

# 1) Analysis

August 8, 2023

## 1 1) Preparación previa

### 1.0.1 Carga de librerías

```
[1]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from IPython.display import display, Markdown
```

### 1.0.2 Lectura del dataset

Se decidió utilizar unos datos de votos legislativos en EEUU, que describe si se votó afirmativamente, negativamente o absteniéndose en 16 leyes.

```
[2]: df_votes = pd.read_csv('https://raw.githubusercontent.com/Agustin-Bulzomi/
↳Projects/main/Programming/Digital%20House%20(Python)/Support%20Files/
↳Project%203/House_votes_1984.csv')
```

### 1.0.3 Vista general

Se corren varias funciones para obtener un resumen general del dataset

```
[3]: df_votes.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 435 entries, 0 to 434
Data columns (total 17 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Class Name                            435 non-null    object
1   handicapped-infants                   435 non-null    object
2   water-project-cost-sharing            435 non-null    object
3   adoption-of-the-budget-resolution     435 non-null    object
4   physician-fee-freeze                  435 non-null    object
5   el-salvador-aid                       435 non-null    object
6   religious-groups-in-schools           435 non-null    object
7   anti-satellite-test-ban              435 non-null    object
8   aid-to-nicaraguan-contras             435 non-null    object
```

```

9      mx-missile                435 non-null    object
10     immigration                435 non-null    object
11     synfuels-corporation-cutback  435 non-null    object
12     education-spending          435 non-null    object
13     superfund-right-to-sue      435 non-null    object
14     crime                      435 non-null    object
15     duty-free-exports           435 non-null    object
16     export-administration-act-south-africa 435 non-null    object
dtypes: object(17)
memory usage: 57.9+ KB

```

```
[4]: df_votes.head(10)
```

```

[4]:   Class Name  handicapped-infants  water-project-cost-sharing \
0  republican                n                y
1  republican                n                y
2   democrat                ?                y
3   democrat                n                y
4   democrat                y                y
5   democrat                n                y
6   democrat                n                y
7  republican                n                y
8  republican                n                y
9   democrat                y                y

      adoption-of-the-budget-resolution  physician-fee-freeze  el-salvador-aid \
0                n                y                y
1                n                y                y
2                y                ?                y
3                y                n                ?
4                y                n                y
5                y                n                y
6                n                y                y
7                n                y                y
8                n                y                y
9                y                n                n

      religious-groups-in-schools  anti-satellite-test-ban \
0                y                n
1                y                n
2                y                n
3                y                n
4                y                n
5                y                n
6                y                n
7                y                n
8                y                n

```

	n	y
aid-to-nicaraguan-contras		
0	n	y
1	n	n
2	n	n
3	n	n
4	n	n
5	n	n
6	n	n
7	n	n
8	n	n
9	y	n

	synfuels-corporation-cutback	education-spending	superfund-right-to-sue
0	?	y	y
1	n	y	y
2	y	n	y
3	y	n	y
4	y	?	y
5	n	n	y
6	n	n	?
7	n	n	y
8	n	y	y
9	n	n	n

	crime	duty-free-exports	export-administration-act-south-africa
0	y	n	y
1	y	n	?
2	y	n	n
3	n	n	y
4	y	y	y
5	y	y	y
6	y	y	y
7	y	?	y
8	y	n	y
9	n	?	?

## 2 2) Análisis exploratorio

Se comienza con el análisis general: ¿cómo se distribuye la variable target? ¿están balanceados los datos?

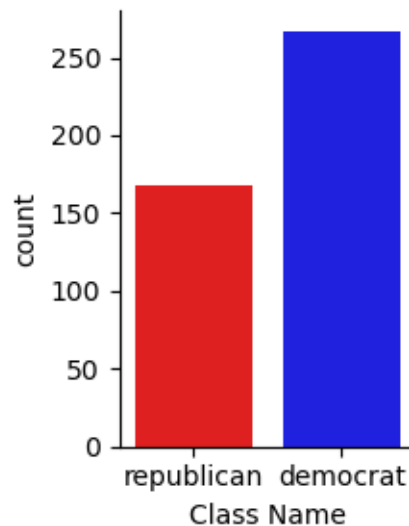
```
[5]: balance = pd.concat([df_votes['Class Name'].value_counts(), df_votes['Class_
    ↳Name'].value_counts(normalize=True).round(2)], axis = 1)
balance.columns = ['Votos', 'Proporción']
balance
```

```
[5]:
```

