AUTOMATED TICKECT CLASSIFICATIO N

GROUP 2

BUSINESS UNDERSTANDING

Financial institutions receive thousands of customer complaints daily.

Manual complaint classification is slow and inefficient.

Customer complaints are submitted as unstructured text, requiring automation for faster processing.

Solution: Implement **Natural Language Processing (NLP)** to classify complaints automatically.

Objective: Automate complaint categorization using NLP & ML.

Business Objective

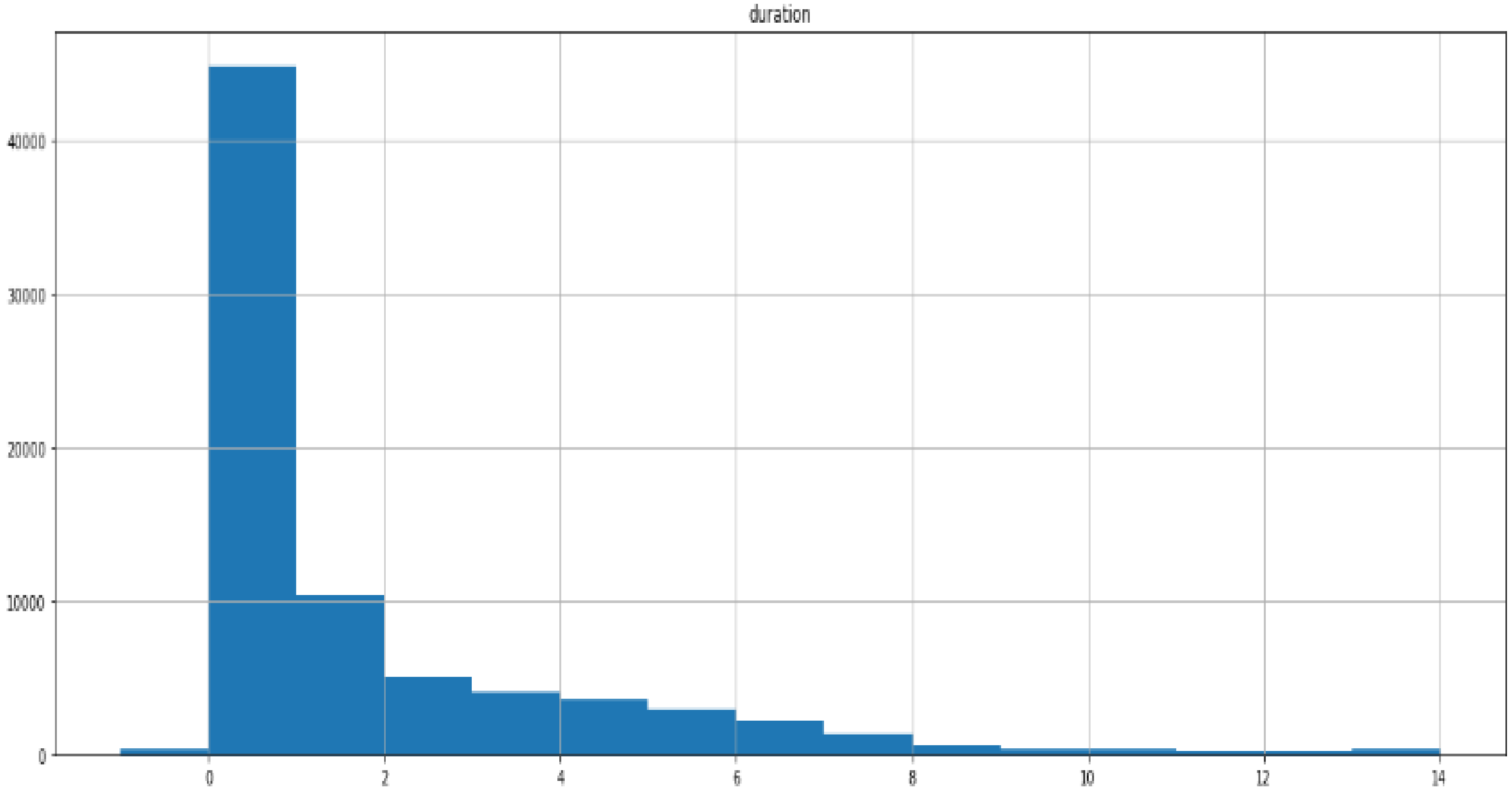
- ❖Develop an aoutomated ticket classification model for customer complaints.
- ❖Categorize complaints into 5 key service areas using Topic Modeling:
  1. Credit Card / Prepaid Card
  2. Bank Account Services
  3. Theft / Dispute Reporting
  4. Mortages / Loans
  5. Others
- ❖Train a supervised learning learning model(e.g.LSTM,)
- ❖Key Benefits:
  1. Faster complaint resolution.
  2. Improved customer experience
  3. Better resource allocation

