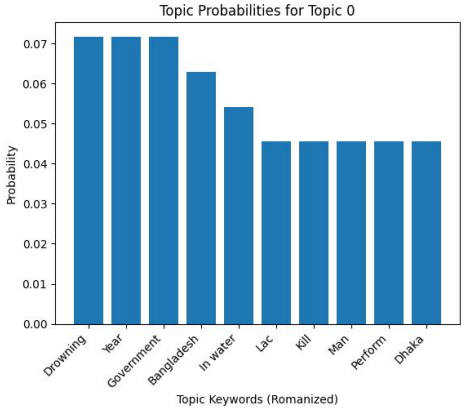
Most trending topic of the day is Topic: 1 with probabilities :

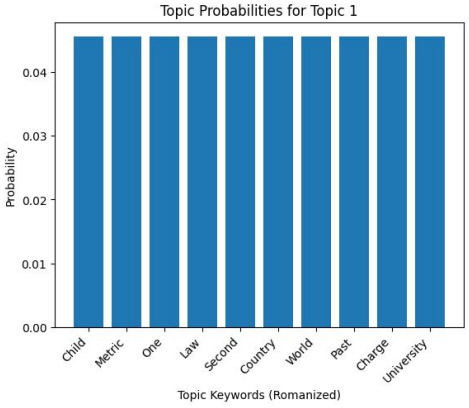
0.103*"বাংলাদেশ" + 0.065*"বছরের" + 0.065*"গত" + 0.065*"তথ্য" + 0.065*"বাংলাদেশের" + 0.053*"প্রধান" + 0.053*"সরকার" + 0.053*"সমাবেশ" + 0.053*"বড়" + 0.053*"আওয়ামী"

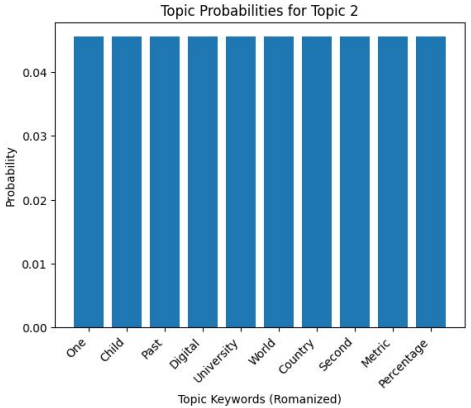
What was the most popular topic during the off peak user hour on July 25, 2023?

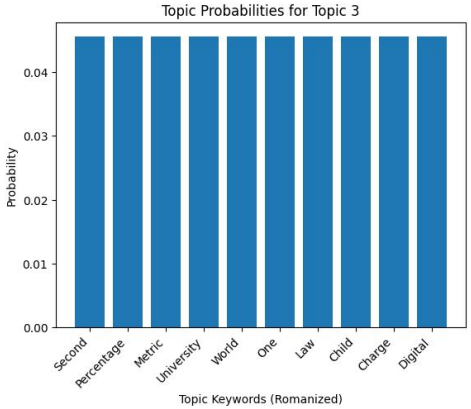


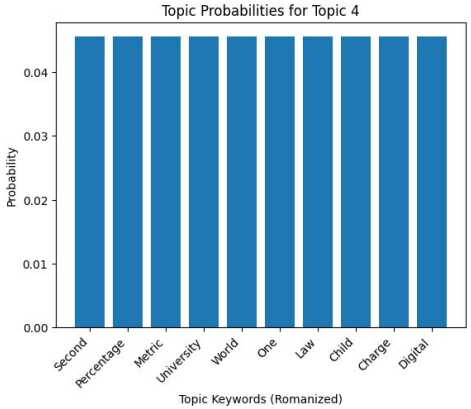
What was the most popular topic during the off peak user hour on July 25, 2023? (Continued)







