Contenido

Introducción

Este repositorio ilustra el proceso de ETL (Extracción, Transformación y Carga) y EDA (Análisis Exploratorio de Datos) correspondiente a la primera fase del proyecto denominado “Senado_Col”. Este proyecto tiene como objetivo ampliar la comprensión de la relación actual en Colombia entre el Poder Ejecutivo y el Senado de la República, que ha mostrado cierta adversidad hacia el ejecutivo.

Pregunta de investigación

La pregunta problematizadora que guía este análisis es:

- **¿Cuáles deberían ser las estrategias más efectivas para que el ejecutivo gane incidencia en el Senado, logrando así mayor eficacia legislativa en los 22 meses que restan de gobierno?**
- Alternativamente: **¿Qué factores inciden en la eficacia o ineficacia legislativa del gobierno Petro, particularmente en el Senado?**

Con este análisis, se busca obtener insights que puedan ser útiles para mejorar la estrategia de relacionamiento del gobierno nacional o facilitar la comprensión de los resultados legislativos hacia el cierre del periodo presidencial en 2026.

Tecnologías: Google Colab Notebook, Excel, Python, Pandas, Numpy, Matplotlib, Seaborn, Re, Openpyxl.

Metodología:

Para abordar la pregunta de investigación, se empleará la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining). Este enfoque incluye las siguientes etapas:

1. **Comprensión del negocio**: Definición de objetivos y preguntas de investigación.
2. **Comprensión de los datos**: Recolección y análisis inicial de datos disponibles.
3. **Preparación de datos**: ETL y limpieza de datos.
4. **Modelado**: Exploración de relaciones entre variables.
5. **Evaluación**: Análisis de insights y recomendaciones.
6. **Implementación**: Creación de un Dash Board interactivo en la siguiente fase.

Dado que la disponibilidad de datos puede limitar el análisis, se considerará información cualitativa para enriquecer el estudio.

Algunas salvedades preliminares

- **Frente a los datos:** El dataset con el que se trabajará, es de elaboración propia y muchas de sus variables han sido obtenidas a partir del análisis de prensa. Adicionalmente, la composición del Senado suele ser dinámica y es bastante probable que algunos de los senadores que tenían credencial en el momento de recolección de los datos hoy ya no ejerzan funciones por una u otra razón. En consecuencia, la calidad del dataset puede verse afectada por estas circunstancias y evidenciar algunas sutiles diferencias con la realidad actual del Senado.
- **Frente a los aspectos teóricos subyacentes:** La intención principal de este proyecto es poner en practica algunas metodologías y tecnologías de a Data Analisys, por lo cual no me detendré en definiciones teóricas. Se dan por sentados entonces conceptos como *división de poderes, iniciativa de gobierno, eficacia legislativa y coalición vencedora mínima*.
- **Frente a los contenidos:** Vale la pena advertir a manera de *disclaimer*, que toda la información allí consignada es de carácter público y es fácilmente accesible a través de búsquedas en Google, y que la intención de este proyecto no ha sido profundizar en ella, sino disponibilizarla y centralizarla para facilitar el análisis.
- **Frente al alcance de esta primera etapa:** Como bien se ha señalado desde el inicio, esta primera etapa tiene como fin realizar la limpieza, transformación y análisis exploratorio de datos, por lo que las visualizaciones serán herramientas complementarias y no centrales. En una segunda etapa se espera construir un Dash Board interactivo con Plotly Dash.

Carga y exploración inicial de datos

Para facilitar el proceso de ETL y EDA se trabajó en Google Colab. En primer lugar, los datos se obtienen de una Base de Datos previamente construida con información de prensa en un documento de Excel con extensión .xlsx, gracias a la ayuda del método ***read_Excel()*** de la librería *Pandas* en Python, y se almacenan en la variable ***df***.

Con los métodos ***info()*** y ***head()*** se obtiene una panorámica general del dataset, a saber:

- 14 columnas.

- 103 registros.
- Tipos de datos: float64(4), int64(1), object(9).
- 5 columnas con un número bastante significativo de registros nulos.

```
<class 'pandas.core.frame.DataFrame'>
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 103 entries, 0 to 102
Data columns (total 14 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   id               103 non-null    int64
 1   nombre           103 non-null    object
 2   partido          103 non-null    object
 3   bancada          103 non-null    object
 4   Duro             94 non-null     object
 5   comision         103 non-null    object
 6   ultima_votacion  103 non-null    float64
 7   perfil           102 non-null    object
 8   entidades        9 non-null      object
 9   departamento     87 non-null     object
 10  debilidades      61 non-null     object
 11  apoyo_rechazo    0 non-null      float64
 12  proyecto_ley     0 non-null      float64
 13  voto             0 non-null      float64
dtypes: float64(4), int64(1), object(9)
```

Con esta panorámica general procedo a realizar la limpieza de datos, en la cual prescindiré de las columnas que no aparecen relevantes para mi análisis, y estandarizaré los registros correspondientes a cada una.

