



Universidad Politécnica
de Madrid



**Escuela Técnica Superior de
Ingenieros Informáticos**

Grado en Ingeniería Informática

Trabajo Fin de Grado

Migración y comparativa de un cuadro de
mandos Power BI sobre datos clínicos;
AccuGest

Autor: Andrés Bravo Francos

Tutor(a): Sergio Paraíso Medina

Madrid, Junio 2023

Este Trabajo Fin de Grado se ha depositado en la ETSI Informáticos de la Universidad Politécnica de Madrid para su defensa.

Trabajo Fin de Grado

Grado en Ingeniería Informática

*Título: Migración y comparativa de un cuadro de mandos Power bi sobre
datos clínicos; AccuGest*

Marzo 2023

Autor: Andrés Bravo Francos

Tutor: Sergio Paraíso Medina

Lenguajes y sistemas informáticos e ingeniería de software

ETS Ingenieros Informáticos

Universidad Politécnica de Madrid

Agradecimientos

Me gustaría hacer un agradecimiento especial a mi familia que me han apoyado en todo momento durante la realización de este trabajo de fin de grado que supone el fin de mi etapa académica, en especial a mi abuela que estoy seguro que estaría muy orgullosa. También quiero agradecer a todos mis amigos que siempre me han apoyado para continuar el esfuerzo que ha supuesto la realización de este proyecto conciliándolo con el trabajo. Por último, quiero agradecer a mi tutor Sergio, por ofrecerme este proyecto conociendo mis conocimientos existentes en la visualización de datos.

Resumen

Este trabajo pretende aportar una mejora y una nueva opción visual al ya existente sistema AccuGest que permite el estudio en mayor profundidad de los datos arrojados por la realización de pruebas TPNI (Test Prenatal No Invasivo).

En un mundo en el que los datos se consideran “el nuevo petróleo” por el valor que aportan y la necesidad que se tiene de “refinarlos” que para que tengan utilidad se enmarca este proyecto. Durante este proyecto vamos a hacer un estudio de las actuales tecnologías que existen en la actualidad, pero vamos a centrarnos principalmente en las que vamos a usar para mejorar el actual cuadro de mandos y para migrar el sistema AccuGest, en concreto mediante Rstudio/Shiny y Power BI. En la sección de mejoras en el actual sistema, se irán añadiendo, diferenciando estas entre mejoras visuales y funcionales, que aportarán tanto una mejor presentación visual como nuevas funcionalidades en estudio de los datos

A lo largo de esta memoria se redactará el proceso que se ha seguido para cumplir los objetivos que más adelante se exponen. Para ello se expondrán los motivos y la intención por los que se toman ciertas decisiones de diseño y de implementación y como se abarcan estas decisiones. Tanto el sistema original como el nuevo estará ligado al uso de filtros para que el usuario pueda interactuar con los datos. Esto le permitirá a este la alteración de las diferentes visualizaciones mediante la aplicación de un filtro temporal en el que poder seleccionar intervalos de tiempo concretos y otro filtro de procedencias con el que poder seleccionar sobre que Centros Hospitalarios se quiere consultar las visualizaciones

Durante la creación del nuevo panel se tratará siempre de aplicar las buenas prácticas que una correcta visualización de datos en un panel debe cumplir y que permitirán futuras incorporaciones de mejoras en el panel de la forma más sencilla. También será una prioridad que el nuevo panel evolucione hacia una carga y transformación de los datos automatizada, de manera que el mantenimiento del nuevo sistema sea mínimo y se actualice periódicamente.

Abstract

This work aims to provide an improvement and a new visual option to the existing AccuGest system, which allows for a deeper study of the data generated by the Non-Invasive Prenatal Testing (TPNI).

In a world where data is considered the "new oil" due to its value and the need to "refine" it to obtain utility. During this project, we will study the current technologies available, but we will mainly focus on those we will use to enhance the current dashboard and migrate the AccuGest system, specifically using RStudio/Shiny and Power BI. In the section on improvements to the current system, visual and functional enhancements will be added, providing both improved visual presentation and new functionalities for data analysis.

Throughout this report, the process followed to achieve the stated objectives will be described. The reasons and intentions behind the design and implementation choices will be outlined, along with the approach taken to address them. Both the original and the new system will be linked to the use of filters, allowing users to interact with the data. This will enable users to alter different visualizations by applying a temporal filter to select specific time intervals, as well as a filter based on hospital centers to choose which visualizations to consult.

During the creation of the new dashboard, the best practices that a data visualization in a dashboard will always be applied, allowing for easy future improvements. It will also be a priority for the new dashboard to evolve towards automated data loading and transformation, minimizing system maintenance and ensuring periodic updates.

Tabla de contenidos

| | |
|--|-----------|
| 1 INTRODUCCIÓN | 10 |
| 1.1 Planteamiento del problema | 10 |
| 1.2 Solución propuesta | 10 |
| 1.3 Objetivos | 11 |
| 1.4 Planificación del trabajo | 11 |
| 2 ESTADO DEL ARTE Y TECNOLOGÍAS EMPLEADAS | 13 |
| 2.1 Herramientas de visualización | 13 |
| 2.1.1 RStudio | 13 |
| 2.1.2 Shiny | 14 |
| 2.1.3 Power BI | 15 |
| 2.2 Sistemas de almacenamiento | 16 |
| 2.2.1 MySQL | 16 |
| 2.3 Fuentes de datos | 17 |
| 3 REQUISITOS Y DISEÑO | 18 |
| 3.1 Requisitos | 18 |
| 3.1.1 Requisitos comunes de cada cuadro de mandos | 18 |
| 3.1.1.1 Filtro temporal | 19 |
| 3.1.1.2 Filtro por procedencia | 19 |
| 3.1.1.3 Tarjetas resumen resultados | 20 |
| 3.1.2 Requisitos propios de cada cuadro de mandos | 21 |
| 3.1.2.1 Requisitos propios AccuGest RStudio/Shiny | 21 |
| 3.1.2.2 Requisitos propios AccuGest PowerBI | 21 |
| 3.2 Diseño | 22 |
| 3.2.1 Diseño mejoras en AccuGest en Rstudio/Shiny | 22 |
| 3.2.2 Diseño panel AccuGest en PowerBi | 24 |
| 3.2.2.1 Diseño migración de datos | 24 |
| 3.2.2.2 Diseño nuevo panel AccuGest | 25 |
| 3.2.2.3 Diseño del flujo de datos | 27 |
| 4 IMPLEMENTACIÓN | 29 |
| 4.1 Implementación mejoras en AccuGest en Rstudio/Shiny | 29 |
| 4.1.1 Modificación carga de datos | 29 |
| 4.1.2 Implementación mejoras en script App.R | 29 |
| 4.1.2.1 Implementación mejoras objeto UI | 30 |
| 4.1.2.2 Implementación mejoras objeto Server | 31 |
| 4.2 Implementación panel AccuGest en Power Bi | 35 |
| 4.2.1 Determinar origen de los datos | 35 |
| 4.2.2 Carga y transformación de los datos en tablas | 35 |
| 4.2.2.1 Carga de datos y transformación en tabla principal | 36 |
| 4.2.2.2 Creación de tablas auxiliares | 39 |
| 4.2.3 Creación nuevo panel AccuGest | 41 |
| 4.2.3.1 Creación de tablas cuadros de texto y botones | 41 |
| 4.2.3.2 Creación de tablas de dimensiones | 42 |
| 4.2.3.3 Administración de relaciones entre tablas | 44 |
| 4.2.3.4 Implementación del panel de visualizaciones según el diseño | 46 |
| 4.2.3.4.1 Implementación panel Inicio | 46 |
| 4.2.3.4.2 Implementación panel Tabla Resultados | 47 |
| 4.2.3.4.3 Implementación panel Gráficos de Distribución Temporal | 49 |
| 5 RESULTADOS Y CONCLUSIONES | 52 |
| 5.1 Resultados de las mejoras en el sistema AccuGest | 52 |
| 5.2 Nuevo panel en Power BI | 53 |
| 5.3 Comparativa entre los dos paneles | 56 |
| 5.3.1 Comparativa de visualización y usabilidad entre los dos paneles | 56 |
| 5.3.2 Análisis comparativo de las herramientas desde la perspectiva del desarrollo | 56 |
| 6 CATALOGACIÓN DENTRO DE LOS OBJETIVOS DE DESARROLLO SOSTENIBLE 2030 | 59 |

| | |
|---------------------------------------|----|
| 7 CONCLUSIONES Y LÍNEAS FUTURAS ----- | 60 |
| 8 BIBLIOGRAFÍA ----- | 60 |

Tabla de ilustraciones

| | |
|---|----|
| Ilustración 1: Diagrama de GANTT del trabajo..... | 12 |
| Ilustración 2 Controlador deslizante de doble entrada para delimitar de manera dinámica el rango temporal..... | 19 |
| Ilustración 3 Controlador de selección de procedencias que se estudiarán en el cuadro de mandos..... | 20 |
| Ilustración 4 Tarjetas creadas a modo resumen de los resultados obtenidos en función del filtro temporal..... | 21 |
| Ilustración 5 Sistema AccuGest desde el que se parte para aplicar mejoras..... | 23 |
| Ilustración 6 Diseño sistema AccuGest desde el que se parte para aplicar mejoras..... | 24 |
| Ilustración 7 Diseño panel inicio sistema AccuGest PowerBI | 25 |
| Ilustración 8 Diseño tooltip en el panel inicio sistema AccuGest PowerBI..... | 26 |
| Ilustración 9 Diseño paneles “Tabla Resultados” “Gráficos de Distribución temporal” sistema AccuGest PowerBI..... | 27 |
| Ilustración 10 Diseño del flujo de datos AccuGest PowerBI..... | 28 |
| Ilustración 11 Creación dataframe “lista_anomalia” y “ lista_anomalia_filtrada”..... | 29 |
| Ilustración 12 Creación dataframe “tiempo_res”..... | 29 |
| Ilustración 13 Ajuste de tamaños en filtro temporal..... | 30 |
| Ilustración 14 Implementación nueva pestaña indicaciones..... | 30 |
| Ilustración 15 Ajuste tamaño tarjetas resumen..... | 31 |
| Ilustración 16 Implementación del output de la pestaña “Anomalías”..... | 32 |
| Ilustración 17 Implementación del output de la pestaña “Indicaciones”..... | 33 |
| Ilustración 18 Implementación del output de la pestaña “Anomalía por indicación”..... | 33 |
| Ilustración 19 Factor de conversión aplicado a anomalías en “Anomalías por hospital”..... | 34 |
| Ilustración 20 Implementación nueva pestaña “Días respuesta”..... | 34 |
| Ilustración 21 Diseño panel inicio sistema AccuGest PowerBI..... | 35 |
| Ilustración 22 Editor Power Query, y receta de los pasos aplicados a los datos..... | 36 |
| Ilustración 23 Función Tipo cambiado en “tpni_entrada”..... | 36 |
| Ilustración 24 Valores erróneos reemplazados en columna en “FECHA_REGISTRO_EXTERNO”..... | 36 |
| Ilustración 25 Creación columna “RESULTADOS” y eliminación columna “RESULTADO”..... | 36 |
| Ilustración 26 Reorganización columnas tabla “tpni_entrada”..... | 37 |
| Ilustración 27 Renombre columnas tabla “tpni_entrada”..... | 37 |
| Ilustración 28 Filas en blanco eliminadas de la tabla “tpni_entrada”..... | 37 |
| Ilustración 29 Tabla “tpni_entrada” tras carga y transformación de los datos..... | 38 |
| Ilustración 30 Nueva tabla “RESULTADOS” con filtrado de columnas..... | 39 |
| Ilustración 31 Nueva tabla auxiliar “ANOMALIAS” a partir de fórmula en lenguaje DAX..... | 39 |
| Ilustración 32 Nueva tabla auxiliar “INDICACIONES” a partir de fórmula en lenguaje DAX..... | 40 |
| Ilustración 33 Nueva tabla auxiliar “ANOMALIA_INDICACION” a partir de fórmula en lenguaje DAX..... | 40 |
| Ilustración 34 Tabla Cuadros de Texto..... | 41 |
| Ilustración 35 Traducción páginas..... | 41 |
| Ilustración 36 Tabla Botones..... | 42 |
| Ilustración 37 Script creación CalendarioAccuGest.csv..... | 43 |
| Ilustración 38 Tabla Dim_Anomalias..... | 43 |
| Ilustración 39 Tabla Lat&Long..... | 44 |
| Ilustración 40 Generación de columna codProcedencia..... | 44 |
| Ilustración 41 Relaciones en la tabla “RESULTADOS”..... | 45 |
| Ilustración 42 Relaciones en la tabla “ANOMALIAS”..... | 45 |
| Ilustración 43 Medida “TOTAL_PRUEBAS” generada mediante DAX..... | 46 |
| Ilustración 44 Creación medida “SUM_CANTIDADES” con formula DAX..... | 49 |

| | |
|--|----|
| Ilustración 45 Filtro de objeto visual para "ANOMALIAS POR HOSPITAL" | 50 |
| Ilustración 46 Creación medida "SUM_CANTIDAD" con formula DAX. | 50 |
| Ilustración 47 Creación medida "T_RESPUESTA" con formula DAX..... | 51 |
| Ilustración 48 Gráfico "Anomalías por Hospital". | 52 |
| Ilustración 49 Nuevo sistema AccuGest mejorado. | 53 |
| Ilustración 50 Nuevo panel "Inicio" Power BI. | 54 |
| Ilustración 51 Nuevo panel "Tabla Resultados" Power BI..... | 54 |
| Ilustración 52 Nuevo filtro de procedencias Power BI..... | 55 |
| Ilustración 53 Nuevo panel "Gráficos de Distribución temporal"..... | 55 |
| Ilustración 54 implementación dataframe "res_out2" | 58 |

1 Introducción

1.1 Planteamiento del problema

En la actualidad, el manejo y análisis de grandes volúmenes de datos clínicos se ha convertido en un desafío creciente en el ámbito sanitario. La toma de decisiones de gestión y la eficiencia de los sistemas sanitarios dependen en gran medida de la capacidad de los profesionales sanitarios para acceder y analizar grandes cantidades de datos.

La visualización de datos es un proceso crucial que permite el análisis, comparación y estudio de información de manera accesible y fácil de entender para los usuarios. Una buena visualización puede simplificar la comprensión de datos complejos y ayudar en la toma de decisiones informadas en diversos temas. Tras todo proyecto de visualización de datos debe haber un buen diseño, un código efectivo y una buena gestión y análisis de los datos.¹

En concreto los datos clínicos que serán los protagonistas y en los que está basado el análisis que se ha realizado serán datos generados por la realización de pruebas de diagnóstico prenatal no invasivo a las gestantes de la Comunidad de Madrid, se utiliza el Plan de Implementación del TPNI (Test Prenatal No Invasivo), que genera cerca de 3.000 pruebas al año.

Con la intención del estudio y tratamiento de estos datos el grupo de Informática Biomédica (GIB) [1] y el grupo de Herencia y Genética del Hospital 12 de Octubre [2] desarrolló la herramienta AccuGest [3], que contiene un cuadro de mandos para la visualización de las pruebas de laboratorio clínico.

El actual cuadro de mandos sobre estos datos clínicos utilizado en el entorno de AccuGest presenta muchas limitaciones, ya que se desarrolló utilizando la herramienta Shiny diseñada en R. Esta herramienta no está diseñada específicamente para el manejo de grandes volúmenes de datos y no cuenta con una gran variedad de visualizaciones que, lo que dificulta la visualización y el análisis de estos por parte del usuario.

1.2 Solución propuesta

En este contexto, surge la necesidad de desarrollar nuevas herramientas que permitan el control y la monitorización automática de grandes volúmenes de datos y presentarlos de una forma visual y clara. Este es el objetivo principal de este Trabajo de Fin de Grado, en el que se presenta la migración y comparativa de un cuadro de mandos sobre datos clínicos, AccuGest.

Esta situación hace que sea necesario plantear una mejora en la visualización de datos para poder tomar decisiones de manera más eficiente y efectiva. La solución propuesta para mejorar la visualización y el análisis de los datos clínicos en AccuGest es la migración a una herramienta más avanzada como Power BI. Esta herramienta permitirá la creación de un nuevo cuadro de mandos que mejorará significativamente la forma en que los datos son presentados y analizados.

Con esta migración, se espera conseguir una herramienta más potente y funcional, capaz de proporcionar una visualización más detallada y completa de los datos, así como de simplificar el proceso de análisis y toma de decisiones. Además, el uso de Power BI proporcionará una significativa mejora en la forma en la que los datos son cargados y presentados al usuario debido al mayor enfoque en la visualización, con interfaces de usuario principalmente basadas en hacer clic para construir gráficos y cuadros de mando.

La migración a Power BI supone una mejora significativa en la capacidad de AccuGest para procesar, analizar y visualizar grandes volúmenes de datos, lo que se traduce en una mayor eficiencia en la gestión del sistema AccuGest.

1.3 Objetivos

El presente trabajo plantea como principales objetivos la mejora de la actual herramienta AccuGest, la migración a Power BI y la comparativa de los dos cuadros de mando. Para la consecución de ello, se plantearon los siguientes objetivos:

- Familiarización con sistema AccuGest: Aprendizaje del lenguaje R y sus herramientas para visualización de datos en Shiny, la familiarización y estudio en profundidad del sistema ACCUGEST, su funcionamiento y estructura de datos y la investigación sobre los datos TPNI (Test Prenatal No Invasivo) utilizados en el cuadro de mandos y sus características.
- Diseño e Implementación de mejoras en Análisis AccuGest: Estudio en profundidad de las necesidades de los usuarios con acceso al sistema AccuGest, un análisis de la mejoraría de la eficiencia y de la usabilidad de la herramienta a la que se llegara tras la migración y el diseño e Implementación de mejoras en AccuGest
- Análisis, migración y despliegue de nueva herramienta: Análisis y comparativa de las características y funcionalidades de los cuadros de mandos en R/Shiny y Power BI, la migración y adaptación del cuadro de mandos AccuGest de R/Shiny a Power BI y la identificación y desarrollo de mejoras en la visualización y análisis de datos en Power BI para el cuadro de mandos.
- Evaluación de resultados: Comparación y evaluación de las mejoras obtenidas tras la migración, un análisis de la usabilidad y satisfacción de los usuarios con la herramienta migrada y la identificación de posibles mejoras y ajustes en la herramienta migrada en función de los resultados obtenidos
- Redacción de la memoria: Estructuración y organización de la memoria, y la redacción y revisión de la memoria.

1.4 Planificación del trabajo

Para completar todos estos objetivos en el tiempo disponible para la realización del Trabajo de Fin de Grado se diseñó un diagrama de Gantt con la lista de tareas y los tiempos para completarlas:

Migración y comparativa de un cuadro de mandos Power BI sobre datos clínicos; AccuGest

TFG ETSIINF
Andrés Bravo Francos

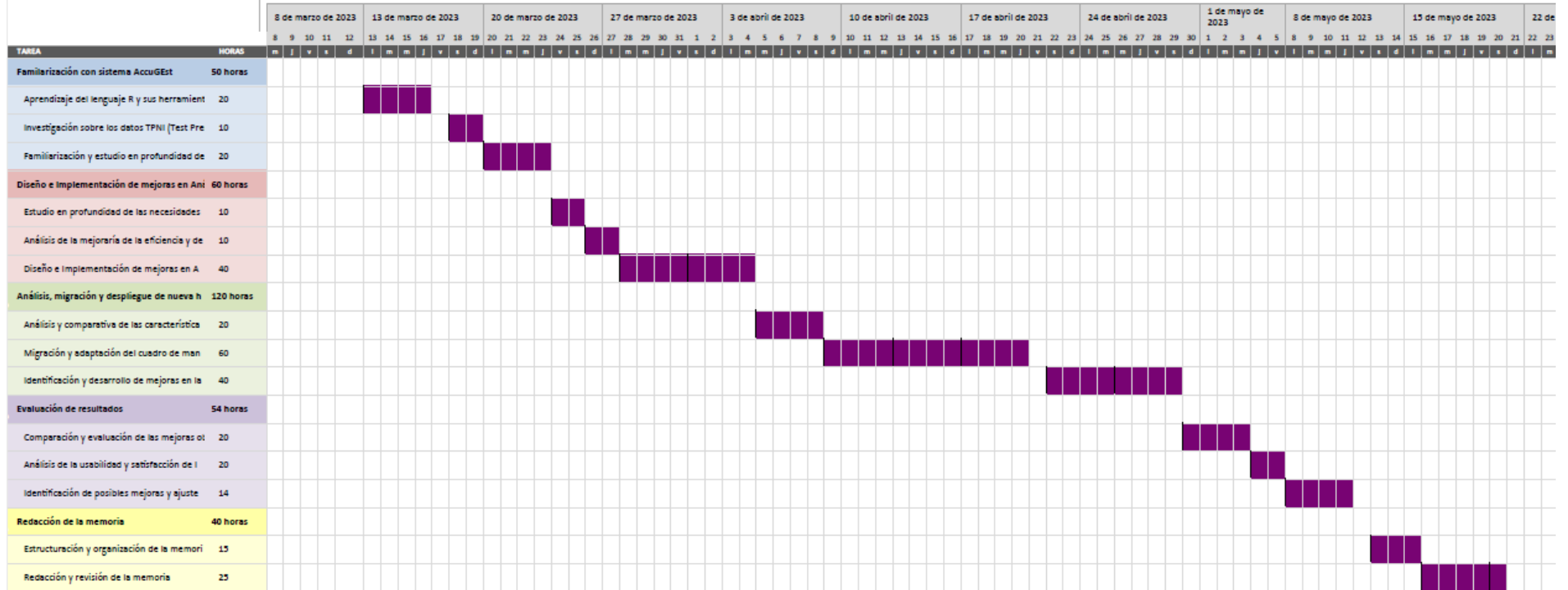


Ilustración 1: Diagrama de GANTT del trabajo

2 Estado del arte y tecnologías empleadas

2.1 Herramientas de visualización

La visualización de datos constituye una disciplina propia dentro del universo de la ciencia de datos[4].

La visualización gráfica de datos juega un papel fundamental en diversos campos como la ciencia, ingeniería, negocios, banca y medio ambiente, entre otros. Actualmente existen numerosas herramientas disponibles para el diseño, desarrollo y comunicación de visualizaciones de datos y a continuación vamos a estudiar algunas de ellas.

