

POKAŻ KOD

Mounted at /content/drive

POKAŻ KOD

▼ Wstępna eksploracja

Z informacji podanych na <https://www.apispreadsheets.com/datasets/121> dowiadujemy się, że tabela przedstawia informację o głosach kongresmenów w 16 różnych, kluczowych głosowaniach.

Oznaczenie:

- y - kongresmen oddał głos za
- n - głos przeciw
- ? - nie zagłosował, wstrzymał się

POKAŻ KOD

	handicapped_infants	water_project_cost_sharing	adoption_of_the_budget_resolution
0	n	y	n
1	n	y	n
2	?	y	y
3	n	y	y
4	y	y	y

POKAŻ KOD

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 435 entries, 0 to 434
Data columns (total 17 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   handicapped_infants                  435 non-null    object
1   water_project_cost_sharing           435 non-null    object
2   adoption_of_the_budget_resolution    435 non-null    object
3   physician_fee_freeze                 435 non-null    object
4   el_salvador_aid                     435 non-null    object
5   religious_groups_in_schools          435 non-null    object
6   anti_satellite_test_ban             435 non-null    object
7   aid_to_nicaraguan_contras           435 non-null    object
8   mx_missile                          435 non-null    object
9   immigration                         435 non-null    object
10  synfuels_corporation_cutback         435 non-null    object
```

```

11 education_spending          435 non-null    object
12 superfund_right_to_sue      435 non-null    object
13 crime                        435 non-null    object
14 duty_free_exports            435 non-null    object
15 export_administration_act_south_africa 435 non-null    object
16 political_party              435 non-null    object
dtypes: object(17)
memory usage: 57.9+ KB

```

Tabela nie zawiera, żadnych braków informacji. Więc od razu przechodzimy do eksploracji danych. Jeszcze przed tym sprawdzając czy tabela nie zawiera, żadnych błędów.

POKAŹ KOD