Limpieza de datos

Con los métodos **dropna() y fillna()** elimino las variables (o columnas) que poseen más de 45 datos nulos y relleno con “desconocido”, las variables que poseen datos nulos por debajo de 45. Para ello creo una copia del df original en la variable df2.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 103 entries, 0 to 102
Data columns (total 10 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   id               103 non-null    int64
 1   nombre           103 non-null    object
 2   partido          103 non-null    object
 3   bancada          103 non-null    object
 4   Duro             94 non-null     object
 5   comision         103 non-null    object
 6   ultima_votacion  103 non-null    float64
 7   perfil           102 non-null    object
 8   departamento     87 non-null     object
 9   debilidades      61 non-null     object
dtypes: float64(1), int64(1), object(8)
memory usage: 8.2+ KB
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 103 entries, 0 to 102
Data columns (total 10 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   id               103 non-null    int64
 1   nombre           103 non-null    object
 2   partido          103 non-null    object
 3   bancada          103 non-null    object
 4   Duro             103 non-null    object
 5   comision         103 non-null    object
 6   ultima_votacion  103 non-null    float64
 7   perfil           103 non-null    object
 8   departamento     103 non-null    object
 9   debilidades      103 non-null    object
dtypes: float64(1), int64(1), object(8)
memory usage: 8.2+ KB
```

Ahora el df2 contiene 10 columnas, con un total de 103 registros, sin datos nulos.

Así mismo, y haciendo uso de los métodos ***select_dtypes()*** con el argumento ***(include="object")*** y el método ***astypes()*** pasando como argumento ***str***, me aseguro de que todas las variables clasificadas como ***"object"*** contengan datos de tipo str.

Esto con el fin de poder aplicar métodos propios de las cadenas de caracteres, como ***strip()*** y ***title()*** para eliminar espacios vacíos al inicio y al final de los registros, y para tratarlos como títulos, al poner las iniciales en mayúscula.

Finalmente, transformo las variables tipo ***"object"*** en ***"category"***, lo que me facilitará su manejo a la hora de crear ***tablas de frecuencias.*** Pero excluyo de esta transformación las variables ***'Nombre', 'Perfil', 'Debilidades',*** dado que son cadenas de caracteres tipo párrafo de carácter descriptivo y que serán necesarias en la fase cualitativa del análisis.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 103 entries, 0 to 102
Data columns (total 10 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   Id               103 non-null    int64
 1   Nombre           103 non-null    object
 2   Partido          103 non-null    category
 3   Bancada          103 non-null    category
 4   Duro             103 non-null    category
 5   Comision         103 non-null    category
 6   Ultima_Votacion  103 non-null    float64
 7   Perfil           103 non-null    object
 8   Departamento     103 non-null    category
 9   Debilidades      103 non-null    object
dtypes: category(5), float64(1), int64(1), object(3)
memory usage: 7.0+ KB
```

Realizo una iteración de tipo ***For*** para seleccionar todas las variables categóricas e imprimir cada una de las categorías contenidas, para establecer si hay registros repetidos o parecidos que puedan reagruparse.

```
for col in df2.select_dtypes(include='category').columns:
    print(f"Columna: {col}")
    print(df2[col].cat.categories)
```

```
Columna: Partido
Index(['Alianza Cristiana', 'Cambio Radical', 'Centro Democrático',
       'Coalición Centro Esperanza', 'Comunes', 'Curules Indígenas', 'De La U',
       'Pacto Histórico', 'Partido Conservador', 'Partido Liberal'],
      dtype='object')
Columna: Bancada
Index(['Gobierno', 'Independiente', 'Oposición'], dtype='object')
Columna: Duro
Index(['Blando', 'Desconocido', 'Dura', 'Duro', 'Flexible'], dtype='object')
Columna: Comision
Index(['Cuarta', 'Primera', 'Quinta', 'Segunda', 'Sexta', 'Séptima',
       'Tercera'],
      dtype='object')
Columna: Departamento
Index(['Antioquia', 'Antioquia, Suroeste', 'Atlantico', 'Atlantico Y  Bogotá',
       'Atlántico', 'Bogotá', 'Bolivar', 'Bolívar', 'Boyacá', 'Bucaramanga',
       'Caldas', 'Cauca', 'Cesar', 'Cordoba', 'Costa Caribe En General',
       'Cucuta', 'Cundinamarca', 'Córdoba', 'Desconocido', 'Exterior',
       'Guajira', 'Huila', 'Magdalena', 'Meta', 'Nariño', 'Norte De Santander',
       'Pereira', 'Risaralda', 'Santander', 'Santander Y Casanare', 'Sucre',
       'Tolima', 'Valle Del Cauca'],
      dtype='object')
```

Como se puede observar, las columnas "partido", "bancada" y comisión ya se encuentran estandarizadas. Procedo a estandarizar las dos restantes categorías con el método ***replace()***.

Departamento

```
Index(['Antioquia', 'Atlántico', 'Bogotá', 'Bolivar', 'Boyacá', 'Bucaramanga',
       'Caldas', 'Cauca', 'Cesar', 'Cordoba', 'Costa Caribe', 'Cucuta',
       'Cundinamarca', 'Córdoba', 'Desconocido', 'Exterior', 'Guajira',
       'Huila', 'Magdalena', 'Meta', 'Nariño', 'Norte De Santander',
       'Risaralda', 'Santander', 'Casanare', 'Sucre', 'Tolima',
       'Valle Del Cauca'],
      dtype='object')
```

Duro

```
Index(['Desconocido', 'Fuerte', 'Flexible'], dtype='object')
```

Así mismo, estandarizo las variables cuantitativas para que pasen de ser ***Float64*** a ***int64,*** creando la función ***conversion_data_type_int*** que usa los métodos ***replace() y astype().*** Se decide crear una función, dado que es posible que a futuro se deban incluir nuevas variables cuantitativas que requieran esta transformación.

Con esto concluyo la limpieza de la data. Exporto el df haciendo uso del método ***.to_excel().***

Paso ahora al *Análisis Exploratorio* de los datos.

