SUMMARY

This study analyzes the recurrence of fines for non-compliance with the General Data Protection Regulation (GDPR) to determine whether sanctions reduce repeated offenses. The database was obtained through scraping from Enforcement Tracker using Octoparse, and processes of cleaning, sector classification, and translation of types of infringement were applied to improve interpretation.

Data controllers were classified by sector, allowing patterns of recurrence in different industries and countries to be identified. The data was segmented according to the number and amount of fines, as well as the nature of the violations. The results were represented in interactive charts to visualize which sectors and countries accumulate more recurrences and which receive more severe sanctions.

In addition, the relationship between the amount of fines and recurrence was assessed, identifying sectors with greater resistance to regulatory compliance. A particular case is the **Telecommunications** sector, which shows high recurrence but does not appear among the sectors with the highest average fines, suggesting that sanctions may not be sufficiently deterrent.

The results provide a basis for evaluating the effectiveness of GDPR fines and their impact on reducing recurrences, offering key information to improve regulatory strategies and ensure more effective compliance with data protection regulations.

INTRODUCTION

Compliance with the General Data Protection Regulation is a fundamental aspect of the European business sector, especially for organizations that handle large volumes of personal data. Despite the established regulatory framework, many companies continue to violate the regulation, raising doubts about the effectiveness of fines as a deterrent tool. Recurrence in sanctions suggests that in some sectors, fines may not be generating structural changes in data management practices.

The results of the analysis suggest that sanctions, although high in some cases, may not be proportional to the benefit that companies obtain by failing to comply with the regulation, especially in sectors highly dependent on data processing. The central hypothesis suggests that certain sectors, such as telecommunications and technology, show higher recurrence due to the nature of their activity, while others may be more sensitive to sanctions and adjust their compliance more quickly.

Through exploratory analysis and analytical models, this study seeks to identify recurrence patterns and assess whether fines have been effective in encouraging regulatory compliance, providing key information to optimize regulatory strategies

OBJECTIVE

The objective of this project is twofold. On one hand, to conduct an exploratory analysis to understand the patterns of fines imposed under the European Union's GDPR. On the other hand, to determine whether fines imposed for GDPR non-compliance have reduced the recurrence of violations and to analyze their impact on different sectors and countries, identifying patterns that can help improve compliance strategies.

1. Analyze recurrence in GDPR sanctions by classifying sanctioned companies according to their sector and country.
2. Assess the relationship between the amount of fines and the frequency of recurrences, identifying whether higher fines have a greater deterrent effect.
3. Determine which sectors show greater resistance to regulatory compliance by comparing the number of sanctions and their impact in different industries.

**HYPOTHESIS**

Fines reduce the recurrence of GDPR violations.

METHODOLOGY

Data Collection:

The data used in this study was extracted from Enforcement Tracker (https://www.enforcementtracker.com/), a database that compiles and categorizes fines for GDPR breaches in EU member states and the United Kingdom. This source was selected because, to date, it is the only publicly accessible database that centralizes this type of information, although the site itself acknowledges that its content may be incomplete due to variability in the publication of sanctions by the data protection authorities of each jurisdiction.

Data extraction was done through scraping with Octoparse, obtaining information on sanctioned controllers, fine amounts, issuance dates, business sectors, and types of violations.

It is worth noting that the source website explicitly allows this practice, as its robots.txt file specifies "Allow: /", indicating that there are no restrictions for automated data collection on the site. This ensures that the process is carried out in compliance with the site's own access rules.

