

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INFORMÁTICA
SAN JOAQUÍN - CHILE



“PROPUESTA DE UNA VISUALIZACIÓN DE DATOS
EFECTIVA, PARA LOS DATOS DE MERCADO PÚBLICO
DURANTE EL PERÍODO DE PANDEMIA”

FRANCISCO ALEJANDRO ABARCA MORAGA

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL EN INFORMÁTICA

Profesor Guía: Jose Luis Martí
Profesor Correferente: Mauricio Olivares Faúndez

Diciembre - 2022

AGRADECIMIENTOS

Quiero expresar mi más sincero agradecimiento a Jose Luis Martí por su valiosa orientación y apoyo durante todo el proceso de investigación y redacción de esta tesis. Su dedicación y paciencia han sido fundamentales para la realización de este trabajo.

También quiero agradecer a los docentes que impartieron la malla curricular que tuve el agrado de cursar, por sus útiles enseñanzas y vivencias. Su aporte ha sido invaluable para la mejora y enriquecimiento de esta tesis.

Por último, pero no menos importante, quiero agradecer a mi familia y amigos por su constante apoyo y comprensión durante este proceso. Gracias por estar siempre a mi lado y por su ánimo en los momentos más difíciles. Este trabajo es también una muestra de mi gratitud hacia todos ustedes.

RESUMEN

Resumen— En este documento se expondrá como se ha realizado un diseño y desarrollo de una propuesta de paneles informativos respecto al estado de las licitaciones existentes hasta octubre de 2021. Lo realizado se segmenta en tres partes principales, la obtención de los datos, la creación de los paneles en cuestión y la validación de lo diseñado contra usuarios. Los resultados obtenidos corroboraron que el diseño presentado en un inicio era efectivo para el propósito presentado. Estos resultados dan hincapié a nuevas ramas para el conjunto de datos trabajado, en donde el trabajo se puede ampliar desde la arquitectura de la solución o hacia la difusión de otros tipos de datos públicos usando el mismo marco de trabajo.

Palabras Clave— Visualización de datos; Paneles de información; Procesamiento de datos; Licitaciones públicas

ABSTRACT

Abstract— In this document we will explain how we have designed and developed a proposal for information panels on the status of existing tenders up to October 2021. The work carried out is divided into three main parts: obtaining the data, creating the panels in question and validating the design against users. The results obtained corroborated that the design initially presented was effective for the purpose presented. These results emphasise new branches for the worked dataset, where the work can be extended from the architecture of the solution or towards the dissemination of other types of public data using the same framework.

Keywords— Data visualisation; Information panels; Data processing; Public tenders; Data visualisation

GLOSARIO

API: Application Programming Interface (Interfaz de Programación de Aplicaciones)

ETL: Extract, Transform & Load (Extraer, Transformar y Cargar)

JSON: JavaScript Object Notation (notación de objeto de JavaScript)

OCDS: Open Contracting Data Standard (Estándar de Datos de Contrataciones Abiertas)

ÍNDICE DE CONTENIDOS

RESUMEN	III
ABSTRACT	III
GLOSARIO	IV
ÍNDICE DE FIGURAS	VII
ÍNDICE DE TABLAS	VIII
INTRODUCCIÓN	1
CAPÍTULO 1: DEFINICIÓN DEL PROBLEMA	2
1.1 Contexto y situación Actual	2
1.2 Información sobre el problema	2
1.3 Objetivos	4
1.3.1 Objetivo General	4
1.3.2 Objetivos Específicos	4
CAPÍTULO 2: MARCO CONCEPTUAL	5
2.1 Visualización de datos	5
2.1.1 Campos utilizados en la visualización de datos	6
2.2 Estándares en visualización de datos	6
2.2.1 Tipos de datos	8
2.2.2 Buenas prácticas	8
2.2.3 Proceso de visualización	9
2.2.4 Métodos de visualización	10
2.3 Herramientas de visualización de datos	13
2.3.1 Herramientas sin utilización de lenguajes de programación	14
2.3.2 Herramientas que utilizan lenguajes de programación	15
2.4 Interacción de usuarios e información	16
2.5 Teoría de la información y canales de comunicación	16
2.6 Modelos para la validación de visualizaciones	17
2.7 Estudios con los datos del problema	21
2.8 Metodologías y flujos de trabajo	21
2.8.1 Kanban	22
2.8.2 Scrum	23
CAPÍTULO 3: PROPUESTA DE SOLUCIÓN	26
3.1 Metodología de trabajo	26
3.2 Etapa 1: Evaluación del conjunto de datos	27

3.3 Etapa 2: Herramientas y tecnologías	29
3.3.1 Comparativa de las herramientas a utilizar	31
3.4 Etapa 3: Extracción, transformación y carga de datos	34
3.5 Etapa 4: Visualizaciones de datos	37
CAPÍTULO 4: VALIDACIÓN DE LA SOLUCIÓN	49
4.1 Métodos de validación	49
4.2 Criterios utilizados	50
4.3 Resultados obtenidos	54
5 CONCLUSIONES	58
REFERENCIAS BIBLIOGRÁFICAS	62

ÍNDICE DE FIGURAS

1	Distintos tipos de visualización de datos	5
2	Interacción en el sistema Gnaeus	7
3	Gráfico de torta multivariado	11
4	Descripción para un gráfico de vela	11
5	Gráfico de líneas informando la distribución de la industria musical	12
6	Gráfico de burbujas correlacionando la ingesta de azúcar y grasas	13
7	Gráfico de coordenadas paralelas comparando atributos entre vehículos	13
8	Treemap sobre las exportaciones de Benin durante 2009	14
9	Sistema de comunicación	17
10	Relación por pares entre los tipos de visualización en las tareas y las métricas de rendimiento. Las flechas muestran que la fuente es significativamente mejor que el objetivo	18
11	Especificación del metamodelo de visualización	19
12	Especificación del contexto de visualización	19
13	Tipos de metodologías organizadas para mostrar las relaciones con la precisión, la generalización y el realismo	20
14	Instanciación y uso de kanban	22
15	Actividades en el marco de trabajo SCRUM	24
16	Instancia en el marco de trabajo SCRUM	24
17	Proceso de una licitación.	28
18	Tablero kanban utilizado.	34
19	Primera versión del proceso ETL.	35
20	Segunda versión del proceso ETL.	37
21	Conjunto de visualizaciones respecto a la ubicación de las licitaciones.	39

22	Panel de ayuda para el conjunto de visualizaciones regional.	41
23	Conjunto de visualizaciones respecto al ámbito monetario.	42
24	Desglose de un mes por región en acción.	43
25	Conjunto de visualizaciones respecto al tipo de licitación.	44
26	Conjunto de visualizaciones respecto al estado de las licitaciones.	45
27	Explorador de datos.	47
28	Menú principal.	48
29	Datos tabulados obtenidos del estudio de campo.	55
30	Datos sobre precisión en el estudio de campo.	56
31	Datos sobre tiempo en el estudio de campo.	57

ÍNDICE DE TABLAS

1	Medidas de efectividad cuantitativas y cualitativas	3
2	Ejemplo de una tabla	10

INTRODUCCIÓN

El trabajo realizado en este documento cae dentro del campo de estudio de la visualización de datos, cuyo fin es la representación de datos en una forma gráfica la cual condensa los datos subyacentes para aumentar la eficiencia de los usuarios, haciendo mejor uso de sus capacidades cognitivas y de percepción. Este campo de estudio será empleado para abordar el tema de las licitaciones públicas sobre los datos expuestos por mercado público hasta el mes de octubre del año 2021. El tema se presenta con mayor interés con posterioridad a la pandemia que afectó al país, y los casos de abuso en ciertas localidades por las posibles situaciones de fraude y precios excesivos por insumos necesarios para copar con la pandemia en curso.

La solución propuesta incluye tanto el tratamiento de datos (extracción, transformación y carga), el diseño de los paneles informativos, la validación de los paneles propuestos con usuarios, la documentación y exposición a canales públicos de lo realizado para fomentar la colaboración del trabajo realizado. Dentro de la metodología se hace uso intensivo de visualizaciones basadas en tareas, método que hace referencia a gráficas con finalidades específicas en mente en su diseño, con literatura reciente que incluye marcos de trabajo al respecto.

Este documento está estructurado en cinco capítulos. El primero define el problema, tanto su contexto como los objetivos que se buscan satisfacer. Con el problema explicado, se dará paso a exponer la literatura actual en un marco conceptual, que define y explica distintas metodologías junto con las referencias necesarias, todo esto en el capítulo 2.

En el tercer capítulo se propone la solución, con la metodología a utilizar, las herramientas, los problemas encontrados y las resoluciones utilizadas para ellos, en conjunto con los paneles de información generados. En el siguiente capítulo se informa del método empleado para validar la solución recién propuesta, así como los resultados obtenidos y las conclusiones de los datos obtenidos sobre la validación realizada.

Finalmente se presentan las conclusiones sobre el trabajo realizado, los alcances y lo que no fue posible realizar dadas distintas limitantes en los propuesto inicialmente. También se identifican los posibles rumbos a tomar desde el trabajo realizado con los temas anexos que se abren del trabajo ejecutado.

CAPÍTULO 1

DEFINICIÓN DEL PROBLEMA

En este capítulo se contextualiza el problema a tratar, sus orígenes y ramificaciones. Además se presentan objetivos los cuales tratan a dar una solución a la cuestión planteada.

1.1. Contexto y situación Actual

Posterior al estallido social ocurrido en nuestro país debido a situaciones políticas, sociales y económicas [Güell, 2019, Charalamby y Andrea, 2020], se produjo un brote epidemiológico no evidenciado desde fines de los años 60 con el virus H2N2 [Henderson *et al.*, 2009].

Este panorama ocasionó que se decretara el estado de emergencia a lo largo del país¹, creando una secuencia de eventos en donde fue necesario utilizar el método de compra directa por sobre las licitaciones, según lo estipulado en la ley 19.886² que regula la contratación de bienes y servicios de más de 800 instituciones administradas por el estado, para compras que sean vistas como necesarias en casos de emergencia o catástrofe, donde estas pueden ser resueltas mediante el método de trato directo.

El uso de este método era para agilizar la obtención de los bienes necesarios para combatir la epidemia en curso, dado que en esta situación era imperante obtener los implementos para la ola de pacientes que eventualmente llegaría a los centros asistenciales. El problema resultante fue el uso indebido de este método para el beneficio económico de terceros, donde el uso de trato directo no tiene la transparencia ni la rigidez de cumplir con una pauta de requisitos, como sí lo tiene una licitación pública de manera normal. Durante el periodo de pandemia se encontraron diversos casos los cuales evidencian³ el nepotismo en la adquisición de trato directo, además de aprovechar el estado caótico del acontecer nacional para el beneficio personal a costa de las arcas fiscales⁴.

1.2. Información sobre el problema

Los datos de las transacciones, sean mediante licitaciones o no, están disponibles para el público general de manera transparente y abierta⁵. y las formas para obtener los datos puede

¹BBC Mundo [Último acceso: 28 Junio 2021]

²El artículo 7 define el trato directo y el artículo 8 define las condiciones o situaciones en las cuales ameritan utilizar el método de trato directo o licitación privada

³Diario Concepción [Último acceso: 28 Junio 2021]

⁴Ciper Chile [Último acceso: 28 Junio 2021]

⁵ChileCompra [Último acceso: 25 Julio 2021]

ser como descarga masiva o mediante una API.

El problema radica en que dichos datos no generan información valiosa para el público general, solo atrae a usuarios especializados los cuales los usan en alguna manera significativa como tareas relacionadas a periodismo de datos o estudios que necesiten los datos como base para generar información relevante. Si a esto se suman los casos fraudulentos del uso de trato directo para compras de insumos, se encuentra con una situación que amerita ser informada, dada la importancia del tema y la posible relevancia en la opinión pública nacional.

En cuanto a las soluciones propuestas o existentes, la más cercana a brindar algo relativamente similar a lo necesario es lo realizado por el Observatorio Fiscal ⁶. Pero al analizar investigaciones que evalúan la efectividad de las visuales [Zhu, 2007], se identifica que estas no son realmente eficaces al utilizar los criterios cuantitativos y cualitativos de evaluación respecto al impacto de una visualización de datos (Ver tabla 1). Además que lo propuesto por el Observatorio Fiscal es una mirada más macro, lo cual puede reducir la funcionalidad de examinación y observación de información para casos puntuales de usuarios específicos.

Otro problema es la escasa documentación respecto al uso del conjunto de datos, si bien existen algunos casos aislados que utilizan los datos e indican el proceso utilizado [Galdames Paredes, 2015]. Esto es un problema que dificulta el proceso investigativo, por lo tanto sería valioso obtener un proceso que brinde un conjunto de datos limpio y pre-procesado con el cual trabajar.

	Medidas cuantitativas	Medidas cualitativas
Precisión	<ul style="list-style-type: none"> • Medir el número de errores interpretativos 	<ul style="list-style-type: none"> • Examinación • Observación • Comparación novato-experta
Utilidad	<ul style="list-style-type: none"> • Medir el número de objetivos logrados • Registrar el número de veces que un diseño de visualizaciones es escogido por usuarios al resolver una tarea 	<ul style="list-style-type: none"> • Examinación • Observación • Comparación novato-experta
Eficiencia	<ul style="list-style-type: none"> • Tiempo para completar una tarea • Movimiento ocular • Medir la curva de aprendizaje 	<ul style="list-style-type: none"> • Examinación • Observación • Comparación novato-experta • Análisis de visualización complejo

Tabla 1: Medidas de efectividad cuantitativas y cualitativas
Fuente: Ying Zhu, Measuring Effective Data Visualization.

En general, el problema radica en que la fuente de datos no es fácilmente disponible al pú-

⁶Observatorio Fiscal [Último acceso: 26 de Julio 2021]

blico general y esto, sumado a la estructura de los datos, los hace difícil de manipular. Para brindar una solución, se debe desarrollar algún tipo de acceso a los datos e informar el cómo están estructurados. Además, proporcionar las herramientas a los usuarios para que sean capaces de realizar análisis propios.

1.3. Objetivos

1.3.1. Objetivo General

Crear un conjunto de visualizaciones de datos las cuales permitan informar, de manera efectiva y simple, a la población sobre su estado, mediante técnicas de representación y síntesis gráfica de información.

1.3.2. Objetivos Específicos

- Crear visualizaciones que brinden información respecto a las licitaciones del periodo, para crear opinión en la población.
- Evaluar las visualizaciones creadas para medir cuantitativamente su efectividad con los usuarios finales.
- Crear procesos y documentación respecto al uso de los datos, para trabajos futuros que puedan surgir de este.

CAPÍTULO 2

MARCO CONCEPTUAL

En este capítulo se detalla la literatura en las disciplinas que participan en el desarrollo de la solución. Ya sean métodos de trabajo, experiencias con el problema o definiciones necesarias para el trabajo posterior.

2.1. Visualización de datos

La visualización de datos es una disciplina muy abierta en cuanto a perspectiva, la cual puede variar desde representación de redes complejas hasta gráficos cuantitativos, lo cual se puede evidenciar en la figura 1. Una definición para este concepto será el siguiente que se basa en tres criterios [Kosara, 2007]:

1. Está basado en datos cuantitativos y cualitativos
2. Resulta en una imagen que es representativa de los datos en el punto anterior
3. Es legible por los consumidores y soporta exploración, examinación y comunicación de los datos provistos

Cabe destacar que el criterio de los datos (cuantitativo y cualitativo) no está definido de manera muy específica, esto, debido a la inmensa cantidad de atributos y características que pueden tener los datos, similar a lo visto en la figura 1.

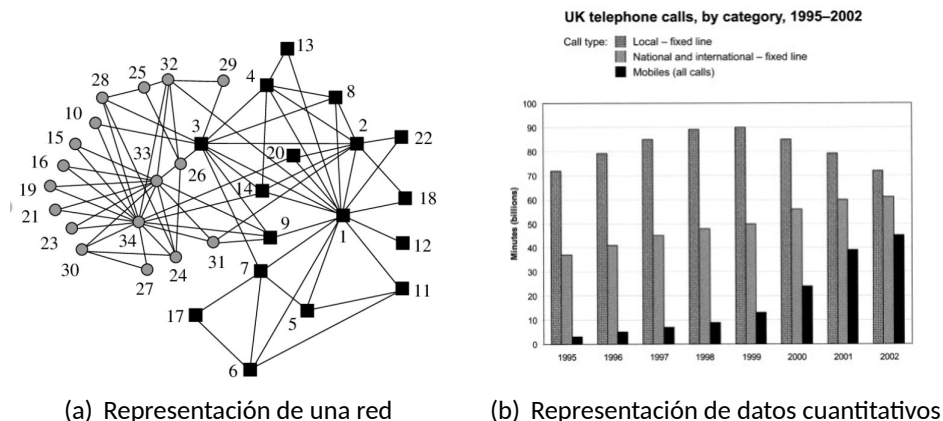


Figura 1: Distintos tipos de visualización de datos

El segundo criterio es un recordatorio para mantener la representatividad de los datos trabajados, esto quiere decir no omitir información crucial, o aún más simple, que la visualización

resultante refleje de una manera fiel la información de los datos utilizados. Es importante este punto al desarrollar la solución, dado que su cumplimiento verificará que no existe una manipulación errónea de los datos y así evitar malentendidos con el mensaje generado por el producto final.

El tercer criterio puede ser utilizado como prueba si el resultado final fue exitoso, dado que si el producto entregado permite exploración de los datos presentados en conjunto de una examinación y comunicación con ellos, la visualización será capaz de mantener al usuario interesado y posiblemente descubra un nuevo conocimiento desde el uso de la información entregada.

2.1.1. Campos utilizados en la visualización de datos

La visualización de datos es una disciplina que utiliza el conocimiento de distintos entornos [Aparicio y Costa, 2015]. Ya sea el uso de principios, conceptos, técnicas y teorías, estos son acoplados con el solo objetivo de transformar los datos en información (y posible nuevo conocimiento). Algunas de las áreas relevantes son las siguientes:

- Programación: Procesos de extracción de datos y transformación de los mismos
- Estadística: Disciplina que busca mediante el análisis numérico obtener explicaciones de fenómenos o poblaciones
- Psicología: El estudio de la percepción de los datos y el impacto que algunos elementos tienen sobre ellos
- Diseño gráfico: Crítico en la creación y diseño de infografías, lo cual es materializado en los resultados finales

2.2. Estándares en visualización de datos

Compartir conocimiento previo ha sido identificado como una de las 10 problemáticas no resueltas en el campo de las visualizaciones [Chen, 2005]. El resolver este problema asegura que los usuarios compartan un piso común con el cual pueden intercambiar experiencias pasadas, esto es consistente con el diseño centrado en usuarios en las interacciones de personas y sistemas informáticos.

Como este problema sigue sin resolver no existe una pauta rígida la cual seguir, pero se han hecho avances en el campo que intentan consolidar un conjunto de teorías, definiciones y prácticas que giran en torno a las disciplinas que utilizan las visualizaciones como un medio para entregar información y conocimiento de los datos. Para adentrar en lo que se ha

intentado llegar a un consenso dentro de la disciplina, debe tener en cuenta las siguientes definiciones [Chen *et al.*, 2008]:

- Datos: Una colección de símbolos
- Información: Datos los cuales son procesados para ser útiles, proveyendo respuesta a preguntas del tipo “¿quién?”, “¿qué?”, “¿dónde?” y “¿cuándo?”
- Conocimiento: Es la aplicación de datos e información, proveyendo respuesta a preguntas del tipo “¿cómo?”

Dentro de la definición de conocimiento se distinguen dos tipos [Nonaka, 2008], **conocimiento explícito** y **conocimiento tácito**. El primero corresponde al que puede ser procesado por un sistema computarizado y el segundo es personal, especializado y solo puede ser extraído por usuarios humanos. Además se han identificado cuatro procesos (internalización, externalización, colaboración y combinación) [Wang *et al.*, 2009], los cuales convierten conocimiento del primer tipo al segundo, generando sistemas de visualización asistidos por conocimiento. Este tipo de sistemas tiene como requisito dos tipos de conocimientos por parte del usuario, conocimiento operacional (el como interactuar con el sistema de visualización) y conocimiento del dominio (como interpretar el contenido visualizado). Si bien un sistema de visualización enfocado en su usabilidad y percepción puede aliviar la necesidad del conocimiento operacional por parte del usuario, el conocimiento del dominio no puede ser reemplazado fácilmente.

Un ejemplo para un sistema de visualización asistido por conocimiento es Gnaeus [Federico *et al.*, 2016], utiliza datos de pacientes para mejorar el trato médico al proveer una comparación de opciones en los tratamientos. Condensa el conocimiento médico en la visualización para brindar al usuario conocimiento del dominio. Una síntesis del modelo utilizado se puede apreciar en la figura 2.

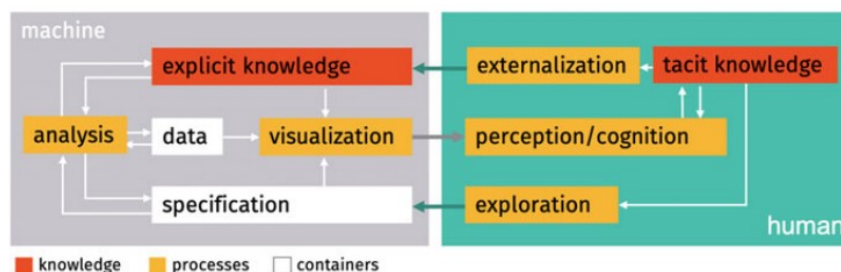


Figura 2: Interacción en el sistema Gnaeus
Fuente: [Federico *et al.*, 2016].

2.2.1. Tipos de datos

El objetivo de la visualización de datos es la transformación de los datos en un formato visual perceptualmente eficiente. Con esta generalización, se generan indicaciones sobre los tipos de datos disponibles para visualizar.

Se ha sugerido que existen dos formas de datos [Ware, 2019], valores y estructuras. Una idea similar es el usar la distinción entre entidades y relaciones. La diferencia es que las entidades son los objetos que se desean visualizar, y las relaciones definen patrones que vinculan a entidades entre sí.

Como resumen se denotan los siguientes tipos de datos:

- Entidades: Estos describen objetos de interés
- Relaciones: Estas forman estructuras que relacionan a distintas entidades. Estas pueden ser tanto físicas como conceptuales.
- Atributos: Características particulares de tanto relaciones como entidades. Estas describen propiedades las cuales no pueden ser generalizadas en una entidad particular.
- Dimensiones de datos: Un atributo puede tener múltiples dimensiones. Esto representa la cantidad necesaria de datos necesarios para describir el atributo referido.

Una buena forma de definir la calidad de los datos es la taxonomía de las escalas utilizadas para definir los valores en los atributos. Un método útil es el definido para modelos estadísticos [Stevens, 1946], que utiliza cuatro niveles de medida: nominal, ordinal, intervalos y magnitud (escalas de razón). Solo tres de estos niveles son utilizados ampliamente, y de una forma levemente distinta dada la influencia de las disciplinas involucradas. Estas clases de datos son las siguientes:

- Datos categóricos: Similar a la clase nominal
- Datos enteros: Similar a la clase ordinal (discretos y ordenados)
- Datos reales (flotantes): Combina las propiedades de las escalas de intervalos y magnitud

2.2.2. Buenas prácticas

La visualización de datos, sobre todo de aquellos multivariados, presentan múltiples desafíos para encontrar una buena representación visual de un problema determinado, lo cual puede ser complejo y no determinista. Lo más inmediato es cómo codificar los atributos en una

imagen cuando se necesita representar más de una dimensión. Para combatir estas dificultades se indica un proceso definido que sirve como esquema para construir visualizaciones de una manera sistemática, en conjunto con distintas técnicas para representar distintos tipos de datos en situaciones particulares.

