



Pontificia Universidad Católica de Chile
Escuela de Ingeniería
Instituto de Ingeniería Matemática y Computacional
IMT2200 – Introducción a la ciencia de datos

Avance de proyecto

Grupo 14

**Benjamin Ulloa Mura - Camila Rojas Abarca - Alejandro Orellana
Contreras - Esteban Vásquez Rivera**
Link del repositorio a usar.

1. Contexto y motivación.

La Toma de decisiones de manera informada y respaldada por datos, es esencial en las diferentes profesiones y oficios en el mundo actual. La capacidad de manejar una alta cantidad de información y de alto valor se ha vuelto indispensable, como también una meta, para garantizar que las decisiones tomadas sean las más adecuadas según el contexto en el que se toman. En este sentido, disponer de datos precisos, confiables y accesibles facilita a profesionales y trabajadores evaluar situaciones pasadas, identificar patrones y prever tendencias futuras, lo que a su vez mejora la eficacia y transparencia en la gestión de recursos y la respectiva toma de decisiones.

Con el objetivo principal de comprender el funcionamiento del Estado de Chile, nuestro proyecto se enfocará en la recopilación y análisis de datos relevantes sobre procesos de contratación pública. Durante nuestra investigación previa, descubrimos la página Open Contracting Partnership, una plataforma que ofrece una vasta cantidad de datos de diferentes países, enfocándose en la transparencia de los procesos de compra estatales. Específicamente, identificamos que esta página alberga información detallada de la Dirección de Compras y Contratación Pública de Chile, disponible en diversos formatos como JSON, Excel y CSV. Nos interesó particularmente comparar estos datos a lo largo de diferentes años para identificar tendencias y evaluar la eficiencia de las contrataciones públicas.

Por lo tanto, nuestra audiencia objetivo son aquellos individuos interesados en examinar las compras realizadas por el Estado pero que carecen de los conocimientos técnicos necesarios para analizar los datos disponibles. Al desarrollar herramientas y visualizaciones accesibles, buscamos facilitar el acceso y la comprensión de esta información, permitiendo que posibles funcionarios y otros interesados puedan tomar decisiones informadas basadas en datos históricos.

2. Objetivos.

Nuestro objetivo principal es permitir que potenciales usuarios puedan analizar y utilizar estos datos de manera accesible para formar opiniones informadas sobre las decisiones económicas del Estado en los años 2018 y 2021.

Adicionalmente, se busca hacer que los datos sean más legibles y accesibles, especialmente para aquellas personas que no tienen los conocimientos técnicos para trabajar directamente con grandes bases de datos. Esto fomentará un acceso más amigable a la información y permitirá una comprensión más profunda del uso de recursos estatales. Al facilitar este flujo de datos, esperamos contribuir al debate público informado sobre la eficiencia del gasto y promover mayor transparencia en la gestión pública.

Finalmente, este proyecto tiene el objetivo de proporcionar herramientas intuitivas para analizar los datos, como lo pueden ser visualizaciones y resúmenes, de manera que usuarios interesados puedan obtener una perspectiva clara y comprensible de las decisiones económicas del Estado chileno.

3. Datos.

Los set de datos a utilizar son un conjunto de documentos CSV presentes en la página mencionada en su momento al inicio de este informe, que contiene datos acerca de las licitaciones y ordenes de comprar realizadas en los años 2018 y 2021, estos documentos son:

- *awards.csv*
- *awards_documents.csv*
- *awards_items.csv*
- *awards_supplier.csv*
- *main.csv*
- *parties.csv*
- *tender_items.csv*
- *parties_additionalIdentifiers.csv*

Cabe mencionar que los datos descargados del 2021 originalmente vienen con los archivos adicionales

- *planning_budget_budget_Breakdown.csv*
- *planning_items.csv*

pero solo se utilizaran los archivos que existen en ambos set de datos. A continuación, se describirán las columnas de cada archivo mencionado con su respectivo significado.

main.csv

1. **initiationType**: es el método por el cual se inicia el proceso de contratación, puede ser *tender* (licitación).
2. **buyer_name**: Nombre del comprador.
3. **tender_tenderPeriod_startDate**: Cuándo inició el periodo en el que los potenciales ofertantes podrán hacer preguntas y solicitudes de aclaración.
4. **tender_tenderPeriod_durationInDays**: Duración del *tender* en días (Float).
5. **tender_enquiryPeriod_endDate**: Fecha de finalización de las consultas de la licitación.
6. **tender_enquiryPeriod_startDate**: Fecha de inicio de las consultas de la licitación.
7. **tender_enquiryPeriod_durationInDays**: El periodo de las consultas de la licitación en días.
8. **tender_value_amount**: El valor total estimado superior de la adquisición. Un valor negativo indica que el proceso de contratación puede implicar pagos del proveedor al comprador.
9. **tender_value_currency**: El tipo de moneda con el que se adquirió.
10. **tender_value_unitOfAccount**: Indica la unidad de la cuenta, por ejemplo: UTM.