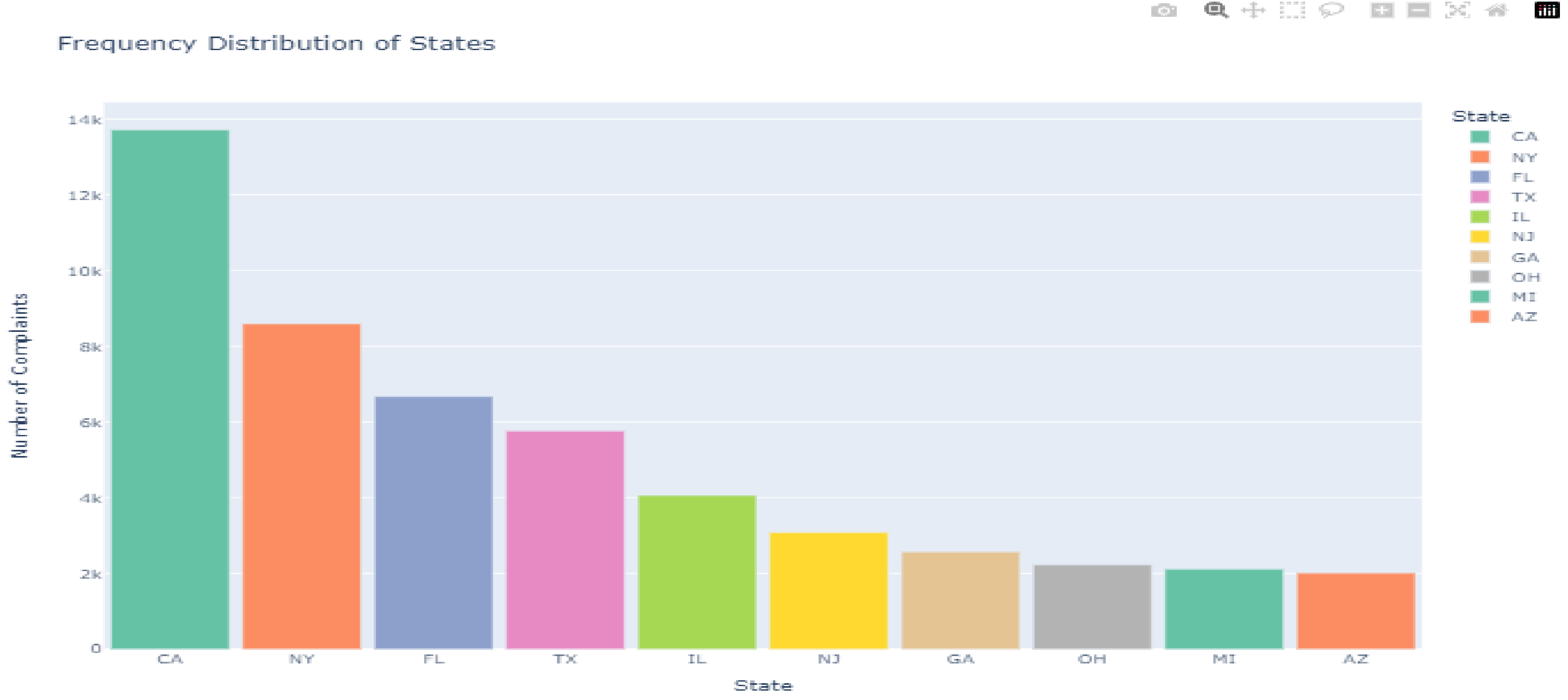
DATA PREPARATION

- **Handling Missing Data:**
  - Removed or imputed missing values to maintain data integrity.
- **Text Cleaning:**
  - Removed stopwords, punctuation, and special characters.
  - Converted text to lowercase for consistency.
- **Tokenization & Vectorization:**
  - Applied **TF-IDF** and **Word2Vec** to transform text into numerical representations.