2.2.3. Proceso de visualización

Dentro de la literatura existen acercamientos de cómo diseñar visualizaciones mediante un proceso bien definido [Chittaro, 2006]. Este proceso contiene seis pasos, los cuales son los siguientes:

1. **Mapeo:** Corresponde a cómo codificar información a una forma visual, mediante características visuales. Un buen desarrollo de este paso produce visuales precisas, cuando existen relaciones específicas entre la información y la descripción visual.
2. **Selección:** Este paso corresponde a seleccionar los datos dentro de los disponibles acorde a la tarea a realizar. Se debe evitar el seleccionar datos innecesarios, puesto que estos pueden llevar a nociones erróneas sobre los datos y el usuario podría tomar malas decisiones por ello.
3. **Presentación:** Este paso refiere a una perspectiva de cómo la información está organizada en el espacio disponible para ser representada.
4. **Interactividad:** Este paso hace referencia a todas las facilidades para organizar, explorar y ordenar dentro de la visualización. La facilidad de uso de estas características mejora drásticamente la capacidad del usuario para explorar, entender e interpretar los datos o información.
5. **Factor humano:** Este paso incluye dos grandes categorías, usabilidad y accesibilidad. Las categorías indican la facilidad de uso para el usuario final y la mejora del uso para usuarios con necesidades especiales. Conocimiento mediante percepción visual y aspectos cognitivos hacen más fácil el diseño de una visualización efectiva [Wang *et al.*, 2009]. Estos factores son práctica común en la interacción de personas y sistemas informáticos.
6. **Evaluación:** Luego de la creación de una interfaz de visualización, el último paso corresponde a la evaluación de la efectividad de lo creado y si el objetivo propuesto fue cumplido. Existen variados desafíos en la evaluación dado el factor humano definido anteriormente [Plaisant, 2004].

Nombre	País	Profesión	Nombre de usuario	Posts
Pedro	Chile	Profesor	pedro_123	15
Juan	Argentina	Contador	juan_arg	21
Diego	Brasil	Abogado	abo_diego	11

Tabla 2: Ejemplo de una tabla

Fuente: Elaboración propia

2.2.4. Métodos de visualización

Al representar los datos en una imagen, existen múltiples técnicas que permiten la representación de los mismos datos que explotan ciertas cualidades del conjunto analizado. En general estas técnicas tienen características genéricas que son fáciles de explotar al tener noción de ellas. Estas representaciones deben ser elegantes, descriptivas e interpretables, en orden para convenir el mensaje al lector de manera efectiva [Magazine, 2007]. Existen múltiples métodos nuevos con sus propios problemas; a continuación se descubren las técnicas con mayor cobertura en la literatura.

Tablas

Las tablas son simples y fáciles de comprender como representación de datos. Tienen un formato estructurado organizado por filas y columnas que brinda relaciones entre las dimensiones. De la forma más estándar, las filas representan registros con un conjunto de valores y las columnas representan variables. Estas entradas no necesariamente están ordenadas. Este método representa los datos de una forma explícita y su versatilidad hace que se use de manera interdisciplinaria. Un ejemplo se puede apreciar en la tabla 2.

Gráficos de torta

Un gráfico de torta es dividido en una cantidad de sectores determinada según la cantidad de entidades a graficar. Estas divisiones son proporcionales a un atributo en común entre las entidades. Cada división es representada normalmente mediante colores e identificadas con una leyenda identificando cada color. La mayor parte del tiempo las divisiones están con su magnitud representada por un porcentaje. Además estos gráficos pueden ser multivariados, expandiendo el nivel de detalle al subdividir las divisiones iniciales, para corresponder a datos relacionados jerárquicamente [Jeff, 2006]. Un ejemplo de gráfico de torta se puede apreciar en la figura 3.

Gráficos de barras

Una de las técnicas para visualizar datos más utilizada en la disciplina, mayormente con datos discretos por sobre continuos. El largo de las barras tiende a representar valores que son usados para comparaciones entre distintos valores discretos. Estas barras pueden estar alineadas tanto horizontal como verticalmente y flotantes (no empezando desde el eje vertical). Una variante especializada de este gráfico corresponde a los gráficos de vela [Iversen y Gergen, 2012], utilizados en el ámbito bursátil para observar la variación del valor de los activos durante un tiempo determinado, en donde cada barra indica un periodo par-

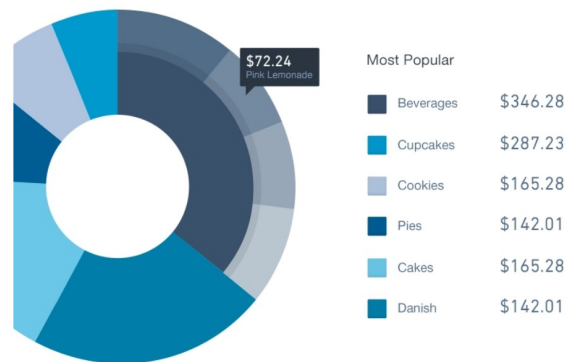


Figura 3: Gráfico de torta multivariado
Fuente: dribbble.com

ticular (normalmente el día completo) y la longitud de la barra señala los valores (alto, bajo, apertura y cierre) que obtuvo el activo en el periodo. Un ejemplo se observa en la figura 4.

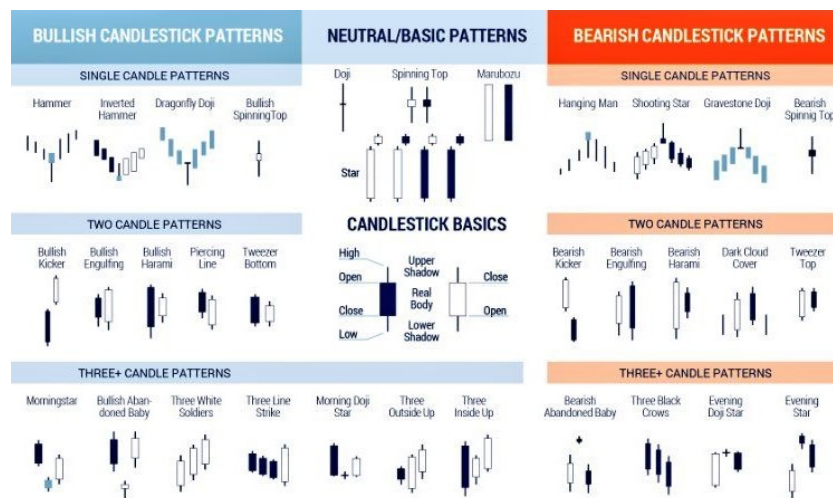


Figura 4: Descripción para un gráfico de vela
Fuente: Forex

Gráficos de líneas

Este tipo de técnica es conocido en distintas disciplinas, en el cual la información es mostrada al conectar puntos y formar líneas. Los puntos son conectados mediante líneas rectas o continuas. Este tipo de gráfico es utilizado en el contexto de mostrar la variación de un atributo durante intervalos de tiempo, ya sea para encontrar tendencias en los datos o ilustrar comportamiento a medida que transcurre el tiempo. Un ejemplo se puede evidenciar en la figura 5.

Gráficos de dispersión

Corresponde a una vista de un conjunto de datos sobre un plano de coordenadas cartesianas, esto para demostrar las relaciones de ambas variables. Esta representación de los datos

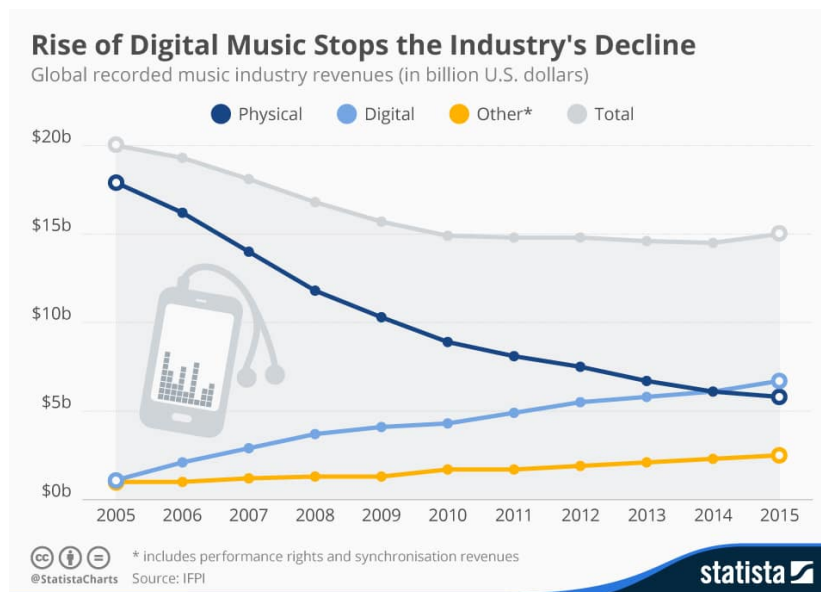


Figura 5: Gráfico de líneas informando la distribución de la industria musical
Fuente: Statista

sirve para corroborar qué tan fuerte son las relaciones entre los datos según las variables dispuestas [Ho, 2012] y si existen datos fuera del rango común dentro de las variables; es útil para identificar tendencias entre variables y *outliers*⁷. Como un subconjunto dentro de esta técnica se encuentran los gráficos de burbujas; es lo mismo que una dispersión pero se añaden una o dos dimensiones adicionales, las cuales corresponden a los colores y tamaños de los puntos en el plano. Un ejemplo de un gráfico de burbujas se puede apreciar en la figura 6

Coordenadas paralelas

Esta es una técnica utilizada para graficar datos individuales a lo largo de múltiples dimensiones. Cada dimensión está relacionada a un eje vertical y cada objeto de datos es mostrado como una conexión de puntos conectados a cada eje vertical [Gemignani, 2010]. Es relativamente útil al intentar visualizar datos con múltiples dimensiones, donde las técnicas nombradas anteriormente fallan. Se menciona en la literatura como las coordenadas paralelas ayudan a reconocer patrones en conjuntos de datos de un tamaño considerable. Un ejemplo de coordenadas paralelas se encuentra en la figura 7.

Treemap

Esta es una técnica utilizada para mostrar datos jerárquicos en la forma de capas de rectángulos anidados [Shneiderman y Plaisant, 1998]. Facilita al usuario el comparar nodos y subnodos en distintas profundidades de la jerarquía, ayudando a reconocer patrones y resultados. Además se habilita el uso de atributos adicionales en formas de color y tamaño de las capas de rectángulos. Un ejemplo de esta técnica se puede observar en la figura 8.

⁷Datos los cuales se comportan distinto a los demás miembros del conjunto

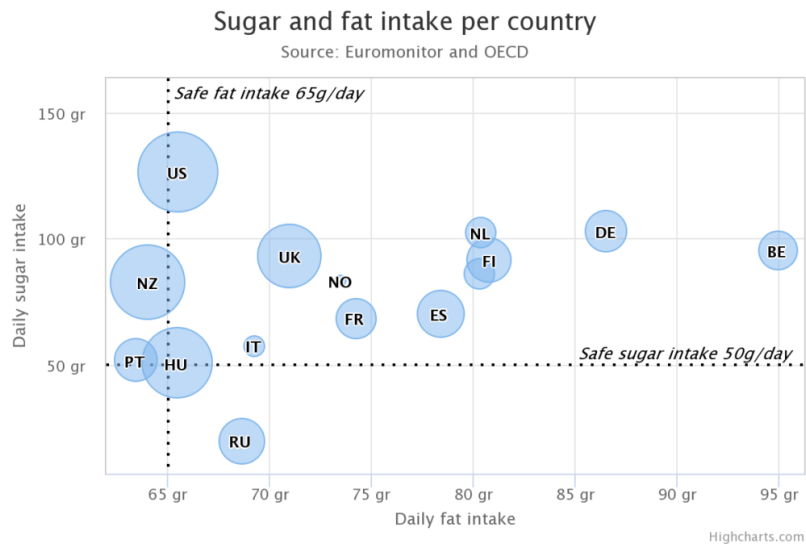


Figura 6: Gráfico de burbujas correlacionando la ingesta de azúcar y grasas
Fuente: OCDE

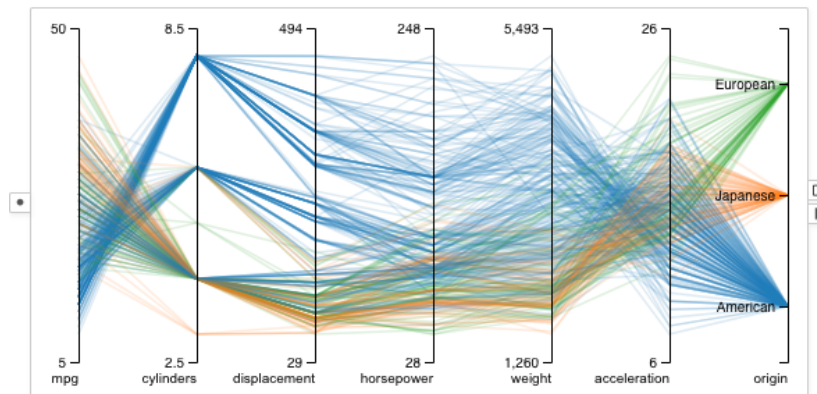


Figura 7: Gráfico de coordenadas paralelas comparando atributos entre vehículos
Fuente: VisFlow

2.3. Herramientas de visualización de datos

En la actualidad las herramientas utilizadas en la visualización de datos, que no incluyen el tratamiento de este, usan dos perspectivas o casos distintos de usos. La primera siendo la visualización pura de los datos y la segunda la exploración de estos. Existen múltiples trabajos [Kistler *et al.*, 2013, Liu *et al.*, 2018, Caldarola y Rinaldi, 2017] que resumen los distintos tipos de herramientas y las catalogan según varias aristas (necesidad de utilizar un lenguaje, especialidad en tipo de visualización, etc).

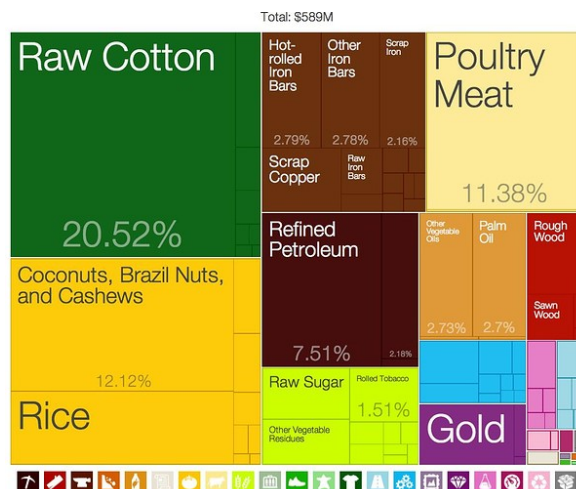


Figura 8: Treemap sobre las exportaciones de Benin durante 2009
Fuente: MetaBase.

2.3.1. Herramientas sin utilización de lenguajes de programación

Este tipo de herramientas son menos complejas para usuarios alejados de las ciencias informáticas, por ende, tienen una menor curva inicial.

Tableau⁸

Software diseñado para la inteligencia de negocios y análisis de datos. Puede utilizar tanto datos locales como remotos en bases de datos, soportando un amplio catálogo de formato de archivos. Provee una interfaz fácil de utilizar al visualizar datos, además de ser interactivo y flexible en su uso. Existe una versión gratis con la cual sólo se puede usar archivos de manera local.

ICharts⁹

ICharts es una aplicación web comercializada, esta integra plataformas de bases de datos las cuales son actualizadas de manera automática. La ventaja de usar datos en tiempo real es el obtener datos que denotan la situación actual en todo momento.

Infogram¹⁰

Infogram es una aplicación web diseñada para crear gráficos informativos. Se pueden integrar varios tipos de formatos para usar como conjunto de datos, desde archivos en posesión del usuario a importar datos desde la nube. Un problema de esta plataforma es que no existe privacidad de los datos a menos que el usuario sea uno de pago. Contiene una interfaz simple de usar, además existen una gran cantidad de ejemplos realizados para usuarios nuevos.

⁸Sitio web de Tableau (<https://www.tableau.com/>)

⁹Sitio web de ICharts (<https://www.icharts.net/>)

¹⁰Sitio web de Infogram (<https://infogram.com/>)

RAWGraphs¹¹

RAWGraph es una plataforma web que puede ser usada de manera directa sin registro. Soporta variados formatos de datos, datos provenientes desde APIs o desde plataformas en la nube. Solo utiliza el navegador local para el procesamiento y visualización de los datos, lo cual usa la librería D3.js de manera local. Es simple de usar, los datos utilizados están seguros dado el comportamiento local de la herramienta.

2.3.2. Herramientas que utilizan lenguajes de programación

Este tipo de herramientas necesitan que el usuario sepa utilizar el lenguaje en el cual fueron escritas las herramientas, por lo tanto hay un requisito o barrera de acceso inherente en ellas.

D3.js¹²

D3.js es una biblioteca de código abierto la cual está escrita en Javascript. Combina HTML y CSS, y los gráficos son generados en formato .svg. Opera mediante peticiones al navegador que utiliza la biblioteca. Cuenta con una comunidad activa y documentación extensa. Existen múltiples biblioteca que realizan trabajos similares a D3.js (Chart.js, Zingchart, Fusioncharts), pero no tienen la madurez de esta.

Gephi¹³

Gephi es un software de código abierto especializado en la visualización de redes y el análisis de estas implementado en java. Una de las virtudes de Gephi es su plataforma altamente interactiva, la cual hace participar al usuario en el proceso de gestionar los espacios para visualizar la red. La herramienta contiene diversos algoritmos utilizados para analizar redes, distribución de los nodos y otras funciones relevantes para las estructuras de grafos.

Ggplot2¹⁴

Es un paquete de software de código abierto para graficar y visualizar datos desde una perspectiva estadística, basado en las bibliotecas gráficas del lenguaje R. Habilita al usuario editar las visuales a un alto nivel de abstracción y a un nivel máximo de libertad modificando la mayoría de las opciones visuales. Contiene una gran cantidad de documentación y una comunidad activa. Se encuentra disponible en varios lenguajes de programación (Python y R) y ha inspirado proyectos similares (Lets-Plot, Chart, gadfly, etc).

Apache ECharts¹⁵

Es una herramienta de visualización de código abierto, pensada para integrar visualizaciones de datos en la web y ser utilizado en múltiples tipos de dispositivos. Contiene múltiples repre-

¹¹Sitio web de RAWGraphs (<https://rawgraphs.io/>)

¹²Sitio web de D3.js (<https://d3js.org/>)

¹³Sitio web de Gephi (<https://gephi.org/>)

¹⁴Sitio web de Ggplot2 (<https://ggplot2.tidyverse.org/>)

¹⁵Sitio web de Apache Echarts (<https://echarts.apache.org/>)

sentaciones distintas para los datos entregados y está en constante desarrollo por parte de la comunidad, además de tener el respaldo de la fundación Apache. Los lenguajes utilizados para crear las visualizaciones son Javascript o Typescript.

2.4. Interacción de usuarios e información

Un factor importante a considerar es la interacción del usuario final con los paneles desarrollados. Esta ha sido un área de estudio desde la concepción de herramientas de analítica para los negocios, pero el auge surgió con el rápido crecimiento de datos generados por los rubros que sufrieron una digitalización en sus procesos y sistemas, en donde el interés se ha centrado en dos inquietudes: ¿Cómo un usuario prefiere o se siente más cómodo frente a un panel de información? y la segunda pregunta se puede resumir con la interrogante ¿Existe algún flujo de trabajo que permita involucrar al usuario en la creación de los paneles de información?

La primera interrogante ha sido estudiada dentro de soluciones comerciales [Shen-Hsieh y Schindl, 2002], tomando en cuenta métricas relativas al uso de los paneles de información para la toma de decisiones, tanto operacionales como administrativas. Otro punto de vista interesante es cómo es la interacción entre personas e información (sea de cualquier tipo), últimamente el enfoque ha cambiado desde una interacción persona-computador a preferir una interacción del tipo persona-información [Marchionini, 2008] abstrayendo el uso del medio en el cual se obtiene.

Para la segunda pregunta, más que un flujo, la literatura recomienda tener en cuenta qué actividades o decisiones de diseño no tomar o intentar evitar [Few, 2006], esto para obtener como resultado visualizaciones más consistentes y efectivas para el público en general. También para este punto se hace énfasis en el uso de la percepción del usuario final, ya sea mediante las habilidades para distinguir tamaño, color, forma (conocidos como los principios de gestalt, los cuales evalúan cómo las personas asocian o disocian los datos presentados).

2.5. Teoría de la información y canales de comunicación

La teoría de la información estudia por una parte la codificación de esta en distintos tipos de secuencias, símbolos u otros tipos, y por otra, contempla el almacenamiento (de manera digital), cuantificación y por sobre todo su comunicación [Gallager, 1968]. Esta idea considera los siguientes entes en el proceso comunicativo: fuente, codificador, canal de ruido, decodificador y destino: un esquema de ejemplo se puede apreciar en la figura 9.

Es útil considerar la teoría de la información, aunque su propósito difiera con el fin del problema atacado, sea esta teoría relativa a cómo se transfiere la información en medios digitales y los métodos de cómo realizarlo de manera efectiva sin perder. Esta teoría puede ser utilizada

como analogía cambiando alguno de sus elementos en el proceso de creación de paneles de información o gráficos de manera general.

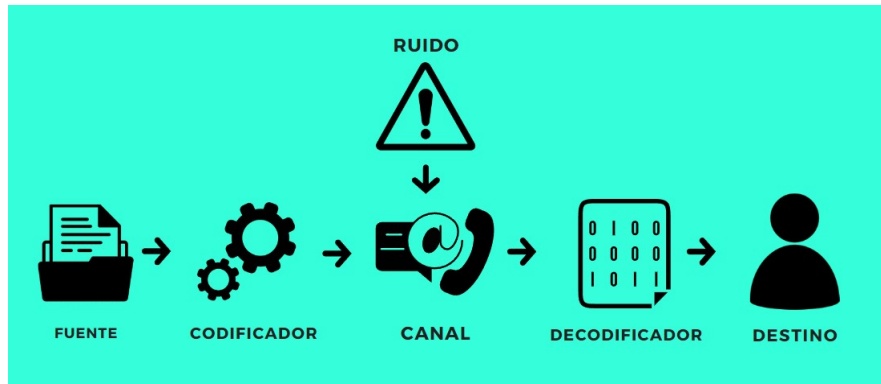


Figura 9: Sistema de comunicación
Fuente: Elaboración propia

La idea general es codificar los datos originales de tal manera que el usuario final sea capaz de recibir la información de manera oportuna. Esta similitud habilita a utilizar conceptos de la teoría [Wang y Shen, 2011] [Chen *et al.*, 2016], por ejemplo los canales de información y el ruido asociado en estos.

