

Instituto Tecnológico y de Estudios Superiores de Occidente

Reconocimiento de validez oficial de estudios de nivel superior según acuerdo secretarial 15018, publicado en el Diario Oficial de la Federación el 29 de noviembre de 1976.

Departamento de Electrónica, Sistemas e Informática

MAESTRÍA EN INFORMÁTICA APLICADA



**ANÁLISIS PARA LA PLANEACIÓN ESTRATÉGICA DE LA PRODUCCIÓN DE
BERRIES EN MÉXICO MEDIANTE LA IMPLEMENTACIÓN DE *BUSINESS*
*INTELLIGENCE***

Tesis para obtener el grado de

MAESTRO EN INFORMÁTICA APLICADA

Presentan: ISC. Pablo de Jesús Kersey Maldonado

Asesora: Dra. Gabriela Calvario Sánchez

San Pedro Tlaquepaque, Jalisco. Enero de 2022

Dedicatoria

A mis papás Paul y Salud, y mi hermano Vicente; son mi pilar, mi fuerza, mi orgullo, y mi razón de ser. Los amo.

A mis abuelos, Nicandro†, Celia, Edward†, Marian†; me regalaron la mejor familia del mundo.

Agradecimientos

Siendo Ingeniero en Sistemas Computacionales, jamás pensé que mi desarrollo profesional fuera a ocurrir dentro del sector agrícola. Me alegro de haber estado equivocado, sin embargo, esto fue gracias al haber encontrado a personas con la disposición y paciencia para forjarme en este rubro. Hace 8 años comencé en este sector cuando tuve la fortuna de conocer a un grupo de profesionistas agrícolas que se convirtieron en mi segunda familia: Juan Carlos Cabrera, Eduardo Camargo, Flavio Lara, Eduardo Tirado y Gerardo Cruz. A todos ustedes, mi mayor agradecimiento por haber tenido la paciencia, la apertura y la confianza de enseñarle a este neófito de las plantas todas las riquezas que nos puede dar el campo y lo importante que es entenderlo, cuidarlo y desarrollarlo.

Después de este primer acercamiento vino una segunda fase donde pude compartir grandes experiencias profesionales, proyectos de gran trascendencia en esta industria, y donde mis primeros aprendizajes tomaron rumbo claro. Esto no hubiera sido posible sin la guía y el apoyo de mis grandes amigos Yadira Godínez, Rafael Vázquez y Mario Uribe, quienes fueron clave en el desarrollo del presente proyecto. A ustedes, mi admiración y agradecimiento por haber estado siempre en el lugar y momento adecuado, personal y profesionalmente.

Es una larga lista de todos aquellos profesionistas, amigos y amigas que me han acompañado en estos años: Cristina Meza, Enot Hueda, Claudio Medrano, Fernando Mendoza, Luis Guerrero, Punit Parmar, Juan Castagnino, y tantos otros que con sus consejos y sus retos se fueron formando valiosas ideas para mejorar el campo mexicano.

A los directivos y colaboradores de la empresas LRDA, gracias por haberme dado la apertura de trabajar junto con ustedes y conseguir los datos para realizar esta propuesta académica.

Agradezco a las autoridades y docentes del ITESO que me acompañaron durante el desarrollo de este proyecto. Al Mtro. Ricardo Salas Mejía, coordinador de la Maestría en Informática Aplicada, por estar siempre atento y disponible para atender mis dudas. A mi directora del Trabajo de Obtención de Grado (TOG), la Dra. Gabriela Calvario Sánchez, quien me dio guía y acompañamiento para hacer realidad este documento. A toda la plantilla docente, en especial al Dr. Humberto Pérez, al Mtro. Roberto Martínez, y al Dr. Iván Esteban Villalón, cuyas asinaturas fueron clave para mi TOG.

Finalmente, gracias al apoyo de mis padres y mi hermano, quienes siempre me han alentado a conseguir mis metas, a no perder el piso, y a buscar cada día ser la mejor versión de mi mismo.

Resumen

México es un país que cuenta con una importante diversidad de productos agrícolas debido a su situación geográfica privilegiada, siendo las berries uno de los productos destacados dentro de las cifras económicas de exportaciones agropecuarias. Para lograr consolidarse como una potencia en la exportación de berries durante todo el año, es necesario que las unidades de producción agrícola cuenten con las capacidades de análisis de información adecuadas para entender su situación actual y desarrollar planes de acción que permitan incrementar sus rendimientos, ayudando a mejorar la calidad de los productos y logrando la rentabilidad requerida para sostener sus planes de crecimiento. La Inteligencia de Negocios (*Business Intelligence* o BI) es una metodología sólida que ha sido usada por diferentes industrias para diseñar su planeación estratégica y tomar las decisiones correctas para alcanzar sus objetivos estratégicos. Este documento pretende demostrar que las capacidades de BI se encuentran al alcance de las unidades de producción agrícola de berries, y que aportan positivamente a superar los retos que enfrentan los productores en México, generando las capacidades necesarias para el análisis e interpretación de información, así como para la conformación de un plan estratégico sustentado en datos.

Tabla de contenidos

Agradecimientos	II
Índice general	V
Índice de figuras	VII
1. Introducción	1
1. Antecedentes	1
2. Definición del problema	2
3. Objetivo general	2
4. Objetivos específicos	3
5. Justificación	3
2. Estado del arte	5
3. Marco teórico	9
1. Producción de berries en México	9
1.1. Fresa	12
1.2. Frambuesa	12
1.3. Zarzamora	12
1.4. Arándano	13
2. Business Intelligence	13
2.1. Fuentes de datos	17
2.2. Almacenamiento de datos	18
2.2.1. Data Warehouse	18
2.2.2. Data Mart	21
2.2.3. ETL	21
2.3. Modelo semántico	22
2.4. Herramientas de visualización y análisis de datos	24
4. Desarrollo Metodológico	27
1. Obtención de datos	28
1.1. Determinación de métricos y dimensiones	29
1.2. Identificación de las fuentes de datos	31
1.3. Selección del tipo de conexión	31
2. Preparación de datos	32
2.1. Proceso de preparación de los datos	33
2.1.1. Definir encabezados	34
2.1.2. Rellenar celdas nulas de las columnas de tipo dimensión	34
2.1.3. Traducir términos al español	34
2.1.4. Eliminar precio de venta	35

2.1.5.	Calcular la fecha de cierre de semana	35
2.1.6.	Separar un DataFrame para volumen y otro para superficie	36
2.1.7.	Asignación de índices	37
2.1.8.	Eliminar superficies nulas y duplicadas del DataFrame de superficie	37
2.1.9.	Guardar los DataFrames para su posterior uso en los modelos	38
3.	Exploración de datos	40
3.1.	Conectando Tableau Desktop a la fuente de datos	40
3.2.	Exploración simple de datos	42
3.2.1.	¿Cuál ha sido la evolución de la agrícola a través de los años?	42
3.2.2.	¿Será que el efecto anterior se presenta en todos los cultivos?	43
3.2.3.	Uso de tableros de control	44
3.3.	Exploración avanzada de datos	45
3.3.1.	¿Cuáles han sido los rendimientos históricos de cada cultivo?	45
3.3.2.	¿Cuáles variedades han impactado el rendimiento promedio en fresa?	46
3.3.3.	¿Cuáles ranchos han impactado el rendimiento promedio en zarzamora?	47
3.3.4.	¿El caso del Rancho2 de zarzamora se debe a todas las variedades o a alguna en específico?	47
3.3.5.	Comentarios de cierre	48
5.	Resultados	50
1.	Resultados	50
2.	Discusión	52
6.	Conclusiones	54
1.	Conclusiones	54
2.	Trabajo futuro	55
	Glosario	56
	Referencias	61