- **Feature Engineering:**
  - Extracted key terms and phrases for better model performance.

Observations

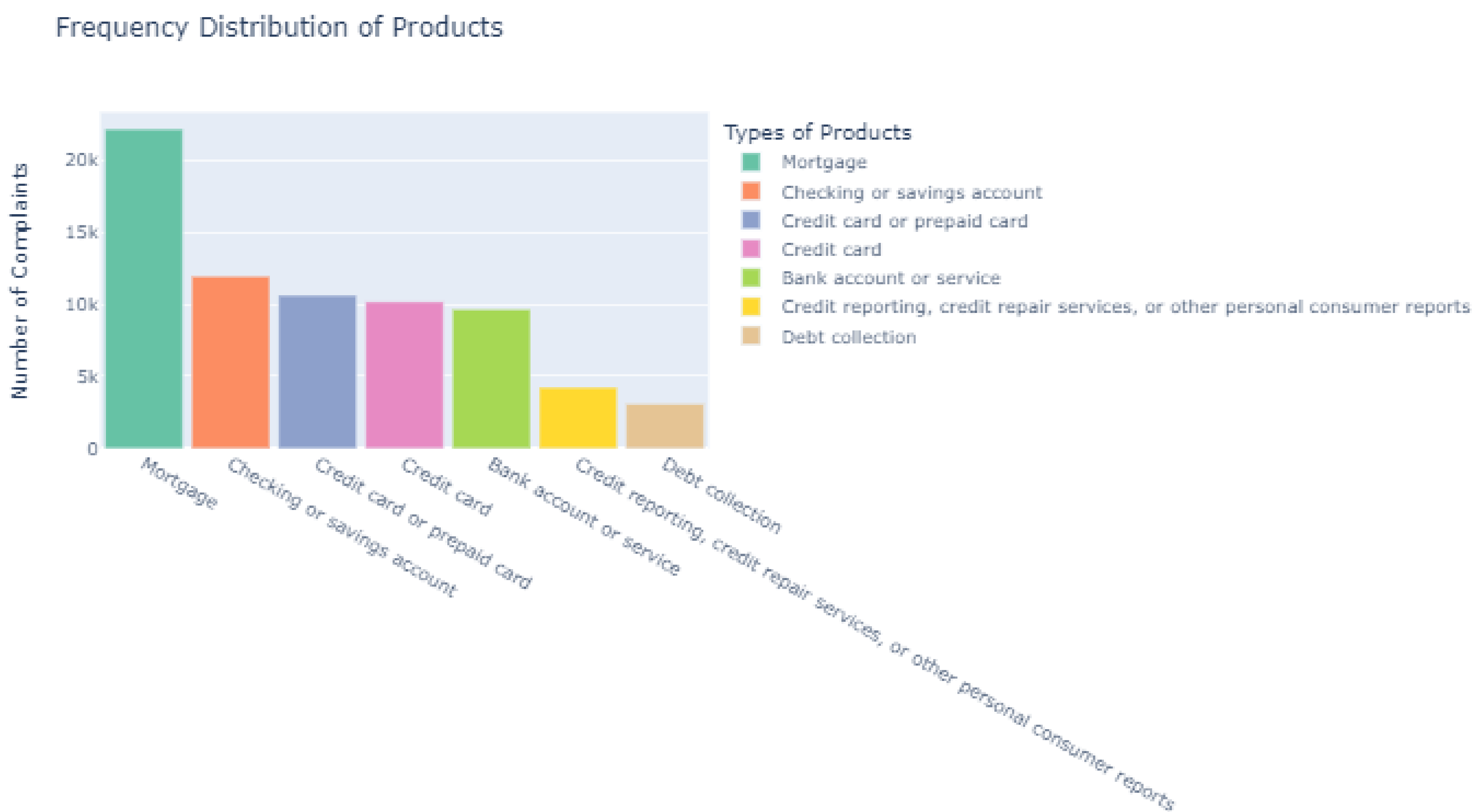
- The histogram shows a right-skewed distribution, with most complaints resolved within a few days (0-1) days. A small number of cases take significantly longer, faster complaints resolution.





