

Procesamiento de Datos Masivos

Tarea 1

1. Consultas

Consulta 1

La primera consulta busca identificar las 5 palabras clave más utilizadas en las intervenciones parlamentarias por cada mes. Primero, en `conteo_palabras`, se agrupan todas las palabras clave asociadas a las intervenciones según el mes en que ocurrieron, contando cuántas veces se mencionó cada palabra en cada mes. Esto se hace uniendo las tablas de intervenciones, palabras clave, y la tabla intermedia que las relaciona. Luego, en `ranking_palabras`, se utiliza la función `ROW_NUMBER()` para asignar una posición o “ranking” a cada palabra según la cantidad de veces que fue mencionada, de forma descendente, dentro de cada mes. Finalmente, se seleccionan únicamente las palabras con un ranking de 5 o menor, es decir, las 5 más frecuentes por mes, ordenando los resultados por mes y posición en el ranking.

```
1 WITH conteo_palabras AS (  
2     SELECT  
3         DATE_TRUNC(intervenciones.fecha, MONTH) AS mes,  
4         keywords.palabra,  
5         COUNT(*) AS conteo  
6     FROM  
7         `proceso-de-datos-454919.tarea1.intervenciones` AS intervenciones  
8     JOIN  
9         `proceso-de-datos-454919.tarea1.intervenciones_keywords` AS  
10        intervenciones_keywords  
11        ON intervenciones.id = intervenciones_keywords.intervencion_id  
12     JOIN  
13        `proceso-de-datos-454919.tarea1.keywords` AS keywords  
14        ON intervenciones_keywords.keyword_id = keywords.id_keyword  
15     GROUP BY  
16        mes, keywords.palabra  
17 ),  
18 ranking_palabras AS (  
19     SELECT  
20         mes,  
21         palabra,  
22         conteo,  
23         ROW_NUMBER() OVER (PARTITION BY mes ORDER BY conteo DESC) AS rank  
24     FROM  
25        conteo_palabras  
26 )  
27 SELECT  
28     mes,  
29     palabra,  
30     conteo  
31 FROM  
32     ranking_palabras  
33 WHERE  
34     rank <= 5
```

```

34 ORDER BY
35 mes, rank;

```

Por ejemplo en enero de 2023, los temas más mencionados fueron relacionados con la política y la constitución. 'Presidente' se destacó con 594 menciones, seguido por 'constitucional' con 367 menciones. Se ha supuesto que en el texto se dirige la palabra o agradece la introducción al presidente o presidenta cuando se empieza la intervención. Por lo tanto, no da una temática real. También puede sugerir un enfoque en la figura presidencial y asuntos constitucionales. 'República' (317 menciones) y 'Chile' (284 menciones) también fueron términos relevantes, lo que podría reflejar debates sobre la situación política del país.

Consulta 2

La segunda consulta calcula una media móvil del número de intervenciones realizadas por cada partido político, agrupadas por mes. Primero, en el CTE llamado intervenciones_mes, se cuenta el número total de intervenciones por partido en cada mes. Para lograrlo, se hace una unión entre las tablas de intervenciones, parlamentarios y partidos, y se agrupa por nombre del partido y mes. Luego, en la consulta principal, se utiliza una ventana temporal para calcular el promedio móvil de las intervenciones, tomando el promedio de las intervenciones en el mes actual y los dos meses anteriores (ROWS BETWEEN 2 PRECEDING AND CURRENT ROW) para cada partido político. Esto permite observar la tendencia de actividad parlamentaria por partido a lo largo del tiempo. Los resultados se ordenan cronológicamente por mes.

```

1 WITH intervenciones_mes AS (
2     SELECT
3         p.nombre_partido AS partido_politico,
4         DATE_TRUNC(i.FECHA, MONTH) AS mes,
5         COUNT(i.ID) AS total_intervenciones
6     FROM `proceso-de-datos-454919.tarea1.intervenciones` i
7     JOIN `proceso-de-datos-454919.tarea1.parlamentarios` pa ON
8         i.PARLAMENTARIO_ID = pa.id_parlamentario
9     JOIN `proceso-de-datos-454919.tarea1.partidos` p ON pa.PARTIDO_ID =
10         p.id_partido
11     GROUP BY p.nombre_partido, mes
12 )
13 SELECT
14     partido_politico,
15     mes,
16     AVG(total_intervenciones) OVER (
17         PARTITION BY partido_politico
18         ORDER BY mes
19         ROWS BETWEEN 2 PRECEDING AND CURRENT ROW
20     ) AS media_movil_intervenciones
21 FROM intervenciones_mes
22 ORDER BY mes;