Índice de figuras

2.1. Objetivos de Desarrollo Sostenible acordados en el 2015	6
2.2. Estrategias de mercado para lograr objetivos estratégicos de las berries en México	7
3.1. Algunos tipos de berries	10
3.2. Riego por goteo	11
3.3. Fresa en macrotúnel	11
3.4. Componentes internos y externos de un BI	14
3.5. Arquitectura de un BI de Salgueiro	16
3.6. Arquitectura de un BI de Rad	16
3.7. Arquitectura de un BI de Rad	17
3.8. Comparación de la estructura del modelo de Kimball y el modelo de Inmon	18
3.9. Comparativa entre Kimball e Inmon.	19
3.10. Comparativa entre modelo estrella vs modelo copo de nieve.	19
3.11. Comparación de tablas de una base de datos relacional y una base de datos dimensional	20
3.12. Esquema en estrella de la tabla de hechos con cuatro tablas de dimensiones	21
3.13. Ejemplo de un ETL en Tableau Prep	22
3.14. Interpretación gráfica de un cubo de datos	24
3.15. Cuadrante Mágico de Gartner para BI y análisis de información	25
3.16. Dashboards de plataformas de análisis y visualización de datos	26
4.1. Proceso de análisis de datos para la toma de decisiones	28
4.2. Arquitectura. Fase de selección de la fuente de datos	29
4.3. Archivos fuente de las últimas 7 temporadas	32
4.4. Extracto de uno de los archivos con sus datos crudos	32
4.5. Esquema de datos original cargado en Python y Pandas	33
4.6. Arquitectura. Fase de preparación de datos usando Python y Pandas	33
4.7. Agregando encabezados a los datos crudos	34
4.8. Tabulando los datos	34
4.9. Traduciendo de inglés a español	35
4.10. Eliminando precio de venta	35
4.11. Calculando la semana	36

4.12. Separando los DataFrames	36
4.13. Asignación de índices	37
4.14. Eliminando superficies repetidas	38
4.15. Arquitectura. Fase de almacenamiento	38
4.16. DataFrame de superficie después de su procesamiento	39
4.17. DataFrame de volumen después de su procesamiento	39
4.18. DataFrame general después de su procesamiento	39
4.19. Arquitectura. Fase de análisis y visualización	40
4.20. Conexión a Excel desde Tableau	41
4.21. Estableciendo la fuente de datos	41
4.22. Estableciendo relaciones en Tableau	42
4.23. Evolución de la agrícola en superficie y volumen.	43
4.24. Evolución de la agrícola en volumen.	43
4.25. Desglose por cultivo de superficie y volumen.	44
4.26. Ejemplo de un tablero de control	45
4.27. Rendimiento histórico por cultivo en kg/ha	46
4.28. Rendimiento histórico por variedad de fresa en kg/ha	47
4.29. Rendimiento histórico por rancho de zarzamora en kg/ha	48
4.30. Rendimiento histórico del Rancho2 en kg/ha	48
5.1. Evolución del BI	53

Capítulo 1

Introducción

Fomentar la agricultura, animar la industria y proteger el comercio son los tres importantes objetos que deben ocupar la atención
— MANUEL BELGRANO, intelectual, economista, periodista y político argentino
(1770 – 1820)

1. Antecedentes

El Atlas Agroalimentario es una publicación mexicana anual creada por la Secretaría de Agricultura y Desarrollo Rural (SAGARPA) a través del Servicio de Información Agroalimentaria y Pesquera (SIAP). Contiene información sobre los principales productos agrícolas, pecuarios y pesqueros de México, como lo es el volumen de producción, el valor comercial y la evolución del comercio exterior [1]. En su versión más reciente, se calcula que, durante el 2018, el valor del comercio exterior anual agroalimentario alcanzó los 34,849 millones de dólares, siendo la producción de **berries** un elemento fundamental representando el 6.5% de esta cifra con un valor de exportación de 2,282 millones de dólares. Estas cifras ubican a las **berries** en el tercer lugar de importancia como producto agroalimentario de exportación solo detrás de la cerveza y el aguacate, y superando a productos como el jitomate y el tequila [2]. Con esto, México se coloca como el cuarto exportador más importante de **berries** a nivel mundial, destacando ser el segundo lugar en fresa y frambuesa [1].

Aunque existen una serie de empresas reconocidas mundialmente y dedicadas a comercializar estos productos, como *Driscoll's*, *Dole* y *Hortifrut*, por mencionar algunas, la labor de la producción recae en los agricultores, ya sean pequeños productores o grandes corporativos de producción. Estos productores afrontan una serie de retos constantes para mantenerse competitivos en el mercado y mantener la **rentabilidad** de sus cultivos. Sin embargo, la producción de **berries** en México presenta retos y oportunidades distintos a aquellos que ostentan otros cultivos como los extensivos (maíz, trigo, sorgo, etc.) o los que se producen en invernaderos de alta tecnología (jitomate, pepino, pimientos, entre otros).

Uno de los puntos débiles en la operación de los productores de berries es su baja capacidad de planeación estratégica, que sigue mayormente basada en el conocimiento empírico del campo, en lugar de estar sustentada en datos confiables y el análisis de información correspondiente.

Como Analista de Negocios durante 6 años en una importante empresa comercializadora de berries, tuve la oportunidad de conversar directamente con decenas de productores, visitar sus campos y entender los retos que afrontan. Varios casos vienen a la mente, como el de un productor que decidió eliminar una variedad de su programa de fresa por tener un bajo desempeño, sin percatarse que era ésta la que le generaba el flujo de efectivo necesario para mantener su operación al inicio de temporada; o también aquellos productores que ceden sus decisiones a la comercializadora sobre qué y cuándo plantar al no estar inmersos en el análisis de información. También están aquellos nuevos productores que vienen de otras industrias y desconocen el manejo del campo, pero que cuentan con el capital para invertir en este giro y el interés en hacerlo.

En resumen, existe una necesidad en los campos de producción de berries de herramientas que permitan el análisis de la información y facilite la toma de decisiones para desarrollar una adecuada planeación estratégica y que les ayude a prepararse adecuadamente para las innovaciones tecnológicas del futuro.

2. Definición del problema

Los productores de berries de México cuentan con capacidades limitadas para recabar y analizar datos de producción que permitan una adecuada planeación estratégica para el suministro de fruta. Estas limitaciones son diversas (equipo de cómputo, procesos, personal); sin embargo, para fines de este documento, se englobarán estas limitaciones como la falta de elementos que permitan visualizar y analizar el estado actual de las operaciones agrícolas. Esta situación los obliga a tomar decisiones basándose en el conocimiento empírico del campo, lo cual compromete la rentabilidad de sus operaciones, la calidad de sus productos y el rendimiento de sus plantas. Situaciones como el cambio climático, la degradación de suelos, la escasez de mano de obra, la necesidad de incrementar el volumen de alimentos y el surgimiento de nuevas amenazas relacionadas con plagas y enfermedades obligan a los productores de México a buscar el mayor aprovechamiento posible de los recursos naturales, humanos y económicos mediante la implementación de las tecnologías más adecuadas para la producción agrícola.

3. Objetivo general

Demostrar la factibilidad de implementar y adaptar metodologías de Inteligencia de Negocios o BI (Business Intelligence, por sus siglas en inglés) en unidades de producción agrícola de berries en México como sustento para el desarrollo de una adecuada planeación estratégica. Esto conlleva el uso de modelos y plataformas de análisis de datos con el fin de

lograr una reducción en tiempos de trabajo y un incremento en la calidad de su información para una toma de decisiones sustentada en datos.

4. Objetivos específicos

1. Mitigar las capacidades limitadas para el análisis de la producción agrícola de los productores al probar que la incorporación del BI es técnicamente viable.
2. Ejemplificar un proceso de procesamiento y análisis de datos a manera demostrativa para corroborar que es posible interpretar datos de campos.

5. Justificación

Generar las capacidades tecnológicas para recabar y analizar información de la producción proporcionará a los productores de **berries** de México las herramientas necesarias para evaluar las mejores alternativas para desarrollar su planeación estratégica. Migrar de una forma de trabajo empírica basada en observaciones a un esquema de trabajo basado en datos y en capacidades de análisis, permitirá que existan registros históricos que ayuden a evaluar la evolución de la **agrícola** en términos de **rendimiento** y **rentabilidad** y, por ende, den soporte a organizar sus operaciones de manera óptima al planear su mezcla de variedades, fechas de plantación y poda, y manejo del campo, entre otros.

La base para lograr las capacidades antes mencionadas es una correcta administración de los datos, lo cual incluye el proceso de adquisición, limpieza, procesamiento y exploración. La implementación de la propuesta a desarrollar en este documento permitirá no sólo que las unidades de producción agrícola cuenten con las capacidades para una correcta planeación estratégica sino que también, como resultado de la implementación de esta propuesta, se beneficien de una optimización en sus procesos de recabación y análisis de datos. Lo anterior se traduce en una reducción de tiempos, un incremento en la eficiencia del personal y una mejora en la calidad de la información resultante.

