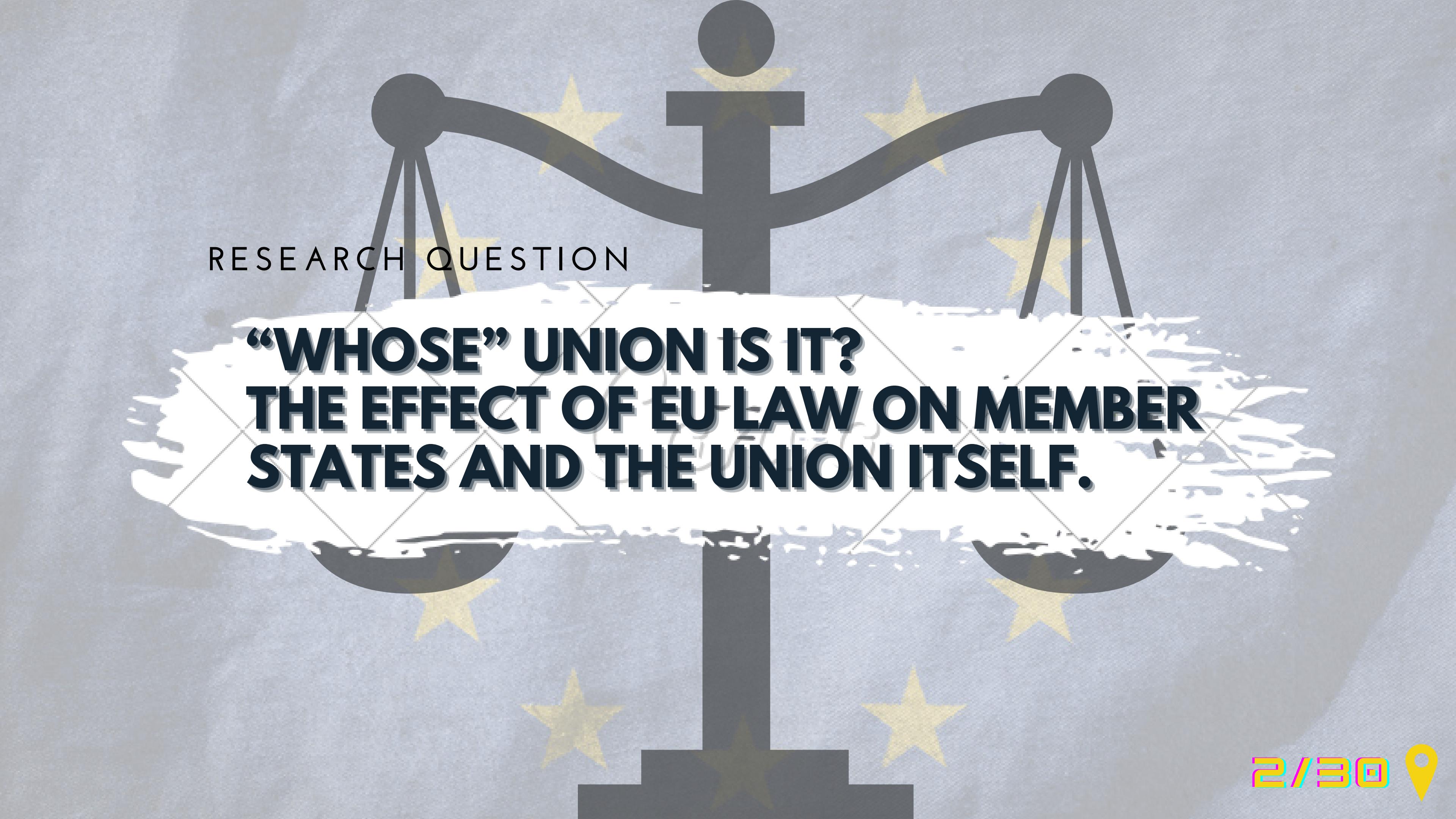


LEGAL DATA ANALYSIS

12 DECEMBER, 2023

MARTA DATCU #100294426,
NURIA BRAMER,
YARA BÖHLEN #100230447



RESEARCH QUESTION

“WHOSE” UNION IS IT? THE EFFECT OF EU-LAW ON MEMBER STATES AND THE UNION ITSELF.





EU LAW AND NATIONAL COURTS



HOW MANY DECISIONS ARE THERE PER YEAR?

WHAT EU INSTRUMENTS ARE USED MOST?

DOES REFERENCING EU LAW AFFECT OUTCOMES?

DID COVID CHANGE ANYTHING?

DO FR AND DE TREAT THE EU DIFFERENTLY?

DIFFERENCE BETWEEN ECJ, CASS & BVERG



THE RESEARCH

DATA

- ECJ Decisions
- Cour de Cassation
- Bundesverfassungsgericht
- 2018-2023



VARIABLES

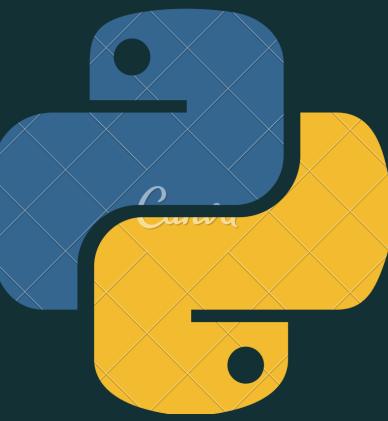
- EU instruments
- Fields of law (themes)
- Text of decisions
- Dates (pre and post COVID)
- Case outcomes

VECTORS OF ANALYSIS

- Decisions per year per court
- Case outcomes per year per court
- EU instruments cited per year and outcome per court

METHODOLOGY & STEPS





METHODOLOGICAL ELEMENTS

PARSING THROUGH XML FILES TO EXTRACT DATA

- ENCODED WITHIN A WEBPAGE ARE SPECIFIC ELEMENTS IN WHICH DATA IS STORED
- EACH ELEMENT HAS A SPECIFIC ATTRIBUTES
 - ATTRIBUTES CONTAIN DATA THAT IS ONLY ACCESSIBLE BY LOOKING AT THE CODE OF THE WEBPAGE
 - EXAMPLE: BGH WEBPAGE CONTAINS THE CLASS ELEMENT “AKTENZEICHEN” UNDER ANOTHER ELEMENT ”SPAN”

```
url_prefix = "https://www.bundesverfassungsgericht.de/SiteGlobals/Forms/  
decision_urls = []  
  
for page_num in range(1,2):  
    response = requests.get(url_prefix.format(page_num))  
    html_content = response.text  
    soup = BeautifulSoup(html_content, 'html.parser')  
    decisions_bverg = soup.find_all('span', class_="aktenzeichen")
```



REGEXES

- THIS IS USED TO EXTRACT A SPECIFIC SEQUENCE OR PATTERN FROM A TEXT
- THE PATTERN ADOPTS A GENERAL FORMAT THAT ALLOWS THE CODE TO EXTRACT ALL DATA THAT HAS THIS SPECIFIC FORMAT
- EXAMPLE: BGH LOOPING THROUGH COLUMNS OF DATE, DECISION NO. URL AND CHAMBER TO EXTRACT DATA THAT MATCHES THE PATTERN

```
df = pd.DataFrame(columns=["Date", "Decision No.", "URL", "Chamber"])

for x in headings:
    date_pattern = r"vom (\d{1,2}\. \w+ \d{4})"
    decision_pattern = r"(\d+ [A-Za-z]+ \d+/\d+)"
    url_pattern = r"https://www.bverfg.de/e/[a-z0-9_]+\.[a-z]+"
    chamber_pattern = r"(Ersten|Zweiten) Senats"

    match = re.search(date_pattern, x)
    date = match.group(1) if match else None

    match = re.search(decision_pattern, x)
    decision_no = match.group(1) if match else None

    match = re.search(url_pattern, x)
    url = match.group() if match else None

    match = re.search(chamber_pattern, x)
    chamber = match.group(1) if match else None
```

LOOPS

- ALLOWS YOU TO RUN THROUGH DATA UNTIL A CERTAIN REQUIREMENT IS FULFILLED AND THE DATA THAT SATISFIES THE REQUIREMENT IS PRINTED
- EXAMPLE: BGH LOOP THAT RUNS THROUGH TEXT OF DECISIONS AND EXTRACTS DATA THAT MATCHES THE SPECIFIC PATTERNS LISTED (*NOT INDICATED HERE*)

```
eu_law_matches = []
for pattern in eu_law_patterns:
    matches = re.findall(pattern, text)
    eu_law_matches.extend(matches)

return eu_law_matches
```



LIST OF LIST METHOD

- THIS METHOD ALLOWS YOU TO STORE MULTIPLE LISTS OF ITEMS INTO ONE BIG LIST (WHICH IS WHAT A DATAFRAME IS)
- EXAMPLE: ECJ LIST OF URLs THAT ARE LOOPED THROUGH TO CREATE SEVERAL COLUMNS THAT ARE THEN ADDED INTO ONE OVERALL DATAFRAME