Para poder acertar en el uso de una determinada herramienta de visualización, será fundamental tener claro el objetivo que se pretende alcanzar. Si el análisis y visualización de los datos se realizará por parte de comunicadores profesionales o personal no técnico de las distintas áreas de negocio, se recomienda utilizar herramientas modernas de Business Inteligente[5] que combina análisis de negocios, minería, visualización, herramientas e infraestructura de datos, además de prácticas recomendadas para ayudar a las empresas a tomar decisiones basadas en los datos. En este caso las herramientas recomendadas son Amazon Redshift, Power BI, Qlik Sense, Tableau, Docker, Ekon Vision[6], etc.

Otro caso, es el del científico de datos o Data Science que se adecua perfectamente al propósito principal que tuvo el sistema AccuGest en un principio. Este tiende a trabajar con dataframes de análisis concretos que generalmente incluyen todos los componentes necesarios para llevar a cabo el trabajo del análisis de datos y la visualización de estos con una curva de aprendizaje más rápida que en otras herramientas. Los entornos más populares hoy en día para este campo son R[7] y Python[8], los dos lenguajes cuentan con librerías nativas para la analítica visual, como son ggplot2, tidyverse o plotly que se comentarán en profundidad más tarde.

Actualmente, en un mundo en el que la cantidad de datos útiles para el análisis permanece en constante crecimiento, es cada día más necesario contar con este tipo de herramientas que facilitan la obtención de valor procedente de los datos y, a partir de este valor, pueda tener lugar una toma de decisiones fundamentadas tanto en el presente como en el futuro de un negocio o cualquier otra actividad que precise de estas herramientas.

Aunque existen muchas herramientas de visualización novedosas, no podemos estudiar todas ellas, por eso nos centraremos en el estudio de las herramientas que van a ser utilizadas en este proyecto, desde la creación del sistema AccuGest hasta la reciente migración a una nueva herramienta.

2.1.1 RStudio

Rstudio[9] es un entorno de desarrollo integrado (IDE) que hace posible el desarrollo de webs y que está especialmente diseñada para trabajar con el lenguaje de programación R. Fue creada y lanzada por el equipo de RStudio Inc. en 2011, y desde entonces es una herramienta muy popular entre científicos de

datos, estadísticos y analistas que buscan realizar análisis y visualizaciones de datos de manera eficiente.

RStudio ofrece un entorno completo para trabajar con R, en el que se incluye un editor de código con características avanzadas como el resaltado de sintaxis y autocompletado, lo cual facilita la implementación de código. Además, proporciona una consola que permite ejecutar comandos de forma interactiva y ver los resultados en tiempo real.

Una de las características que distingues a Rstudio del resto de IDE es su capacidad para generar visualizaciones interactivas y gráficos de alta calidad. Para ello utiliza librerías y paquetes especializados como ggplot2, plotly, etc. Los usuarios pueden crear gráficos personalizados y explorar los datos de manera visual. Esta herramienta está diseñada de forma que cumple dos objetivos que eran necesarios para diseñar el primer sistema AccuGest, en primer lugar, busca facilitar la curva de aprendizaje de nuevos usuarios que eligen R como opción y está pensada también para proporcionar herramientas de alta productividad a usuarios más avanzados.[10]

Además de sus capacidades de análisis y visualización, RStudio ofrece herramientas para la gestión de proyectos, la colaboración y la presentación de resultados. La integración con sistemas de control de versiones, la generación de informes reproducibles con RMarkdown y la creación de aplicaciones interactivas con Shiny son de hecho algunas de las muchas funcionalidades que han hecho que Rstudio sea una herramienta poderosa, versátil y una elección popular en profesionales que buscan un entorno de desarrollo intuitivo y eficiente para sus tareas de análisis de datos.

2.1.2 Shiny

Shiny [11] es una plataforma desarrollada por RStudio que hace posible la creación gráficos interactivos y de aplicaciones web interactivas mediante R y Python. Se lanzó en 2012 y desde entonces ha experimentado un crecimiento significativo en popularidad y adopción.

Todas las aplicaciones creadas con Shiny se comportan de forma similar siguiendo una misma estructura, están compuestas por dos componentes principales: la Interfaz de Usuario (UI) y el Server. La UI es el objeto que contiene los widgets para recibir el input del usuario y mostrar los outputs, mientras que el Server recibe los inputs del UI y con ellos genera outputs dentro de este que serán los que al final se representen en la aplicación creada.

Entre las cualidades más importantes que podemos destacar de Shiny se encuentran:

Facilidad de uso: Shiny está diseñado para ser accesible tanto para usuarios principiantes como para usuarios más avanzados en R. Al proporcionar una estructura intuitiva junto a una amplia gama de funciones y herramientas, se facilita el desarrollo de aplicaciones interactivas.

Interactividad: Shiny permite crear aplicaciones web dinámicas que tienen como objetivo final, que los usuarios puedan interactuar con los datos y visualizaciones. Esto sin duda proporciona una experiencia mucho más

inmersiva al usuario y le brinda la posibilidad de explorar y analizar los datos de forma personalizada.

Flexibilidad: Shiny ofrece una amplia flexibilidad en cuanto al tipo de aplicaciones que cualquier usuario podrá crear. Permite integrar de forma más o menos complejas visualizaciones, gráficos, tablas y otros elementos interactivos, y proporciona opciones que permitirán personalizar la apariencia y el diseño final de la aplicación.

Compatibilidad con R: Al estar basado en Python y en R, Shiny permitirá al usuario utilizar todas las capacidades y librerías disponibles que están disponibles en este lenguaje de programación para el análisis y procesamiento de datos. Esto incluye una amplia gama de funciones estadísticas y de visualización, lo que brinda a los usuarios la capacidad de realizar análisis avanzados en sus aplicaciones.

Comunidad y soporte: Shiny cuenta con una comunidad de usuarios activa y con desarrolladores que comparten sus recursos, experiencias a la hora de resolver problemas, tutoriales y paquetes adicionales. Además, RStudio proporciona una documentación lo suficientemente detallada, con ejemplos y un soporte técnico que tiene como intención ayudar a los usuarios a aprovechar al máximo la plataforma.

2.1.3 Power BI

Utilizaremos como herramienta de visualización de creación del nuevo panel Power BI, ya que es una de las herramientas líder en el análisis empresarial y al estar desarrollada por Microsoft ofrece visualizaciones de gran calidad interactivas integradas en informes y paneles. Inicialmente se desarrolló en 2010 bajo el nombre de "Project Crescent", aunque no fue lanzada públicamente hasta el 2011. En 2013 la herramienta cambió de nombre, renombrándola, así como Power BI y finalmente su lanzamiento oficial fue en julio de 2015[12]. Ofrece una inmensa variedad de potentes objetos de visualización de datos e informes interactivos, ese es uno de los principales motivos por los que se ha convertido en una opción popular para las organizaciones que buscan aprovechar al máximo sus datos y obtener información significativa para la toma de decisiones. Entre sus características, destacan tanto las opciones de seguridad para proteger los datos de análisis como el etiquetado de confidencialidad y el control de acceso en tiempo real.

Actualmente, la herramienta Power BI está dividida en dos partes fundamentales: los servicios de BI basados en "cloud" o también dicho en la nube, que se conocen como "Power BI Services", y una interfaz de escritorio llamada "Power BI Desktop". A través de Power BI Services, los usuarios podrán acceder a sus datos desde cualquier lugar mediante la nube, con este servicio Microsoft facilita el trabajo colaborativo entre varios usuarios y el acceso en tiempo real a los informes y paneles. Por otro lado, una aplicación de escritorio llamada Power BI Desktop, que permite a los usuarios diseñar y personalizar sus informes de forma más detallada, ofreciendo una interfaz intuitiva y flexible.

Además de estas capacidades, Power BI proporciona opciones que han sido motivo de su reconocimiento hoy en día, en concreto herramientas para la preparación de datos, lo que permite a los usuarios limpiar, transformar y modelar los datos antes de su análisis. También ofrece funcionalidades que

permitirán al usuario identificar patrones, tendencias y relaciones ocultas en los datos.

La herramienta también incluye capacidades de análisis avanzado, como el modelado de datos, el filtrado interactivo y la capacidad de realizar cálculos complejos utilizando el lenguaje DAX (Data Analysis Expressions) [13], un lenguaje de expresiones de fórmulas que se usa en Analysis Services, Power BI y Power Pivot en Excel. Las fórmulas DAX abarcan funciones, operadores y valores para realizar cálculos avanzados y consultas en los datos de las tablas y columnas relacionadas de los modelos de datos tabulares. Esto permite a los usuarios explorar los datos en profundidad y descubrir patrones, tendencias y relaciones ocultas en la información.

En cuanto a la escalabilidad y el rendimiento, podemos decir que Power BI está diseñada una manera que la convierte en una de las herramientas líder en el manejo de grandes volúmenes de datos y ofrece una respuesta rápida a las consultas y visualizaciones. La herramienta cuenta además con opciones algo más complejas de optimización y ajuste de rendimiento que garantizan al usuario una experiencia fluida incluso con conjuntos de datos complejos.

2.2 Sistemas de almacenamiento

Es fundamental contar con sistemas de almacenamiento eficaces y suficientes para la estructuración y gestión de la información a la hora de diseñar un cuadro de mando de datos clínicos. Examinaremos algunas tecnologías de almacenamiento populares en esta sección, concentrándonos en MySQL[14], Redis[15] y Oracle DBMS[16].

2.2.1 MySQL

MySQL es un sistema de gestión de bases de datos muy utilizado en la industria. Su popularidad se debe a que es un sistema de código abierto, lo que facilita su acceso y soporte por una comunidad enorme. MySQL destaca por su compatibilidad con el lenguaje SQL, esto permite una migración desde otros sistemas de bases de datos de forma sencilla. Además, su arquitectura sigue el modelo cliente-servidor, esto mejora el rendimiento y ofrece funciones como procedimientos almacenados, lo que permite una mayor eficiencia al realizar la tarea de procesamiento de tablas y en actualización de datos. La seguridad en las operaciones es otra de las características más destacada de MySQL que garantiza la exactitud de los datos

2.2.2 Redis

Redis conocida por su sencillo uso, es otra de las herramientas puntero actualmente en el almacenamiento en memoria de código abierto, cuenta con una arquitectura que se basa en un modelo clave-valor, esto permite un rendimiento extremadamente rápido. Cada pieza de información se asocia con una clave única, esta clave tiene como objetivo actuar como identificador único para acceder y recuperar los datos almacenados en el sistema Redis. El “valor” no es más que la información que se guarda y se asocia con la clave. Es por ello, que la utilización de este modelo clave-valor permite una gran flexibilidad en la gestión y manipulación de la información. Redis ofrece diversas estructuras de

datos, entre ellas podemos destacar listas, conjuntos y mapas de hash. La utilización de estas estructuras permite a los usuarios una gran flexibilidad en la gestión y manipulación de la información. Además, cuenta con una gran capacidad de persistencia, esto permite guardar los datos en disco, lo que garantiza que la información incluso en un hipotético caso de reinicio del sistema sea duradera.

2.2.3 Oracle DBMS

Por último, vamos a hablar de Oracle DBMS, un sistema de administración de bases de datos multi-modelo, en este caso de uso comercial ampliamente utilizados en entornos empresariales. Ofrece un conjunto completo de herramientas y funcionalidades para el almacenamiento, gestión y análisis de datos. Oracle DBMS destaca por su escalabilidad, seguridad y confiabilidad, esto lo hace un sistema capaz de manejar grandes volúmenes de información de una manera muy eficiente. Además, es importante que destaquemos la que ofrece capacidades avanzadas de consulta y análisis, permitiendo realizar consultas complejas y generar informes detallados. Mantiene un permanente enfoque en la integridad y consistencia de los datos, junto con su robustez y soporte técnico, esto lo convierte en una de las mejores opciones para aplicaciones que requieren altos estándares de rendimiento y disponibilidad.

2.3 Fuentes de datos

En este proyecto, es esencial comprender de dónde provienen los datos que utilizamos para mostrar la información en el sistema AccuGest. Las fuentes de datos son como los lugares de donde sacamos la información, que luego organizaremos y relacionaremos para poder verla en la herramienta de visualización.

Es importante asegurarse, que los datos se obtienen de manera correcta y que así estos sean precisos y confiables en el sistema AccuGest. Para eso, utilizamos diferentes fuentes, como bases de datos de los hospitales y sistemas externos de análisis clínico. Se eligen cuidadosamente las fuentes y se modifican para tener datos fiables en el sistema.

Luego, se lleva a cabo un proceso para preparar los datos y que estos se ajusten al formato que el sistema AccuGest necesita para mostrar los datos. Durante esta etapa, limpiamos y organizamos los datos para que sean coherentes y se puedan mostrar correctamente.

Es importante destacar que elegir y configurar las fuentes de datos adecuadas es fundamental para obtener resultados precisos y relevantes en el sistema AccuGest. También nos aseguramos de actualizar regularmente los datos para tener la información más actualizada para los usuarios.

En resumen, las fuentes de datos son los lugares de donde sacamos la información para mostrar en la herramienta de visualización utilizada en el sistema AccuGest.

3 Requisitos y diseño

3.1 Requisitos

El propósito principal de este proyecto es mejorar el sistema AccuGest, un cuadro de mandos desarrollado en R y Shiny que se enfoca en los resultados de los Test Prenatales No Invasivos (TPNI) realizados en los hospitales de la Comunidad de Madrid. El objetivo es migrar este cuadro de mandos a otra herramienta de visualización y realizar una comparativa entre ambos.

- Utilización de una misma fuente de datos: Ambos cuadros de mandos deben utilizar la misma fuente de datos, la cual da forma al actual sistema AccuGest. Esto garantiza la consistencia y coherencia de la información utilizada en ambas herramientas.
- Seguir el flujo de datos establecido en AccuGest: El flujo de datos existente desde la puesta en marcha del sistema AccuGest será mantenido en los nuevos cuadros de mandos. Esto implica la recopilación y procesamiento de muestras y datos de múltiples sistemas involucrados en la gestión de datos para los TPNI.
- Visualización clínica con filtros temporales y por procedencia: Ambos cuadros de mandos deben permitir la aplicación de filtros temporales y de procedencia por selección del usuario. Los filtros temporales permitirán analizar los datos en diferentes intervalos de tiempo, mientras que los filtros por procedencia permitirán realizar análisis específicos para distintos Centros hospitalarios

3.1.1 Requisitos comunes de cada cuadro de mandos

El flujo de los datos será común en ambos cuadros de mando, seguirá de la misma manera el curso que ha seguido desde la puesta en marcha del sistema AccuGest. La gestión de datos en un sistema centralizado para una región extensa como la Comunidad de Madrid involucra a múltiples sistemas que se encargan de recopilar y procesar muestras y datos con el fin de obtener resultados precisos y confiables.

Desde el Hospital 12 de Octubre se reciben todos los datos relacionados con las peticiones TPNI de 22 centro públicos de la Comunidad de Madrid a través de un gestor de peticiones unificado con acceso mediante LDAP (Lightweight Directory Access Protocol) para aquellos usuarios autorizados desde CESUS[17] Comunidad de Madrid.

Las mujeres embarazadas que son estudiadas se identifican a través del Código de Identificación Personal Autonómico (CIPA), cuyos datos demográficos se agregan al sistema mediante una consulta a la base de datos CIBELES[18]. Estos datos, junto con la información específica de la solicitud (como el número del tubo de extracción), se incorporan al sistema corporativo SIL, que a su vez envía una solicitud al sistema departamental de Genética utilizando el estándar de mensajería HL7-FIHR los datos, un estándar de interoperabilidad desarrollado por HL7 (conjunto de estándares Health Level 7) diseñado para permitir el intercambio electrónico de datos de atención médica entre diferentes sistemas en la industria de la atención médica [19].

En este punto, se añaden los datos resultados del análisis terciario de la secuenciación de genoma completo realizada como consecuencia del TPNI, que se incorporan al sistema departamental SIL para generar un informe final. Este

informe final se guarda tanto en el sistema departamental SIL como en el sistema unificado para que pueda ser consultado por los diferentes Centros.

Como resultado, todos los datos completos de cada mujer embarazada se registran en el sistema departamental SIL, que es la fuente primaria de datos que alimenta el panel de control actual y alimentará el nuevo panel de control.

Como estudiante de la Universidad Politécnica de Madrid no cuento con ese acceso, aunque no será necesario puesto que como origen de los datos para la visualización de datos genómicos se tomará un archivo en formato.xls generado desde el SIL departamental del Servicio de Genética del Hospital 12 de Octubre.

3.1.1.1 Filtro temporal

Dentro de este contexto de requisitos comunes para ambos cuadros de mando, será necesario incorporar de igual manera un filtro temporal que sea útil y se aplique a todos los objetos visuales del nuevo panel. Este filtro debe permitir delimitar un rango de tiempo específico para los obtener los cálculos y la visualización de los datos en función de este.

El filtro temporal dentro del actual sistema AccuGest consiste en un widget de control, que permite al usuario seleccionar un rango temporal deseado como puede observarse en la “Ilustración 2”. Para ello, se utiliza un deslizador de doble entrada que establece los límites del rango temporal de los datos que se mostrarán en el panel. Las fechas límite que se presentan en los extremos del deslizador se generan automáticamente a partir de los datos brutos descargados del SIL departamental.

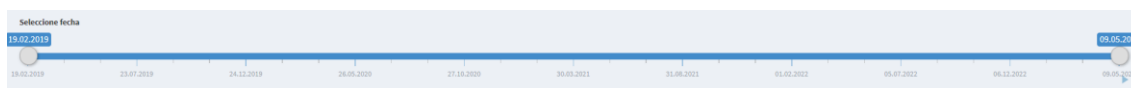


Ilustración 2 Controlador deslizante de doble entrada para delimitar de manera dinámica el rango temporal

La incorporación de este filtro temporal brinda al usuario la posibilidad de adaptarse a un intervalo de tiempo particularmente relevante dentro de los datos disponibles en el cuadro de mando. Esto facilita la visualización y análisis de los datos en función de un rango temporal seleccionado, lo que puede ayudar a identificar patrones, tendencias y cambios significativos en los datos a lo largo del tiempo.

Cabe destacar que este filtro temporal debe ser una característica común en todos los objetos visuales del cuadro de mando, tanto en el cuadro de mando existente como en el nuevo. Su inclusión en ambos cuadros de mando, el ya existente desarrollado en R/Shiny y el nuevo desarrollado en PowerBi, garantiza una experiencia de usuario coherente y facilita la exploración y comprensión de los datos a lo largo del tiempo.

3.1.1.2 Filtro por procedencia

Ambos cuadros de mando deben incluir un filtro de procedencias que sea aplicable a todos los objetos visuales del nuevo cuadro de mando. Este filtro de procedencias tiene como objetivo permitir al usuario seleccionar y filtrar los datos en función de las distintas procedencias o hospitales de origen.

En el existente sistema AccuGest desarrollado en R/Shiny, se optó por utilizar un filtro de selección como método para llevar a cabo un filtro de procedencias como puede observarse en la “Ilustración 3”. Las opciones de selección en el filtro deben incluir todos los Centros públicos de la Comunidad de Madrid que han realizado peticiones TPNI, estos datos se obtendrán de la información obtenida de los datos brutos descargados del SIL Departamental. Actualmente, los hospitales seleccionados inicialmente en el filtro se basan en los seis hospitales con mayor cantidad de pruebas realizadas durante el periodo de tiempo establecido por el filtro temporal, que inicialmente abarcará todas las pruebas realizadas a lo largo del tiempo registrado. El filtro de procedencias debe ser configurable y permitir al usuario realizar selecciones personalizadas de hospitales.

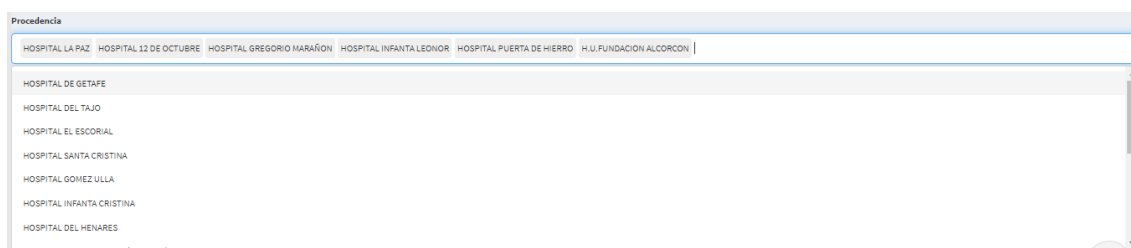


Ilustración 3 Controlador de selección de procedencias que se estudiarán en el cuadro de mandos.

El fin de este filtro es el de ayudar a identificar patrones específicos o tendencias en los datos provenientes de diferentes hospitales. Además, como es una característica común en todos los objetos visuales del cuadro de mando, tanto en el existente sistema AccuGest existente como en el nuevo desarrollado en Power BI, se garantiza una experiencia coherente y facilita la exploración y comprensión de los datos relacionados con las diferentes procedencias o hospitales.

3.1.1.3 Tarjetas resumen resultados

Es necesario crear unas tarjetas a modo de resumen en el nuevo sistema AccuGest de forma similar a lo que se ha hecho en el sistema existente como se observa en la Ilustración 4. Estas tarjetas tienen la función de resumir los resultados a grande escala, sin llegar a profundizar en ellos de manera que incluyan:

Número de pruebas totales: muestras el número de pruebas realizadas a partir de las peticiones solicitadas desde cada centro y registradas en la base de datos.

Número de casos positivos: Muestra la cantidad de casos con resultados positivos y su porcentaje en relación al total de casos en el rango temporal seleccionado.

Número de casos no concluyentes: Muestra la cantidad de casos que no han concluido y su porcentaje en relación con el total de casos en el rango temporal seleccionado.

Número de procedencias: Muestra el número de centros que participan en el programa para realizar TPNI (Test Prenatal No Invasivo) de Aneuploidías en la Comunidad de Madrid.

Estas tarjetas resumen deben ser dinámicas y ajustarse exclusivamente al rango temporal definido por el filtro temporal y no se deben filtrar por procedencia, ya que su objetivo principal es brindar un resumen general de los resultados en función del tiempo. De esta manera, los valores mostrados en las tarjetas se actualizarán automáticamente según el período de tiempo seleccionado por el usuario y deben formar parte del panel de inicio.

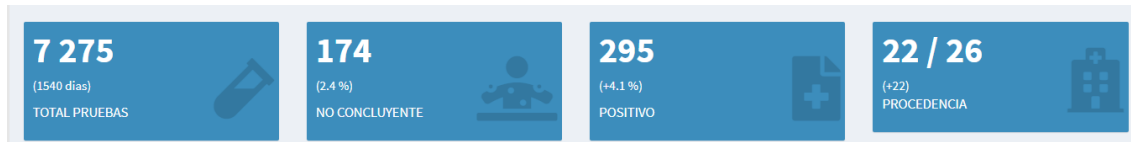


Ilustración 4 Tarjetas creadas a modo resumen de los resultados obtenidos en función del filtro temporal.