```

handicapped_infants          [n, ?, y]
water_project_cost_sharing   [y, n, ?]
adoption_of_the_budget_resolution [n, y, ?]
physician_fee_freeze         [y, ?, n]
el_salvador_aid              [y, ?, n]
religious_groups_in_schools   [y, n, ?]
anti_satellite_test_ban      [n, y, ?]
aid_to_nicaraguan_contras    [n, y, ?]
mx_missile                   [n, y, ?]
immigration                   [y, n, ?]
synfuels_corporation_cutback  [?, n, y]
education_spending           [y, n, ?]
superfund_right_to_sue       [y, ?, n]
crime                         [y, n, ?]
duty_free_exports            [n, y, ?]
export_administration_act_south_africa [y, ?, n]
political_party               [republican, democrat]
dtype: object

```

POKAŹ KOD

Zamieniam dane typu str na numeryczne aby łatwiej je było analizować.

- -1 to głos przeciw
- 0 powstrzymanie się od głosu
- 1 głos za

W ostatniej kolumnie:

- 0 republikanin
- 1 demokrata

POKAŹ KOD

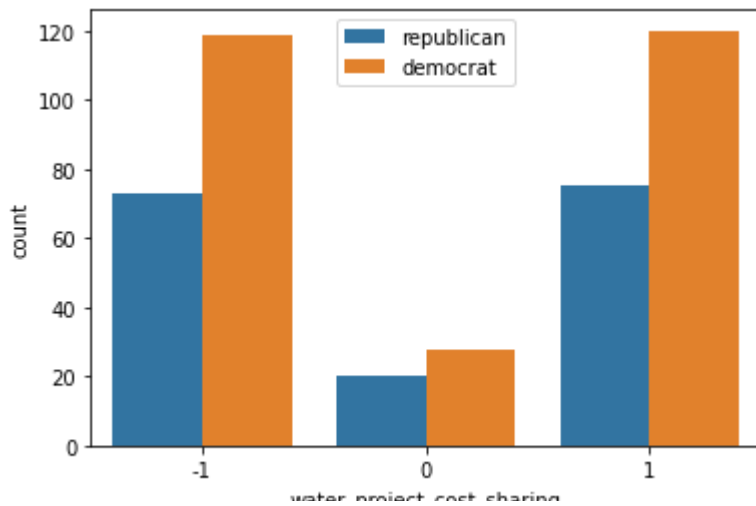
```
<function matplotlib.pyplot.show>
```



Co można zauważyć to, że w naszych danych jest o ok. 100 więcej demokratów. Poza dwoma głosowaniami w których była duża liczba wstrzymań od głosów to głosy rozkładają się mniej więcej równo co może wskazywać na głosowanie według partyjnych podziałów.

Zobaczmy jak rozdzielają się te głosowania które odznaczają się dużą liczbą głosów powstrzymanych. Ciekawe jest to, że głosy nie rozkładają się jasno wobec podziału na partię. A w sprawie pierwszego przedstawionego niżej głosowania wyglądałoby jakby przynależność partyjna nie miała, żadnego znaczenia. Było to na tyle zaskakujące, że aż sprawdziłem dokładne wartości.

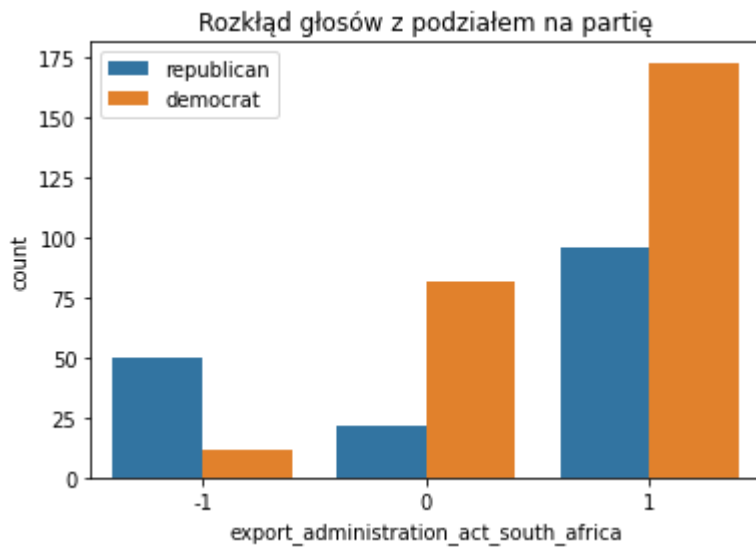
POKAŻ KOD



POKAŹ KOD

```
water_project_cost_sharing  political_party
-1                          0              73
                             1             119
 0                          0              20
                             1              28
 1                          0              75
                             1             120
Name: water_project_cost_sharing, dtype: int64
```

POKAŹ KOD



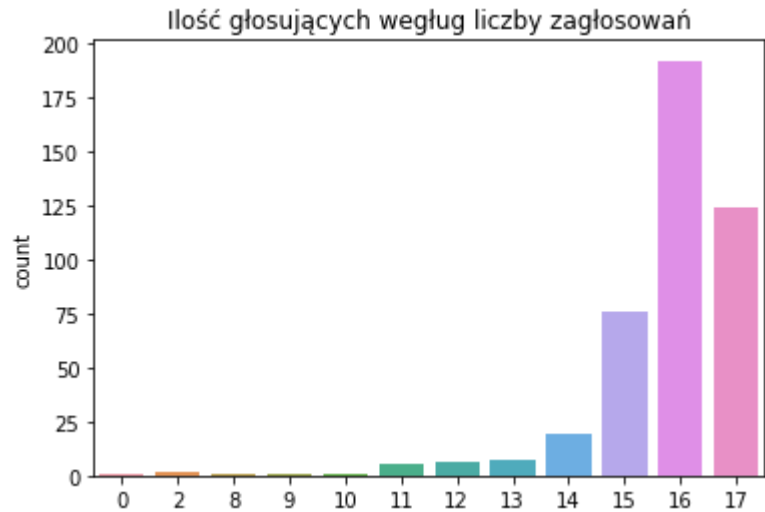
Sprawdzam czy istnieją jakieś *outliery* które w tym wypadku są np powstrzymywaniem się od głosu za każdym razem. Mamy jednego takiego kongresmena

POKAŹ KOD

handicapped_infants	water_project_cost_sharing	adoption_of_the_budget_resolutior
248	0	0

POKAŻ KOD

