Dashboard de indicadores de ventas

Joan Sebastián Iragorri Barreto, Diana Carolina Cárdenas, Álvaro Mauricio Zamora Roa ,Eseit, Bogotá, Especialización en Big Data y Analítica de datos.

*Resume*n – En este trabajo se muestra todo el proceso que se realizó para la creación de un dashboard de seguimiento de ventas el cual es generado mediante un proceso automatizado de ETL y con actualización en tiempo real, el cual se incluirá graficas en lenguaje R el cual fue aprendido en el curso.

*Palabras clave*

1. *Dashboard: Herramienta utilizada para gestionar un conjunto de datos, generando indicadores de gestión mostrándolos de manera visual mediante gráficos.*
2. ***KPI: Indicador de clave de actuación o indicadores de calidad de negocio.***
3. ***Power bi: Herramienta de visualización de informes web***
4. ***Gráfica: Tipo de representación de datos visuales***
5. ***Indicador: Pequeños medidores que comunican el estado de un valor único.***
6. ***SQL: Lenguaje estructurado para la manipulación masiva de dato***
7. ***ETL: Sistema de extracción, transformación y carga de datos de forma programada***
8. ***R: Lenguaje de programación orientado a manipulación de datos y generación de gráficas***
9. ***Informe: Expresión oral o escrita sobre el estado de una cosa o acción***
10. ***CSV: Formato de archivos el cual se caracteriza por estar almacenado en un archivo de texto con un delimitador para separar la información***
11. ***GIT: Sistema de control de versiones y trabajo colaborativo para trabajar en cualquier entorno de desarrollo.***

*Abstract—* *This work shows the entire process that was carried out to create a sales tracking board which is through an automated ETL process and updated in real time.*

# **Introducción**

En las empresas actuales existen diversas formas de presentar los resultados que se obtienen, ya sea por medio de informes de forma escrita, oral o digital. Los avances tecnológicos han permitido que por medio de herramientas ofimáticas como por ejemplo el paquete de office de Microsoft sean utilizadas para mostrar los resultados ya que son de fácil utilización y se pueden compartir a un cierto número de personas de manera simultánea (correo electrónico).

En este caso puntual donde va enfocado el proyecto que es una empresa de Ventas, la cual desea desarrollar un proceso de automatización de informes a partir de sus datos generados mediante archivos planos de texto sin ningún tipo de tratamiento ni limpieza. Por lo cual se realizará todo un proceso para limpiar estos datos, alimentar un servidor, diseñar un dashboard con los datos más relevantes encontrados y generar una automatización completa para que el informe se actualice de forma independiente y no requiera de un usuario.

# **Objetivos (general y Específicos)**

**Objetivo general:** Crear un dashboard permita realizar seguimiento de todos los indicadores (KPI) que maneja la empresa de ventas mediante gráficos estadísticos en la herramienta Power Bi.

**Objetivos específicos:**

* Realizar proceso de limpieza, alistamiento, depuración, cargue de las bases de datos entregadas en CSV.
* Crear un proceso ETL que permita realizar la automatización del proceso.
* Utilizar el lenguaje de programación R para manipular datos y generación de gráficas estadísticas que permitan analizar y encontrar información de valor para la toma de decisiones

# **METODOLOGIA**

La empresa que nos solicita el modelo de automatización nos realizó entrega de los archivos donde maneja la información la cual se actualiza de manera diaria y se almacena en archivos planos en formato CSV.



Fig 1. Archivos CSV Entregados

Nuestro equipo de trabajo recibió los archivos y se creó un repositorio en GIT para manejar y controlar las versiones de los archivos a generar, adicionalmente para trabajar de forma segmentada en el proyecto, el cual tendrá una fase de ingeniería de datos y otra parte de análisis y desarrollo del dashboard.