```

Los resultados muestran la media de intervenciones por un mes y sus dos anteriores. Por ejemplo, el Partido Socialista de Chile muestra una media superior de intervenciones para marzo de 2023, con un valor de 430. Por otro lado, el partido con menos intervenciones se trata del Partido Ecologista Verde. En general, para todas las fechas consideradas, el partido Renovación Nacional y Unión Demócrata Independiente, tienen en general más intervenciones.

Consulta 3

Para la siguiente consulta se busca identificar el tema más frecuente en las intervenciones parlamentarias de cada partido político durante cada trimestre. Primero, en una etapa inicial, se agrupan las intervenciones según el partido, el trimestre en que se realizaron y la palabra clave asociada a cada intervención, contando cuántas veces fue mencionada cada una. Para esto, se conectan las tablas de intervenciones, parlamentarios, partidos políticos, y palabras clave a través de sus relaciones correspondientes. Luego, en la parte principal de la consulta, se selecciona el tema con mayor número de menciones por partido y trimestre. Esto se logra comparando, dentro de cada grupo, cuál fue el tema con la frecuencia más alta. El resultado final es un listado donde, para cada trimestre, aparece el tema principal tratado por cada partido, ordenado por fecha y nombre del partido.

```
1  WITH temas_trimestre AS (
2      SELECT
3          p.nombre_partido AS partido_politico,
4          DATE_TRUNC(i.FECHA, QUARTER) AS trimestre,
5          k.palabra AS tema,
6          COUNT(k.id_keyword) AS frecuencia
7      FROM `proceso-de-datos-454919.tarea1.intervenciones` i
8      JOIN `proceso-de-datos-454919.tarea1.parlamentarios` pa ON
9      i.PARLAMENTARIO_ID = pa.id_parlamentario
10     JOIN `proceso-de-datos-454919.tarea1.partidos` p ON pa.PARTIDO_ID =
11     p.id_partido
12     JOIN `proceso-de-datos-454919.tarea1.intervenciones_keywords` ik ON
13     i.ID = ik.intervencion_id
14     JOIN `proceso-de-datos-454919.tarea1.keywords` k ON ik.keyword_id =
15     k.id_keyword
16     GROUP BY p.nombre_partido, trimestre, k.palabra
17 )
18 SELECT
19     partido_politico,
20     trimestre,
21     tema AS tema_principal
22 FROM temas_trimestre t1
23 WHERE frecuencia = (
24     SELECT MAX(t2.frecuencia)
25     FROM temas_trimestre t2
26     WHERE t2.partido_politico = t1.partido_politico
27     AND t2.trimestre = t1.trimestre
28 )
29 ORDER BY trimestre, partido_politico;
```

Durante el período analizado (2023-2024), se observa que el tema del 'presidente/presidenta' domina de forma transversal en casi todos los partidos, mostrando el interés constante por la figura presidencial y liderazgo político. Sin embargo, hay matices relevantes:

- El Partido Ecologista Verde muestra una gran variedad de temas, lo que refleja una agenda diversa que va desde temas medioambientales hasta telecomunicaciones y gestión municipal.
- Movimiento Amarillos por Chile toma fuerza especialmente en 2024, con una explosión temática vinculada a constitución, democracia, seguridad y gobernabilidad, indicando una expansión activa en su discurso.
- Partidos tradicionales como Renovación Nacional, UDI y el Partido Socialista también repiten temas como “presidente”, “municipalidad” o “indicaciones”, apuntando a una actividad estable pero menos temática que la de partidos emergentes.