Resumen estadístico

Genero los resúmenes estadísticos de las distintas variables.

Variables cuantitativas:

Llamo el método ***describe()*** con el argumento ***int*** para obtener el resumen estadístico de las variables cuantitativas.

	Id	Ultima_Votacion
count	103.000000	103.000000
mean	53.524272	98512.349515
std	30.836331	88749.111499
min	1.000000	6200.000000
25%	27.500000	60500.000000
50%	53.000000	85000.000000
75%	80.500000	135461.000000
max	106.000000	875554.000000

En la tabla anterior se puede observar un resumen estadístico básico de las variables categóricas continuas. La media, la desviación estándar, la máxima, la mínima y los rangos Intercuartiles (que incluye la mediana correspondiente al Q2 o segundo cuartil).

No me interesa la columna ID dado que se trata de un simple consecutivo. Mi atención se dirige a la variable ***Ultima_Votación.***

En ella se evidencia a primera vista la existencia de un dato que parece ser un *Outlier* o valor atípico. Si la media está en 98.512 votos, un valor de 875.554, casi un millón de votos, no deja de sobresalir. Sin embargo, al ser perfectamente posible por tratarse de una votación popular, no haré más que tomar nota y continuar, pues podría resultar importante para el análisis cualitativo. Pero en principio no afecta en nada esta parte del proyecto.

Variables cualitativas:

Llamo el método ***describe()*** con el argumento ***category*** para obtener el resumen estadístico de las variables cualitativas.

	Partido	Bancada	Duro	Comision	Departamento
count	103	103	103	103	103
unique	10	3	3	7	28
top	Pacto Histórico	Oposición	Fuerte	Primera	Desconocido
freq	17	45	66	23	16

En ella se observa que la variable *Partido* cuenta con 10 categorías, la variable *Bancada* con 3, *Duro* con 3, *Comisión* con 7, *Departamento* con 28. Y se pueden observar cuales son las categorías con mayor frecuencia por cada una de las variables.

Aquí se encuentra información valiosa para tomar en cuenta. Pese a que el "partido" Pacto Histórico (en realidad se trata de una coalición, pero decidí codificarlo como partido, en la nota se explican las razones) es el mayoritario y afín al gobierno, la bancada mayoritaria es de oposición. También se encuentra que 66 de los congresistas se caracterizan por tener una posición Fuerte y que de la mayoría desconocemos sus departamentos de mayor incidencia.

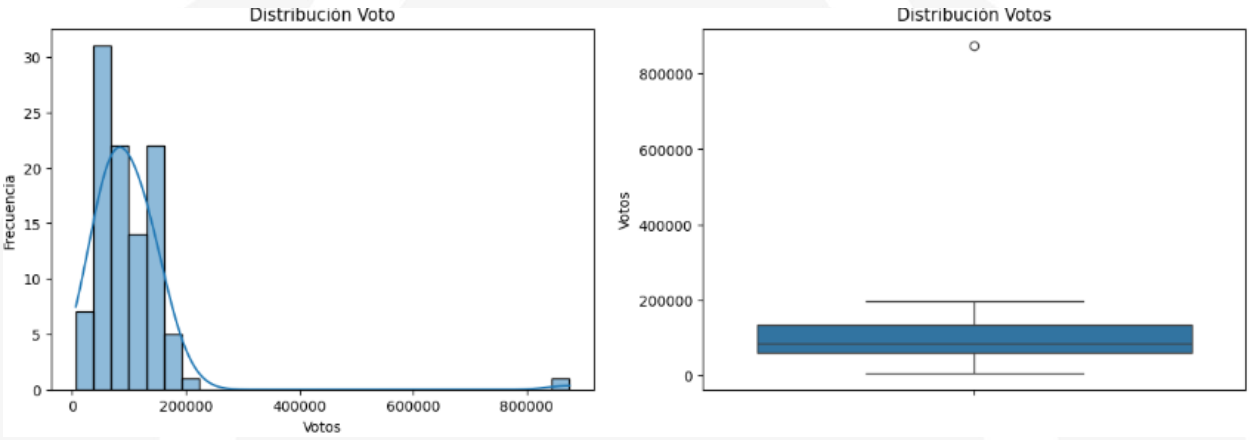
NOTA: Si bien lo adecuado sería la creación de un code book donde se explicara el contenido y alcance de cada una de las categorías, baste aquí con algunas aclaraciones.

1. Las variables *Bancada y Duro*, y la respectiva categorización de los senadores en estas, se construyeron a partir de un análisis cualitativo que buscaba reflejar la realidad del comportamiento de los congresistas, más que su identificación en una coalición específica o discurso.
   En Bancada se clasifican como de gobierno aquellos congresistas que, aún declarados en independencia o de oposición, han sido bastante proclives a apoyar iniciativas legislativas de gobierno. Así mismo, Duro, hace referencia a qué tan Fuerte o Flexible es su posición, lo que permitirá identificar si vale la pena intentar modificarla o no.
2. Por su parte, la variable *Departamento*, hace alusión no al origen del parlamentario sino a la zona geográfica donde tiene mayor influencia política.
3. Finalmente, en la variable partido se usan como categorías unas veces el nombre de coaliciones, otras veces los nombres de partidos, pues las primeras, si bien son la sumatoria de varias personerías jurídicas, han asumido una dinámica muy parecida a las de los partidos. Es el caso del Pacto Histórico y de la Alianza Cristiana, por ejemplo.