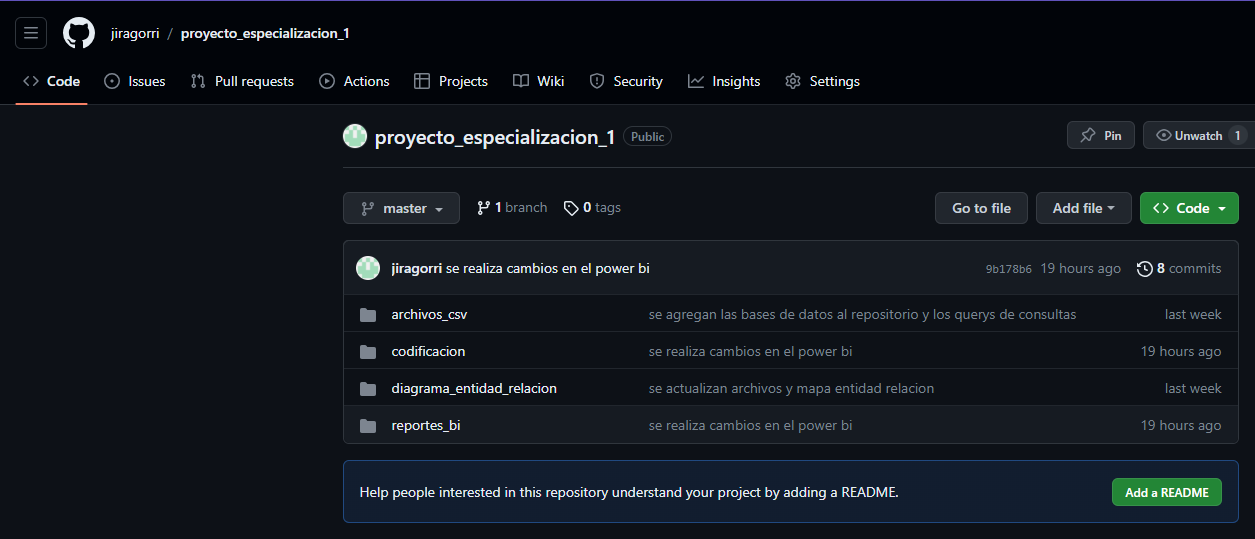


Fig 2. Creación del repositorio

Por medio de la herramienta ETL llamada Talend ESB se realizó el proceso de la creación de la automatización, limpieza y cargue de las bases a un servidor de bases de datos, que para este proyecto se implementó sobre SQL Server.

La cual toma cada uno de los archivos csv los carga, realiza la depuración para identificar que tipo de datos es y luego los transforma en un tipo de dato que el servidor de SQL server reconozca

Diagrama

Descripción generada automáticamente

Fig 3. Creación de ETL proyecto

La cual toma cada uno de los archivos CSV los carga, realiza la depuración para identificar qué tipo de datos es y luego los transforma en un tipo de dato que el servidor de SQL server reconozca, adicionalmente el ingeniero de datos define en cada base de datos las llaves primarias y foráneas.

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Fig 4. Transformación de datos y definición de claves primarias

Luego de la correspondiente configuración de la ETL, se configura para correr el proceso y realice el cargue de los archivos planos al servidor. Teniendo en cuenta que no son muchos registros para iniciar, carga la información de manera eficiente y rápida, adicional el servidor que almacenará la base de datos de acuerdo a las configuraciones establecidas en Talend ya tiene los datos listos para procesar y trabajar.

Interfaz de usuario gráfica, Aplicación, Word

Descripción generada automáticamente

Fig 5. Cargue de bases de datos en servidor SQL Server

Teniendo las bases cargadas, se realiza el diseño del modelo entidad relación de las bases de datos, con la finalidad de establecer una relación entre ellas y poder realizar unificación de un reporte que abarque todas las necesidades de la empresa de ventas.

Diagrama

Descripción generada automáticamente

Fig 6. Modelo entidad - relación

Ya estableciendo la relación entre las bases de datos, se procede a realizar las consultas sql necesarias para armar nuestra tabla que tenga toda la información necesaria para construir el dashboard, la cual se genera por medio de 2 querys, el primero generando una vista para la tabla de fechas y el segundo para construir el informe final.

En el query de fechas se realiza un segundo proceso de transformación ya que la fecha que viene en el archivo plano trae una hora que no necesitamos y no da valor a los datos, adicional utilizamos esa fecha para traer datos como año, mes, día, nombreDia, nombreMes, Trimestre. Esto con la finalidad de poderlos usar como comparativos o filtros mediante segmentaciones de datos en nuestro informe final.