TOKENISATION

- THIS ALLOWS YOU TO SPLIT UP TEXT INTO SMALLER BITS OF TEXT (EX. A SENTENCE INTO INDIVIDUAL WORDS)
- THIS ALLOWS YOU TO WORK ON THE INDIVIDUAL UNITS
- EXAMPLE: ECJ

```
urls = [
    'https://curia.europa.eu/juris/documents.jsf?page=1&nat=or&mat=or&pcs=0or&jur=C&for=&jge=&dates=%',
    'https://curia.europa.eu/juris/documents.jsf?page=2&nat=or&mat=or&pcs=0or&jur=C&for=&jge=&dates=%',
    'https://curia.europa.eu/juris/documents.jsf?page=3&nat=or&mat=or&pcs=0or&jur=C&for=&jge=&dates=%',
    'https://curia.europa.eu/juris/documents.jsf?page=4&nat=or&mat=or&pcs=0or&jur=C&for=&jge=&dates=%',
    'https://curia.europa.eu/juris/documents.jsf?page=5&nat=or&mat=or&pcs=0or&jur=C&for=&jge=&dates=%',
    'https://curia.europa.eu/juris/documents.jsf?page=6&nat=or&mat=or&pcs=0or&jur=C&for=&jge=&dates=%'
]

dfs = []

for url in urls:
    response = requests.get(url)

    if response.status_code == 200:
        html_content = response.text

        soup = BeautifulSoup(html_content, 'html.parser')
        table = soup.find('table', class_='detail_table_documents')

        if table:
            column_aff_data = table.find_all('td', class_='table_cell_aff')
            column_aff_values = [data.get_text(strip=True) for data in column_aff_data]

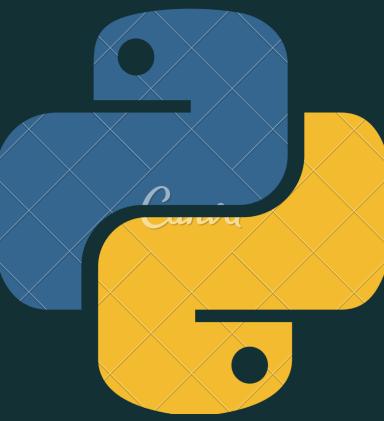
            df_Case = pd.DataFrame({'Case': column_aff_values})
            df_Case = df_Case[df_Case['Case'] != ''] #filter out empty cells
            df_Case.reset_index(drop=True, inplace=True)

            dfs.append(df_Case)
```

```
stopwords_german = set(stopwords.words('german'))
text_content = re.sub(r'\W+', ' ', text_content)
text_content = re.sub(r'\d+', ' ', text_content)
tokens = word_tokenize(text_content.lower())
filtered_tokens = [token for token in tokens if token not in stopwords_german]
preprocessed_text = " ".join(filtered_tokens)
```



STEP 1: CREATION OF 3 DATAFRAMES COUR DE CASSATION, BUNDESGERICHTSHOF AND EUROPEAN COURT OF JUSTICE



	CASS	BverfG	ECJ
extraction method	API method	parsing xmlfiles	parsing xmlfiles
data frame	df_france	df	df_ECJ_final
variables	columns: "score", "highlights", "id", "jurisdiction", "chamber", "number", "numbers", "ecli", "formation", "publication", "decision_date", "solution", "type", "summary", "themes", "keyword"	"date", "Decision N°", "URL", "CHamber", "EU law"	"dates", "case outcome", "subject", "Number"





RESULTS



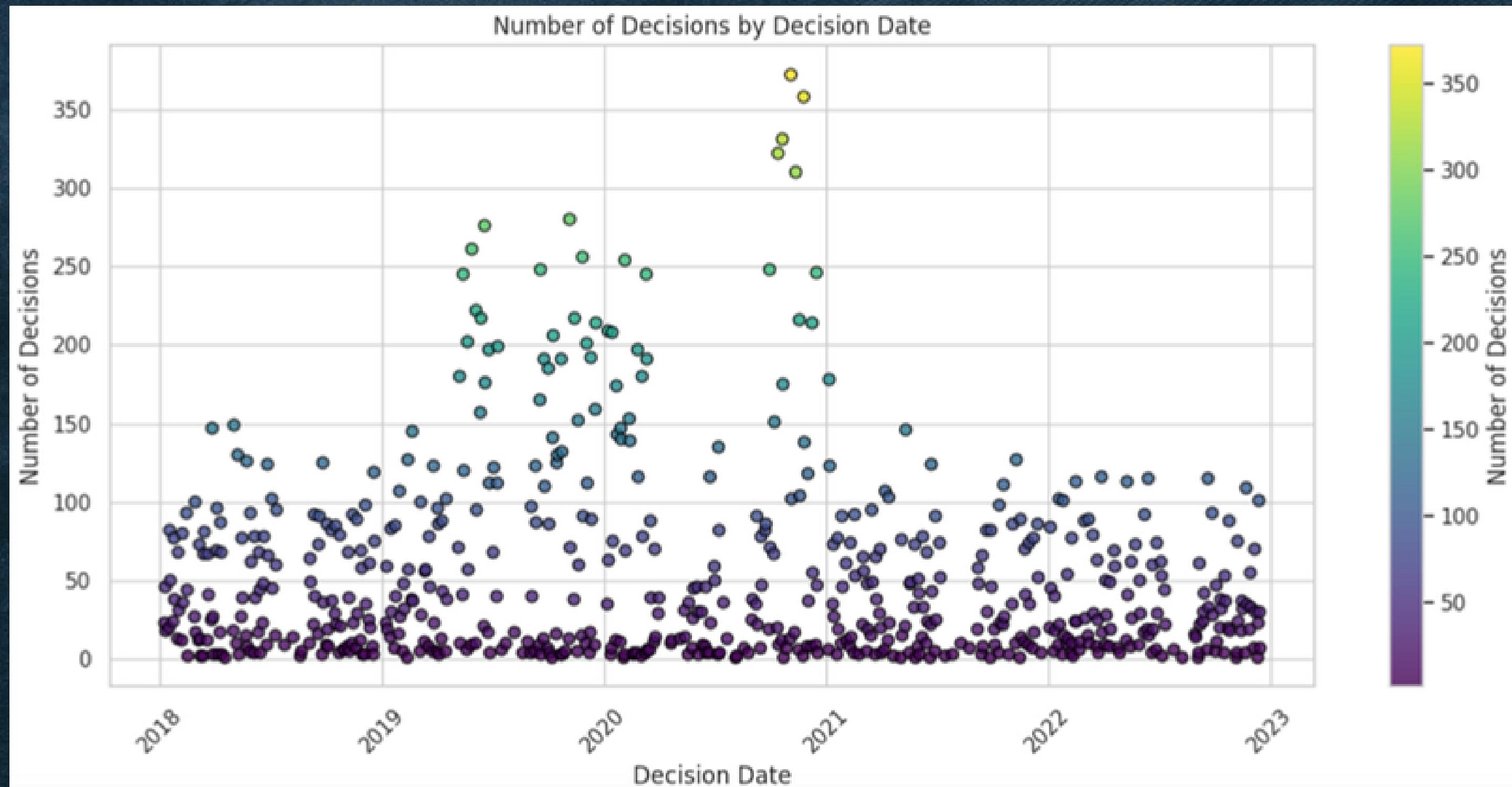


THEMES:

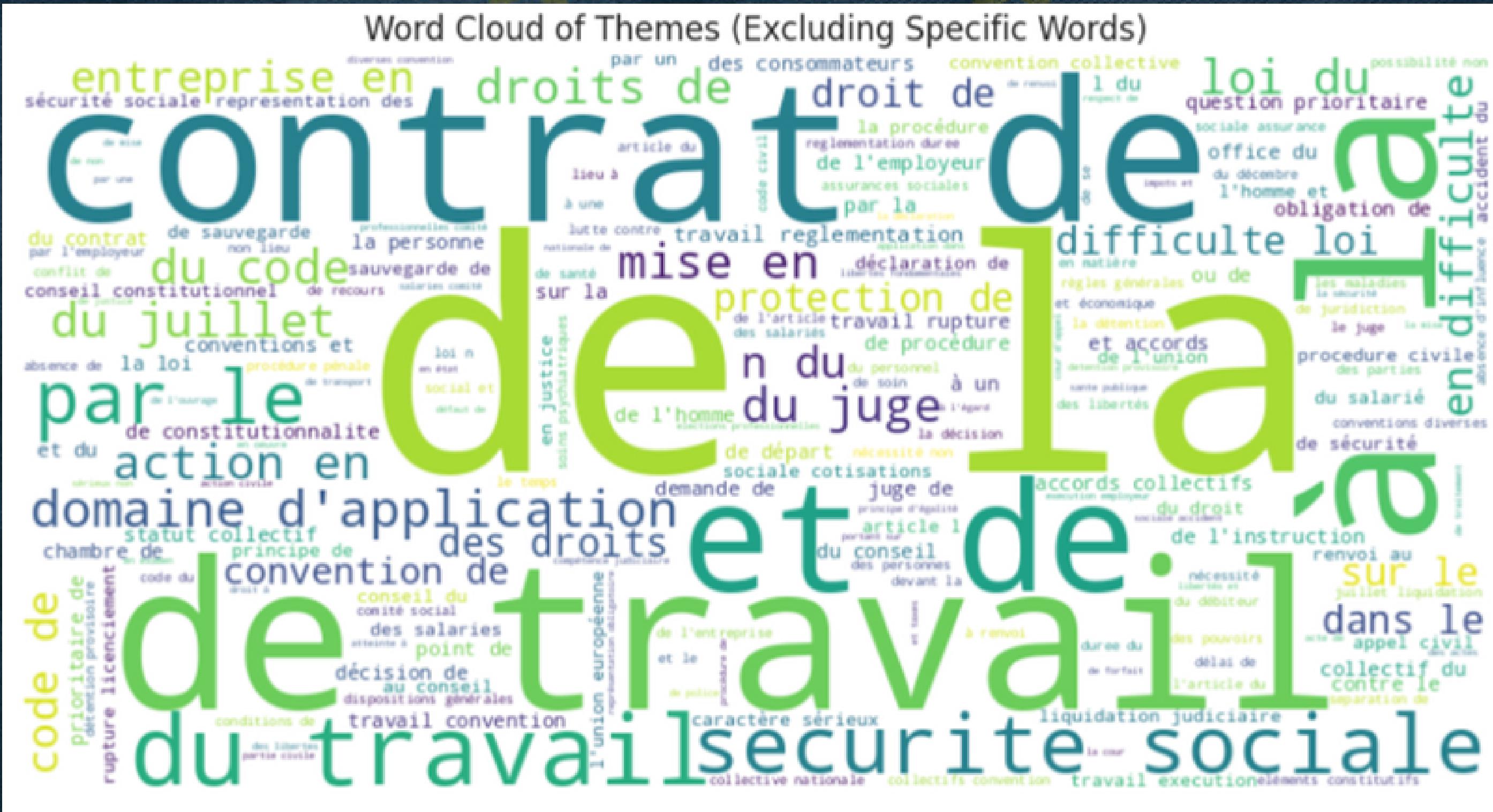
SURVEILLANCE, CARS, CREDIT INSTITUTES, SELF-DETERMINATION, BROADCAST, FUNDAMENTAL RIGHTS,
GOVERNMENT BONDS, GREECE, FEDERAL STATES,
ASSISTED SUICIDE



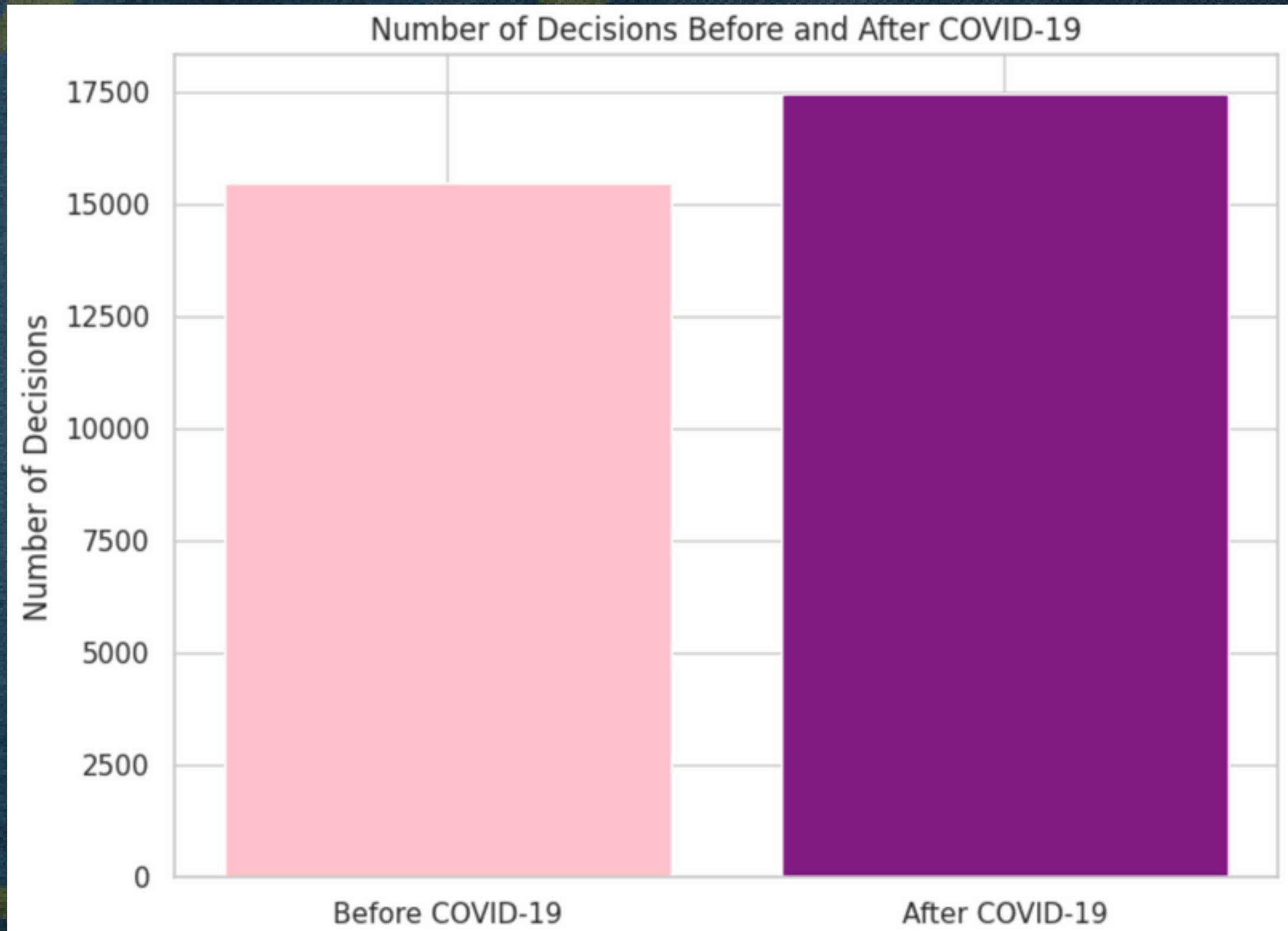
THE DISTRIBUTION OF DECISIONS OVER THE 5 YEARS



