



Mejoramiento de los servicios de inteligencia microbiológica de la empresa oKuo

Esteban Valencia Yepes

Trabajo de grado presentado para optar al título de Ingeniero Bioquímico

Asesores

David Andrés Gómez Ríos, PhD Ingeniería química

Sebastián Vélez Pérez, Zootecnista

Universidad de Antioquia

Facultad de Ingeniería

Ingeniería Bioquímica

El Carmen de Viboral, Antioquia, Colombia

2021

Cita	Valencia Yepes [1]
Referencia Estilo IEEE (2020)	[1] E. Valencia Yepes, “Mejoramiento de los servicios de inteligencia microbiológica de la empresa oKuo”, Trabajo de grado profesional, Ingeniería Bioquímica, Universidad de Antioquia, El Carmen de Viboral, Antioquia, Colombia, 2021.



Biblioteca Seccional Oriente (El Carmen de Viboral)

Repositorio Institucional: <http://bibliotecadigital.udea.edu.co>

Universidad de Antioquia - www.udea.edu.co

Rector: John Jairo Arboleda Céspedes.

Decano/Director: Jesús Francisco Vargas Bonilla.

Jefe departamento: Lina María González Rodríguez

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

Agradecimientos

Gracias a mi Alma Máter por haberme brindado tanto conocimiento durante toda la trayectoria del pregrado, gracias a mis asesores Sebastián Vélez y David Andrés Gómez quienes me acompañaron en esta práctica empresarial que fue de mucho crecimiento personal y profesional. Gracias a mi familia y amigos por haberme apoyado en el transcurso de la carrera, hicieron de esta etapa de mi vida algo excepcional.

TABLA DE CONTENIDO

- LISTA DE FIGURAS Y TABLAS.....4
- RESUMEN.....5
- ABSTRACT6
- I. INTRODUCCIÓN7
- II. OBJETIVOS.....9
- III. MARCO TEÓRICO9
- IV. METODOLOGÍA12
 - A. Revisión general y familiarización con la plataforma Qlik Sense, mongo DB y el proceso de ETL.....13
 - B. Revisión y descarga de información de las plataformas para un análisis detallado en Python13
 - C. Reestructuración de la arquitectura del código en Qlik Sense remota de datos.....14
 - D. Replantear la estructura de los datos presentados en mongo DB.....14
 - E. Complementar proceso de la ETL.....15
 - F. Pruebas e implementación.....17
 - G. Aseguramiento de la calidad y mejora continua17
- V. RESULTADOS Y ANÁLISIS17
- VI. CONCLUSIONES23

REFERENCIAS	24
-------------------	----

LISTA DE FIGURAS

Fig. 1. Metodología	12
Fig. 2. Nuevas fórmulas empleadas para el cálculo total del índice.....	18
Fig. 3. Hoja principal del índice antes de los cambios.....	18
Fig. 4. Nueva Hoja principal del oKuo scope TM	19
Fig. 5. Esquema general oKuo scope TM	21
Fig. 6. Vista de informe general en la plataforma.....	22

LISTA DE TABLAS

Tabla 1. Comparativo antes y después de mejoras	20
---	----

RESUMEN

oKuo es una empresa que hace parte de la Alianza Iluma, su propósito está en nutrir el bienestar de las personas, a través del aseguramiento de la inocuidad de los alimentos, contando con más de 20 años en el mercado. Su propuesta de valor está basada en tecnología, el conocimiento y la innovación. Son productores de aditivos de alta tecnología, que se adaptan a los nuevos retos del mercado. Cuentan con un portafolio amplio de productos como inhibidores, antioxidantes, acidificantes, desinfectantes y otros que aportan en la mejora de la productividad en la cadena de producción de proteína de origen animal.

La empresa oKuo, en cumplimiento de su propuesta de valor, ofrecen un catálogo de servicios de vigilancia sanitaria y aseguramiento de la inocuidad, que provee información a sus clientes de su desempeño microbiológico con base en datos históricos, experiencia y conocimiento acumulados, muestreo y análisis microbiológicos. Para cumplir con su propuesta de valor, la compañía ha desarrollado herramientas informáticas para la gestión y análisis de la información.

Dentro de este portafolio de servicios se ha desarrollado una plataforma que permite visualizar de forma interactiva y a distintos niveles el estado microbiológico de las empresas de los clientes. Esta plataforma involucró en su desarrollo la visualización y cálculos algorítmicos en el software de analítica de datos Qlik Sense (QS), el almacenamiento de los datos en MongoDB y el proceso de Extracción Transformación y Carga o ETL por sus siglas en inglés, de los datos se implementó con Python. Por ende, cada uno de los procesos en el backend es una posible fuente de errores, que se suma a los posibles problemas originados en la estructura inicial del algoritmo de cálculo como son la definición de ponderados, zonas de proceso, tratamientos de desinfección, entre otros, que deben integrar el conocimiento de la ingeniería de procesos.

En el siguiente trabajo se ha realizado un proceso de aseguramiento de la calidad en el cual se ha revisado, corregido y reestructurado la arquitectura de los algoritmos empleados en QS tanto para la visualización como para los cálculos, que además de permitir un mejor mantenimiento del

código permite la rápida implementación de nuevos análisis, gráficos y modelos. Adicionalmente, se ha analizado el proceso de ETL, logrando modificarlo de forma que se puedan recuperar una mayor cantidad de datos valiosos de los informes microbiológicos de los clientes y ajustando la forma en la que estos se cargan a la base de datos, lo que permitió una mejora importante en la flexibilidad y facilidad al momento de consultar los datos desde QS.