2.6. Modelos para la validación de visualizaciones

En cuanto a validación de visualizaciones, existen dos ramas de investigación. La primera corresponde en indagar cuáles visualizaciones son preferidas o permiten realizar ciertas tareas específicas a los usuarios cuando son solicitadas [Saket *et al.*, 2018]. Esta primera rama es llamada **efectividad en base a tareas**, en donde se verifica cómo se comportan distintas visualizaciones en diferentes contextos y evalúa su efectividad en una escala cuantitativa. Las tareas a validar son usualmente las siguientes:

- Encontrar anomalías
- Encontrar agrupaciones
- Encontrar correlaciones
- Calcular un valor derivado (agregación de los datos)
- Caracterizar la distribución de los datos
- Encontrar los extremos
- Filtrar
- Ordenar
- Determinar el rango de los datos
- Obtener un valor específico

Además estas tareas tienen tres métricas asociadas. La primera considera qué tan **precisa** es la visualización en la tarea, esto quiere decir una interpretación correcta de los datos expuestos, respecto a los usuarios que realizan la prueba de la tarea. La segunda métrica mide el **tiempo** en realizar la tarea consultada al usuario. Y la última métrica es respecto a la **preferencia del usuario**, que debe ser considerada ya que es un buen dato al crear visualizaciones para el público general y tiende a mejorar los paneles de información con un sentimiento de familiaridad.

En la figura 10 se encuentra un resumen de lo indagado respecto a la efectividad en base a tareas.

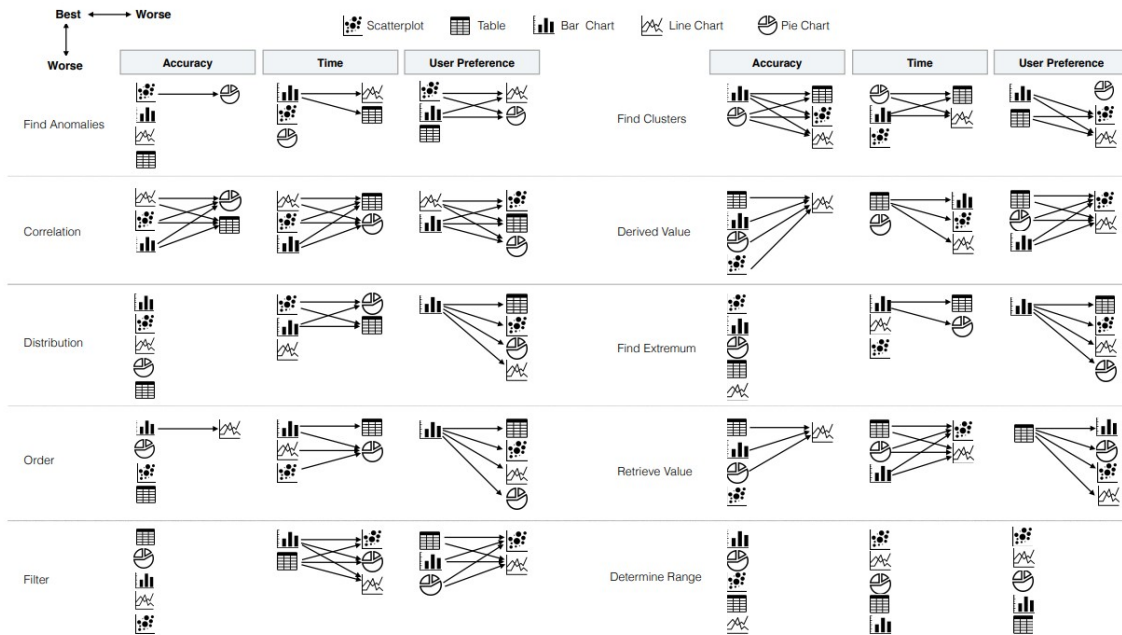


Figura 10: Relación por pares entre los tipos de visualización en las tareas y las métricas de rendimiento. Las flechas muestran que la fuente es significativamente mejor que el objetivo
Fuente: [Saket et al., 2018]

La segunda rama de investigación considera marcos de trabajo para definir o llegar a una visualización útil de lo que se desea mostrar al usuario. Un marco de trabajo propuesto [Lavalle et al., 2019] basado en requerimientos se basa en especificaciones que resumen lo necesario para llegar a un objetivo. Este modelo se puede apreciar en la figura 11, el cual representa las necesidades u objetivos, así como las tareas necesarias y el tipo de análisis necesario para responder a las tareas.

Además del modelo existe la especificación técnica que indica el qué consultar o preguntar al comenzar a realizar una visualización de interés. Esto ayuda a usuarios inexpertos brindar un contexto a la visualización, separado en dos ámbitos: uno respecto al usuario (características respecto al objetivo, las interacciones necesarias, etc) y el otro en relación a los datos

(la dimensionalidad, cardinalidad, etc). Un ejemplo de especificación del contexto se puede observar en la figura 12.

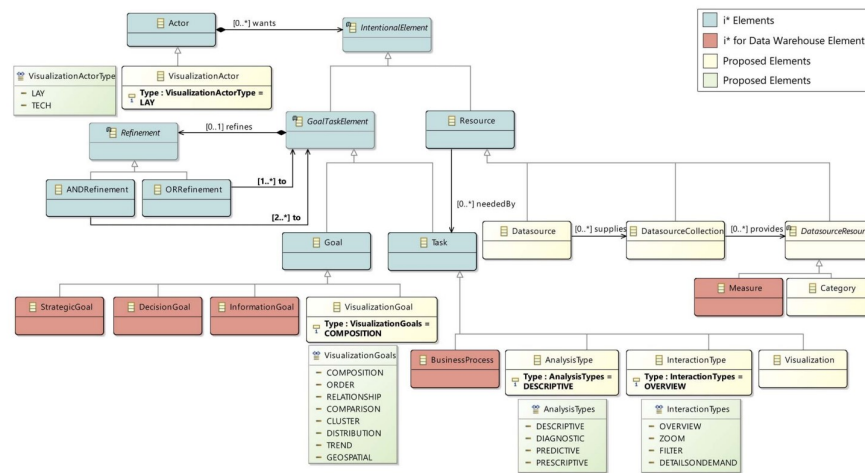


Figura 11: Especificación del metamodelo de visualización
Fuente: [Lavalle et al., 2019]

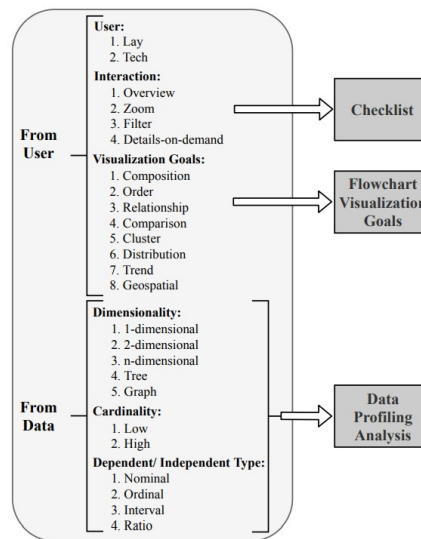


Figura 12: Especificación del contexto de visualización
Fuente: [Lavalle et al., 2019]

La idea de este modelo es crear un flujo de trabajo más rígido y evitar tener que comenzar desde cero en la creación, sea tanto del contexto o el modelo de metadatos, puesto que ya queda una pauta predefinida de lo realizado, y lo que no funcionó en un momento dado para volver a ser evaluado nuevamente.

Otro rama anexa de estudio interesante, es respecto a los métodos utilizados en la evaluación de las visualizaciones [Carpendale, 2008]. Esta área se centra en la validación de las gráficas

con usuarios para crear un marco de trabajo el cual se basa en métodos cuantitativos para la realización de pruebas y comprobación de las gráficas. Además se da pie a estudiar la naturaleza perceptual del usuario con distintos métodos, los factores que son evaluados en los usuarios son tres:

- **Generalidad:** Un resultado es generalizable en la medida que es aplicable a otros individuos.
- **Precisión:** Un resultado es preciso en la medida en que se puede definir sobre las medidas que se tomaron y sobre el control de los factores que no se pretendían estudiar.
- **Realismo:** Un resultado se considera realista en la medida en que el contexto en el que se ha estudiado es como el contexto en el que se utilizará.

En la figura 13 se puede apreciar las metodologías que se practican para evaluar según los criterios mencionados anteriormente.

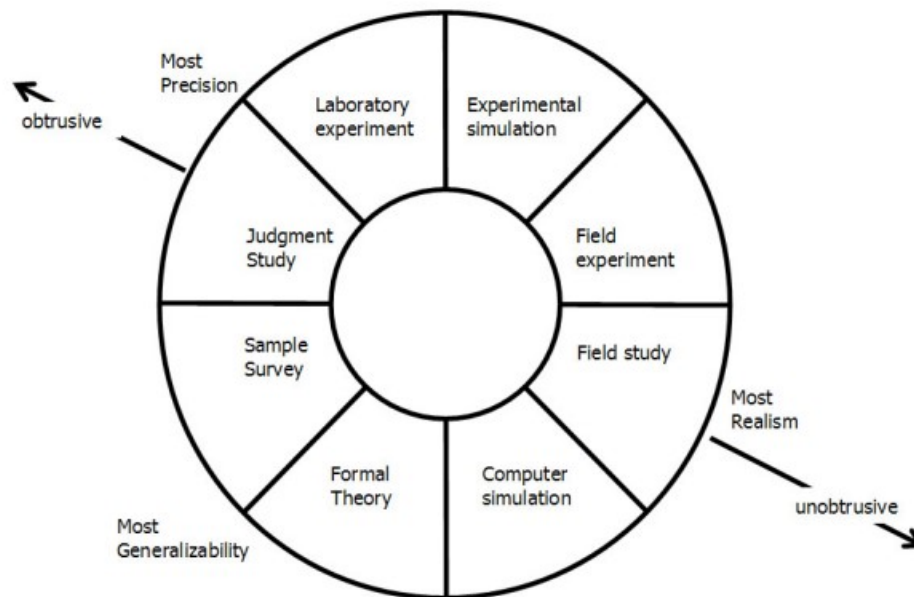


Figura 13: Tipos de metodologías organizadas para mostrar las relaciones con la precisión, la generalización y el realismo

Fuente: [Carpendale, 2008]

Además de los factores a evaluar, también se debe considerar el enfoque de la evaluación, si esta será de tipo cuantitativa o cualitativa. Aunque la mayoría de los estudios llaman a realizar evaluaciones de tipo cualitativa y el enfoque cuantitativo como un apoyo sobre lo ya realizado [Ellis y Dix, 2006] (haciendo acotación a una famosa frase encontrada en la oficina de Einstein, "Everything that can be counted does not necessarily count; everything that counts cannot necessarily be counted").

2.7. Estudios con los datos del problema

Dentro de las investigaciones y estudios realizados con el tema propuesto, lo más público y de conocimiento general se ha realizado como una iniciativa sin fines de lucro con el observatorio fiscal¹⁶. Esta idea comienza en 2016 con el fin de generar métricas y gráficas digeribles para la población, resumiendo sobre el uso de las arcas fiscales a un nivel de detalle bajo. Este trabajo es un muy buen primer avance, dado el alcance y visión que tiene sobre el producto final entregado; lo que difiere con lo propuesto es sobre el nivel de detalle el cual no parece ser suficiente para usuarios más avanzados o con consultas particulares.

Otros trabajos hacen uso de una parte de las licitaciones; para ser más específico, las órdenes de compra realizadas por la central de abastecimiento (CENABAST). Esta central influye en el abastecimientos de todo tipo de insumos médicos utilizados tanto por el sector público como privado, manteniendo el inventario y la demanda prevista bajo control. Los estudios realizados se hacen cuanto a un nivel de efectividad de la institución [Galdames Paredes, 2015] o bien a un nivel operacional [Madariaga Torres, 2021]. Este tipo de estudios dista de lo que se ha propuesto, ya que evalúa el rendimiento y las posibles mejoras a una institución particular, más que el brindar información sobre lo realizado.

Un estudio reciente sobre el conjunto a trabajar, se basa en analizar de manera estadística el conjunto de datos [Velásquez Pizarro, 2022]. Propone un análisis multivariable en el cual compara órdenes de compra con solo un ítem y más de uno, con la finalidad de realizar una reducción de dimensionalidad.

Con los estudios postulados anteriormente se puede apreciar que se han atacado distintos ángulos, pero todos difieren ya sea en la finalidad o la granularidad del mismo.

2.8. Metodologías y flujos de trabajo

Existen múltiples métodos de trabajo efectivos para desarrollar software [Despa, 2014] y llevar un versionado de este [Loeliger y McCullough, 2012].

Dadaa las circunstancias del trabajo a realizar, ciertas metodologías quedan fuera de una consideración inicial como sea el desarrollo en cascada, puesto que es poco viable planificar y esperar hasta el final del desarrollo por el resultado obtenido.

A continuación se mencionan metodologías para el desarrollo de software viables para la tarea en cuestión, ya sea por la modalidad propuesta o que los requisitos fundamentales se alinean con los objetivos a conseguir.

¹⁶Sitio web del observatorio fiscal

2.8.1. Kanban

Kanban [Anderson y Carmichael, 2016] es un método ágil de desarrollo basado en balancear las demandas con las capacidades disponibles. Este método se originó para mejorar la eficiencia en procesos de manufactura en una planta de autos Toyota. Con el tiempo se observó la versatilidad del modelo y empezó a ser adaptado por equipos de desarrollo de software (en 2004 múltiples áreas de Microsoft obtuvieron buenos resultados al aplicar el modelo).

El modelo contempla una pizarra la cual es dividida verticalmente, donde cada separación vertical representa un estado o acción. En esta pizarra se agregan tarjetas o cartas que representan tareas a realizar. Con una representación de la pizarra y las tarjetas se puede evidenciar el estado del proyecto. Un ejemplo de la representación del modelo en uso se encuentra en la figura 14.

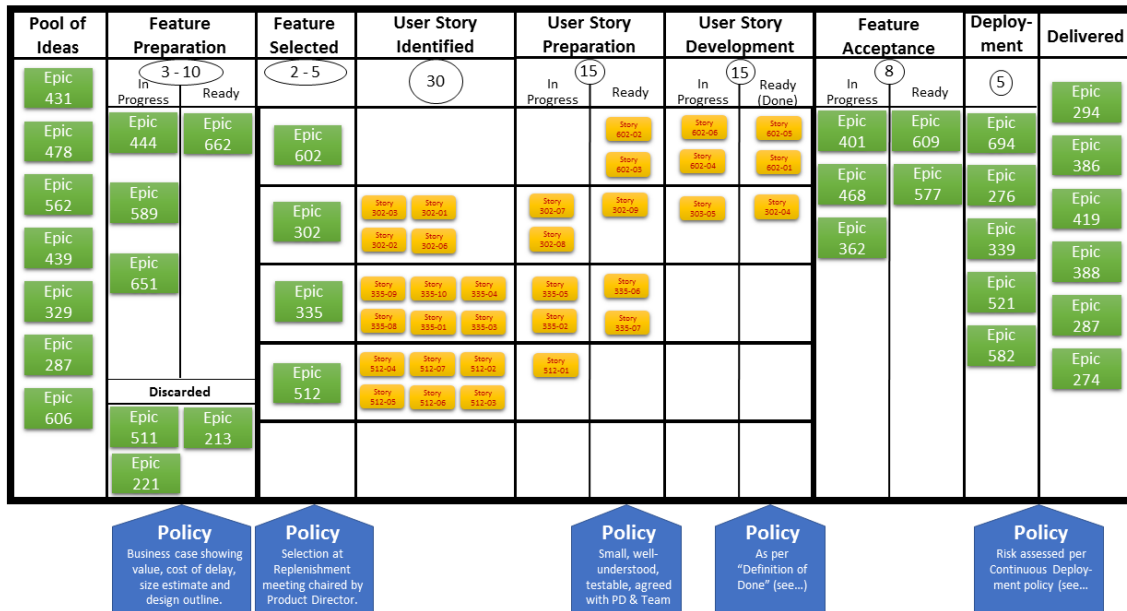


Figura 14: Instanciación y uso de kanban
Fuente: Wikimedia

Además el modelo contempla restricciones a la cantidad de tarjetas que pueden estar en el tablero en un momento dado, esto es para balancear la carga asignada. Estas restricciones son asignadas respecto a las condiciones del equipo, las tareas a realizar y las etapas utilizadas en la pizarra, y pueden variar en el tiempo según el rendimiento del equipo y las métricas de evaluación utilizadas.

El flujo de trabajo con kanban es muy simple e intuitivo, las tareas y sus estados son visibles para todos los participantes del proyecto. Además, la subdivisión e identificación de tareas permite el obtener métricas para saber la capacidad del equipo que utiliza el método, así

como su eficiencia.

Algunas de las métricas utilizadas por el modelo de desarrollo son las siguientes:

- **Tiempo de ciclo:** Métrica que indica cuánto demora una tarjeta en pasar de un estado al siguiente en promedio, que puede ser respecto a una tarea o columna en particular.
- **Tiempo de entrega:** Es el tiempo utilizado desde que se instauran las tareas y estas son realizadas hasta su etapa final; esta es la métrica que el usuario final esta más interesado, normalmente.
- **Flujo acumulativo:** Esta métrica es graficada usualmente e indica el general del proyecto en el tiempo sumando las tareas completadas por etapa de manera acumulativa.

El método kanban es recomendado para equipos con pocos integrantes y no es necesario tener experiencia con métodos ágiles de desarrollo para ser utilizado de manera óptima, debido a la simpleza del flujo de trabajo.

2.8.2. Scrum

Scrum es un método para administrar el desarrollo en proyectos, en particular desarrollos de software [Schwaber, 1997]. Este método está diseñado para equipos reducidos (10 o menos integrantes) en donde se realizan iteraciones sobre el trabajo establecido para hacer revisiones o modificaciones a la planificación inicial.

La idea principal es el desarrollo continuo; con la naturaleza iterativa e incremental del flujo de trabajo propuesto se obtiene un producto en las etapas tempranas del proyecto, el cual tiene un alto valor para el usuario final.

El método es muy estructurado en las prácticas y actividades a realizar, pero lo suficiente flexible para integrar lo necesario por el equipo que utilizará Scrum. Las iteraciones derivadas en el desarrollo son nombradas *sprint*, que tienen por finalidad dar un intervalo fijo de tiempo que se utiliza para las tareas propuestas y revisar si se necesitan modificaciones al término del periodo.

Las actividades propuestas se encuentran definidas en la figura 15.

Los eventos indicados son utilizados para monitorizar el avance del proyecto y cambiar el foco del mismo de ser necesario. Además, con cada iteración es posible identificar puntos de mejora en el proceso de desarrollo o de ejecución de Scrum. Este ciclo de desarrollo y *feedback* es fundamental en el marco de trabajo propuesto por Scrum y a su vez, contempla personas dedicadas a monitorear que el flujo de trabajo (*scrum master*) y las instancias de mejora que este provee se estén aprovechando.

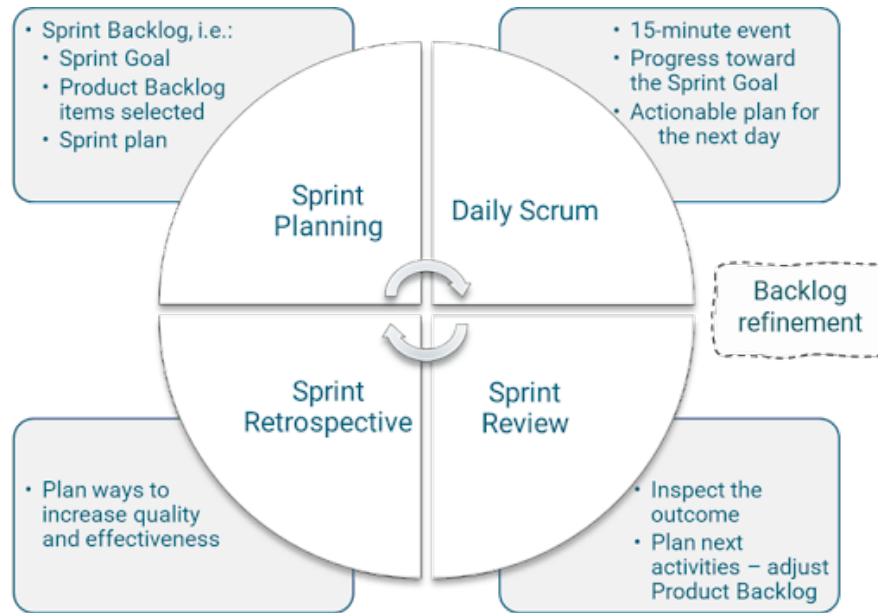


Figura 15: Actividades en el marco de trabajo SCRUM
Fuente: Wikimedia

Una instancia del proceso completo del marco de trabajo se puede apreciar en la figura 16

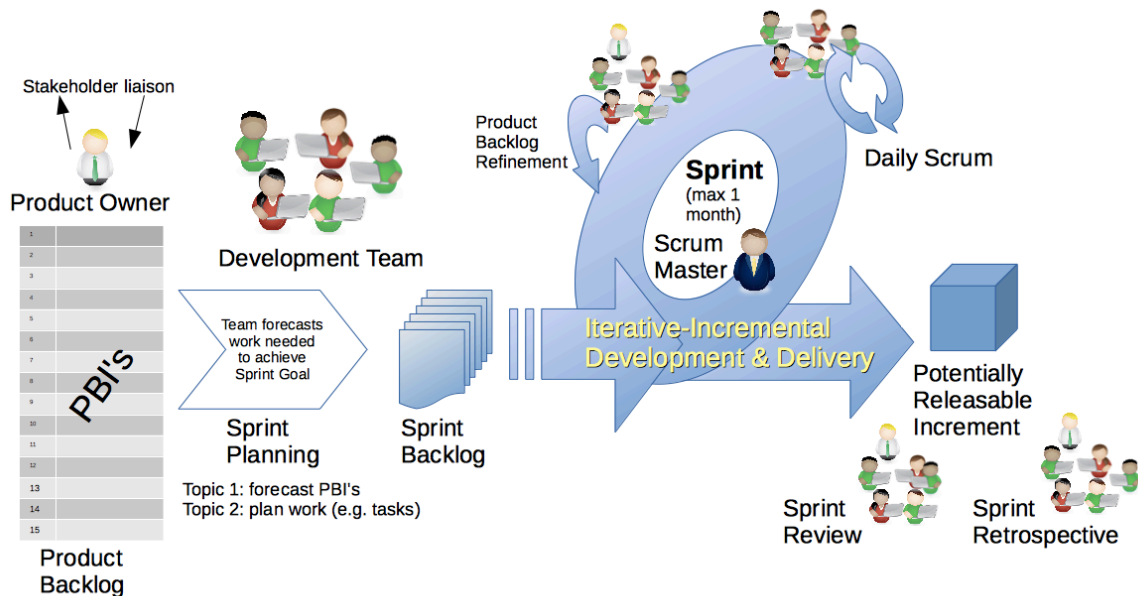


Figura 16: Instancia en el marco de trabajo SCRUM
Fuente: Wikimedia

Este método de trabajo necesita un alineamiento de los participantes sobre la metodología ágil de desarrollo [Fowler *et al.*, 2001], por lo que no es fácilmente asimilada y es neces-

rio un intervalo de tiempo para adaptarse. El modelo contempla la mejora continua, la cual permite las instancias de mejora en el equipo de ser necesario y habilita que los eventos se robustezcan a modo que los participantes obtienen mayor experiencia con el método.