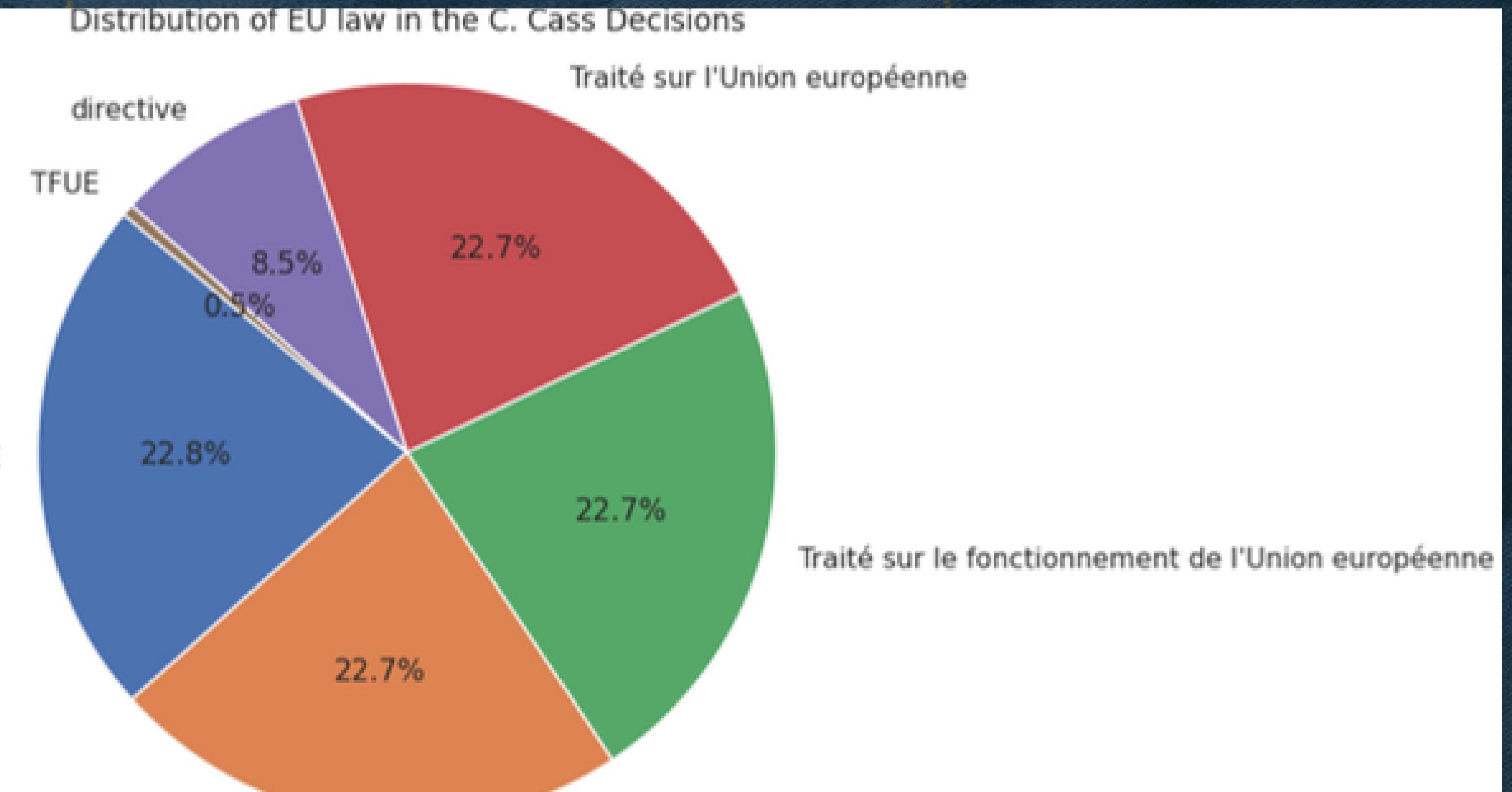
WORD CLOUD ON THE MOST COMMON SUBJECTS



THERE HAS BEEN A SLIGHT SURGE IN THE NUMBER OF DECISIONS CONCERNING EU LAW FOLLOWING THE COVID-19 PANDEMIC. THIS TREND APPEARS LINKED TO THE SURGE IN LEGISLATIVE ACTIVITY POST-PANDEMIC, NOTABLY THE IMPLEMENTATION OF A SERIES OF EU LAWS (THE EU RECOVERY PLAN, MODIFICATIONS IN THE FIELDS OF HEALTH, HEALTH ADMINISTRATION, AND LABOUR LAW;).



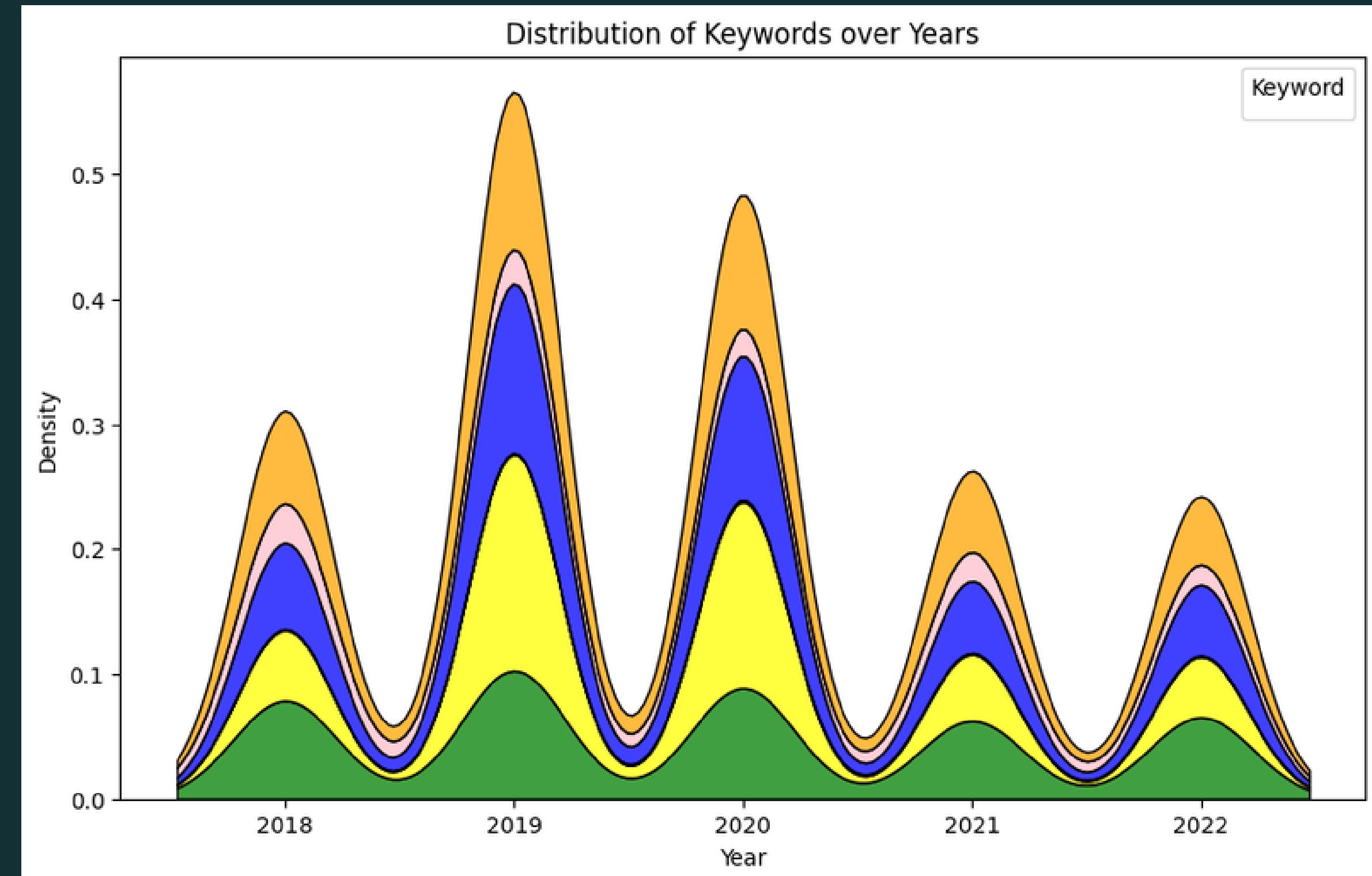
DISTRIBUTION OF THE TYPES OF EU LAW MENTIONED IN THE DECISION



THE CODE GENERATES A KDE
(KERNEL DENSITY ESTIMATE) PLOT
USING SEABORN'S KDEPLOT(),
UTILIZING THE 'YEAR' AS THE X-
AXIS, 'KEYWORD' AS THE HUE,
AND STACKS MULTIPLE DENSITIES
WHILE FILLING THE AREAS UNDER
THE CURVES. THE RESULTING
PLOT VISUALIZES THE DENSITY
DISTRIBUTION OF EU LEGAL
INSTRUMENTS OVER THE YEARS.

LEGEND:

RÈGLEMENT UE = ORANGE
DIRECTIVE = PINK
CHARTE DES DROITS
FONDAMENTAUX DE L'UE= BLUE
TRAITÉ SUR L'UNION EUROPEENNE
=GREEN
TRAITÉ SUR LE FONCTIONNEMENT
DE L'UNION EUROPÉENNE=
YELLOW
TFEU = PURPLE