***Palabras clave* — Índice de inocuidad, análisis de datos, Qlik sense, aseguramiento de calidad.**

ABSTRACT

oKuo is a company that is part of the Iluma Alliance, its purpose is to nurture the well-being of people, through ensuring the safety of food, counting on more than 20 years in the market. Its value proposition is based on technology, knowledge, and innovation. They are producers of high-tech additives, which adapt to new market challenges. They have a wide portfolio of products such as inhibitors, antioxidants, acidifiers, disinfectants, and others that contribute to the improvement of productivity in the production chain of animal protein.

The oKuo company, in compliance with its value proposition, offers a catalogue of health surveillance and safety assurance services, which provides information to its customers about their microbiological performance based on historical data, accumulated experience and knowledge, microbiological sampling and analysis. To meet its value proposition, the company has developed computer tools for information management and analysis. Within this portfolio of services, a platform has been developed that allows users to visualize the microbiological status of their clients interactively and at different levels. This platform involved in its development the visualization and algorithmic calculations in the data analytics software Qlik sense (QS), the storage of the data in MongoDB and the process of Extraction Transformation and Load or ETL data was implemented with Python. Each of the processes in the data layer (backend) is a possible source of errors, which is added to the possible problems originated in the initial structure of the calculation algorithm such as the definition of weights, process zones, disinfection treatments, among others, that should integrate the knowledge of process engineering.

In the following work, a quality assurance process has been carried out in which the architecture of the algorithms used in QS for both visualization and calculations has been revised, corrected and restructured, which in addition to allowing better maintenance of the code allows the rapid implementation of new analysis, graphics and models. In addition, the ETL process has been analyzed, modifying the process so that more valuable data can be retrieved from customers' microbiological reports and adjusting the way in which they are uploaded to the database, which allowed for a significant improvement in the flexibility and ease of accessing data from QS.

Keywords — **Safety index, data analysis, Qlik sense, quality assurance.**

I. INTRODUCCIÓN

oKuo es una empresa que mediante la unión de tecnología, conocimiento e innovación busca mitigar los riesgos de inocuidad en la cadena de producción de proteína. Con este propósito, la empresa incursiona en la creación de una herramienta que denomina 'oKuo scopeTM', la cual está disponible en la plataforma de análisis de datos Qlik Sense (QS). El oKuo scope es una herramienta de gestión que ofrece una visualización gráfica y estadística para la trazabilidad y toma de decisiones asertivas, óptimas y oportunas para el proceso de la producción de proteína animal.

El funcionamiento del oKuo scopeTM depende principalmente de 3 procesos, cada uno asociado a una plataforma o herramienta. La parte de la visualización y algunos de los cálculos se lleva a cabo en QS, plataforma de pago desarrollada para trabajo en ciencias de datos y los modelos de inteligencia de negocios. QS se alimenta de datos, por lo cual esta herramienta permite conectarse a gran variedad de plataformas de almacenamiento de datos. En este caso se usa MongoDB para la gestión de datos de los clientes que son utilizados por QS. MongoDB es una base de datos de tipo no relacional, la cual tiene grandes ventajas frente a bases de datos relacionales, como la de permitir cambiar de forma fácil y rápida la estructura de la base de datos a medida que estos sea necesario. A MongoDB se ingresan los datos de microbiología de los clientes, los cuales provienen de los análisis microbiológicos reportados por los laboratorios contratados por las empresas.

Cada cliente almacena información diferente de acuerdo con sus necesidades, en diferentes formatos y plataformas, por lo que, para ingresar esta información a la base de datos, se requiere realizar un proceso denominado Extract transform and Load (ETL). El proceso de ETL se realiza con el lenguaje de programación Python, en este se le da el debido tratamiento a cada archivo de cada cliente con el fin de estandarizar la información que quedará almacenada en MongoDB. Cada uno de estos procesos se convierte en un segmento de la cadena en la cual se pueden generar errores. Adicional a lo anterior, la lógica inicial sobre la que está basado el oKuo scopeTM podría variar a medida que las necesidades de los clientes, la visión de oKuo sobre la importancia de ciertos procesos en planta respecto a la microbiología y la estructura la de plataforma como tal vayan cambiando.

La visión inicial y general sobre el proceso, la lógica microbiológica para los cálculos, la ETL, la estructura de la base de datos y el modelo en QS, están cada uno realizados por personas diferentes, de las cuales sólo una es experta en microbiología, y ninguna conoce el funcionamiento de todos los eslabones de la cadena que se requieren para el funcionamiento del oKuo scopeTM.

Durante esta práctica, se realizó un proceso de aseguramiento de la calidad, de forma integral en los principales procesos que conforman el oKuo scopeTM, primero entendiendo la lógica microbiológica sobre la que está estructurada la plataforma y el funcionamiento de cada uno de los procesos. A partir de eso se hace una revisión y corrección del modelo en QS, con lo cual se logran detectar y reparar errores importantes en el propio modelo de QS, pero también en la extracción, en la transformación y en la carga de datos, con lo cual se almacenaron datos de forma incorrecta en la base de datos que alimenta a QS.

