

Project Milestone 2

Group 6: Search Wizards

Exploration

We downloaded the [eCFR Title 12](#) dataset and did some preliminary exploration on a [Colab notebook](#). First, we imported the dataset into Colab and inspected the first 5-10 records to get a general feel and understanding of the data and its structure. We evaluated these records and looked at both the text and metadata for any potential issues. Then we did other general best practices such as checking for null values.

Evaluation

- The data is very solid, there are no immediate or apparent red flags such as missing data.
- Each record represents a section in the eCFR Title 12 which we determined was the smallest meaningful part of the overall document. Each section is represented as a string of raw text and metadata associated with that section as a key-value pair. Within the metadata are more key-value pairs representing various attributes of the metadata, mainly hierarchical information, used to situate and locate the section. This may be very useful but might require extra work to “unflatten” and extract for each section **[1, 2, 3]**
- Potential challenges with the dataset:
 - **Complexity and length** - lengthy strings and complex legal jargon
 - **Hierarchy** - nested dictionaries may be difficult to work with when extracting and analyzing attributes for each section
 - **Formatting** - some sections have slight formatting issues when translated to raw text, such as subsections prefixed by commas, <code> and links.
- Additionally, we used BERTopic to take a look at topic modeling using the first 1000 rows of the dataset. We expected that the topics would roughly reflect the various chapters of eCFR Title 12, but found that it wasn't the case. Just checking the first 10 records we can see that various sections from Chapter 2 were assigned different classes **[4, 5]**

Questions

- What applications are we going to use with this dataset? What will we be predicting?
- How important is the metadata? Are some more important than others?

Next steps

Decide on what problem we want to solve, which model we want to use, and why.

[1] First 10 records of the dataset

	text	meta
0		{'chapter': ['II', 'II', 'II', 'II', 'II', 'II']...}
1	(a) (1) Compliance with the requirements of th...	{'chapter': ['II'], 'chapter_title': ['CHAPTER...']}
2	(a) <E T="04">Federal Register. The Board publ...	{'chapter': ['II'], 'chapter_title': ['CHAPTER...']}
3	(a) <E T="04">Federal Register. The Committee ...	{'chapter': ['II'], 'chapter_title': ['CHAPTER...']}
4	(a) A Bank shall make long-term advances only ...	{'chapter': ['XII'], 'chapter_title': ['CHAPTE...']}
5	(a) A Bank shall require each member to mainta...	{'chapter': ['XII'], 'chapter_title': ['CHAPTE...']}
6	(a) A Bank shall require the borrower to certi...	{'chapter': ['XII'], 'chapter_title': ['CHAPTE...']}
7	(a) A Board-regulated institution described in...	{'chapter': ['II'], 'chapter_title': ['CHAPTER...']}
8	(a) A Board-regulated institution that is an a...	{'chapter': ['II'], 'chapter_title': ['CHAPTER...']}
9	(a) A Farm Credit Bank or agricultural credit ...	{'chapter': ['VI'], 'chapter_title': ['CHAPTER...']}