```
Text(0.5, 1.0, 'Ilość głosujących według liczby zagłosowań')
```



POKAŻ KOD

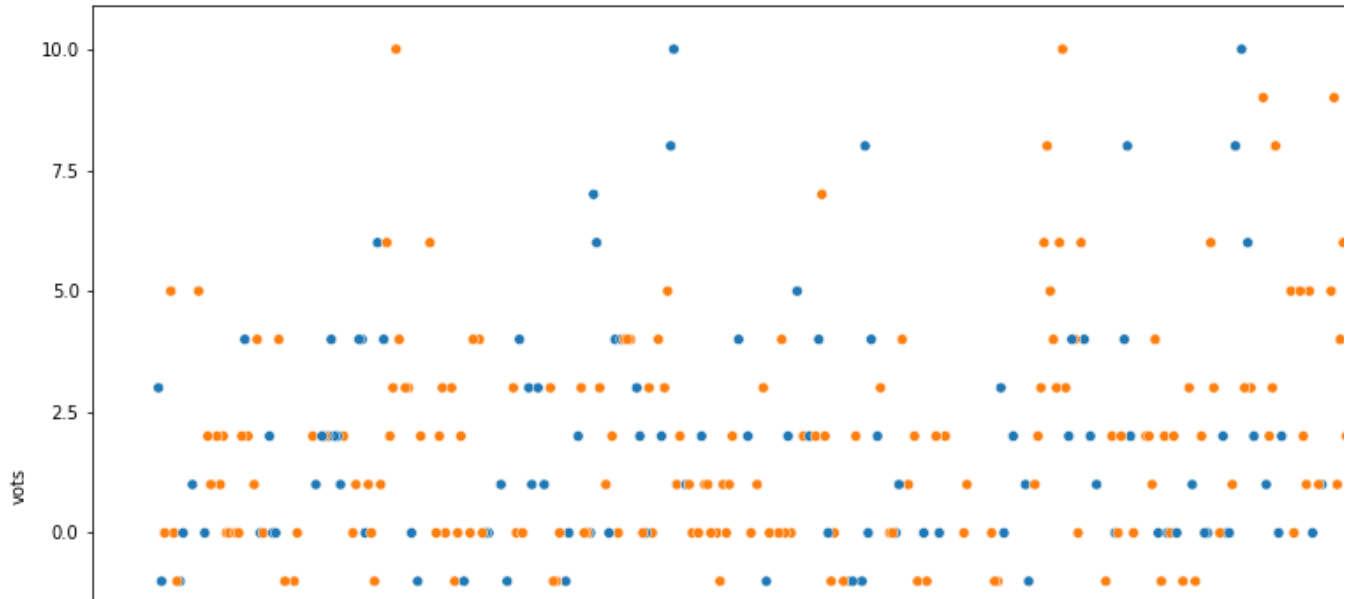
```
[ 2  8  2  0 10  9]
```

▼ Rozkład głosów poszczególnych kongresmenów

Wykres ten miał za zadanie sprawdzić

POKAŻ KOD

POKAŻ KOD

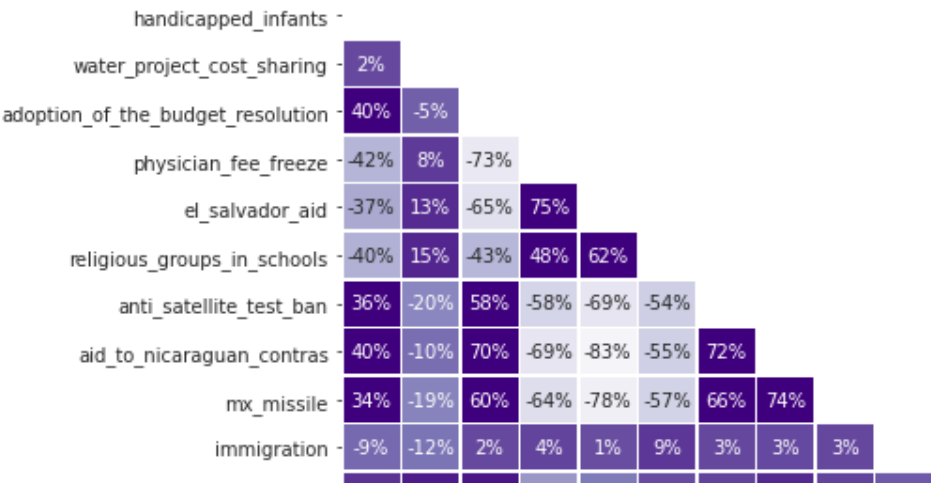


POKAŹ KOD

	handicapped_infants	water_project_cost_sharing	adoption_of_the_budget_resolutor
256	-1	-1	-1
47	1	-1	1
94	1	-1	1

POKAŹ KOD

Macierz korelacji



POKAŻ KOD



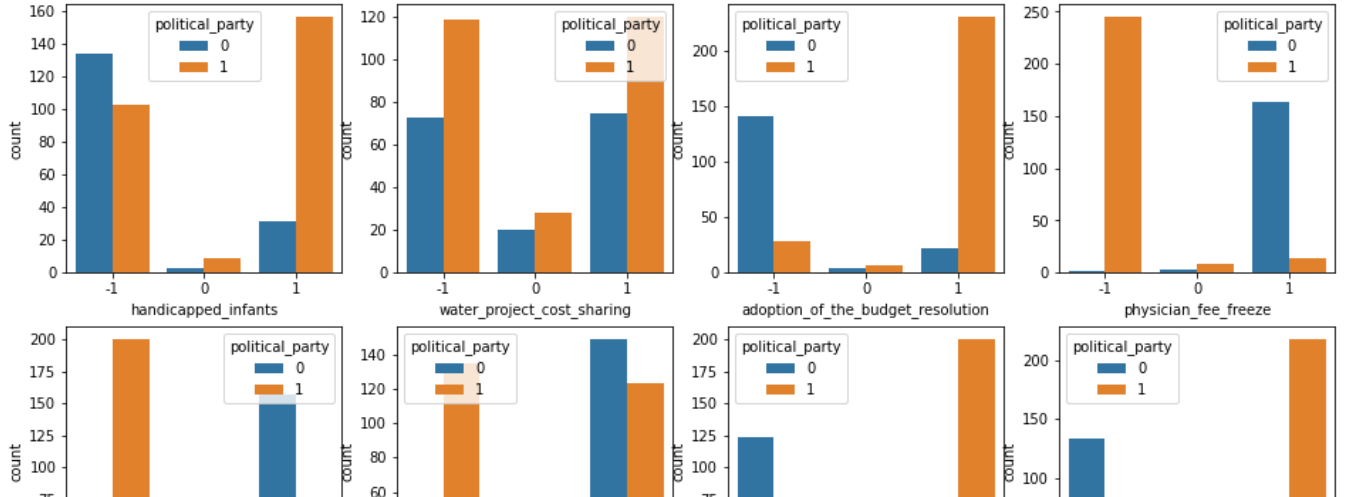
```

IndexError                                Traceback (most recent call last)
<ipython-input-18-4038b9180011> in <module>()
      2
      3 for i in range(0, len(df.columns)):
----> 4     sns.countplot(x=df.columns[i],hue=df.columns[-1], data=df, ax=axes[i//4,
i%4])
      5 plt.legend('republican', 'democrat')
      6 plt.show()

```

IndexError: index 4 is out of bounds for axis 0 with size 4

SEARCH STACK OVERFLOW



POKAŻ KOD



POKAŻ KOD



```
df_rep = grra(df, 0)
```

```
df_dem = grra(df, 0)
```

```
df_rep.index
```