Este proceso de revisión y corrección se realizó varias veces hasta alcanzar un producto óptimo que pudiera ser presentado al cliente. Para una parte de la revisión de los datos de MongoDB, se usó Python integrando librerías de manipulación de datos y cálculos matemáticos como Pandas y Matplotlib, el resto de la revisión y corrección se realizó usando la propia plataforma QS y MongoDB compass en su propia sintaxis. Gracias al trabajo realizado, se logró obtener un producto de calidad, el cual gracias a la nueva arquitectura en el código y en la cadena del proceso, se pueden realizar modificación de forma mucho más rápida y eficiente, haciendo del oKuo scopeTM uno producto mucho más flexible y fácil de mantener ante posibles cambios.

II. OBJETIVOS

A. Objetivo general

Desarrollar soluciones para el procesamiento, análisis y gestión de información de análisis microbiológicos, inocuidad de procesos y control de calidad sanitario que la empresa oKuo provee a clientes corporativos de la industria de alimentos.

B. Objetivos específicos

- Detectar errores de tipo estadístico y proponer mejoras en la gestión de la plataforma Qlik Sense que almacena datos de diferente naturaleza microbiológica y química en relación con la inocuidad de procesos y medios de control recomendados e implementados.
- Incorporar nueva información, herramientas de análisis estadístico y presentación de datos en la plataforma y sus bases de datos asociados que permitan un mejor análisis de la información microbiológica.
- Realizar recomendaciones precisas que permitan a las empresas mejorar su control sanitario, además de apoyar a la empresa con la realización de informes sobre el estado y el comportamiento microbiológico de las diferentes plantas de algunos clientes.

III. MARCO TEÓRICO

QS es una plataforma que permite generar tableros en los que se visualizan de forma didáctica, organizada y de forma sencilla datos empresariales, con el fin de extraer información con alto valor que permitan a las organizaciones realizar informes, crear presentaciones y extraer información importante de forme sencilla y rápida. QS permite crear dashboards de forma sencilla al permitir seleccionar entre varios tipos de gráficos, seleccionando la información que se quiere presentar y visualizando al instante. QS emplea el paradigma de programación funcional, lo que

permite emplear comandos para realizar transformaciones de forma sencilla sobre los datos cargados en la plataforma. (Qlik Sense | Data Analytics Platform, n.d.)

En el caso de OKuo, la ruta para los análisis estadísticos y los diferentes gráficos ofrecidos en la plataforma Qlik Sense se presentan en documentos internos de la empresa, en la cual se establecen los criterios de microbiológicos de interés según el Instituto Colombiano Agropecuario (ICA), además de la parametrización que realiza a los datos según ubicación del análisis, el tipo de microorganismo analizado y el tipo de muestra.

Los datos de los análisis microbiológicos que son ingresados por los diferentes clientes de OKuo se almacenan en MongoDB. MongoDB es una base de datos de documentos de código abierto construida sobre una arquitectura de escala horizontal. En esta, en lugar de almacenar datos en tablas de filas o columnas como bases de datos SQL, cada fila en una base de datos MongoDB es un documento descrito en JSON. Las bases de datos de documentos son extremadamente flexibles, permitiendo variaciones en la estructura de los documentos y permitiendo el almacenamiento de documentos que están parcialmente completos. (Why Use MongoDB & When to Use It? | MongoDB, n.d.)

Para el análisis de los datos, se emplea el lenguaje de programación Python, en cual se ha convertido en una de las herramientas para el análisis de datos más importantes del momento, contando con librerías especializadas para tal fin que permiten la manipulación y análisis de grandes cantidades de datos de forma potente y flexible. La librería más importante empleada es pandas, la cual es una herramienta de manipulación y análisis de datos de código abierto rápida, potente, flexible y fácil de usar, construida sobre el lenguaje de programación Python. (Pandas - Python Data Analysis Library, n.d.)