← → C 🔒 enforcementtracker.com/robots.txt

```
User-agent: *
Allow: /
```
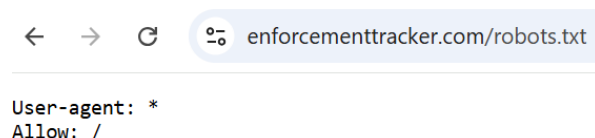
Regarding the date column, data was found in different formats (year only, month and year, full date) and some records without information. To standardize these values, only the year was extracted, ensuring consistency in the temporal analysis.

```python
bbdd['Date_of_Decision'] =  bbdd['Date_of_Decision'].apply(lambda x: str(x)[:4])
```

Other fields also contained missing data, which were filled with the value "Unknown" following the methodology of Enforcement Tracker itself, which classifies certain cases in this way due to a lack of official information. However, in the fine amount column, missing values were not filled in to avoid distorting financial results.

```python
bbdd['ControllerProcessor'] =bbdd['ControllerProcessor'].fillna('Desconocido')
bbdd['Sector'] =bbdd['Sector'].fillna('Desconocido')
bbdd['Type'] =bbdd['Type'].fillna('Desconocido')
bbdd['Date_of_Decision'] = bbdd['Date_of_Decision'].replace('nan', 'Desconocido')
```

The categories for types of infringement were translated and abbreviated to facilitate data visualization and understanding without altering the original GDPR-based categorization.

```python
traducciones = {
    'Non-compliance with general data processing principles': 'Principios Generales',
    'Insufficient technical and organisational measures to ensure information security': 'Medidas Insuficientes',
    'Insufficient legal basis for data processing': 'Base Legal Insuficiente',
    'Insufficient fulfilment of data subjects rights': 'Derechos de los Interesados',
    'Insufficient cooperation with supervisory authority': 'Cooperación Insuficiente',
    'Insufficient fulfilment of information obligations': 'Información Insuficiente',
    'Insufficient fulfilment of data breach notification obligations': 'Notificación de Brechas',
    'Desconocido': 'Desconocido',
    'Insufficient involvement of data protection officer': 'No Participación del DPO',
    'Insufficient data processing agreement': 'Acuerdo de procesamiento Insuficiente'}

bbdd['Type'] = bbdd['Type'].replace(traducciones)
```

Finally, the extracted database was stored in XLSX format, enabling its later processing and analysis.

Variables:

The original data obtained from Enforcement Tracker includes information on fines imposed since the GDPR came into force, that is, between 2018 and 2024. The variables available in the original database are: ETid (unique fine identifier), Country (country where the sanction was imposed), Date of Decision (date of the sanction), Fine_€ (fine amount), Controller/Processor (name of the sanctioned company), Quoted Art. (violated GDPR article), Type (type of infringement), and Source (information source).

During data processing, the Quoted Art. and Type columns were removed as they did not add value to the recurrence analysis. For the Sector variable, which was not included in the original database, a manual classification was performed in batches of 50 rows, using artificial intelligence to assign sectors based on the company name and country of origin. Errors were manually

detected and corrected to improve classification accuracy. A total of 191 distinct sectors were identified in the database.

Statistical Analysis:

A descriptive analysis of the data was carried out using statistical and visualization tools. For processing and analysis, Python was used with libraries such as Pandas for data manipulation. Flourish was used to create interactive visualizations, and Power BI was used to develop a dashboard for easier data exploration.