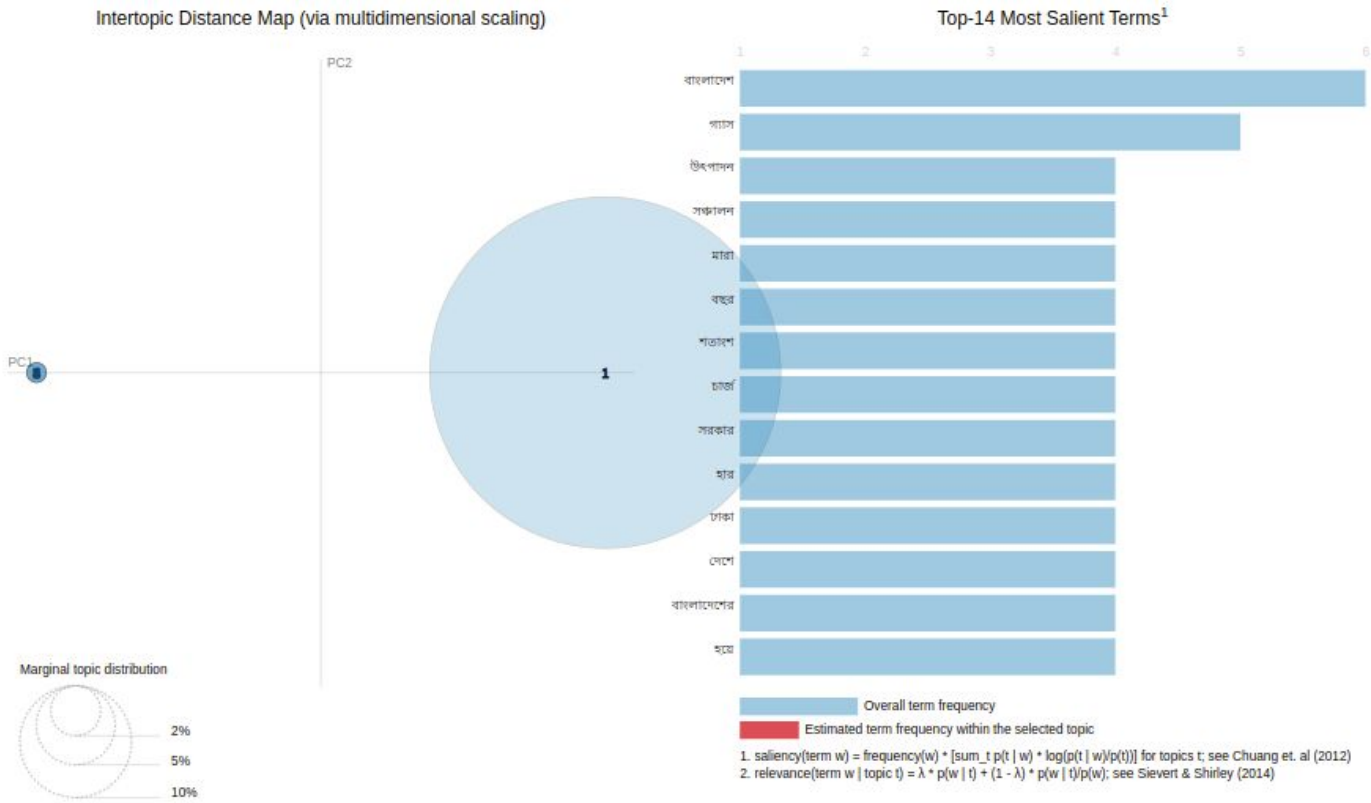




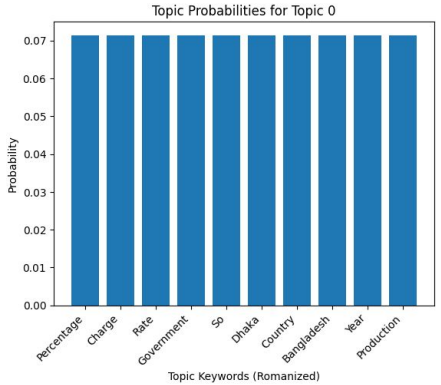
Most trending topic of the day is Topic: 0 with probabilities :

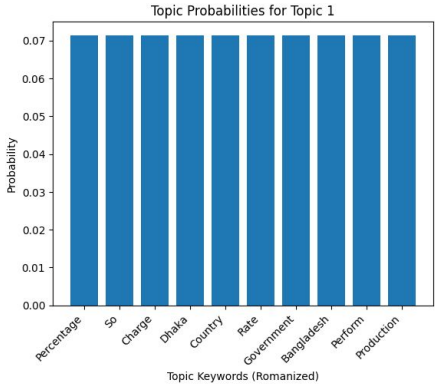
```
0.072*"ডুবে" + 0.072*"বছর" + 0.072*"
সরকার" + 0.063*"বাংলাদেশ" + 0.054*"
পানিতে" + 0.045*"লাখ" + 0.045*"মারা" +
0.045*"মানুষ" + 0.045*"সঞ্চালন" +
0.045*"ঢাকা"
```

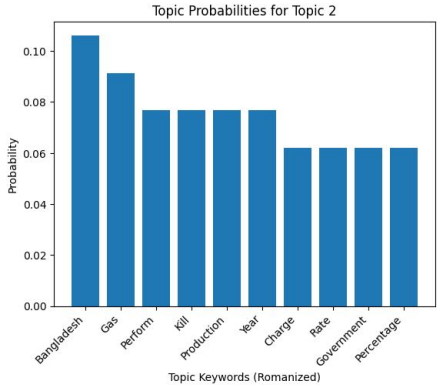
What was the most popular topic during the off peak user hour on July 26, 2023?

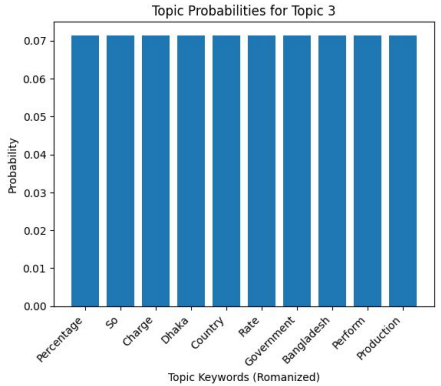


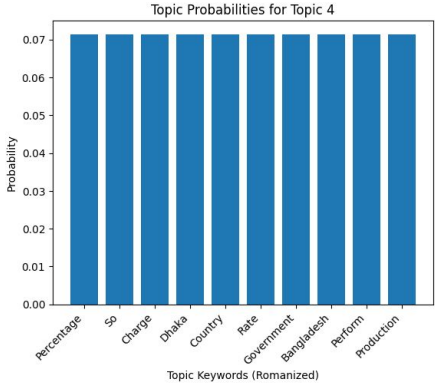
What was the most popular topic during the off peak user hour on July 26, 2023? (Continued)