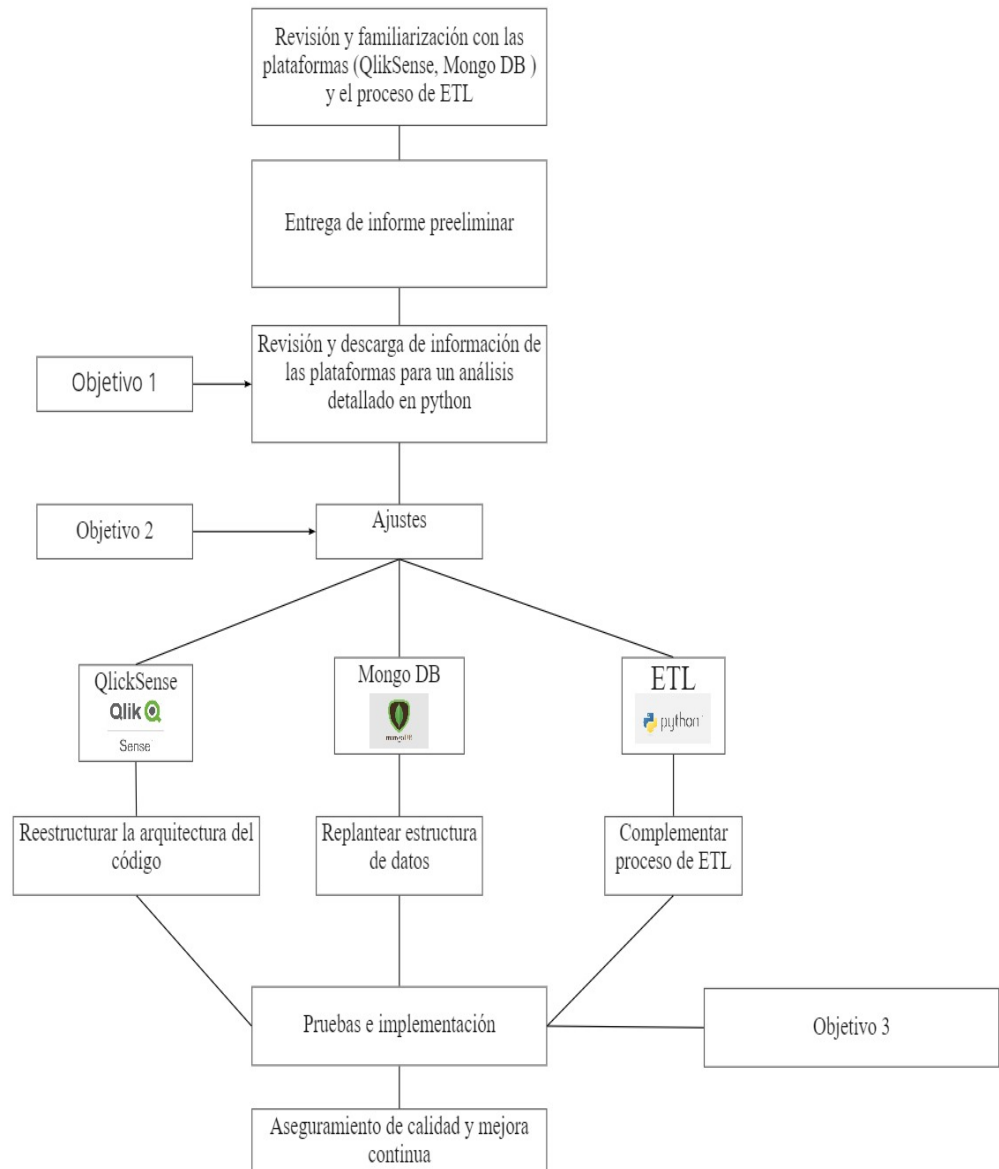
El riesgo microbiológico en inocuidad está relacionado directamente con microorganismos patógenos, siendo los más relevantes en la industria de producción de proteína animal: *Salmonella spp*, *Clostridium perfringens*, *E.coli* patogénica, *Listeria monocytogenes* y *Campylobacter spp*. El recuento estándar para *aerobios mesófilos*, *enterobacterias*, *coliformes totales*, *mohos* y *levaduras*, *esporas Clostridium sulfito reductor* y *E.coli* biotipo 1 es usado para: (i) determinar el estado

higiénico de los equipos de procesamiento y del ambiente, (ii) comprender la ecología microbiana del ambiente de procesamiento, (iii) validar o verificar la limpieza y el saneamiento, (iv) verificar los pasos de control del proceso y (v) evaluar el riesgo de contaminación posterior a procesamiento, porque son microorganismos indicadores que afectan la calidad y no la inocuidad. Y la segunda categorización está relacionada con el proceso en términos de control de proceso, materia prima y producto terminado.

El índice de inocuidad es obtenido de la calificación ponderada de microorganismos patógenos y microorganismos indicadores y del proceso, teniendo mayor peso los microorganismos patógenos y en su orden el control de proceso, la materia prima y por último el producto terminado bajo un enfoque preventivo, donde el mayor control debe ser realizado en etapas anteriores y durante la elaboración del alimento para que el riesgo de contaminación por microorganismos patógenos o alteración por microorganismos indicadores sea menor o dentro de un valor aceptable permitido en el producto final.

IV. METODOLOGÍA

En la figura 1 se presentan las etapas en las que se dividió la metodología de este trabajo.



miro

Figura. 1. Metodología

A. Revisión general y familiarización con la plataforma Qlik Sense, mongo DB y el proceso de ETL

Para la familiarización inicial con la plataforma se realizaron capacitaciones con el product manager, la consultora experta en inocuidad de oKuo y el value hunter de asimatrix, empresa que presta los servicios y recursos de ingeniería de software y analítica de datos, quienes enseñaron los fundamentos y todo el proceso general que soportan el funcionamiento de la plataforma.

- Qlik Sense: Para la revisión general de la plataforma se realizó la comprobación de todos los elementos visuales, indagando en el correcto funcionamiento de gráficos, botones y filtros, además de buscar errores generales como nombres incorrectos, campos erróneos en los gráficos, estadísticas mal planteadas, colores indebidos y estructuras incorrectas.
- Mongo DB: En primer lugar, se analiza la estructura de la base de datos, entendiendo las funciones de cada una, el origen de los datos y cómo estas se integran en la información final que llega a Qlik Sense
- ETL: a) Extract: De forma general, se revisa el proceso de extracción, el origen y los diferentes datos de los archivos, así como qué información es la más relevante y como es el proceso de extracción de esta mediante Python.
b) Transform: El proceso de transformación de la información se realiza en Python, se revisan los algoritmos empleados para este proceso y su funcionamiento.
c) Load: Este proceso se lleva a cabo en Python y los datos se cargan a la base de datos mongo DB. En este proceso no se realizaron ajustes debido a recomendaciones de la empresa.

B. Revisión y descarga de información de las plataformas para un análisis detallado en Python

Se descarga la información contenida en Qlik Sense en formato Excel, esta se ingresa a Python por medio de la librería pandas, con el fin de realizar análisis más detallados y simulaciones de algunos gráficos y estadísticos y así asegurar su correcto funcionamiento, además de comparar posteriormente esta información con la contenida en la base de datos para comprobar que no se hubieran generado errores en la carga desde mongo DB hacia Qlik Sense.