	Votos	Proporción
democrat	267	0.61
republican	168	0.39

```
[6]: # Se crea una paleta con los típicos colores partidarios, ya que será usada a
      ↪ futuro
color_partidos = ['red', 'blue']

g = sns.catplot(x='Class Name',
                data=df_votes, kind="count", palette=color_partidos,
                height=3, aspect=.8);
```



Hay un importante desbalance considerando que casi 2 tercios de los legisladores son demócratas y 1 tercio republicano. Este dato es importante a la hora de contrastar votos. A continuación se procederá con el análisis más específico, ley por ley. Para ello, se crea una nueva columna con el nombre Target ya que es necesaria para realizar análisis con una variable numérica

```
[7]: df_votes['Target'] = np.where(df_votes['Class Name'] == 'democrat', 1, 0)
```

```
[8]: laws = ['handicapped-infants', 'water-project-cost-sharing',
             'adoption-of-the-budget-resolution', 'physician-fee-freeze',
             'el-salvador-aid', 'religious-groups-in-schools',
             'anti-satellite-test-ban', 'aid-to-nicaraguan-contras', 'mx-missile',
             'immigration', 'synfuels-corporation-cutback', 'education-spending',
             'superfund-right-to-sue', 'crime', 'duty-free-exports',
             'export-administration-act-south-africa']

for law in laws:
    numero = laws.index(law) + 1
```

```

display(Markdown('## {} {} {} {} \n #### {}'.format("Ley N°", numero, ":", "\n",
↪law, "Cantidad de votos afirmativos, negativos y abstenciones:")))
porcentajes = df_votes[law].value_counts()
print(porcentajes)
print() # Un espacio para darle legibilidad
display(Markdown('#### {} '.format("Porcentaje de votos según partido:")))
table = pd.concat([100*(df_votes.groupby(law)['Target'].mean().round(2)),
↪100* (1 - (df_votes.groupby(law)['Target'].mean().round(2)))], axis = 1)
table.columns = ["Democrat", "Republican"]
display(table)
display(Markdown('#### {} '.format("Gráfico de la proporción:")))
grafico = sns.catplot(x = law, hue = "Class Name", data = df_votes,
↪kind="count", palette = color_partidos, height = 3, aspect = .8);
plt.show(grafico)
print() # Un espacio para darle legibilidad

```

## 2.1 Ley N° 1 : handicapped-infants

#### Cantidad de votos afirmativos, negativos y abstenciones:

```

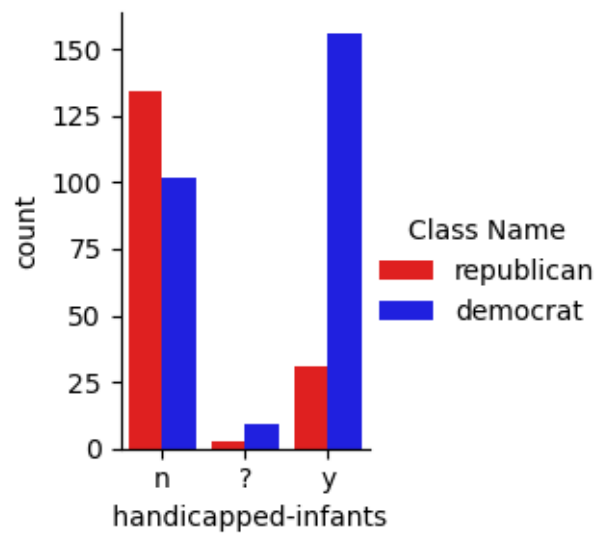
n      236
y      187
?       12
Name: handicapped-infants, dtype: int64

```

Porcentaje de votos según partido:

	Democrat	Republican
handicapped-infants		
?	75.0	25.0
n	43.0	57.0
y	83.0	17.0

Gráfico de la proporción:



## 2.2 Ley N° 2 : water-project-cost-sharing

#### Cantidad de votos afirmativos, negativos y abstenciones:

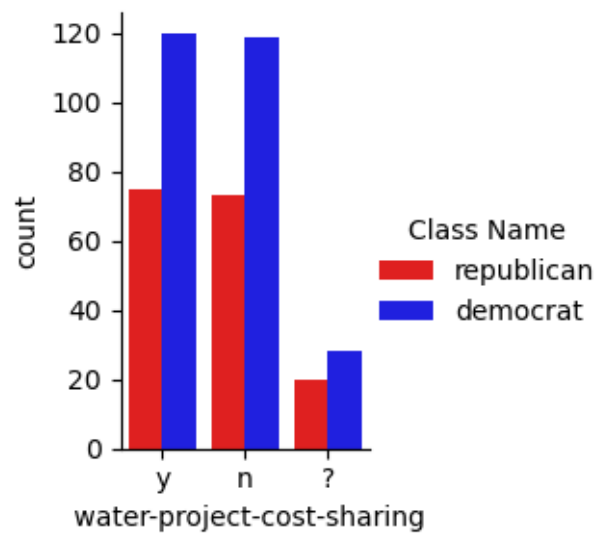
y 195  
n 192  
? 48

Name: water-project-cost-sharing, dtype: int64

Porcentaje de votos según partido:

	Democrat	Republican
water-project-cost-sharing		
?	58.0	42.0
n	62.0	38.0
y	62.0	38.0

Gráfico de la proporción:



### 2.3 Ley N° 3 : adoption-of-the-budget-resolution

#### Cantidad de votos afirmativos, negativos y abstenciones:

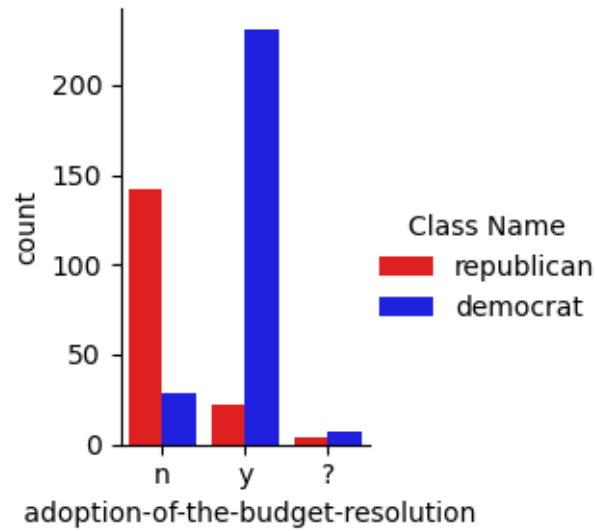
y 253  
n 171  
? 11

Name: adoption-of-the-budget-resolution, dtype: int64

Porcentaje de votos según partido:

	Democrat	Republican
adoption-of-the-budget-resolution		
?	64.0	36.0
n	17.0	83.0
y	91.0	9.0

Gráfico de la proporción:



## 2.4 Ley N° 4 : physician-fee-freeze

#### Cantidad de votos afirmativos, negativos y abstenciones:

```
n    247
y    177
?     11
```

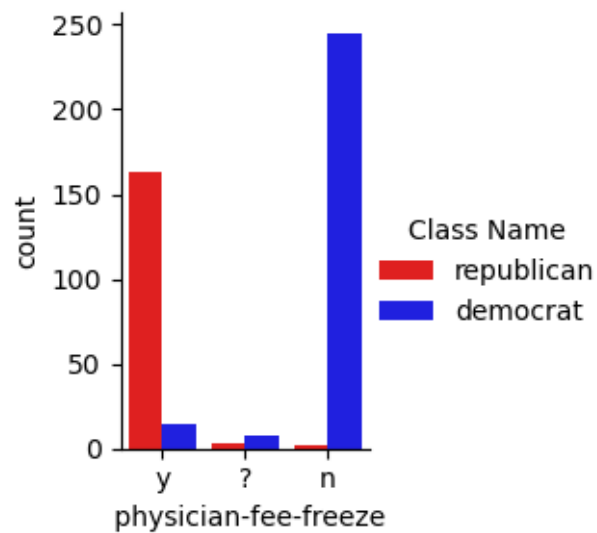
Name: physician-fee-freeze, dtype: int64