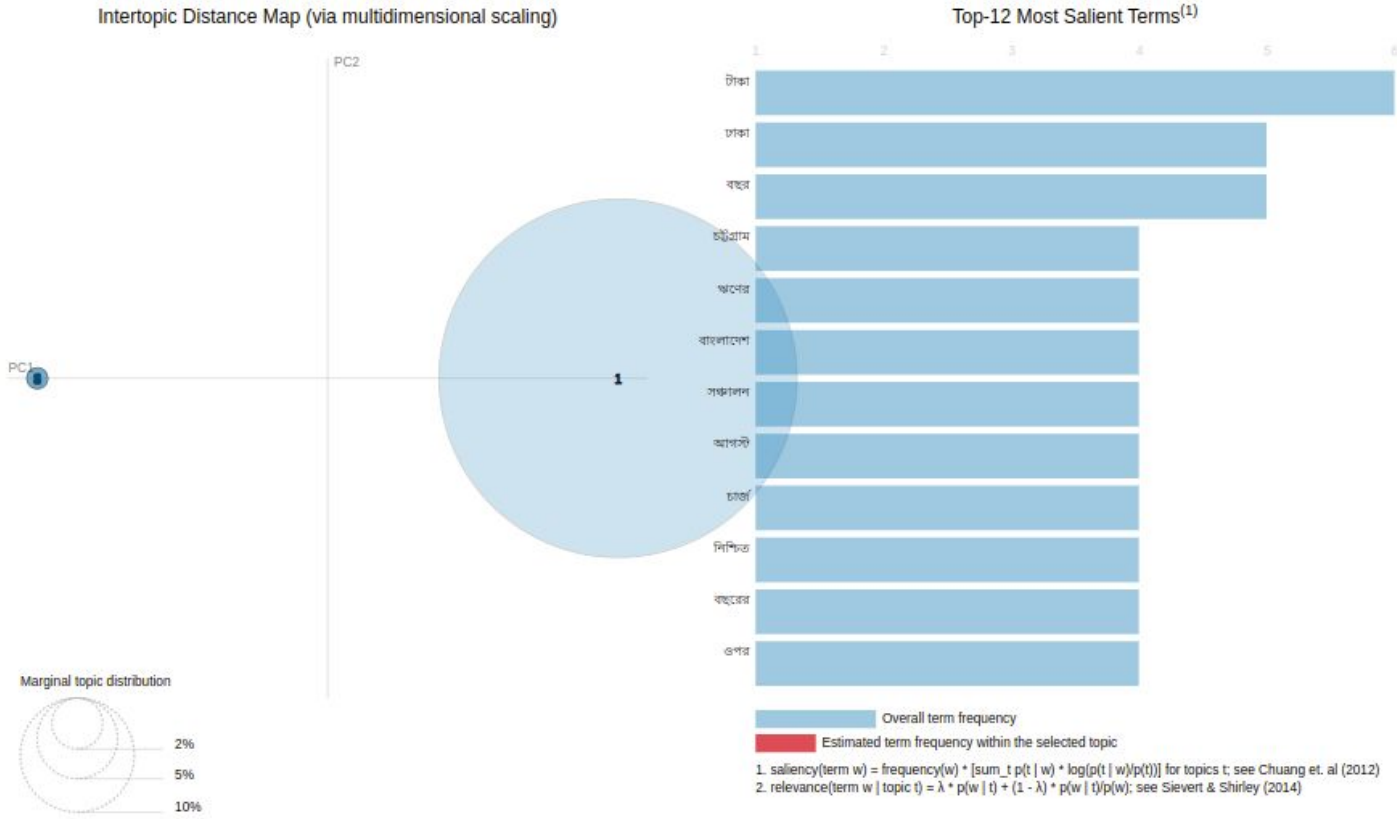




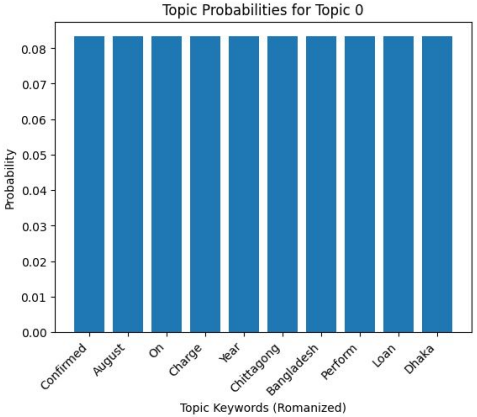
Most trending topic of the day is Topic: 2 with probabilities :

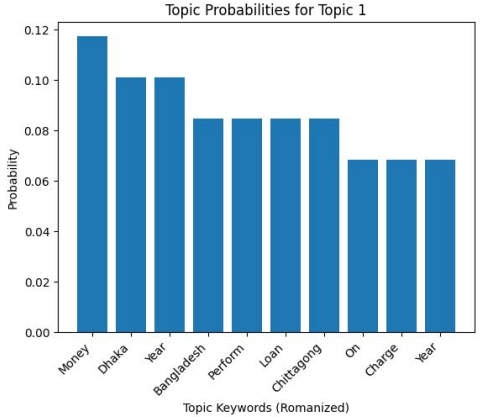
0.106*"বাংলাদেশ" + 0.091*"গ্যাস" + 0.077*"সঞ্চালন" + 0.077*"মারা" + 0.077*"উৎপাদন" + 0.077*"বছর" + 0.062*"চার্জ" + 0.062*"হার" + 0.062*"সরকার" + 0.062*"শতাংশ"

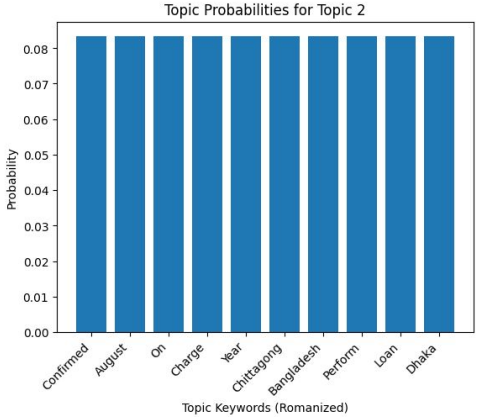
What was the most popular topic during the off peak user hour on July 27, 2023?

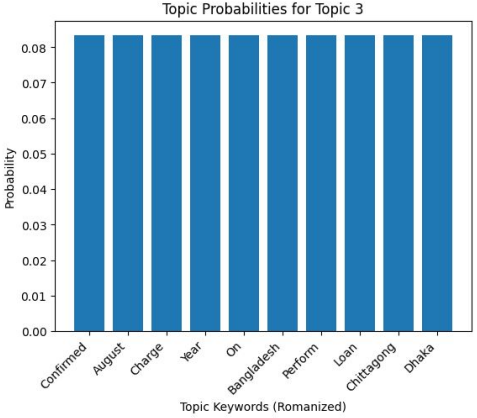


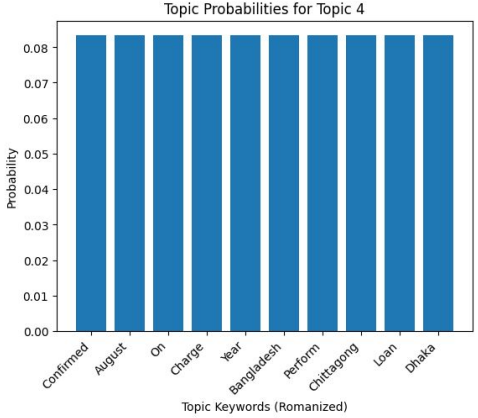
What was the most popular topic during the off peak user hour on July 27, 2023? (Continued)