Para la revisión de los análisis de mongo BD, al igual que con QS, se usan principalmente Python y la librería Pandas, pero en este caso se comparan los datos almacenados en la base de datos con

los datos de los archivos enviados por los clientes para el proceso de ETL con el fin de determinar errores en este proceso.

C. Reestructuración de la arquitectura del código en Qlik Sense

Se indagó en la arquitectura del código, encontrando una estructura dispersa, en la cual cada gráfico y análisis usaba sus propias fórmulas, eso hacía que realizar modificaciones, crear nuevos gráficos y corregir errores fuera un trabajo extenso y complejo, por lo que añadir un nuevo campo a un gráfico de barras se convertía en un trabajo que podía llevar hasta quince días, ya que añadir este campo tenía que realizarse en la mayoría de las otras fórmulas siendo necesario hacer comprobaciones para cada una de ellas. Para resolver este problema, se empleó un esquema de fórmulas heredadas, en el cual las fórmulas del primer nivel se reutilizan en el segundo, las del segundo nivel en el tercero y así sucesivamente hasta la última fórmula más general.

Mientras se realizaba la reestructuración del código y teniendo en cuenta lo encontrado en la primera revisión general de los datos y las fórmulas con las simulaciones en Python, se realizaron algunos ajustes a las fórmulas, mayormente en el apartado que recalcula los ponderados, ya que estos no tenían realmente en cuenta los análisis que hacían parte de una muestra, sino que tomaban valores promedio, lo cual permitía que muestras sin algunos análisis tomaran valores del promedio y asignaran un valor para ese análisis en esa muestra.

D. Replantear la estructura de los datos presentados en mongo DB

Para el oKuo scopeTM se crearon varias bases de datos, entre ellas se encuentra una base de datos que almacena los límites microbiológicos (limits) y otra que almacena los registros con cada análisis, sus ponderados, los valores de los análisis entre otros (oKuo DB).

Los límites microbiológicos son los valores máximos que definen si un análisis de un microorganismo en una muestra cumple con un criterio de inocuidad, este criterio se asigna respecto a varias variables que, en general, son las normativas para alimentos a las oKuo recomienda ajustarse, pero estos límites pueden cambiar de acuerdo a las necesidades de cada empresa y a las actualizaciones que se den en las normas. Es por eso que es importante poder

actualizar estos límites de forma fácil y rápida en todo momento. Esto no se podía realizar con la estructura que se manejaba en las bases de datos de este momento, ya que los límites almacenados en la base de datos *limits* se asignaban a los registros en la propia base de datos, y si se realizaba algún cambio se debía recargar nuevamente toda la base de datos para ajustar todos los registros, proceso lento, laborioso y para nada recomendable dentro de las buenas prácticas para el manejo de bases de datos. Así, se realizaron ajustes que permiten a QS conectarse directamente a la base de datos *limits*, para desde allí asignar su límite a cada registro, esto permite que para cualquier cambio o actualización de los límites, sólo sea necesario realizar una modificación a la base de datos de *limits*, base de datos mucho más pequeña y manejable que la base de datos *oKuo db*, además de permitir agregar más especificidad al momento de asignar un límite, ya que desde QS se puede asignar un límite específico, a una empresa específica, para una zona específica de un microorganismo específico por ejemplo.

E. Complementar proceso de la ETL

Luego de realizado todo este proceso, se siguieron realizando comprobaciones en el código, agregando elementos visuales en la plataforma QS, revisando las bases de datos y apoyando el proceso de la ETL para mantener un producto en las mejores condiciones posibles para el cliente.

En el proceso de extracción de la información, se deben comentar los archivos que mandan los clientes, generalmente en formato Excel, estos comentarios van dirigidos al personal de Asimetrix que realiza este proceso, con el fin de guiarlos para lograr extraer toda la información relevante de forma correcta. Para este proceso se recomendó capturar campos adicionales como fecha de toma de muestra, fecha de recepción y fecha de análisis, para revisar el comportamiento de los valores microbiológicos de las muestras cuando estos tiempos varían, además de capturar variables como temperatura de la muestra, nombre del técnico que realiza el muestreo, nombre del analista de laboratorio y observaciones, lo que permite realizar análisis más específicos sobre el comportamiento de las muestras.

Sobre el proceso de la extracción se encontraron algunas prácticas que debían ser corregidas ya que podrían afectar los resultados del índice, entre estas cuando aparecía un valor como <10 el dato pasaba como 10 y no como 9 o 9.9, aspecto importante sí el límite microbiológico de ese análisis coincide con 10, lo que haría que una muestra con valor <10 se procesara como si

hubiera cumplido con el valor microbiológico y por consiguiente con la norma para ese microorganismo o con las especificaciones definidas por el cliente.

Un proceso importante dentro de la transformación de los datos es el de reasignar un nombre a una muestra, esto se realiza con el objetivo de estandarizar los nombre de las muestras que se presentan en la plataforma QS, como por ejemplo cuando un cliente puede entregar dentro de su formato muestras con nombres como maíz Colombia, maíz Perú, maíz amarillo y maíz, es importante crear un nuevo campo para todas estas muestras con el nombre de maíz para ser presentadas de forma limpia y concreta dentro del índice (aunque dentro QS aún se tiene la posibilidad de ver el nombre real de la muestra).

Este proceso consta de varios pasos, el primero es el de un etiquetado manual, en el que se revisa el nombre de la muestra (llamada muestraReal) y se le asigna un nuevo nombre (muestra), luego de este paso inicial se guardan estos nombres en una librería, que ayuda a automatizar el proceso, así cuando llega una nueva muestraReal, esta primero se busca en la librería, y si hay un nombre guardado para esta, se le asigna y así no es necesario volverlo a fijar manualmente.