Observations :

Most complaints originate from CA, NY, FL. These are the three states with most complains



Frequency Distribution of Products
Number of Complaints
0
5k
10k
15k
20k
Mortgage
Checking or savings account
Credit card or prepaid card
Credit card
Bank account or service
Credit reporting, credit repair services, or other personal consumer reports
Debt collection
Types of Products
Mortgage
Checking or savings account
Credit card or prepaid card
Credit card
Bank account or service
Credit reporting, credit repair services, or other personal consumer reports
Debt collection

Observation on Distribution of product

Mortgage-related complaints are the highest, exceeding 20,000 cases.

Checking/savings accounts, credit cards, and prepaid cards also have a significant number of complaints, each around 10,000.

Bank account services, credit reporting, and debt collection have fewer complaints in comparison.

The distribution is uneven, with mortgage-related issues being the most frequent concern.

MODELING APPROACH

- ❖ Data Preprocessing:
  - o Cleaned text data.
  - o Tokenized text and TF-IDF vectorization for feature selection.
- ❖ Unsupervised Learning: Topic Modeling (LDA & NMF)
- ❖ Supervised Learning: (LSTM)

Topic Modeling with NMF.

Topic 1: Bank Account Services

Topic 2: Credit Card / Prepaid Card

Topic 3: Mortgages/Loans

Topic 4: Theft/Dispute Reporting

Topic 5: Others

	Word 1	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7	Word 8	Word 9	Word 10
Topic 1	account	bank	check	money	fund	chase	wa	deposit	branch	day
Topic 2	credit	card	report	inquiry	chase	account	score	company	limit	bureau
Topic 3	loan	mortgage	home	chase	modification	property	year	wa	rate	letter
Topic 4	charge	card	chase	transaction	dispute	wa	fraud	claim	merchant	purchase
Topic 5	payment	balance	fee	month	statement	wa	day	time	date	auto