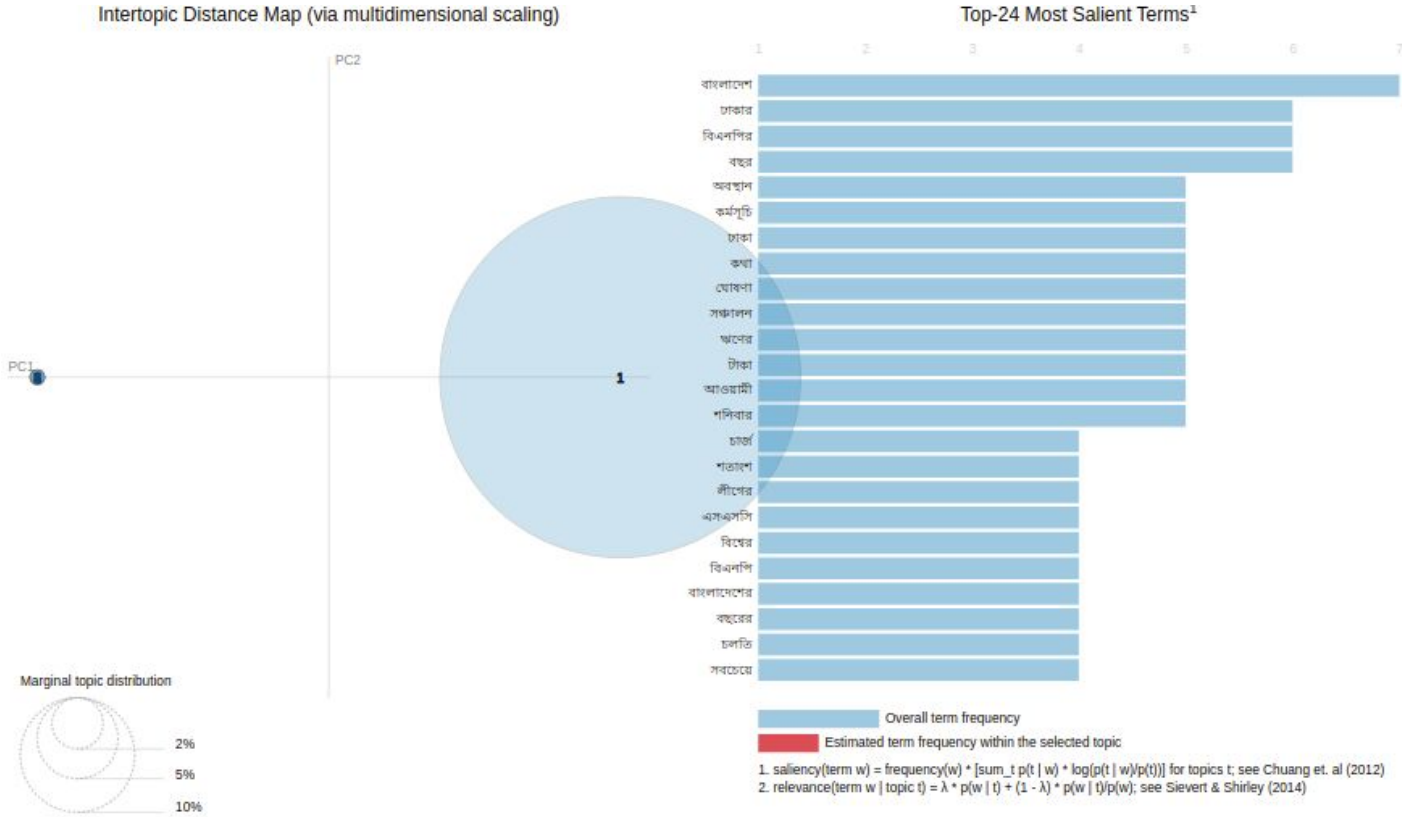




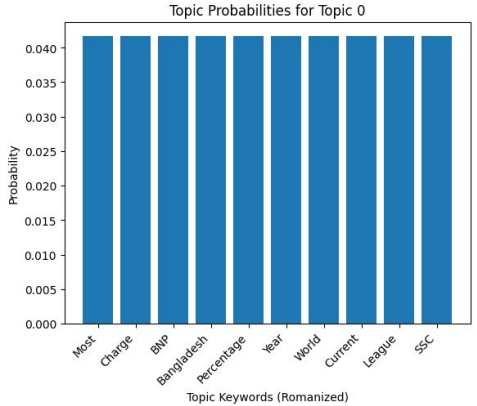
Most trending topic of the day is Topic: 1 with probabilities :

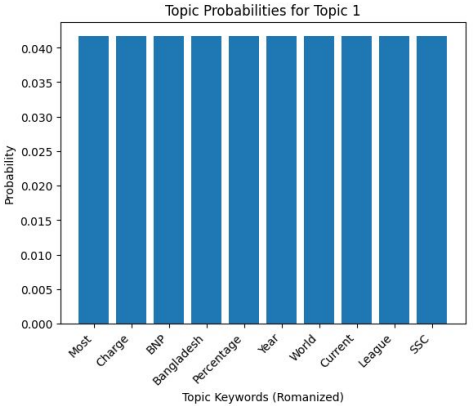
```
0.117*"টাকা" + 0.101*"ঢাকা" + 0.101*"
বছর" + 0.085*"বাংলাদেশ" + 0.085*"
সঞ্চালন" + 0.085*"ঋণের" + 0.085*"চট্টগ্রাম"
+ 0.068*"ওপর" + 0.068*"চার্জ" +
0.068*"বছরের"
```

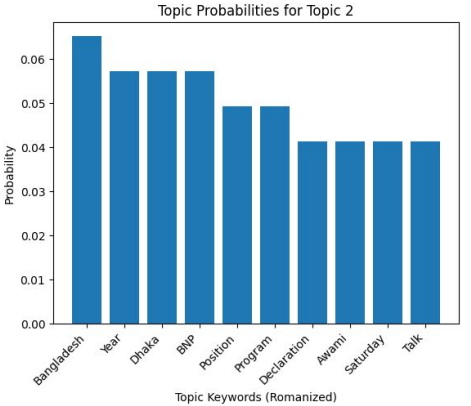
What was the most popular topic during the off peak user hour on July 28, 2023?

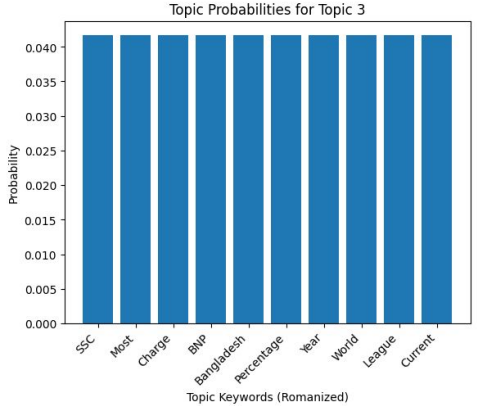


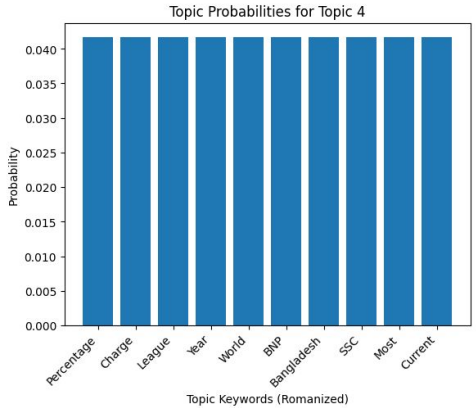
What was the most popular topic during the off peak user hour on July 28, 2023? (Continued)