Visualización exploratoria de datos

Dado que esta visualización es de carácter exploratorio y su utilidad será exclusivamente ofrecer información acerca de la distribución de las distintas variables y la utilidad de este dataset para el objetivo del proyecto, no cuidaré aquí mucho los asuntos estéticos. Esa tarea contará con especial dedicación en otra fase del proyecto en la que se construya el Dash Board interactivo.

Gráfico 1. Distribución de frecuencias Ultima_Votación



Fuente: datos de compilación propia a partir de prensa.

En este histograma, así como en el gráfico de cajas y bigotes, se puede observar con mayor claridad lo que se mencionó más arriba: hay un dato que está por encima de los 800 mil votos; resulta significativo y habrá que tenerlo en consideración para el análisis cualitativo.

Así mismo, encuentro que los resultados electorales más frecuentes para los senadores están en el rango de los 40 mil (poco más de 30 senadores) y los 90 mil votos (más de 20 senadores). De otro lado, en el gráfico de cajas y bigotes, se observa que la mediana está en torno a los 80 mil votos. Esta información podría llegar a ser relevante en el análisis cualitativo, dado que a mayor votación es posible que las posiciones de los senadores sean menos flexibles, en la medida en que debe cuidar su reputación y mostrarse coherente con su electorado.

Continúo con las variables cualitativas, iniciando con la columna “Partido”. Para ello, empezamos por crear la tabla de frecuencias con ayuda del método ***value_counts().*** También creo un diccionario para

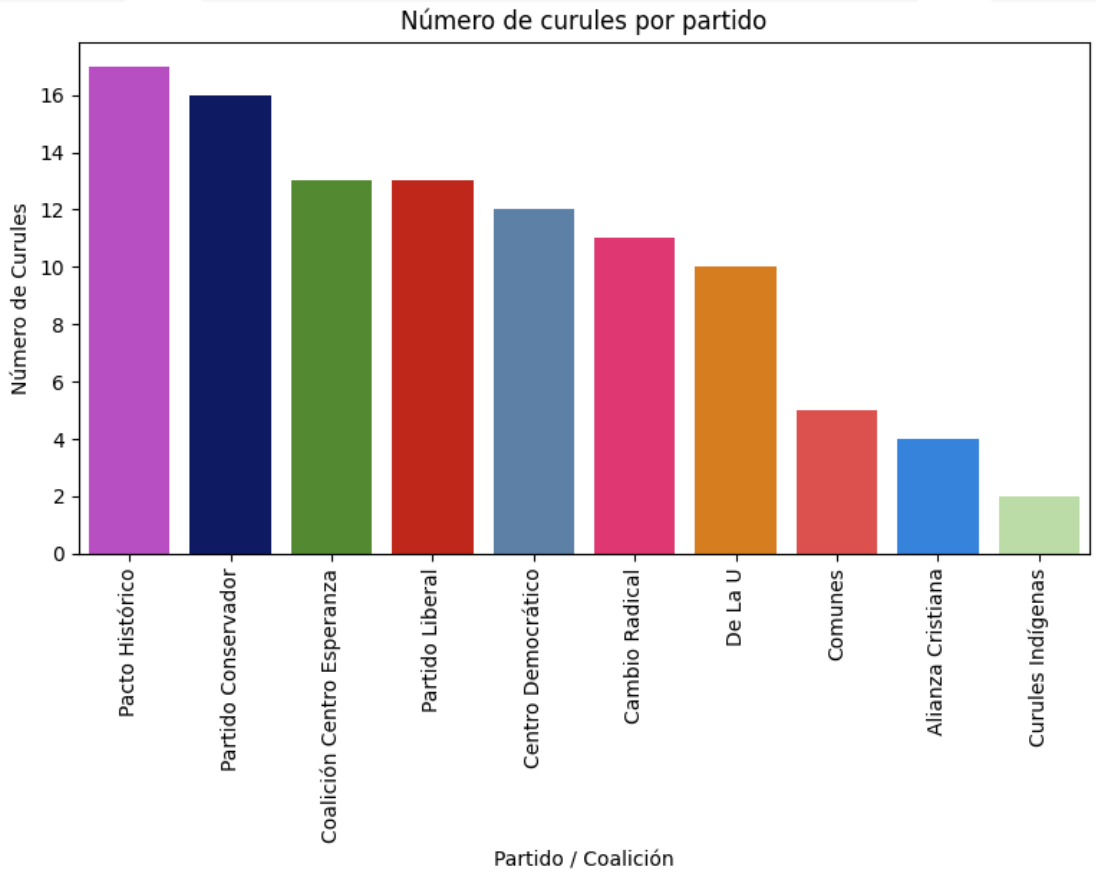
asignar colores a cada partido y posteriormente usarlos en el gráfico de barras con el argumento ***palette=*** del método **barplot()** de **Seaborn**.

	Partido	Frecuencia
0	Pacto Histórico	17
1	Partido Conservador	16
2	Coalición Centro Esperanza	13
3	Partido Liberal	13
4	Centro Democrático	12
5	Cambio Radical	11
6	De La U	10
7	Comunes	5
8	Alianza Cristiana	4
9	Curules Indígenas	2

```
color_partido = {
    "Pacto Histórico" : "#C63CD6",
    "Partido Conservador" : "#001170",
    "Coalición Centro Esperanza" : "#509922",
    "Partido Liberal" : "#DB0F00",
    "Centro Democrático" : "#5080B2",
    "Cambio Radical" : "#FA1B68",
    "De La U" : "#F57E00",
    "Comunes" : "#F43A36",
    "Alianza Cristiana" : "#1A82F5",
    "Curules Indígenas" : "#B9E59E"
}
```

Genero 2 gráficos para representar la distribución del congreso por partido. El primero es un gráfico de barras.