Topic Modeling with LDA

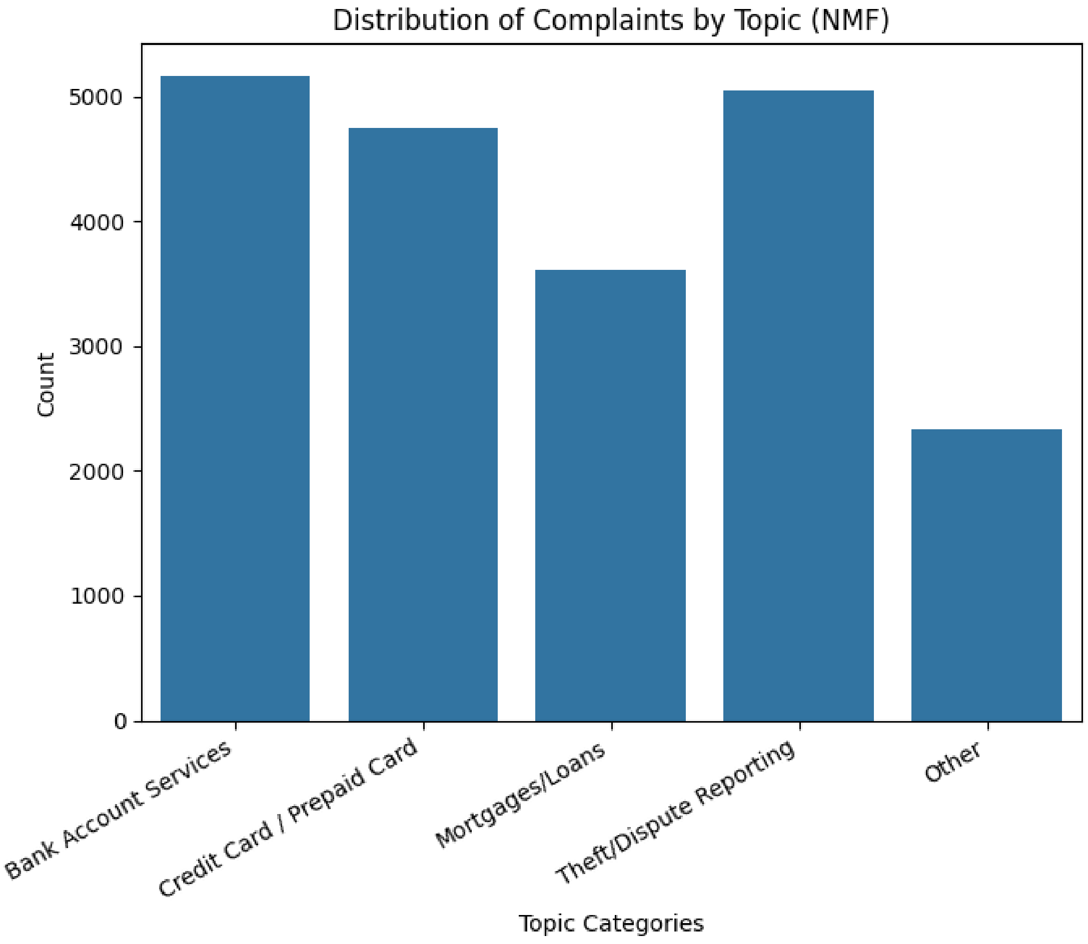
LDA Topics Identified:

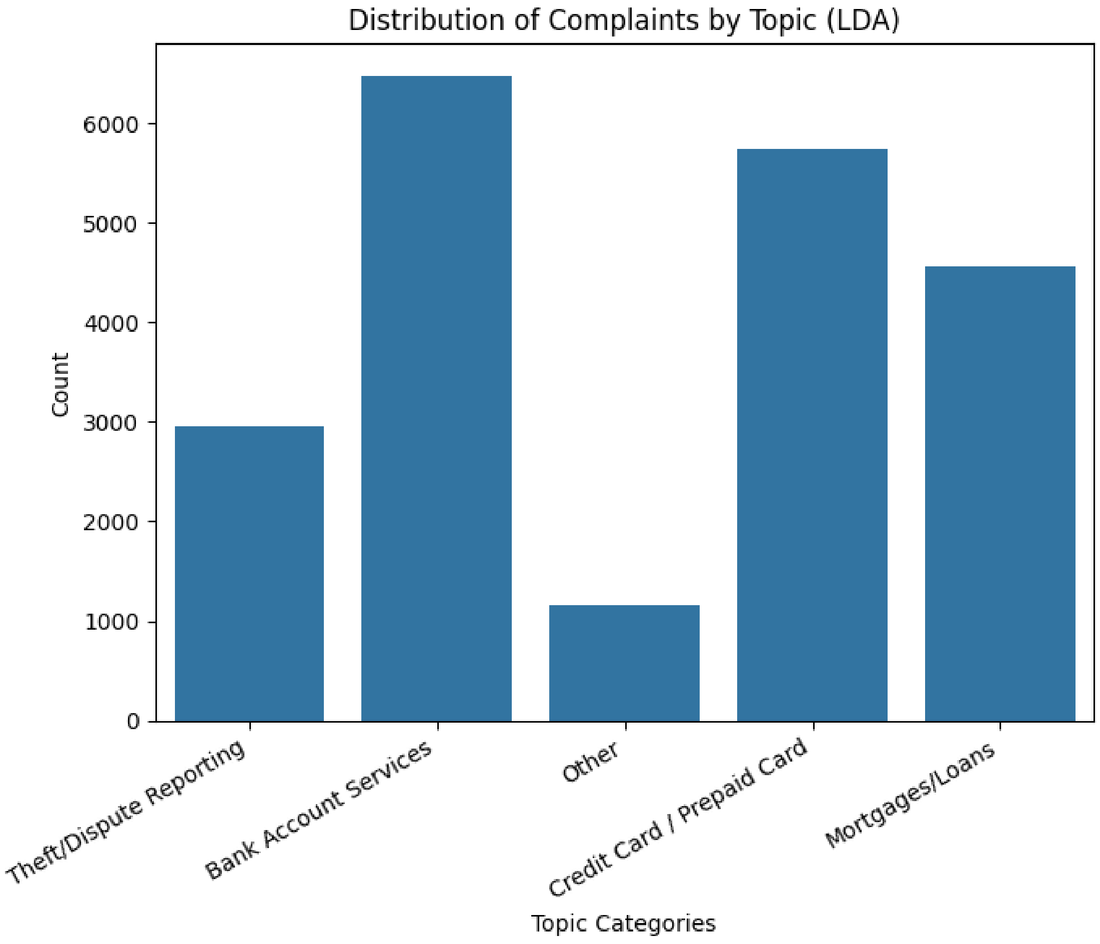
- Similar to NMF with slight variations
- **Main Difference:** "Others" category was less defined