3.1.2 Requisitos propios de cada cuadro de mandos

3.1.2.1 Requisitos propios AccuGest RStudio/Shiny

- Mejorar el procesamiento de los datos: El objetivo es optimizar los algoritmos diseñados para realizar el análisis de los resultados de los TPNI. Se debe buscar mejorar la eficiencia en el procesamiento de los datos clínicos estudiados.
- Mejorar la interfaz de usuario: La mejora del sistema AccuGest debe proporcionar una interfaz fácil de usar y comprender para el usuario. Se debe buscar la coherencia las funcionalidades originales y el nuevo diseño y las nuevas funcionalidades del sistema.
- Lograr una visualización más clara y precisa de los actuales datos: El nuevo sistema tiene como requisito mejorar las visualizaciones de los datos del sistema.
- Añadir funcionalidades de filtrado mejoradas: Se deben incorporar, en la medida de lo posible nuevas funcionalidades de filtrado que mejoren la experiencia y la interpretación de los datos. Estas funcionalidades deben permitir a los usuarios aplicar filtros más precisos y personalizados en los datos.

3.1.2.2 Requisitos propios AccuGest PowerBI

Los requisitos propios para AccuGest PowerBI se basan en replicar el sistema original de RStudio/Shiny y agregar valor mediante las capacidades que ofrece la herramienta PowerBI:

- Replicar el sistema original de RStudio/Shiny y añadir valor con PowerBI: El nuevo panel en AccuGest PowerBI debe replicar las funcionalidades y características del sistema original desarrollado en RStudio/Shiny. Esto garantizará la continuidad y coherencia en la experiencia de uso para los usuarios que estén familiarizados con el sistema anterior. Además, se debe aprovechar el poder y las capacidades adicionales que ofrece PowerBI para agregar un valor añadido al sistema.
- Mejorar el diseño actual de AccuGest: El diseño del panel en AccuGest PowerBI se mejorará con el objetivo de ofrecer una experiencia visual atractiva que se adapte más al término “userfriendly”. Esto implica utilizar una mejor disposición de los elementos visuales, utilizar una

gama de colores que encaje con los datos clínicos que se vas a analizar y un diseño que se adapte a diferentes dispositivos y tamaños de pantalla.

- Añadir mayor posibilidad de filtrado de datos y mejorar la facilidad de uso: Se deben implementar funcionalidades de filtrado más avanzadas en las visualizaciones en AccuGest para permitir a los usuarios realizar un análisis más profundo de los datos. Estos filtros deben abarcar diferentes dimensiones para el estudio de los TPNI, como el tiempo, la procedencia, las indicaciones, las anomalías, entre otros.
- Utilizar tooltips en las herramientas para mejorar la experiencia de uso: Se deben incorporar tooltips descriptivos en las diferentes visualizaciones del sistema. Estos tooltips aportaran información sobre los datos analizados en cada objeto visual, ayudando al usuario a comprender mejor cada una de las visualizaciones.

3.2 Diseño

3.2.1 Diseño mejoras en AccuGest en Rstudio/Shiny

Vamos a detallar las mejoras de diseño propuestas para el sistema AccuGest existente, en su versión basada en Rstudio/Shiny. Estas mejoras van a centrarse tanto en los aspectos visuales como funcionales con el objetivo de mejorar la experiencia de usuario y la interpretación de los datos.

A simple vista, el sistema se encuentra tal y como aparece en la Ilustración 5 contaba con varias posibles mejoras. Dividiremos el diseño de mejoras en dos tipos, mejoras visuales y mejoras funcionales

- Mejoras visuales:
 - Mejora en la visualización de los filtros: Se realizará una mejora en la presentación visual de los dos filtros principales. Se ajustará el diseño para evitar que se corten en la parte izquierda y se asegurará que haya suficiente espacio entre el borde izquierdo y los filtros. Como puede observarse el campo fecha en el borde superior izquierdo no es completamente visible, y lo mismo pasa con el filtro de selección de procedencia.
 - Ajustar tamaño de la cuarta tarjeta resumen: Se realizará una mejora visual en la cuarta tarjeta resumen que indica el número de Centros hospitalario que forman parte de este estudio. Se ajustará el tamaño de la tarjeta para que tenga las mismas dimensiones que las otras tarjetas, creando una apariencia simétrica y coherente en el diseño.
 - Mejoras en los gráficos de distribución temporal: se realizarán correcciones y ajustes en los gráficos de distribución temporal, dentro de las opciones que permite de la herramienta. El objetivo es mejorar la interpretación de estos gráficos, especialmente en la primera pestaña del panel de gráficos de distribución temporal que como se observa en la Ilustración 5, la interpretación de esos datos resulta muy compleja.

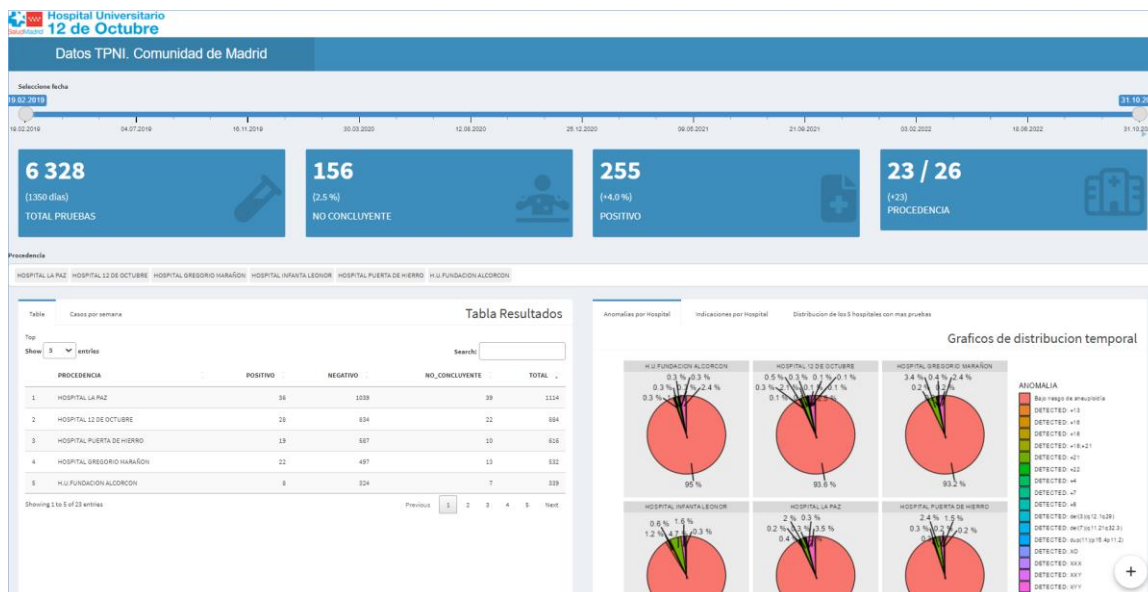


Ilustración 5 Sistema AccuGest desde el que se parte para aplicar mejoras

Si bien las mejoras visuales propuestas no implican directamente una mejora en la funcionalidad, son de extrema necesidad. Estas mejoras tienen un impacto significativo en la experiencia del usuario y en la comprensión de los datos presentados. Un diseño visual diseñado correctamente y atractivo juega un papel crucial en la usabilidad de cualquier aplicación.

- Mejoras funcionales:

Añadir tres nuevas pestañas en el panel "Tabla Resultados" como puede observarse en la Ilustración 6 y una nueva pestaña en el panel de gráficos de distribución temporal:

- Pestaña "Anomalías": Se añadirá una tabla con el desglose de las anomalías detectadas en el estudio (actualmente solo disponible en la sección de gráficos) junto a un filtro de selección múltiple que permita filtrar los datos por anomalía y que así a los usuarios puedan explorar los resultados en función de diferentes anomalías seleccionadas. Esta tabla permitirá visualizar de manera detallada las anomalías identificadas en el análisis, a la vez que se observa el gráfico "Anomalías por Hospital".
- Pestaña "Indicaciones": Se creará una tabla desglosando las indicaciones relacionadas con el estudio junto un filtro de selección múltiple para filtrar los datos por indicación. Esto permitirá a los usuarios explorar los resultados en función de las diferentes indicaciones.
- Pestaña "Anomalías por Indicación": Se habilitará una pestaña que permita seleccionar una anomalía específica mediante un filtro, y mostrará el número de indicaciones relacionadas ordenadas de manera descendente que han dado lugar a resultados positivos en dicha anomalía.
- Pestaña "Tiempo Respuesta": Esta pestaña mostrará el tiempo de respuesta por centro, representando la diferencia entre la fecha de solicitud y la fecha de informe. Esta información permitirá evaluar los tiempos de respuesta en cada centro y realizar un seguimiento de la eficiencia del proceso.

Las mejoras funcionales propuestas para el sistema AccuGest en su versión basada en RStudio/Shiny tienen como objetivo aumentar la exploración de resultados específicos dentro del sistema. Aunque, al contrario que los anteriores vistos en este mismo apartado pueden no tener un impacto directo en el aspecto visual, son fundamentales para mejorar la funcionalidad y la usabilidad de la aplicación.

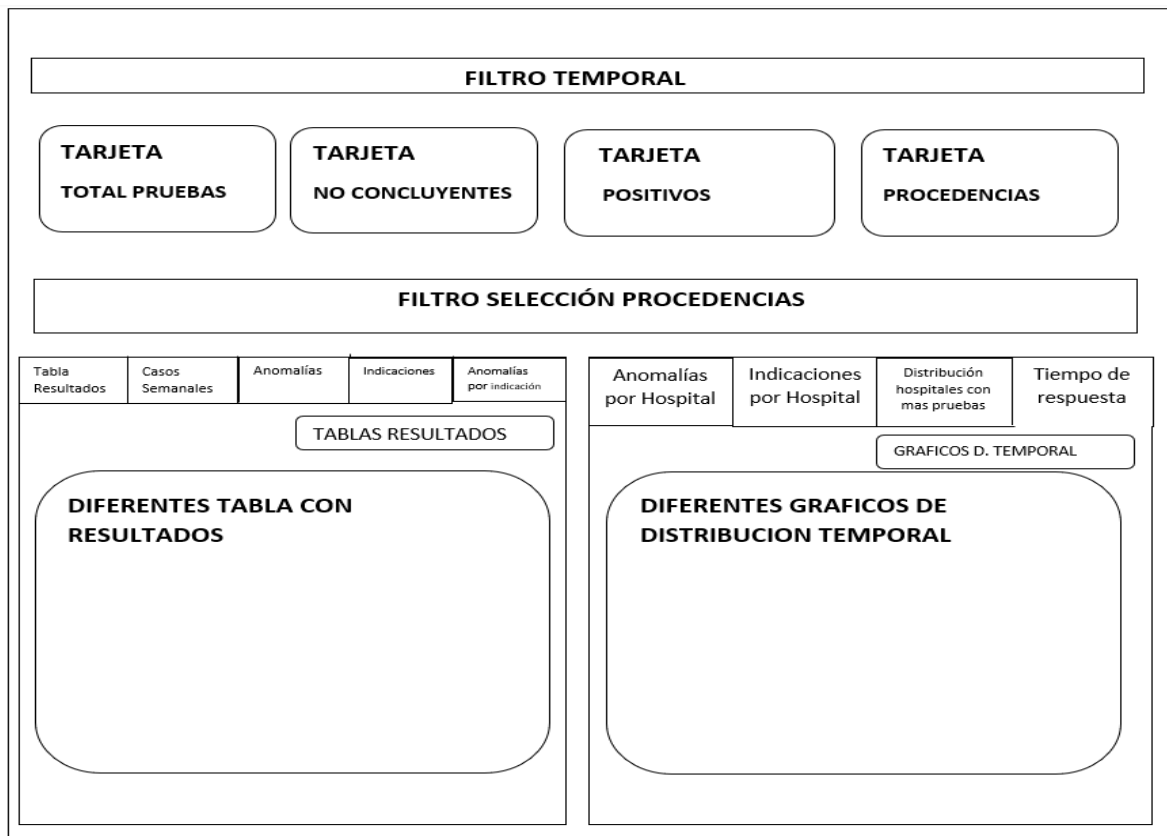


Ilustración 6 Diseño sistema AccuGest desde el que se parte para aplicar mejoras.

3.2.2 Diseño panel AccuGest en PowerBi

En Power BI, el diseño de paneles se basa en la utilización de páginas y botones para organizar y presentar los datos de manera efectiva y fácil para el usuario. Las páginas son como pestañas que permiten mostrar diferentes conjuntos de visualizaciones y análisis de datos, y se navega entre ellas a través de botones.

3.2.2.1 Diseño migración de datos

En el diseño de migración de datos, se priorizará la facilidad de mantenimiento del nuevo sistema, buscando minimizar las tareas necesarias para su correcto funcionamiento. Para ello, se tratará de aprovechar grandes capacidades que Power Bi ofrece para la gestión y análisis de datos.

El proceso de migración se basará en el archivo XLS[20] facilitado que contiene los datos a ser migrados. Este archivo será utilizado como fuente de datos por Power BI, permitiendo la importación y transformación de los datos de manera automatizada.

Además, de utilizar Power BI como herramienta principal, se utilizará la herramienta desarrollada en su interior, Power Query para llevar a cabo el proceso de Extracción, Transformación y Carga. Con Power Query, se definirán una serie de pasos aplicados y transformaciones necesarios para preparar los datos antes de cargarlos en las visualizaciones. Estos pasos se pueden aplicar de forma repetible y automática, lo que asegura la consistencia y la calidad de los datos en el nuevo sistema AccuGest.

3.2.2.2 Diseño nuevo panel AccuGest

Se ha decidido orientar el diseño del nuevo panel de AccuGest en PowerBI de forma que esta nueva herramienta se divida en tres subpaneles principales en los que la navegación de páginas entre ellos se realice para que el usuario pueda moverse por el desde una primera vez y de forma intuitiva.

Un primer panel inicio será la primera visión que el usuario tendrá del nuevo sistema AccuGest. Este tiene la intención de resumir los datos obtenidos en el sistema ofreciendo la posibilidad de verlo de la manera original, es decir a partir de tarjetas resumen y a partir de un mapa interactivo. Como se observa en la Ilustración 7, en la esquina superior derecha, se incluirán dos botones: uno para acceder a los resultados en tablas y otro para ver los resultados en gráficos de distribución temporal.

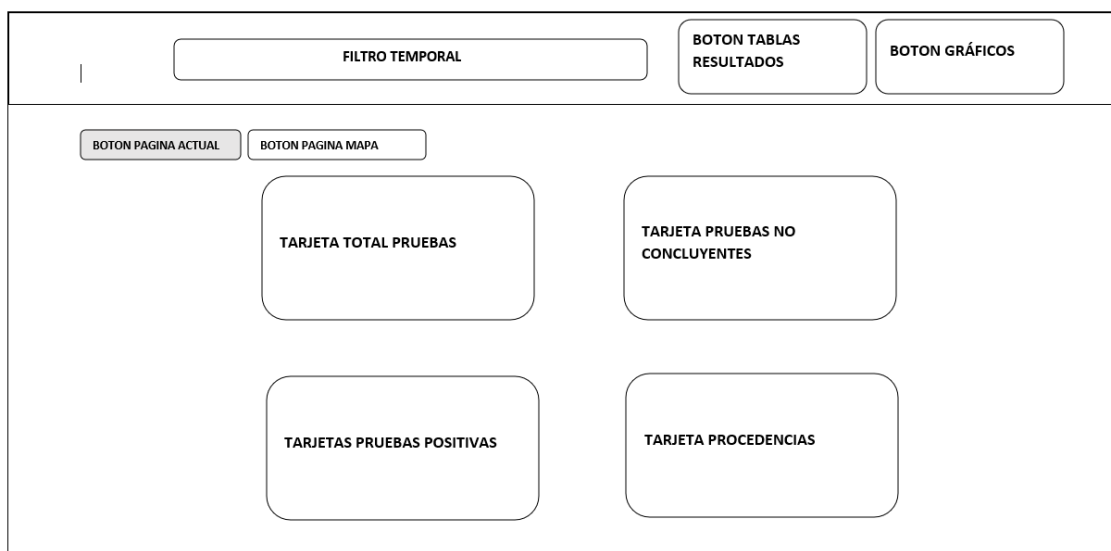


Ilustración 7 Diseño panel inicio sistema AccuGest PowerBI

En la vista de tarjetas resumen, se presentarán los valores clave y los indicadores principales de los datos en forma de resumen. Estas tarjetas estarán asociadas al filtro temporal, lo que permitirá que los valores se actualicen automáticamente en función del rango temporal seleccionado.

Por otro lado, en la vista del mapa interactivo, se representará un punto por cada procedencia de los datos. Se está considerando la posibilidad de añadir un tooltip que proporcione un resumen de los resultados del centro hospitalario al pasar el cursor sobre cada punto en el mapa, como puede observarse en la Ilustración 8.

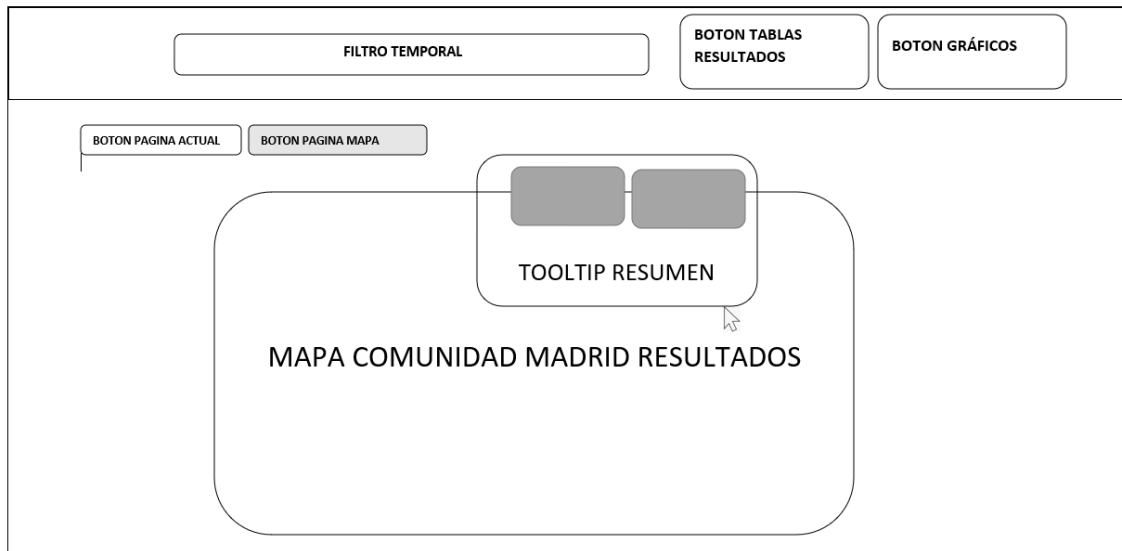


Ilustración 8 Diseño tooltip en el panel inicio sistema AccuGest PowerBI.

Los otros dos subpaneles del sistema AccuGest en Power BI, se mantendrá un diseño parecido al descrito anteriormente, a excepción de un pequeño conjunto de botones situados en la parte superior derecha, que permitirá al usuario acceder rápidamente a todas las tablas o a todos los gráficos del sistema en cualquier momento. Dicha navegación de páginas tiene la intención de garantizar una navegación fluida y una exploración eficiente de los datos.

Los gráficos y objetos visuales presentes en estos subpaneles seguirán los requisitos propios y comunes de cada panel, tal como se ha mencionado en apartados anteriores. Los objetos visuales estarán sujetos al filtro de procedencia y al filtro temporal, lo que permitirá al usuario manipular estos filtros en cualquier momento para realizar un estudio más detallado de los datos proporcionados por el sistema. Además, se contempla la posibilidad de añadir pestañas dentro de cada página, en caso de que sea necesario mostrar diferentes visualizaciones que pertenezcan al mismo estudio de los datos.

En la Ilustración 9 se puede observar cómo se estructuran los subpaneles de las tablas resultados y gráficos de distribución normal y cómo se implementa la navegación de páginas y pestañas dentro de ellos.

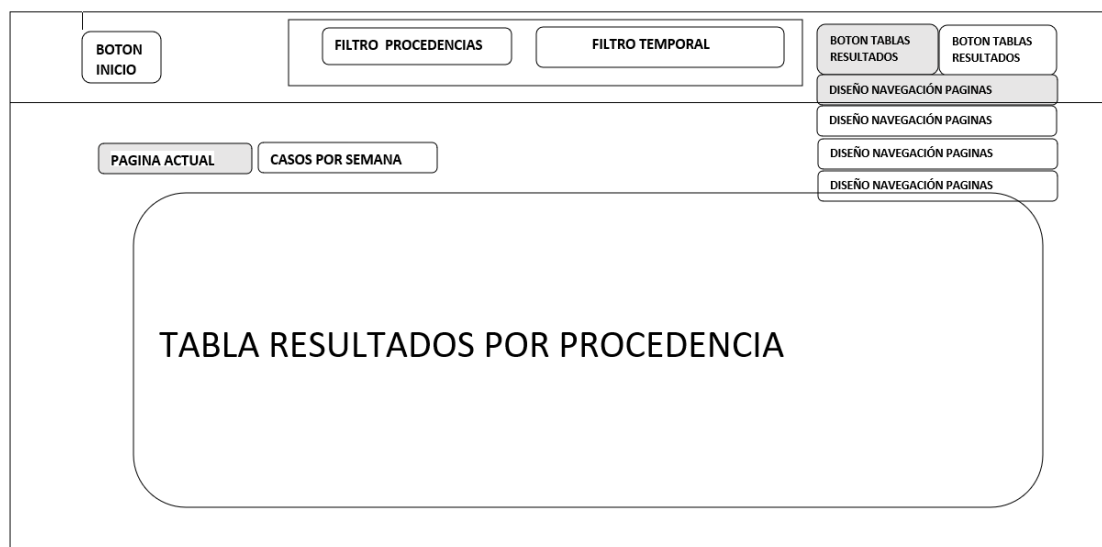


Ilustración 9 Diseño paneles “Tabla Resultados” “Gráficos de Distribución temporal” sistema AccuGest PowerBI.