El problema con este proceso es que se estaba realizando por medio de una librería en Python llamada fuzzywuzzy, esta librería asigna un valor de similitud entre dos palabras, así, si una palabra supera un umbral de similitud con otra, la librería realiza la asignación. Esta librería se emplea con el fin de que si una muestraReal, por ejemplo, viene con el nombre maíz ecuador y dentro de la librería hay una muestraReal con el nombre maíz ecuadar, se pueda asignar el nombre de muestra como maíz y no tener que fijar el nombre de muestra para maíz ecuadar manualmente, ya que un cliente puede mandar 60.000 registros y esto se vuelve muy difícil de procesar manualmente, pero en este proceso el valor de tolerancia escogido para el fuzzywuzzy era muy bajo, lo que permitía que muestras sin suficiente semejanza se asignaran, por ejemplo que para maíz ecuador se asignaran nombres como zaranda. Para esto se reajustó el tipo de algoritmo de emparejamiento empleado por la librería fuzzywuzzy y su porcentaje de tolerancia para evitar que muestras con semejanza muy lejana pudieran ser emparejadas.

F. Pruebas e implementación

En este proceso se recolectaron todos los cambios realizados en una nueva aplicación en QS, la cual fue sometida a pruebas por todo el equipo para verificar su robustez y que todos los cambios realizados fueran efectivos. De este proceso se obtuvieron resultados satisfactorios por lo que está pasó a ser la nueva aplicación que se usa actualmente para realizar presentaciones a los clientes.

G. Aseguramiento de la calidad y mejora continua

Luego de realizado todo este proceso, se siguieron realizando comprobaciones en el código, agregando elementos visuales en la plataforma QS, revisando las bases de datos y apoyando el proceso de la ETL para mantener un producto en las mejores condiciones posibles para el cliente.

V. RESULTADOS Y ANÁLISIS

En Qlik Sense se logra obtener un producto de nivel profesional, en el cual todos los cálculos, filtros, gráficos funcionan de forma correcta, además de crear una nueva arquitectura del código que permite una implementación de nuevos gráficos, cálculos y modificación del programa mucho más fácil y rápida, permitiendo pasar de un tiempo de espera de 15 días, a un día en la implementación de un nuevo nivel en los gráficos.

En la figura 2 se logran observar 6 funciones que inician con IIT, estas son todas las nuevas fórmulas necesarias para realizar el 98% de los cálculos y la creación de gráficos dentro del oKuo scope™, estas funciones están relacionadas de forma jerárquica, por lo que el cambio de una fórmula en un nivel inferior se verá reflejado consecuentemente en las fórmulas de los niveles superiores. Estas 6 fórmulas reemplazan las 25 fórmulas empleadas anteriormente, las cuales no tenían una estructura jerárquica y cualquier cambio tenía que ser realizado para todas las demás fórmulas.

Las figuras 3 y 4 presentan una comparación del estado inicial y final de la hoja principal del índice, en las cuales se logra observar una interfaz más completa en la que se puede visualizar mucha más información. (El oKuo scope antiguo fue eliminado por completo, por lo que no fue posible capturar imágenes del estado anterior para algunas comparativas)