Comparison: NMF had better topic separability

	Word 1	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7	Word 8	Word 9	Word 10
Topic 1	card	chase	charge	wa	credit	fraud	transaction	account	claim	number
Topic 2	loan	chase	mortgage	payment	wa	home	bank	time	year	property
Topic 3	account	bank	chase	wa	check	money	fund	day	branch	time
Topic 4	credit	card	account	chase	payment	balance	report	wa	month	time
Topic 5	chase	wa	dispute	fee	time	charge	car	day	merchant	company

DISTRIBUTION OF COMPAINTS BY LDA & NMF





Observations on Topic Consistencies Between NMF and LDA

From the two distribution plots, we can observe some consistencies despite the differences in the topic modeling approaches:

- Bank Account Services – This topic is the most dominant in both NMF and LDA, suggesting strong clustering around banking-related complaints.

- Credit Card / Prepaid Card – This category also appears as a significant topic in both models, showing consistent identification of credit card-related issues.

- Mortgages/Loans – Both models assign a comparable number of complaints to this topic, reflecting a shared pattern in complaint distribution.

- Theft/Dispute Reporting – While there is some variation in frequency, this category is clearly identified in both models.

- Other Category – The "Other" category differs in size between the models, but it still represents a smaller portion of the dataset in both cases.

Topic Consistency Observations

Similarities:

- Bank Account Services & Credit Cards were strong topics in both models
- Mortgages/Loans & Theft/Dispute categories were consistent

Differences:

- "Others" category was more ambiguous in LDA

Model perfomance Metrics

LSTM Classification Results:

NMF-Based Model: 90% Accuracy

LDA-Based Model: 88% Accuracy

Key Observations:

NMF-based model performed better due to clearer topic clusters

"Others" category had the lowest performance in both models

Model Evaluation

Precision: How many predicted complaints were correct

Overall Accuracy: 90%

Recall: How many actual complaints were identified correctly

F1-score: Balance between Precision & Recall

CONCLUSION

- **Best Approach:** NMF + LSTM for automated classification
- **Key Takeaways:**
  - NMF produced clearer topic separation, leading to better classification
  - LDA struggled with overlapping categories

RECOMENDATION

Use NMF for Topic Extraction (Better coherence, clearer clusters)

Improve "Others" Category Labeling (Refine definitions for better classification)

Deploy NMF-based BiLSTM Model (Automate complaint classification)

Implement Continuous Learning (Retrain model periodically for accuracy improvement)

Next Steps

Implementation Plan: Integrate the model into customer service workflows

Monitor Performance: Regular evaluation & retraining

Q&A: Open for discussion