En resumen, el diseño del panel AccuGest en Power BI trata de proporcionar una estructura sencilla y organizada, en la que prima una navegación intuitiva para el usuario. Los elementos visuales deben estar ligados a los filtros de ubicación y tiempo, lo que permite a los usuarios explorar los datos con mayor detalle y comprender mejor los resultados del sistema AccuGest.

3.2.2.3 Diseño del flujo de datos

El flujo de datos en el sistema AccuGest en Power BI (Ilustración 10) se ha diseñado de manera sencilla y eficiente. Los datos clínicos para estudiar serán extraídos del archivo "TPNI_total2.xls", que actuará como origen de los datos. Power BI ofrece la capacidad de utilizar una amplia variedad de fuentes de datos, en nuestro caso seleccionaremos como fuente de datos una hoja Excel . Además, se utilizará la herramienta nativa Power Query [21] para realizar las transformaciones necesarias en los datos.

El proceso de modelado de datos se llevará a cabo, incluyendo la generación de relaciones entre las diferentes tablas que se utilizarán en el sistema, así como la creación de tablas, columnas y medidas calculadas para el análisis de los datos. Power BI proporciona herramientas no demasiado complejas para realizar estas tareas de modelado y estructuración de los datos.

Una vez que los datos han sido transformados y modelados, se procederá a la representación visual de los mismos utilizando las diversas visualizaciones que ofrece Power BI. Finalmente, el sistema AccuGest PowerBI será desplegado a través de la plataforma online de PowerBI, lo que permitirá un acceso fácil y seguro a los usuarios. La herramienta online de PowerBI ofrece opciones de colaboración y compartición de informes, lo que facilita el intercambio de datos y resultados entre los diferentes usuarios del sistema.

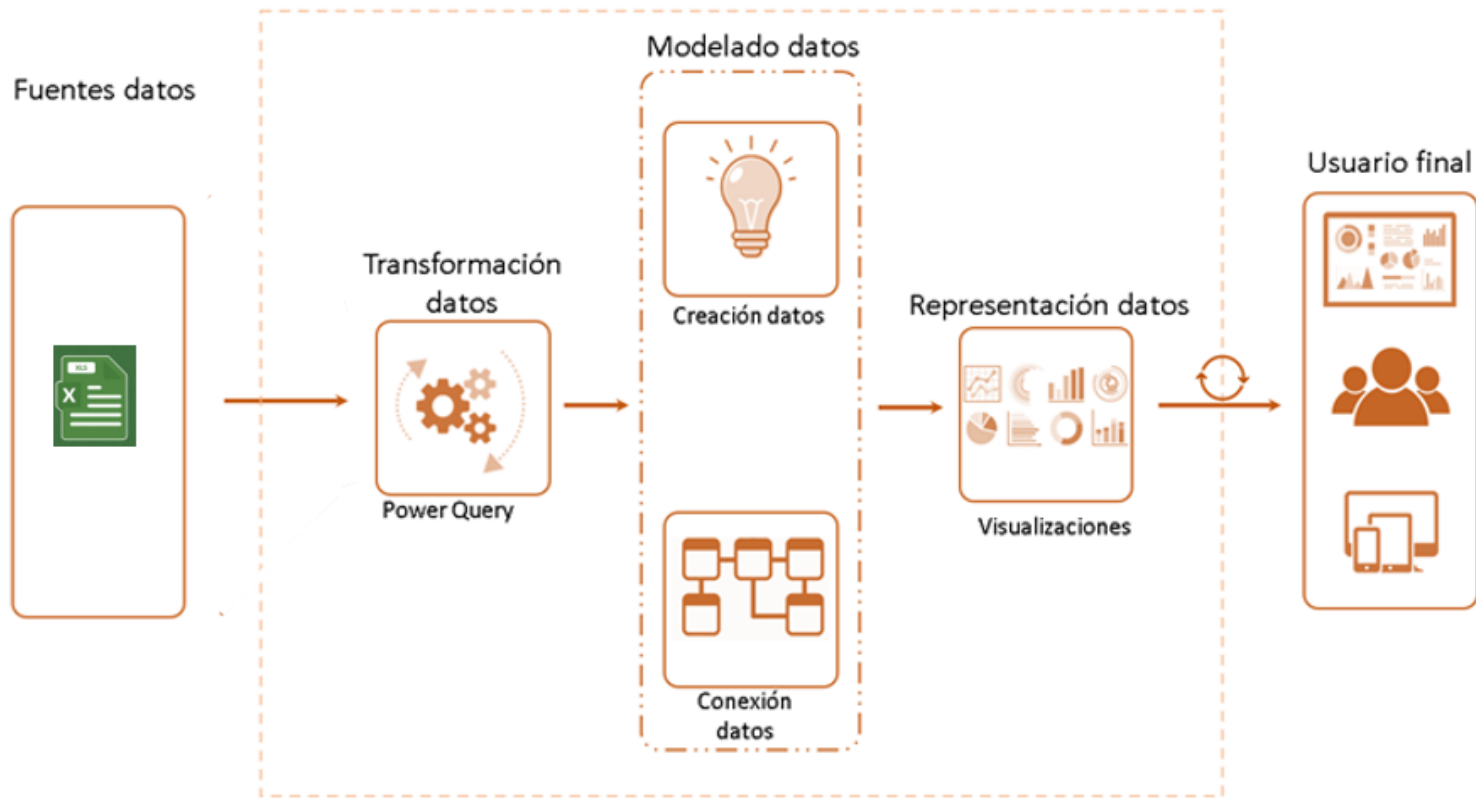


Ilustración 10 Diseño del flujo de datos AccuGest PowerBI.

4 Implementación

4.1 Implementación mejoras en AccuGest en Rstudio/Shiny

4.1.1 Modificación carga de datos

La carga y preparación de los datos se lleva a cabo en el archivo CargaDatos.R, este ha sufrido ligeras modificaciones para poder llevar a cabo la implementación de mejoras en el código fuente de la app. Las líneas añadidas han sido las siguientes:

- Se agregaron las líneas al nuevo código CargaDatos.R que se observan en la Ilustración 11. Estas líneas de código se encargan de crear un nuevo dataframe “lista de anomalías” a partir de los datos de la tabla “tpni_entrada”, eliminando los valores nulos y excluyendo ciertas anomalías específicas. En concreto las anomalías "Bajo riesgo de aneuploidía", "INVALIDATED", "NO CONCLUYENTE". La intención de esta transformación era la de poder entender realmente que se representaba en los gráficos originales, y si se podía prescindir de alguna de las existentes.

```
lista_anomalia <- tpni_entrada %>%
  select(13) %>%
  distinct() %>%
  drop_na()

exclusiones <- c("Bajo riesgo de aneuploidía", "INVALIDATED", "NO CONCLUYENTE")
lista_anomalia_filtrada <- subset(lista_anomalia, !ANOMALIA %in% exclusiones)
```

Ilustración 11 Creación dataframe “lista_anomalia” y “ lista_anomalia_filtrada”.

- Se genera un nuevo dataframe a partir de las líneas de la Ilustración 12 llamado "tiempo_res" a partir de las columnas "PROCEDENCIA", "DIAS_INFORME" y "CANTIDAD" de “tpni_entrada”. Sirve para poder generar la nueva pestaña “tiempo de respuesta” que represente, por centro, la diferencia entre la fecha de solicitud y la fecha de informe.

```
tiempo_res <- tpni_entrada %>%
  select(PROCEDENCIA, DIAS_INFORME, CANTIDAD)
```

Ilustración 12 Creación dataframe “tiempo_res”.

No se precisan más cambios en el archivo CargaDatos.R, ya se consideran suficientes los ya existentes para implementar las nuevas mejoras del sistema AccuGest en Rstudio/Shiny.

4.1.2 Implementación mejoras en script App.R

En un script Shiny, el objeto ui (interfaz de usuario) define la estructura y los componentes visuales de la aplicación web. Especifica cómo se verá y cómo interactuará el usuario con la aplicación “shinyApp(ui, server)”. Por otro lado, el objeto server define el comportamiento y la lógica de la aplicación. Contiene la programación que maneja las interacciones del usuario y los outputs que el usuario llama en el objeto ui.

4.1.2.1 Implementación mejoras objeto UI

Todas las mejoras que se añaden en el código fuente aparecen reflejadas en el apartado “3.2.1 Diseño mejoras en AccuGest en Rstudio/Shiny” por lo que en este apartado nos limitaremos a explicar donde y como se añaden dichas mejoras del objeto ui.

- En primer lugar, como se comenta en el apartado citado, se ajustan los tamaños de los filtros principales que son protagonistas en el panel, el filtro temporal y el filtro de procedencias. Los dos se ajustan de la misma manera, y se toma como ejemplo el filtro temporal (Ilustración 13) para explicar como se realiza la transformación. Se modifica el tamaño del campo “width” del “sliderInput” (objeto visual que facilita la herramienta Shiny para añadir un panel deslizable) para que ocupe el “98%” del tamaño que ocupa el objeto. Se modifica también el campo “style” de manera que el valor 20px establece el espaciado interno a la izquierda y 15px establece el espaciado interno a la derecha. Conseguimos con estos cambios que el filtro deje de estar pegado a los bordes.

```
fluidRow(  
  # slider  
  column(  
    12,  
    sliderInput(  
      "timeSlider",  
      label = "Seleccione fecha",  
      min = min(tpni_entrada$FECHA_RECEPCION),  
      max = max(tpni_entrada$FECHA_RECEPCION),  
      value = c(min(tpni_entrada$FECHA_RECEPCION), max(tpni_entrada$FECHA_RECEPCION)),  
      width = "98%",  
      timeFormat = "%d.%m.%Y",  
      animate = animationOptions(loop = TRUE)  
    ),  
    class = "slider",  
    style = 'padding-left:20px; padding-right:15px',  
    height = "5vh"  
  )  
)  
: _ . . .
```

Ilustración 13 Ajuste de tamaños en filtro temporal.

- Se añaden nuevas pestañas en el objeto ui como se indica en el diseño, todas ellas comparten la misma manera de implementarse, y se usará la implementación de la pestaña “Indicaciones” (Ilustración 14) como ejemplo para explicar la implementación del resto de pestañas, ya que es un ejemplo de creación de una pestaña completa, con un filtro individual asociado al objeto que representa.

```
#AÑADIDO PANEL TABLA INDICACIONES  
tabPanel(  
  "Indicaciones",  
  pickerInput("in_indicaciones", "Filtrar por indicacion:",  
    choices = unique(as.character(lista_indicaciones$INDICACION)),  
    options = pickerOptions(style = "top: auto; bottom: 100%; string",  
      dropdownAlignRight = NULL, selectAllText = NULL),  
    multiple = TRUE,  
    selected = unique(lista_indicaciones$INDICACION)[1:3]),  
  
  dataTableOutput("tabla_indicaciones"),  
  height = "40vh",  
  selected = TRUE  
)  
: _ . . .
```

Ilustración 14 Implementación nueva pestaña indicaciones.

Este panel incluye los siguientes elementos:

- Un filtro de selección (pickerInput), nombrado como "Filtrar por indicación:", que permite al usuario filtrar los datos por diferentes indicaciones. El parámetro "choices" establece las opciones disponibles en el cuadro de selección, obtenidas de la columna "INDICACION" de la variable "lista_indicaciones", y se asegura de que solo aparezcan valores únicos. El parámetro "options" permite personalizar el estilo filtro desplegable. Con el parámetro "múltiple" establecido en "TRUE", se consigue que el filtro sea de selección múltiple. Por último, el parámetro "selected" define las opciones preseleccionadas por defecto, tomando las tres primeras opciones de la columna "INDICACION" de "lista_indicaciones".
- El código "dataTableOutput ("tabla indicaciones")" genera una tabla de datos con el nombre "tabla indicaciones". El parámetro "height = "40vh"" establece la altura de la tabla en un 40% del tamaño de la ventana del navegador (viewport height). Por último, el parámetro "selected = TRUE" indica que esta pestaña estará seleccionada de forma predeterminada al cargar la página. Como todas las pestañas tienen el campo "selected = TRUE" se cargará como primera página la primera declarada.
- El resto de las pestañas implementadas ("Anomalías", "Anomalías detectadas por indicación" y "Tiempo de respuesta") se dan por explicadas porque siguen el mismo proceso de implementación que la pestaña "Indicaciones" o un aún más sencillo, es decir con menos opciones. La única diferencia destacable será que a la hora de generar las pestañas se añade "plotlyOutput" como uno de los parámetros de la función tabPanel() en vez del anteriormente usado "dataTableOutput".

4.1.2.2 Implementación mejoras objeto Server

Actuando de manera coherente con el apartado anterior, las mejoras que se añaden en el código fuente aparecen reflejadas en el apartado "3.2.1 Diseño mejoras en AccuGest en Rstudio/Shiny" por lo que en este apartado nos limitaremos a explicar donde y como se añaden dichas mejoras del objeto server.

Error visual tarjetas resumen: Para solucionar el error visual de tamaño en la cuarta tarjeta resumen ha bastado con sustituir la línea original ("`<h5 style='display:inline'>(%+d)</h5>`") por la de la Ilustración 15 en la lista de figuras que se añaden en el UI.

El cambio resuelve la desalineación al hacer que el contenido de la figura desalineada se expanda verticalmente y ocupe más espacio, alineándose con las demás figuras.

```

"PROCEDENCIA" = HTML(paste(
  format(data$PROCEDENCIA, big.mark = " "),
  "/ 26", |
  sprintf("<h5>(%+d)</h5>", data_new$new_PROCEDENCIA)
))
)

```

Ilustración 15 Ajuste tamaño tarjetas resumen.

- Implementación nueva pestaña “Anomalías” en “Tabla Resultados” (Ilustración 16)

```
#TABLA ANOMALIAS
sidebyside_anomalia2 <- reactive({
  tpmi_a <- sidebyside_anomalia()
  # Transformar la matriz de 2 dimensiones en un marco de datos
  df <- as.data.frame(tpmi_a, stringsAsFactors = FALSE)
  names(df) <- c("PROCEDENCIA", "ANOMALIA", "CANTIDAD")
  # Filtrar por anomalias seleccionadas en el filtro
  df <- df[df$ANOMALIA %in% input$in_anomalia, ]
  return(df)
})

output$tabla_anomalias <- renderDataTable({
  df <- sidebyside_anomalia2()
  if (is.null(df)) return(NULL)

  df <- df %>% select(PROCEDENCIA, ANOMALIA, CANTIDAD)
  datatable(df, options = list(
    order = list(list(3, 'desc')),
    pageLength = 10,
    searching = FALSE,
    lengthChange = FALSE,
    initComplete = JS('function(setting, json) { ; }')
  ))
})
```

Ilustración 16 Implementación del output de la pestaña “Anomalías”.

- Antes de esta implementación, la función `sidebyside_anomalia` ya estaba implementada. Esta función realiza el cálculo de las anomalías utilizando los datos filtrados de entrada y realiza un resumen por procedencia y anomalía, sumando la cantidad correspondiente.
 - La función `sidebyside_anomalia2` se utiliza para transformar la matriz de 2 dimensiones generada por “`sidebyside_anomalia()`” en un dataframe y luego filtrar los datos por las anomalías seleccionadas y añadirlo como input del “`pickerInput`” del “`tabPanel`” del objeto `ui`, de manera que se realiza un filtrado adicional de los datos para incluir solo las filas que correspondan a las anomalías seleccionadas.
 - El output “`tabla_anomalias`” se utiliza como respuesta a la llamada a la `tabla_anomalias` del `tabPanel` “Anomalías”. En este output, se realiza el formateo del dataframe resultante y se utiliza la función “`renderDataTable`” para mostrar la tabla de anomalías. Se aplican opciones diferentes opciones de ordenamiento, longitud de página, búsqueda y cambio de longitud de página.
- Implementación nueva pestaña “Indicaciones” (Ilustración 17) en “Tabla Resultados”:
 - Se repite el mismo proceso, que en la pestaña “Anomalías”, pero utilizando las funciones `sidebyside_indicación()` (anteriormente implementada), y las creadas `sidebyside_indicación2()` y un output “`tabla_indicaciones`” se utiliza como respuesta a la llamada a la `tabla_anomalias` del `tabPanel` “Indicaciones”. Funcionan exactamente de la misma manera, pero en vez de desarrollar las funciones orientándolas hacia la columna “ANOMALIA” se orientan hacia “INDICACION”. Como sigue el mismo funcionamiento, se entiende que la explicación de la implementación queda reflejada en el punto anterior.

```

#TABLA INDICACIONES
sidebyside_indicacion2 <- reactive({
  tpni_a <- sidebyside_indicacion()
  # Transformar la matriz de 2 dimensiones en un marco de datos
  df2 <- as.data.frame(tpni_a, stringsAsFactors = FALSE)
  names(df2) <- c("PROCEDENCIA", "INDICACION", "CANTIDAD")

  df2 <- df2[df2$INDICACION %in% input$in_indicaciones, ]

  return(df2)
})
output$tabla_indicaciones <- renderDataTable({
  df <- sidebyside_indicacion2()
  if (is.null(df)) return(NULL)
  df <- df %>% select(PROCEDENCIA, INDICACION, CANTIDAD)
  datatable(df, options = list(
    order = list(list(3, 'desc')),
    pageLength = 10,
    searching = FALSE,
    lengthChange = FALSE,
    initComplete = JS('function(setting, json) { ; }')
  ))
})

```

Ilustración 17 Implementación del output de la pestaña “Indicaciones”.

- Implementación nueva pestaña “Anomalías por indicación” (Ilustración 18) en “Tabla Resultados”:
 - En primer lugar, se realiza un filtrado de los datos de entrada (utilizando la función `tpni_entrada_filtrado_sin_max_hos`) que ya estaba implementada en el sistema. Esta función devuelve la tabla "tpni_entrada" filtrada para con los valores que hay seleccionados en el filtro de selección de procedencias.
 - A continuación, se agrupan los datos por anomalía e indicación y se realiza un resumen, calculando la suma total de pruebas para cada combinación de anomalía e indicación. Los resultados se ordenan por anomalía y luego por indicación. Después, se filtran los datos por las anomalías seleccionadas y se añaden como input del “pickerInput” del “tabPanel” del objeto ui de forma que se realiza un filtrado adicional de los datos para incluir solo las filas que correspondan a las anomalías seleccionadas.

```

# TASA DE DETECCIÓN ANOMALÍAS
anomalias_tasa_deteccion <- reactive({
  tpni_a <- na.omit(tpni_entrada_filtrado_sin_max_hos()) %>%
  filter(RESPUESTA == "POSITIVO") %>%
  group_by(ANOMALIA, INDICACION) %>%
  summarise(total_pruebas = sum(CANTIDAD))%>%
  arrange(ANOMALIA, INDICACION)
  if (!is.null(input$in_anomalia_por_indicacion)) {
    tpni_a <- tpni_a[tpni_a$ANOMALIA %in% input$in_anomalia_por_indicacion, ]
  }

  return(tpni_a)
})

```

Ilustración 18 Implementación del output de la pestaña “Anomalía por indicación”.

- Mejora pestaña “Anomalías por Hospital” en “Gráficos de distribución temporal”:

La implementación de esta mejora supuso una gran parte del tiempo empleado en la implementación de mejoras del cuadro AccuGest, y por desgracia no se encontraron demasiadas soluciones que mejoraran la interpretación de los datos a través de una nueva visualización. Finalmente, como única solución al problema planteado en el diseño de la pestaña “Anomalías por Hospital”, se opta por aplicar un factor de conversión a los datos de anomalías que se dibujan en la implementación original del output de la pestaña

(Ilustración 19). De forma que, si ANOMALIA es "Bajo riesgo de aneuploidía", entonces IMPR se calcula como PERC multiplicado por 0.1 y para cualquier otro valor de ANOMALIA, se calcula como PERC multiplicado por 1.9. Realizando este pequeño cambio en la implementación original se facilita la visión e interpretación de los datos representados en el objeto visual.

```
tpni_a <- tpni_a %>%
  group_by(PROCEDENCIA) %>%
  mutate(
    IMPR = ifelse(is.na(ANOMALIA), PERC,
                  ifelse(ANOMALIA == "Bajo riesgo de aneuploidía",
                        PERC * 0.1, PERC * 1.9))) %>%
  mutate(IMPR = IMPR / sum(IMPR) * 100)
```

Ilustración 19 Factor de conversión aplicado a anomalías en “Anomalías por hospital”.

- Implementación nueva pestaña “Tiempo de respuesta” (Ilustración 20) en “Gráficos de distribución temporal”:
 - Para la creación del output que recibirá el objeto ui de la shinyApp se define una primera función `sidebyside_respuesta()`. Esta se encarga de filtrar la tabla que se genera con la función `tpni_entrada_filtrado_sin_max_hos()`, para seleccionar las columnas “PROCEDENCIA” y “DIAS_INFORME”. Agrupa la tabla por “PROCEDENCIA”, calcula la media de los valores de “DIAS_INFORME” para cada grupo de “PROCEDENCIA” y ordena la tabla en orden ascendente por la columna “DIAS_INFORME”.
 - Como output de la pestaña, se renderiza un gráfico de barras interactivo en el que se filtran los datos de `sidebyside_respuesta` según las procedencias seleccionada en el filtro. Luego, se utiliza la función “`plot_ly()`” para crear el gráfico de barras, donde se establece “DIAS_INFORME” como eje x y “PROCEDENCIA”, reordenada por “DIAS_INFORME”(los Centros que más días de respuesta tengan aparecerán primero) como eje y. El tipo de gráfico se configura como 'bar' y la orientación como 'h'. Se añaden también anotaciones que dan nombre al título del gráfico y se configuran las propiedades de las fuentes.

```
#DIAS RESPUESTA
sidebyside_respuesta <- reactive({
  tpni_a <- tpni_entrada_filtrado_sin_max_hos() %>%
    select(PROCEDENCIA, DIAS_INFORME) %>%
    group_by(PROCEDENCIA) %>%
    summarise(DIAS_INFORME = mean(DIAS_INFORME))
  # Ordenar de forma ascendente por la media de respuesta
  tpni_a <- tpni_a[order(tpni_a$DIAS_INFORME, decreasing = FALSE), ]
  return(tpni_a)
})
output$tiempo_respuesta <- renderPlotly({
  filtered_data <- sidebyside_respuesta() %>%
    filter(PROCEDENCIA %in% input$in_procedencia)
  plot_ly(data = filtered_data, x = ~round(DIAS_INFORME, digits = 2),
          y = ~reorder(PROCEDENCIA, DIAS_INFORME),
          type = 'bar', orientation = 'h',
          text = ~paste(round(DIAS_INFORME, digits = 2), "días")) %>%
    add_annotatons(text = "Tiempo de respuesta medio por centro",
                  xref = "paper", yref = "paper",
                  x = 0.5, y = 1.1, showarrow = FALSE,
                  font = list(size = 16)) %>%
    layout(list(title = "Días de respuesta por procedencia",
                xaxis = list(title = "Días de respuesta", tickfont = list(size = 12)),
                yaxis = list(title = " "),
                width = 650, height = 350 )
  })
```

Ilustración 20 Implementación nueva pestaña “Días respuesta”.