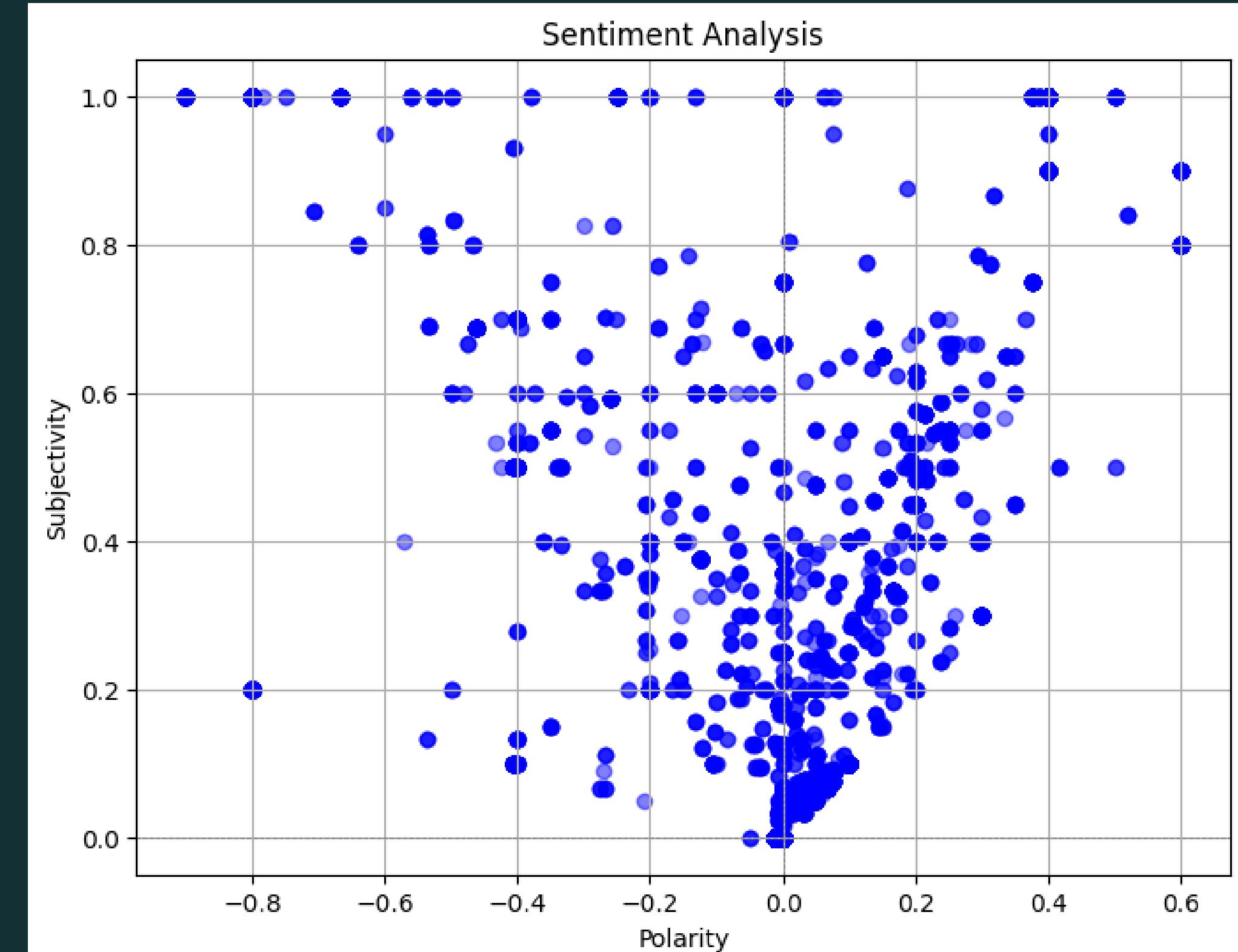


SENTIMENT ANALYSIS

HERE WE USED THE TEXTBLOB LIBRARY TO ANALYZE THE SENTIMENT (POLARITY AND SUBJECTIVITY) OF FRENCH TEXT OF THE SUMMARY OF EACH DECISION.

FOR US: (I) WE REPLACED ANY NAN VALUES IN THE 'SUMMARY' COLUMN OF THE DATAFRAME DF WITH AN EMPTY STRING. (II) THE SUMMARY ANALYSIS ITERATES THROUGH EACH ROW OF THE DATAFRAME (DF) AND EXTRACTS THE TEXT FROM THE 'SUMMARY' COLUMN. IT THEN USES TEXTBLOB TO CONVERT THE TEXT INTO A TEXTBLOB OBJECT, WHICH IS NECESSARY FOR SENTIMENT ANALYSIS.

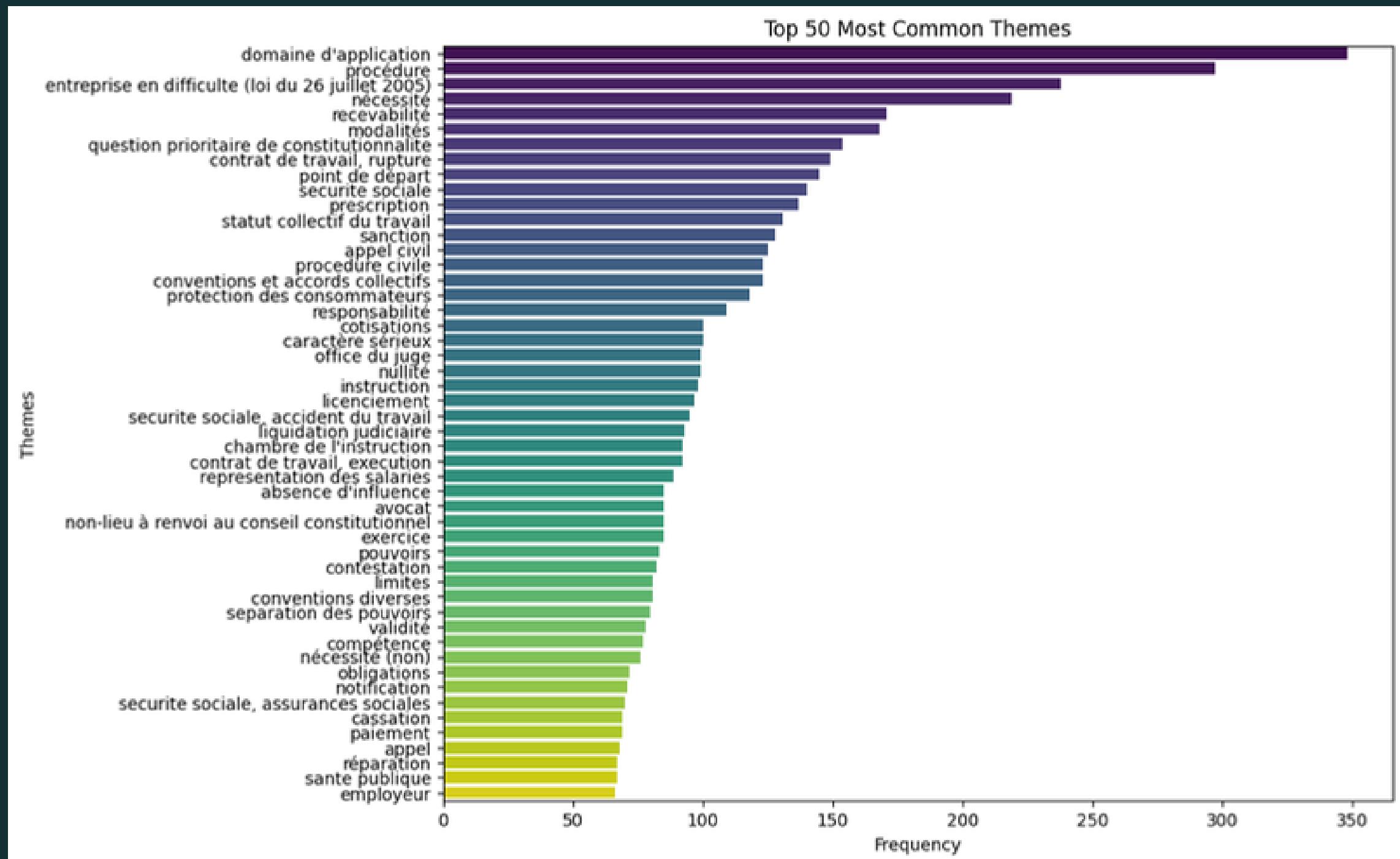
THE SENTIMENT (POLARITY AND SUBJECTIVITY) OF EACH TEXT IS COMPUTED USING TEXTBLOB'S SENTIMENT METHOD. (III) THE SENTIMENT SCORES ARE APPENDED INTO THE "SENTIMENT_SCORE" LIST AND THEN A "SENTIMENT_DF" DATAFRAME IS CREATED



I. THEMES

WHAT ARE THE MOST COMMON FRENCH LAW AREAS (THEMES) THAT REFERENCE EU LAW?

- A. OVERALL THEME IS CONTRACT AND WORK-PLACE LAW
- I. “ENTERPRISE EN DIFFICULTÉ”, “CONTRAT DE TRAVAIL RUPTURE”, “STATUT COLLECTIF DU TRAVAIL”, “REPRÉSENTATION DES SALARIÉS”.
- B. “SANTÉ PUBLIQUE”
- I. COVID: TRAVEL WAS HEAVILY RESTRICTED DURING THE PANDEMIC FOR REASONS OF NATIONAL PUBLIC HEALTH
 - II. RESTRICTIONS HAD TO STAY IN COMPLIANCE WITH FREEDOM OF MOVEMENT, WHICH EXPLAINS WHY A MATTER OF NATIONAL PUBLIC HEALTH WOULD BE HEAVILY LINKED WITH EU LAW