Porcentaje de votos según partido:

	Democrat	Republican
physician-fee-freeze		
?	73.0	27.0
n	99.0	1.0
y	8.0	92.0

Gráfico de la proporción:





## 2.5 Ley N° 5 : el-salvador-aid

#### Cantidad de votos afirmativos, negativos y abstenciones:

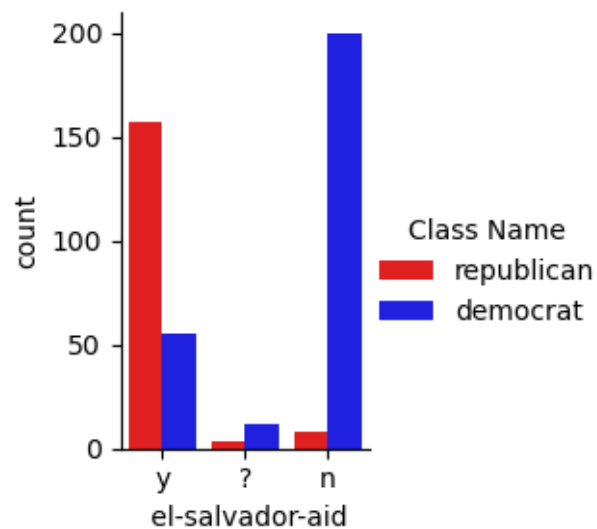
```
y      212
n      208
?       15
```

Name: el-salvador-aid, dtype: int64

Porcentaje de votos según partido:

	Democrat	Republican
el-salvador-aid		
?	80.0	20.0
n	96.0	4.0
y	26.0	74.0

Gráfico de la proporción:



## 2.6 Ley N° 6 : religious-groups-in-schools

#### Cantidad de votos afirmativos, negativos y abstenciones:

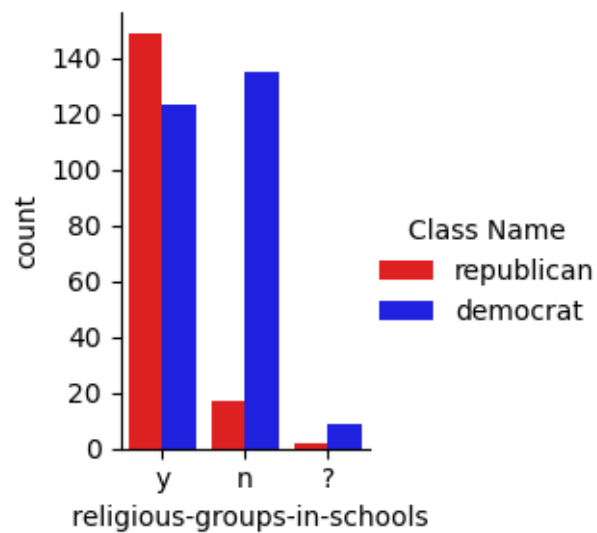
y 272  
n 152  
? 11

Name: religious-groups-in-schools, dtype: int64

Porcentaje de votos según partido:

	Democrat	Republican
religious-groups-in-schools		
?	82.0	18.0
n	89.0	11.0
y	45.0	55.0

Gráfico de la proporción:



## 2.7 Ley N° 7 : anti-satellite-test-ban

#### Cantidad de votos afirmativos, negativos y abstenciones:

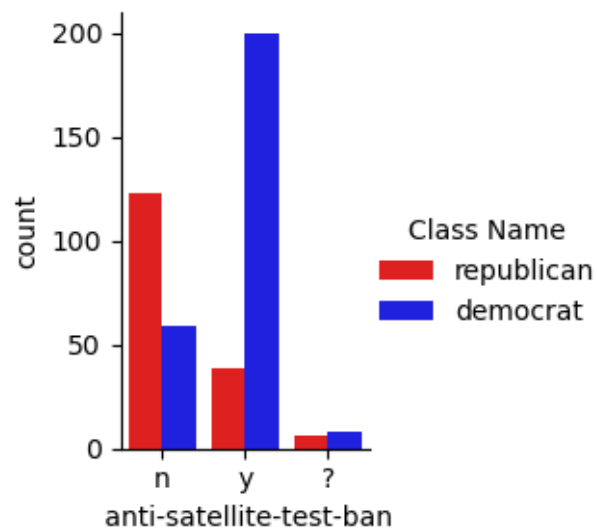
```
y      239
n      182
?       14
```