4.2 Implementación panel AccuGest en Power Bi

4.2.1 Determinar origen de los datos

Debido a que este caso de estudio trata datos clínicos, como estudiante de la Universidad Politécnica de Madrid no tengo acceso a las bases de datos de donde se obtiene la información. La idea principal de este proyecto es la automatizar completamente el nuevo sistema AccuGest creado en PowerBI, y esto será posible en el momento en el que desde el Hospital de Genética del Hospital 12 de Octubre, el responsable del mantenimiento del sistema añada la misma fuente de datos utilizada en el sistema original dentro de la herramienta PowerBI. Se debe añadir el servidor, la base de datos, añadir las credenciales necesarias y por último añadir la tabla que genera el SIL departamental del Servicio de Genética TPNI_total2.xls, que se ha utilizado para la creación de ambos sistemas AccuGest. Toda esta información quedará reflejada en unas instrucciones minuciosamente detalladas alojadas en un archivo PDF (InstruccionesDesplegarAccuGest Power BI) en la carpeta contenedora del proyecto.

En primer lugar, se añadieron como origen de los datos libro de Excel, en concreto el archivo TPNI_total.xls (Ilustración 21).

```
let
    Origen = Excel.Workbook(File.Contents("C:\Users\abfra\OneDrive\Documents\Accugest_10112022\TPNI_total2.xls"), null, true),
```

Ilustración 21 Diseño panel inicio sistema AccuGest PowerBI.

Dicho archivo viene con unos campos de interés previamente seleccionados: Petición, Indicación, Id. Procedencia, Procedencia, Servicio, Fecha de Solicitud, Fecha de Recepción de la muestra, Mes de recepción, Semana de Recepción, Días de recepción (periodo desde la petición hasta la recepción en nuestro Centro), Fecha del Informe, Días en informar, Anomalía, Observaciones y Cantidad. Todos estos datos se cargarán y en ellos se aplicarán unas determinadas transformaciones.

4.2.2 Carga y transformación de los datos en tablas

Este proceso se ha diseñado de manera que se realicen todas las transformaciones necesarias en la tabla principal denominada "tpni_entrada". La tabla "tpni_entrada" se utiliza como base para obtener tablas auxiliares según las necesidades de cada visualización.

Para realizar todo el proceso de ETL (Extract Transform Load)(REF) utilizaremos Power Query , que permite la extracción, transformación y carga de datos. Además, sigue un proceso de recetas (Ilustración 22), en las que se crean unas "instrucciones" que los datos deben atravesar. Las recetas son las instrucciones que Power Query sigue para transformar los datos. Cada vez que se actualicen los datos, la tabla sigue el mismo proceso independientemente de los nuevos datos que contenga la fuente, lo que garantiza la consistencia de los datos.

| DN | FECHA_INFORME | DIAS_INFORME | RESULTADO | ANOMALIA |
|----|---------------|--------------|--|----------------------------|
| 1 | 22/02/2019 | 3 | Cribado negativo de aneuploidias en los cromosomas estudiados. | Bajo riesgo de aneuploidia |
| 2 | 22/02/2019 | 3 | Cribado negativo de aneuploidias en los cromosomas estudiados. | Bajo riesgo de aneuploidia |
| 3 | 22/02/2019 | 3 | Cribado NO CONCLUYENTE en los cromosomas estudiados. | INVALIDATED |
| 4 | 22/02/2019 | 3 | Cribado negativo de aneuploidias en los cromosomas estudiados. | Bajo riesgo de aneuploidia |
| 5 | 22/02/2019 | 3 | Cribado negativo de aneuploidias en los cromosomas estudiados. | Bajo riesgo de aneuploidia |
| 6 | 22/02/2019 | 3 | Cribado negativo de aneuploidias en los cromosomas estudiados. | Bajo riesgo de aneuploidia |
| 7 | 22/02/2019 | 3 | Cribado negativo de aneuploidias en los cromosomas estudiados. | Bajo riesgo de aneuploidia |
| 8 | 22/02/2019 | 3 | Cribado negativo de aneuploidias en los cromosomas estudiados. | Bajo riesgo de aneuploidia |
| 9 | 22/02/2019 | 3 | Cribado negativo de aneuploidias en los cromosomas estudiados. | Bajo riesgo de aneuploidia |
| 10 | 22/02/2019 | 3 | Cribado negativo de aneuploidias en los cromosomas estudiados. | Bajo riesgo de aneuploidia |
| 11 | 22/02/2019 | 3 | Cribado negativo de aneuploidias en los cromosomas estudiados. | Bajo riesgo de aneuploidia |
| 12 | 22/02/2019 | 3 | Cribado negativo de aneuploidias en los cromosomas estudiados. | Bajo riesgo de aneuploidia |

Ilustración 22 Editor Power Query, y receta de los pasos aplicados a los datos.

4.2.2.1 Carga de datos y transformación en tabla principal

- Cambiamos y asignamos a cada columna el tipo de datos al que pertenecen sus datos, para garantizar que los datos se almacenen correctamente como se observa en la Ilustración 23.

```

#"Tipo cambiado" = Table.TransformColumnTypes("#Encabezados promovidos",{{"PETICION", type text}, {"INDICACION", type text}, {"PROCEDENCIA", type text}, {"SERVICIO", type text}, {"FECHA_REGISTRO_EXTERNO", type date}, {"FECHA_RECEPCION", type date}, {"MES_RECEPCION", Int64.Type}, {"SEMANA_RECEPCION", Int64.Type}, {"DIAS_RECEPCION", Int64.Type}, {"FECHA_INFORME", type date}, {"DIAS_INFORME", Int64.Type}, {"RESULTADO", type text}, {"ANOMALIA", type text}, {"CANTIDAD", Int64.Type}}),

```

Ilustración 23 Función Tipo cambiado en "tpni_entrada".

- Se detectan errores en fechas, en primer lugar, se estudia la posibilidad de eliminar esos valores erróneos, pero a partir de la columna "PETICION" en la que sus dos primeros dígitos indican la fecha (XXDFXXXX) en que se añade la fila a la tabla podemos detectar la fecha real en la que se introdujo y se reemplazan valores específicos en la columna "FECHA_REGISTRO_EXTERNO" (Ilustración 24) asegurando consistencia en los datos.

```

#"Valor reemplazado" = Table.ReplaceValue("#Tipo cambiado",#date(2023, 12, 23),#date(2021, 12, 23),Replacer.ReplaceValue,{"FECHA_REGISTRO_EXTERNO"}),
#"Valor reemplazado1" = Table.ReplaceValue("#Valor reemplazado",#date(2024, 12, 23),#date(2021, 12, 23),Replacer.ReplaceValue,{"FECHA_REGISTRO_EXTERNO"}),
#"Valor reemplazado2" = Table.ReplaceValue("#Valor reemplazado1",#date(2025, 12, 23),#date(2021, 12, 23),Replacer.ReplaceValue,{"FECHA_REGISTRO_EXTERNO"}),
#"Valor reemplazado3" = Table.ReplaceValue("#Valor reemplazado2",#date(2022, 12, 23),#date(2021, 12, 23),Replacer.ReplaceValue,{"FECHA_REGISTRO_EXTERNO"}),

```

Ilustración 24 Valores erróneos reemplazados en columna en "FECHA_REGISTRO_EXTERNO".

- Se agrega nueva columna "RESULTADOS" (Ilustración 25) que depende del contenido de la columna "RESULTADO". Esta tarea se realiza fila por fila y se asigna "NEGATIVO" si contiene la palabra "negativo", se asigna "POSITIVO" si contiene la palabra "positivo" o se asigna "NO CONCLUYENTE" si aparece la palabra "concluyente", en caso contrario se deja la fila como nula. Posteriormente eliminamos la columna antigua "RESULTADO" que ya no será de utilidad.

```

#"Columna condicional agregada" = Table.AddColumn("#Valor reemplazado3", "RESULTADOS",
each if Text.Contains([RESULTADO], "negativo") then "NEGATIVO" else if Text.Contains([RESULTADO], "positivo")
then "POSITIVO" else if Text.Contains([RESULTADO], "CONCLUYENTE") then "NO CONCLUYENTE" else null),
#"Columnas quitadas" = Table.RemoveColumns("#Columna condicional agregada",{"RESULTADO"}),

```

Ilustración 25 Creación columna "RESULTADOS" y eliminación columna "RESULTADO".

- Se reorganizan las columnas en el orden deseado para facilitar la visualización y comprensión de los datos, y con la intención de respetar el formato que la tabla tenía en el anterior sistema (Ilustración 26).

```
##"Columnas reordenadas" = Table.ReorderColumns(##"Columnas quitadas",{ "PETICION", "INDICACION", "PROCEDENCIA", "SERVICIO",
"FECHA_REGISTRO_EXTERNO", "FECHA_RECEPCION", "MES_RECEPCION", "SEMANA_RECEPCION", "DIAS_RECEPCION", "FECHA_INFORME", "DIAS_INFORME",
"RESULTADOS", "ANOMALIA", "CANTIDAD"}),
```

Ilustración 26 Reorganización columnas tabla "tpni_entrada".

- Se reemplazan valores específicos en la columna "PROCEDENCIA" y "INDICACION" para estandarizar la nomenclatura. En concreto se transforman todos los valores de en la columna "PROCEDENCIA" cuyo nombre sea "HOSPITAL DEL SURESTE" por "HOSPITAL SURESTE" y "PROYECTO ILLUMINA" por "Proyecto ILLUMINA". Estos errores se producen porque a la hora de realizar la petición hay campos que se generan automáticamente como INDICACION, y otros tienen que introducirse a mano lo que puede generar errores (Ilustración 27).

```
##"Valor reemplazado4" = Table.ReplaceValue(##"Columnas reordenadas","HOSPITAL DEL SURESTE","HOSPITAL SURESTE",Replacer.ReplaceText,{"PROCEDENCIA"}),
##"Valor reemplazado5" = Table.ReplaceValue(##"Valor reemplazado4","PROYECTO ILLUMINA","Proyecto ILLUMINA",Replacer.ReplaceText,{"INDICACION"}),
```

Ilustración 27 Renombre columnas tabla "tpni_entrada".

- Por último, se eliminan las filas que contienen valores en blanco o nulos en cualquiera de sus columnas para obtener datos completos y consistentes (Ilustración 28).

```
##"Filas en blanco eliminadas" = Table.SelectRows(##"Valor reemplazado6", each not List.IsEmpty(List.RemoveMatchingItems(Record.FieldValues(_), {"", null})))
```

Ilustración 28 Filas en blanco eliminadas de la tabla "tpni_entrada".

Aunque PowerBI cuenta con herramientas suficientes como para observar y filtrar todos los elementos de la tabla, copiamos y cargamos la tabla generada en Microsoft Excel (Ilustración 29) y comprobamos que todos los datos sean los correctos. Una vez confirmado cerramos el editor y aplicamos los cambios en Power Query.

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N |
|----|----------|-------------------------------------|--------------------|---------------------------|------------|------------|---|---|---|-----------|---|-----------|----------------------------|---|
| 1 | PETICION | INDICACION | PROCEDENCIA | SERVICIO | FECHA_REGI | FECHA_RE | | | | FECHA | | RESULTAD | ANOMALIA | |
| 2 | 19DF0040 | Cribado Combinado 1º trimestre | HOSPITAL LA PAZ | INGEMM | 11/02/2019 | 19/02/2019 | 2 | 8 | 8 | 22/02/201 | 3 | NEGATIVO | Bajo riesgo de aneuploidía | 1 |
| 3 | 19DF0041 | Cribado Combinado 1º trimestre | HOSPITAL LA PAZ | INGEMM | 11/02/2019 | 19/02/2019 | 2 | 8 | 8 | 22/02/201 | 3 | NEGATIVO | Bajo riesgo de aneuploidía | 1 |
| 4 | 19DF0042 | Cromosomopatía anterior | HOSPITAL LA PAZ | INGEMM | 11/02/2019 | 19/02/2019 | 2 | 8 | 8 | 22/02/201 | 3 | NO CONCLU | INVALIDATED | 1 |
| 5 | 19DF0043 | Cribado Combinado 1º trimestre | HOSPITAL LA PAZ | INGEMM | 14/02/2019 | 19/02/2019 | 2 | 8 | 5 | 22/02/201 | 3 | NEGATIVO | Bajo riesgo de aneuploidía | 1 |
| 6 | 19DF0044 | Cribado Combinado 1º trimestre | HOSPITAL LA PAZ | INGEMM | 14/02/2019 | 19/02/2019 | 2 | 8 | 5 | 22/02/201 | 3 | NEGATIVO | Bajo riesgo de aneuploidía | 1 |
| 7 | 19DF0045 | Cribado Combinado 1º trimestre | HOSPITAL LA PAZ | INGEMM | 13/02/2019 | 19/02/2019 | 2 | 8 | 6 | 22/02/201 | 3 | NEGATIVO | Bajo riesgo de aneuploidía | 1 |
| 8 | 19DF0046 | Cribado Combinado 1º trimestre | HOSPITAL LA PAZ | INGEMM | 13/02/2019 | 19/02/2019 | 2 | 8 | 6 | 22/02/201 | 3 | NEGATIVO | Bajo riesgo de aneuploidía | 1 |
| 9 | 19DF0047 | Edad + Ausencia de IRC | HOSPITAL LA PAZ | INGEMM | 14/02/2019 | 19/02/2019 | 2 | 8 | 5 | 22/02/201 | 3 | NEGATIVO | Bajo riesgo de aneuploidía | 1 |
| 10 | 19DF0048 | Cromosomopatía anterior | HOSPITAL LA PAZ | INGEMM | 14/02/2019 | 19/02/2019 | 2 | 8 | 5 | 22/02/201 | 3 | NEGATIVO | Bajo riesgo de aneuploidía | 1 |
| 11 | 19DF0049 | Cromosomopatía anterior | HOSPITAL LA PAZ | INGEMM | 18/02/2019 | 19/02/2019 | 2 | 8 | 1 | 22/02/201 | 3 | NEGATIVO | Bajo riesgo de aneuploidía | 1 |
| 12 | 19DF0050 | Cribado Combinado 1º trimestre | HOSPITAL LA PAZ | INGEMM | 18/02/2019 | 19/02/2019 | 2 | 8 | 1 | 22/02/201 | 3 | NEGATIVO | Bajo riesgo de aneuploidía | 1 |
| 13 | 19DF0051 | Cribado Combinado 1º trimestre | HOSPITAL LA PAZ | INGEMM | 18/02/2019 | 19/02/2019 | 2 | 8 | 1 | 22/02/201 | 3 | NEGATIVO | Bajo riesgo de aneuploidía | 1 |
| 14 | 19DF0052 | Edad + Ausencia de IRC | HOSPITAL 12 DE OCT | GENETICA | 12/02/2019 | 19/02/2019 | 2 | 8 | 7 | 22/02/201 | 3 | NEGATIVO | Bajo riesgo de aneuploidía | 1 |
| 15 | 19DF0053 | Cribado Combinado 1º trimestre | HOSPITAL 12 DE OCT | GENETICA | 19/02/2019 | 19/02/2019 | 2 | 8 | 0 | 22/02/201 | 3 | NEGATIVO | Bajo riesgo de aneuploidía | 1 |
| 16 | 19DF0054 | Cribado Combinado 1º trimestre | HOSPITAL 12 DE OCT | GENETICA | 15/02/2019 | 19/02/2019 | 2 | 8 | 4 | 22/02/201 | 3 | NEGATIVO | Bajo riesgo de aneuploidía | 1 |
| 17 | 19DF0055 | Contraindicación de prueba invasiva | HOSPITAL 12 DE OCT | GENETICA | 14/02/2019 | 19/02/2019 | 2 | 8 | 5 | 22/02/201 | 3 | NEGATIVO | Bajo riesgo de aneuploidía | 1 |
| 18 | 19DF0056 | Edad + Ausencia de IRC | HOSPITAL 12 DE OCT | GENETICA | 19/02/2019 | 19/02/2019 | 2 | 8 | 0 | 22/02/201 | 3 | NEGATIVO | Bajo riesgo de aneuploidía | 1 |
| 19 | 19DF0057 | Cribado Combinado 1º trimestre | HOSPITAL DE GETAFE | GINECOLOGIA Y OBSTETRICIA | 21/02/2019 | 26/02/2019 | 2 | 9 | 5 | 04/03/201 | 6 | POSITIVO | DETECTED: +21 | 1 |
| 20 | 19DF0058 | Cribado Combinado 1º trimestre | HOSPITAL DE GETAFE | GINECOLOGIA Y OBSTETRICIA | 21/02/2019 | 26/02/2019 | 2 | 9 | 5 | 04/03/201 | 6 | NEGATIVO | Bajo riesgo de aneuploidía | 1 |
| 21 | 19DF0059 | Cribado Combinado 1º trimestre | HOSPITAL DE GETAFE | GINECOLOGIA Y OBSTETRICIA | 21/02/2019 | 26/02/2019 | 2 | 9 | 5 | 04/03/201 | 6 | NEGATIVO | Bajo riesgo de aneuploidía | 1 |
| 22 | 19DF0060 | Cribado Combinado 1º trimestre | HOSPITAL 12 DE OCT | GENETICA | 21/02/2019 | 26/02/2019 | 2 | 9 | 5 | 04/03/201 | 6 | NEGATIVO | Bajo riesgo de aneuploidía | 1 |
| 23 | 19DF0061 | Cribado Combinado 1º trimestre | HOSPITAL 12 DE OCT | GENETICA | 21/02/2019 | 26/02/2019 | 2 | 9 | 5 | 04/03/201 | 6 | NEGATIVO | Bajo riesgo de aneuploidía | 1 |
| 24 | 19DF0062 | Edad + Ausencia de IRC | HOSPITAL 12 DE OCT | GENETICA | 19/02/2019 | 26/02/2019 | 2 | 9 | 7 | 04/03/201 | 6 | NEGATIVO | Bajo riesgo de aneuploidía | 1 |
| 25 | 19DF0063 | Cribado Combinado 1º trimestre | HOSPITAL LA PAZ | INGEMM | 20/02/2019 | 26/02/2019 | 2 | 9 | 6 | 04/03/201 | 6 | NEGATIVO | Bajo riesgo de aneuploidía | 1 |
| 26 | 19DF0064 | Cribado Combinado 1º trimestre | HOSPITAL LA PAZ | INGEMM | 22/02/2019 | 26/02/2019 | 2 | 9 | 4 | 04/03/201 | 6 | NEGATIVO | Bajo riesgo de aneuploidía | 1 |
| 27 | 19DF0065 | Cribado Combinado 1º trimestre | HOSPITAL LA PAZ | INGEMM | 22/02/2019 | 26/02/2019 | 2 | 9 | 4 | 04/03/201 | 6 | NEGATIVO | Bajo riesgo de aneuploidía | 1 |

Ilustración 29 Tabla "tpni_entrada" tras carga y transformación de los datos.

4.2.2.2 Creación de tablas auxiliares

Se necesita utilizar la tabla principal “tpni_entrada” para poder representar los datos de las tarjetas resumen, el mapa con los datos de las pruebas resumidos y tablas resultado. Para poder generar correctamente estos datos y poder generar más tarde tablas auxiliares se opta por generar una nueva tabla “RESULTADOS” a partir de un duplicado de la tabla “tpni_entrada” a la que se le aplica unas determinadas transformaciones que se explican a continuación.

Junto a la creación de la tabla “RESULTADOS” crearemos varias tablas auxiliares:

- Tabla RESULTADOS: A partir de un duplicado de la tabla “tpni_entrada” junto a una transformación en la que seleccionamos exclusivamente las columnas que consideramos necesarias para crear posteriormente las respectivas visualizaciones generamos la nueva tabla “RESULTADOS” (Ilustración 30).

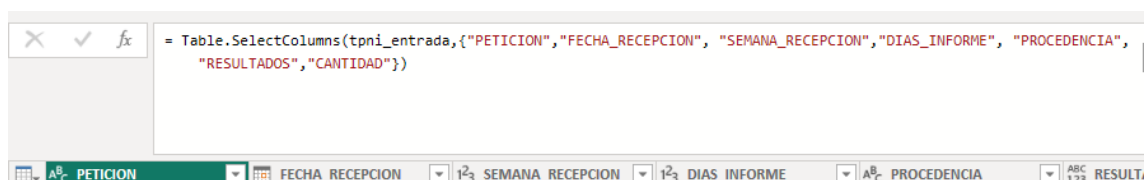


Ilustración 30 Nueva tabla “RESULTADOS” con filtrado de columnas.

- Tabla ANOMALIAS: Se crea una nueva tabla calculada a partir del lenguaje DAX, el código utiliza la función CALCULATETABLE para generar una tabla partir de la tabla original "tpni_entrada" y unas determinadas condiciones (Ilustración 31):

```
1 ANOMALIAS =
2   CALCULATETABLE (
3     SUMMARIZE (
4       FILTER (
5         tpni_entrada,
6         tpni_entrada[ANOMALIA] <> "" &
7         tpni_entrada[FECHA_RECEPCION]
8       ),
9       tpni_entrada[FECHA_RECEPCION],
10      tpni_entrada[PROCEDENCIA],
11      tpni_entrada[ANOMALIA],
12      "CANTIDAD", SUM ( tpni_entrada[CANTIDAD] )
13    ),
14    ALLSELECTED ( tpni_entrada )
15  )
16
17
```

Ilustración 31 Nueva tabla auxiliar “ANOMALIAS” a partir de fórmula en lenguaje DAX.

- En primer lugar, la función FILTER filtra las filas de la tabla "tpni_entrada" que cumplen dos condiciones, la primera condición es que el campo "ANOMALIA" no esté vacío (" $<> ''$ "), lo que significa que solo se seleccionarán las filas que tienen información de anomalías que son válidas y no están vacías.

