

# LEGAL DATA ANALYSIS PROJECT

Presented By CHIOCCHIA Mathilde,  
ENDERLIN Léo, SILBERREISS Faustine

12/12/2023

Damin Charlotin

# Greetings! Let the adventure commence!

Our problematic: What type of offenses (theft, misappropriation of corporate assets, discrimination, moral harassment, sexual molestation, manslaughter) were subject to the most appeals before the Cour de cassation in France between January 1st 2019 and January 1st 2023?

## Outline:

- Introduction
- I. Data collection
- II. Analysis of the results
- III. Unexplored biases

# Goal: readability analysis

- We focus on the readability of texts.
- We need to consider whether articles are clearer or less clear in relation to each other.
- We consider judgments handed down between January 1, 2019 and January 1, 2023

**Our starting hypothesis** is that the least clear articles (with the lowest readability) are referred to more often, since they leave more room for interpretation by the parties, lawyers and judges.

## Offenses taken into account:

- Theft (article 331-1 of the French penal code)
- Sexual assault (article 222-22 of the French penal code)
- Corruption (article 433-1 of the French penal code)
- Moral harassment (article L. 1152-1 of the Labor Code)
- Misappropriation of corporate assets (article L. 231-3 of the French Commercial Code)
- Manslaughter (article 221-6 of the French penal code)
- Discrimination (article 225-1 of the French penal code)
- Assault and battery (article 222-7 of the French penal code)

# I. Data Collection

Importation of the useful plugins

```
import pandas as pd
import re
import numpy as np
from requests_oauthlib import OAuth2Session
import requests
import json
import matplotlib.pyplot as plt
```

Firstly, we specify the credentials necessary for connecting to the API.

```
# OAuth2 authentication parameters
API_HOST = "https://sandbox-api.aife.economie.gouv.fr"
TOKEN_URL = 'https://sandbox-oauth.aife.economie.gouv.fr/api/oauth/token'
client_id = "b800ee81-4b08-4b7c-9cb6-67f51535888e"
client_secret = "0adff97c-722d-40aa-bbd6-3e32a08c6ae3"
```

```
# Obtain the token
tokencall = requests.post(TOKEN_URL, data={"grant_type": "client_credentials",
"client_id": client_id, "client_secret": client_secret, "scope": "openid"})
token = tokencall.json()
client = OAuth2Session(client_id, token=token)
```

```
# List of legal articles to search
articles_de_loi = [
    '311-1 du code pénal',      # Theft
    '222-22 du code pénal',    # Sexual Assault
    '433-1 du code pénal',      # Corruption
    'L. 1152-1 du code du travail', # Moral Harassment
    'L. 231-3 du code de commerce', # Misuse of Corporate Assets
    '221-6 du code pénal',      # Involuntary Manslaughter
    '225-1 du code pénal',      # Discrimination
    '222-7 du code pénal'       # Assault and Battery
]
```

```
# Base URL for the search
base_urls = {
    'ca': "https://sandbox-api.piste.gouv.fr/cassation/judilibre/v1.0/search?que",
    'cc': "https://sandbox-api.piste.gouv.fr/cassation/judilibre/v1.0/search?que",
}
```

```
# DataFrame to store the results
if_resultats = pd.DataFrame(columns=["decision_id", "jurisdiction", "texte_decision", "article_loi"])
```

```
# Function to add a line to the DataFrame
def ajouter_ligne_df(df, decision_id, jurisdiction, texte_decision, article):
    new_row = pd.DataFrame({"decision_id": [decision_id], "jurisdiction": [jurisdiction], "texte_decision": [texte_decision], "article_loi": [article]})
    return pd.concat([df, new_row], ignore_index=True)
```

```

# Browse through result pages for the Court of Appeal and the Court of Cassation
for key, base_url in base_urls.items():
    page = 1
    while True:
        resp = client.get(base_url + str(page))
        if resp.status_code == 200:
            try:
                results = resp.json()["results"]
                for result in results:
                    decision_id = result["id"]
                    jurisdiction = result.get("jurisdiction", "").lower()
                    decision_resp = client.get(f"https://sandbox-api.piste.gouv.fr/cassation/judilibre/v1.0/decision?id={decision_id}")
                    if decision_resp.status_code == 200:
                        try:
                            decision_resp_json = decision_resp.json()
                            texte_decision = decision_resp_json.get("text", "")

                            # Check for the presence of legal articles
                            for article in articles_de_loi:
                                if re.search(article.replace(' ', '\s*'), texte_decision, re.S|re.I):
                                    df_resultats = ajouter_ligne_df(df_resultats, decision_id, jurisdiction, texte_decision, article)
                        except json.JSONDecodeError:
                            print(f"JSON decoding error for decision {decision_id}")
                    except json.JSONDecodeError:
                        print(f"JSON decoding error for page {page}")
                page += 1