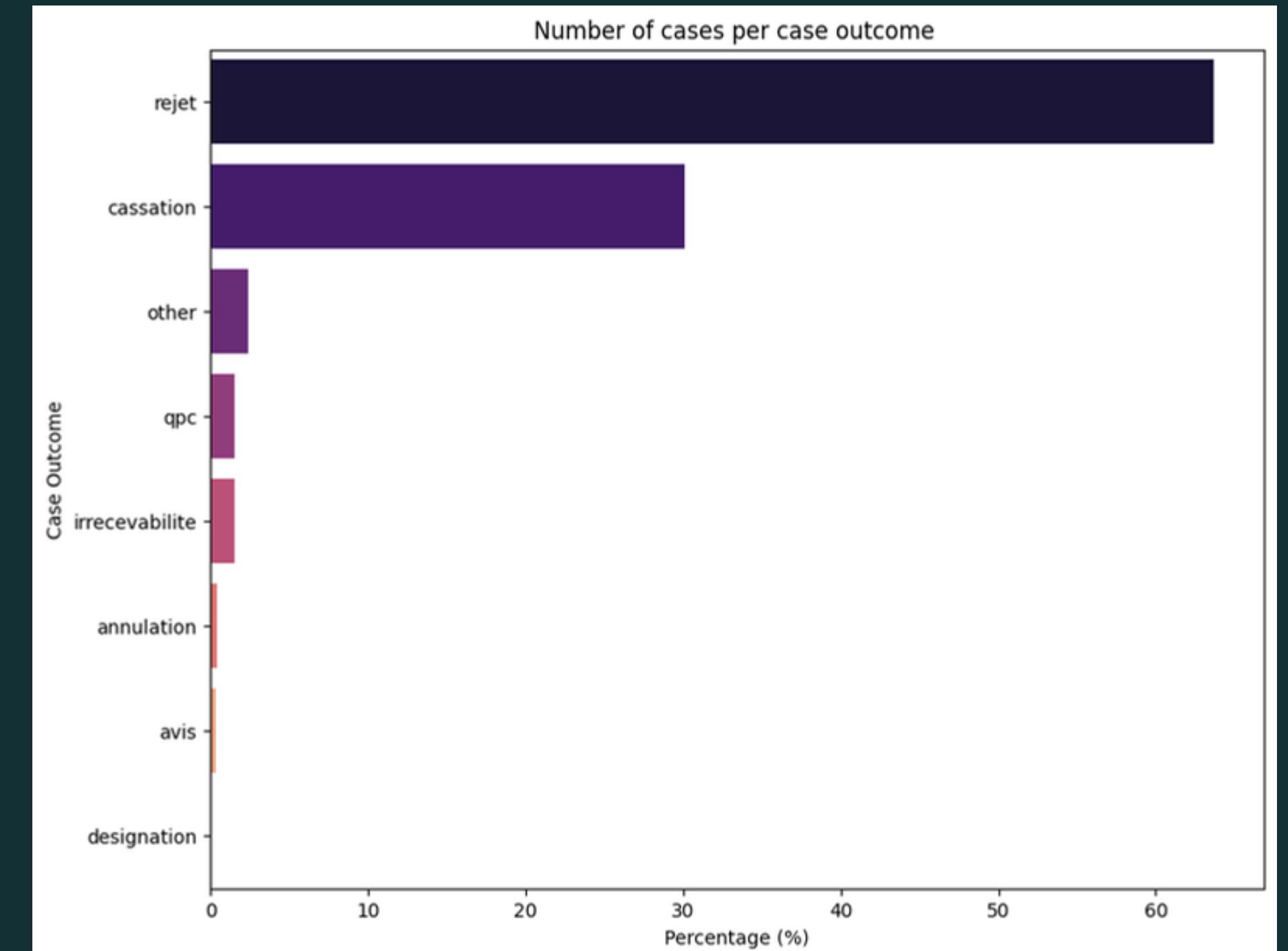


II. CASE OUTCOMES

A. DOES REFERENCING EU LAW HAVE AN IMPACT ON HOW CASES ARE DECIDED?

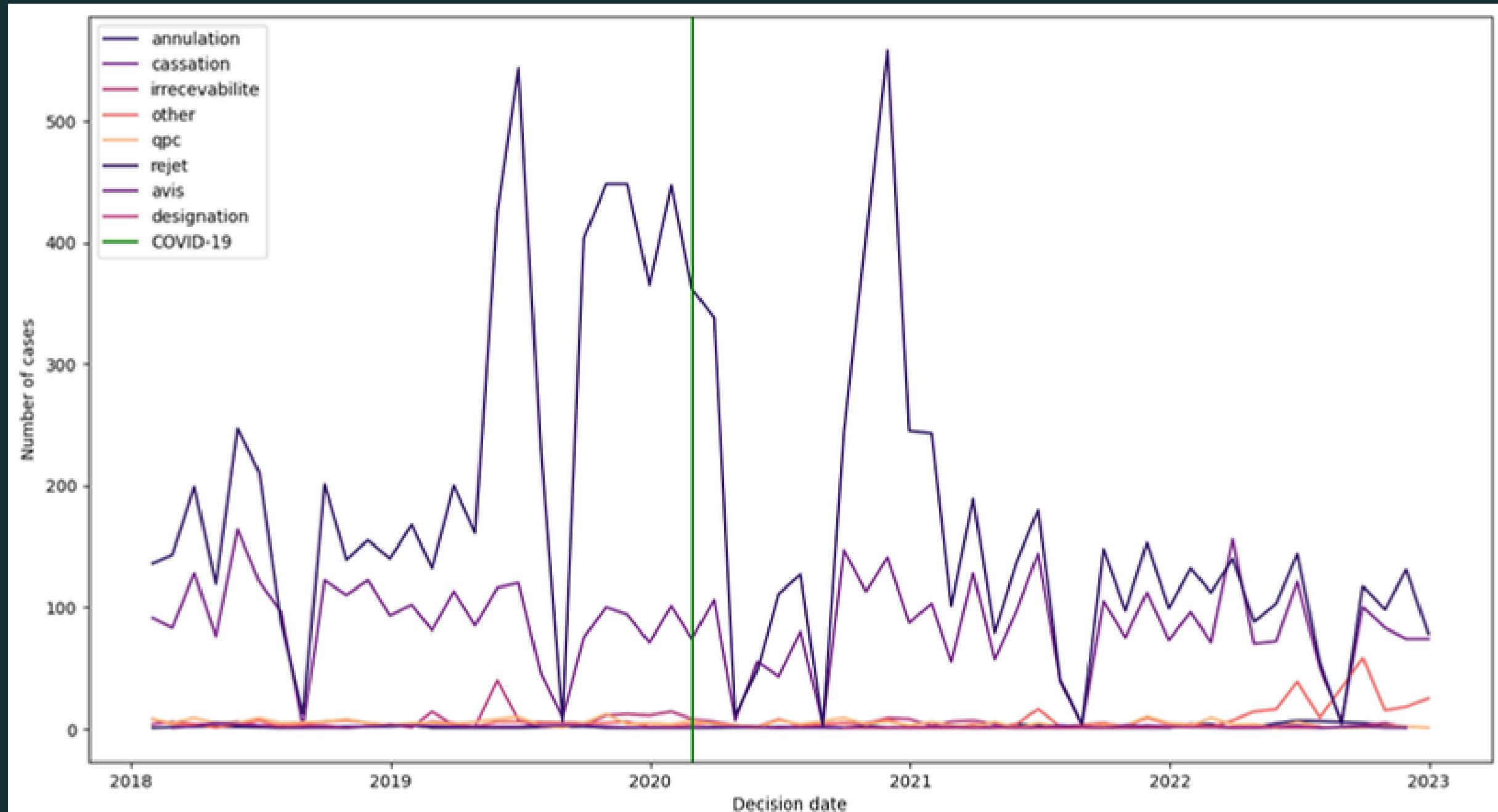
1.13 DIFFERENT TYPES OF OUTCOME, DECIDED TO LIMIT THESE AND GROUPED OUTCOMES WITH FEWER RESULTS INTO ALREADY EXISTING CATEGORY "OTHER"

- A. MAJORITY: DISMISSAL, CASSATION
- B. "QPC" IS 4TH LARGEST GROUP OF OUTCOMES (BUT REMAINS SMALL PROPORTIONALLY)
 - I. ROLE OF COUR DE CASSATION IN SEIZING THE CONSEIL CONSTITUTIONNEL ON A MATTER OF CONSTITUTIONALITY
 - II. QUESTION OF EU SUPRA-CONSTITUTIONALITY AND NATIONAL COURTS' JURISDICTIONS



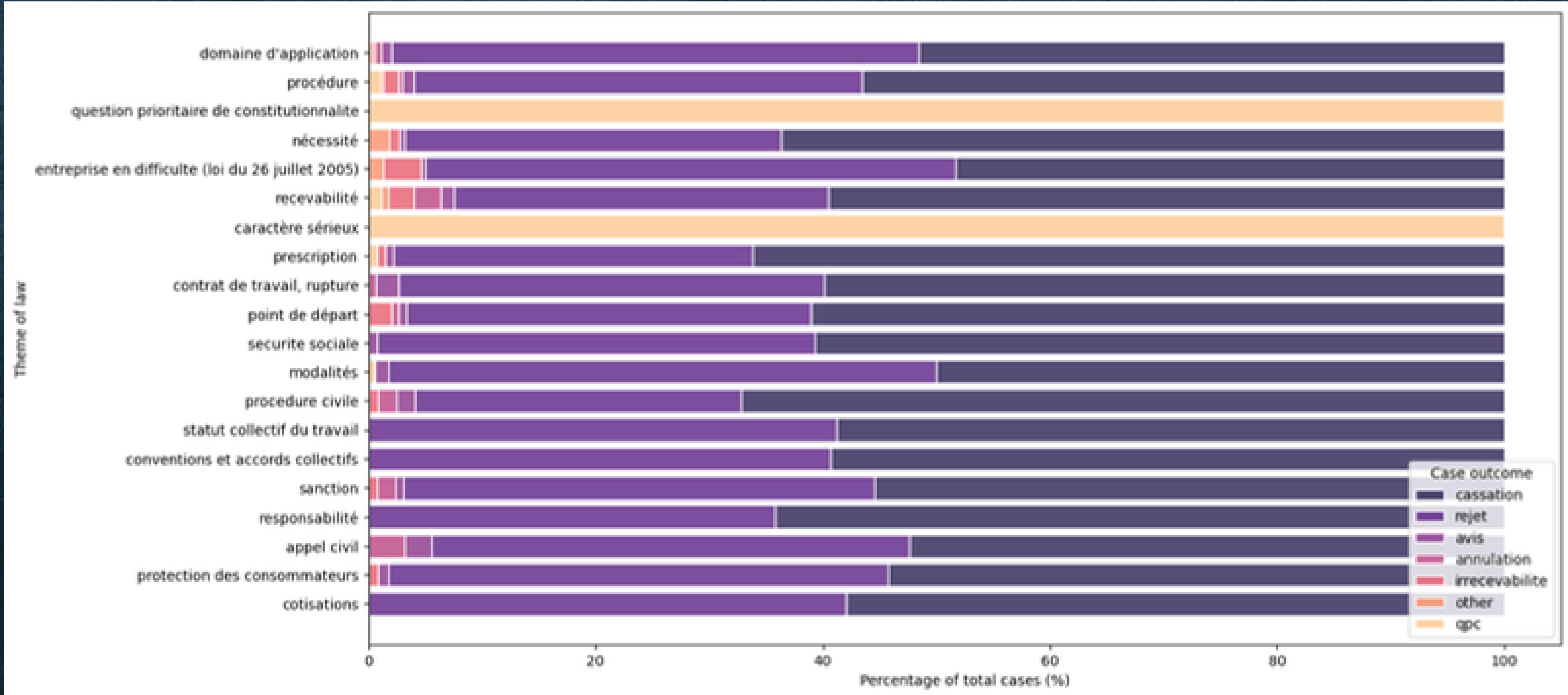
II. CASE OUTCOMES HOW ARE CASES DECIDED OVER TIME?