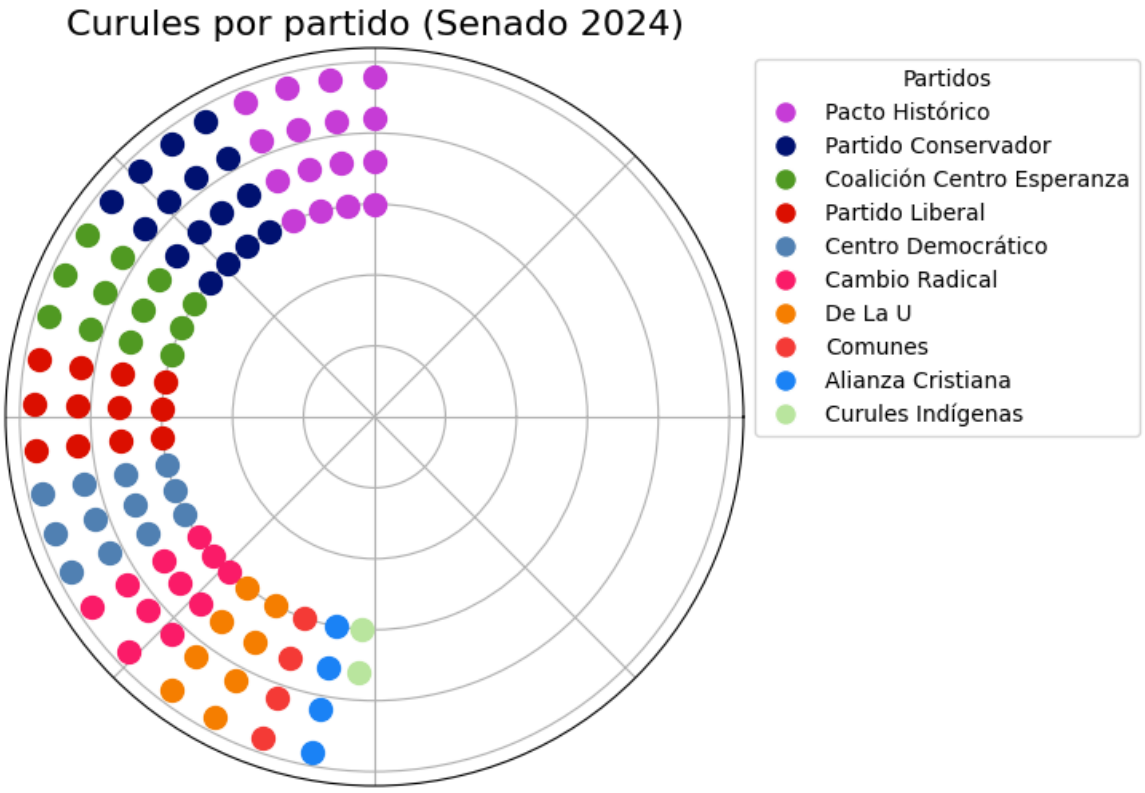
Gráfico 2. Distribución de frecuencias Partido



Fuente: datos de compilación propia a partir de prensa.

Gráfico 2. Hemiciclo distribución de frecuencias Partido

El segundo es un gráfico de hemiciclo que cumple exactamente la misma función que el anterior, pero a mi modo de ver es un poco más ilustrativo.

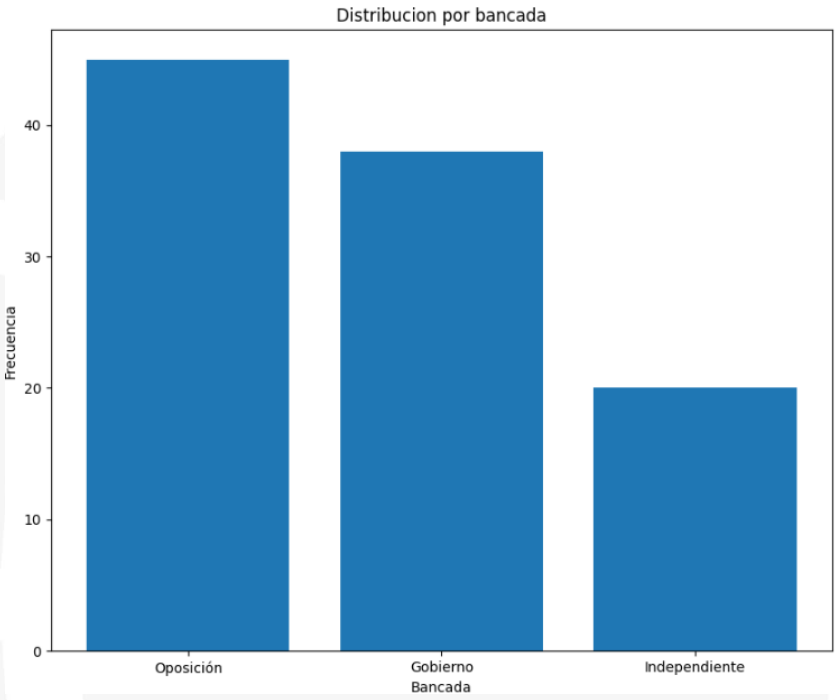


Fuente: datos de compilación propia a partir de prensa.