```

```

        elif resp.status_code == 416:
            # End of page range, exit the loop
            break
        else:
            print(f"Request failed with status code: {resp.status_code}")
            break

```

```

df_resultats_ca = df_resultats[df_resultats['jurisdiction'] == 'ca']
df_resultats_cc = df_resultats[df_resultats['jurisdiction'] == 'cc']

```

Create a DataFrame to store the summary

```

summary = pd.DataFrame(columns=["Article_Loi", "Nombre_Decisions_ca", "Nombre_Decisions_cc", "Percentage of Cassation Appel"])

```



# Our results

	Article_Loi	Nombre_Decisions_ca	Nombre_Decisions_cc	\
0	311-1 du code pénal	0	11	
1	222-22 du code pénal	0	25	
2	433-1 du code pénal	0	0	
3	L. 1152-1 du code du travail	0	57	
4	L. 231-3 du code de commerce	0	0	
5	221-6 du code pénal	0	68	
6	225-1 du code pénal	0	1	
7	222-7 du code pénal	0	3	

# Our results

	Percentage of Cassation Appeals	Percentage_Decisions_cc	readability_score
0	0.0	6.666667	29.181818
1	0.0	15.151515	78.017812
2	0.0	0.0	123.083333
3	0.0	34.545455	0.000000
4	0.0	0.0	113.409091
5	0.0	41.212121	0.000000
6	0.0	0.606061	0.000000
7	0.0	1.818182	45.315789

« Le vol est la soustraction frauduleuse de la chose d'autrui ».

« Est puni de dix ans d'emprisonnement et d'une amende de 1 000 000 €, dont le montant peut être porté au double du produit tiré de l'infraction, le fait, par quiconque, de proposer sans droit, à tout moment, directement ou indirectement, des offres, des promesses, des dons, des présents ou des avantages quelconques à une personne dépositaire de l'autorité publique, chargée d'une mission de service public ou investie d'un mandat électif public, pour elle-même ou pour autrui :

1° Soit pour qu'elle accomplisse ou s'abstienne d'accomplir, ou parce qu'elle a accompli ou s'est abstenue d'accomplir, un acte de sa fonction, de sa mission ou de son mandat, ou facilité par sa fonction, sa mission ou son mandat ;