Por último, hay que destacar la realidad de las operaciones **agrícolas** de **berries** en México y lo variado que son los esquemas de inversión y la disponibilidad de recursos. Así como existen productores de más de 1,000 hectáreas en producción, también existen pequeños productores de menos de 10 hectáreas. Las capacidades tecnológicas y los recursos de las diversas unidades de producción agrícolas serán variadas. Este documento no pretende proponer una **plataforma** única para recabar y analizar datos. El enfoque está orientado a la adaptabilidad y flexibilidad para que una metodología de trabajo basada en los principios de los análisis de datos pueda ser implementada en cualquier operación agrícola que cuente con equipos de cómputo, sin importar el nivel de sofisticación que se tenga en cuanto a sistemas contables, **ERPs** o sistemas de índole similar. Se demostrará a través de este documento que es posible adaptar estas metodologías, permitiendo que incluso operaciones de menor tamaño se puedan beneficiar de análisis detallados de sus

datos volviéndolas así más competitivas, mejorando sus capacidades de toma de decisiones, optimizando sus recursos y sacando provecho de los beneficios que otorga un buen esquema de análisis de datos.

Capítulo 2

Estado del arte

Datos, datos, datos. No puedo hacer ladrillos sin arcilla

— SHERLOCK HOLMES, personaje ficticio creado por Sir Arthur Conan Doyle,
escritor y médico británico (1859 - 1930)

Las próximas décadas presentan un reto para la alimentación mundial. Acorde a cifras de la FAO, se prevé que la población mundial aumente a 9,100 millones de personas [3]. Plantear la necesidad de prepararse para esta situación llevó a los 193 Estados miembros de las Naciones Unidas a adoptar la Agenda 2030 para el Desarrollo Sostenible, el cual incluye 17 Objetivos de Desarrollo Sostenible, conocidos como ODS, y 169 metas asociadas a estos [4]. Esta Agenda pretende lograr un desarrollo sostenible en el cual tanto la alimentación como la [agricultura](#), los medios para la subsistencia de las personas y la gestión de los recursos naturales sean tratados como una cuestión única y no como situaciones aisladas [5].

Aunque los 17 objetivos son igualmente importantes, el ODS 9 destaca, no sólo por su relevancia intrínseca, sino por el papel fundamental que desempeña para dar soporte a los demás objetivos. La industria, innovación e infraestructura son fundamentales para lograr un desarrollo sostenible, empoderar a la sociedad, fomentar una mayor estabilidad social y conseguir poblaciones más resistentes al [cambio climático](#) [7]. Para poder lograr incrementar el volumen de la producción agrícola mundial en un 70 %, porcentaje necesario para cubrir la demanda prevista [3], se requiere un incremento en la [superficie](#) cultivada de 70 millones de hectáreas, así como una mejora en el [rendimiento](#) de los cultivos [8].

Las Tecnologías de la Información Agrícola, o AIT por sus siglas en inglés, han logrado un notorio desarrollo en los últimos 20 años. y han demostrado ser una herramienta directamente ligada a la mejora de [productividad](#) y un mecanismo indirecto en facilitar la toma de decisiones para los productores [9]. La importancia de los datos y su administración se reconoce ya como un nuevo recurso estratégico de gran valor potencial [10], que ha sido tema fundamental en foros como el “21st Century Smart Agriculture” celebrado en el 2015 en Estados Unidos [11]. Conceptos como Big Data, Internet of Things y [Machine Learning](#) comienzan a formar parte del léxico agrícola. Uno de los conceptos trascendentales es el denominado [Business Intelligence](#) (BI), el cual apoya a las empresas [agrícolas](#) al fortalecer



Figura 2.1: Objetivos de Desarrollo Sostenible acordados en el 2015. [6]

su potencial productivo y su eficiencia técnica gracias al efectivo soporte que provee en temas de administración, análisis de información, planeación y toma de decisiones [12]. Es necesario implementar un BI en una organización cuando se busque un crecimiento, se identifiquen deficiencias en los procesos para lograrlo, se tenga problemas con la concentración de información, o se tenga incertidumbre en la toma de decisiones [13]. Estos aspectos se pueden encontrar dentro de los retos de la administración de los ranchos comerciales de producción de *berries* en México, uno de los productos más importantes para la economía en México y enfoque principal de este documento.

Las *berries* son también llamadas frutas finas, frutillas o frutos del bosque y están caracterizados por su tamaño pequeño y colores brillantes; esta categorización incluye a la fresa, la frambuesa, la zarzamora y la mora azul, en ocasiones también llamado arándano, pero que no debe ser confundido con el arándano rojo [14]. Son productos de gran relevancia económica para México, ya que ocupan el tercer lugar dentro de los productos agrícolas más exportados por el país [2]. Se calcula que el valor de exportación de estos productos en el 2018 fue de \$2,282 millones de dólares, tan solo detrás de la cerveza y el aguacate [1]. Se estima que la *superficie* en producción para el mismo año rondaba las 44 mil hectáreas, con una producción total de 800 mil toneladas para mercados nacionales e internacionales [2]. Su producción se encuentra mayormente localizada en cinco estados de la República Mexicana, siendo estos Michoacán, Jalisco, Baja California, Estado de

México y Puebla [14].

A pesar de su importancia en la economía nacional, las *berries* no se encuentran exentas de los retos a los que se refiere la FAO. La frambuesa reportó en 2018 una caída del 3.2% en su *rendimiento*, mientras que el resto de los cultivos mostró un incremento por debajo del 1.5% [1]. Estos valores provocan que el incremento en volumen provenga de un aumento en la *superficie* productiva, lo cual no es sostenible debido a que la *superficie* destinada para actividades agrícolas es limitada [8]. Algunos de los retos que enfrentan los productores mexicanos son la alta demanda de mano de obra, la responsabilidad social, la migración interna, la alta dependencia con Estados Unidos y la sostenibilidad, así como la necesidad de programas de mejora genética de variedades para que produzcan con un uso menor de recursos hídricos y sean más resistentes a *plagas* [15]. Estas situaciones en conjunto incitaron a que la Secretaría de Agricultura y Desarrollo Rural (SAGARPA) emitiera en el 2017 un documento titulado Planeación Agrícola Nacional 2017-2030: Frutas del Bosque Mexicanas, el cual marca como objetivo estratégico el consolidar a México como potencia exportadora de frutos del bosque a través de un esquema óptimo en materia fitosanitaria y el impulso logístico para llegar a nuevos mercados en Estados miembros de la Unión Europea y Asia [16]

ESTRATEGIA	PAÍS(ES)	¿CÓMO?
CONSOLIDAR	Estados Unidos y Canadá	Mantener comercio libre de arancel, ofreciendo producto de alta calidad y desarrollar esquemas de protección de propiedad intelectual que permitan diferenciar los mercados y acceder a los segmentos de mayor precio.
	Japón y Estados miembros de la Unión Europea	Expandir el flujo de exportaciones a estos países, así como aprovechar el proceso de Modernización del Pilar Comercial del Acuerdo Global entre México y la Unión Europea (Acuerdo de Asociación Económica, Concertación Política y Cooperación entre la Comunidad Europea y los Estados Unidos Mexicanos [TLCUEM]). Será similar el caso con Reino Unido en una negociación de un TLC como resultado de su salida de la Unión Europea.
EXPANDIR	Suiza	Se sugiere a los productores impulsar la creación de una solicitud de reducción o eliminación arancelaria para las subpartidas 0810.20 y 0811.20, como parte del protocolo correspondiente al TLC entre México y los países de la Asociación Europea de Libre Comercio.
DESARROLLAR	Australia	Se recomienda la solicitud de peticiones de reducciones arancelarias para la exportación de frutos del bosque, ya sea a través de la negociación del denominado TPP 11 (Tratado de Asociación Transpacífico sin Estados Unidos), de su incorporación a la Alianza del Pacífico como Estados asociados o con base en un tratado de libre comercio bilateral.

Figura 2.2: Estrategias de mercado para lograr objetivos estratégicos de las *berries* en México. [16]

Lo anterior describe la importancia que tiene para México el buscar soluciones que den soporte a una mejor toma de decisiones en busca de un mejor *rendimiento* a través de la identificación de las variedades más adecuadas, fechas de plantación y poda, regiones geográficas, marcos de plantación y manejos agronómicos. Existen algunas empresas en el mercado internacional que ya se encuentran trabajando en uno o más de estos aspectos. Tres *plataformas* que destacan por su enfoque agrícola son SXagro (India), Prospera (Israel) y rNET (México). Estas *plataformas*, aunque no son las únicas en el mercado, ofrecen ejemplos de la oportunidad de negocio que representa la *agricultura* para las áreas de tecnologías de la información, ya sea para recabar datos de campo o para analizarlos.