1. FILTERED SOLUTIONS AGAINST DECISION DATES, TO SEE IF THERE WAS A CHANGE IN OUTCOMES OVER TIME (GROUPED DATES PER MONTH)
2. MAJORITY OF CASES ARE DISMISSED
 - A. 2019?
3. COVID 19 - Affected NUMBER OF DECISIONS BUT NOT OUTCOMES THEMSELVES EXCEPT FOR MASIVE INITIAL SPIKE IN DISMISSALS



II. CASE OUTCOMES

IS THERE A RELATIONSHIP BETWEEN THE DIFFERENT AREAS OF FRENCH LAW THAT REFERENCE EU LAW AND THE CASE OUTCOME?



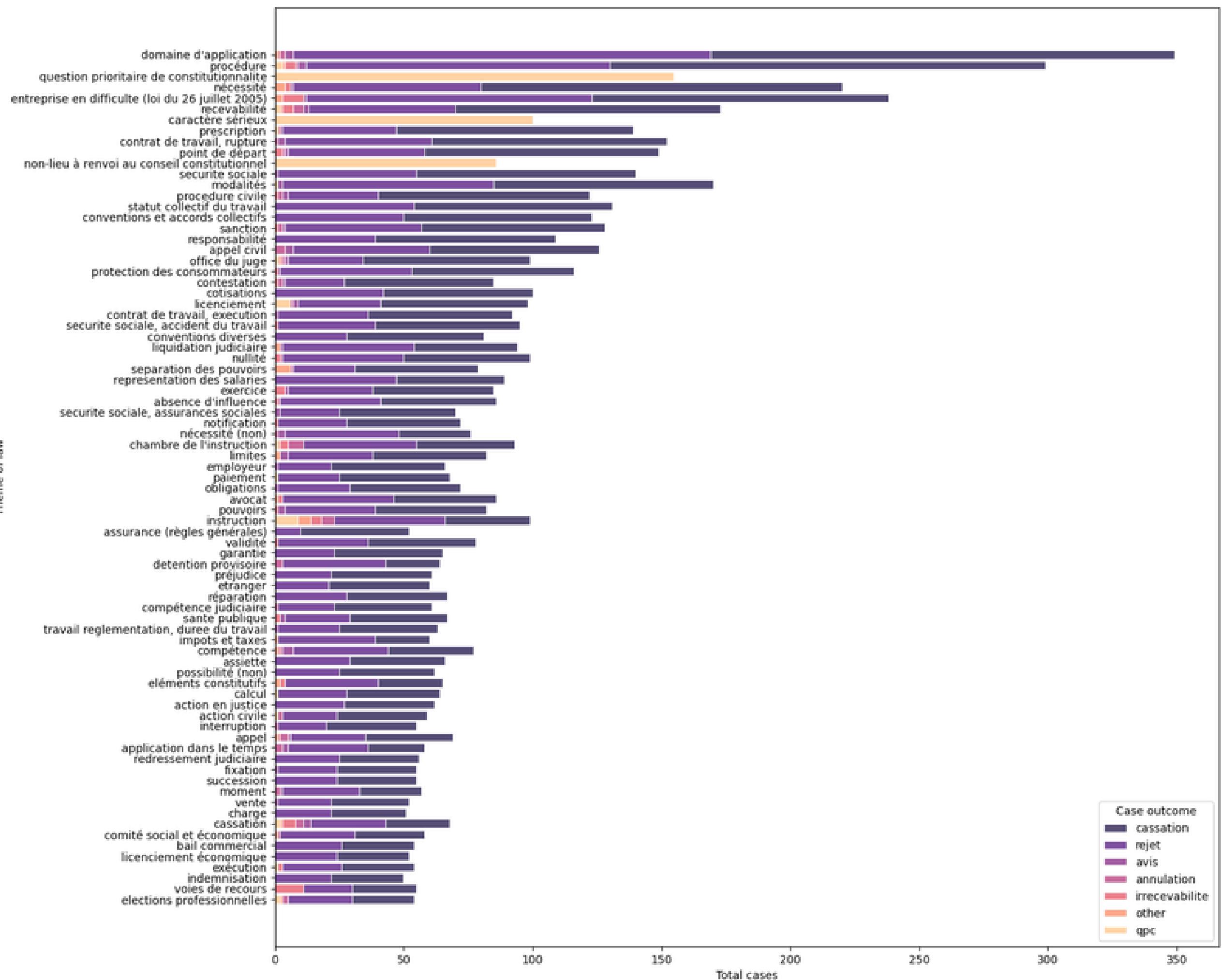
1. FILTERED AND UNIQUE THEMES AND FILTERED CASE OUTCOMES. EACH THEME WAS THEN CROSS-REFERENCED WITH THE DIFFERENT OUTCOMES, ALLOWING US TO COUNT HOW MANY TIMES A SPECIFIC THEME LED TO A SPECIFIC SOLUTION OVER TIME

A. FOR INSTANCE, "DOMAINE D'APPLICATION" LED TO 179 CASSATIONS, 162 DISMISSALS, ETC.

2. "LICENCIEMENT" AND "QPC"

A. EU LAW HAS BETTER PROTECTION FOR UNFAIR DISMISSALS

B. EU LAW AND ITS STATUS OF SUPRA-CONSTITUTIONALITY AND COMPATIBILITY WITH FRENCH LAW



III. REFERENCES TO EU LAW

HAVE THE REFERENCES TO PARTICULAR EU INSTRUMENTS CHANGED OVER TIME?

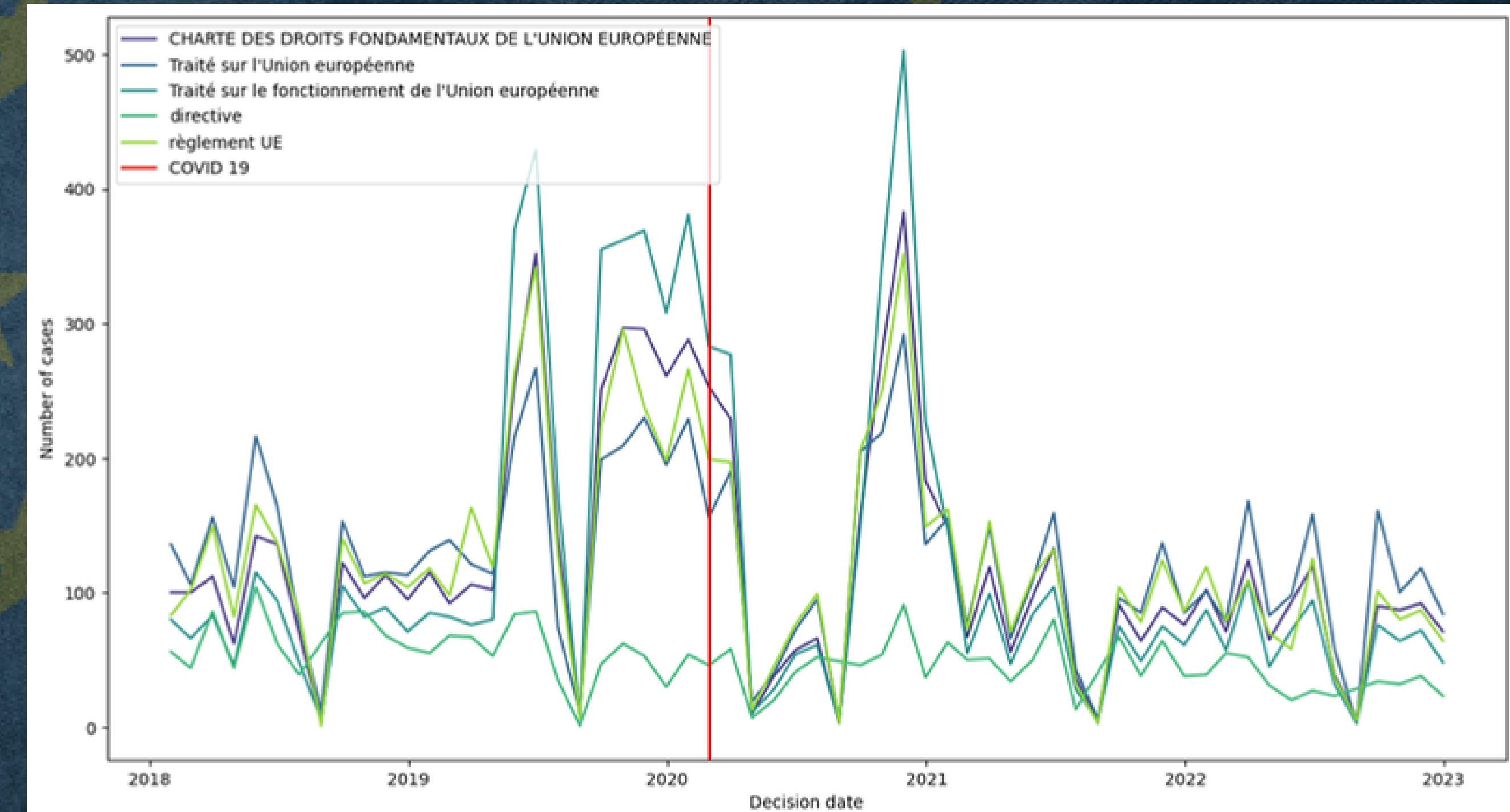
1. EU INSTRUMENTS OVER TIME (GROUPED BY MONTH)