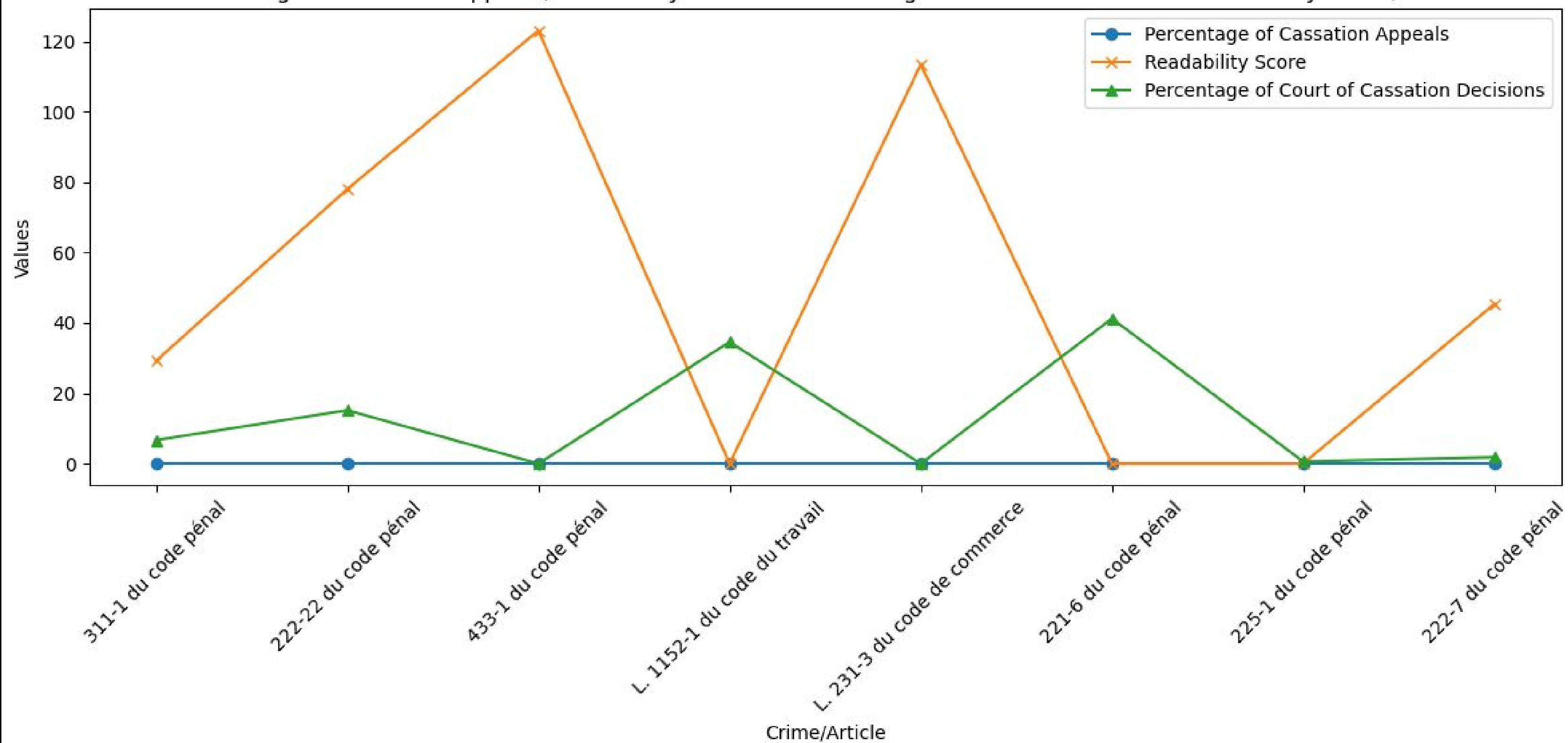
2° Soit pour qu'elle abuse, ou parce qu'elle a abusé, de son influence réelle ou supposée en vue de faire obtenir d'une autorité ou d'une administration publique des distinctions, des emplois, des marchés ou toute autre décision favorable.

Est puni des mêmes peines le fait de céder à une personne dépositaire de l'autorité publique, chargée d'une mission de service public ou investie d'un mandat électif public qui sollicite sans droit, à tout moment, directement ou indirectement, des offres, des promesses, des dons, des présents ou des avantages quelconques, pour elle-même ou pour autrui, pour accomplir ou avoir accompli, pour s'abstenir ou s'être abstenue d'accomplir un acte mentionné au 1° ou pour abuser ou avoir abusé de son influence dans les conditions mentionnées au 2°.

La peine d'amende est portée à 2 000 000 € ou, s'il excède ce montant, au double du produit de l'infraction, lorsque les infractions prévues au présent article sont commises en bande organisée ».



Percentage of Cassation Appeals, Readability Score and Percentage of Court of Cassation Decisions by Crime/Article



# Unexplored biases



1

## Procedural complexity

time limit, burden of proof...

2

## Media coverage

Alters the company's focus on the business and its progress

3

## The topicality

Controversial subjects never before dealt with...

4

## Social status of defendants/victims

White-collar crime would involve people with more means, who could therefore pay for justice for longer.

5

## Missing data

Not all decisions, especially appeals, appear on API.



*Thank  
you!*

**TALK TO US**

CHIOCCHIA Mathilde

ENDERLIN Léo

SILBERREISS Faustine