The dashboard has two tabs, one for general data exploration and another to display the geographic distribution of the data.
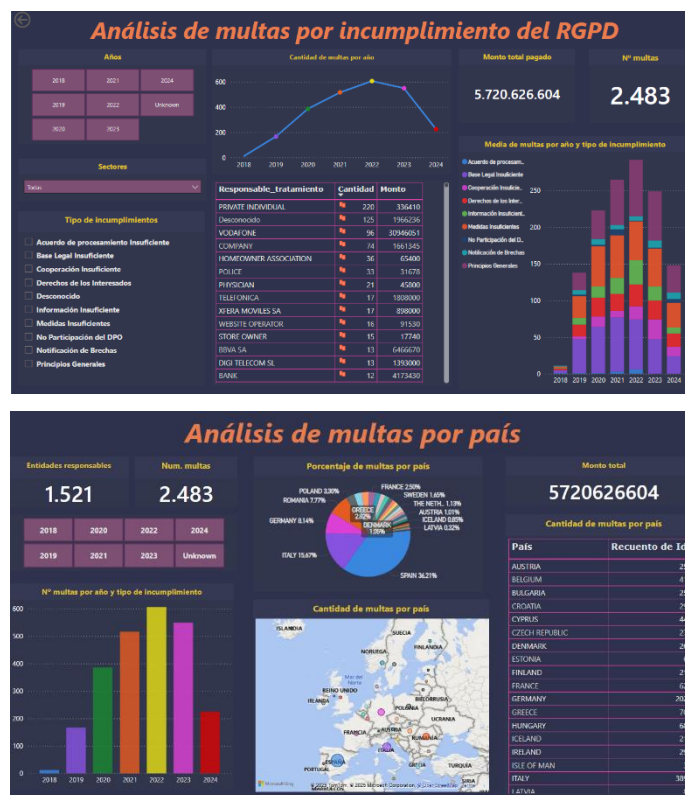
RESULTS

A quick look at the data shows that the number of fines imposed grew steadily from 2018 to 2022, with a significant drop in 2023 and 2024. While this could provide a quick answer, it does not give a complete or accurate response to our hypothesis, so a deeper analysis is required.
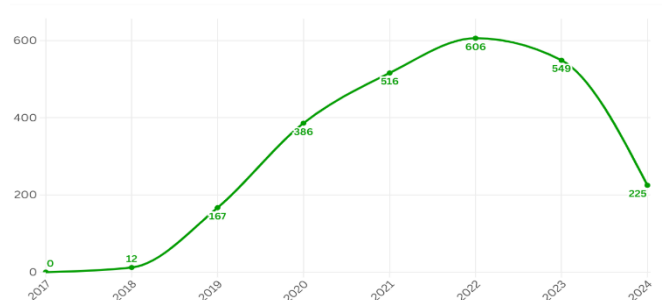
The database contains 8 columns and 2,483 records, identifying 1,522 controllers. These will be the main focus of the study, with special attention to those that have reoffended.

For this study, reoffenders are defined as controllers that have received more than one fine. Applying this criterion, it is observed that out of the 1,522 identified controllers, only 137 have reoffended, representing just over 9% of the total. This suggests that in most cases, fines have been effective, while in a small percentage they have not produced a deterrent effect.

However, beyond the numbers, it is essential to understand the nature of these recurrences. It is necessary to analyze who the reoffenders are, the sectors in which they operate, how much they have paid in fines, and where they have been sanctioned. These factors will help determine whether recurrence is due to sector-specific characteristics, business strategies, or the lack of sufficiently strict regulatory mechanisms.

DISCUSSION

Analysis of Recurrences: Entities with the Most and the Fewest Sanctions
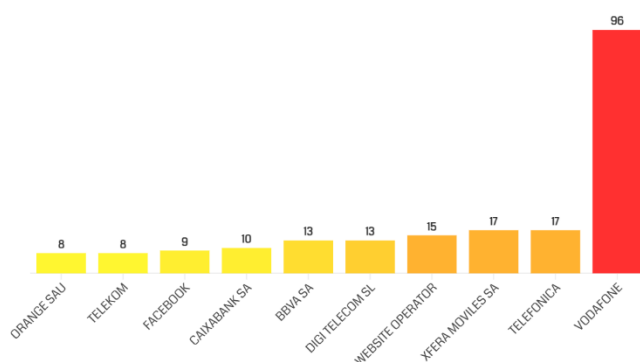
To identify patterns in fine recurrence, controllers with the highest and lowest number of recorded sanctions were analyzed. Due to anonymization applied by authorities in some cases, it was necessary to exclude entities whose names could not be specified, as they group multiple controllers under the same category and cannot be studied in depth.