Como método para desarrollar software ha mostrado ser eficiente [Hossain *et al.*, 2009] (en medida que se sigan las reglas y eventos propuestos), pero contempla un periodo de adaptación que puede ser extenso dependiendo de los participantes y su experiencia con metodologías hábiles [López-Martínez *et al.*, 2016]. Además, este marco de trabajo es utilizado mayormente por grupos de personas, dado que para el uso personal existen muchos eventos sin valor agregado y el *feedback* se ve afectado sin tener perspectivas distintas.

CAPÍTULO 3

PROPUESTA DE SOLUCIÓN

En este capítulo se describe el cómo se implementa la solución que satisface los objetivos propuestos. Se hallan las tecnologías y la toma de decisiones que justifican su uso en el diseño de la solución.

3.1. Metodología de trabajo

Dentro de las etapas importantes a realizar en el proyecto, existen cinco a desarrollar. La primera etapa se trata de evaluar el conjunto de datos, tanto la extracción de los datos como el procesamiento necesario (transformaciones y carga). Esta etapa es fundamental para el trabajo posterior, dado que según el origen y formato de los datos se determinan las herramientas y métodos a utilizar para el desarrollo de la arquitectura que utilizada en la solución. Aunque sea una etapa más bien exploratoria, esta define en gran parte el trabajo y las decisiones posteriores según las conclusiones que se tomen sobre el conjunto de datos y su estructura particular.

Luego de tener una idea sobre el conjunto de datos a utilizar, se proponen herramientas que ayuden con las tareas posteriores, esto, tanto para el proceso ETL (extracción, transformación y carga de datos) como el de creación de las visualizaciones. Las decisiones son tomadas respecto a la exploración inicial del conjunto de datos, con el fin de elegir las tecnologías que brindan las características necesarias para la manipulación de los datos según la estructura de estos, y las técnicas que brinden el mejor entorno para comunicar la información resultante de las visualizaciones propuestas. Además, en este instante de la solución se considera un método con el cual trabajar a continuación en el desarrollo de la solución, esto para conseguir un procedimiento al proponer un marco de trabajo y conjunto de herramientas.

Como tercer etapa está la manipulación del conjunto de datos. Este es el más relevante al desarrollo pues alimenta todo lo realizado a continuación; y es el etapa que toma más tiempo y recursos de planificación puesto que los procesos de limpieza de datos y documentación de lo extraído necesitan varias iteraciones, hasta que se consiga un resultado final lo suficientemente robusto para continuar con el proceso de visualización. El proceso de ETL conlleva mucho ensayo y error, dado que es un proceso exploratorio e iterativo; es muy importante darle el tiempo a este desarrollo puesto que los errores u omisiones en esta etapa, que no sean debidamente resguardados, tendrán consecuencias mayores en los pasos posteriores, atrasando el desarrollo considerablemente.

Con los datos revisados y estructurados con el ETL se puede proceder al cuarto etapa, que considera la creación de las visualizaciones y otorgar todo lo resultante al dominio público para que se produzca un entorno colaborativo al desarrollo, si tiene un buen nivel de adopción

por parte del público. Esta etapa toma lo planificado en el etapa dos a nivel de herramientas y tecnologías definidas, y el tercer etapa a nivel de datos a utilizar para el diseño de las gráficas en cuestión. Como se ha mencionado anteriormente, los errores en el etapa anterior retrasan enormemente el avance de esta etapa dado que, por cada error del ETL, este se debe volver a formular, y lo más probable es que se produzcan cambios en las estructuras de datos ya utilizadas. Las plataformas para compartir lo desarrollado pueden ser elegidas luego del término de esta etapa, debido a que lo realizado será sometido a pruebas antes de ser publicado; por lo tanto no hay una motivación a compartir lo realizado si esto no se ha comprobado efectivo.

Como último etapa se procede a validar lo realizado con una muestra representativa de personas, mediante un conjunto de pruebas. Este trabajo será realizado en dos partes como se mencionó en el marco conceptual; se medirá cualitativamente lo realizado para obtener una perspectiva de los usuarios finales, y cuantitativamente para verificar que las visualizaciones planteadas cumpla con los objetivos propuestos. Esta dualidad es propuesta en el marco conceptual priorizando la generalización y precisión.

3.2. Etapa 1: Evaluación del conjunto de datos

Al evaluar el conjunto de datos debe intentar responder ciertas preguntas sobre la colección que esta analizando. Lo primero que se debe averiguar es el origen de los datos así como su formato.

Para satisfacer la información sobre el origen de los datos se debe proporcionar el contexto de estos. En el proceso de las licitaciones públicas (y en su medida las privadas también) se requiere una amplia gama de información, desde qué se necesita obtener con la licitación (ya sea un bien o servicio) y bajo qué condiciones se propone para que los interesados que decidan entregar una oferta por ello puedan hacerla bajo una base común de información.

Una licitación tiene normalmente cinco partes representadas en la figura 17; cada parte contiene información fundamental sobre el proceso que ha sido estandarizado y utilizado por gobiernos de todo el mundo. El estándar utilizado es el OCDS ¹⁷ o Open Contracting Data Standard (Estándar de Datos de Contratación Abierta); este considera una estructura de datos en formato I-JSON definido en el RFC7493. Este nivel de documentación sobre la estructura del archivo proporcionado desde el estándar lo hacen fácil de utilizar y entender en un periodo corto de tiempo, con lo cual la interrogante de los orígenes y el formato de los datos es zanjada rápidamente.

Sabiendo cómo el estándar es utilizado en el mundo y el formato en el cual la información es transportada, falta verificar si el conjunto que se necesita sigue los mismos patrones dictados por la literatura. Al investigar los canales de Mercado Público llama la atención que existen tres maneras de obtener la información. La primera es mediante una opción de descarga

¹⁷Sitio OCDS (<https://standard.open-contracting.org/latest/en/schema/reference/>)

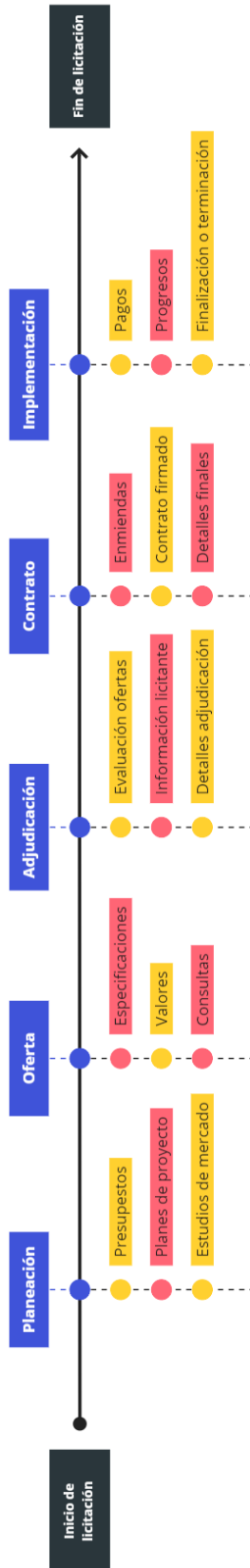


Figura 17: Proceso de una licitación.
Fuente: Elaboración propia.

masiva, la cual propone descargar la información de un periodo, pero sin mayor referencia de cómo está estructurado el archivo ofrecido. Como segunda y tercera opciones existen dos API las cuales difieren en su funcionamiento, una de ellas corresponde al estándar OCDS y la otra es mediante un acceso por *ticket* que debe ser solicitado con un formulario.

El propósito es utilizar un trabajo que sirva de base para otros, por lo cual se optó por la API con el estándar OCDS; los archivos de descarga masiva pueden simplificar el proceso pero tienden a ser más manuales en su uso, impidiendo automatizar un proceso de extracción en el futuro, en contraste con la utilización de una API.

Al indagar el funcionamiento de la API expuesta en el sitio de Mercado Público se identifica que para obtener las licitaciones (el mismo caso con los tratos directos y convenios marco), primero se debe obtener un listado el cual tiene un largo máximo (dado con los parámetros *offset* y *limit* de la consulta) de 1000 registros, por lo cual se deben hacer múltiples visitas a la API para obtener los enlaces. Además, se debe consultar por un mes y año específico, dado que no se existe una acción singular que obtenga el año completo.

Con el listado de las licitaciones, se debe iterar sobre este dado que el listado brinda los enlaces particulares de cada licitación para obtenerlo desde la API. Este proceso iterativo puede resultar complejo dada la cantidad de licitaciones y tratos directos a analizar y obtener desde la API. La complejidad no es mucho a un nivel de solución, pero sí de obtención de la información, la cual se puede ver afectada si se rechaza el acceso por muchas consultas de manera simultánea o en un periodo particular.

Al intentar utilizar unos de los enlaces proveídos por el listado se obtiene un archivo JSON con la información sobre el estado y propiedades de la licitación, trato directo o convenio marco consultado. Con esta información se puede avanzar a diseñar una solución, puesto que se sabe el medio por el cual se obtendrá la información, el formato en el cual viene y el esquema que emplea para desglosar la información.

3.3. Etapa 2: Herramientas y tecnologías

La arquitectura de la solución a desarrollar consta de tres partes. La primera incluye todo lo relevante a la obtención de los datos, esto va desde la extracción de los datos y la transformación de los mismos. No necesariamente la herramienta de extracción debe ser la misma que la utilizada para el paso transformativo, como a su vez no necesariamente una sola herramienta debe ser utilizada en un solo paso.

Si bien los datos serán extraídos desde una API, existen muchos métodos para la extracción inicial de datos. Lo más plausible sería utilizar algún lenguaje de programación el cual obtenga el listado y con este iterar sobre los enlaces obtenidos y descargar la información. El problema de utilizar este enfoque monolítico es la poca optimización de las bibliotecas en los lenguajes de programación para crear solicitudes HTTPS, las cuales no están pensadas para el

volumen necesario. Este paso sería más sensato realizarlo con herramientas pensadas para la carga y descarga de datos como lo serían *cURL* o *wget*. El paso de extracción de los enlaces, al no tener el mayor impacto, puede ser desarrollado en un lenguaje de programación de *scripting* cuya velocidad de desarrollo es mayor.

Sobre la manipulación y carga de los datos, existen muchas soluciones dependiendo del contexto en la cual se encuentra la arquitectura propuesta. Como se mencionó anteriormente, la estructura de los datos es un archivo JSON por cada licitación o trato directo obtenido en el paso de extracción. La manipulación puede ser realizada con lenguajes de programación con bibliotecas del tipo Pandas en Python y dplyr en R. También hay programas especializados en crear flujos de datos para usuarios que no tienen mucha experiencia creando código como RapidMiner, Tableau Prep o Hitachi Pentaho (Pentaho siendo la más difícil de instanciar). Una opción también es realizar la carga antes de la manipulación, esto se realiza cargando los datos en alguna base de datos previo a la manipulación; en donde la base de datos se especializa en el uso de estructuras de datos tipo JSON en la cual se puedan realizar las operaciones de forma nativa como lo son MongoDB o CouchDB.

La desventaja de utilizar un lenguaje de programación en la manipulación es el posible formato de las estructuras y que tan robustas son las bibliotecas a utilizar para la lectura de los archivos JSON. Al contrario, las bases de datos que aceptan este formato de manera nativa proveen una base en la cual trabajar, en donde el esfuerzo sería realizar la carga inicial de los archivos dado su volumen (esto puede ser complementado por un lenguaje de programación que tenga soporte para un conector a la base de datos). Como se menciona, la ventaja inherente de las bases de datos es el tener que preocuparse de la carga de los datos y conocer bien el lenguaje en el cual se realizan las manipulaciones, dado que la mayoría de las bases de datos que tienen soporte nativo de JSON tienden a ser NOSQL, y estas por lo general utilizan lenguajes distintos a SQL para crear el flujo de manipulación de los datos.

Otro punto a analizar en el uso de herramientas es en el campo de las visualizaciones a diseñar, luego de obtener el conjunto de datos; en este campo las soluciones son muy variadas, en donde el factor más importante es qué tan a gusto se encuentra la persona que va a desarrollar las gráficas con realizar código en algún lenguaje de programación. Esto se debe a que existen muchas soluciones en las cuales se requiere de escribir el código para obtener las visualizaciones, como el caso de D3.js y Matplotlib entre otros. Una de las desventajas de estas herramientas es que se utiliza mucho tiempo en la curva de aprendizaje de la biblioteca, aún si se conoce el lenguaje de programación; además, los objetivos apuntan a que lo realizado sea compartido con el público general lo cual no es una prioridad o no está previsto como parte de la solución en estas herramientas.

Cuando el usuario solo necesita concentrarse en la creación de las visualizaciones es recomendable utilizar algún tipo de herramienta de inteligencia de negocios, pensadas en crear gráficas o paneles de información con la idea de compartirlas dentro del negocio para la propagación de datos y conocimiento dentro de la organización, sin tener mayor conocimiento sobre lenguajes de programación (además su curva de aprendizaje tiende a ser muy amigable). El ámbito colaborativo de estas herramientas es muy importante para el trabajo en

cuestión pues se alinea con los objetivos a realizar. Aplicaciones de este tipo son Tableau, Qlik y PowerBI, en donde Tableau sería la más recomendable al tener una licencia educativa ofreciendo el producto completo, y poder publicar lo desarrollado en Tableau Public sin mayor coste al usuario o a quien desarrolla la solución.

3.3.1. Comparativa de las herramientas a utilizar

Ya con las posibles alternativas de herramientas a utilizar en el desarrollo, se procede a comparar las ventajas y desventajas que traen consigo al ser integradas a la arquitectura de la solución.

Extracción de datos desde la API

La extracción de los datos se puede dividir en dos partes, la primera es la extracción de la lista de enlaces dado un periodo en particular y la segunda es la propia descarga de información de cada uno de los enlaces proporcionados en el listado.

Para la primera parte de la extracción es importante discutir cuál lenguaje de programación es más apto para esta tarea, que si una herramienta o programa de terceros. Esto, debido a que solo se realizarán peticiones HTTP a una API y la cantidad de información no es muy relevante en volumen (50000 enlaces en promedio para convenios marco por periodo, 30000 para tratos directos y 8000 para licitaciones, en donde el máximo por petición son 1000 enlaces). Dada la baja intensidad del trabajo a realizar en esta actividad, el único cuello de botella posible está en la conexión de la API, por lo tanto la velocidad del lenguaje a utilizar no es relevante. Bajo estas condiciones el candidato ideal sería un lenguaje de *scripting* el cual permita crear una solución rápidamente, por lo cual la opción más razonable es Python.

En cuanto a la segunda parte de la extracción hay una variable de volumen a considerar, dado que la descarga por API de una gran cantidad de solicitudes tiene desafíos propios. Lo primero es el balance de las peticiones en un periodo determinado para no ser bloqueado por el servidor que ofrece los datos; en lenguajes de programación esto se puede realizar de manera temporal o con multiprocesamiento, pero a costa de aumentar la complejidad del problema sin un mayor beneficio (en un principio se utilizó este método pero con malos resultados los cuales serán discutidos en la sección 3.6).

Otra alternativa es el utilizar una consola UNIX y las herramientas disponibles en ella, dado que una de las principales funcionalidades es el poder realizar múltiples tareas en donde el poder no viene de las herramientas en el sistema, pero sí de la comunicación y trabajo entre ellas. Una opción en este ámbito sería *wget* en conjunto con *xargs*, en donde *xargs* es capaz de utilizar un archivo con los enlaces y entregarlos a *wget* el cual solicitará la información al servidor. Entre las múltiples ventajas de *wget* se encuentra la resiliencia a errores, dado que es posible pasar múltiples veces el listado a descargar, pero solo descargará lo restante (en caso de problemas de conexión o bloqueo por parte del servidor) o que haya tenido cambios

en la información entregada por el servidor. Con las ventajas mencionadas, el uso de UNIX se acerca más a las necesidades de la tarea en cuestión tanto en volumen como en facilidad de uso, dado que el desarrollo es un simple *script* de *bash* alimentado por un archivo con los enlaces a consultar.

Carga y manipulación de los datos

Como se mencionó anteriormente, tiene más sentido realizar la carga y luego la manipulación de los datos, esto para utilizar las ventajas de ciertas tecnologías como las bases de datos no relacionales. El porqué utilizar este tipo de herramienta sobre un proceso clásico de extracción, transformación y carga, es dado que el procesar el volumen de información no es óptimo en un lenguaje de programación sin tener que desarrollar e implementar métodos y bibliotecas especializadas para la gran cantidad de datos a utilizar, sin considerar las posibles transformaciones que se deseen realizar a los datos. Por estos motivos las ventajas de utilizar una base NoSQL, como MongoDB, superan las desventajas de tener que instalar una instancia de la base para utilizarla.

También existe la posibilidad de usar herramientas especializadas en la manipulación y carga de datos, como lo son Tableau Prep, RapidMiner, Talend o Pentaho. Las desventajas de estas herramientas son variadas, desde necesitar una licencia como lo es en el caso de Tableau (si se desean conectar bases de datos como fuente, pero con archivos tipo csv basta con la versión gratis) o RapidMiner, o necesitar una instancia de algún sistema operativo, en especial para ejecutar la herramienta en el caso de Pentaho. Pero una de las desventajas mayores, es la cantidad de recursos computacionales necesarios para el proceso de carga y transformación de datos dado el volumen de archivos en el conjunto en cuestión; esto es no muy relevante en el ámbito empresarial, pero al realizar la arquitectura de la solución no se está considerando equipo especial o dedicado para la solución.

Volviendo a la idea de utilizar una base de datos no relacional, esto sin duda simplificará la solución, dado que sólo es relevante la carga de los datos mediante la ayuda de un programa en algún lenguaje de programación (en donde MongoDB tiene un amplio soporte), y el flujo de transformación de datos también puede ser realizado mediante la misma conexión desde el lenguaje de programación, pero el trabajo será realizado por la base de datos. El único hincapié en el uso de MongoDB y es la posible curva de aprendizaje de su lenguaje MQL y JSON, dado que se encuentra con una base no relacional y esta no necesariamente utiliza SQL para las consultas a realizar en el proceso de transformación de datos.

Visualizaciones de datos

Para el proceso de creación de visualizaciones de datos, hay tres importantes factores a considerar: facilidad de uso, personalización de las visuales y capacidad de compartir lo producido. La facilidad de uso es relativa a qué tan compleja es la curva de aprendizaje para cada herramienta; en particular, programas de inteligencia de negocios como Tableau o Qlik son más simples de utilizar desde un comienzo, dada la interactividad de la herramienta de ver instantáneamente lo realizado. En contraparte, las herramientas que se basan en bibliotecas

de lenguajes de programación, como matplotlib o datapane, tienen una curva de aprendizaje mucho más elevada sin considerar el dominio del desarrollador en el lenguaje de programación que las bibliotecas utilizan.

Sobre las características en la personalización de las visualizaciones no hay una mayor distinción entre las herramientas; puede existir una visualización que sea más difícil de generar sobre otra herramienta, pero no existen muchos casos de una visualización relevante que conviva en una herramienta particular. Para este punto no hay mayores discrepancias entre las alternativas, el nivel de personalización es el adecuado para la solución que se desea brindar.

En cuanto a la capacidad de compartir lo desarrollado, existe una brecha entre las herramientas. Las bibliotecas en los lenguajes de programación deben ser exportadas en algún formato para ser alojadas en algún servidor, esto genera gastos adicionales en la solución. Al contrario, las herramientas de inteligencia de negocios incluyen alguna manera de exportar lo desarrollado a algún portal en el donde pueden ser visualizadas. Un caso particular es Tableau, el cual proporciona Tableau Public para la difusión gratuita de lo desarrollado supliendo este pilar relevante a los objetivos propuestos anteriormente.

Dadas las ventajas comparativas importantes de Tableau como visualizador de las gráficas realizadas sobre las herramientas propuestas, se propone como la herramienta que mejor se acopla a las características necesarias y cumple con los objetivos propuestos.

Flujo de trabajo

El flujo de trabajo representa cómo se desarrollan los objetivos y qué métodos de trabajo se utilizarán para crear la solución final. Dado que es un trabajo unipersonal, tiene mucho sentido el utilizar el método kanban para la administración del proyecto y tareas que se desprenden de este, puesto que un método como Scrum tiene muchas actividades enfocadas a un equipo, en contraste a la naturaleza individual de este trabajo.

Una muestra de lo utilizado en el desarrollo se puede ver en la figura 18, en donde se aprecian las columnas utilizadas, las tareas a desarrollar y el estado en el cual se encuentran. Este espacio de trabajo fue creado en Trello (Miro también permite una funcionalidad similar pero más básica, esto basado a experiencias personales con ambas plataformas). Esta plataforma permite crear pizarras interactivas para el desarrollo personal o en equipos según las preferencias elegidas por el administrador de la pizarra.

Las tareas o etapas importantes a realizar ya fueron descritos tanto en las secciones 3.1 y 3.2; en la figura 19 el flujo ideal de los acontecimientos en el desarrollo, aunque en caso de imprevistos una etapa puede ser postergada o removida del flujo. Dada la linealidad del trabajo que se realizará los errores, a lo más, pueden tener un cambio en las tecnologías o herramientas a utilizar más que las tareas a concretar.

En cuanto a avances y descubrimientos estos serán dialogados con el profesor guía de ma-

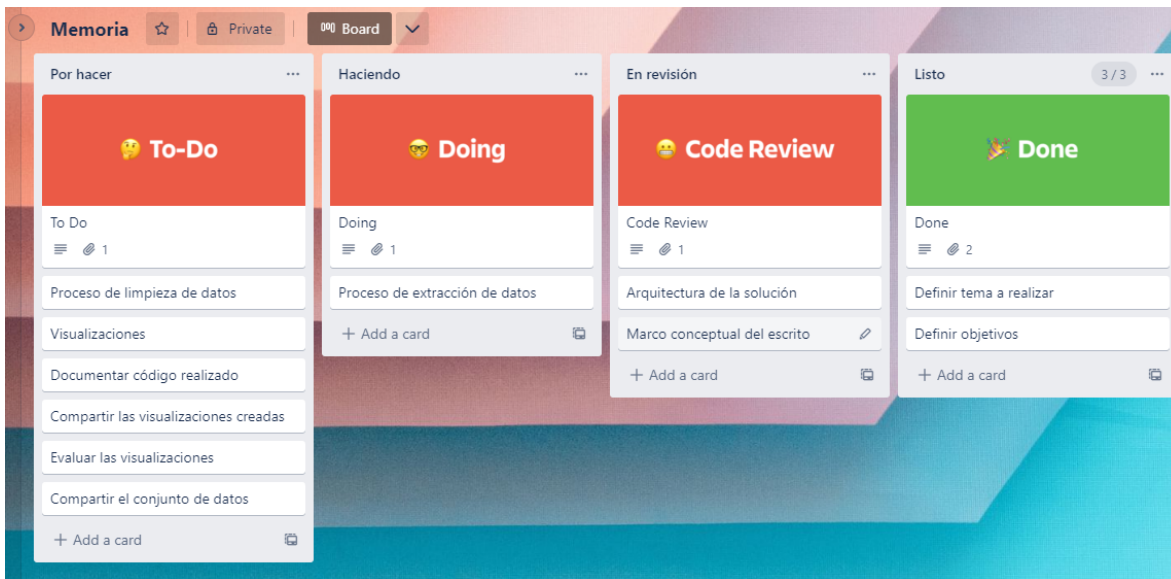


Figura 18: Tablero kanban utilizado.
Fuente: Elaboración propia.

nera periódica, a través de reuniones (planificadas de manera semanal o bisemanal los días martes) para discutir que la perspectiva de lo desarrollado se alineen con los objetivos propuestos desde un comienzo. Estas reuniones son muy importantes, pues es una de las pocas ocasiones de retroalimentación a lo desarrollado, por la naturaleza personal del trabajo.

