

Procesamiento de Datos Masivos

Tarea 1 - Resumen Ejecutivo

1. Objetivo del Proyecto

Analizar las intervenciones de los parlamentarios en el Congreso de Chile para identificar las principales temáticas discutidas, cómo han evolucionado en el tiempo y cuál ha sido el rol de los partidos políticos en dichas discusiones.

2. Metodología

- **Preprocesamiento de Datos:** Se trabajó con archivos ‘.csv’ de intervenciones parlamentarias y datos de parlamentarios. Limpieza de datos: corrección de errores tipográficos. Extracción automática de keywords desde la columna ‘TEXTO PRINCIPAL’, usando el modelo KeyBert.
- **Modelamiento de Datos:** Diseño de un modelo relacional normalizado con entidades: Parlamentarios, Partidos, Intervenciones y Keywords. Implementación del modelo en BigQuery.
- **Consultas Analíticas:** Identificación de las 5 temáticas más tratadas por mes. Media móvil de intervenciones por partido político (ventana de 3 meses). Temas principales por partido en cada trimestre. Top 3 de temáticas tratadas por partido en un mes específico.
- **Visualización y Análisis:** Se utilizaron herramientas de visualización (Python + Matplotlib) para presentar los hallazgos de forma clara e intuitiva.

3. Principales Hallazgos

A lo largo de los meses, los términos más discutidos fueron aquellos relacionados con la figura presidencial, seguramente sesgados por una mala extracción de las keywords, temas constitucionales, y reformas legislativas. Sin embargo, hubo un cambio hacia temas más locales como ”municipalidad” temas educativos hacia mediados de 2023 y 2024. Esto refleja la dinámica política y los cambios en la agenda pública, dependiendo de los eventos y prioridades del gobierno y el parlamento.

4. Conclusión

Este análisis permite entender mejor el comportamiento legislativo de los parlamentarios y sus partidos, ofreciendo transparencia sobre los intereses y prioridades del Congreso. Además, entrega una herramienta valiosa para la ciudadanía, periodistas y tomadores de decisiones.

5. Herramientas Utilizadas

Python (pandas, spaCy / nltk para NLP), BigQuery, SQL (con Window Functions), Herramientas de visualización