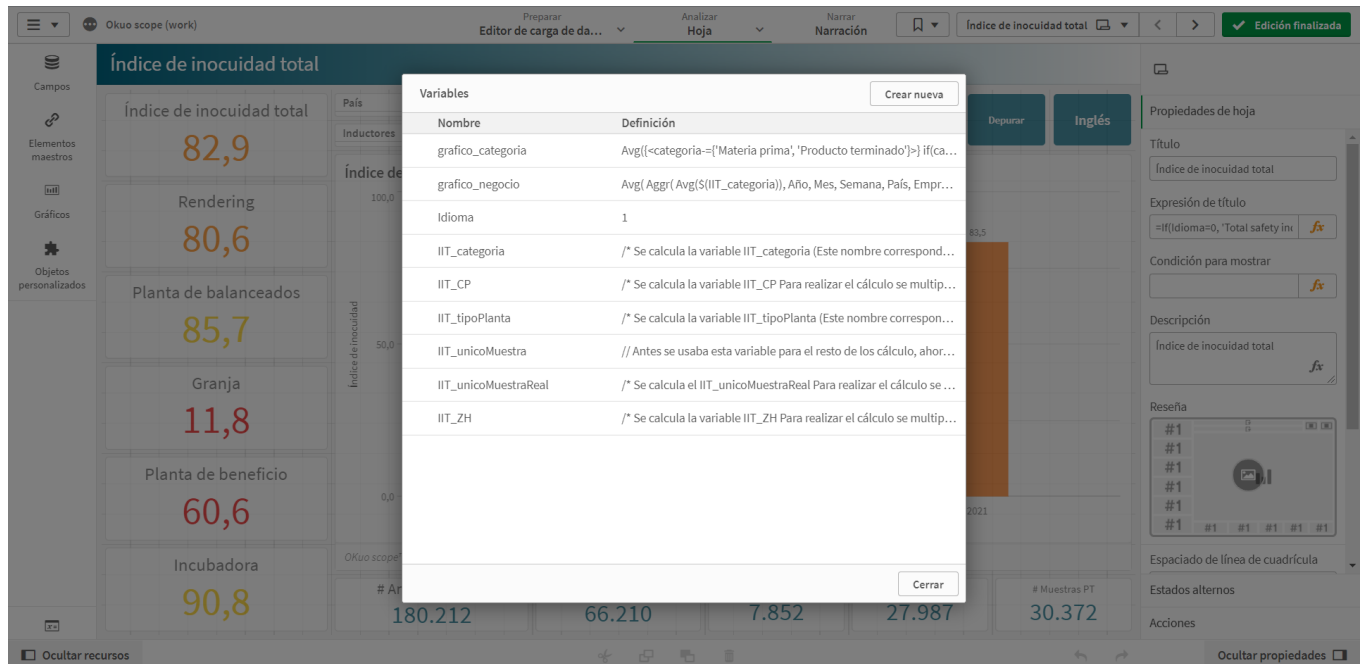


Figura 2. Nuevas fórmulas empleadas para el cálculo total del índice



Figura 3. Hoja principal del índice antes de los cambios

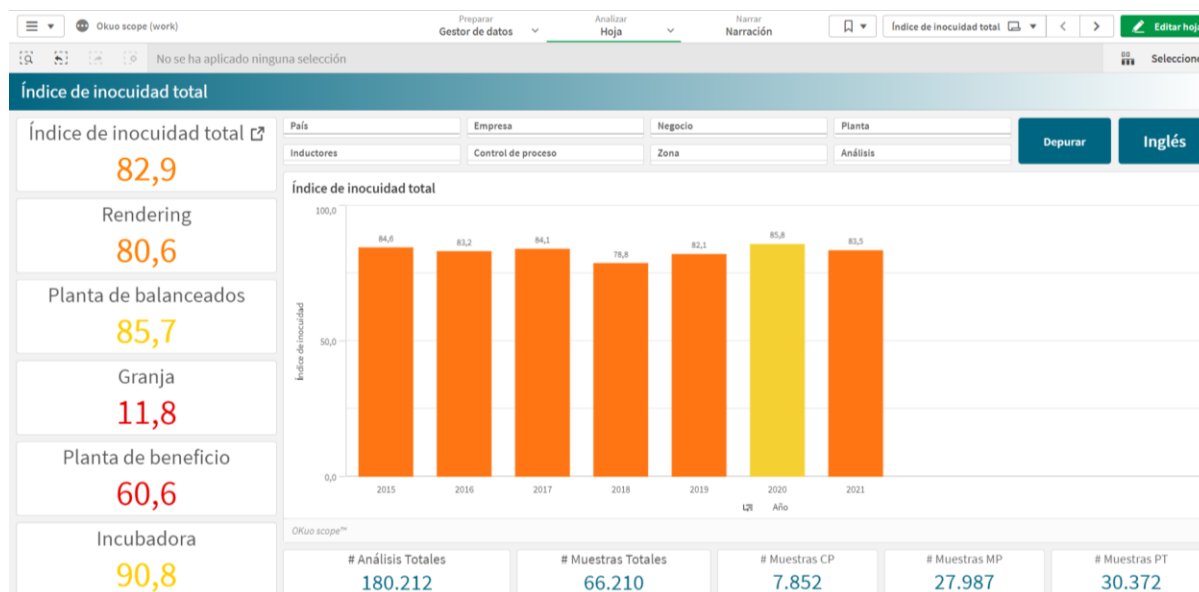


Figura 4. Nueva Hoja principal del oKuo scope™

MongoDB

Desde la base de datos, se logró crear una nueva estructura para conectarse con QS, esta estructura consiste en realizar una conexión directa desde la base de datos limits con QS, para allí emplear las funciones map y applymap con el fin de realizar la asignación de los límites a los registros de la base de datos oKuo scope™ de acuerdo a los criterios definidos en limits, lo cual permite una mayor flexibilidad en cuanto a la definición de los límites por medio de los cuales se realizan los cálculos de los valores de inocuidad. Esto es importante, ya que, aunque estos límites fueron definidos en base a las normas que rigen las plantas de alimentos, y la experiencia en microbiología de la empresa oKuo, los clientes pueden solicitar que su información sea tratada en base a unos límites y ponderados diferentes de acuerdo con sus necesidades particulares.

ETL

- **Extracción**

En el proceso de extracción de los datos, se encontraron varios errores al no tratar de forma adecuada los archivos procedentes de los clientes, esto fue solucionado luego de discutir con la experta en inocuidad de la empresa sobre la necesidad de capturar más datos de los archivos y de como tratar de forma correcta cada uno de los campos en los archivos, capturando nuevas columnas como las fechas de toma de muestras y de análisis.

- **Transformación**

Para el proceso de transformación se especificaron los criterios para el tratamiento de los valores microbiológicos en caso datos como <10 , $>10^5$. Además de reestructurar los parámetros de la librería fuzzywuzzy permitiendo la correcta asignación de el nombre muestra a la muestra real.

- **Carga**

Al momento de la carga de datos, se encontraron errores como datos duplicados y datos viejos con un mal proceso de carga y extracción que se combinaban con los datos nuevos y bien cargados, lo que provocaba que los cálculos en QS no fueran correctos tuvieran un comportamiento extraño y no fuera acertados. A partir de esta corrección y de las demás mencionadas anteriormente, la información presentada en el oKuo scopeTM pudo ser considerada como una representación acertada sobre el estado microbiológico de las plantas de los clientes.