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Fig 7. Generación consulta SQL para generar vista de fechas

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Fig 8. Generación consulta SQL para generar vista de informe

Teniendo como resultado una tabla de datos final lista para ser cargada en power bi, en donde realizaremos todo el proceso de análisis, generación de KPI y gráficas que nos permitan realizar comprensión de los datos generando información de valor

El proceso de carga a power bi se realiza mediante la herramienta de vinculación de datos que ya incluye la herramienta

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

Fig 9. Conexión en power bi a SQL Server

Luego power bi nos permite seleccionar la base de datos con la cual se utilizarán los datos para trabajar, en este caso utilizaremos la vista que se generó para trabajar con todos los datos unificados

Interfaz de usuario gráfica, Aplicación, Tabla

Descripción generada automáticamente

Fig 10. Selección de los datos transformados para trabajar

De forma paralela se realiza conexión de la base de datos a RStudio para el análisis de la base de datos y generación de gráficos de valor con los conceptos aprendidos en la clase.

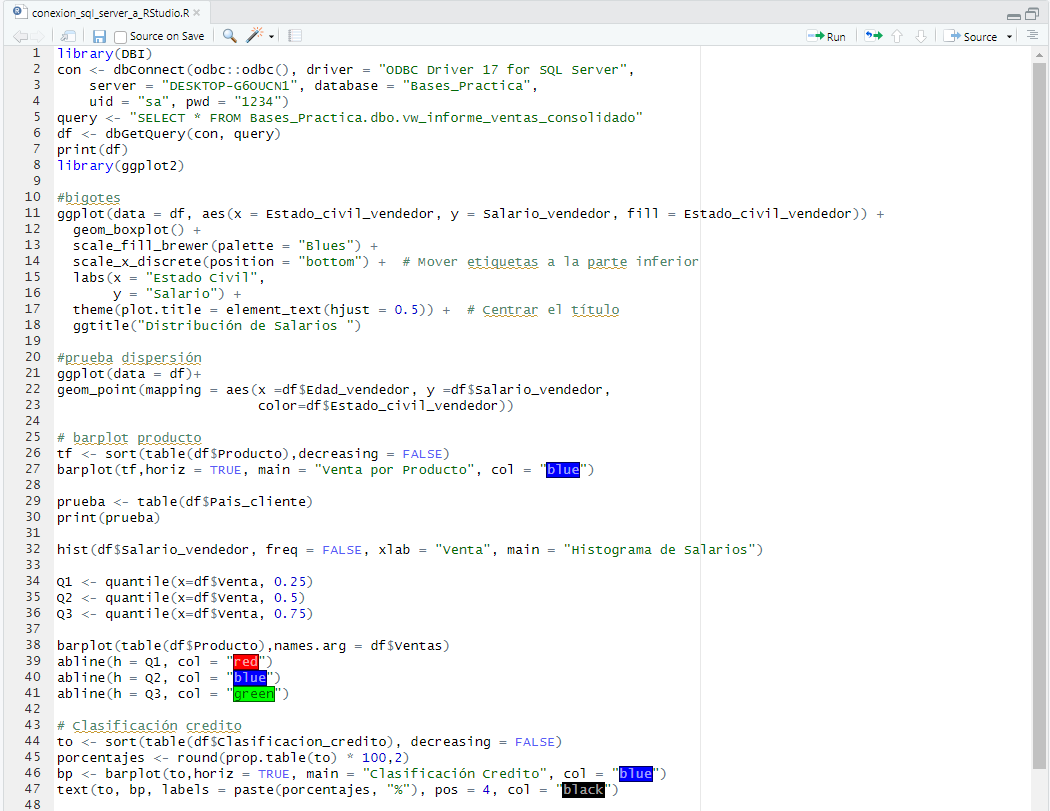


Fig 11. Codificación en RStudio

Finalmente, se procede a realizar toda la construcción del tablero de control mediante power bi, en el cual se realizó previamente el análisis de la base de datos para determinar que datos debíamos representar y que tipo de gráficos podíamos utilizar, por lo que se utilizó diferentes tipos de gráficos y de igual forma se usaron tres gráficos que diseñamos previamente mediante RStudio