A continuación se describe brevemente el enfoque de cada uno de estos sistemas:

- SXagro. **Plataforma** para el campo que se ofrece como servicio (SaaS) y que proporciona datos y análisis en tiempo real mediante el uso de datos obtenidos de manera satelital y de campo [17].
- Prospera. **Plataforma** especializada en el análisis de información cuya misión es transformar a la producción agrícola en un recurso predecible y optimizado. El producto que ofrece permite recolectar, digitalizar y analizar grandes volúmenes de datos para que los productores puedan controlar y optimizar su producción, haciendo uso de **Machine Learning**, **Redes Neuronales** y **Deep Learning**. Algunos de los temas que abordan son calidad, planeación de la producción, seguimiento a **rendimientos**, y **plagas** y enfermedades [18].
- rNet. Empresa mexicana que desarrolla sistemas informáticos a la medida. Cuenta con distintas aplicaciones para sectores específicos, incluyendo el agrícola. Se basa en el desarrollo de módulos a través de **Odoo**, el cual es una **suite** de aplicaciones de gestión empresarial de código abierto [19].

Es importante señalar que la producción de **berries** es un rubro aparte dentro de la **agricultura** en México con condiciones muy particulares como lo son: el uso de **macrotúnel**, su temporalidad, y las características fenológicas. Por ello se requiere de un enfoque muy específico, diferente de aquel usado en un cultivo extensivo o en la producción en invernaderos de alta tecnología.

Por último, aunque no fue posible encontrar algún estudio específico sobre la situación actual del BI en la **agricultura** mexicana, Tyrychtr publica los resultados de un censo realizado a 500 **agricultores** en la República Checa, de las cuales 135 completaron la encuesta correctamente donde se destaca que solo el 1% de las **agriculturas** cuenta con un sistema de BI, contrastando con el 40-50% que dicho país tiene a nivel nacional incluyendo todas las industrias [12]. Si suponemos una situación similar en México, donde la mayor parte de las empresas de **berries** son familiares, se puede asumir el área de oportunidad y el impacto que se puede tener para mejorar la toma de decisiones en un producto que es fundamental para la economía mexicana y, al mismo tiempo, parte de la solución para ayudar a lograr la producción agrícola necesaria para alimentar a la población mundial en el futuro próximo.

Capítulo 3

Marco teórico

La agricultura, para un hombre honorable y de alto espíritu, es la mejor de todas las ocupaciones y artes por medio de las cuales un hombre puede procurarse el sustento.

— JENOFONTE, Historiador, militar y filósofo griego (431 a. C. - 354 a. C.)

1. Producción de berries en México

La importancia en el mercado de **berries** ha venido en aumento y ha cobrado gran peso en el mercado agroalimentario mediante la oportunidad que representan los consumidores de ingresos medios y altos dispuestos a pagar un precio considerable por productos que se adecúen a los nuevos estilos de vida, por ejemplo familias pequeñas y con poco tiempo para preparar alimentos [14]. Dentro de las principales razones que explican el interés actual en dichos frutos son su alta **rentabilidad**, el rápido retorno sobre la inversión, el uso intensivo de mano de obra, la versatilidad en su consumo y la capacidad de exportación [20]. En el 2019, al volverse el tercer producto de exportación agrícola más importante para México [2], resulta fundamental contar con los elementos técnicos y tecnológicos que ayuden a mejorar la calidad, confiabilidad, **rendimiento** y **rentabilidad** de estos productos.

Aunque no es la finalidad de este documento el proporcionar demasiado detalle sobre los aspectos agronómicos de estos cultivos (manejo nutricional, fisiología de las plantas, manejo cosecha y post-cosecha, entre otros), sí es necesario describir algunos de los términos y condiciones clave para poder entender la magnitud de los retos y oportunidades que presenta el negocio de la producción de **berries** en México. El desconocer estos aspectos puede llevar al falso supuesto de que este tipo de industria puede adoptar medidas tecnológicas idénticas a las ya existentes en otros tipos de productos agrícolas como el maíz, el agave o el aguacate, o a otras industrias de tipo industrial o manufacturera.

La palabra **berry** engloba a una gran gama de productos, como las **aronías**, las **grosellas**, el **saúco**, o el **saskatun** [21], sin embargo, en México los términos como **berry**, **baya**, **frutos rojos** o **frutillas** son usados para describir a cuatro productos agrícolas: la fresa, la

frambuesa, la zarzamora y el arándano o mora azul. Estos y otros tipos de berry se pueden apreciar en la Figura 3.16. Además, es importante recalcar que el mercado de *berries* se refiere a la producción y comercialización de lo que se conoce como producto en fresco; es decir, que no se encuentra procesado ni transformado, por lo que su consumo es directo. Mercados alternos al fresco se les conoce como canería, *fruta* de proceso, o simplemente proceso, conceptos que tienen distinto significado en las diversas regiones de México, pero que al final se refieren al uso de las *berries* transformadas en otros productos como lo son la mermelada, el yogurth o los jugos, entre otros.

Además, hay dos aspectos que son fundamentales en la producción de *berries*: el desarrollo varietal y los sistemas de producción.

Respecto al desarrollo varietal, es importante resaltar que las distintas empresas e incluso instituciones educativas inmersas en la producción de *berries* buscan generar ventajas competitivas mediante la obtención de un fruto que sea de mayor calidad (sabor, apariencia, condición), de mayor facilidad de producción, resistente a plagas y enfermedades, con mayor *vida de anaquel* y que generen un mayor margen de ganancia al productor. Es por esto que diversos entes mantienen un programa genético de mejora constante para la producción de nuevas y mejores variedades, ya sea a través de programas públicos o privados. Cada una de estas variedades, además de diferenciarse en aspectos como el tamaño, color y sabor del fruto, tienen requerimientos específicos para su producción, como lo son sus esquemas nutricionales, la altura sobre el nivel del mar, el clima, el tipo de suelo, la luminosidad, entre otras.



Figura 3.1: Algunos tipos de berries. Obtenida de https://www.freepik.com/premium-vector/fresh-berries-list-with-names_4427767.htm



Figura 3.2: Riego por goteo. Foto de <https://agrotendencia.tv/agropedia/el-cultivo-de-la-fresa-o-frutilla/>

Debido a lo anterior, los sistemas de producción deben estar también en constante actualización para dar respuesta a las necesidades de las nuevas variedades que estarán presentes en el mercado. A causa de las características climáticas de México, la producción de *berries* ha obligado a buscar alternativas que ayuden a ofrecer un producto de calidad y que, al mismo tiempo, sea rentable. Esto se ha logrado con la implementación de la *agricultura de precisión*, el cual es un sistema empleado para analizar y controlar la variación espacio-temporal del terreno y el cultivo, manejando las variables y administrando eficientemente los insumos, como agua o fertilizante, para minimizar los recursos invertidos y los riesgos agroalimentarios, al mismo tiempo que se maximiza la producción [22].

Algunas de estas variables incluyen temperatura, humedad relativa, nutrientes, entre otros. Un ejemplo de esto son los sistemas de riego por goteo, los cuales permiten conducir el agua mediante una red de tuberías y aplicarla a los cultivos a través de emisores que entregan pequeños volúmenes de agua periódicamente en forma de gotas y por medio de goteros [23]. Otra tecnología utilizada es el uso de *macrotúnel*, los cuales protegen al cultivo de las condiciones adversas del medioambiente, reduciendo daños por heladas, el porcentaje de frutas deformes, el daño por granizo y el provocado por la incidencia de lluvias [24]. No se debe confundir el uso del *macrotúnel* con los invernaderos tradicionales, ni con los invernaderos de alta tecnología, los cuales son estructuras para la producción más avanzadas y que en México son usadas principalmente en la producción de pepino, jitomate y pimientos.



Figura 3.3: Fresa en macrotúnel. Foto de <https://www.tunnel-tek.com/es/macrotuneles>

Todas estas tecnologías y sistemas de producción han sido cruciales para el desarrollo de la producción de berries en México. Sin embargo, como se mencionó anteriormente, esta categoría incluye múltiples cultivos que, aunque existen en la naturaleza, no todos forman parte de los programas comerciales y de exportación. Es importante destacar que, hasta este momento, son cuatro los productos considerados como berries desde el punto de vista de la comercialización: la fresa, frambuesa, zarzamora, y arándano.