- La segunda condición es que el campo "FECHA_RECEPCION" tenga un valor, lo que implica que solo se considerarán las filas que tengan una fecha de recepción. Con esto aseguramos que los cálculos sean correctos y no se utilicen filas que no contienen valores para hacer los cálculos.
 - La función SUMMARIZE realiza un resumen de los datos filtrados. En esta función, se especifican las columnas que se desean incluir en la tabla auxiliar, que son "FECHA_RECEPCION", "PROCEDENCIA", "ANOMALIA" y una columna calculada llamada "CANTIDAD" que contiene la suma de los valores en la columna "CANTIDAD". La función ALLSELECTED se utiliza para mantener el contexto de selección en la tabla original al realizar los cálculos en la tabla "ANOMALIAS".
- Tabla INDICACIONES: Esta tabla se calcula de forma similar a la anterior. Esto tiene sentido ya que se pretende obtener un mismo tipo de visualizaciones, pero en base a otra columna, la columna INDICACIONES. El funcionamiento de la fórmula en DAX será exactamente el mismo que en la anterior tabla (Ilustración 32).

```

1 INDICACIONES =
2   CALCULATETABLE (
3     SUMMARIZE (
4       FILTER (
5         tpni_entrada,
6         tpni_entrada[INDICACION] <> "null" &
7         tpni_entrada[FECHA_RECEPCION]
8       ),
9       tpni_entrada[FECHA_RECEPCION],
10      tpni_entrada[PROCEDENCIA],
11      tpni_entrada[INDICACION],
12      "CANTIDAD", SUM ( tpni_entrada[CANTIDAD] )
13    ),
14    ALLSELECTED ( tpni_entrada )
15  )
16
17

```

Ilustración 32 Nueva tabla auxiliar "INDICACIONES" a partir de fórmula en lenguaje DAX.

- Tabla ANOMALIA_INDICACION: Por último, se crea la última tabla auxiliar, esta utiliza de nuevo la misma función DAX que las usadas en las anteriores, añadiendo en esta consulta las dos columnas "ANOMALIA" E "INDICACION"(Ilustración 33).

```

1 ANOMALIA_INDICACION = CALCULATETABLE (
2   SUMMARIZE (
3     FILTER (
4       tpni_entrada,
5       tpni_entrada[ANOMALIA] <> "" &
6       tpni_entrada[FECHA_RECEPCION]
7     ),
8
9     tpni_entrada[FECHA_RECEPCION],
10    tpni_entrada[PROCEDENCIA],
11    tpni_entrada[ANOMALIA],
12    tpni_entrada[INDICACION],
13    "CANTIDAD", SUM ( tpni_entrada[CANTIDAD] )
14  ),
15  ALLSELECTED ( tpni_entrada )
16 )
17

```

Ilustración 33 Nueva tabla auxiliar "ANOMALIA_INDICACION" a partir de fórmula en lenguaje DAX.

Se diferencia entre la tabla "RESULTADOS" y el resto de las tablas auxiliares ya que estas últimas se generan a partir del lenguaje DAX, y la tabla "RESULTADOS" se genera a partir de POWER QUERY.

4.2.3 Creación nuevo panel AccuGest

4.2.3.1 Creación de tablas cuadros de texto y botones

Como parte de las buenas prácticas en el diseño de visualizaciones de datos, se ha optado por crear tablas de títulos en el proceso de implementación de los cuadros de texto y botones. A pesar de existir la posibilidad en PowerBI de añadir en la propia herramienta títulos por cada botón o título de una visualización, se opta por la creación de estas tablas. Si bien no parece una práctica útil para el rango de público que tiene este cuadro de mandos, si lo será si en algún momento se decide expandir, y surge la necesidad de querer traducir los cuadros de mandos a otros idiomas.

| | A | B | C | D | E | F |
|----|----------|----|----------------|----------------------------------|-----------|--------|
| 1 | ACCUGEST | id | id_texto | dim | idioma_id | idioma |
| 2 | ACCUGEST | 1 | INI_ACCUGEST | DATOS TPNI. COMUNIDAD DE MADRID | 1 | es |
| 3 | ACCUGEST | 2 | INI_PANEL_1 | TOTAL PRUEBAS | 1 | es |
| 4 | ACCUGEST | 3 | INI_PANEL_2 | NO CONCLUYENTE | 1 | es |
| 5 | ACCUGEST | 4 | INI_PANEL_3 | POSITIVO | 1 | es |
| 6 | ACCUGEST | 5 | INI_PANEL_4 | PROCEDENCIAS | 1 | es |
| 7 | ACCUGEST | 6 | INI_PANEL_5 | Tablas Resultados | 1 | es |
| 8 | ACCUGEST | 7 | INI_PANEL_6 | Gráficos distribucion temporal | 1 | es |
| 9 | ACCUGEST | 8 | RES_PANEL_1 | RESULTADOS POR PROCEDENCIA (POS | 1 | es |
| 10 | ACCUGEST | 9 | DESPLA_PANEL_1 | SELECCIONE PROCEDENCIA | 1 | es |
| 11 | ACCUGEST | 10 | TABLA_PANEL_4 | ANOMALÍAS DETECTADAS POR INDICAC | 1 | es |

Ilustración 34 Tabla Cuadros de Texto.

En primer lugar, la tabla “CuadrosdeTexto” como se puede observar en la (Ilustración 34) está diseñada de manera que contenga una única columna clave (“id”) que la identifique de manera única. Además, se añade una columna (“id_texto”) para ubicar de forma sencilla el titulo (“dim”). Por último, como se ha comentado y como propósito de añadir títulos teniendo en cuenta futuras traducciones, se añaden dos últimas columnas “idioma_id” y “idioma”, en caso de querer cambiar de idioma únicamente deberían añadirse las mismas filas que ya existen, pero con el campo “dim” traducido al idioma deseado y un número asociado a otro idioma (por ejemplo, en ingles el campo “idioma_id” será 2 y el campo “idioma” será “en”). Se añade a demás una tabla “DimIdioma” con los posibles idiomas a traducir. Se deja implementado en el nuevo panel la arquitectura para una posible traducción futura, en la que únicamente habrá que añadir las nuevas traducciones en la tabla “CuadrosdeTexto” como ya se ha indicado y seleccionar un idioma en el campo “Filtros de todas las páginas” como se observa en la Ilustración 35.

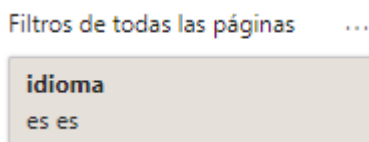


Ilustración 35 Traducción páginas.

En segundo lugar, se crea una tabla “Botones” (Ilustración 36) que sirve para dar título a todos los botones del nuevo panel y que tiene un formato diferente al resto de tablas anteriores. El identificador de cada título de un botón es el nombre de cada columna, en ella se encuentra un título único para cada botón. De forma similar a la explicada en párrafos anteriores, en caso de necesitar traducir el panel, bastaría con añadir esta vez una nueva fila en la tabla “Botones”, cambiando el campo “id_Idioma” por el idioma deseado (es, en...) y cambiar el contenido de cada columna identificativa de títulos por la traducción al idioma deseado.

| | A | B | C | D | E | F | G | H | I |
|---|-----------|------------|-----------|------------|-----------|-----------|-----------|-----------|-----------|
| 1 | id_Idioma | Button_Ini | Button_Ta | Button_Fil | Button_Gr | Button_Ve | Button_An | Button_In | Button_An |
| 2 | es | INICIO | RESULTAD | SELECCION | GRÁFICOS | VER MAPA | TABLA AN | TABLA INC | ANOMALIA |

Ilustración 36 Tabla Botones.

4.2.3.2 Creación de tablas de dimensiones

Continuando las buenas prácticas recomendadas en el diseño de visualizaciones de datos, esto ha dado lugar a la creación de tablas de dimensiones. Estas tablas de dimensiones se utilizan para mejorar la eficiencia en el uso de filtros y se consideran elementos fundamentales para lograr visualizaciones de calidad. Han sido diseñadas para representar aspectos clave de la información, permitiendo una exploración y filtrado eficiente de los datos aportando mayor versatilidad al estudio de los datos por parte del usuario.

A continuación, se detallará la creación de las seis tablas de dimensiones principales: “CalendarioAccuGest”, “Dim_Anomalías”, “Dim_Hospitales”, “Dim_Indicaciones”, “Lat&Long” y la ya comentada en el anterior punto DimIdioma, cuyo funcionamiento ya ha sido descrito

- CalendarioAccuGest: Se crea tabla de dimensiones (figura ()) contiene información relacionada con todas las posibles fechas desde que, incluyendo los campos Fecha, Año, Mes, FechaMes, NumAñoMes, Semana. Permite realizar análisis y filtrados basados en el tiempo de manera efectiva. Además, se añaden columnas que no se usan en el panel actual como la columna “Mes”, pero pueden ser de utilidad en futuras mejoras del sistema.

Esta tabla se añade a PowerBI a partir de un archivo “.csv”, generado mediante un script (Calendario.py) creado en Python, se muestra y se explica en la Ilustración 37 en el propio código. Tras ejecutarlo se carga la tabla en la herramienta como nuevo origen de datos, específicamente como Texto o CSV.

```

1 import csv
2 import sys
3 from datetime import datetime, timedelta
4 import locale
5 import calendar
6
7 # Establecer el idioma local a español
8 locale.setlocale(locale.LC_TIME, 'es_ES')
9 # Obtener los argumentos de la línea de comandos
10 csv_filename = sys.argv[1]
11 end_date_str = sys.argv[2]
12 # Convertir la fecha final a un objeto datetime
13 end_date = datetime.strptime(end_date_str, '%d/%m/%Y')
14 # Establecer la fecha inicial en el 01/01/2019
15 current_date = datetime(2019, 1, 1)
16 # Abrir el archivo CSV para escribir en él
17 with open(csv_filename, 'w', newline='') as csv_file:
18     # Crear el objeto escritor CSV
19     csv_writer = csv.writer(csv_file)
20     # Escribir la primera fila de encabezado
21     csv_writer.writerow(['Año', 'Mes', 'Fecha', 'Fecha Mes', 'NúmeroMes'])
22     # Loop a través de todos los días desde el 01/01/2019 hasta la fecha final
23     while current_date <= end_date:
24         # Obtener el nombre del mes en español
25         mes = calendar.month_name[current_date.month]
26         # Escribir la fila para la fecha actual
27         csv_writer.writerow([
28             current_date.year,
29             current_date.month,
30             current_date.strftime('%d/%m/%Y'),
31             current_date.strftime('%A, %d de {} de %Y').format(mes),
32             current_date.strftime('%Y%m')
33         ])
34         # Avanzar al siguiente día
35         current_date += timedelta(days=1)
36

```

Ilustración 37 Script creación CalendarioAccuGest.csv.

- Dim_Anomalias: Se crea una nueva tabla dimensión (Ilustración 38), que contiene únicamente dos columnas, una “id_ANOMALIA” que sirve como identificador exclusivo para cada anomalía, y otra “ANOMALIA” que será la información que aparece en los filtros con los nombres de cada anomalía. La columna “id_ANOMALIA” se relacionará con las tablas auxiliares y la tabla “RESULTADOS”, y se explicará de que forma en siguientes apartados. La creación de esta dimensión proporciona una clasificación detallada que permite identificar y analizar las anomalías presentes en el sistema.

| id_ANOMALIA | ANOMALIA |
|-------------|----------------------------|
| 1 | Bajo riesgo de aneuploidía |
| 2 | INVALIDATED |
| 3 | DETECTED: +21 |
| 4 | DETECTED: XOY |
| 5 | DETECTED: XO |
| 6 | DETECTED: +13 |

Ilustración 38 Tabla Dim_Anomalias.

- Dim_Indicaciones, Dim_Hospitales: Se generan de forma idéntica a la anterior dimensión cambiando los nombres de la columna anterior “ANOMALIA” por los nombres de cada indicación u hospital existentes y asignándoles de igual forma un identificador exclusivo que servirá para relacionarlo con las tablas de datos del panel.
- Lat&Long: Se genera una tabla que une los diferentes Centros públicos hospitalarios de la comunidad de Madrid como se observa en la Ilustración 39, junto a su latitud y longitud añadidas manualmente con Google Maps [22]. Se aplican unas sencillas transformaciones, de forma similar a como se hizo en anteriores transformaciones en tablas, es decir utilizando el método de pasos aplicados que PowerBI ofrece. Estas transformaciones tienen como objetivo separar la latitud y longitud para su posterior inserción en los campos de un objeto visual llamado “Mapa Interactivo”

| id_procedencia | PROCEDENCIA | LAT&LONG | Latitud | Longitud |
|----------------|---------------------------|---|------------------|---------------------|
| 1 | HOSPITAL LA PAZ | 40.481255357301315, -3.6862255308589864 | 40,4812553573013 | -3.6862255308589864 |
| 2 | HOSPITAL 12 DE OCTUBRE | 40.376193336069946, -3.6990067942246108 | 40,3761933360699 | -3.6990067942246108 |
| 3 | HOSPITAL DE GETAFE | 40.31275714909554, -3.7423014341987977 | 40,3127571490955 | -3.7423014341987977 |
| 4 | HOSPITAL GREGORIO MARAÑÓN | 40.41967166865756, -3.671386910131402 | 40,4196716686576 | -3.671386910131402 |

Ilustración 39 Tabla Lat&Long.

4.2.3.3 Administración de relaciones entre tablas

Las relaciones entre tablas en Power BI desempeñan un papel crucial en la creación de modelos de datos eficientes y en la correcta generación de visualizaciones. Estas relaciones permiten establecer conexiones lógicas entre las tablas, lo que facilita la combinación y el análisis de datos de diferentes fuentes.

En este apartado vamos a explicar como se ha afrontado la relación entre las distintas tablas de nuestro sistema. Como contamos con una sola fuente de datos principal no han surgido apenas problemas en la relación entre esta y el resto de las tablas. La forma que Power BI tiene para relacionar tablas entre sí, es muy sencilla, y se divide en páginas. Una primera página común a todas las relaciones en las que se encuentran todas las tablas y sus relaciones, y un número ilimitado de páginas en las que simplificar estas relaciones. Desde un primer momento se ha buscado aplicar al diseño del modelado de datos un modelo de estrella. Este tipo de modelo simplifica y optimiza las consultas permitiendo que la elaboración de informes y la realización de un análisis en los datos sea mucho más sencillo. Además, facilita la comprensión y navegación de los datos, mejorando la experiencia del usuario al explorar la información. A continuación, se explica como se han relacionado las distintas tablas:

- Tabla “RESULTADOS”: La relación de la tabla “RESULTADOS” (Ilustración 30), generada como un duplicado, y a la que se le realizaron unos determinados cambios anteriormente descritos en el apartado “4.2.2.2 Creación de modelos auxiliares”, se relaciona con la tabla “Dim_Hospitales” de forma “Varios a Uno” entre las siguientes columnas:
 - “RESULTADOS [codProcedencia]”- Dim_Hospitales[id_procedencia]”.

La columna codProcedencia se crea a partir de una columna calculada con una función en DAX como se observa en la Ilustración 40, en la que se asocian “id_procedencia” exclusivos de cada Centro de la tabla “Dim_Hospitales” a partir de la función SWITCH. Esta función busca en la columna que se le pasa como parámetro los diferentes campos que se le pasan como parámetros y en caso de encontrarlo le asigna un número. Recordemos que la función de relacionar estas tablas se explica en el apartado anteriormente mencionado.

```

1 codProcedencia =
2 SWITCH (
3     RESULTADOS[PROCEDENCIA],
4     "HOSPITAL LA PAZ", 1, "HOSPITAL 12 DE OCTUBRE", 2, "HOSPITAL DE GETAFE", 3, "HOSPITAL GREGORIO MARAÑÓN", 4,
5     "HOSPITAL DEL TAJO", 5, "HOSPITAL INFANTA LEONOR", 6, "HOSPITAL EL ESCORTAL", 7, "HOSPITAL SANTA CRISTINA", 8,
6     "HOSPITAL PUERTA DE HIERRO", 9, "HOSPITAL GOMEZ ULLA", 10, "HOSPITAL INFANTA CRISTINA", 11, "HOSPITAL DEL HENARES", 12,
7     "HOSPITAL SEVERO OCHOA (LEGANES)", 13, "HOSPITAL INFANTA SOFIA", 14, "HOSPITAL DE MOSTOLES", 15,
8     "HOSPITAL DE FUENLABRADA", 16, "HOSPITAL PRINCIPE DE ASTURIAS", 17, "H.U.FUNDACION ALCORCON", 18,
9     "HOSPITAL CLINICO SAN CARLOS", 19, "HOSPITAL SURESTE", 20, "HOSPITAL UNIVERSITARIO DE TORREJON", 21,
10    "HOSPITAL RAMON Y CAJAL", 22, "HOSPITAL DEL SURESTE", 23,
11    BLANK ()
12 )

```

Ilustración 40 Generación de columna codProcedencia.

Además, cuenta con otra relación entre “Varios a Varios” en ambas direcciones. Inicialmente esta relación esta desactivada y si activará manualmente en la creación de las medidas que se explicarán en el siguiente apartado. Esta relaciona las columnas:

- “RESULTADOS[FECHA_RECEPCION]” – “CalendarioAccuGest [Fecha]”

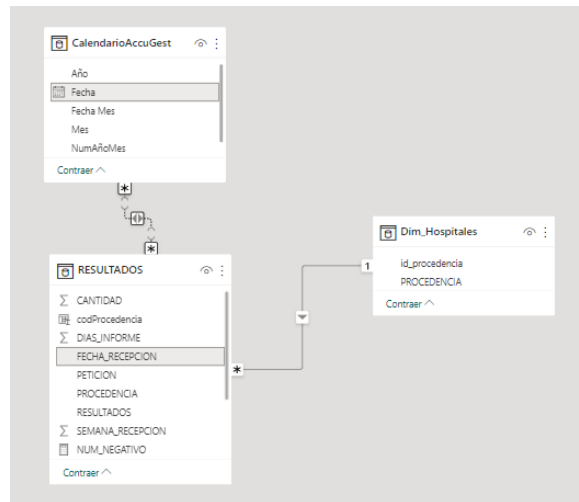


Ilustración 41 Relaciones en la tabla “RESULTADOS”.

- Relación tabla “ANOMALIAS”, “INDICACIONES” e “ANOMALIAS_INDICACIONES”: La relación de las tres tablas auxiliares funciona de la misma manera. Como ejemplo para explicar como funcionan las relaciones entre las tres tablas se va a explicar la relación de tabla “ANOMALIAS” (Ilustración 42) con las tablas “Dim_Hospitales”, “Dim_Anomalias” y “CalendarioAccuGest”. La única diferencia que existe con las relaciones de la tabla “RESULTADOS” es que además se añade una relación con la dimensión creada para anomalías. Funciona exactamente de la misma forma que funcionaba la relación entre “RESULTADOS” y “Dim_Hospitales”, se crea una columna “codProcedenciaAnom” adicional y otra columna generada “codAnomalia” con la misma fórmula DAX usada anteriormente, pero asignando los números de asociados de anomalías de la tabla “Dim_Anomalias” para poder relacionarlas entre sí.

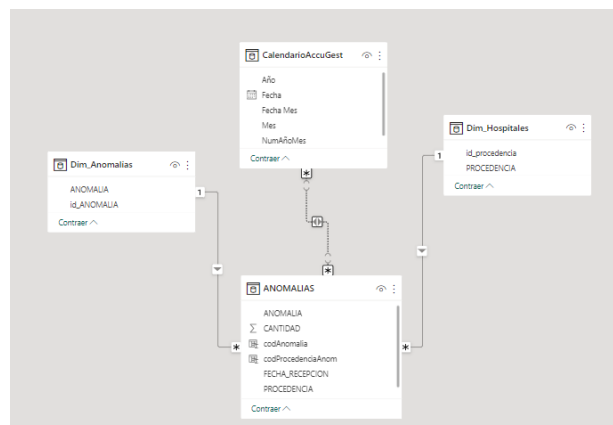


Ilustración 42 Relaciones en la tabla “ANOMALIAS”.

4.2.3.4 Implementación del panel de visualizaciones según el diseño

Para realizar la implementación del nuevo sistema AccuGest se ha seguido el diseño anteriormente realizado en el apartado “3.2.2.2 Diseño nuevo panel AccuGest”, para poder realizar una explicación de como se ha implementado se irán contrastando con las diferentes figuras que aparecen a lo largo del apartado. Además, se explicarán las medidas generadas por cada tabla a la vez que se muestran las interpretaciones en cada uno de los siguientes apartados y no se expondrán en un apartado diferente con la intención de que se entienda perfectamente el proceso seguido para la creación de las nuevas visualizaciones.

Los títulos de todas las tarjetas, botones y objetos visuales se añaden utilizando las tablas “Botones” y “CuadrosdeTexto” siguiendo el propósito ya mencionado anteriormente. A continuación, se presentan las nuevas visualizaciones del sistema AccuGest.

4.2.3.4.1 Implementación panel Inicio

Como se puede observar en las figuras () el panel “Inicio” del sistema AccuGest cuenta dos pestañas distintas, la primera expone un resumen a grande escala de los resultados obtenidos a partir de la fuente de los datos del sistema. La manera de hacerlo es a partir de tarjetas resumen como se explica profundamente en el apartado “3.2.2.2 Diseño nuevo panel AccuGest”.

La manera de obtener estas tarjetas ha sido a partir de medidas, una por cada tarjeta. Una medida no es más que una expresión calculada que proporciona un valor basado en los datos existentes de una tabla. Cada visualización es una tarjeta, y en el campo de cada una de las tarjetas se añade una de las medidas. Se utilizan cuatro medidas distintas

- Medida “TOTAL_PRUEBAS” (Ilustración 43): Calcula la suma de la columna CANTIDAD de la tabla RESULTADOS, activando la relación entre las columnas FECHA_RECEPCION de RESULTADOS y Fecha de la tabla CalendarioAccuGest utilizando para ello la fórmula USERRELATIONSHIP que, para evitar problemas de ambigüedad entre las relaciones de distintas tablas, activa la relación de forma puntual solo para calcular el valor de la medida.

```
1 TOTAL_PRUEBAS = CALCULATE(SUM(RESULTADOS[CANTIDAD]),USERRELATIONSHIP(RESULTADOS[FECHA_RECEPCION],CalendarioAccuGest[Fecha]))
```

Ilustración 43 Medida “TOTAL_PRUEBAS” generada mediante DAX.

- Medida “NUM_POSITIVO”, “NUM_NEGATIVO” y “NUM_NOCONCLUYENTE”: La medida NUM_POSITIVO calcula la suma de la columna CANTIDAD de la tabla RESULTADOS, filtrando solo aquellos registros donde la columna RESULTADOS es igual a "POSITIVO", y utiliza la relación entre las columnas FECHA_RECEPCION de RESULTADOS y Fecha de la tabla CalendarioAccuGest. Las tres se calculan de forma similar, solo se debe cambiar el resultado a filtrar en la consulta (“NEGATIVO” o “NO CONCLUYENTE”).