- Hay presencia de temas más específicos como “hirsch” (Partido Acción Humanista), “lagos” (PPD), o “provoste” (RN), lo que muestra el protagonismo de figuras políticas particulares en ciertos momentos.

En resumen, aunque la palabra ‘presidente’ sigue siendo el más repetido, hay señales de diversificación en la agenda de partidos más pequeños o emergentes, mientras que los grandes mantienen una línea más constante.

Consulta 4

La última consulta trata de obtener los tres temas más mencionados por cada partido político en un mes, se decidió tomar de ejemplo enero de 2023, ya que es la fecha más reciente. En una primera fase, se agrupan todas las intervenciones de ese mes, relacionándolas con el partido político correspondiente y las palabras clave asociadas, contando cuántas veces fue mencionada cada palabra. Posteriormente, se utiliza una función de ordenamiento (RANK()) que asigna una posición a cada tema según su frecuencia, dentro de cada partido. Finalmente, se filtran únicamente los tres temas mejor posicionados para cada partido, es decir, aquellos con mayor cantidad de menciones. El resultado presenta los tres temas más relevantes para cada partido en enero de 2023, junto con el número de veces que fueron tratados, organizados alfabéticamente por partido político.

```

1  WITH temas_por_mes AS (
2      SELECT
3          p.nombre_partido AS partido_politico,
4          DATE_TRUNC(i.FECHA, MONTH) AS mes,
5          k.palabra AS tema,
6          COUNT(k.id_keyword) AS frecuencia
7      FROM `proceso-de-datos-454919.tarea1.intervenciones` i
8      JOIN `proceso-de-datos-454919.tarea1.parlamentarios` pa ON
9      i.PARLAMENTARIO_ID = pa.id_parlamentario
10     JOIN `proceso-de-datos-454919.tarea1.partidos` p ON pa.PARTIDO_ID =
11     p.id_partido
12     JOIN `proceso-de-datos-454919.tarea1.intervenciones_keywords` ik ON
13     i.ID = ik.intervencion_id
14     JOIN `proceso-de-datos-454919.tarea1.keywords` k ON ik.keyword_id =
15     k.id_keyword
16     WHERE DATE_TRUNC(i.FECHA, MONTH) = '2023-01-01'
17     GROUP BY partido_politico, mes, tema
18 ),
19 ranking_temas AS (
20     SELECT
21         partido_politico,
22         mes,
23         tema,
24         frecuencia,
25         RANK() OVER (
26             PARTITION BY partido_politico, mes
27             ORDER BY frecuencia DESC
28         ) AS rank_tema
29     FROM temas_por_mes
30 )
31 SELECT
32     partido_politico,
33     tema AS tema_principal,
34     frecuencia
35 FROM ranking_temas
36 WHERE rank_tema <= 3
37 ORDER BY partido_politico, rank_tema;

```

El análisis muestra que los temas más recurrentes están fuertemente ligados a figuras de autoridad política ('presidente/a') y términos legislativos ('constitucional', 'indicacion', 'republica'), lo que refleja el foco institucional del discurso partidario.

Destaca el Partido Frente Amplio con una fuerte presencia en temas de educación, mientras que Renovación Nacional y la UDI concentran menciones sobre presidencia y constitucionalidad. El Partido Demócrata Cristiano resalta con frecuencia en torno a 'provoste' y el tema 'autista', lo cual podría indicar una coyuntura específica.

Partidos más pequeños, como el Partido Humanista, muestran una dispersión temática amplia pero con baja frecuencia, lo que puede reflejar menor visibilidad o foco comunicacional. En tanto, el Partido Comunista repite intensamente conceptos como 'honorables' y 'autista', lo que también sugiere temas polémicos o estratégicos.

En suma, los datos reflejan diferencias claras en las prioridades discursivas de cada partido, con algunos concentrados en debates institucionales y otros en figuras o temas sociales específicos.