Con kanban como método para administrar el avance del proyecto, las reuniones periódicas como instancia de retroalimentación sobre lo desarrollado y perspectiva futura para modificaciones o afirmaciones en el trabajo a realizar, y las tecnologías definidas para la utilización dentro del proyecto, se obtuvo un marco de trabajo con bases claras el cual permitió comenzar a desarrollar la solución a un nivel más técnico.

3.4. Etapa 3: Extracción, transformación y carga de datos

El proceso de ETL utilizado en este trabajo pasó por dos versiones importantes, dadas las dificultades encontradas, tanto en el proceso de extracción como el de transformación, al tener que trabajar con archivos JSON.

Como primer avance se diseñó una solución monolítica con Python, la cual consideraba todas las etapas desarrolladas en esta herramienta, como se puede apreciar en la figura 19, a excepción de la generación de visualizaciones la cual consideró el uso de Tableau. Este modelo resultó ser inviable por algunos percances que se encontraron en el desarrollo; el primero respecto a la velocidad en la cual se obtenían los archivos JSON desde los enlaces proporcionados por la API, debido a que el servidor parece contar con una cuota en la conexión

proveída. Esto resultó en errores de las conexiones e inhabilitaba la descarga por un tiempo determinado (una de las soluciones probadas para este caso es el uso de una VPN), cual hace que la etapa de extracción tome más tiempo del aceptable para la cantidad de datos, o llanamente que no sea posible extraer los datos en su totalidad.

Proceso ETL V1

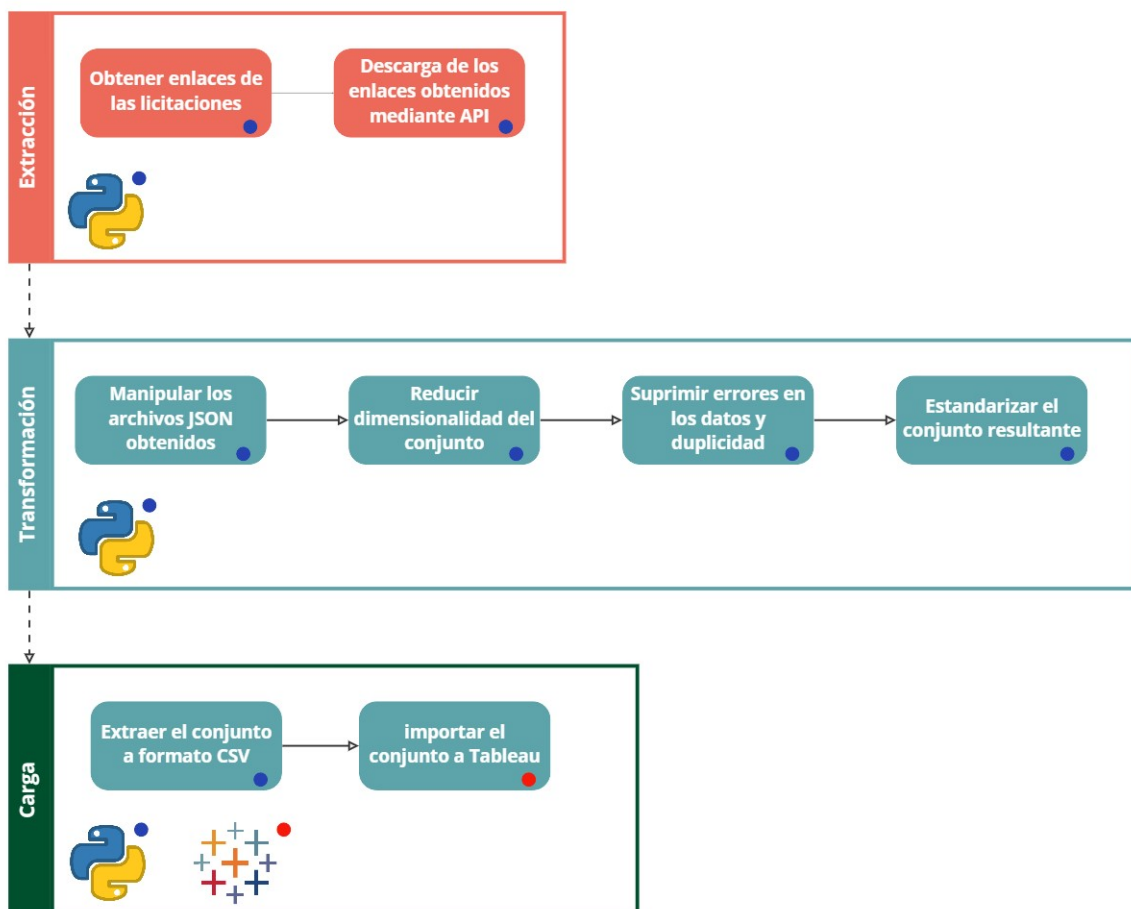


Figura 19: Primera versión del proceso ETL.

Fuente: Elaboración propia.

Como segunda incidencia en el progreso de la solución se encontró la manipulación de los archivos JSON en Python. El problema se deriva en la cantidad de archivos a transformar previa carga, en donde, aproximadamente, dos millones de archivos JSON deben ser leídos en Python antes de realizar cualquier operación. Esto es computacionalmente un cuello de botella en el proceso, lo cual puede demorar un par de horas en tan solo leer el contenido de un mes. Además, se debe considerar la complejidad del esquema analizado, el cual contiene arreglos en ciertas columnas las cuales deben ser aplanadas en un proceso que aumenta la dificultad de crear un esquema de datos consistente, al tener un rango variable de columnas entre archivos.

Para solventar el asunto sobre el tiempo utilizado en la descarga de los archivos JSON posterior a la obtención de los enlaces, se utilizaron las herramientas que proveen los sistemas operativos UNIX. La combinación utilizada fue `cat`, `tr`, `xargs` y `wget`; con estas herramientas se genera un *script* que toma los enlaces obtenidos desde la API y son enviados por UNIX *pipelines* de una herramienta a otra.

El funcionamiento del *script* utilizado es el siguiente; con `cat` se toma un mes en particular y se pasa por un *pipeline* al comando `tr`, en el cual se eliminan los carros de retorno en caso de haberlos en el archivo inicial, esto es para casos que realicen la descarga por la API mediante sistemas MS-DOS. Las entradas sanitizadas son enviadas a `xargs` con dos argumentos más el comando a repetir, el primero (-n) indica la cantidad de argumentos al comando a repetir y el segundo (-P) la cantidad máxima de procesos simultáneos. El comando a repetir en `xargs` es `wget`, con este comando se descarga la información relevante y se agregan las banderas `-nc` para no repetir la descarga del mismo archivo y `-nv` para evitar el *log* innecesario del comando.

Con este *script* se resguardan algunas funcionalidades importantes. La primera es continuar la descarga desde cualquier punto de un archivo objetivo en particular, esto dada la bandera `-nc` proporcionada a `wget`. La segunda funcionalidad importante es la cantidad de conexiones simultáneas indicada en el argumento `-P`, esta sirve para evitar problemas con el servidor sobre el límite de conexiones. Adicionalmente existe la facilidad de ejecutar múltiples veces el mismo comando desde una consola UNIX, con el fin de descargar la información de varios meses simultáneamente.

Para mitigar el problema en el tiempo de transformación de los archivos JSON, la opción más factible era integrar al diseño una base de datos que sea capaz de utilizar este tipo de documento de manera nativa. Con este dilema la alternativa más adecuada era una base de datos no relacional; dentro de ellas la que tiene el perfil buscado es MongoDB, pues tiene soporte nativo para la carga de archivos JSON y no necesita un esquema previo sobre los datos. El esquema se crea sobre la marcha al cargar los archivos lo cual remueve la complejidad del esquema dentro del modelo. Además, MongoDB contiene múltiples conectores para realizar la carga desde el lenguaje seleccionado originalmente no teniendo que cambiar demasiado el esquema, tan sólo integrar la herramienta a la solución ya planteada realizando la transformación mediante consultas en el lenguaje de MongoDB (MQL).

Al integrar dichas dos soluciones, el modelo cambió en las tecnologías utilizadas, pero no demasiado en el flujo inicial. Esto puede ser apreciado en la figura 20, en donde al integrar nuevas herramientas y tecnologías sobre el desarrollo monolítico propuesto inicialmente, se reducen los problemas mencionados con anterioridad.

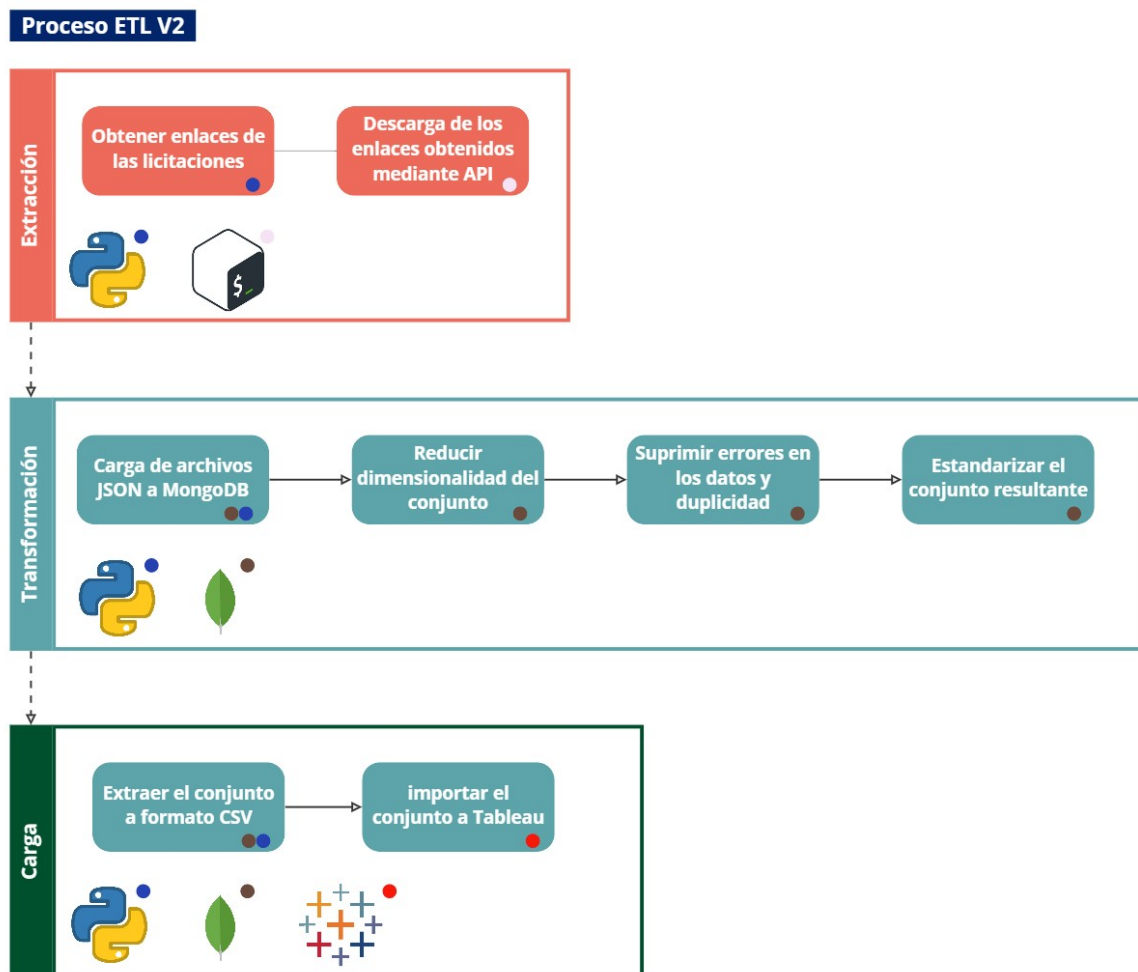


Figura 20: Segunda versión del proceso ETL.
Fuente: Elaboración propia.

3.5. Etapa 4: Visualizaciones de datos

El proceso de creación de visualizaciones se centra de primera manera, en los datos disponibles y las dimensiones que pueden ser utilizadas para brindar información a los usuarios finales. De estas variables disponibles, cuatro se sobrepone sobre las demás en su capacidad de segmentar los datos en perspectivas distintas, siendo estas últimas la localización de la licitación, el aspecto monetario, los actores relacionados y el estado de la licitación.

Estas perspectivas se utilizaron como pilares para realizar preguntas en torno a ellas, y de esta manera crear visualizaciones de datos que puedan responder las interrogantes respecto a las dimensiones en cuestión. Por lo tanto se trabajó en torno a visualizaciones enfocadas en tareas, las cuales tienen como finalidad responder preguntas específicas, que revuelven respecto a temas relacionados con las dimensiones mencionadas anteriormente.

Para cada visualización, además de seguir una tarea en particular, se propuso un contexto de uso el cual es mencionado en [Lavalle *et al.*, 2019], esto con la intención de estandarizar la creación de las visualizaciones de datos según el uso o función brindada dentro de la solución. Este contexto toma en cuenta tanto la participación del usuario en forma de interacción como de los datos para ser expresados en una infografía. El modelo a utilizar para la creación de las visualizaciones puede observarse en la figura 11, y el modelo del contexto a utilizar en las visualizaciones es representado en la figura 12.

Definiendo un poco más en detalle el contexto a utilizar, los análisis para esta solución son una descomposición sobre el diagnóstico de los datos; las visualizaciones no intentan predecir el uso de las licitaciones o su asignación en el futuro, ni un análisis prescriptivo de cómo actuar frente ciertos escenarios.

Para cada visualización o conjunto de visualizaciones mostradas en las figuras 21 a la 28, se indagan con anterioridad los siguientes atributos como contexto para la creación de visualizaciones de datos:

- **Objetivo:** La interrogante que se intenta resolver (relacionar, comparar, tendencia, etc).
- **Interacción:** Lo que el usuario tratará de realizar (filtrar, buscar detalles, resumir, etc).
- **Dimensionalidad:** Define la cantidad de variables a visualizar; este parámetro filtra las gráficas disponibles para la visualización evaluada.
- **Cardinalidad:** La cantidad de items a mostrar en pantalla; su función es similar a la dimensionalidad en el contexto de la visualización.
- **Dependencia de la métrica:** Declara el tipo de la variable a mostrar y si depende de otras variables para ser representada.

La primera variable a analizar es respecto a la localización de las licitaciones respecto al ente que las publica. En este contexto interesó resolver las dudas respecto a una región particular en función de las otras variables mencionadas. El resultado final se puede apreciar en la figura 21, en donde se encuentran cuatro visualizaciones (dos tablas y dos barras apiladas), filtros en la parte derecha del *dashboard*, navegación en la esquina inferior derecha y una imagen de la región seleccionada en la esquina superior derecha.

Con este conjunto de visualizaciones se pueden satisfacer varias interrogantes. Con las tablas superiores del panel se obtienen los entes que más participan dentro de la región seleccionada, ya sea tanto como en rol de consumidor (tabla de la izquierda) o como proveedor de servicios y bienes (tabla de la derecha). Estas tablas pueden ser modificadas por el usuario, a nivel de métrica (se puede seleccionar el desglose por cantidad de licitaciones o por el monto de ellas) y según la cantidad de participantes a mostrar en ambas tablas (parámetro 'Top N')

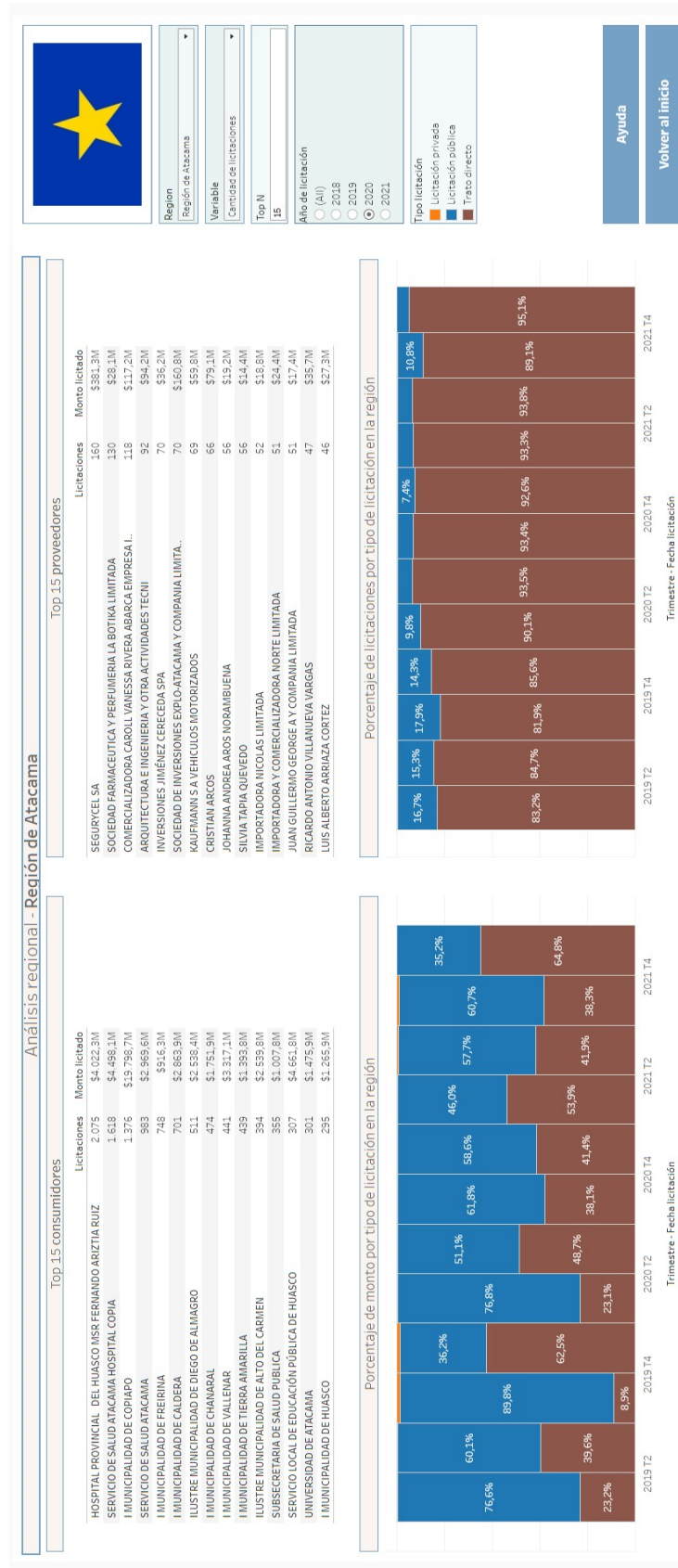


Figura 21: Conjunto de visualizaciones respecto a la ubicación de las licitaciones. Fuente: Elaboración propia.

en los controles de la derecha). Con esta información se pueden obtener actores importantes dentro de la región consultada, tanto actores del gobierno como terceros, que proveen las licitaciones.

En cuanto a las barras apiladas de la mitad inferior del *dashboard*, estas proveen la información respecto al monto y cantidad de licitaciones relativas al total durante los trimestres del conjunto de datos. Además, las barras están segmentadas en colores para identificar el tipo de licitación para cada trimestre.

En general, este *dashboard* provee la información sobre qué actores participan mayormente y la topología de las licitaciones en periodos trimestrales para una región en particular. Como punto de interés en el diseño se agrega un botón de ayuda en el apartado de navegación. Este *dashboard* incluye descripciones de los gráficos utilizados, el uso de los parámetros y filtros disponibles para el usuario en el conjunto de visualizaciones; figura 22. El *dashboard* de ayuda tiene como finalidad familiarizar más prontamente al usuario con las funcionalidades propuestas por el conjunto de visualizaciones presentadas.

Al cambiar el foco de análisis al ámbito monetario de las licitaciones, se obtiene el conjunto de visualizaciones mostrado por la figura 23. En este panel se encuentran tres visualizaciones, dos gráficos de líneas en la sección izquierda y una de barras en la parte derecha. El gráfico de líneas superior desglosa el valor promedio de una licitación por año, en donde cada línea representa a un año en específico. El gráfico inferior representa la cantidad de licitaciones por rango monetario interpretado por las líneas del gráfico. Las leyendas de ambos gráficos se encuentran en el panel de la derecha, en donde también existe un selector de tipo de licitación. El gráfico de barras denota el total de dinero utilizado en un mes en particular para el tipo de licitación seleccionado; al sobreponer el cursor en alguna barra se desglosa el monto por región, como se ejemplifica en la figura 24.

El panel, además hereda, los controles de navegación y ayuda del panel anterior, tanto para mantener la interfaz homogénea como estandarizar el diseño utilizado. Las visualizaciones representadas en las figuras 23 y 24 responden las preguntas de cuánto dinero se está utilizando (tanto a un nivel macro y regional), cuánto es el uso promedio de una licitación y cómo es la segmentación de los valores en el total de las licitaciones. Todo esto es aplicable para cada tipo de licitación con el selector en el panel de la derecha, en donde la selección es ratificada por el título ubicado en la sección superior del panel.

Utilizando otra perspectiva se encuentra el tipo de licitación utilizada como método de obtención en los servicios o bienes necesarios. Para este fin se crea el panel representado en la figura 25, que contiene cuatro tablas y un gráfico de líneas. El gráfico de líneas representa los tipos de licitaciones y su estado de aprobación, mientras que cada línea muestra la evolución de la métrica de aprobación o cancelación de un tipo de licitación por periodo trimestral.