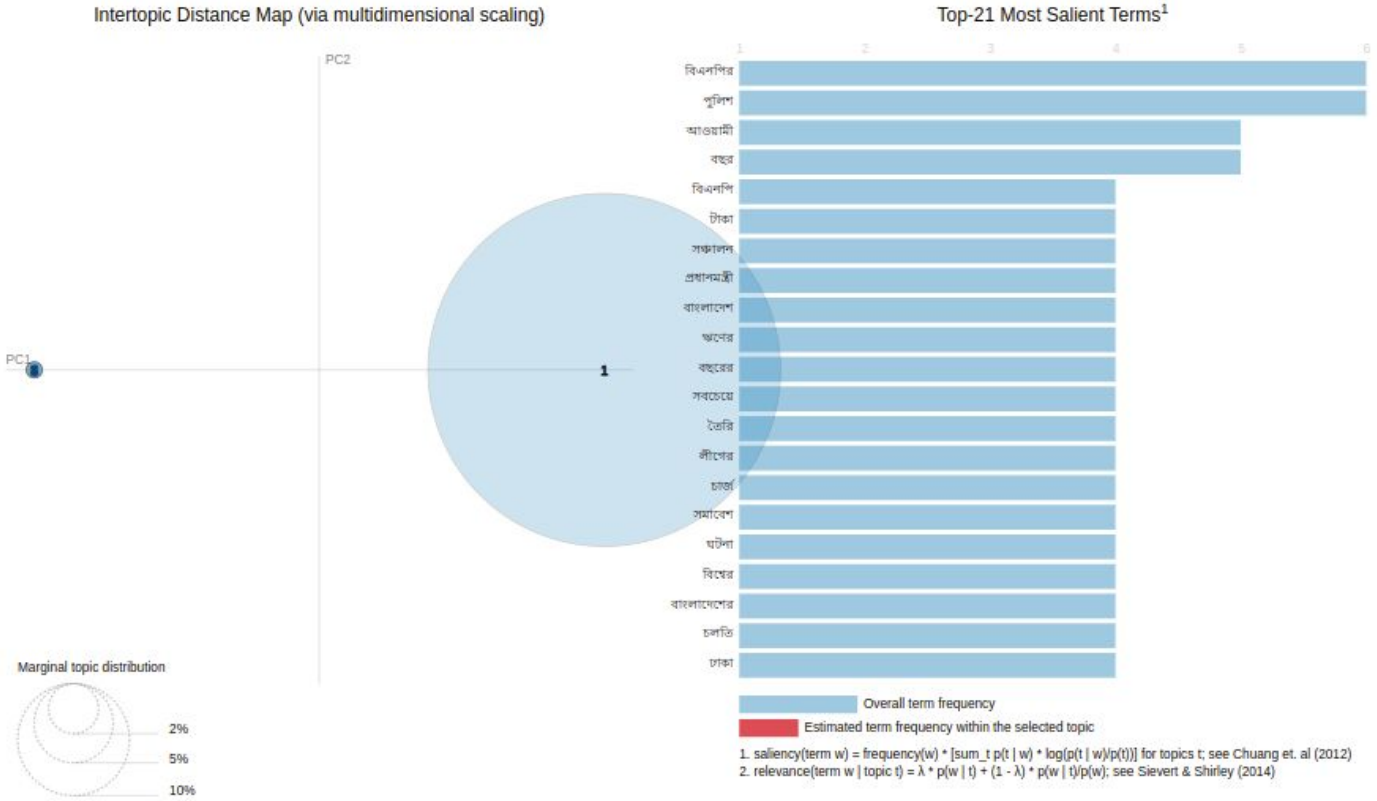




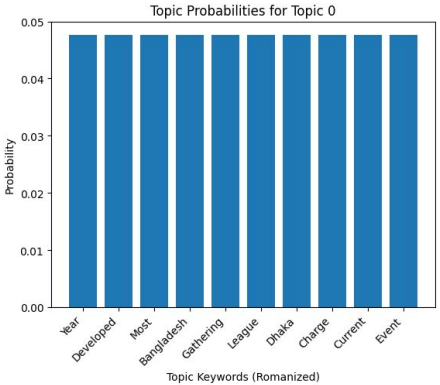
Most trending topic of the day is Topic: 2 with probabilities :

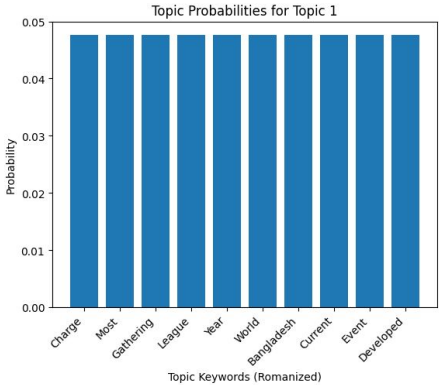
```
0.065*"বাংলাদেশ" + 0.057*"বছর" +
0.057*"ঢাকার" + 0.057*"বিএনপির" +
0.049*"অবস্থান" + 0.049*"কর্মসূচি" +
0.041*"ঘোষণা" + 0.041*"আওয়ামী" +
0.041*"শনিবার" + 0.041*"কথা"
```

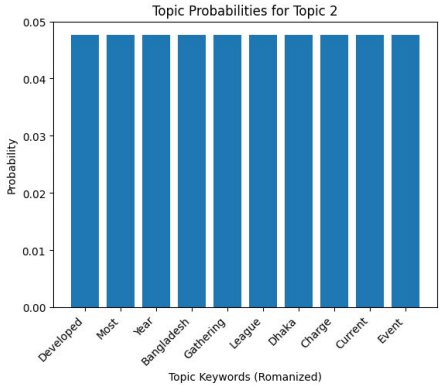
What was the most popular topic during the off peak user hour on July 29, 2023?