En la segunda pestaña del panel “Inicio” se selecciona como objeto visual un mapa interactivo. Para que se genere la visualización que teníamos pensada

diseñar, rellenamos los campos del objeto visual. En el campo “Latitud” añadimos la columna “Lat&Long [Latitud]” y en el campo “Longitud” se añade la columna “Lat&Long [Longitud]” de la misma forma. Esto genera un mapa lleno de burbujas, que representan cada una de las procedencias que se registran en el sistema. Además, a las burbujas se les asigna un tamaño a través del campo “Tamaño de burbuja” utilizando la medida calculada anteriormente “TOTAL_PRUEBAS”. Como se puede observar en la Ilustración 50 cada burbuja tiene un color, esto se debe a que en el campo “Leyenda” se añade la columna PROCEDENCIAS de la tabla dimensión “Dim_Hospitales” que al estar relacionada con la tabla “RESULTADOS” genera el mismo resultado que si añadiéramos directamente la columna “RESULTADOS[PROCEDENCIA]”.

Además, el panel como se planteó en el diseño se filtra por un slider de todas las FECHAS_RECEPCION registradas en la fuente de datos, y permite al usuario un estudio más personalizado de las visualizaciones. Este filtro tiene como campo la columna “Fecha” de la tabla “CalendarioAccuGest” relacionada con la tabla “RESULTADOS”.

Como puede observarse en la “Ilustración 50” se ha incorporado un tooltip personalizado que aparece al posicionar el cursor en cualquiera de las burbujas generado a base de tarjetas en una página independiente a las que se les da un formato de diseño. El tooltip intenta recrear la información de la página anterior, pero mostrando la información exclusiva de cada Centro. Para añadir esta nueva página “tooltip_mapa” debemos seleccionar en el campo “Información sobre herramientas” la propia página.

Añadir tooltips a las visualizaciones proporciona una forma interactiva de mostrar información adicional sin ocupar espacio adicional en la pantalla, lo que mejora la experiencia del usuario.

Si pulsamos sobre el botón situado en la parte superior derecha con un icono personalizado de una tabla, accedemos al siguiente panel.

4.2.3.4.2 Implementación panel Tabla Resultados

Tras acceder al nuevo panel “Tabla Resultados” llegamos a la primera de un conjunto de páginas que construirán este segundo subpanel del nuevo sistema AccuGest. En primer lugar, puede observarse en la figura () aparece en la esquina superior derecha una navegación de páginas creadas utilizando numerosos botones, con el objetivo de poder navegar de forma rápida e intuitiva por las distintas páginas del panel “Tabla Resultados”. Hay que remarcar que en cualquier momento se podrá volver al panel “Inicio” sencillamente pulsando sobre el botón “Inicio” situado en la esquina superior izquierda. El panel está dividido en cuatro páginas implementadas:

- La primera página de este panel llamada “RESULTADOS PRUEBAS” tiene a su vez dos pestañas:
 - Una primera pestaña “RESULTADOS TOTALES” que cuenta con un nuevo objeto visual, en concreto una tabla generada con las mismas medidas calculadas en el anterior panel, y que

se pueden consultar como han sido creadas en el anterior apartado “4.2.3.4.2 implementación panel Inicio”. Utiliza como campos la columna Dim_Hospitales [PROCEDENCIA], y las tres medidas “NUM_NEGATIVO”, “NUM_NOCONCLUYENTE” y “NUM_POSITIVO”.

Esta tabla tiene como propósito replicar la Tabla Rank del sistema original, y se ordena descendientemente por los Centros con más pruebas realizadas.

- En la segunda pestaña del panel Tabla Resultados, aparece la ya existente visualización “CASOS POR SEMANA” del anterior sistema implementado en Rstudio/ Shiny. Esta nueva implementación de la visualización cuenta con varias mejoras desde un punto de vista analítico. Se implementa nuevamente añadiendo como campos de un gráfico de columnas agrupadas las medidas “TOTAL_PRUEBAS”, “NUM_NEGATIVO”, “NUM_NOCONCLUYENTE” y “NUM_POSITIVO”. Esta página cuenta con dos filtros personalizados, uno que permite filtrar los casos por semana por años y otro tipo de slider que permite seleccionar el rango deseado de semanas.
- La segunda página titulada “TABLA ANOMALIAS” representa un nuevo objeto visual, en concreto una tabla con el desglose de las anomalías detectadas por Centro. Esta compuesta por tres columnas generadas a su vez por las columnas Dim_Hospitales [PROCEDENCIA] relacionada con la tabla auxiliar “ANOMALIAS”, Dim_Anomalias [ANOMALIA] relacionada también con la misma tabla auxiliar y por último la columna “CANTIDAD” de la tabla auxiliar “ANOMALIAS”, en concreto los datos se resumen como “suma de ANOMALIAS[CANTIDAD]”. Este modo de resumir la columna funciona de manera que se suman todas las interacciones de cantidad de la tabla “ANOMALIAS” en las que se repite el campo de la procedencia y la anomalía. En la página aparece un filtro de selección múltiple del campo “ANOMALIA” que permite filtrar la tabla por las anomalías que el usuario desea estudiar.
- La tercera página titulada “TABLA INDICACIONES” funciona exactamente de la misma manera que la anterior página, pero asociada a la tabla “INDICACIONES”. En este caso se repite el proceso, pero utilizando como campo protagonista la columna INDICACION[INDICACION].
- La cuarta y última página tiene como intención visualizar de manera específica y detallada las anomalías relacionadas con cada indicación médica. Esto permite anomalías asociadas a diferentes tipos de indicaciones. En el sistema AccuGest desarrollado en Rstudio se ha integrado esta misma información, pero gracias a las enormes posibilidades de PowerBI, se consigue que en el nuevo panel se pueda interpretar de mejor forma los datos. Tras una investigación, se decide implementar esta página importando un nuevo objeto visual llamado “Sankey”, un diagrama de flujo donde el ancho de la serie es proporcional a la cantidad del flujo.

El objeto visual utiliza Dim_Indicaciones [INDICACION] como campo “Origen”, Dim_Anomalias [ANOMALIA] como campo “Destino” y la cantidad de flujo viene dada por la medida “SUM_CANTIDADES” como se observa en la Ilustración 44. La medida "SUM_CANTIDADES" calcula la suma de la columna "CANTIDAD" de la tabla "ANOMALIA_INDICACION", agrupando los datos por las columnas "INDICACION" y "ANOMALIA". Además, se establece una relación de usuario entre las columnas "FECHA_RECEPCION" de "ANOMALIA_INDICACION" y "Fecha" de "CalendarioAccuGest".

```
1 SUM_CANTIDADES = CALCULATE(SUM(ANOMALIA_INDICACION[CANTIDAD]),GROUPBY(ANOMALIA_INDICACION,
ANOMALIA_INDICACION[INDICACION],ANOMALIA_INDICACION[ANOMALIA]),USERRELATIONSHIP(ANOMALIA_INDICACION
[FECHA_RECEPCION],CalendarioAccuGest[Fecha]))
```

Ilustración 44 Creación medida “SUM_CANTIDADES” con formula DAX.

Como última implementación, se añade un filtro del campo “INDICACIÓN”, con el propósito de poder investigar sobre las diferentes indicaciones que dan como resultado las diferentes anomalías.

En todas las páginas aparece un filtro de procedencias común a todas las páginas del panel “TABLA_RESULTADOS”. Como puede observarse en la Ilustración 52 es un filtro de selección que aparece tras pulsar sobre el botón, a través de la herramienta “Marcadores” de POWER BI. Cada vez que se pulsa sobre el botón “SELECCIONE PROCEDENCIAS” se activa un marcador, que deja visibles en el panel el filtro de selección de procedencias, y al pulsar fuera del filtro de selección se activa otro marcador que regresa a la vista del panel. A su vez todas las páginas del panel, excepto en la segunda pestaña de la primera página, aparece el filtro temporal que se estableció como requisito de diseño, y que afecta a todos los objetos visuales de las páginas.

Si pulsamos sobre el botón situado en la parte superior derecha con un icono personalizado de un gráfico, accedemos al último panel.

4.2.3.4.3 Implementación panel Gráficos de Distribución Temporal

El tercer panel tiene el mismo formato que el del anterior, pero en este caso se representan los datos como gráficos de distribución temporal, lo que da nombre al subpanel “Gráficos de Distribución Temporal”. Podemos dividir el subpanel en cuatro páginas distintas:

- La primera página del subpanel tiene como nombres “ANOMALIAS POR HOSPITAL” y representa la cantidad de anomalías encontradas a lo largo del tiempo por cada procedencia. Desde un principio se intenta representar como se hacía en el sistema original, es decir, utilizando el filtro de procedencias para dibujar simultáneamente todos los gráficos de quesitos seleccionados. Power BI permite realizar esto importando nuevos objetos visuales, pero son productos premium que solo se permiten usar durante un periodo de prueba de un mes. Esto no es posible, ya que uno de los requisitos fundamentales del proyecto es que el sistema sin la necesidad de que un técnico este permanentemente pendiente de los cambios que puedan producirse en el panel. Por ese motivo se toma la decisión de representar

un solo gráfico de quesitos simultáneamente y permitir al usuario elegir, sobre la procedencia a estudiar seleccionando una de las opciones de un filtro de única selección de procedencia situado arriba del objeto visual. El filtro temporal se mantiene, y permite observar la información en el rango de fechas que el usuario desee.

Otra de las decisiones tomadas, es la de eliminar la anomalía “Bajo riesgo de aneuploidía” a partir de un filtro de objeto visual como se observa en la Ilustración 45. Gracias a esta decisión se consigue el objeto visual sea más fácil de interpretar y se estudien solo las anomalías cuyo resultado aparece como “POSITIVO” en la fuente de datos.

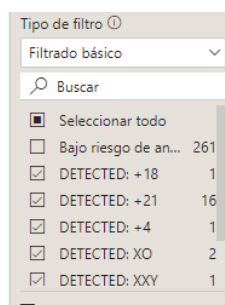


Ilustración 45 Filtro de objeto visual para “ANOMALIAS POR HOSPITAL”.

El objeto visual utiliza dos campos para representar los valores. En primer lugar, se añade Dim_Anomalias [ANOMALIA] como campo leyenda y una medida “SUM_CANTIDAD” (Ilustración 46) que calcula la suma de la columna "CANTIDAD" de la tabla "ANOMALIAS".

```
1 SUM_CANTIDAD = CALCULATE(SUM(ANOMALIAS[CANTIDAD]))
```

Ilustración 46 Creación medida “SUM_CANTIDAD” con formula DAX.

Como última implementación se añade un tooltip personalizado, que permita mostrar información adicional del objeto visual al pasar el cursor sobre él.

- La segunda página funciona exactamente de la misma manera que la anterior página, pero asociada a la tabla “INDICACIONES”. En este caso se repite el proceso, pero utilizando como campo protagonista la columna INDICACION[INDICACION].
- La tercera página se nombra como “PRUEBAS POR HOSPITAL” representa los Centros que encabezan la realización de pruebas genéticas en la Comunidad de Madrid por año. Se utiliza un gráfico de áreas, la medida “TOTAL_PRUEBAS” anteriormente calculada en la tabla “RESULTADOS” como eje y CalendarioAccuGest [Año] como eje x y Dim_Hospitales [PROCEDENCIA] como leyenda. Además, se añade un filtro personalizado al objeto visual de tipo “TopN” que muestra los 6 elementos que encabezan el ranking de Centros con más pruebas.

- La cuarta página funciona de manera similar a la explicada detalladamente en el anterior apartado. Adopta el nombre de “TIEMPO DE RESPUESTA”, está formada por un gráfico de barras agrupadas representa por centro, la diferencia entre la fecha de solicitud y la fecha de informe. Una vez realizado este análisis calculado a partir de la medida “T_RESPUESTA” (Ilustración 47) se añade como campo en el eje y las procedencias y como eje x la medida. La medida se obtiene calculando la media de la columna “DIAS_INFORME” que representa la diferencia entre la fecha de solicitud y la fecha de informe, agrupando las columnas RESULTADOS[PROCEDENCIA] y RESULTADOS[DIAS_INFORME].

```
1 T_RESPUESTA = CALCULATE(AVERAGE(RESULTADOS[DIAS_INFORME]),GROUPBY(RESULTADOS,RESULTADOS[PROCEDENCIA],  
RESULTADOS[DIAS_INFORME]))
```

Ilustración 47 Creación medida “T_RESPUESTA” con formula DAX.

5 Resultados y conclusiones

5.1 Resultados de las mejoras en el sistema AccuGest

Tras la implementación de nuevas mejoras en el sistema original AccuGest desarrollado en Rstudio/ Shiny procedemos a comentar los resultados obtenidos con estas. Se han llevado a cabo mejoras visuales en el sistema AccuGest, esto incluye mejoras en la visualización de los filtros, ajustando los bordes para una apariencia más estética y organizada. Además, se ha realizado un renombramiento de columnas en algunas secciones del sistema. Por ejemplo, en la pestaña "Table" se ha renombrado la columna "NO_CONCLUYENTE" por "CONCLUYENTE", lo que brinda una mayor claridad en la interpretación de los datos. También se ha logrado ajustar el tamaño de las tarjetas resumen, asegurando que todas tengan un tamaño uniforme y una apariencia más equilibrada en el sistema.

Se realizó un profundo estudio de posibilidades de mejora en la pestaña "Anomalías por Hospital". Como resultado, se decidió aplicar un factor de conversión a los datos, como se explica en el punto "4.1.2.2 Implementación mejoras objeto Server". Esta mejora ha permitido una mejor visualización e interpretación de las anomalías por centro, proporcionando información más precisa y útil para los usuarios (Ilustración 48).

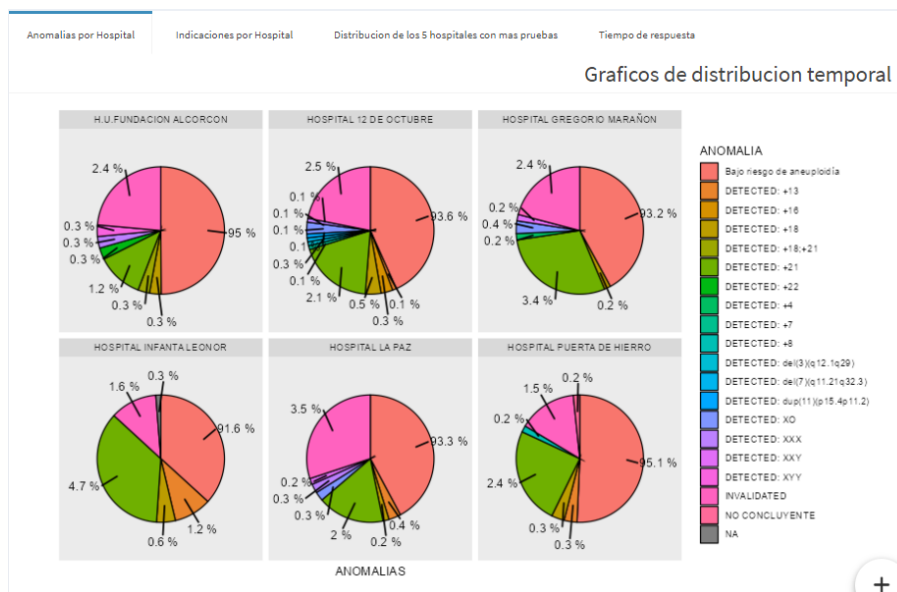


Ilustración 48 Gráfico "Anomalías por Hospital".

Por otra parte, se han realizado mejoras funcionales en el sistema AccuGest. Entre ellas, se ha creado tres nuevas pestañas en la sección inferior izquierda denominada "Tabla Resultados". Estas pestañas son "Anomalías", "Indicaciones" y "Anomalías por indicación". Cada una de estas pestañas funciona de manera similar, proporcionando un filtro personalizado que permite filtrar la tabla y presentar los resultados relacionados con la columna que se está estudiando en cada pestaña. Estas mejoras han ampliado la funcionalidad del sistema, brindando a los usuarios más opciones para analizar y explorar los datos.

Adicionalmente, se ha creado una nueva pestaña llamada "Tiempo de respuesta", como se puede observar en la "Ilustración 49". Esta pestaña utiliza un gráfico de barras agrupadas para representar, por centro hospitalario, la diferencia entre la fecha de solicitud y la fecha de informe. Esta visualización permite una mejor comprensión y seguimiento del tiempo de respuesta de cada centro, lo que facilita la identificación de posibles áreas de mejora en cuanto a eficiencia y tiempos de entrega de los informes.

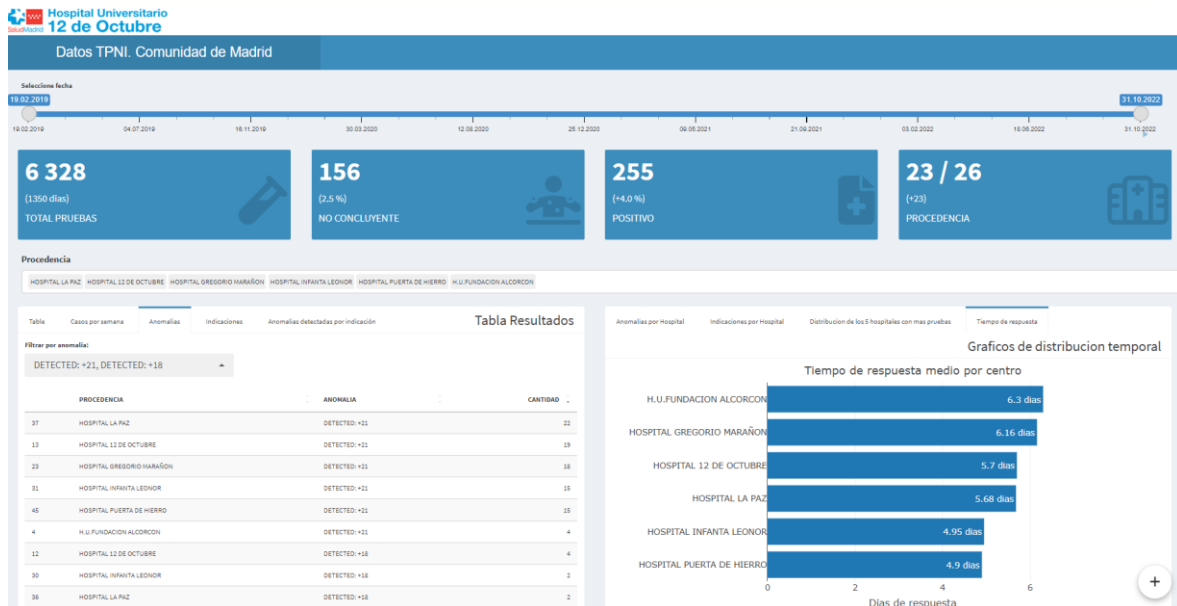


Ilustración 49 Nuevo sistema AccuGest mejorado.

5.2 Nuevo panel en Power BI

El primer subpanel, denominado "Inicio", consta de dos pestañas. La primera pestaña replica las tarjetas del panel anterior, como se muestra en la "Ilustración 50". Además, se consideró importante mostrar esta información de forma individual por centro. Sin embargo, utilizar el mismo filtro de procedencias que se utiliza en los paneles posteriores generaría una incoherencia con el panel anterior, ya que el inicio del panel mostraría todos los centros seleccionados, a diferencia del filtro de procedencias en el sistema AccuGest de RStudio, que muestra solo 6 centros seleccionados. Por lo tanto, se tomó la decisión de crear una nueva pestaña dentro del subpanel que represente un mapa de la Comunidad de Madrid, con una burbuja por cada centro. Además, se añadió un tooltip personalizado que resume la información de los resultados obtenidos por centro al pasar el cursor sobre las burbujas del mapa.

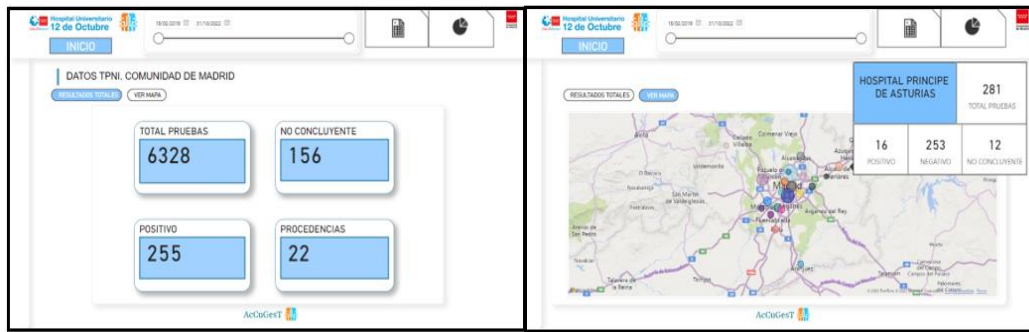


Ilustración 50 Nuevo panel "Inicio" Power BI.

Pulsando el botón situado en la esquina superior derecha que muestra un icono de una hoja con una tabla de resultados, se accede al segundo subpanel del sistema "Tabla Resultados". Al acceder a este subpanel, se despliega un menú de navegación de páginas. El propósito de este menú es permitir un acceso rápido e intuitivo a la misma información en forma de gráfico, replicando el funcionamiento del sistema original. La "Ilustración 51" muestra las 4 páginas que conforman el subpanel "Tabla Resultados". En la segunda pestaña de la primera página, "RESULTADOS PRUEBAS", se han añadido dos filtros uno de año y otro semanal, para realizar un estudio más detallado y personalizado de los resultados por semana. En la cuarta página, se muestra una tabla de anomalías por indicación con un botón que al pulsarlo muestra un gráfico con un diagrama de flujo, donde el ancho de la serie es proporcional a la cantidad del flujo. Se implementó este gráfico para mejorar la comprensión de esta nueva página. En las 4 páginas de este subpanel, se aplican filtros temporales y de procedencias, como se planteó en el diseño. El filtro temporal, como se muestra en la "Ilustración 52", despliega un filtro de selección que ocupa el tamaño completo del panel, permitiendo seleccionar las procedencias. Al hacer clic en cualquier área de la pantalla fuera del filtro, se vuelve a la página que se estaba estudiando antes de pulsar sobre el filtro, pero con la nueva selección de procedencias aplicada a los objetos visuales.