Entities with the highest number of violations include large corporations in the telecommunications, technology, and financial sectors such as Vodafone, Xfera Moviles SA, Telefonica, BBVA SA, Digi Telecom SL, Caixabank SA, Facebook, Orange SAU, Telekom, and Google.

On the other hand, among the entities with the fewest sanctions are smaller organizations and more diverse sectors. The analysis of low recurrence presented an additional challenge, as 42 entities have received 2 fines, making it impossible to establish a clear ranking.
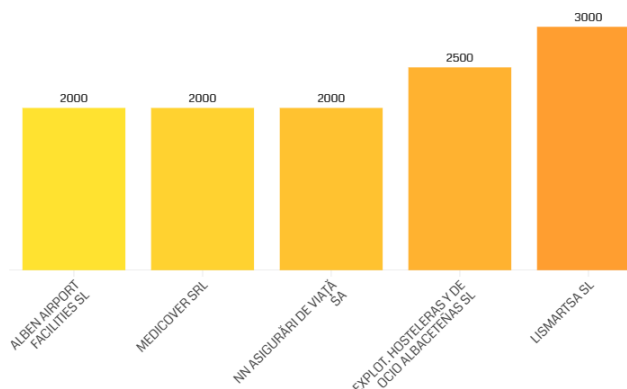
To address this, cases were ordered by the total amount of fines, and a representative sample was selected, consisting of 5 entities with the lowest fines and 5 with the highest within this category.

The study shows that the impact of sanctions varies depending on the sector and business structure of each entity. While some large companies in telecommunications and technology have reoffended multiple times, other similar-sized companies have only been sanctioned once, suggesting that fines can be effective in certain contexts. This indicates that recurrence is not solely dependent on the amount of the fine, but also on the business model and reliance on personal data.
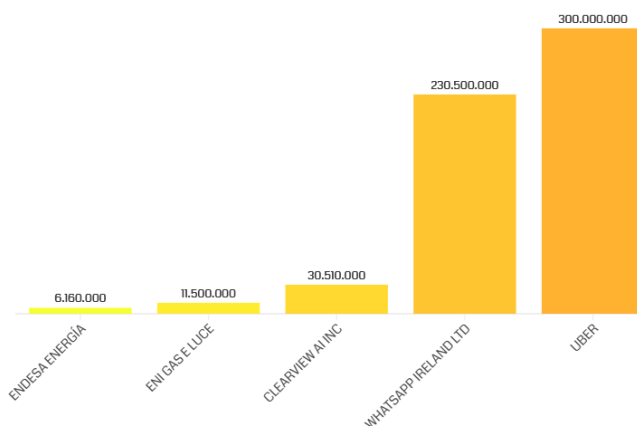
Comparison between Entities with the Highest and Lowest Fines

This analysis identified the 10 reoffenders with the highest accumulated fines for GDPR non-compliance and compared them with those with the lowest fines.

The companies with the highest fines include large tech companies such as Facebook, Amazon, TikTok, LinkedIn, Uber, WhatsApp Ireland Ltd, Google, Enel Energia, Clearview AI Inc, and Criteo. These companies have been fined millions, largely due to the magnitude of the personal data they process and the global impact of their services.
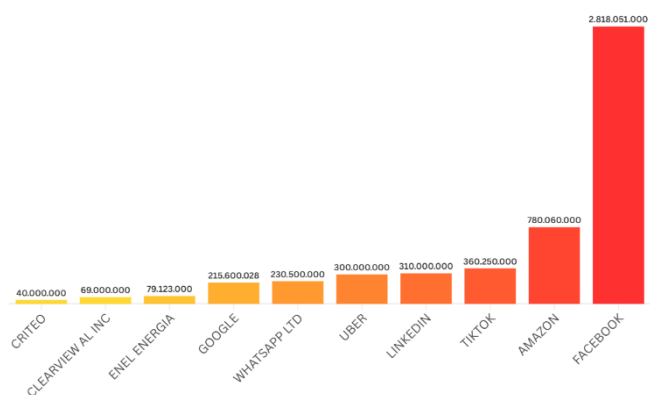
Reoffenders with the lowest fines include CBHNOS SL, "Representative of a Local Government," Directorate of Social and Child Welfare Institutions of the Ferencvaros District of Budapest, MH Vilaseca SL, Carrozados Tecai SL, Y Otro Mas CB, Super 24H Los Rosales SL, Euro Doner Kebab, Inmur Joyeros SL, and "Health Care Worker."