Las tablas de la mitad inferior del panel indican la proporción del monto utilizado contra el dispuesto en las licitaciones por los n actores (tabla de la derecha para solicitantes e izquierda para proveedores) que acumulan más licitaciones o dinero licitado. Tanto la cantidad de

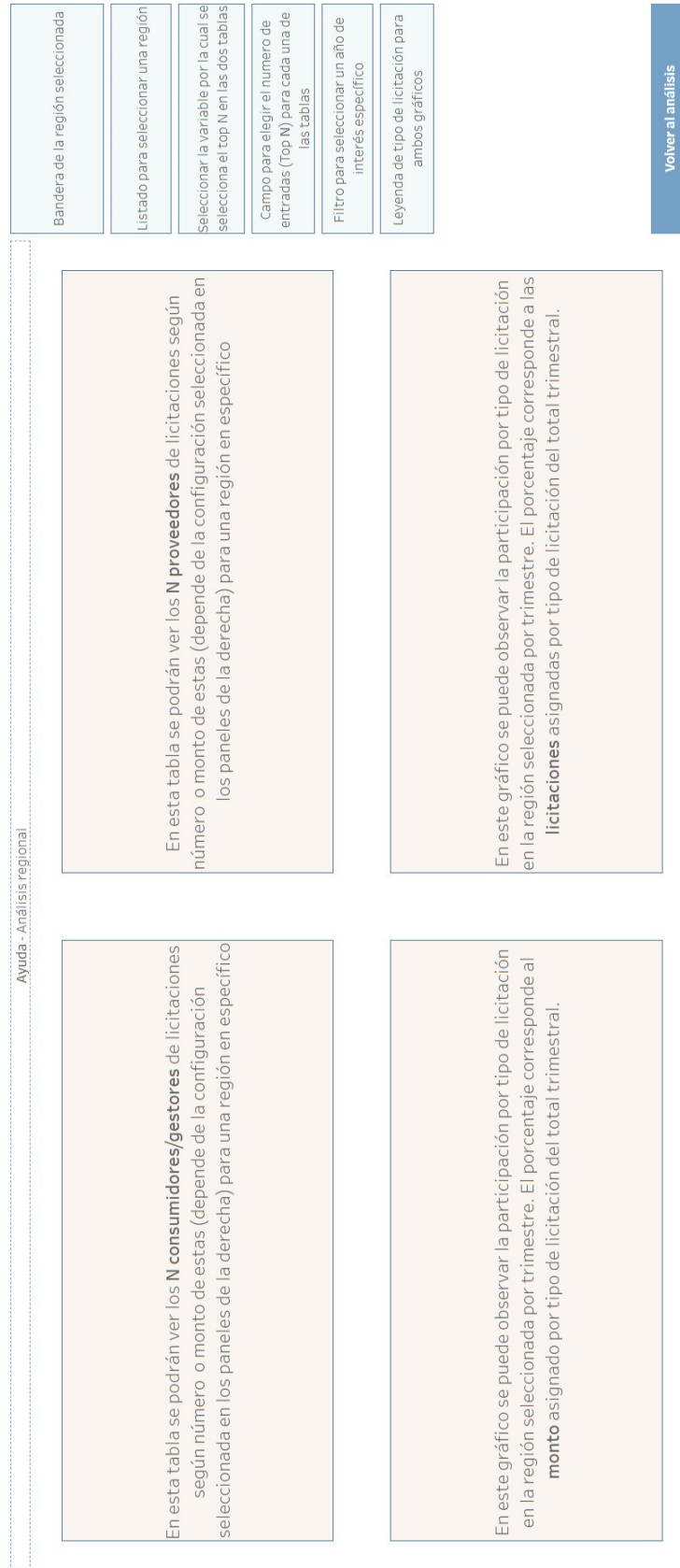


Figura 22: Panel de ayuda para el conjunto de visualizaciones regional.
Fuente: Elaboración propia.

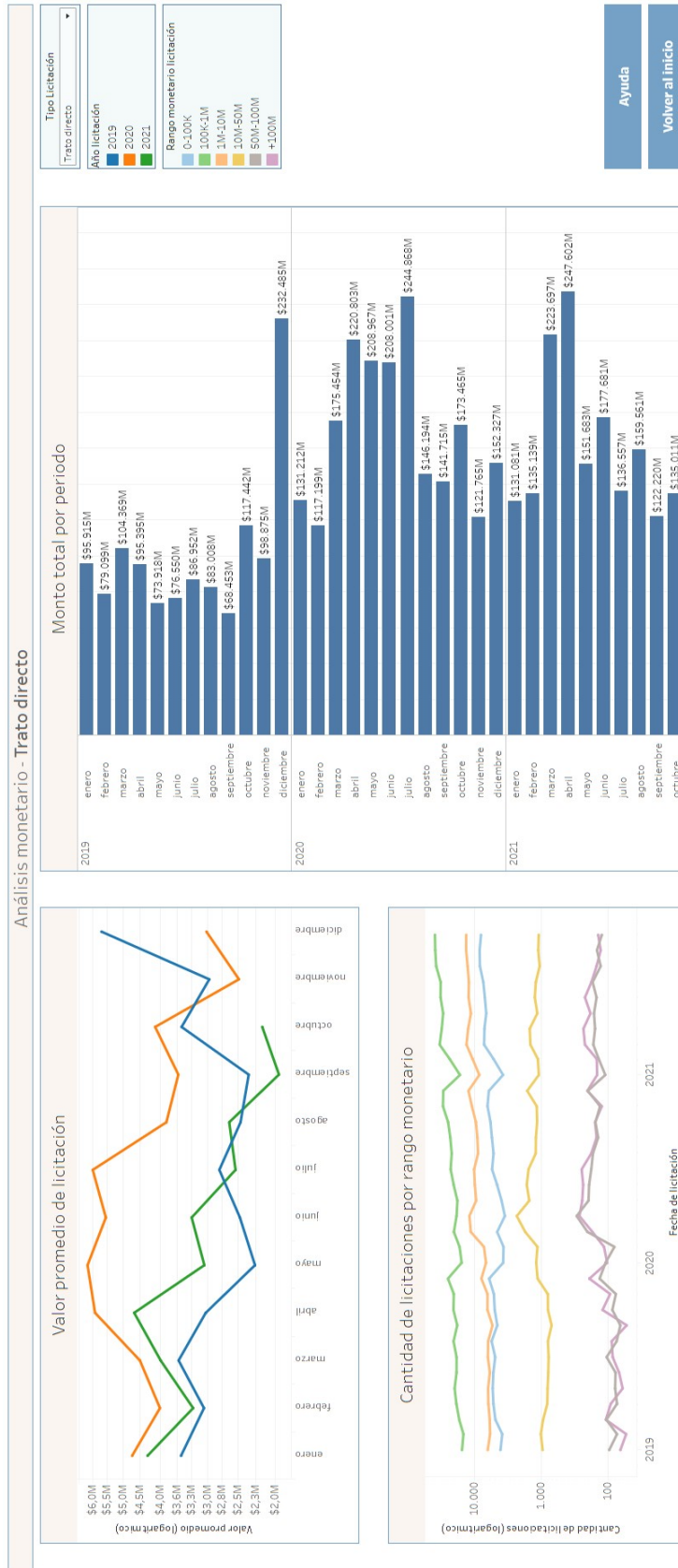


Figura 23: Conjunto de visualizaciones respecto al ámbito monetario.
Fuente: Elaboración propia.

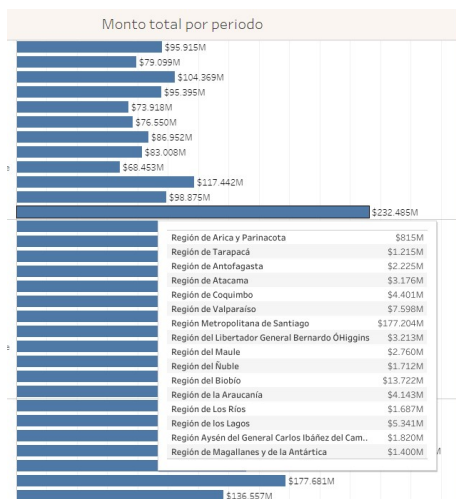


Figura 24: Desglose de un mes por región en acción.

Fuente: Elaboración propia.

elementos en la lista como la métrica discriminante son seleccionadas en los menús proporcionados a la derecha. En cuanto a las tablas de la mitad superior, una indica el ratio licitado por trimestre por tipo de licitación y la otra desglosa el porcentaje de licitaciones según la cantidad de proveedores de estas por tipo de licitación y año aludido.

Con este conjunto de gráficas, con el tipo de licitación como piedra angular, se responden las dudas en cuanto a rendimiento de los actores más utilizados, cómo evoluciona el porcentaje de aprobación en las licitaciones creadas según su tipo y cómo el *ratio* licitado sobre el propuesto evoluciona en el tiempo, indicando si los distintos procesos de licitación varían entre ellos.

Como última gran segmentación de los datos a analizar, el interés cae hacia el estado de las licitaciones en el conjunto de datos. En este panel se recopilan cuatro métricas y dos mapas de árbol como se aprecia en la figura 26. Los cuadros de métricas entregan la información del promedio de días en la etapa de postulación en el par superior, y los días promedio entre la postulación y la asignación en el par inferior. Las métricas de la derecha indican las cifras para las licitaciones privadas y en la sección de la izquierda se obtiene el valor de las licitaciones públicas (el trato directo no tiene un proceso de etapas ya que es asignado, no licitado).

En cuanto a los árboles de la parte inferior, estos representan la cantidad de licitaciones según su estado de postulación y asignación por año; en la parte derecha se encuentran los datos para las licitaciones privadas y la izquierda corresponde a las licitaciones públicas. Existe un selector de regiones en los controles de la derecha así como la leyenda de colores para la lectura de los diagrama de árboles.

Este panel argumenta sobre la situación de las licitaciones según su fecha de creación y región particular de ser necesario. Además, entrega indicadores generales de cuanto demora en promedio ciertas etapas del proceso licitatorio respecto a los datos del conjunto.

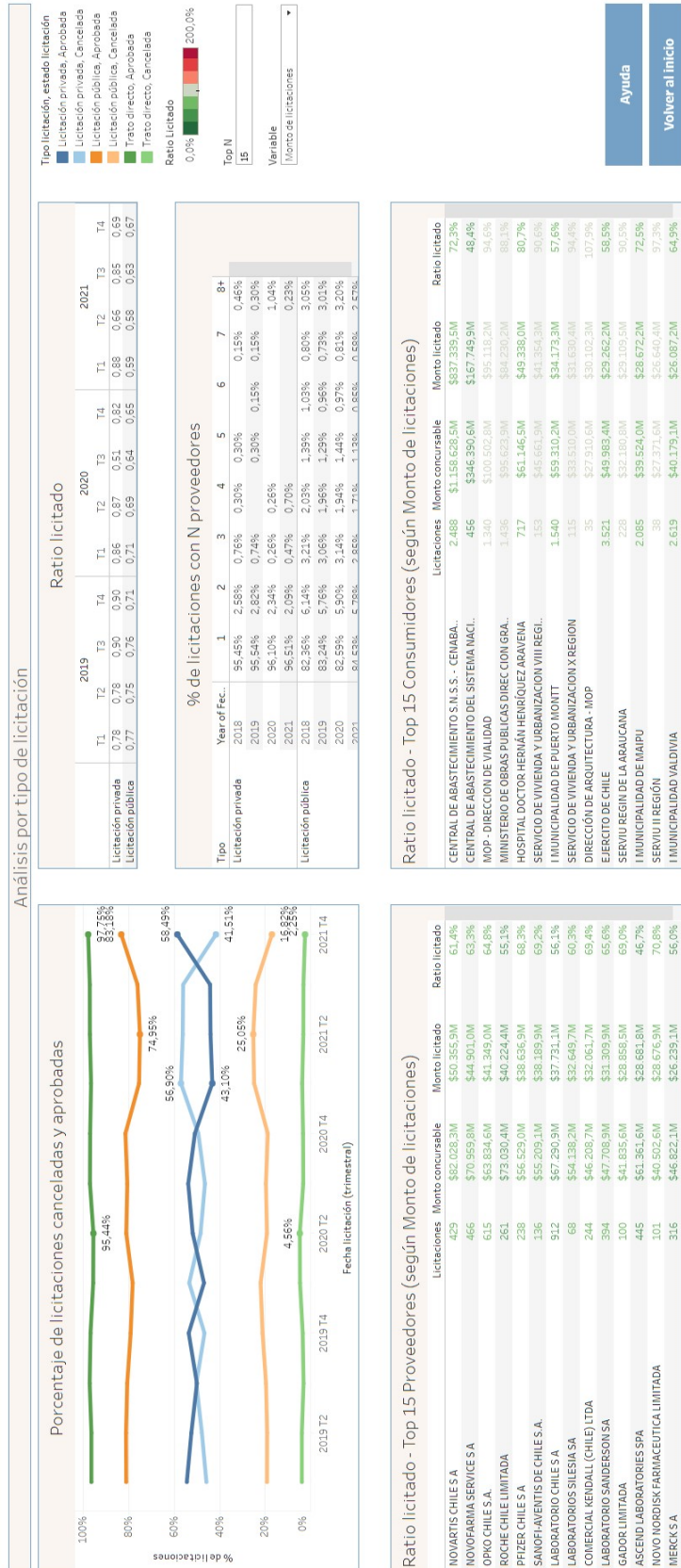


Figura 25: Conjunto de visualizaciones respecto al tipo de licitación. Fuente: Elaboración propia.



Figura 26: Conjunto de visualizaciones respecto al estado de las licitaciones.
Fuente: Elaboración propia.

Finalmente, se diseña un *dashboard* final para la exploración de los datos. Este *dashboard* considera mostrar la información sin tratamiento, la cual puede ser tanto filtrada como el usuario estime conveniente, y descargada con un botón habilitado para esto. Las funcionalidades descritas se pueden apreciar en la figura 27, donde hay un *dashboard* que cumple con una funcionalidad de exploración y descubrimiento de la información, ya sea para usuarios más avanzados ou otros que necesiten contrastar el conjunto utilizado con los datos proveídos por las fuentes oficiales.

Los distintos paneles son conectados entre sí mediante los controles de navegación. En cada *dashboard* existe un botón con la etiqueta 'Volver al inicio', que tiene como función llevar al usuario a una vista que incluye las rutas para los paneles comentados anteriormente. Además, contiene una sección para ir hacia el repositorio del conjunto de datos¹⁸ y el repositorio de código utilizado para la etapa de procesamiento de datos¹⁹, lo que puede ser apreciado en la figura 28.

¹⁸Enlace de Kaggle hacia los conjuntos de datos (<https://www.kaggle.com/franciscoabarca/datasets>)

¹⁹Enlace hacia el repositorio en github (https://github.com/Albireolmma/Memoria_INF)

Explorador de datos

Ocid	Título Licitación	Consumidor	ID Consumidor
ocds-7042ns-4371-371-LE21	ADQUISICION BANO MODULAR CONTAINER	MUNICIPALIDAD DE SAN VICENTE MUNICIPALIDAD DE SAN VICENTE	CL-MP-5287
ocds-7042ns-3810-63-LE21	SEGUINDO LLAMADO ADQUISICION DE MATERIALES	MUNICIPALIDAD DE PICHILEMU MUNICIPALIDAD DE PICHILEMU	CL-MP-4730
ocds-7042ns-2081-3-LE21	SUMINISTRO DE CHIPS PIRAC 2021 DPP COLCHAGUA	DELEGACION PRESIDENCIAL PROVINCIAL DE COLCHAGUA DELEGACION PRESIDENCIAL PROVINCIAL DE COLCHAGUA	CL-MP-3076
ocds-7042ns-1035-38-LE21	CURSO CAPACITACION	CORP NACIONAL FORESTAL VI OFICINA REGIONAL Y PROV. CACHAPOAL	CL-MP-2031
ocds-7042ns-1035-36-LE21	MANTENCIÓN SALA DE REUNIONES CONAF VI REGION	CORP NACIONAL FORESTAL CORP NACIONAL FORESTAL	CL-MP-2031
ocds-7042ns-3944-38-LE21	LICITACION INSTALACION CENTRAL TELEFONICA Y VOCEO	MUNICIPALIDAD DE PERALILLO MUNICIPALIDAD DE PERALILLO	CL-MP-4864
ocds-7042ns-2678-17-LE21	PRODUCCION EVENTOS CEREMONIA DIA DE LA EDUCACION	MUNICIPALIDAD DE MOSTAZAL MUNICIPALIDAD DE MOSTAZAL	CL-MP-3667
ocds-7042ns-3656-149-LE21	SUMINISTRO MATERIALES DE OFICINA MUNICIPALIDAD DE REQUINOA	MUNICIPALIDAD DE REQUINOA ILUSTRE MUNICIPALIDAD DE REQUINOA	CL-MP-4576
ocds-7042ns-997-130-LE21	3 CONTENEDORES DE 20 PIES USADOS PARA USO DE BODEGA PUESTO EN RANCAGUA	MINISTERIO DE OBRAS PUBLICAS DIRECCION GRAL DE OOP P D CYF DIRECCION DE VALIADU - VI REGION - DIREC.	CL-MP-1993
ocds-7042ns-4449-39-LE21	ADQUISICION DE LIBROS PARA TELETON	SERVICIO NACIONAL DEL PATRIMONIO CULTURAL SUBDIRECCION DE BIBLIOTECAS PUBLICAS	CL-MP-5365
ocds-7042ns-2271-109-LE21	ADQUISICION DE TOTEM TERMOMETRO INFRARROJO CON DISPENSADOR ALCOHOL GEL	DIRECCION GENERAL DE AERONAUTICA CIVIL UNIDAD DE COMPRAS ZONAL CENTRAL	CL-MP-3265
ocds-7042ns-1743-77-LE21	COORDINADOR MUNICIPAL CHILE CREEE CONTIGO	ILUSTRE MUNICIPALIDAD DE LUTICHE ILUSTRE MUNICIPALIDAD DE LUTICHE	CL-MP-2739
ocds-7042ns-1402-18-LE21	INSUMOS Y ACCESORIOS COMPUTACIONALES AREA DE INFORMATICA GENDARMERIA REGION DE OHIGGINS	DIRECCION GENERAL DE GENDARMERIA DE CHIL REGION DE OHIGGINS	CL-MP-2398
ocds-7042ns-4768-24-LE21	REPARACION CAMION RECOLECTOR IRRCT81.	ILUSTRE MUNICIPALIDAD DE LA ESTRELLA ILUSTRE MUNICIPALIDAD DE LA ESTRELLA	CL-MP-5683
ocds-7042ns-4768-24-LE21	SERVICIOS BANQUETERIA	ILUSTRE MUNICIPALIDAD DE LA ESTRELLA ILUSTRE MUNICIPALIDAD DE LA ESTRELLA	CL-MP-5683
ocds-7042ns-2401-113-LE21	COCTAIL CEREMONIA POR AÑOS DE SERVICIO	MUNICIPALIDAD DE RANCAGUA MUNICIPALIDAD DE RANCAGUA	CL-MP-3393
ocds-7042ns-7307-36-41-LE21	JUGUETES PARA CELEBRACION COMUNAL NAVIDAD 2021	ILUSTRE MUNICIPALIDAD DE PAREDES ADQUISICIONES	CL-MP-730736
ocds-7042ns-5296-23-LE21	SUMA. MED. COMP. ORALES. INYECTABLES Y OTROS SEPT.	SERVICIO DE SALUD HOSPITAL DE RENGO UNIDAD DE FARMACIA	CL-MP-6239
ocds-7042ns-3729-78-LE21	JUGUETES DE NAVIDAD 2021	MUNICIPALIDAD DE CHIMBARONGO SECLA	CL-MP-4649
ocds-7042ns-3656-148-LE21	SUMINISTRO DE TINTAS Y TONER PARA LAS DISTINTAS DEPENDENCIAS MUNICIPALES/REQUINOA	MUNICIPALIDAD DE REQUINOA ILUSTRE MUNICIPALIDAD DE REQUINOA	CL-MP-4576
ocds-7042ns-4769-18-LE21	COMBUSTIBLE Y LUBRICANTES CESFAM LA ESTRELLA 2021	ILUSTRE MUNICIPALIDAD DE LA ESTRELLA DEPARTAMENTO SALUD	CL-MP-5684
ocds-7042ns-3656-147-LE21	ADQUISICION DE ADORNOS PARA ARBOL NAVIDEÑO. REQUINOA 2021	MUNICIPALIDAD DE REQUINOA ILUSTRE MUNICIPALIDAD DE REQUINOA	CL-MP-4576
ocds-7042ns-3478-20-LE21	INSUMOS VETERINARIOS PARA PLANI MASCOTA PROTEGIDA	ILUSTRE MUNICIPALIDAD DE QUINTA DE TILCOCO ILUSTRE MUNICIPALIDAD DE QUINTA DE TILCOCO	CL-MP-4399
ocds-7042ns-2401-109-LE21	CAJAS DE COLACION	MUNICIPALIDAD DE RANCAGUA MUNICIPALIDAD DE RANCAGUA	CL-MP-3393
ocds-7042ns-1590-29-LE21	CONSULTORIA PARA LA ACTUALIZACION DEL DIAGNOSTICO ORGANIZACION DE TRANSVERSALIZACION DEL ENFOQUE DE GENERO EN...	SUBSECRETARIA DEL MINISTERIO DE AGRICULT SUBSECRETARIA DE AGRICULTURA	CL-MP-2586
ocds-7042ns-7307-36-40-LE21	INSUMOS PARA PROGRAMA AUTOCONSUMO GESTION 2021	ILUSTRE MUNICIPALIDAD DE PAREDES ADQUISICIONES	CL-MP-730736
ocds-7042ns-3838-113-LE21	ADQUISICION DE MATERIAL DE FEBRERIA	MUNICIPALIDAD DE SANTA CRUZ MUNICIPALIDAD DE SANTA CRUZ	CL-MP-4758
ocds-7042ns-3810-60-LE21	PRODUCCION ACTIVIDADES CELEBRANDO LA VIDA RURAL	MUNICIPALIDAD DE PICHILEMU MUNICIPALIDAD DE PICHILEMU	CL-MP-4730
ocds-7042ns-3810-59-LE21	ARRIENDO STANDERIA EMPRENDIMIENTO ARTESANAL NOVI	MUNICIPALIDAD DE PICHILEMU MUNICIPALIDAD DE PICHILEMU	CL-MP-4730
ocds-7042ns-3764-44-LE21	AUDITORIA EXTERNA PARA MUNICIPALIDAD NANCAGUA	MUNICIPALIDAD DE NANCAGUA MUNICIPALIDAD DE NANCAGUA	CL-MP-4684
ocds-7042ns-3656-146-LE21	REMODELACION OFICINA DE ALCALDIA DE REQUINOA	MUNICIPALIDAD DE REQUINOA MUNICIPALIDAD DE REQUINOA	CL-MP-4576
ocds-7042ns-3656-145-LE21	CONTRATACION DE PRODUCCION TECNICA Y ARTISTICA PARA LA CELEBRACION DE ANIVERSARIO REQUINOA 2021	MUNICIPALIDAD DE REQUINOA ILUSTRE MUNICIPALIDAD DE REQUINOA	CL-MP-4576
ocds-7042ns-2564-45-LE21	DIA DEL FUNCIONARIO MUNICIPAL	MUNICIPALIDAD DE CHEPICA MUNICIPALIDAD DE CHEPICA	CL-MP-3555
ocds-7042ns-837-119-LE21	SERVICIO RESTAURACION LIBRERO MUSEO DR. CARLOS YBAR	SERVICIO MEDICO LEGAL SERVICIO MEDICO LEGAL	CL-MP-1833
ocds-7042ns-837-117-LE21	ADQUISICION VITRINAS PARA MUSEO DR. CARLOS YBAR	MUNICIPALIDAD DE LAS CABRAS MUNICIPALIDAD DE LAS CABRAS	CL-MP-5725
ocds-7042ns-4810-20-LE21	SUMINISTRO DE LIMPIEZA DE OSMAS SÉPTICAS	ILUSTRE MUNICIPALIDAD DE GRANEROS ILUSTRE MUNICIPALIDAD DE GRANEROS	CL-MP-5578
ocds-7042ns-4862-27-LE21	INSTALACION DE UN ESTANQUE SEMIRREMOLOQUE	GOBIERNO REGIONAL VI REGION DEL LI BERTADOR B O'HIGGINS GOBIERNO REGIONAL VI REGION DEL LI BERTA...	CL-MP-5290
ocds-7042ns-4374-12-LE21	ADQUISICION DE UN ESTANQUE SEMIRREMOLOQUE	MUNICIPALIDAD DE NANCAGUA MUNICIPALIDAD DE NANCAGUA	CL-MP-4684
ocds-7042ns-3764-45-LE21	II LLAMADO MOBILIARIO CENTRO DE LA MUJER SERNAMEG	MUNICIPALIDAD DE REQUINOA ILUSTRE MUNICIPALIDAD DE REQUINOA	CL-MP-4576
ocds-7042ns-3656-143-LE21	SUMINISTRO SERVICIO DE TRANSPORTE PARA EL PROGRAMA CHILE CREEE CONTIGO REQUINOA	MUNICIPALIDAD DE REQUINOA ILUSTRE MUNICIPALIDAD DE REQUINOA	CL-MP-4576
ocds-7042ns-1050283-26-LE21	SERVICIO DE INSTALACION Y MONTAJE	MUNICIPALIDAD DE SAN VICENTE MUNICIPALIDAD DE SAN VICENTE	CL-MP-1050283
ocds-7042ns-565-40-LE21	ADQ. DE MAQUINAS DE EJERCICIOS PROGRAMA PROMOCION	ILUSTRE MUNICIPALIDAD DE SAN FERNANDO ILUSTRE MUNICIPALIDAD DE SAN FERNANDO	CL-MP-1561
ocds-7042ns-3954-11-LE21	CAMARAS DE VIGILANCIA PARA PERALILLO	MUNICIPALIDAD DE PERALILLO SECRETARIA DE PLANIFICACION COMUNAL	CL-MP-4874
ocds-7042ns-3374-27-LE21	ADQUISICION DE MATERIALES DE CONSTRUCCION PARA REPARACION DE CUADRAS DEL REGIMIENTO Nº 19 "COLCHAGUA"	COMANDO DE APOYO A LA FUERZA REGIMIENTO Nº 19 "COLCHAGUA"	CL-MP-4295
ocds-7042ns-997-126-LE21	3.649 M3 DE BASE GRANULAR CBP-100%, PARA CAMINO RPH-804 "CRUCE H-784- H-804 (LAS QUISCAS)" SECTOR KM 0,000 A KM...	MINISTERIO DE OBRAS PUBLICAS DIRECCION GRAL DE OOP P D CYF DIRECCION DE VALIADU - VI REGION - DIREC.	CL-MP-1993
ocds-7042ns-3996-86-LE21	INSUMOS DE ASEO	MUNICIPALIDAD DE COLTAUCO DEPARTAMENTO DE SALUD	CL-MP-4916

Búsqueda por OCID

Búsqueda por título

Búsqueda por ID Consum...