awards.csv

1. **_link_main**: Identificador que puede hacer referencia a un conjunto más amplio de datos asociados con el *id*.
2. **title**: El título o nombre corto de la licitación, que describe de manera resumida el proyecto o servicio contratado.
3. **status**: El estado de la licitación o del premio, que puede indicar si la licitación fue exitosa (*active*) o fallida (*unsuccessful*).
4. **value_amount**: El monto asociado con el premio o contrato, que indica el valor monetario en la unidad de cuenta indicada.
5. **value_currency**: La moneda en la que se expresa el valor del contrato.
6. **description**: Una descripción más detallada de la licitación o el servicio contratado. Incluye información específica sobre la naturaleza del contrato o la tarea realizada.

awards_items.csv

1. **description**: Una descripción detallada del artículo o servicio dentro del contexto de la adjudicación o contrato.
2. **quantity**: La cantidad del artículo o servicio adjudicado, es decir, cuántos de ese bien o servicio se están comprando o contratando.
3. **unit_name**: El nombre de la unidad de medida en la que se expresa la cantidad (por ejemplo, unidades, kilogramos, etc.).
4. **unit_value_amount**: El valor unitario del artículo o servicio, que indica el precio de una unidad de ese bien o servicio.
5. **unit_value_currency**: La moneda en la que se expresa el valor unitario (por ejemplo, CLP, USD).
6. **classification_id**: Un identificador único para clasificar el artículo o servicio adjudicado dentro de algún esquema de clasificación.
7. **classification_scheme**: El esquema o sistema de clasificación utilizado para identificar el artículo o servicio.
8. **_link_main**: Un enlace o identificador principal que conecta esta información con otros datos o sistemas.

awards_suppliers.csv

1. **_link_main**: Identificador principal que conecta información con otros datos.
2. **name**: Nombre de la persona relacionada a la licitación.

tenders_items.csv

1. **_link_main**: Identificador principal que conecta información con otros datos.
2. **description_items**: Descripción de lo que es la consulta/licitación.
3. **quantity**: Cantidad de *items*.
4. **unit_name**: Nombre general de lo que es el *item*.
5. **classification_id**: El id de la licitación.
6. **classification_uri**: El URI de este paquete que lo identifica de forma única en el mundo.
7. **classification_scheme**: El esquema o lista de códigos del que se toma el código de clasificación.

parties.csv y parties_additionalIdentificadores

1. **_link_main_x:**
2. **id:** La identificación utilizada para hacer referencias cruzadas a esta parte desde otras secciones del comunicado.
3. **name:** Un nombre común para esta organización u otro participante en el proceso de contratación.
4. **roles:** El(los) rol(es) de la parte en el proceso de contratación, utilizando la lista de códigos abierta *partyRole*.
5. **address_region:** La región.
6. **identifier_id:** El identificador de la organización en el esquema seleccionado.
7. **identifier_scheme:** El esquema o registro del que se toma el identificador.
8. **identifier_legalName:** El nombre legalmente registrado de la organización.
9. **address_countryName:** El nombre del país.
10. **legalName:** El nombre legalmente registrado de la organización.

4. Preguntas de investigacion.

Teniendo ya en consideracion los datos a utilizar en conjunto a los objetivos que se buscan en este proyecto, es posible abordar preguntas clave relacionadas con las licitaciones públicas. Se investigarán las relaciones entre variables como los montos adjudicados y los tipos de procedimientos de contratación, y se explorarán factores que influyen en la duración de las adjudicaciones. También se buscará inferir qué tipo de empresa tiene más éxito en ciertos sectores ademas de tratar de identificar patrones entre categorías de licitaciones. Dicho esto último, planteamos las siguientes 5 preguntas de manera directa con tal de ser una actual guía de como continuar este proyecto a futuro.

- ¿Existe una correlación entre el monto adjudicado y el tipo de procedimiento de contratación utilizado?
- ¿Qué factores influyen en la duración de los procesos de adjudicación de licitaciones?
- ¿Es posible predecir qué tipo de empresa (por tamaño) gana más contratos en ciertos sectores?
- ¿Qué categorías de licitaciones muestran similitudes en cuanto a montos y tiempos de adjudicación?
- ¿Cómo se distribuyen geográficamente las contrataciones públicas según el tipo de proveedor?

5. Diseño tentativo.

El diseño tentativo del análisis abarcará diversos métodos computacionales y estadísticos, adaptados a las preguntas de investigación planteadas. En primer lugar, se realizará una limpieza de datos exhaustiva para garantizar la calidad y consistencia de la información. Se categorizarán los datos, permitiendo el clasificar los tipos de ítems en función de su servicio, como social, salud, educación y seguridad.

En paralelo, se crearán dataframes agrupados por mes para analizar promedios de duración de los procesos de licitación. Esto facilitará la identificación de patrones y tendencias en la duración de los procesos según los meses en que se inician. También se explorarán correlaciones entre variables, como la duración de los procesos en relación con el mes y el tipo de licitación en términos de rango monetario.

La visualización de datos será una parte esencial del análisis. Se emplearán gráficos para mostrar la frecuencia de concesiones por mes y los porcentajes de éxito de las licitaciones clasificadas por categoría. Esto permitirá a los usuarios comprender mejor las dinámicas de las contrataciones públicas en Chile, facilitando una interpretación más accesible de los datos y contribuyendo a una toma de decisiones informada. En resumen, se combinarán técnicas de limpieza, categorización y visualización para proporcionar un análisis integral de las licitaciones públicas.