1.1. Fresa

Este fruto proviene de la planta de la fresa, la cual es de tipo **herbáceo** con sistema radicular fasciculado y compuesto de raíces y raicillas, una altura menor a los 50 cm, numerosas hojas originadas en una corona o rizoma muy corto a nivel de suelo, con flor blanca, cáliz de 5 piezas hendidas, 5 pétalos redondeados y numerosos estambres y pistilos [25]. Las alternativas para su plantación son el uso de planta a raíz desnuda o de un **plug** de planta de fresa. El fruto es rojo, tiene sabor dulce y presenta un aroma característico, es recolectado manualmente y debe cosecharse cada tercer día [25]. En el 2018, México produjo 654 mil toneladas de fresa, ubicándose como el tercer país más importante para la producción con 14 mil hectáreas en producción, un **rendimiento** promedio de 47.9 toneladas por hectárea y recaudando \$700 millones de dólares en exportaciones [1].

1.2. Frambuesa

El fruto proviene de un arbusto de tallo subterráneo corto que genera ramas aéreas de dos años de duración [16]. La **fruta** es convexa y rugosa, con forma similar a la piña y de color rojo o amarillento [16]. Típicamente es plantada mediante el uso de raíz, aunque también es posible utilizar **plugs** de frambuesa. En el 2018, se logró en México una producción de 130 mil toneladas, consiguiendo así el segundo lugar a nivel mundial, con alrededor de 7,100 hectáreas productivas, un **rendimiento** promedio de 18.2 ton/ha y un valor de \$750 millones de dólares en exportaciones [1].

1.3. Zarzamora

La zarzamora proviene de una planta arbustiva con tallos de 3 a 4 metros de largo de crecimiento erecto al principio, pero que se van tumbando con el tiempo; son angulosos y pueden tener espinas [16]. El fruto es carnoso, formado por numerosos frutitos esféricos apiñados cada uno con un huesillo de color rojizo al principio y negro en la maduración, con sabor dulce y aromático [16]. El **plug** de zarzamora es lo más usado para su plantación. Con una producción de 287 mil toneladas en el 2018, México se posiciona como el principal productor mundial de zarzamora, con más de 13 mil hectáreas en producción, un **rendimiento** de 1.3 ton/ha y \$405 millones de dólares en exportaciones [1].

1.4. Arándano

Proveniente de un arbusto caducifolio de hasta 60 cm de altura, el arándano o mora azul es un fruto de forma esférica que típicamente mide de 1 a 2 cm de diámetro, aunque recientes variedades logran un tamaño superior, de color azul intenso a suave y recubierto por una serosidad característica importante que aporta a su *vida de anaquel* [16]. Se cuenta con *plugs* de distintos tamaños para su plantación. Es un cultivo más nuevo para México que el resto, por lo que con sus 40 mil toneladas se logró una cuarta posición mundial en el 2018, con alrededor de 3,600 hectáreas en producción, un *rendimiento* de 11.1 ton/ha y \$306 millones de dólares en exportaciones [1].

2. Business Intelligence

En un Congreso de productores de *berries* celebrado en el 2019 en Cancún, Quintana Roo, Mary Shelman, Ex-Directora del Programa de Agronegocios de Harvard y Fundadora del Mary Shelman Group, remarcaba que toda empresa en la actualidad, sin importar su giro, es una empresa de tecnología. Las organizaciones del siglo XXI dependen de sus capacidades de recolección de datos y de análisis de información para poder tomar decisiones fundamentadas. La Inteligencia de Negocios, *Business Intelligence* o BI, es una de las múltiples metodologías vigentes que las organizaciones pueden implementar como soporte para su planeación estratégica.

El BI es considerado un Sistema de Información (SI). Silva [26] compila diversas definiciones formales de lo que es un SI, las cuáles ayudan a describir al SI como un conjunto de elementos que interactúan entre sí mediante el almacenamiento, procesamiento y distribución de información, con el fin de dar soporte a las actividades de la empresa, apoyando la toma de decisiones, y desarrollando estrategias para lograr los objetivos organizacionales.

Curto y Conesa definen al BI como el conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización [27]. Es un concepto amplio que abarca tecnologías tanto internas como externas. La empresa de investigación de mercado, Forrester Research, publicó un diagrama que resume las distintas tecnologías que participan en un BI (ver Figura 3.4).

Como se puede apreciar en la Figura 3.4, la cantidad de conceptos relacionados con el uso de BI en las organizaciones es amplia. No todos estos estarán presentes en todas las implementaciones, y distintos autores presentan diversas arquitecturas. Salgueiro, Carrión y González destacan que la arquitectura de un BI incluye 4 elementos fundamentales: el almacenamiento de datos (*data warehouse* y *data marts*), las funciones de analítica empresarial (*business analytics*), los sistemas de gestión de *rendimiento* y las interfaces de usuario [29]. Rad presenta una arquitectura alterna más compacta e igualmente válida [30]. Para comprender más a detalle cómo se conforma un sistema de BI se usará como referencia la arquitectura provista por Rad, agregando solamente el concepto de *Data Mart* que no se encuentra incluido en ésta. De esta manera, el resto del capítulo se dividirá en cuatro

- **Preparación de datos.** Consiste en la limpieza y procesamiento de los datos para lograr un nivel de estandarización y homogeneidad necesario para que el BI opere de manera congruente.
- **Almacenamiento para análisis.** Dependiendo de la finalidad de la organización, podrá ser necesario o no contar con Data Marts. Algo importante de recalcar es que la existencia de un Data Warehouse orientado al análisis es crítico en un BI. Mientras que las bases de datos transaccionales priorizan la velocidad de carga de datos, las bases de datos analíticas se enfocan más en la lectura y disponibilidad de los datos.
- **Análisis y visualización.** Existe un amplio catálogo de herramientas gratuitas y de paga en el mercado que permiten analizar los datos resultantes luego de un procesamiento y almacenamiento. El propósito de estas herramientas es facilitar al usuario final la interpretación de los mismos, facilitando la toma de decisiones.

Es fundamental recordar que el propósito de este proyecto está orientado a demostrar la que es posible introducir conceptos de BI en el ramo agrícola, reconociendo que existen operaciones agrícolas de diversos tamaños y capacidades económicas. Por esta razón, esta arquitectura muestra varias tecnologías a manera de ejemplo, y no se pretende en ningún momento recomendar alguna por encima de otra, ni adentrarse en temas de costos de licenciamiento o implementación.

Bajo esta lógica, de aquí en adelante se usarán tecnologías variadas con la intención de hacer una demostración más conceptual que una verdadera propuesta de implementación. Algunas aplicaciones que se usarán en este documento son:

- **Python.** Es un lenguaje de programación de alto nivel que ha sido muy bien recibido para trabajar cuestiones relacionadas con la obtención, limpieza, procesamiento, exploración, y análisis de datos. Hay 2 razones principales para usar este software dentro de este proyecto: la practicidad que provee para poder demostrar paso a paso la evolución en un proceso de limpieza de datos, y el hecho de que forma parte de la oferta académica de la maestría a través de la materia de Programación para Análisis de Datos.
- **Pandas.** Es una librería desarrollada para el análisis y manipulación de datos, y que permite integrarse con Python. Se seleccionó esta herramienta bajo la misma lógica del uso de Python.
- **Tableau Desktop.** Es un software enfocado en el análisis y visualización de datos. Aunque Tableau es de costo, a diferencia de Python y Pandas, su versatilidad y simplicidad para ser usado por usuarios con un bajo conocimiento de programación lo hace atractivo para demostrar los puntos establecidos en los objetivos de este proyecto. Existen otras aplicaciones en el mercado como Power BI, Looker, y QlikView, y cualquiera de estos es una alternativa técnicamente viable; sin embargo, se tomó la

decisión de tomar esta vía debido a la experiencia previa del autor en este software y la disponibilidad de la licencia.