2. THE TRAITÉ DU FONCTIONNEMENT DE L'UE IS MOST REFERENCED IN THE DECISIONS

3. COVID:
 A. DROP IN DECISIONS RENDERED
 B. INCREASED REFERENCES TO CHARTER OF FUNDAMENTAL RIGHTS OF THE EU MID-2020 (PROPORTIONALLY)

C. FREEDOM OF MOVEMENT

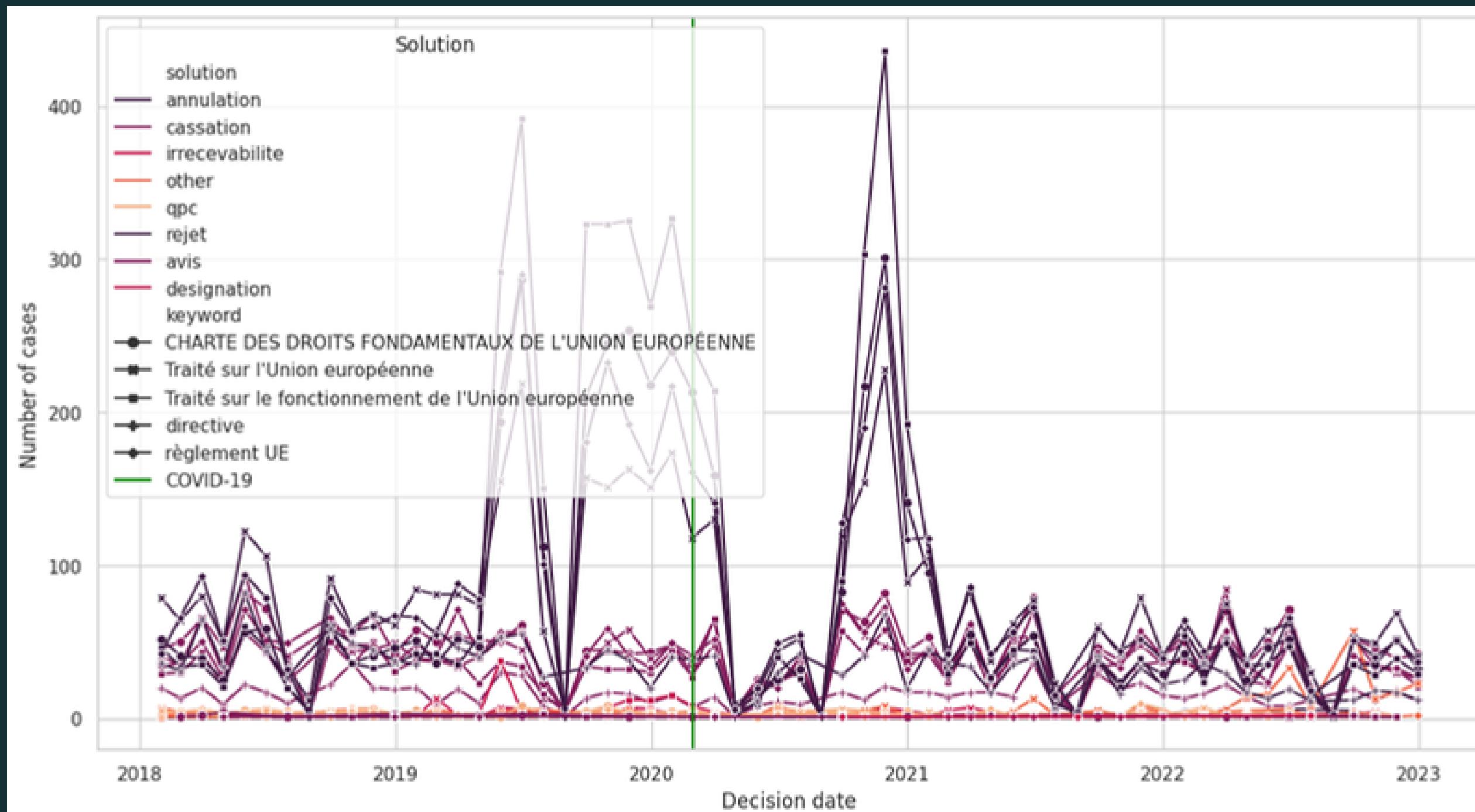
4. NB: ONLY FILTERED DATA FOR 5 EU INSTRUMENTS



III. REFERENCES TO EU LAW

IS THERE A RELATIONSHIP BETWEEN THE REFERENCE TO SPECIFIC EU INSTRUMENTS IN DECISIONS AND THE OUTCOME OF THE DECISION OVER TIME?

1. EU INSTRUMENTS AGAINST FILTERED SOLUTIONS OVER TIME (GROUPED BY MONTH) - SO WE ARE PLOTTING CASES COLOURED BY KEYWORD, BUT THE SAME CASE CAN HAVE MULTIPLE KEYWORDS
2. NUMBER OF TIMES A PARTICULAR OUTCOME OCCURS WHEN A SPECIFIC EU INSTRUMENT IS REFERRED TO IN THE DECISION.
 - I. GENERAL TREND: DISMISSAL IS MOST LIKELY REGARDLESS OF EU INSTRUMENT CITED
 - II. EXCEPTION: CHARTER OF FUNDAMENTAL RIGHTS WAS MOST CITED EU DOCUMENT IN CASSATIONS OUTCOME IN LATE 2020
 - I. IMPACT OF COVID: CLAIMS OF VIOLATIONS OF INDIVIDUAL RIGHTS (ESP. FREEDOM OF MOVEMENT), WHICH ARE ENSHRINED IN THE CHARTER



RESEARCH & CHALLENGES LIMITATIONS



LIMITATIONS & CHALLENGES

1. CODING IN THREE LANGUAGES
(VARIABLES + PYTHON COMMANDS)
2. DATA NOT COMPARABLE: DIFFERENT
INSTITUTIONAL ROLE OF CASS AND
BVERFG (QUANTITY OF DECISIONS)
3. LIMITED POSSIBILITY TO SCRAPE ECJ
4. COURTS HAVE DIFFERENT METHODS
OF PRESENTING LEGAL DATA AND
CITING EU LAW



QUALITATIVE ANALYSIS & CONCLUSIONS



CONCLUSION

- I. GENERALLY SEE GREATER NUMBER OF REFERENCES TO SPECIFIC EU DIRECTIVES IN FRENCH DECISIONS, WHILE GERMAN LAW REFERS TO EU LAW WIHTIN GERMAN NATIONAL LEGISLATION (TRANSPOSED IN NATIONAL LEGISLATION).
- II. THE CENTRAL THEME FOR FRANCE IS LABOUR LAW, WHILST GERMANY ONLY REFERS TO NEW DEVELOPMENTS IN THE EU WHEN REFERRING TO EU LAW.
- III. COVID IMPACT ON COURTS: WHILE THE IMPACT CAN BE SEEN IN THE COUR DE CASSATION, PARTICULARLY ON NUMBER OF DECIDED CASES AND A SLIGHT CHANGE IN THE USE OF EU-INSTRUMENTS (GREATER NUMBER OF REFERENCES TO THE CHARTER OF FUNDAMENTAL RIGHTS), THE DATA FRAMES FOR THE BVERFG AND ECJ ARE TOO SMALL TO REALLY NOTICE AN EFFECT (DUE TO DIFFERENT INSTITUTIONAL COMPOSITION OF CASS & BVERFG).





THANK YOU!