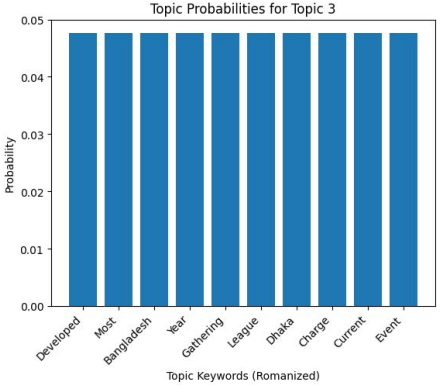


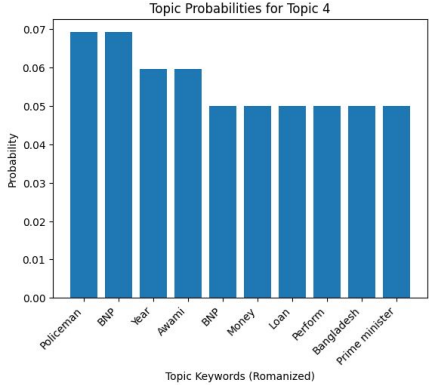
What was the most popular topic during the off peak user hour on July 29, 2023? (Continued)







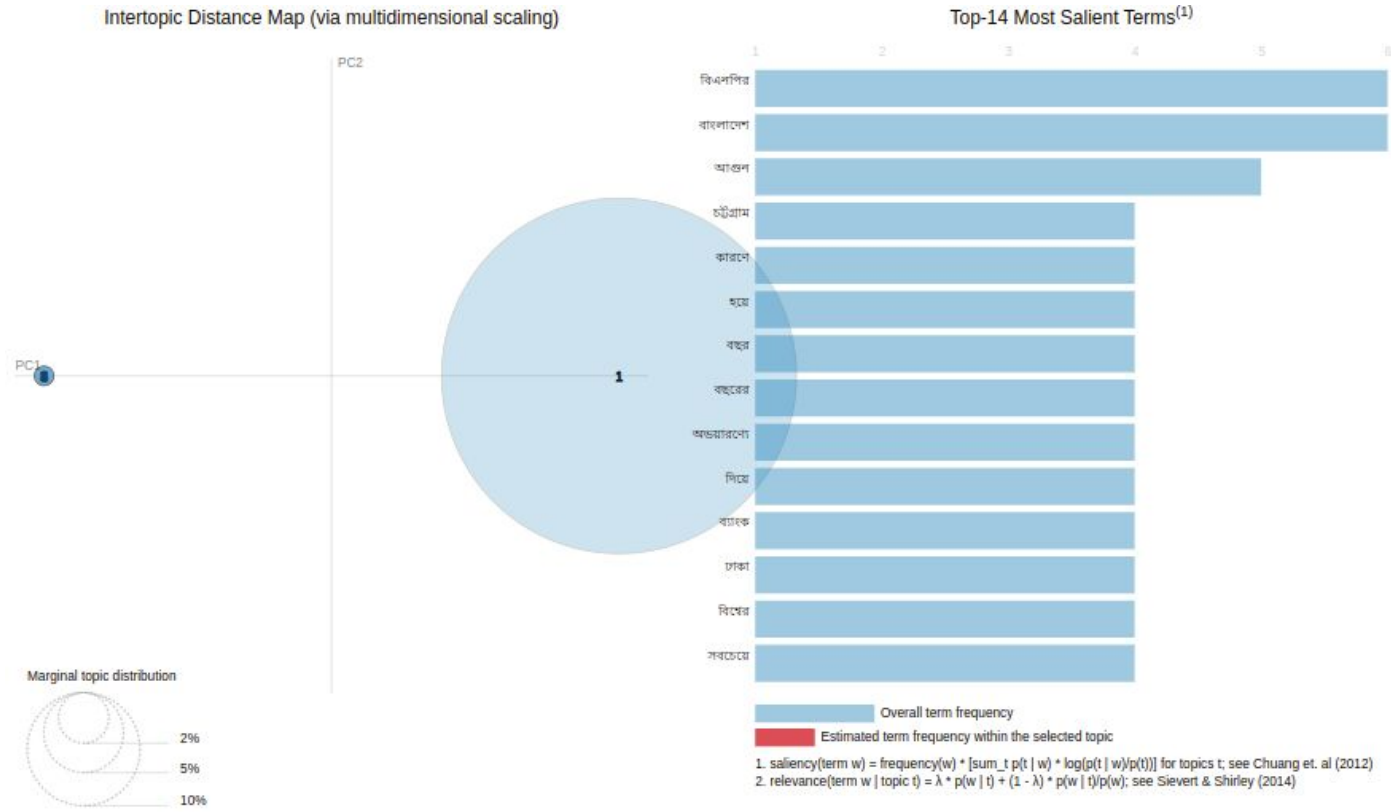




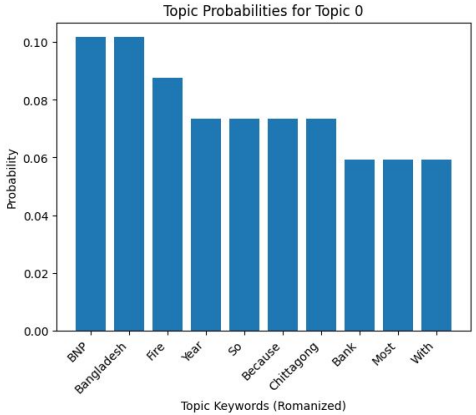
Most trending topic of the day is Topic: 4 with probabilities :

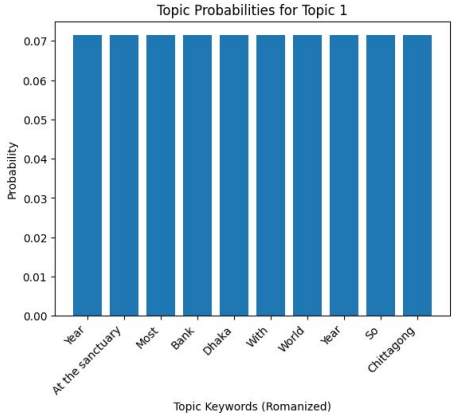
```
0.069*"পুলিশ" + 0.069*"বিএনপির" +
0.060*"বছর" + 0.060*"আওয়ামী" +
0.050*"বিএনপি" + 0.050*"টাকা" +
0.050*"ঋণের" + 0.050*"সঞ্চালন" +
0.050*"বাংলাদেশ" + 0.050*"প্রধানমন্ত্রী"
```

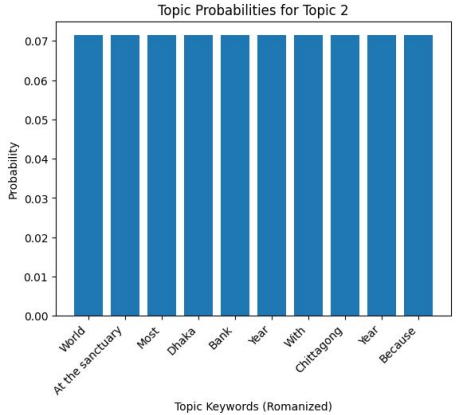
What was the most popular topic during the off peak user hour on July 30, 2023?