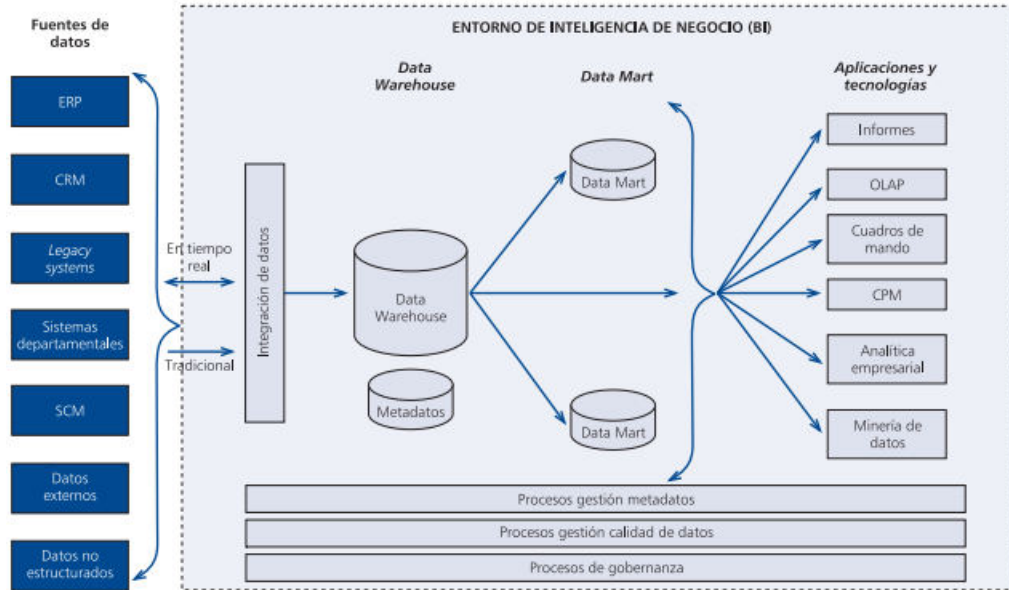


Figura 3.5: Arquitectura de un BI de Salgueiro. [29]

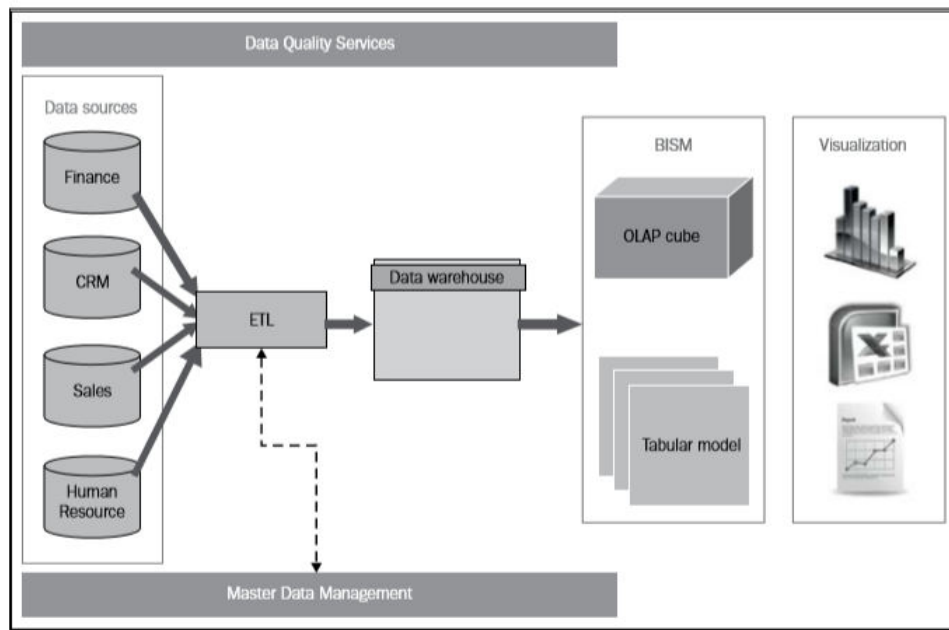


Figura 3.6: Arquitectura de un BI de Rad. [30]

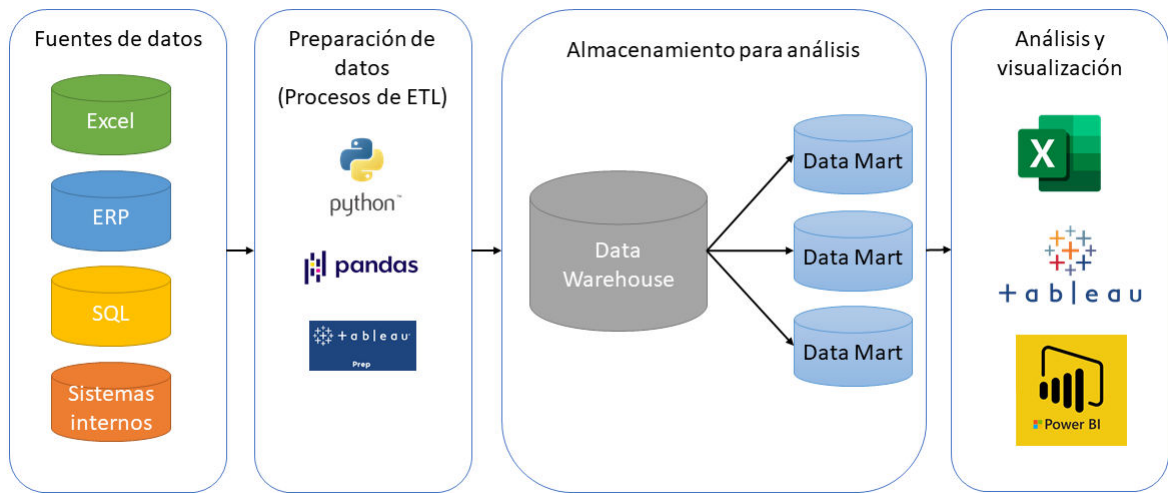


Figura 3.7: Arquitectura propuesta.

2.1. Fuentes de datos

Un sistema de BI requiere de un paso preliminar que se encuentra fuera de su arquitectura, y es la forma de obtención de datos. Las fuentes son diversas y tienen distintos grados de formalidad, pueden incluir el uso de sistemas de **ERP**, **CRM**, **SCM**, servidores externos, sistemas externos, **RFID** e incluso datos estructurados y no estructurados [29] (aunque no es el propósito del proyecto ahondar en estos conceptos, una breve definición de los mismos se incluye dentro del glosario al final del documento). Esto genera una complejidad para el BI al momento de cargar los datos al data warehouse, ya que es necesario contemplar que cada fuente tendrá estructuras de datos distintas, nomenclaturas, tamaños, lenguajes, entre otros [31].

Para ejemplificar parte de esta problemática, suponga que una organización cuenta con 3 sistemas: **ERP**, **CRM**, **SCM**. Todos hacen uso del producto denominado “fresa” para sus operaciones; sin embargo, el **ERP** lo nombra como STRAWBERRY, el **CRM** trabaja en español y lo denomina FRESA, mientras que el **SCM** usa nombres cortos y se refiere al mismo producto como STRAW. El BI requiere que, durante la obtención de datos, se estandarice esta nomenclatura para operar adecuadamente; para ello, hace uso de herramientas llamadas **ETL** que se verán más adelante en este documento. Cano menciona que el número de fuentes de información es distinto entre organizaciones, teniendo una media de 8 bases de datos en grandes corporaciones, aunque llega a haber hasta 50 [31].

A continuación se enlistan algunos de los problemas más comunes al momento de trabajar en la obtención de datos:

- Fuentes de datos: las fuentes son de distinto origen y usan distintas tecnologías con condiciones específicas para poder extraer los datos.
- Nomenclatura: es común encontrar que no existen procesos y acuerdos para cómo se nombran y/o catalogan los datos de una organización-

- Tipos de dato: es posible encontrar que un mismo dato se almacene diferente en las distintas fuentes. Por ejemplo, un número puede registrarse como tipo cadena en una fuente y tipo numérico en otra.

2.2. Almacenamiento de datos

Una vez que las fuentes de datos han sido identificadas, es necesario almacenar estos datos para su posterior análisis. Para esto, existen 3 conceptos fundamentales:

2.2.1. Data Warehouse

Es una base de datos corporativa en la que se integra información previamente depurada desde distintas fuentes; es decir, la información ya viene homogénea y es fiable [32]. Está construida con el propósito de atender las necesidades de análisis de datos y **reporteo**, por lo que no es recomendable usar bases de datos normalizadas como las requeridas en sistemas transaccionales tradicionales [30]. En su lugar, se recomienda el uso de modelos en estrella (star schema) o en copo de nieve (snowflake schema) [32], los cuales reducen el tiempo de respuesta e incrementan el desempeño de las consultas para reportes y análisis [30].