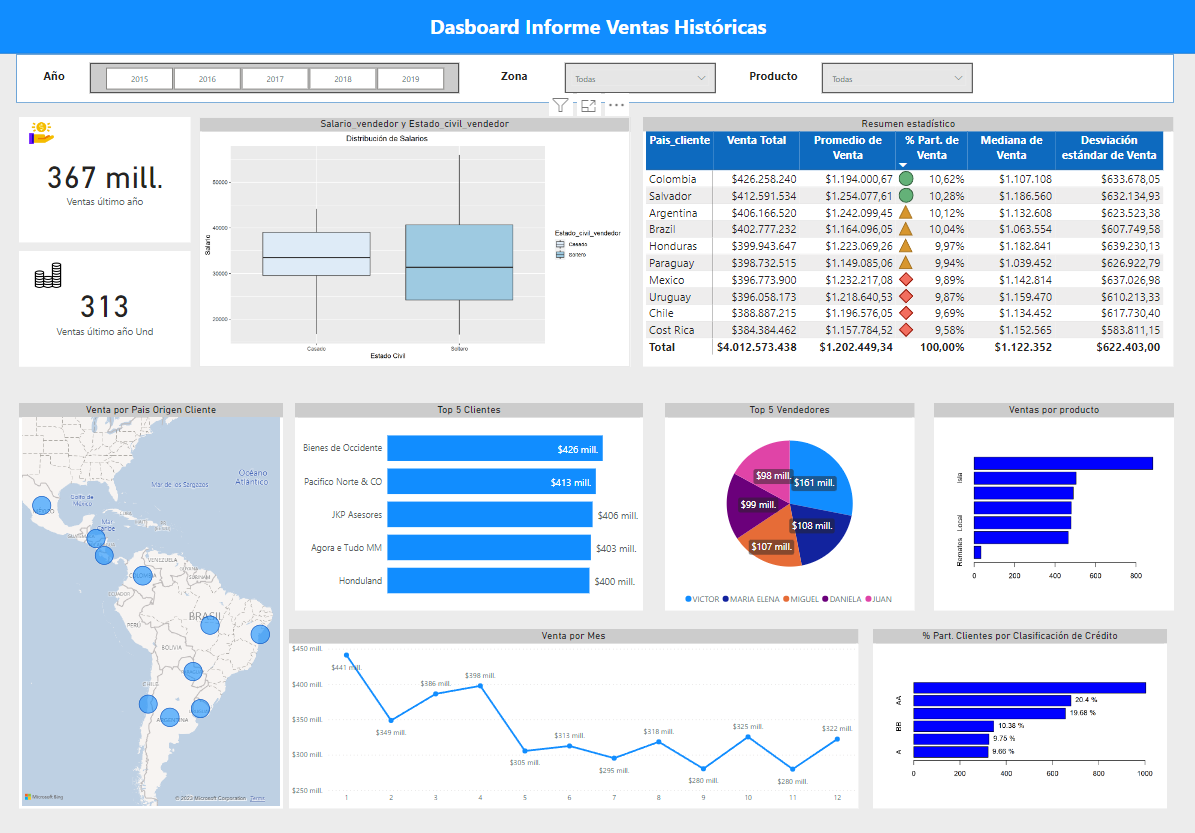


Fig 11. Diseño final Dashboard

# **CONCLUSIONES**

Por medio de los programas utilizados en este trabajo se identificó que se puede crear un sistema automatizado de información que permita conocer los indicadores en tiempo real, todo con un análisis previo de la unidad de negocio a utilizar, el estudio de las fuentes de información y las bases de datos y las herramientas que nos permiten realizar todos estos procesos de análisis de datos.

Adicionalmente se identificó que la estadística es indispensable para la creación de tableros de control debido a que por medio de este estudio se puede contar por medio de los datos y las gráficas los diferentes aspectos analizados para una mayor comprensión y la ayuda para la toma de decisiones.

# REFERENCIAS

[1] SQL DE CERO A EXPERTO:

<https://cursos.a2capacitacion.com>

[2] R Y R STUDIO DE CERO A MASTER DATA:

<https://cursos.a2capacitacion.com>