De aquí en adelante se realizan las tablas de frecuencias y los respectivos gráficos de barras para explorar la distribución de las variables restantes.

Gráfico 3. Distribución de frecuencias Bancada

	Bancada	Frecuencia
0	Oposición	45
1	Gobierno	38
2	Independiente	20



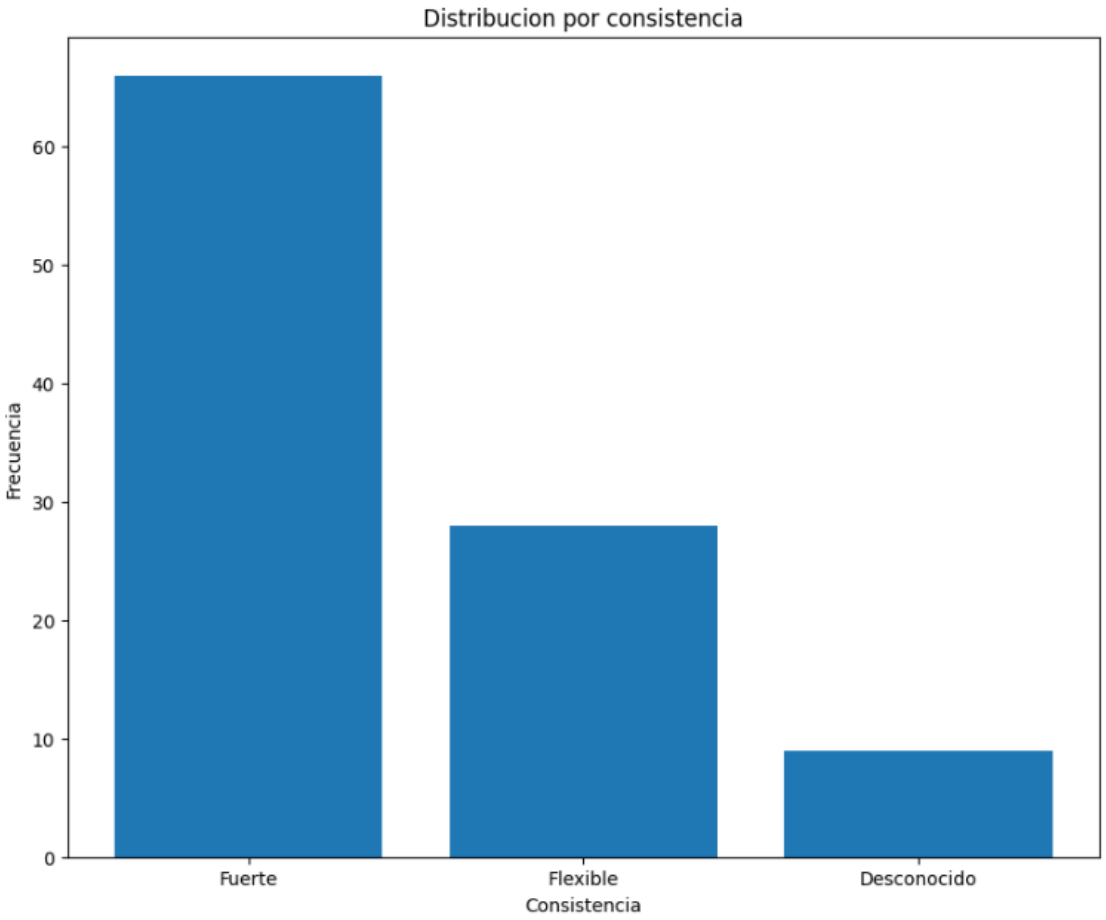
Fuente: datos de compilación propia a partir de prensa.

Al observar la distribución de esta variable construida a partir del análisis cualitativo del comportamiento de los senadores, se puede evidenciar que, si bien el Pacto Histórico “es partido” mayoritario, cuando los senadores se agrupan en función de su comportamiento, la bancada de oposición supera a la bancada de gobierno por 7 votos.

Por su parte, 20 senadores han actuado en independencia y podría llegar a ser más proclives a apoyar iniciativas de gobierno en algún momento. Pero para intentar evidenciar esto, quizá resulte conveniente más adelante explorar la posible correlación entre las variables Bancad y Duro (Consistencia) que a continuación se analiza de manera independiente.

Gráfico 3. Distribución de frecuencias Consistencia (Duro)