Name: anti-satellite-test-ban, dtype: int64

Porcentaje de votos según partido:

	Democrat	Republican
anti-satellite-test-ban		
?	57.0	43.0
n	32.0	68.0
y	84.0	16.0

Gráfico de la proporción:



## 2.8 Ley N° 8 : aid-to-nicaraguan-contras

#### Cantidad de votos afirmativos, negativos y abstenciones:

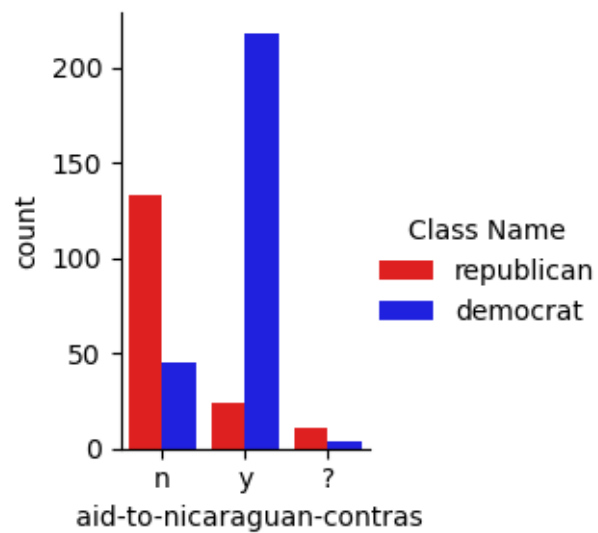
y 242  
n 178  
? 15

Name: aid-to-nicaraguan-contras, dtype: int64

Porcentaje de votos según partido:

	Democrat	Republican
aid-to-nicaraguan-contras		
?	27.0	73.0
n	25.0	75.0
y	90.0	10.0

Gráfico de la proporción:



## 2.9 Ley N° 9 : mx-missile

#### Cantidad de votos afirmativos, negativos y abstenciones:

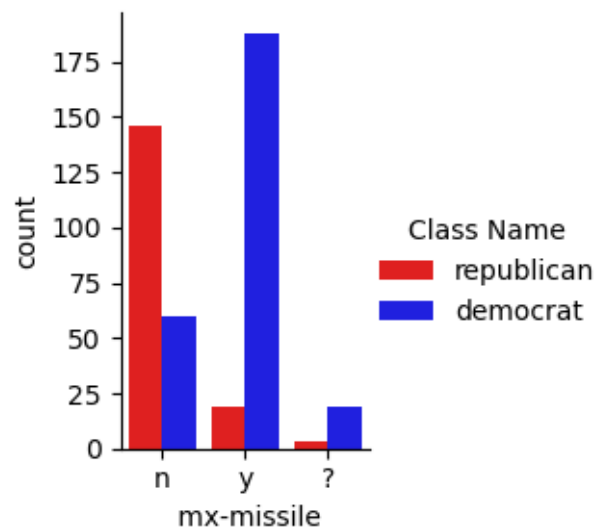
y 207  
n 206  
? 22

Name: mx-missile, dtype: int64

Porcentaje de votos según partido:

	Democrat	Republican
mx-missile		
?	86.0	14.0
n	29.0	71.0
y	91.0	9.0

Gráfico de la proporción:



## 2.10 Ley N° 10 : immigration

#### Cantidad de votos afirmativos, negativos y abstenciones:

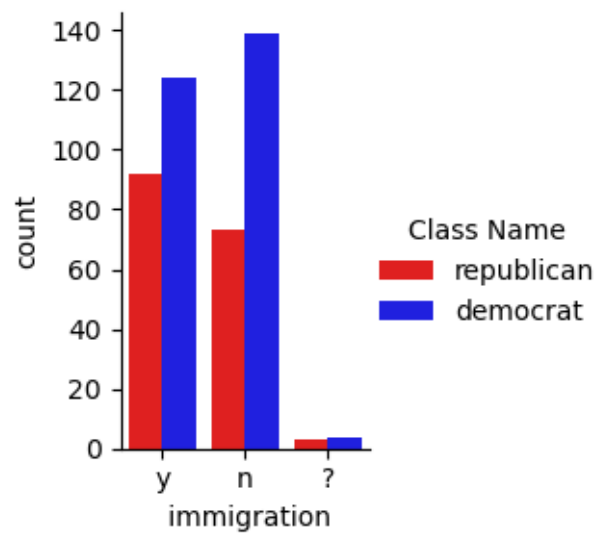
```
y    216
n    212
?      7
```

Name: immigration, dtype: int64

Porcentaje de votos según partido:

	Democrat	Republican
immigration		
?	57.0	43.0
n	66.0	34.0
y	57.0	43.0

Gráfico de la proporción:



## 2.11 Ley N° 11 : synfuels-corporation-cutback

#### Cantidad de votos afirmativos, negativos y abstenciones:

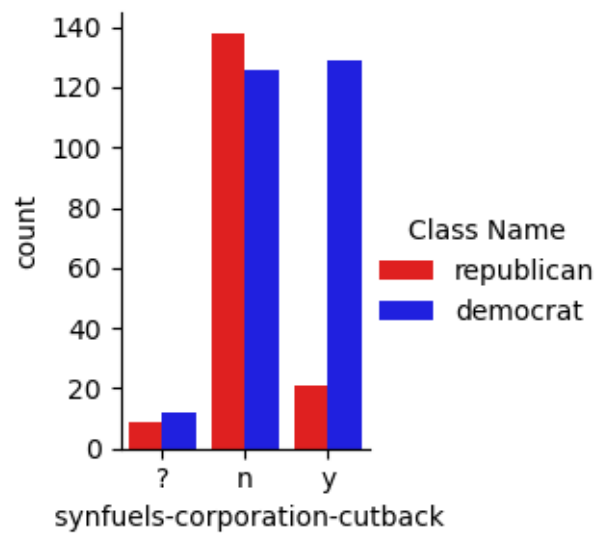
n 264  
y 150  
? 21

Name: synfuels-corporation-cutback, dtype: int64

Porcentaje de votos según partido:

	Democrat	Republican
synfuels-corporation-cutback		
?	57.0	43.0
n	48.0	52.0
y	86.0	14.0

Gráfico de la proporción:



## 2.12 Ley N° 12 : education-spending

#### Cantidad de votos afirmativos, negativos y abstenciones:

```
n    233
y    171
?     31
```

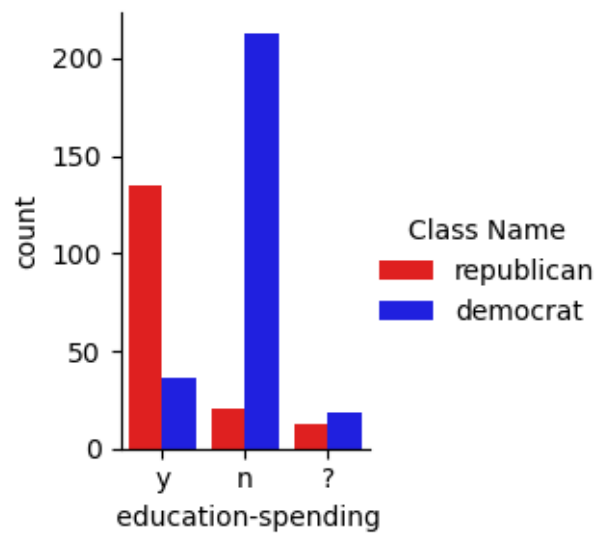
Name: education-spending, dtype: int64

Porcentaje de votos según partido:

	Democrat	Republican
education-spending		
?	58.0	42.0
n	91.0	9.0
y	21.0	79.0

Gráfico de la proporción:





### 2.13 Ley N° 13 : superfund-right-to-sue

#### Cantidad de votos afirmativos, negativos y abstenciones:

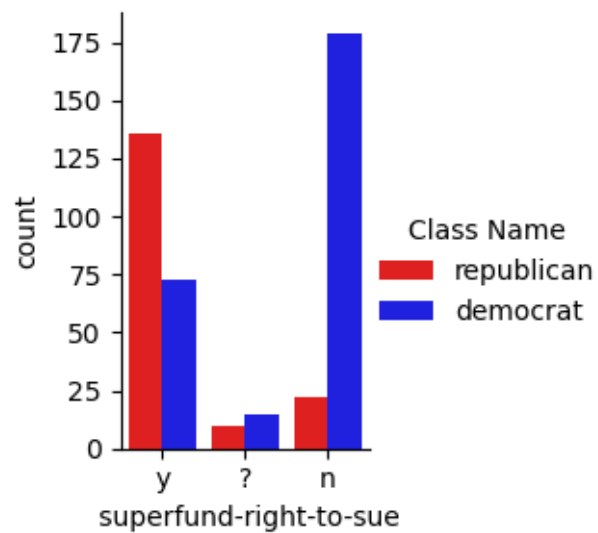
y 209  
n 201  
? 25

Name: superfund-right-to-sue, dtype: int64

Porcentaje de votos según partido:

	Democrat	Republican
superfund-right-to-sue		
?	60.0	40.0
n	89.0	11.0
y	35.0	65.0

Gráfico de la proporción:



## 2.14 Ley N° 14 : crime

#### Cantidad de votos afirmativos, negativos y abstenciones:

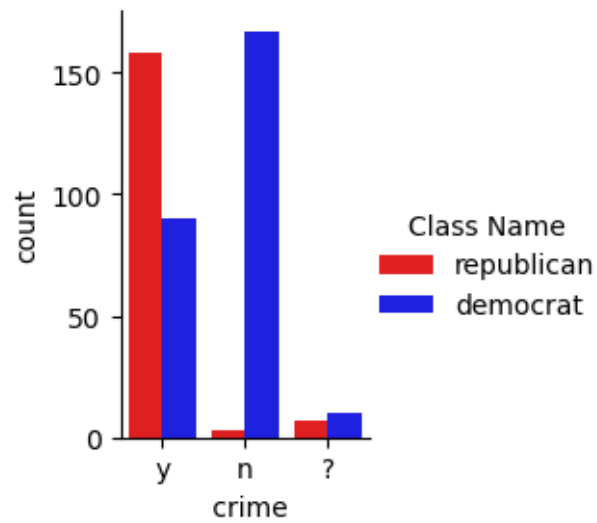
```
y    248
n    170
?     17
```

Name: crime, dtype: int64

Porcentaje de votos según partido:

	Democrat	Republican
crime		
?	59.0	41.0
n	98.0	2.0
y	36.0	64.0

Gráfico de la proporción:



## 2.15 Ley N° 15 : duty-free-exports

#### Cantidad de votos afirmativos, negativos y abstenciones:

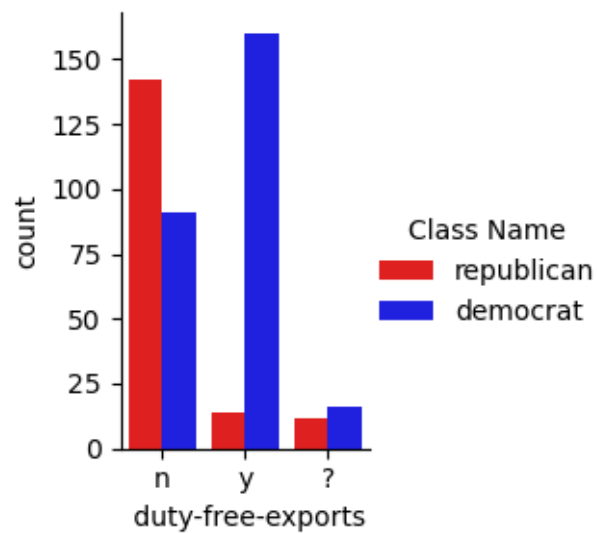
n 233  
y 174  
? 28

Name: duty-free-exports, dtype: int64

Porcentaje de votos según partido:

	Democrat	Republican
duty-free-exports		
?	57.0	43.0
n	39.0	61.0
y	92.0	8.0

Gráfico de la proporción:



## 2.16 Ley N° 16 : export-administration-act-south-africa

#### Cantidad de votos afirmativos, negativos y abstenciones:

y 269

? 104

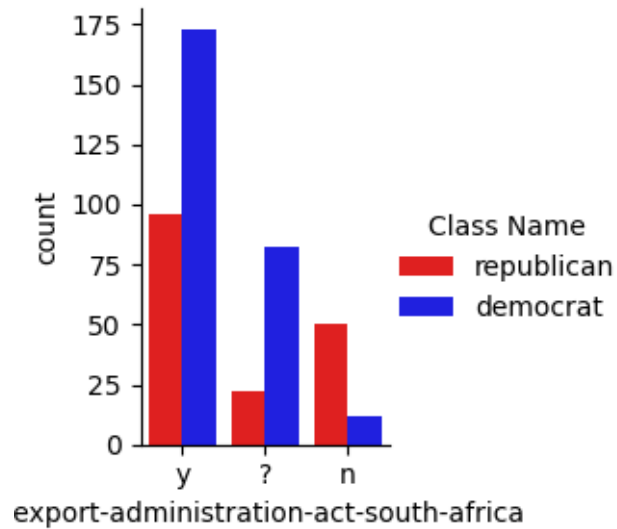
n 62

Name: export-administration-act-south-africa, dtype: int64

Porcentaje de votos según partido:

	Democrat	Republican
export-administration-act-south-africa		
?	79.0	21.0
n	19.0	81.0
y	64.0	36.0

Gráfico de la proporción:



### 3 3) Exportación

Se prepara la base de datos antes de ser exportada para aplicar modelos en otra notebook. Ya la variable target tiene su columna numérica homónima. Ahora se procede a reemplazar los valores string de los votos por otros numéricos (n = -1, ? = 0, y = 1)

```
[9]: votos_replace = df_votes.replace(to_replace = ["n", "?", "y"], value = [-1, 0, 1])
      votos_replace.head(10)
```

```
[9]:
```

	Class Name	handicapped-infants	water-project-cost-sharing	\
0	republican	-1	1	
1	republican	-1	1	
2	democrat	0	1	
3	democrat	-1	1	
4	democrat	1	1	
5	democrat	-1	1	
6	democrat	-1	1	
7	republican	-1	1	
8	republican	-1	1	
9	democrat	1	1	

	adoption-of-the-budget-resolution	physician-fee-freeze	\
0	-1	1	
1	-1	1	
2	1	0	
3	1	-1	

4		1	-1
5		1	-1
6		-1	1
7		-1	1
8		-1	1
9		1	-1

	el-salvador-aid	religious-groups-in-schools	anti-satellite-test-ban \
0	1	1	-1
1	1	1	-1
2	1	1	-1
3	0	1	-1
4	1	1	-1
5	1	1	-1
6	1	1	-1
7	1	1	-1
8	1	1	-1
9	-1	-1	1

	aid-to-nicaraguan-contras	mx-missile	immigration \
0	-1	-1	1
1	-1	-1	-1
2	-1	-1	-1
3	-1	-1	-1
4	-1	-1	-1
5	-1	-1	-1
6	-1	-1	-1
7	-1	-1	-1
8	-1	-1	-1
9	1	1	-1

	synfuels-corporation-cutback	education-spending \
0	0	1
1	-1	1
2	1	-1
3	1	-1
4	1	0
5	-1	-1
6	-1	-1
7	-1	-1
8	-1	1
9	-1	-1

	superfund-right-to-sue	crime	duty-free-exports \
0	1	1	-1
1	1	1	-1
2	1	1	-1

3	1	-1	-1
4	1	1	1
5	1	1	1
6	0	1	1
7	1	1	0
8	1	1	-1
9	-1	-1	0

	export-administration-act-south-africa	Target
0	1	0
1	0	0
2	-1	1
3	1	1
4	1	1
5	1	1
6	1	1
7	1	0
8	1	0
9	0	1

Se exporta la base de datos

```
[10]: votos_replace.to_csv('votos_final.csv', index = False)
```