[2] Example of text

(a) (i) Compliance with the requirements of this part shall be enforced under - , (1) Section 8 of the Federal Deposit Insurance Act, by the appropriate Federal banking agency, as defined in section 3(g) of the Federal Deposit Insurance Act (12 U.S.C. 1813(g)), with respect to - , (A) National banks, federal savings associations, and federal branches and federal agencies of foreign banks; (B) Member banks of the Federal Reserve System (other than national banks), branches and agencies of foreign banks (other than Federal branches, Federal Agencies, and insured state branches of foreign banks), commercial lending companies owned or controlled by foreign banks, and organizations operating under licenses issued pursuant to the International Banking Act of 1978 (12 U.S.C. 1861-1866); (C) State member banks of the Federal Reserve System and state branches of foreign banks; (ii) The Federal Credit Union Act (12 U.S.C. 1751 et seq.), by the Administrator of the National Credit Union Administration (National Credit Union Administration Board with respect to any Federal credit union); (iii) The Federal Aviation Act of 1958 (49 U.S.C. 40101 et seq.), by the Secretary of Transportation, with respect to any air carrier or foreign air carrier subject to that act; and, (iv) The Securities Exchange Act of 1934 (15 U.S.C. 78a et seq.), with respect to any issuer, security holder, or other person subject to the jurisdiction of the Commission under the Securities Exchange Act of 1934 (15 U.S.C. 78a et seq.) in this part or otherwise defined in the Federal Deposit Insurance Act (12 U.S.C. 1813(c)). Nothing shall have the meaning given to them in paragraph (i)(b) of the Federal Deposit Insurance Act of 1978 (12 U.S.C. 1813). (b) (A) Additional powers. (1) For the purpose of the exercise by any agency referred to in paragraphs (a)(i)(1) through (a)(i)(4v) of this section of its power under any statute referred to in those paragraphs, a violation of this part is deemed to be a violation of a requirement imposed under that statute. (2) In addition to its powers under any provision of law specifically referred to in paragraphs (a)(i)(1) through (a)(i)(4v) of this section, each of the agencies referred to in paragraph (a)(i) of this section has all the powers conferred upon it by the provisions of law referred to in paragraph (a)(i) of this section, except that enforcement of the requirements imposed under this title is specifically granted to another government agency under paragraphs (a)(i)(1) through (a)(i)(4v) of this section, and subject to subtitle B of the Consumer Financial Protection Act of 2010, the Federal Trade Commission has the authority to enforce such requirements. For the purpose of the exercise by the Federal Trade Commission of its functions and powers under the Federal Trade Commission Act, a violation of this part shall be deemed a violation of a requirement imposed under the Federal Trade Commission Act. All of the functions and powers of the Federal Trade Commission under the Federal Trade Commission Act shall apply to the jurisdiction of the Federal Trade Commission under this part, regardless of whether that person is engaged in commerce or meets any other jurisdictional tests under the Federal Trade Commission Act.

[3] Example of meta

```
{'chapter': ['II'],
'chapter_title': ['CHAPTER II - FEDERAL RESERVE SYSTEM (CONTINUED)'],
'subchapter': ['A'],
'subchapter_title': ['SUBCHAPTER A - BOARD OF GOVERNORS OF THE FEDERAL RESERVE SYSTEM (CONTINUED)'],
'part': ['235'],
'part_title': ['PART 235 - DEBIT CARD INTERCHANGE FEES AND ROUTING (REGULATION II)'],
'section': ['235.9'],
'section title': ['$ 235.9 Administrative enforcement.']}
```

[4] BERTopic generated topics for the first 10 records

	Document	Topic	Name
0		-1	-1_the_of_or_to
1	(a) (1) Compliance with the requirements of th...	2	2_of_the_and_bank
2	(a) <E T="04">Federal Register. The Board publ...	0	0_the_of_or_to
3	(a) <E T="04">Federal Register. The Committee ...	0	0_the_of_or_to
4	(a) A Bank shall make long-term advances only ...	-1	-1_the_of_or_to
5	(a) A Bank shall require each member to mainta...	9	9_capital_bank_the_undercapitalized
6	(a) A Bank shall require the borrower to certi...	-1	-1_the_of_or_to
7	(a) A Board-regulated institution described in...	-1	-1_the_of_or_to
8	(a) A Board-regulated institution that is an a...	4	4_capital_ratio_boardregulated_buffer
9	(a) A Farm Credit Bank or agricultural credit ...	8	8_agricultural_farm_credit_bank

[5] Chapters for the first 10 records

```

0    {'chapter': ['II', 'II', 'II', 'II', 'II', 'II...
1    {'chapter': ['II'], 'chapter_title': ['CHAPTER...
2    {'chapter': ['II'], 'chapter_title': ['CHAPTER...
3    {'chapter': ['II'], 'chapter_title': ['CHAPTER...
4    {'chapter': ['XII'], 'chapter_title': ['CHAPTE...
5    {'chapter': ['XII'], 'chapter_title': ['CHAPTE...
6    {'chapter': ['XII'], 'chapter_title': ['CHAPTE...
7    {'chapter': ['II'], 'chapter_title': ['CHAPTER...
8    {'chapter': ['II'], 'chapter_title': ['CHAPTER...
9    {'chapter': ['VI'], 'chapter_title': ['CHAPTER...
Name: meta, dtype: object

```