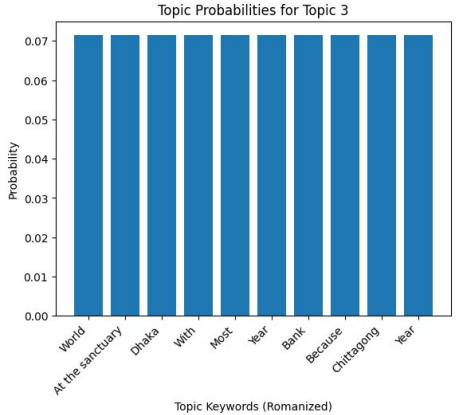


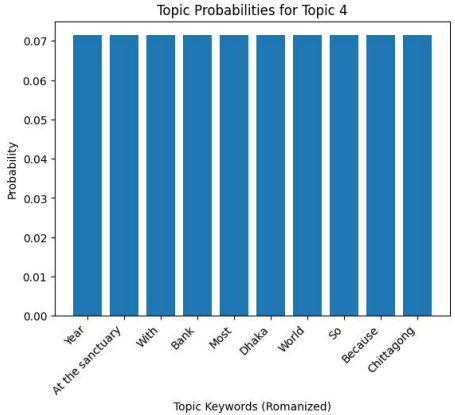
What was the most popular topic during the off peak user hour on July 30, 2023? (Continued)







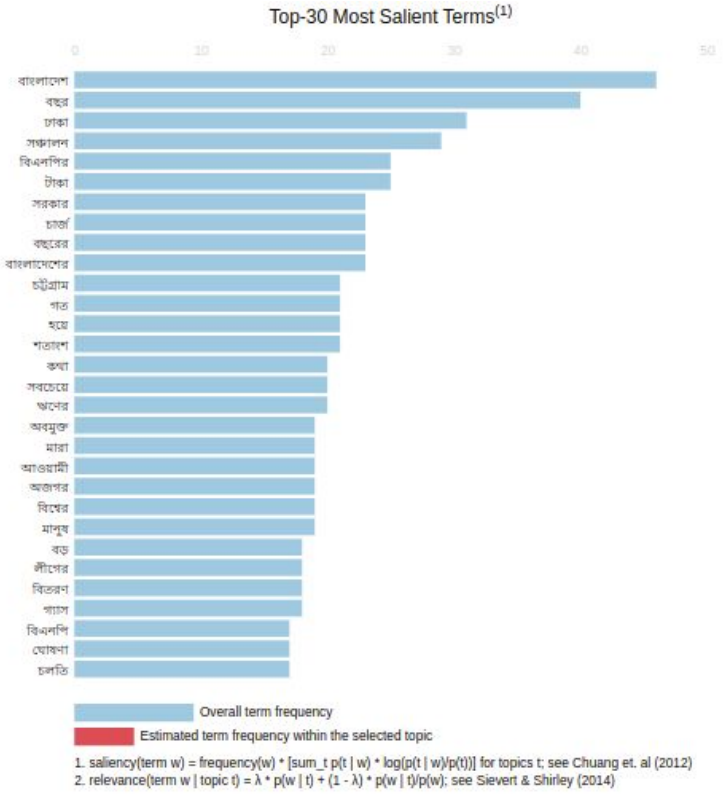




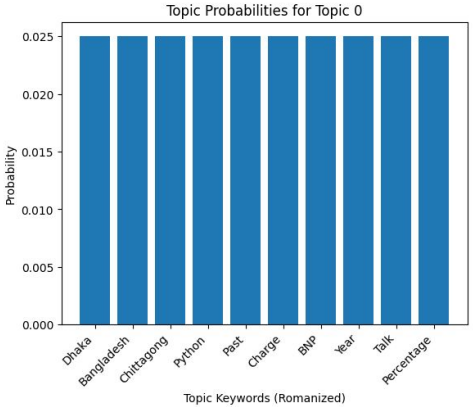
Most trending topic of the day is Topic: 0 with probabilities :

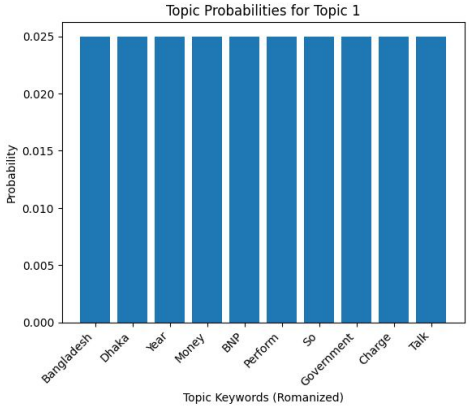
```
0.102*"বিএনপির" + 0.102*"বাংলাদেশ" +
0.088*"আগুন" + 0.073*"বছর" +
0.073*"হয়ে" + 0.073*"কারণে" + 0.073*"
চট্টগ্রাম" + 0.059*"ব্যাংক" + 0.059*"
সবচেয়ে" + 0.059*"দিয়ে"
```

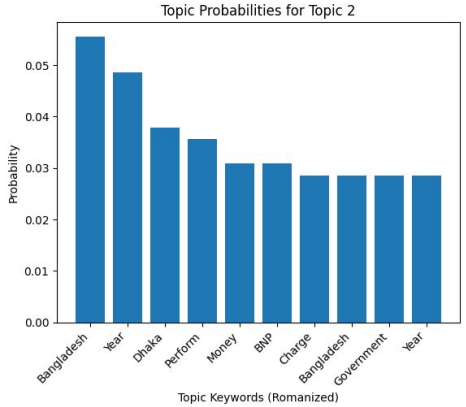
What is the most popular topic during the off peak user hour of this week?

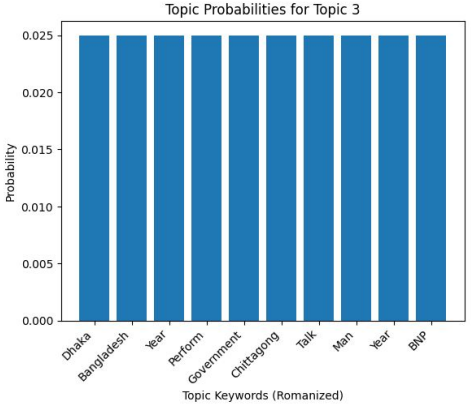


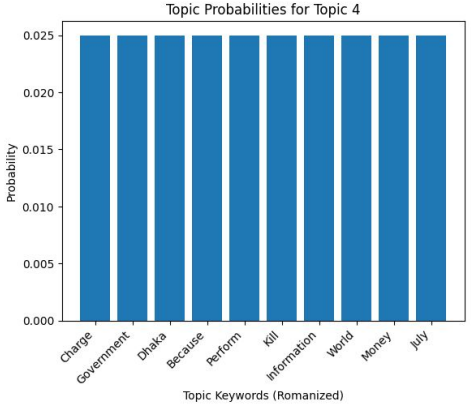
What is the most popular topic during the off peak user hour of this week? (continued)











Most trending topic of the day is Topic: 2 with probabilities :

```
0.056*"বাংলাদেশ" + 0.049*"বছর" +
0.038*"ঢাকা" + 0.036*"সঞ্চালন" +
0.031*"টাকা" + 0.031*"বিএনপির" +
0.029*"চার্জ" + 0.029*"বাংলাদেশের" +
0.029*"সরকার" + 0.029*"বছরের"
```

Descriptions

Kazi Shadman Sakib
4th Year Undergraduate Student

Department of Computer Science & Engineering
University of Dhaka

Dataset Description

The datasets were extracted from The Daily Star **Bangla** website using Python scripts.