El todo el proceso de aseguramiento de la calidad, se añadieron algunas características oKuo scopeTM que no fueron mencionadas anteriormente, en esta tabla se mencionan estos cambios y se resumen los mencionados anteriormente

Tabla 1. Comparativo antes y después de las mejoras

Antes	Después
Los cálculos llegan hasta el nivel 4 dentro de la jerarquía del índice (Inductores)	Ahora los cálculos se realizan hasta el nivel 5 de la jerarquía (Negocios)
Los gráficos descienden hasta el nivel de mes	Se incluye el nivel semana en los gráficos
El índice sólo cuenta con la opción en español	Se añade un botón que traduce la hoja principal del índice a inglés
Sólo se presentan los datos de todas las fuentes en el índice	Se añade un botón que filtra sólo las fuentes confiables con buena información
Cada gráfico y elemento visual tiene sus propias fórmulas	Todas las fórmulas se relacionan de forma jerarquita, por lo que la misma fórmula se puede usar para varios gráficos

<p>No se tomar varios campos como fechas de toma de muestras, fechas de recepción y fecha de análisis en la ETL</p>	<p>Se capturan más campos de los archivos de los clientes en la ETL</p>
<p>La asignación de los nombres de las muestras en la ETL se realiza con un algoritmo y un límite de coincidencia muy alto que permite malas asignaciones</p>	<p>El nuevo algoritmo y los nuevos límites de coincidencia permiten la correcta asignación de los nombres de las muestras en la ETL</p>
<p>La estructura de la base de datos no permite una asignación dinámica de los límites de los valores microbiológicos</p>	<p>La nueva estructura en mongo DB permite una asignación rápida y flexible para los límites microbiológicos</p>
<p>Se cuenta con gráficos hasta control de proceso</p>	<p>Se añade la hoja informe general donde se presentan valores por país, empresa, negocio, planta, inductores y control de proceso (Figura 5 y 6)</p>

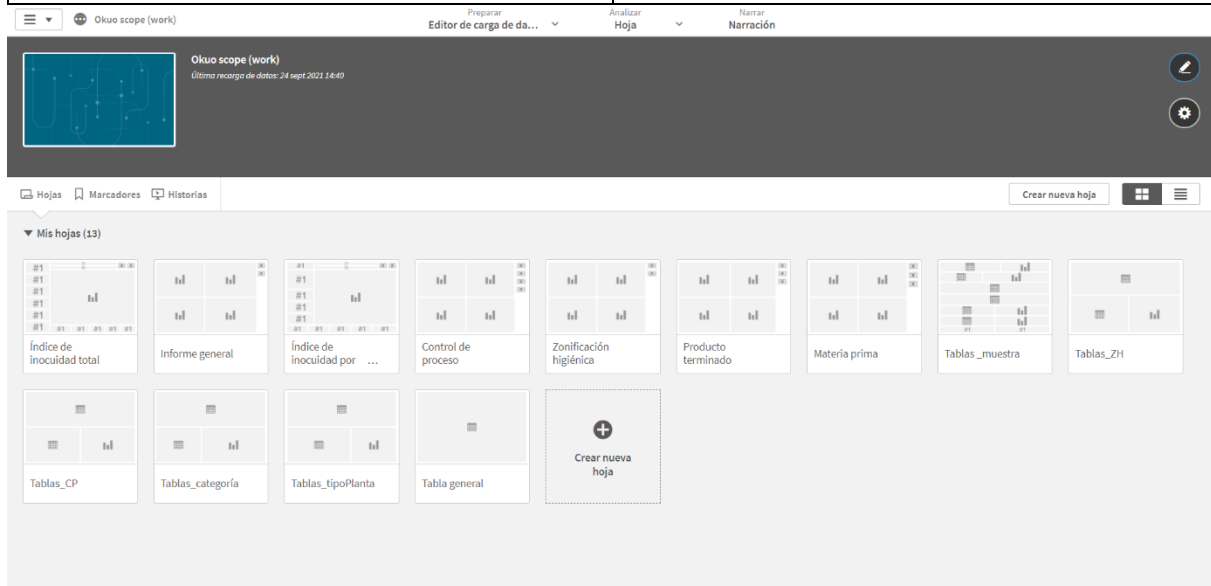


Figura 5. Esquema general oKuo scope™



Figura 6. Vista de informe general en la plataforma

VI. CONCLUSIONES

- Se lograron detectar y corregir los errores de tipo estadístico en la estructura del código, los cuales se encontraban localizados y relacionados con todas las partes que componen el oKuo scope™.
- Se aplicaron mejoras en la gestión de la plataforma QS creando una nueva arquitectura que optimiza la adición de nuevos cálculos, gráficos y corrección de errores.
- Se incorporó nueva información al modelo, modificando el proceso de la ETL para extraer y almacenar nuevos datos en MongoDB, lo permitirá a corto plazo implementar nuevos análisis y estadísticos en QS.
- Al conseguir un oKuo scope™ con cálculos e información correcta, se hizo posible, crear informes para algunas empresas, sobre el estado microbiológico de su negocio, permitiendo dar una idea global de las acciones que se deben tomar a corto y largo plazo para incrementar la inteligencia microbiológica en cada parte de sus plantas

REFERENCIAS

- [1] Pandas - Python Data Analysis Library. (n.d.). Retrieved June 8, 2021, from <https://pandas.pydata.org/>
- [2] Qlik Sense | Data Analytics Platform. (n.d.). Retrieved June 8, 2021, from <https://www.qlik.com/us/products/qlik-sense>
- [3] Why Use MongoDB & When to Use It? | MongoDB. (n.d.). Retrieved September 5, 2021, from <https://www.mongodb.com/why-use-mongodb>