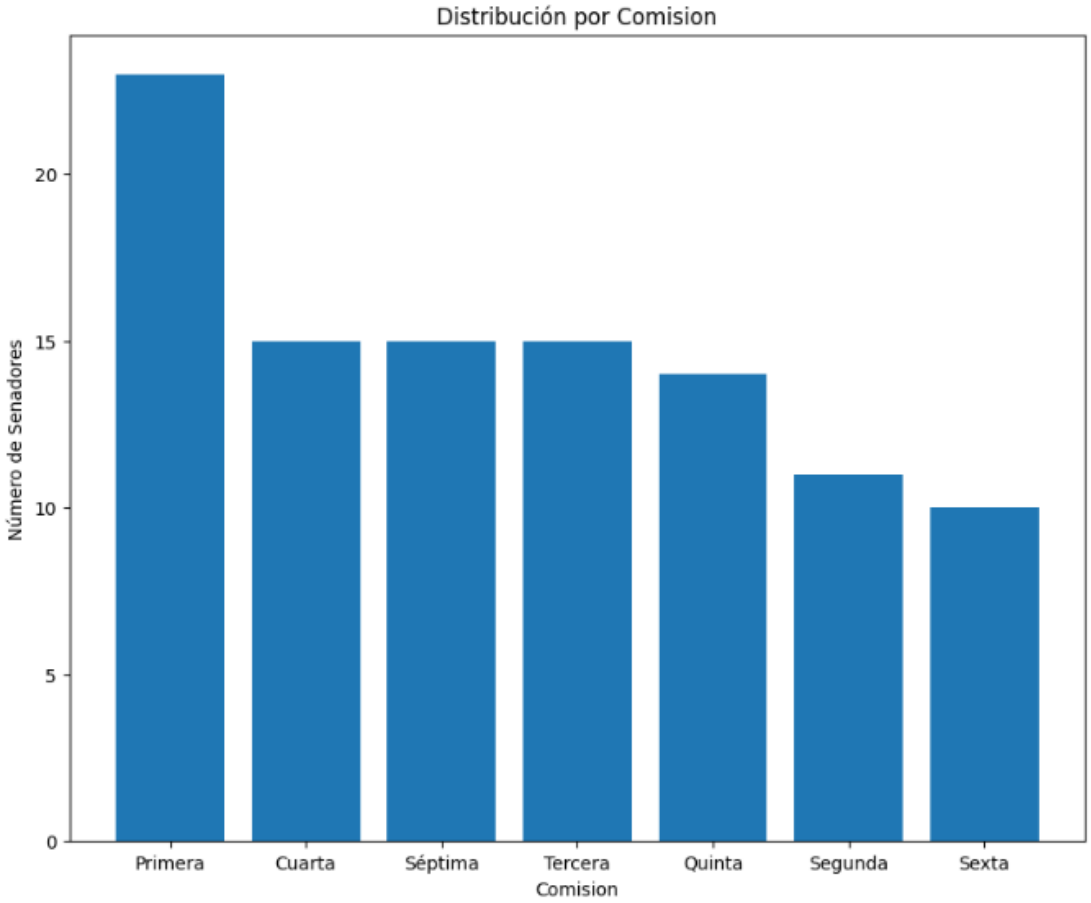
	Consistencia	Frecuencia
0	Fuerte	66
1	Flexible	28
2	Desconocido	9



Fuente: datos de compilación propia a partir de prensa.

Aquí se encuentra que la mayoría de los congresistas tienen una posición claramente definida y que suelen ser consistentes en el comportamiento de apoyar o distanciarse de las iniciativas de gobierno. Esta podría llegar a ser una medida relevante, pero requerirá sustentarse cuantitativamente a partir de los datos de la manera en que los congresistas han votado entre 2022 y 2024. Esto implicará una nueva recopilación de datos.

Gráfico 4. Distribución de frecuencias Comisión



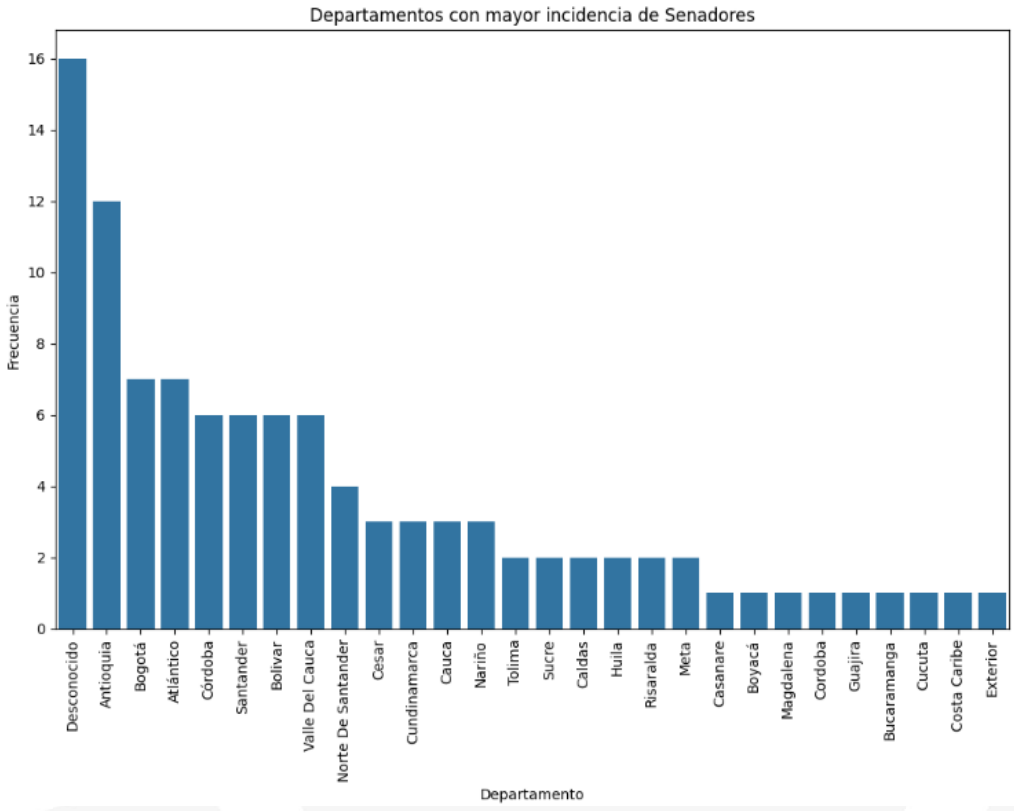
	Comision	Frecuencia
0	Primera	23
1	Cuarta	15
2	Séptima	15
3	Tercera	15
4	Quinta	14
5	Segunda	11
6	Sexta	10

Fuente: datos de compilación propia a partir de prensa.