Region

Región de Libertador...

Tipo Licitación

(All)

Licitación privada

Licitación pública

Trato directo

Fecha Licitación

01-01-2020 29-10-2021

Exportar datos

Ayuda

Volver al inicio

Figura 27: Explorador de datos. Fuente: Elaboración propia.



Figura 28: Menú principal.
Fuente: Elaboración propia.

CAPÍTULO 4

VALIDACIÓN DE LA SOLUCIÓN

En este capítulo se define un marco para validar que las funcionalidades propuestas son cumplidas por la solución diseñada. Además, se pone en práctica el marco ideado, en donde se analizan los resultados obtenidos.

4.1. Métodos de validación

Algunos métodos y prácticas sobre la validación de las visualizaciones fueron expuestas en la sección 2.7. De estos trabajos es posible concluir que los métodos para validar la funcionalidad de una gráfica, se basan en verificar que la visualización proveída satisface las acciones necesarias para obtener respuestas a preguntas sobre los datos, y si el usuario es capaz de obtener un análisis útil sobre los datos propuestos y como estos han sido representados por el diseño de la visualización seleccionada.

Usando la literatura expuesta en el marco conceptual es posible utilizar una mezcla de dos tipos de validación, el primero corresponde a la utilización de gráficas particulares para cierto tipo de tareas [Saket *et al.*, 2018], hay datos conocidos sobre precisión, tiempo y preferencia para los pares tarea-gráfico, por lo cual para una cierta finalidad es posible considerar estas evaluaciones y crear paneles informativos con una ventaja empírica sin tener que iterar desde un comienzo con usuarios.

Dicho método simplifica la elección de gráficos a seleccionar dada la finalidad o tarea que el usuario debe satisfacer, como ejemplo para la búsqueda de correlaciones los gráficos de líneas y dispersión son los que obtienen mejores métricas de precisión y tiempo utilizado; sin embargo la preferencia de los usuarios tiende a ser mayor para la versión de líneas sobre un gráfico de dispersión. La utilización de esta información empírica tiene sus defectos, puesto que no estima visualizaciones interactivas por lo cual deben ser consideradas como directrices principales pero no una receta específica de como crear la solución. Un resumen de las relaciones entre gráficos y tareas según métrica de importancia es representado por la figura 10.

El segundo método es aún más pragmático y de cierta forma complementario al primero, pues propone distintas prácticas con los usuarios finales para corroborar que las tareas a resolver por las visualizaciones son satisfechas en donde se constatan tres (la mayoría de veces dos de las tres) aristas: precisión, realismo o generalidad [Carpendale, 2008]. Estas aristas se ven en mayor o menor proporción según el método utilizado como se demuestra en la figura 13.

Cada tipo de verificación en este método se basa en alguna de las aristas propuestas, en este

caso es de interés conocer de manera precisa que las funcionalidades se cumplen según el diseño propuesto. Además se discute sobre la validez cualitativa sobre los datos cuantitativos de ciertos métodos; como lo cualitativo es altamente interpretativo según quien dirija el experimento, se dará importancia a los métodos cualitativos que sean medibles y repetibles en el tiempo. Además, como contraste el primer método con su naturaleza empírica, a continuación considera la opinión de los usuarios dentro de las métricas aportando de cierta manera una variable cuantitativa (dado que hace referencia a la percepción de los usuarios por sobre las métricas de tiempo y precisión).

Dentro de los métodos cualitativos deben ser considerados los que sean menos susceptibles al ruido de los participantes, dado que los recursos para disponer de grandes muestras no está al alcance. Por lo tanto, se opta por un estudio de campo, en donde se enfrentará a los usuarios a resolver una cantidad de tareas que corresponden a las funcionalidades de las gráficas o elementos del diseño creado, para simular el caso de uso más realista posible en donde no habrá intervención sobre las acciones de los usuarios.

En conjunto ambos métodos deberían abarcar los puntos de validación que requieren las gráficas diseñadas en la sección anterior, para validar que las funciones con las cuales fueron diseñadas en mente sean satisfechas. El estudio de campo será realizado con una muestra de 17 personas, la demografía de esta muestra corresponde a individuos entre 23 y 32 años de los cuales 14 fueron a instituciones de educación superior y el resto ya finalizaron su educación media. El grupo es homogéneo respecto a sus habilidades analíticas y conocimientos sobre tecnologías de la información.

4.2. Criterios utilizados

Para el primer método los criterios utilizados son dado según la tarea a representar gráficamente. De estas tareas se consideran principalmente diez:

- **Localizar:** el usuario es capaz de encontrar algo que ya conoce y lo indica señalándolo o describiéndolo.
- **Identificar:** similar a localizar, pero el usuario es capaz de localizar un elemento sin conocerlo previamente.
- **Distinguir:** el usuario es capaz de distinguir entre los diferentes elementos de los conjuntos de datos.
- **Categorizar:** el usuario es capaz de identificar divisiones de categorías de artículos mediante objetos visuales.
- **Agrupar:** el usuario es capaz de encontrar un grupo de artículos (si los hay). Pueden ser identificados por artículos que están vinculados o agrupados.

- **Distribuir:** el usuario es capaz de caracterizar la distribución de los valores de un atributo en el conjunto de los datos casos.
- **Ordenar:** el usuario puede indicar el orden de los elementos mostrados según alguna métrica.
- **Comparar:** el usuario puede comparar entidades similares entidades o conjuntos de elementos diferentes.
- **Asociar:** el usuario es capaz de establecer relaciones entre los elementos mostrados.
- **Correlacionar:** dados dos atributos del conjunto de datos, el usuario es capaz de determinar si existe una relación entre los valores de esos atributos.

Definido el par de tarea y métrica que se desee estimar se utilizará una gráfica adecuada a lo referenciado en la literatura [Saket *et al.*, 2018]. Esto considera la opción empírica de la evaluación al utilizar la literatura existente como primer acercamiento a una validación de los resultados existente sobre precisión, tiempo y preferencias de los usuarios ya interrogados con anterioridad. Este criterio tiene la finalidad de evaluar la toma de decisiones sobre el diseño utilizado, al tomar en cuenta las tareas y acciones como variables en conjunto al conocimiento previo.

Respecto al estudio de campo, este evalúa múltiples facetas de los paneles generados, pero principalmente con énfasis en las tareas utilizadas como criterio del método anterior. Estas tareas serán evaluadas por individuo y por panel, el énfasis es cuantitativo, en donde se mide el tiempo en realizar la tarea y la cantidad de intentos como medida de precisión (con 7 intentos por pregunta). El medio para la recopilación de información es mediante una entrevista directa con los individuos, donde se les brinda el control absoluto para interactuar con los paneles y se les solicita ejecutar tareas para evaluar los criterios mencionados sin ningún tipo de obstrucción.

Las tareas evaluadas fueron las siguientes:

- **Conjunto de visualizaciones respecto a la ubicación de las licitaciones** (figura 21)
 - (Localizar) ¿Puedes identificar algún dato o información antes de asimilar la información?
 - (Identificar) ¿Puedes identificar al proveedor con mayor cantidad de licitaciones o monto licitado?
 - (Distinguir) ¿Puedes distinguir entre los tipos de licitaciones en los gráficos disponibles?
 - (Categorizar) ¿Puedes identificar las categorías de los tipos de licitaciones en los gráficos disponibles?

- (Agrupar) ¿Puedes buscar un trimestre en particular para una métrica de los gráficos?
 - (Distribuir) ¿Crees que los gráficos demuestran como se distribuye un tipo de licitación en el tiempo?
 - (Ordenar) ¿Puedes identificar un orden en los datos de las tablas?
 - (Comparar) ¿Puedes comparar la información entre tipos de licitación?
 - (Asociar) ¿Puedes identificar las relaciones entre los datos representados?
 - (Correlacionar) ¿Puedes describir una relación entre la fecha y el tipo de licitación?
- **Conjunto de visualizaciones respecto al ámbito monetario (figura 23)**
- (Localizar) ¿Puedes identificar algún dato o información antes de asimilar la información?
 - (Identificar) ¿Puedes identificar el periodo con el mayor monto asociado?
 - (Distinguir) ¿Puedes distinguir entre los años de licitaciones en los gráficos disponibles?
 - (Categorizar) ¿Puedes identificar las categorías de los segmentos monetarios en los gráficos disponibles?
 - (Agrupar) ¿Puedes buscar un año en particular para los montos asociados?
 - (Distribuir) ¿Crees que los gráficos demuestran como se distribuyen los rangos en el tiempo?
 - (Ordenar) ¿Puedes identificar un orden en los datos?
 - (Comparar) ¿Puedes comparar la información de dos años distintos?
 - (Asociar) ¿Puedes identificar las relaciones entre los datos representados?
 - (Correlacionar) ¿Puedes describir una relación entre la fecha y el valor promedio de las licitaciones?

■ **Conjunto de visualizaciones respecto al tipo de licitación** (figura 25)

- (Localizar) ¿Puedes identificar algún dato o información antes de asimilar la información?
- (Identificar) ¿Puedes identificar el tipo de licitación con mayor o peor porcentaje de aprobación?
- (Distinguir) ¿Puedes distinguir entre monto concursable y monto licitado?
- (Categorizar) ¿Puedes identificar los distintos tipos de licitaciones y estados?
- (Agrupar) ¿Puedes buscar un trimestre en particular para obtener su ratio licitado?
- (Distribuir) ¿Crees que los gráficos demuestran como se distribuyen los rangos en el tiempo?
- (Ordenar) ¿Puedes identificar un orden en los datos?
- (Comparar) ¿Puedes comparar la aprobación de distintos tipos de licitaciones?
- (Asociar) ¿Puedes identificar las relaciones entre los datos representados?
- (Correlacionar) ¿Puedes describir una relación entre el tipo de licitación?

■ **Conjunto de visualizaciones respecto al estado de las licitaciones** (figura 26)

- (Localizar) ¿Puedes identificar algún dato o información antes de asimilar la información?
- (Identificar) ¿Puedes identificar la duración en una etapa para un tipo de licitación?
- (Distinguir) ¿Puedes distinguir entre los tipos de licitaciones?
- (Categorizar) ¿Puedes identificar los distintos tipos de licitaciones y estados?
- (Agrupar) ¿Puedes buscar un estado en particular?
- (Distribuir) ¿Crees que los gráficos demuestran como se distribuyen las licitaciones en el tiempo?
- (Ordenar) ¿Puedes identificar un orden en los datos?
- (Comparar) ¿Puedes comparar la información entre tipo de licitaciones y estados similares?
- (Asociar) ¿Puedes identificar las relaciones entre los datos representados?
- (Correlacionar) ¿Puedes describir una relación entre el tipo de licitación?

En base al trabajo de observar cómo los usuarios utilizaron la solución, se obtienen las métricas para validar lo diseñado. En la siguiente sección se analizan los datos obtenidos y la evaluación de las gráficas de manera cuantitativa, respecto al uso de los individuos que fueron parte del estudio de campo.

4.3. Resultados obtenidos

Los resultados obtenidos del estudio de campo pueden ser observados en la figura 29. En esta se presentan los datos de los 17 individuos, en donde se aprecian los datos obtenidos por panel analizado y tarea solicitada, así como el tiempo utilizado en realizar la tarea y la cantidad de intentos necesarios para llegar a una respuesta correcta.

En la figura 30 se aprecian los datos de precisión para el estudio en cuestión. Se puede analizar de esta tabla, que el porcentaje de precisión entre paneles es muy similar, la variación se puede dar entre tareas, pero en general cada tarea promedia por sobre un 80 % de precisión en los paneles estudiados. Este indicador se puede interpretar que, para obtener una respuesta satisfactoria, en una tarea específica el usuario toma en promedio tan solo un error para resolver la tarea solicitada.

Respecto al tiempo utilizado en las tareas se puede observar la dispersión en la figura 31, muy similar al análisis en el criterio de precisión, hay variaciones en las tareas entre los distintos paneles, pero en general los tiempos utilizados tienden a ser homogéneos. Esta regularidad se ve representada en las distribuciones al agrupar los datos de los paneles o por tarea. En general un usuario para finalizar una tarea demora entre 26 a 37 segundos, con una mediana de 32 segundos entre los distintos paneles.

Con los datos analizados se puede estimar que los porcentajes de precisión y tiempo utilizados por los usuarios en realizar las tareas propuestas, con el diseño presentado son satisfactorios y dentro de los márgenes de error para una solución que tiene como objetivo entregar información a una gran cantidad de público de la manera más general posible.

Con estos resultados se puede comentar que la solución propuesta abarca los objetivos que se han presentado, y los concluye con la validación de las métricas obtenidas de los usuarios al utilizar la solución diseñada con los objetivos en mente.

Panel	Estado de licitación	Item	SUB-001		SUB-002		SUB-003		SUB-004		SUB-005		SUB-006		SUB-007		SUB-008		SUB-009		SUB-010		SUB-011		SUB-012		SUB-013		SUB-014		SUB-015		SUB-016		SUB-017							
			Inte...	Tie...	Inte...	Tie...	Inte...	Tie...	Inte...	Tie...	Inte...	Tie...	Inte...	Tie...	Inte...	Tie...	Inte...	Tie...	Inte...	Tie...	Inte...	Tie...	Inte...	Tie...	Inte...	Tie...	Inte...	Tie...	Inte...	Tie...	Inte...	Tie...	Inte...	Tie...	Inte...	Tie...						
Monetario	Agrupar	Asociar	0	27	1	30	0	33	0	35	1	31	1	22	0	36	0	29	0	29	0	36	0	29	0	29	0	29	0	37	1	35	2	41	1	23	2	36				
		Categorizar	2	36	0	24	2	33	1	39	0	37	0	27	0	23	0	37	0	28	0	34	1	34	1	28	0	37	0	32	1	38	0	43	2	39	0	35				
		Comparar	0	33	2	20	0	31	1	42	2	35	2	49	1	40	4	28	0	40	1	31	0	30	0	30	0	30	0	18	1	38	1	38	0	42	1	29	0	29		
		Correlacion..	0	19	2	41	0	36	2	26	1	35	1	31	1	19	1	19	1	34	1	24	1	24	1	24	1	28	1	28	1	24	2	34	1	34	1	31	1	31		
		Distinguir	2	26	3	38	2	23	1	26	0	32	2	40	2	29	0	28	1	34	1	24	1	24	1	26	1	27	2	27	2	25	1	26	2	46	0	20	0	20		
		Distribuir	0	32	0	30	1	37	1	24	1	31	1	28	0	26	2	35	2	24	1	38	1	35	1	35	1	37	4	37	4	46	0	33	0	27	0	35	0	35		
		Identificar	1	26	1	45	0	42	0	31	1	36	2	49	1	24	2	34	1	47	1	21	1	47	1	21	1	31	1	31	1	29	0	25	1	37	2	46	0	46		
		Localizar	1	21	2	43	1	28	0	32	1	24	2	41	1	22	3	37	0	37	0	38	0	31	2	46	0	22	2	25	1	41	0	36	0	32	4	27	0	33		
		Ordenar	2	33	0	35	0	29	2	32	1	34	2	31	1	22	3	31	0	43	2	31	0	30	2	31	0	22	1	47	0	23	1	23	0	26	1	33	0	33		
		Regional	Agrupar	Asociar	1	29	0	28	2	37	1	25	0	26	1	36	2	23	0	28	1	24	0	22	0	22	0	18	0	18	0	31	0	25	1	46	0	26	1	35	0	35
				Categorizar	2	34	2	34	0	23	0	30	2	42	1	38	2	31	1	26	1	39	2	43	2	40	1	40	1	40	1	40	2	42	0	21	1	26	4	37	0	33
				Comparar	1	41	1	34	2	31	0	37	1	22	1	29	1	30	1	30	2	47	2	38	2	40	1	28	2	28	2	52	2	44	0	23	0	30	1	35	0	35
				Correlacion..	1	34	1	38	2	43	0	20	1	40	2	37	1	42	1	30	3	46	2	32	0	28	2	47	2	42	2	33	5	43	2	33	0	30	1	39	0	39
				Distinguir	0	26	0	30	1	46	1	32	2	22	2	40	0	32	0	31	1	29	0	26	0	26	0	22	2	22	2	35	1	28	1	37	2	25	2	34	0	34
				Distribuir	2	25	2	35	2	33	0	24	2	34	0	21	2	39	0	25	2	42	2	34	1	34	0	31	2	35	1	30	2	32	2	36	2	36	2	43	0	43
				Identificar	0	26	0	27	1	39	0	33	1	30	0	33	2	33	2	33	2	33	2	30	1	32	0	42	0	33	1	49	1	26	1	23	0	26	1	31	0	31
				Localizar	0	24	1	24	0	26	1	37	2	48	1	27	2	36	2	39	3	38	0	25	2	24	1	31	2	44	1	44	1	33	2	35	1	33	0	29	0	29
Ordenar	0			47	1	36	1	30	2	33	2	45	1	27	2	51	1	29	1	31	1	35	0	31	1	35	0	28	1	32	1	21	4	54	4	31	1	22	0	46		
Tipo de licitación	Agrupar			Asociar	0	26	2	28	0	30	2	43	0	39	2	29	0	24	0	30	1	35	2	36	0	33	1	27	1	40	0	28	0	18	2	42	1	34	0	35	0	35
				Categorizar	2	39	2	34	2	49	0	32	1	29	1	21	1	22	0	22	0	40	1	31	4	27	0	25	0	25	0	26	2	40	0	26	2	28	2	40	0	40
				Comparar	0	26	2	40	1	32	2	28	2	40	2	41	1	29	0	30	0	23	2	41	0	22	4	38	0	20	2	39	0	33	1	36	2	29	0	35	0	35
				Correlacion..	1	25	2	47	2	50	2	21	2	44	0	26	0	27	1	26	1	33	1	34	0	23	0	22	4	38	0	34	1	31	0	24	1	34	0	33	0	33
				Distinguir	2	36	0	26	2	41	2	26	3	50	1	34	1	34	1	39	1	24	2	37	2	37	2	24	2	39	2	32	2	36	0	24	0	28	1	26	0	26
				Distribuir	1	38	1	43	2	36	1	31	2	30	0	34	1	24	3	35	1	32	0	27	2	42	0	24	0	24	0	18	2	31	0	27	0	41	5	38	0	38
				Identificar	2	38	0	31	2	21	2	26	2	35	2	32	0	43	2	32	1	30	0	29	2	37	2	23	2	35	2	44	0	37	2	35	2	35	2	35	0	35
				Localizar	0	30	2	34	0	37	2	33	2	34	0	34	1	31	1	40	1	24	0	35	2	35	2	35	2	28	2	20	0	36	1	32	1	34	1	43	0	43
		Ordenar	2	38	2	32	1	25	1	24	0	28	2	35	0	26	1	35	0	28	0	30	1	35	0	42	1	37	1	37	1	29	2	31	4	28	2	34	0	35		
		Tipo de licitación	Agrupar	Asociar	2	27	1	38	1	31	1	26	2	33	1	25	2	29	2	30	2	36	3	39	0	28	0	26	0	26	1	38	0	31	1	30	1	34	0	35	0	35
				Categorizar	1	26	2	37	1	28	2	34	1	37	1	28	1	21	1	38	2	34	0	30	2	20	1	48	0	23	1	27	1	32	1	41	0	31	1	37	0	37
				Comparar	0	24	0	17	2	36	0	34	2	42	0	33	1	41	2	32	2	37	4	36	1	31	0	18	1	34	0	41	1	30	1	34	0	31	0	36	0	36
				Correlacion..	1	31	1	32	0	34	0	20	0	24	1	24	1	30	1	40	1	28	2	26	1	35	0	30	2	39	1	36	0	26	2	36	2	36	2	45	0	45
				Distinguir	1	36	2	19	3	25	2	27	2	35	3	30	0	27	2	41	2	40	2	36	2	31	2	42	2	42	2	42	0	27	0	23	2	43	2	23	0	23
				Distribuir	0	26	1	37	0	26	1	37	1	36	0	35	2	33	2	43	1	21	1	29	2	30	2	46	2	38	2	34	1	34	0	30	1	34	0	30	1	35
				Identificar	1	36	0	29	2	35	1	32	0	21	2	34	0	39	0	29	2	20	2	26	2	20	2	37	2	39	2	21	0	37	0	35	2	43	2	42	0	42
				Localizar	0	37	0	28	0	32	0	23	1	28	0	32	2	29	3	21	2	36	0	22	1	37	2	27	2	39	2	39	2	36	0	31	2	21	1	27	0	27
Ordenar	0			26	2	31	1	38	1	45	2	28	2	37	2	30	1	31	1	25	1	38	1	33	1	34	1	34	0	33	2	28	0	26	0	32	2	32	0	32		

Figura 29: Datos tabulados obtenidos del estudio de campo.
Fuente: Elaboración propia.