This group is mainly composed of small businesses, government entities, individuals, and social sectors, whose violations seem related to administrative errors or minor deficiencies in compliance.

The contrast between both groups shows a large gap in financial impact and the type of sanctioned entities. While large corporations receive fines that can exceed hundreds of millions of euros, fines for small businesses and local bodies are significantly lower. None of the entities with the lowest fines have reoffended, suggesting that the relative impact of fines may be greater in smaller organizations, while in large tech companies fines may not be a sufficient incentive to change behavior.
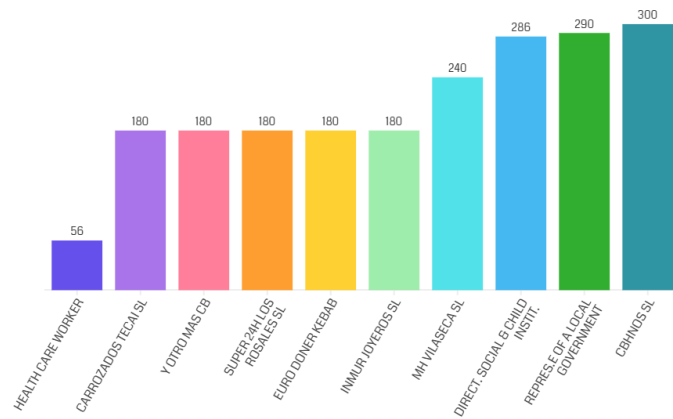
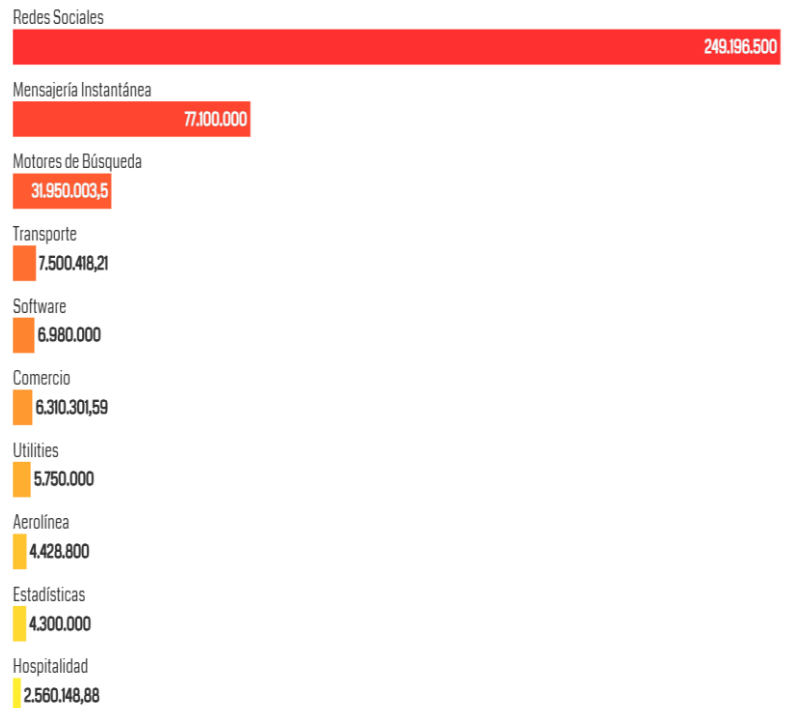
Recurrence Analysis by Sector