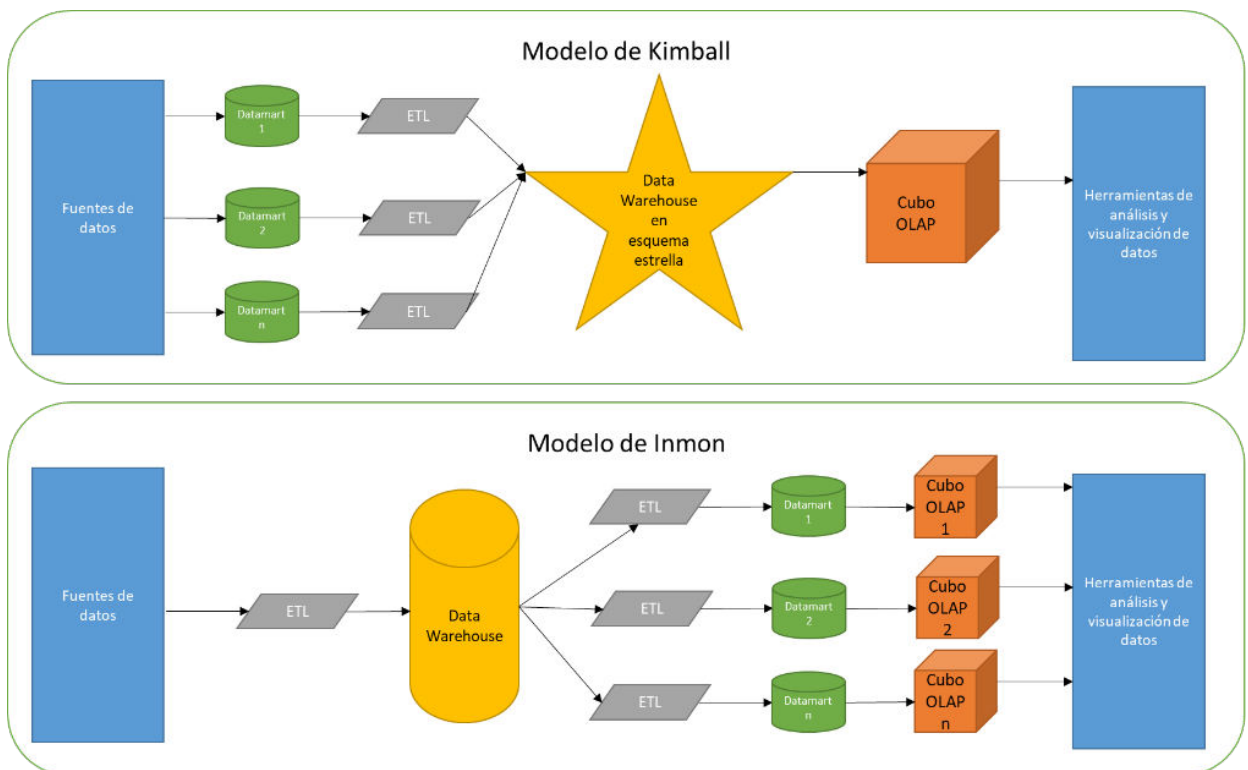


Figura 3.8: Comparación de la estructura del modelo de Kimball y el modelo de Inmon. Diseño propio.

Al diseñar un data warehouse existen dos modelos: el de Ralph Kimball y el de Bill Inmon. Ambos manejan los mismos términos, pero con una descripción diferente de lo que es un data warehouse y un data mart. Las ventajas que presenta Kimball como son el bajo costo, la rapidez de implementación y un menor requerimiento de personal especializado [30], lo hacen una mejor alternativa para un proyecto demostrativo como éste.

Parámetro	Inmon	Kimball
Construcción del Data Warehouse	Alto consumo de tiempo	Bajo consumo de tiempo
Mantenimiento	Simple	Complejo, redundante, y sujeto a revisiones
Costo	Altos costos iniciales, con fases subsecuentes más económicas	Bajos costos iniciales, con costos bajos también en fases subsecuentes
Tiempo	Altos tiempos de implementación	Bajos tiempos de implementación
Nivel de habilidad	Requiere equipo de especialistas	De más fácil acceso para incursionadores
Requerimientos de integración de datos	Nivel corporativo	Nivel departamental

Figura 3.9: Comparativa entre Kimball e Inmon.

Respecto a los modelos de datos, tanto de estrella como el de copo de nieve son válidos. Se usará el primero debido a que facilita el entender la construcción de un sistema de BI. Esta decisión se fundamenta en la Figura 3.10. El modelo en estrella presenta una mayor simplicidad y, aunque que el mantenimiento es una desventaja, no se prioriza por encima de las ventajas como lo son la existencia de tablas únicas para cada dimensión y la practicidad de consultas más simples, menos uniones, y su menor dependencia de las formas normales.

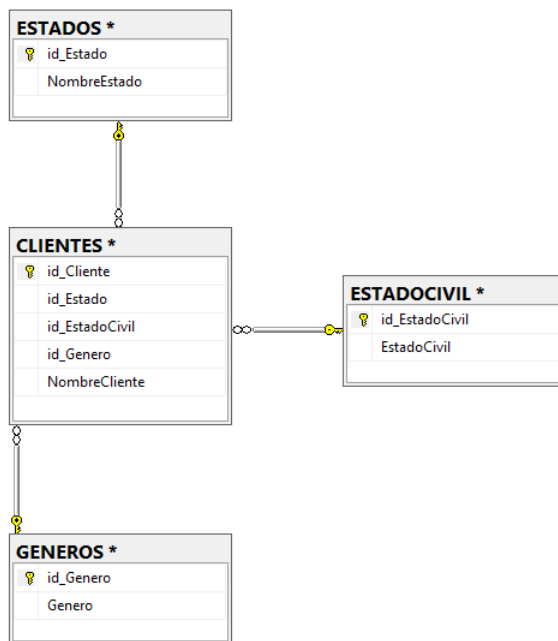
Parámetro	Modelo de copo de nieve	Modelo de estrella
Normalización	3ra forma normal	Denormalización de la 2da forma normal
Uniones	Alto número de uniones (joins)	Bajo número de uniones (joins)
Facilidad de uso	Consultas complejas y difíciles de entender	Consultas simples y fáciles de entender
Desempeño	Altos tiempos de ejecución por la cantidad de llaves foráneas	Bajos tiempos de ejecución por la cantidad de llaves foráneas (baja o nula)
Mantenimiento	Fácil de dar mantenimiento por la ausencia de redundancia	Difícil de dar mantenimiento por la presencia de redundancia
Dimensiones	Múltiples tablas para una sola dimensión	Tablas únicas para cada dimensión

Figura 3.10: Comparativa entre modelo estrella vs modelo copo de nieve.

Una vez definido que se usará el modelo de Kimball basado en estrella, es necesario diseñar el data warehouse, el cual es una base de datos dimensional que está compuesta de dos tipos de tablas: *fact tables* (tablas de hechos) y *dimension tables* (dimensiones) [30].

1. Tablas de dimensiones. Son las que almacenan una serie de atributos que podemos usar para agrupar, dividir o filtrar información [32]. Contrario a lo que ocurre con tablas normalizadas, las tablas de dimensiones suelen contener una gran cantidad de columnas o encabezados. Mientras que en una base de datos relacional se tendrían distintas tablas para llevar un catálogo de clientes, una base de datos dimensional da prioridad a la rapidez de consulta y busca agrupar la mayor cantidad de atributos posible en una sola tabla, aun cuando se corre el riesgo de perder la integridad de la información.

Diseño de base de datos relacional



Diseño de base de datos dimensional



Figura 3.11: Comparación de tablas de una base de datos relacional (izq) y una base de datos dimensional (der). Diseño propio.

2. Tablas de hechos. Son tablas que representan un proceso de negocio, como pueden ser las compras o las ventas, y que están compuestas por los métricos o valores numéricos y sus asociaciones a las dimensiones mediante los identificadores [32]. Esta combinación entre tablas de hechos y tablas de dimensiones es las que da origen al concepto de esquema en estrella, ya que cada tabla de hechos está relacionada a una o más tabla de dimensiones.