Panel

Item	Regional	Monetario	Tipo de licitación	Estado de licitación	Total general
Localizar	84.9%	82.4%	84.9%	82.4%	83.6%
Identificar	79.0%	89.1%	83.2%	85.7%	84.2%
Distinguir	79.8%	87.4%	75.6%	79.8%	80.7%
Categorizar	86.6%	88.2%	83.2%	89.9%	87.0%
Agrupar	88.2%	91.6%	83.2%	88.2%	87.8%
Distribuir	82.4%	79.8%	84.0%	86.6%	83.2%
Ordenar	83.2%	80.7%	84.0%	83.2%	82.8%
Comparar	82.4%	83.2%	85.7%	85.7%	84.2%
Asociar	83.2%	79.8%	87.4%	86.6%	84.2%
Correlacionar	86.6%	77.3%	86.6%	82.4%	83.2%
Total general	83.6%	84.0%	83.8%	85.0%	84.1%

Figura 30: Datos sobre precisión en el estudio de campo.
Fuente: Elaboración propia.

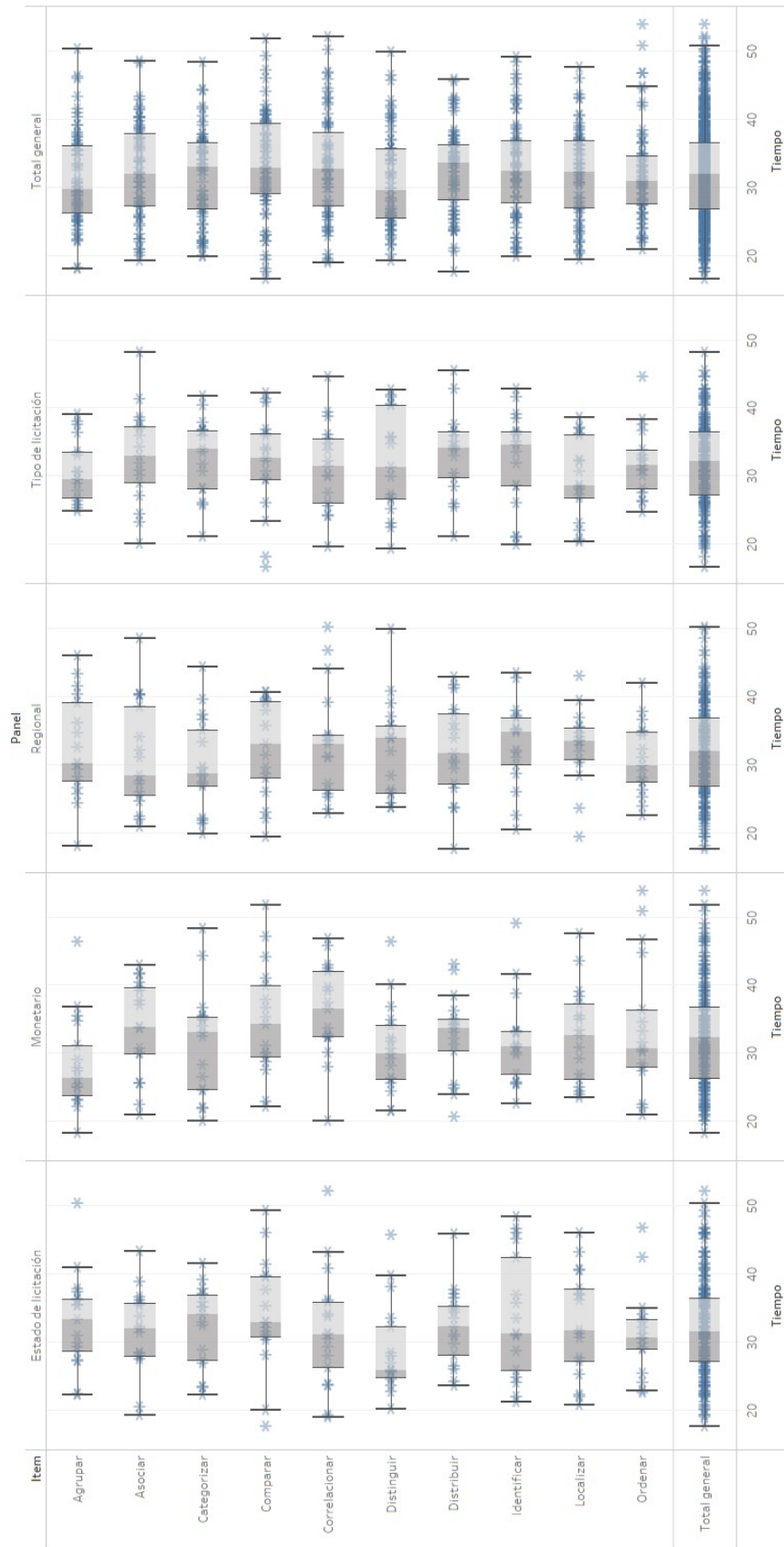


Figura 31: Datos sobre tiempo en el estudio de campo.
Fuente: Elaboración propia.

CONCLUSIONES

Dentro de lo realizado en este trabajo se obtiene una variedad de enseñanzas para futuros trabajos en esta área, tanto a un nivel de organización personal y mejores prácticas para coordinar un proyecto de esta magnitud a un nivel personal, como también el marco de trabajo propuesto y utilizado en el desarrollo de la solución al problema presentado. Es importante lo aprendido y experimentado en cuanto a organización personal, lo que no tiene una mayor importancia al comienzo del proyecto, pero se va haciendo imperante la necesidad de organizarse a medida que incrementan las tareas por resolver y para cumplir con los objetivos propuestos. De esta manera la utilización de un método para disponer de los recursos disponibles se hace indispensable, además ayuda a documentar sobre lo trabajado y permite realizar estimaciones de cuánto tiempo necesita el trabajo restante por realizar o evaluaciones para entregas futuras.

Otro punto importante a destacar dentro de las experiencias, es el uso de un marco de trabajo propuesto en la literatura y llevado al ámbito práctico. Si bien el marco de trabajo expuesto es utilizado, este es retocado para cumplir ciertas condiciones o limitaciones inherentes del trabajo. Lo principal de esta experiencia es que las metodologías en la teoría deben ser vistas como consejos en el diseño de la solución por sobre un conjunto de reglas a seguir, sin ningún criterio por el contexto del problema que se desea resolver. Aún así, el marco de trabajo expuesto es fácilmente reproducible, con el problema de que la validación de las soluciones tiende a utilizar una gran parte del tiempo empleado dado que se deben verificar todas las tareas a realizar por el usuario final. Si bien esto puede ser reducido a un esquema más nuclear y tan solo realizar corroboración en ciertas tareas de interés de quien está diseñando una solución para un problema más acotado, pero por la generalidad de los objetivos este no fue el caso.

Adicionalmente, en lo relacionado en la ejecución de la solución, la mejor manera encontrada para probar la factibilidad del diseño propuesto es realizar múltiples pruebas de conceptos en la fase temprana del proyecto. Se desaprovechó tiempo en crear una solución con una cantidad limitada de datos en un inicio y no simulando la carga completa, debiendo cambiar partes de la arquitectura originalmente ideada inicialmente en medio del desarrollo. Esto pudo ser fácilmente remediado al realizar pruebas de manera constante simulando el ambiente final de uso.

En cuanto a una práctica rescatable dentro de las decisiones tomadas, el uso de MongoDB para hacer uso de archivos JSON en una base de datos documental fue muy acertada. Este cambio facilitó el flujo de trabajo y se integró muy bien en los procesos ETL, dado que la herramienta en cuestión es relativamente fácil de aprender a utilizar y no requiere un gran esfuerzo o tiempo para dominarla. En otras palabras, aunque la herramienta puede tener algunas características o funcionalidades avanzadas, todavía es relativamente fácil de aprender a utilizar y no impide que se pueda realizar un buen uso de ella. Otra decisión acertada en la arquitectura es el uso de Tableau, herramienta que permitió disponer la solución crea-

da en un panel directamente en la web sin un costo adicional; lo único a considerar es que si se necesita conectar a fuentes distintas a archivos de textos (bases de datos o almacenamiento en la nube), se debe utilizar una licencia la cual es entregada y renovada de manera anual por Tableau (actualmente Salesforce tiene propiedad sobre la herramienta) con tan solo acreditar que el usuario es estudiante.

Los resultados obtenidos del proceso de validación definido y ejecutado con los usuarios que probaron la solución diseñada en un estudio de campo, fueron favorables de tal modo que se puede afirmar con confianza que las tareas para las cuales fueron diseñados los paneles son cumplidas a cabalidad. De esta experiencia híbrida de utilizar literatura como guía inicial en la creación de las gráficas dependiendo de la tarea a realizar y su posterior validación con un ejercicio con usuarios reduce de cierta forma la toma de decisiones del desarrollador de las visualizaciones, esto puede ser tanto positivo como negativo dependiendo del contexto, pero de igual manera hay un elemento cuantitativo de qué tan efectivos son los *dashboards* creados al finalizar tanto el proceso de desarrollo y validación. En esta instancia no hubo un proceso iterativo de desarrollo, puesto que los resultados de la validación confirmaron la funcionalidad de las gráficas creadas, por lo que futuros trabajos utilizando este marco de trabajo pueden variar si no se obtienen las métricas y resultados en primera instancia de la validación.

Lo obtenido por el estudio no tiende a demostrar demasiada desviación en su distribución, esto a pesar de no ser una muestra de gran magnitud (17 usuarios en este caso). Es recomendable en caso de obtener resultados cuantitativos muy variados verificar que las preguntas o tareas solicitadas a los usuarios sean comunicadas en un lenguaje claro, en lo posible crear un set de prueba previo al estudio a realizar. Se nota en ciertos usuarios que en el primer conjunto de preguntas no tienen muy claro algunas de ellas y sus acciones, pero luego de una iteración estos tienden a mejorar su precisión y tiempo de respuesta. En estos casos es recomendable volver a realizar la prueba en otra ocasión y analizar a qué se deben las variaciones para proponer soluciones según el contexto.

Otro punto interesante a considerar, es el integrar métodos cualitativos a la validación, ya sea integrar el contexto en la estimación del diseño de las gráficas o la evaluación de la experiencia de usuario a un nivel similar en la arquitectura en el diseño de las páginas web. Aunque altamente subjetivo, estos métodos pueden brindar perspectivas interesantes o que llanamente no son aparentes al desarrollador, y que pueden contribuir de gran manera al producto final al añadir distintos criterios de terceros a la solución.

Como aplicaciones del trabajo efectuado, la más importante es la fundación de un marco de trabajo para el conjunto de datos, el cual no tiene mucha documentación más que la propuesta por el organismo que define las directrices de OCDS. Más allá de una definición de la estructura de datos, el crear un método para simplificar la extracción y una solución para tratar los datos son un aporte para los interesados en seguir una rama de análisis respecto a los datos del conjunto referenciado.

Otra contribución es a los usuarios directos de los *dashboards* desarrollados, ya que estos

plasman el uso directo de los recursos fiscales, los cuales siempre serán de interés a la población afligida, tanto con un fin informativo como uno fiscalizador. El poder facilitar información que no está cómodamente a la disposición de los usuarios en un modo organizado y resumido tiene una gran utilidad para el ciudadano que desea conocerla, pero no tiene los conocimientos para poder obtenerla.

Una población informada de los proyectos en progreso y el estado de estos, es de un alto impacto en una sociedad que puede acceder a información de manera inmediata con los avances tecnológicos de las últimas décadas. Esta rama es de mucho interés por el público general, el ser humano tiene una condición por querer conocer y entender lo que lo rodea, por lo tanto es de suma atención la exposición de datos que provean contexto o explicaciones a ciertas acciones. Por lo tanto es recomendable seguir esta rama, ya sea con este conjunto de datos u otro nuevo, para incluir a un público general a nuevos análisis y oportunidades de conocimiento. Los datos son tan útiles mientras sean mostrados de manera sincera y efectiva a quien tenga el provecho de acercarse a conocerlos, pero es nuestro cometido poder proveer estas instancias de acercar o disponibilizar la información que no siempre está disponible al público general, mediante estas instancias para crear un aporte a ciertos sectores que no lo tendrían de otra manera.

Las posibles aristas para la continuación de este trabajo son variadas. La más inmediata, es la automatización del proceso ETL para mantener la fuente de datos del extracto actualizada; con esta funcionalidad se podría crear reportería que esté siempre actualizada al día. Esta arista tiene diferentes posibles acercamientos, como por ejemplo, el uso de la nube para realizar una arquitectura que refleje produzca los datos en un repositorio público y habilitando a los usuarios para crear sus análisis de manera personalizada.

Adicionalmente, el aspecto de la difusión de la solución es de interés para este trabajo, el cual crea una solución pero dentro de su desarrollo no considera un plan de difusión. Si bien lo desarrollado se encuentra disponible en su totalidad, la difusión en distintos medios es de vital importancia dado el alcance que necesita o por lo cual fue creada la solución en un inicio. Sería de interés un trabajo el cual analice los métodos de difusión más importantes, o bien crear una herramienta para difusión de proyectos de esta categoría, tanto para encontrar a posibles usuarios de prueba para validar los diseños de las soluciones o para que los productos finales tomen tracción con el público de interés.

Otra línea de ampliación para el trabajo actual es utilizar lo ya creado y evaluar bajo criterios distintos, esto con la finalidad de verificar si el marco de trabajo es validado para ser tratado como guía válida para futuros trabajos probada de una manera más general, y no tan de nicho como fue en este trabajo. Con este acercamiento se puede crear un aporte a futuros trabajos que necesiten un método para validar sus gráficas o paneles de información y una metodología que haya sido contrastada bajo múltiples parámetros y contextos para confiar en ella.

Una idea para avanzar en esta dirección es integrar y documentar el uso cuantitativo y cualitativo en el trabajo de validación, así como crear las evaluaciones de los *dashboards* bajo

un estándar que pueda ser ajustado a distintos tipos de trabajos, logrando el agnosticismo del contexto en el desarrollo se puede crear un marco de trabajo aplicable a distintas disciplinas que necesiten evaluar sus gráficos y la efectividad de estos frente a una amalgama de usuarios.

El aporte de la carrera en esta propuesta está inherente en muchas de las disciplinas utilizadas, desde la formación en herramientas de inteligencia de negocio, hasta el uso de lenguajes de programación para la automatización de las tareas en el ETL. La carrera de ingeniería civil informática proveyó la formación de las habilidades necesarias, para diseñar e implementar una solución, la cual independiente de los posibles errores iniciales en la fase de diseño, fue posible concretarla mediante el conocimiento especializado sobre distintos aspectos en tecnologías relevantes. Este conocimiento especializado, que es optado por cada estudiante de manera autónoma, permite al alumnado integrarse en una rama del vasto ecosistema informático e indagar en profundidad. El tener la posibilidad de ahondar una derivación específica en nuestro rubro, permite que el interés se focalice en temas interesantes que van más allá de cumplir con simples asignaturas y ramos impartidos, en donde reúnen tanto a profesores como alumnos con una inclinación particular.

Este ambiente es altamente productivo en los últimos años de carrera, sobre todo considerando que en el rubro laboral de la informática existe una superespecialización de los perfiles que son requeridos, y el ambiente mencionado ayuda en gran parte a un estudiante a identificar su interés en todo el ecosistema disponible.

REFERENCIAS BIBLIOGRÁFICAS

- [Anderson y Carmichael, 2016] Anderson, D. J. y Carmichael, A. (2016). *Essential kanban condensed*. Blue Hole Press.
- [Aparicio y Costa, 2015] Aparicio, M. y Costa, C. J. (2015). Data visualization. *Communication design quarterly review*, 3(1):7-11.
- [Caldarola y Rinaldi, 2017] Caldarola, E. G. y Rinaldi, A. M. (2017). Big data visualization tools: a survey. *Research Gate*.
- [Carpendale, 2008] Carpendale, S. (2008). Evaluating information visualizations. En *Information visualization*, pp. 19-45. Springer.
- [Charalamby y Andrea, 2020] Charalamby, M. y Andrea, P. (2020). Aproximaciones a las causas del estallido social de Chile, octubre 2019. B.S. thesis, PUCE-Quito.
- [Chen, 2005] Chen, C. (2005). Top 10 unsolved information visualization problems. *IEEE computer graphics and applications*, 25(4):12-16.
- [Chen et al., 2008] Chen, M., Ebert, D., Hagen, H., Laramée, R. S., Van Liere, R., Ma, K.-L., Ribarsky, W., Scheuermann, G., y Silver, D. (2008). Data, information, and knowledge in visualization. *IEEE computer graphics and applications*, 29(1):12-19.
- [Chen et al., 2016] Chen, M., Feixas, M., Viola, I., Bardera, A., Shen, H.-W., y Sbert, M. (2016). *Information theory tools for visualization*. AK Peters/CRC Press.
- [Chittaro, 2006] Chittaro, L. (2006). Visualizing information on mobile devices. *Computer*, 39(3):40-45.
- [Despa, 2014] Despa, M. L. (2014). Comparative study on software development methodologies. *Database Systems Journal*, 5(3):37-56.
- [Ellis y Dix, 2006] Ellis, G. y Dix, A. (2006). An explorative analysis of user evaluation studies in information visualisation. En *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*, pp. 1-7.
- [Federico et al., 2016] Federico, P., Amor-Amorós, A., y Miksch, S. (2016). A nested workflow model for visual analytics design and validation. En *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, pp. 104-111.
- [Few, 2006] Few, S. (2006). *Information dashboard design: The effective visual communication of data*. O'Reilly Media, Inc.
- [Fowler et al., 2001] Fowler, Martin and Highsmith, Jim and others (2001). The agile manifesto. *Software development*, 9(8):28-35.

- [Galdames Paredes, 2015] Galdames Paredes, F. J. (2015). Análisis econométrico del abastecimiento de medicamentos en el mercado público vía licitaciones, convenios marco y cenabast.
- [Gallager, 1968] Gallager, R. G. (1968). *Information theory and reliable communication*, volumen 588. Springer.
- [Gemignani, 2010] Gemignani, Z. (2010). Better know a visualization: Parallel coordinates.
- [Güell, 2019] Güell, P. (2019). El estallido social de Chile: piezas para un rompecabezas. *Mensaje*, 68(685):8-14.
- [Henderson et al., 2009] Henderson, D. A., Courtney, B., Inglesby, T. V., Toner, E., y Nuzzo, J. B. (2009). Public health and medical responses to the 1957-58 influenza pandemic. *Biosecurity and bioterrorism: biodefense strategy, practice, and science*, 7(3):265-273.
- [Ho, 2012] Ho, K. (2012). Scatter plot and correlation coefficient. *Anaesthesia and intensive care*, 40(4):730.
- [Hossain et al., 2009] Hossain, E., Babar, M. A., y Paik, H.-y. (2009). Using scrum in global software development: a systematic literature review. En *2009 Fourth IEEE International Conference on Global Software Engineering*, pp. 175-184. Ieee.
- [Iversen y Gergen, 2012] Iversen, G. R. y Gergen, M. (2012). *Statistics: The conceptual approach*. Springer Science & Business Media.
- [Jeff, 2006] Jeff, C. (2006). Multi-level pie charts. URL: <http://www.neoformix.com/2006/MultiLevelPieChart.html>.
- [Kistler et al., 2013] Kistler, S. J., Evergreen, S., y Azzam, T. (2013). Toolography. *New Directions for Evaluation*, 2013(139):73-84.
- [Kosara, 2007] Kosara, R. (2007). Visualization criticism-the missing link between information visualization and art. En *2007 11th International Conference Information Visualization (IV'07)*, pp. 631-636. IEEE.
- [Lavalle et al., 2019] Lavalle, A., Maté, A., Trujillo, J., y Rizzi, S. (2019). Visualization requirements for business intelligence analytics: a goal-based, iterative framework. En *2019 IEEE 27th International Requirements Engineering Conference (RE)*, pp. 109-119. IEEE.
- [Liu et al., 2018] Liu, J., Tang, T., Wang, W., Xu, B., Kong, X., y Xia, F. (2018). A survey of scholarly data visualization. *Ieee Access*, 6:19205-19221.
- [Loeliger y McCullough, 2012] Loeliger, J. y McCullough, M. (2012). *Version Control with Git: Powerful tools and techniques for collaborative software development*. "O'Reilly Media, Inc."

- [López-Martínez *et al.*, 2016] López-Martínez, J., Juárez-Ramírez, R., Huertas, C., Jiménez, S., y Guerra-García, C. (2016). Problems in the adoption of agile-scrum methodologies: A systematic literature review. En *2016 4th international conference in software engineering research and innovation (conisoft)*, pp. 141–148. IEEE.
- [Madariaga Torres, 2021] Madariaga Torres, S. M. (2021). Machine learning para predecir volúmenes operacionales de las líneas de negocio de cenabast.
- [Magazine, 2007] Magazine, S. (2007). Data visualizations: modern approaches. *Online*) <http://www.smashingmazine.com>. *com/2007/08/02/data-visualisation-modern-approaches.html* (Accessed 23/04/10).
- [Marchionini, 2008] Marchionini, G. (2008). Human–information interaction research and development. *Library & Information Science Research*, 30(3):165–174.
- [Nonaka, 2008] Nonaka, I. (2008). The knowledge-creating company (harvard business review classics). *Harvard Business School Pr*, ISBN.
- [Plaisant, 2004] Plaisant, C. (2004). The challenge of information visualization evaluation. En *Proceedings of the working conference on Advanced visual interfaces*, pp. 109–116.
- [Saket *et al.*, 2018] Saket, B., Endert, A., y Demiralp, Ç. (2018). Task-based effectiveness of basic visualizations. *IEEE transactions on visualization and computer graphics*, 25(7):2505–2512.
- [Schwaber, 1997] Schwaber, K. (1997). Scrum development process. En *Business object design and implementation*, pp. 117–134. Springer.
- [Shen-Hsieh y Schindl, 2002] Shen-Hsieh, A. y Schindl, M. (2002). Data visualization for strategic decision making. En *Case studies of the CHI2002| AIGA Experience Design FORUM*, pp. 1–17.
- [Shneiderman y Plaisant, 1998] Shneiderman, B. y Plaisant, C. (1998). Treemaps for space-constrained visualization of hierarchies.
- [Stevens, 1946] Stevens, S. (1946). On the theory of scales of measurement.
- [Velásquez Pizarro, 2022] Velásquez Pizarro, A. I. (2022). Análisis estadístico multivariante de licitaciones de compra del gobierno de Chile.
- [Wang y Shen, 2011] Wang, C. y Shen, H.-W. (2011). Information theory in scientific visualization. *Entropy*, 13(1):254–273.
- [Wang *et al.*, 2009] Wang, X., Jeong, D. H., Dou, W., Lee, S.-w., Ribarsky, W., y Chang, R. (2009). Defining and applying knowledge conversion processes to a visual analytics system. *Computers & Graphics*, 33(5):616–623.
- [Ware, 2019] Ware, C. (2019). *Information visualization: perception for design*. Morgan Kaufmann.

[Zhu, 2007] Zhu, Y. (2007). Measuring effective data visualization. En *International Symposium on Visual Computing*, pp. 652–661. Springer.