Previous findings suggest that recurrence in GDPR sanctions may be more influenced by the company's sector of activity than by its size. Entities with the most violations tend to be concentrated in specific sectors, indicating that certain business areas face greater challenges in regulatory compliance and have greater exposure to repeated sanctions.
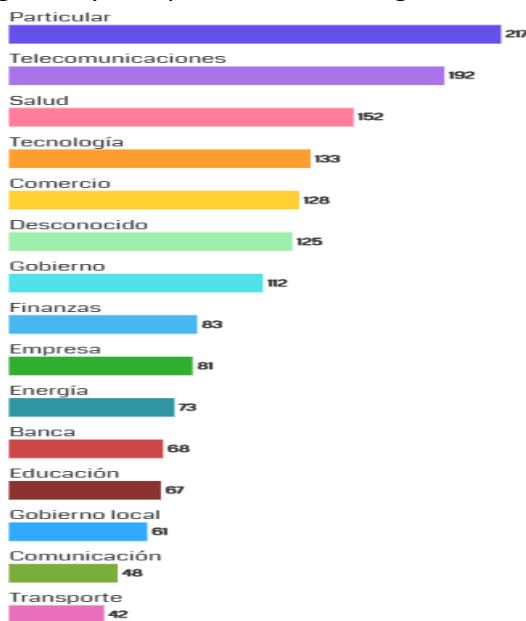
The data shows that recurrence in fines appears closely related to the sector to which the company belongs, suggesting that sectors such as Telecommunications, Technology, and Health tend to accumulate more sanctions due to the type and volume of data they handle, increasing the risk of non-compliance. However, not all companies within these sectors reoffend at the same rate, indicating that recurrence also depends on internal compliance policies and the ability to adapt after a sanctionThe severity of fines is not always linked to the number of sanctions. As noted earlier, companies with few recurrences, such as WhatsApp, have received significantly higher fines than others with multiple minor violations, suggesting that regulators consider the seriousness of the violation and its impact on user privacy.

However, some companies reoffend without fines seeming to generate changes in behavior, raising questions about the effectiveness of fines as a deterrent tool.

A particular case is the Telecommunications sector, which, despite being the second most sanctioned, does not appear among the sectors with the highest average fines. This suggests that although telecommunications companies frequently violate the rules, the economic sanctions they receive are not as severe as in sectors such as Social Media or Search Engines. This difference could be explained by the nature of the infringement or the type or amount of data affected, but the fact that it remains one of the sectors with the most recurrences suggests that current fines may not be generating the necessary impact to improve compliance.

Sector	Amount
Redes Sociales	249.196.500
Mensajería Instantánea	77.100.000
Motores de Búsqueda	31.950.003,5
Transporte	7.500.418,21
Software	6.980.000
Comercio	6.310.301,59
Utilities	5.750.000
Aerolínea	4.428.800
Estadísticas	4.300.000
Hospitalidad	2.560.148,88

CONCLUSION

The analysis of fines imposed for GDPR violations suggests that although the total number of sanctions has decreased in recent years, recurrence remains present in certain sectors and companies. Only 9% of sanctioned entities have reoffended, suggesting that in most cases fines have been effective in preventing new violations. However, some companies, especially in the telecommunications and technology sectors, have accumulated multiple sanctions, raising doubts about whether fine amounts are truly deterrent or whether these companies simply absorb them as an operating cost.

Another key finding is that recurrence seems more related to the sector of activity than to company size. Half of the fines are concentrated in just 10 of the 191 sectors analyzed, suggesting that certain areas have greater difficulty complying with regulations. Additionally, although the number of fines imposed has decreased, many reoffending companies continue to commit the same type of violation, indicating that the issue is not just the existence of fines, but whether they are capable of generating changes in compliance behavior.

In conclusion, <u>GDPR fines have succeeded in reducing recurrence in most cases</u>, but in certain sectors, especially those highly dependent on personal data processing, sanctions do not seem sufficient to ensure effective long-term compliance.