Daily Datasets (Each Day's News): These datasets contain news from different topics for each day, ranging from **26 to 31** news items. The news topics vary from day to day. This variation in topics and the number of news items per day contributes to the **heterogeneity** of these datasets. Each dataset exhibited a size variation spanning from **10 kB to 26 kB**.

Weekly Dataset (7 Days' News): This dataset combines news from all 7 days, encompassing a broader range of topics and potentially a larger number of news items. The mixture of news topics across the week further adds to the **heterogeneity** of this dataset. Each dataset exhibited a size variation spanning from **76 kB to 175 kB**.

Model Description

Data Collection and Preprocessing :

- The Bengali stemmer function (**'bengali_stem(word)'**) is defined to remove common suffixes from words for stemming.
- The **'normalize(text)'** function is defined to preprocess and normalize the text data, converting it to lowercase and tokenizing it while removing stopwords and numerical values.
- The code reads the Bengali dataset from a file and applies text normalization to the entire text content.

Topic Modeling with LDA:

- The normalized text is tokenized, and a dictionary is created from the tokens using the **corpora.Dictionary** class.
- A bag of words representation (corpus) is generated using the created dictionary for the LDA model input.
- The **perform_lda_topic_modeling** function is defined to execute LDA topic modeling using the **Gensim** library. The function takes the corpus, dictionary, number of topics, and number of passes as parameters.
- Inside the **perform_lda_topic_modeling** function, the LDA model is trained on the provided corpus and dictionary using the **gensim.models.LdaModel** class. The top keywords with probabilities for each topic are printed using **lda_model.show_topics**.

Model Description (Continued)

After extensive experimentation and parameter tuning, I have refined the following predefined variables to optimize the performance of the Latent Dirichlet Allocation (LDA) model for topic modeling in the context of Bengali news articles:

- **num_topics = 5**: This predefined variable represents the number of distinct topics that the LDA model aims to identify within the dataset. Through careful adjustment, I have settled on five topics, ensuring a balance between granularity and coherence in the extracted themes.
- **passes = 80**: The predefined variable passes is indicative of the number of iterations employed during the convergence process of the LDA model. By setting it to 80, I have established a convergence threshold that enables the model to sufficiently capture the underlying topic patterns while avoiding overfitting.
- **num_words = 10**: This predefined variable controls the quantity of top keywords associated with each identified topic that are displayed. After iterative experimentation, I have chosen to display 10 keywords per topic. This value facilitates meaningful insights while maintaining concise topic representations.

Model Performance

The LDA model has successfully processed the dataset of Bengali news articles from The Daily Star Bangla and identified **five distinct topics** based on the analysis of the text content. Each topic is represented by a set of top keywords along with their corresponding probabilities. The associated probabilities indicate the relevance of each keyword within its respective topic.

The model's performance in conducting topic extraction from Bengali news articles **showcases its efficacy in identifying distinct subject clusters**. Furthermore, it has the capability to retrieve relevant and frequently discussed topics along with their corresponding keywords for each dataset. This applies equally to distinct datasets encompassing daily news, as well as separate datasets focusing on peak user hour news, off-peak user hour news, weekly news, peak user hour weekly news, and off-peak user hour weekly news.

Model Performance (Continued)

Evaluation of Model Performance across Three Distinct Datasets :

	Weekly News	Peak User Hour Weekly News	Off Peak User Hour Weekly News
Best Topic	4	3	2
Upper bound of the most relevant topics based on keyword probabilities	0.131, 0.131, 0.092, 0.085, 0.082, 0.082, 0.075, 0.075, 0.064, 0.061	0.060, 0.047, 0.039, 0.037, 0.034, 0.033, 0.031, 0.031, 0.030, 0.029	0.056, 0.049, 0.038, 0.036, 0.031, 0.031, 0.029, 0.029, 0.029, 0.029
Lower bound of the most relevant topics based on keyword probabilities	0.083, 0.083, 0.083, 0.083, 0.083, 0.083, 0.083, 0.083, 0.083, 0.083	0.027, 0.027, 0.027, 0.027, 0.027, 0.027, 0.027, 0.027, 0.027, 0.027	0.025, 0.025, 0.025, 0.025, 0.025, 0.025, 0.025, 0.025, 0.025, 0.025

Findings

1. The findings suggest a pattern such as specific topics in the weekly dataset related to Awami League and BNP, discussed in the last 3-4 days, continue to have a notable impact on the overall dataset of the past 7 days. This influence is reminiscent of the final results, indicating a recurrent focus on discussions concerning Awami League and BNP over the span of the preceding week.

2. Notably, certain topics consistently experience heightened discussion and engagement throughout 7 days duration. These prominently recurring topics, accompanied by their associated keywords, are subsequently integrated into the final results (for all the datasets). This strategic inclusion of recurrently emphasized subjects sheds light on their enduring significance within the dataset.

3. The meticulously fine-tuned values of the predefined variables (num_topics, passes and num_words) enhance the model's ability to discern relevant and coherent topics from Bengali news articles, thereby contributing to an informed understanding of the dataset's content.

4. The emergence of redundant topics is a consequence of setting the predefined variable "num_topics" to a value exceeding 5. In cases where the predefined variable "passes" is set significantly below 80, the model convergence may not reach an optimal state. The presence of redundant keywords arises when the assigned value to "num_words" surpasses 10. This phenomenon can be attributed to the inherent data heterogeneity, which impacts the efficacy of topic extraction and the model's convergence behavior.

Thank You

Any Questions?

Kazi Shadman Sakib
4th Year Undergraduate Student

Department of Computer Science & Engineering
University of Dhaka