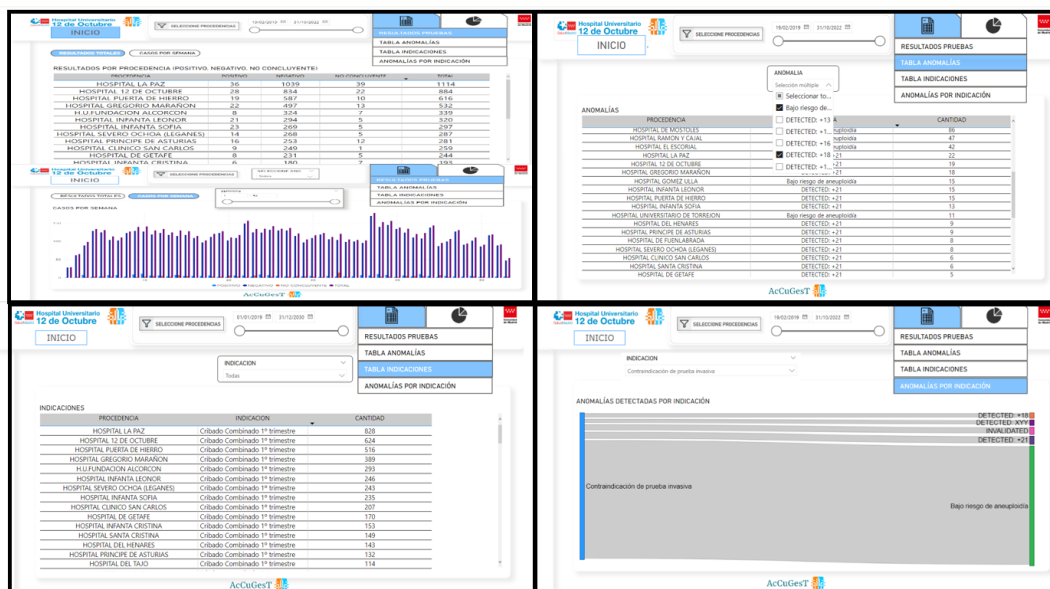


Ilustración 51 Nuevo panel "Tabla Resultados" Power BI.

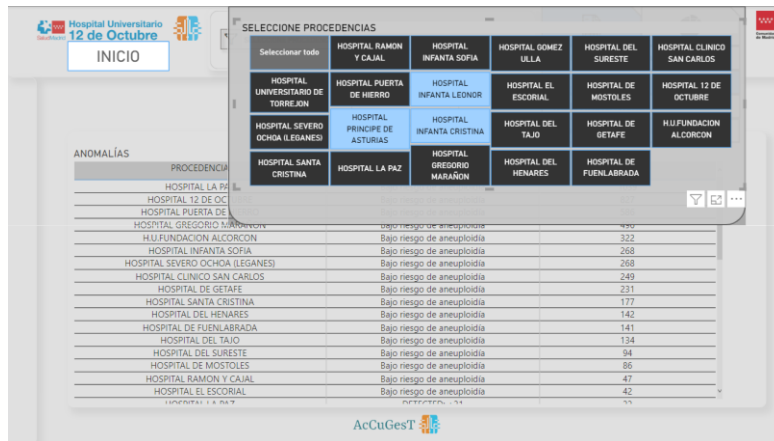


Ilustración 52 Nuevo filtro de procedencias Power BI.

Por último, se presenta el subpanel "Gráficos de Distribución temporal" "Ilustración 53", que replica la pestaña "Gráficos de Distribución Temporal" del panel AccuGesT mejorado en RStudio. Cabe destacar que en las dos primeras páginas que forman este panel ("ANOMALÍAS POR HOSPITAL" e "INDICACIONES POR HOSPITAL"), no fue posible crear tantos gráficos de quesito como selecciones marcadas en el filtro de procedencias, no existen o al menos no se han encontrado formas gratuitas de replicar el diseño que se utilizaba en el anterior sistema AccuGesT. Por tanto, se toma la decisión tanto en las dos páginas de eliminar el filtro de procedencias, y crear un filtro de única de selección de procedencias que permita al usuario estudiar los datos que los objetos visuales presentan a partir una única procedencia para estudiar. Las siguientes dos páginas si cuentan con el filtro de selección de procedencias planteado en el diseño del panel y funcionan de la misma manera que en el sistema replicado.

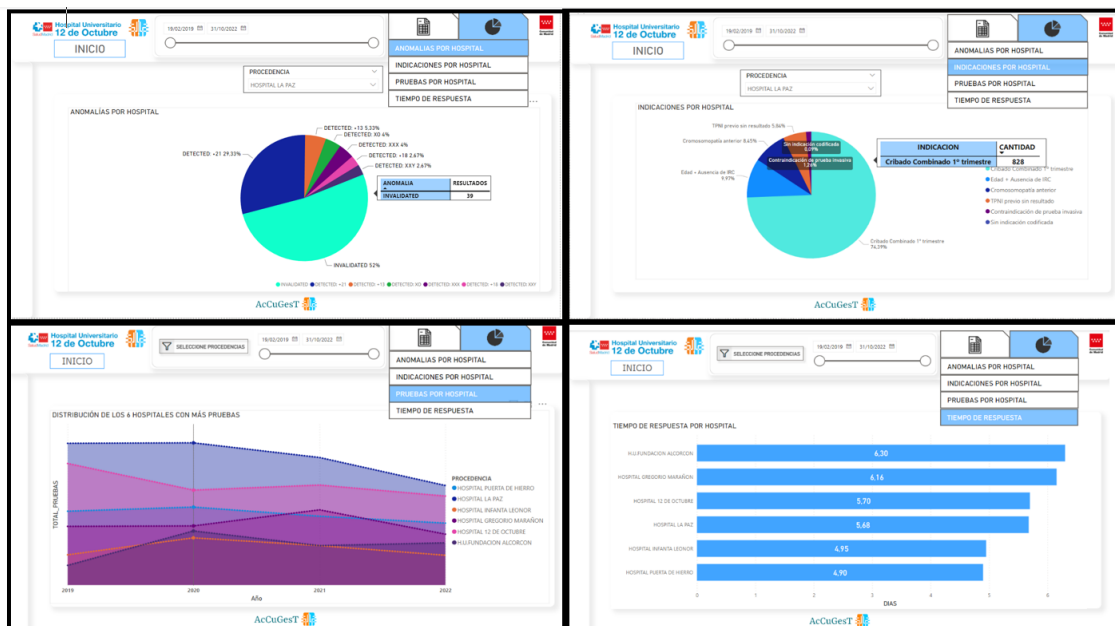


Ilustración 53 Nuevo panel "Gráficos de Distribución temporal".

5.3 Comparativa entre los dos paneles

5.3.1 Comparativa de visualización y usabilidad entre los dos paneles

En esta sección, se analizarán los aspectos que cada una de las herramientas aporta al usuario, centrándonos en la facilidad de uso, la flexibilidad, las capacidades de visualización, el rendimiento y la usabilidad.

En cuanto a la facilidad de uso, tanto el panel desarrollado en RStudio como el panel en Power BI ofrecen una experiencia bastante similar para los usuarios. Ambos paneles están diseñados de manera intuitiva y no requieren conocimientos especializados para poder manejar la herramienta.

Sin embargo, es importante destacar que el panel desarrollado en Power BI ofrece una visualización mejorada y más intuitiva en comparación con el panel en RStudio. Además, proporciona una mayor personalización y profundidad en el estudio de los datos. Tomaremos como ejemplo la pestaña "Casos por Semana" ilustrada en las "Ilustraciones 5 e 51". En el panel desplegado en RStudio, se permite profundizar en la visualización a través de una herramienta que permite hacer zoom y moverse por el panel, proporcionada por la función `plotlyOutput`. Si comparamos esta herramienta con la posibilidad de filtrado que ofrece Power BI y que se ha desarrollado, queda evidente que la pestaña "Casos por Semana" en Power BI proporciona un mayor entendimiento de los datos y sobre todo una mayor facilidad en el uso de las herramientas para explorarlos en detalle.

En conclusión, aunque ambas herramientas ofrecen una facilidad de uso similar para los usuarios, el panel desarrollado en Power BI destaca por su visualización mejorada, más intuitiva y con una mayor capacidad de personalización. Es capaz de proporcionar una mayor profundidad en el estudio de los datos. Esto representa una clara ventaja para los usuarios que buscan una herramienta que les permita analizar y comprender los datos de manera más efectiva.

5.3.2 Análisis comparativo de las herramientas desde la perspectiva del desarrollo

Si bien el lenguaje R y el uso de la librería Shiny permite a los usuarios con poca experiencia desplegar un cuadro de mando de forma sencilla y en poco tiempo, Power BI ofrece esa misma capacidad y, además, agrega un valor añadido en la visualización de los datos que se desean presentar. Ambas herramientas no presentan una gran dificultad en su uso, pero es innegable que Microsoft brinda más soporte y ayuda en el uso de sus herramientas en comparación con Shiny, a pesar de que Shiny cuenta con el apoyo de una comunidad activa.

RStudio es una herramienta poderosa que permite crear cuadros de mando, pero a medida que se requiere una mayor personalización y complejidad en la visualización de datos, Power BI ofrece una ventaja significativa. Microsoft invierte día a día en el desarrollo de Power BI para proporcionar una experiencia más completa y numerosos recursos de soporte. Esto incluye una gran cantidad

de documentación, tutoriales, comunidades en línea y servicios de asistencia al usuario.

Aunque Shiny cuenta con una comunidad activa que comparte conocimientos y brinda soporte, la diferencia en el respaldo proporcionado por Microsoft es notable. Esto facilita a los desarrolladores la exploración y explotación de las capacidades de visualización de Power BI de manera más eficiente.

Además, es importante mencionar que el análisis de datos mediante R puede resultar más complejo en comparación con la utilización de Power Query y el lenguaje propio de Power BI, DAX. Para ilustrar esta diferencia, vamos a presentar un caso de ejemplo que compara la complejidad en el desarrollo de una visualización utilizando ambas herramientas.

La tabla “Table” que aparece en la Ilustración 5, utiliza un dataframe que ya estaba implementado “res_out2” (Ilustración 52), que se obtiene a partir de una serie de operaciones en el dataframe tpni_entrada para obtener el dataframe res_out2. En primer lugar, se agrupa el dataframe por la fecha de recepción, semana de recepción, procedencia y resultado, y se calcula la frecuencia agrupada. A continuación, se realiza una transformación pivotando los datos para crear columnas con los valores de los resultados. Después, se reemplazan los valores faltantes por ceros en el dataframe resultante. Luego, se calcula la suma total por filas, creando una nueva columna llamada “Total”. Por último, se realiza una suma acumulada de los totales por procedencia en una nueva columna llamada “Acumulada”. En resumen, el código pasa por los pasos de agrupación, transformación, reemplazo de valores, cálculo de sumas y suma acumulada para generar el dataframe resultante. A todo esto, habría que sumarle, llamar a al dataframe en un output, crear un tabPanel en el objeto ui que llame al output, y algunas modificaciones más para poder finalmente representarlo en el panel.

Este mismo proceso en Power BI, se lleva a cabo a partir de la tabla “RESULTADOS” que se explica en el apartado “4.2 Implementación panel AccuGest en Power Bi”. Utiliza además las medidas explicadas en el mismo apartado “NUM_NEGATIVO”, “NUM_POSITIVO”, “NUM_NOCONCLUYENTE” y “TOTAL_PRUEBAS”. Para crear la misma visualización que en el anterior panel se añade un objeto visual seleccionando “Tabla” en el campo “Visualizaciones” y se añaden las cuatro medidas, y una columna procedencias, para generar la misma tabla como se ve en la ilustración perteneciente a la creación del subpanel “Tabla Resultados”.

En conclusión, el desarrollo de visualizaciones es más sencillo en Power BI en comparación con la herramienta que utiliza la librería Shiny. Mientras que en la herramienta anterior se requiere realizar una serie de operaciones en el código para obtener el dataframe final, en Power BI se puede lograr la misma visualización de manera más directa y eficiente utilizando medidas y seleccionando las columnas necesarias. Power BI proporciona una interfaz intuitiva y fácil de usar, donde los usuarios pueden seleccionar visualizaciones predefinidas y agregar medidas y columnas con unos pocos clics. Esto simplifica el proceso de creación de visualizaciones y reduce la necesidad de realizar operaciones manuales en el código. En resumen, Power BI ofrece una

experiencia más fluida y accesible para el desarrollo de visualizaciones, lo que facilita la generación de informes y paneles interactivos.

```
res_out <- tpni_entrada %>% group_by(FECHA_RECEPCION, SEMANA_RECEPCION,PROCEDENCIA, RESULTADO) %>%
summarise( FREC_GROUP = sum(CANTIDAD))

res_out2<-res_out %>%
  pivot_wider(names_from = RESULTADO, values_from = FREC_GROUP) #pivotamos filas para hacer columnas
res_out2[is.na(res_out2)] <- 0 #sustituye NA por 0
res_out2$Total = rowSums (res_out2[, 4:6]) #suma por filas

#suma acumulada de nuevos casos por tipo de resultado y procedencia
#(EN COLUMNAS, SUMA DE TOTALES, NO DESGLOSADA POR TIPO DE RESULTADO)

res_out2 <- res_out2 %>%
  group_by(`PROCEDENCIA`) %>%
  mutate(Acumulada = cumsum(Total)) %>%
  ungroup()
```

Ilustración 54 implementación dataframe "res_out2".

6 Análisis de impacto

Vamos a catalogar este proyecto en el marco de los siguientes Objetivos de Desarrollo Sostenible (ODS) [23] establecidos por las Naciones Unidas:

- El proyecto actual se alinea con el Objetivo de Desarrollo Sostenible número 3, que busca promover la salud y el bienestar. Mediante el uso de datos clínicos y la creación del cuadro de mandos AccuGest en Power BI, se logra recopilar, organizar y analizar información para comprender los datos generados a partir de los TPNI. En conclusión, el proyecto cumple el objetivo de mejorar la salud y el bienestar de las personas mediante el uso de datos clínicos y herramientas tecnológicas avanzadas.
- El proyecto actual se alinea también con el Objetivo de Desarrollo Sostenible número 9, que intenta promover la industrialización inclusiva y sostenible y fomentar la innovación. En este caso, se ha llevado a cabo utilizando herramientas de visualización para desarrollar el cuadro de mandos AccuGest. Esto tiene como objetivo mejorar la infraestructura de datos y el diseño de un anterior panel desarrollado en otra herramienta. Al utilizar Power BI como solución tecnológica, se impulsa la capacidad de recopilar, analizar y compartir datos clínicos de manera más eficiente, lo que a su vez contribuye a mejorar la gestión y la toma de decisiones en el ámbito de la salud.
- El proyecto actual se alinea con el Objetivo de Desarrollo Sostenible número 17, centrado en las alianzas para lograr los objetivos. Durante la elaboración del proyecto ha sido importante de la colaboración entre profesionales de la salud, un responsable en la toma de decisiones y usuarios que utilicen nuevas tecnologías para la visualización de datos.

7 Conclusiones y líneas futuras

La mejora del panel y la migración a otra herramienta, en concreto Power BI, ha conseguido mejorar el antiguo sistema, tanto de forma visual como funcional. Estos eran objetivos necesarios para que poder dar esta parte del proyecto como exitosa. Además, el nuevo panel generado ofrece otro diseño del sistema AccuGest, proporcionando otra forma de visualizar los datos en mayor profundidad a partir de otra herramienta más utilizada en el ámbito de la visualización de datos. Se consiguió replicar el funcionamiento de los objetos visuales a partir de un filtro temporal y de un filtro de procedencia, replicando los del anterior modelo. Además, en la pestaña “Casos por semana” se consigue generar filtros suficientes que permiten realizar un estudio y un seguimiento más personalizado de todos los resultados obtenidos en las pruebas realizadas. Se ha conseguido también añadir un mapa personalizado con todas las procedencias que aparecen en los datos, y hacer un breve resumen de los resultados de las pruebas a partir de un tooltip personalizado para cada procedencia.

Algunas mejoras que se podrían introducir de cara a nuevas versiones ampliadas del panel ya existente podrían venir a partir de la extracción de nuevos datos procedentes de la fuente de origen de los datos. En concreto de la columna “CIPA” que se elimina en la extracción del .xls generado para el estudio de los datos. Esto permitiría hacer un estudio mucho más personalizado de cada paciente, obteniendo por ejemplo el número real de pacientes cuyo resultado ha sido “NO CONCLUYENTE”, ya que actualmente si un mismo paciente ha recibido como resultado de la indicación realizada más de un resultado “NO CONCLUYENTE” se puede identificar correctamente para cuantos pacientes no ha sido posible obtener un resultado válido. Otra posible mejora podría ser encontrar una forma de representar las pestañas “Anomalías por Hospital” e “Indicaciones por Hospital” de manera que funcionen ligados al mismo filtro de procedencias con el que se relacionan el resto de los objetos visuales. Esto ofrecería al usuario la posibilidad de estudiar en estas páginas varias procedencias al mismo tiempo y no hacerlo de forma individualizada.

A título personal, este proyecto me ha permitido conocer nuevas herramientas de visualización de las que no tenía referencias, y a aprender cuales son las bases fundamentales en la implementación de un código analítico.

8 Bibliografía

- [1] "Observatorio de I+D+i UPM". Universidad Politécnica de Madrid. <https://www.upm.es/observatorio/vi/index.jsp?pageac=grupo.jsp&idGrupo=322#:~:text=El%20Grupo%20de%20Informática%20Biomédica,un%20gran%20número%20de%20colaboradores>. (accedido el 22 de marzo de 2023).
- [2] "Genética". Hospital Universitario 12 de Octubre. <https://www.comunidad.madrid/hospital/12octubre/profesionales/servicios-centrales/genetica> (accedido el 22 de marzo de 2023).
- [3] Welcome to Shiny Server! <https://rserver.h12o.es/AccuGest/> (accedido el 22 de marzo de 2023).
- [4] "datos.gob.es". Datos abiertos del Gobierno de España – Datos.gob.es. <https://datos.gob.es/es/blog/la-importancia-de-la-visualizacion-grafica-de-los-datos> (accedido el 31 de marzo de 2023).
- [5] "¿Qué es Business Intelligence o inteligencia de negocios?" Tableau. <https://www.tableau.com/es-es/learn/articles/business-intelligence#:~:text=El%20Business%20Intelligence,%20BI%20o,decisiones%20basadas%20en%20los%20datos>. (accedido el 31 de abril de 2023).
- [6] "COMPARATIVA: 10 herramientas de analítica de datos". Revista Byte TI. <https://revistabyte.es/comparativa/comparativa-analitica-de-datos/> (accedido el 9 de abril de 2023).
- [7] "R (lenguaje de programación) - Wikipedia, la enciclopedia libre". Wikipedia, la enciclopedia libre. [https://es.wikipedia.org/wiki/R_\(lenguaje_de_programación\)](https://es.wikipedia.org/wiki/R_(lenguaje_de_programación)) (accedido el 9 de abril de 2023).
- [8] "Python - Wikipedia, la enciclopedia libre". Wikipedia, la enciclopedia libre. <https://es.wikipedia.org/wiki/Python> (accedido el 9 de abril de 2023).
- [9] "Posit". Posit. <https://posit.co/> (accedido el 15 de abril de 2023).
- [10] "R: The R Project for Statistical Computing". <https://www.r-project.org/conferences/useR-2011/abstracts/180111-allairejj.pdf> (accedido el 15 de abril de 2023).
- [11] "Shiny". Shiny. <https://shiny.posit.co/> (accedido el 15 de abril de 2023).
- [12] "Power BI - Wikipedia, la enciclopedia libre". Wikipedia, la enciclopedia libre. https://es.wikipedia.org/wiki/Power_BI (accedido el el 15 de abril de 2023).
- [13] "Información general sobre DAX - DAX". Microsoft Learn: Build skills that open doors in your career. <https://learn.microsoft.com/es-es/dax/dax-overview> (accedido el 20 de abril de 2023).
- [14] "MySQL". MySQL. <https://www.mysql.com/> (accedido el 20 de abril de 2023).
- [15] "Expanding Redis Database Trigger Features | Redis". Redis. <https://redis.com/blog/database-trigger-features> (accedido el 20 de abril de 2023).
- [16] "Database Management Cloud | Oracle España". Oracle | Cloud Applications and Cloud Platform. <https://www.oracle.com/es/manageability/database-management/> (accedido el 20 de abril de 2023).
- [17] "APISCAM". APISCAM CESUS. <http://apiscam.blogspot.com/search/label/CESUS> (accedido el 25 de abril de 2023).

[18] "Comunidad de Madrid - CIBELES - madrid.org". Comunidad de Madrid. http://www.madrid.org/wleg_pub/secure/normativas/contenidoNormativa.jsf?opcion=VerHtm&nmnorma=6655&eli=true#no-back-button (accedido el 25 de abril de 2023).

[19] "HL7-FHIR" ¿Qué es HL7 FHIR? . <https://www.tibco.com/es/reference-center/what-is-hl7-fhir> (accedido el 25 de abril de 2023).


[20] "¿Qué es un archivo XLS?" online-convert.com. <https://www.online-convert.com/es/formato-de-archivo/xls> (accedido el 25 de abril de 2023).

[21] "¿Que es Power Query? - Power Query". Microsoft Learn: Build skills that open doors in your career. <https://learn.microsoft.com/es-es/power-query/power-query-what-is-power-query> (accedido el 15 de mayo de 2023).

[22] "Google Maps". Google. <https://www.google.es/maps/?hl=es> (accedido el 25 de mayo de 2023).

[23] "Ministerio de Derechos Sociales y Agenda 2030 - Agenda 2030". Ministerio de Derechos Sociales y Agenda 2030 - Web del Ministerio de Derechos Sociales y Agenda 2030. <https://www.mdsocialesa2030.gob.es/agenda2030/index.htm> (accedido el 25 de mayo de 2023).

Este documento esta firmado por



| | |
|-------------------------------|---|
| Firmante | CN=tfgm.fi.upm.es, OU=CCFI, O=ETS Ingenieros Informaticos - UPM, C=ES |
| Fecha/Hora | Sun Jun 11 22:29:32 CEST 2023 |
| Emisor del Certificado | EMAILADDRESS=camanager@etsiinf.upm.es, CN=CA ETS Ingenieros Informaticos, O=ETS Ingenieros Informaticos - UPM, C=ES |
| Numero de Serie | 561 |
| Metodo | urn:adobe.com:Adobe.PPKLite:adbe.pkcs7.sha1 (Adobe Signature) |