Aquí aparecen algunas anomalías respecto de la información oficial publicada por el Senado de la República, donde las frecuencias por comisión varían sutilmente. Habrá que realizar una revisión más

exhaustiva de los datos, porque el tipo de comisión relacionado con la posición, flexibilidad o inflexibilidad de los congresistas, podría ser útil a la hora de observar iniciativas legislativas que tengan trámite en cada una de las comisiones, según la temática correspondiente a cada comisión constitucional permanente.

Gráfico 5. Distribución de frecuencias Departamento



	Departamento	Frecuencia
0	Desconocido	16
1	Antioquia	12
2	Bogotá	7
3	Atlántico	7
4	Córdoba	6
5	Santander	6
6	Bolivar	6
7	Valle Del Cauca	6
8	Norte De Santander	4
9	Cesar	3
10	Cundinamarca	3
11	Cauca	3
12	Nariño	3
13	Tolima	2
14	Sucre	2
15	Caldas	2
16	Huila	2
17	Risaralda	2
18	Meta	2
19	Casanare	1
20	Boyacá	1
21	Magdalena	1
22	Cordoba	1
23	Guajira	1
24	Bucaramanga	1
25	Cucuta	1
26	Costa Caribe	1
27	Exterior	1

Fuente: datos de compilación propia a partir de prensa.

Finalmente, la variable Departamento, que mide la incidencia política de congresistas en las distintas regiones, podría ser de interés dado que algunos territorios suelen ser más críticos del gobierno nacional que otros y así suelen expresarlo en las urnas. Sin embargo, la calidad de los datos de esta variable requiere una revisión dado que hay muchos registros desconocidos.

También se podría evaluar si existe alguna correlación entre los congresistas, la región y su posición política, para considerar dicho indicador a la hora de ponderar los congresistas hacia los que debería enfocarse la estrategia del gobierno.

Insights

- **Eficiencia Legislativa y Composición del Senado:** A pesar de que el partido Pacto Histórico es el mayoritario, la bancada opositora tiene una mayor cantidad de senadores. Esto sugiere que, aunque el ejecutivo tiene al partido mayoritario de su lado, la oposición puede dificultar la implementación de su agenda legislativa. Este hallazgo resalta la necesidad de una estrategia más sólida de diálogo y negociación por parte del gobierno.
- **Votaciones y Comportamiento de los Senadores:** La variable Ultima_Votación presenta un outlier significativo (875,554 votos), que podría ser un indicativo de una votación especial o un fenómeno electoral inusual. Este dato es importante ya que podría influir en la percepción pública y en el comportamiento de otros senadores, sugiriendo que un alto número de votos podría correlacionarse con mayor rigidez en la posición política de un congresista.
- **Posiciones Políticas y Consistencia:** La mayoría de los senadores tienen una posición política claramente definida (categorizada como "Duro" o "Flexible"). Esto implica que los senadores que se inclinan a apoyar o oponerse a las iniciativas del gobierno lo hacen de manera consistente. Este insight es crucial para el ejecutivo, ya que la identificación de senadores más flexibles podría ofrecer oportunidades para el diálogo y la negociación.
- **Impacto de la Comisión y la Región:** La exploración de la variable Comisión revela ciertas anomalías respecto a la información oficial.

Esto podría indicar inconsistencias en la clasificación de los senadores según sus comisiones, que son importantes en la evaluación de iniciativas legislativas. Analizar cómo las posiciones de los senadores varían según su comisión podría proporcionar información valiosa sobre dónde enfocar los esfuerzos del ejecutivo.

- **Desconocimiento Regional:** La variable Departamento muestra una cantidad significativa de registros desconocidos. Esto sugiere que algunos senadores podrían no estar suficientemente representando a sus regiones, lo que puede influir en sus decisiones legislativas. Abordar este aspecto podría ayudar al gobierno a identificar senadores que son críticos o más proclives a apoyar sus iniciativas, según su influencia regional.
- **Correlaciones Potenciales:** Se plantea la necesidad de explorar correlaciones entre variables, como la relación entre la bancada (gobierno/oposición) y la consistencia de las posiciones (Duro/Flexible). Esto puede ayudar a identificar patrones de comportamiento que podrían ser estratégicamente relevantes para el gobierno al tratar de ganar apoyo legislativo.
- **Revisión de Datos y Calidad:** La revisión de la calidad de los datos es fundamental, especialmente en las variables que presentan anomalías. Se recomienda realizar una verificación exhaustiva para garantizar que los análisis futuros se basen en datos precisos y completos.

Conclusión:

Dado que la variable dependiente que se busca explicar es “eficacia legislativa del ejecutivo en el Senado de la República de Colombia”, con los datos recabados hasta el momento no es posible realizar un análisis de correlación estadístico que nos permita formular un modelo predictivo, para identificar las oportunidades de mejora.

En consecuencia, se requieren recolectar otros datos como el comportamiento de los senadores en las iniciativas legislativas impulsadas por el gobierno en las dos legislaturas anteriores. Ésta variable sí podría compararse con la variable Duro y Coalición, para establecer un indicador a través del cual se pueda valorar cuáles serían los parlamentarios con mayor probabilidad de apoyar iniciativas de

gobierno, frente a los cuales deberían centrarse los esfuerzos del ejecutivo. Sin embargo, esta parte del análisis será incorporado en otra etapa del proyecto gracias a la flexibilidad que permite la metodología CRISP-DM.