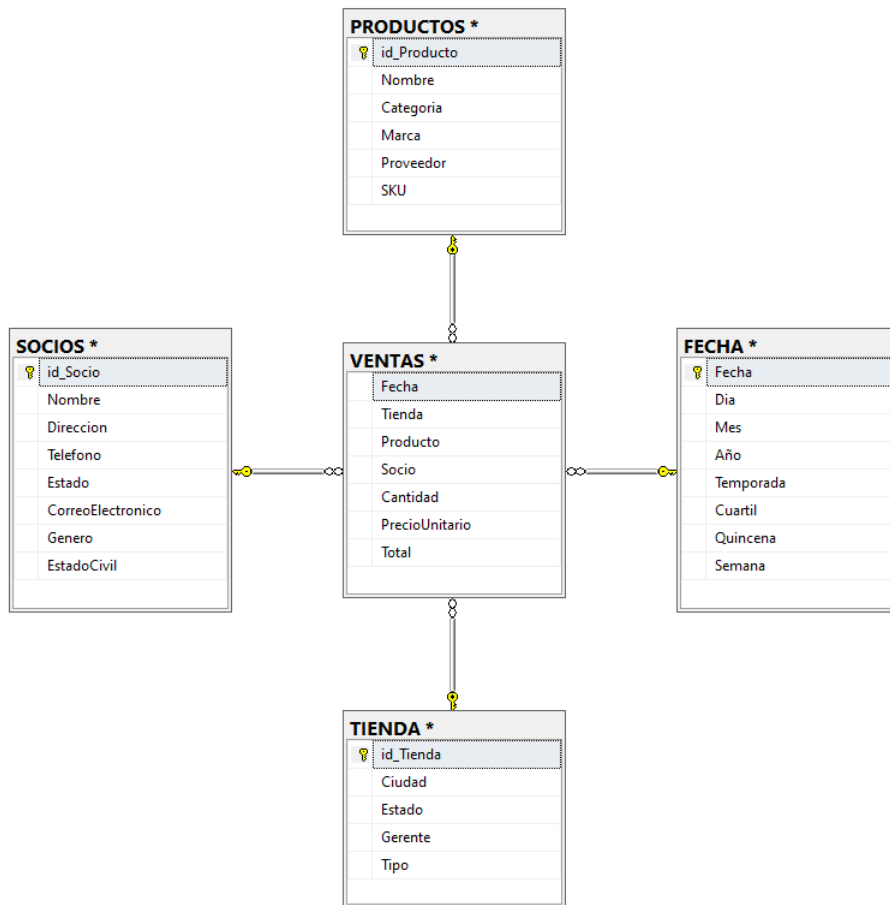


Figura 3.12: Esquema en estrella de la tabla de hechos (Ventas) con cuatro tablas de dimensiones (Socios, Productos, Tienda y Fecha). Diseño propio.

2.2.2. Data Mart

Un data mart es también una base de datos, con la diferencia que, desde la perspectiva de Kimball, está orientada a un proceso de negocio o área específica de la organización, por lo que el conjunto de data marts constituye el data warehouse [32]. En caso de no requerir diversos data marts en proyectos pequeños, o donde no es necesaria una división departamental, es posible mantener solamente el data warehouse.

2.2.3. ETL

Para alimentar el sistema de BI con datos homogéneos es necesario que estos pasen por un proceso de limpieza. Este proceso es conocido como **ETL**. Las siglas de este concepto corresponden a tres actividades, las cuales son extracción (Extract), transformación (Transformation) y carga (Load). La finalidad de usar **ETL** consiste en poder extraer infor-

mación desde distintas fuentes, hacer las transformaciones necesarias para que se adecúen al data warehouse y cargarlas dentro de éste [30]. El primer paso, extracción, requiere del conocimiento y las capacidades para poder conectarse a fuentes que manejen distintas lógicas, como pueden ser archivos de Excel, bases de datos de SQL, XML, Web Services, entre otros.

Una vez extraídos, los datos requerirán de una serie de transformaciones como lo son su limpieza (corrección ortográfica, resolución de conflictos, elementos vacíos, estandarización), la combinación de datos de múltiples fuentes y la eliminación de datos duplicados, lo cual conlleva una posible reingeniería de los procesos de negocio para mejorar la calidad de esos datos [33]. Con esta fase, los datos estarán listos para ser insertados dentro del data warehouse para su posterior manipulación. Existen distintas herramientas en el mercado que permiten diseñar y ejecutar los procesos de ETL, como lo son SQL Server Integration Services de Microsoft (SSIS), DataStage de Oracle o Tableau Prep de la compañía del mismo nombre. Basados en el modelo de Kimball, los datos ahora se encuentran ya en el data warehouse listos para ser consultados y para darles sentido durante su interpretación.

Los procesos de ETL no deben confundirse con la creación de una base de datos. Aunque algunas aplicaciones para implementar ETLs permiten también crear un esquema de base de datos, la finalidad del ETL es simplemente tomar información de una fuente de datos, transformarla, y cargarla en otra fuente, la cual se recomienda ya se encuentre construida.

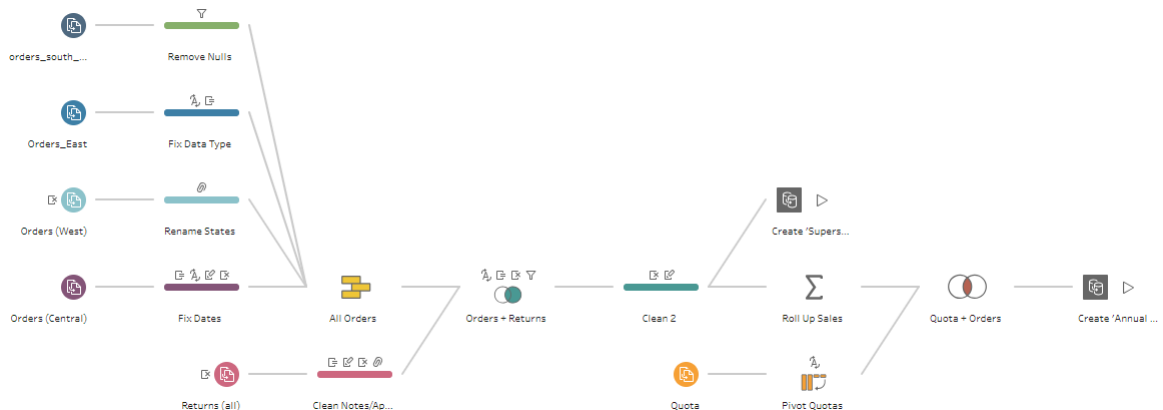


Figura 3.13: Ejemplo de un ETL en Tableau Prep. Impresión de pantalla tomada del ejemplo de Superstore de Tableau.

2.3. Modelo semántico

Dentro de la arquitectura de Rad, el modelo semántico incluye lo relacionado con las capacidades de analítica empresarial — o BA por sus siglas en inglés *Business Analytics* — la cual proporciona los modelos y los procedimientos de análisis al BI e implica el seguimiento de los datos y su posterior análisis para el logro de las ventajas competitivas [29]. Dicho en otras palabras, es aquí donde la lógica de negocios se inserta a los datos para poder interpretarlos y lograr que agreguen valor a los procesos de toma de decisiones.

Hay diversas soluciones comerciales disponibles que nos ayudan en esto, como lo es SQL Server Analysis Services de Microsoft (SSAS, por sus siglas en inglés).

Ahora bien, es importante resaltar que existen aplicaciones que permiten ignorar o saltar esta fase de la arquitectura y conectarse directamente al data warehouse para su consulta. Herramientas como PowerBI, Tableau o QlikView pueden hacer uso de sus capacidades integradas para simular los procesos que ocurren dentro del modelo semántico al conectarse de forma directa al data warehouse, o bien, pueden trabajar de forma integral con herramientas como SSAS. Esto presenta algunas ventajas y desventajas, como se desglosa a continuación:

Ventajas de omitir el modelo semántico:

- Demanda menos conocimiento técnico de bases de datos.
- Reduce los tiempos de implementación.
- Evita incurrir en costos de sistemas como SSAS.

Desventajas de omitir el modelo semántico:

- Incrementa el riesgo de redundancia de datos.
- Complica que se homologuen conceptos y cálculos a través de la organización.
- El mantenimiento se vuelve más complejo y poco administrable.

Regresando al modelo semántico, la arquitectura de Rad muestra dos posibles elementos para analizar la información: el modelo multidimensional, también conocido como cubos de datos, y el modelo tabular. Esta sección dentro del modelo existe porque aun cuando el data warehouse fue construido mediante el modelo en estrella, sigue siendo una base de datos relacional, por lo que no tiene la rapidez de consulta requerida [30]. Tanto el modelo multidimensional como el tabular ayudan a generar modelos basados ya sea en archivos o en memoria, que permitan un acceso más veloz al momento de consultar los datos.

El cubo de datos, modelo multidimensional o cubo OLAP (On-Line Analytical Processing) es un almacenamiento basado en archivos que carga los datos del data warehouse en un modelo tipo cubo, el cual contiene información descriptiva (dimensiones) y celdas (hechos y medidas). Un cubo de datos se puede entender de la siguiente manera:

En la Figura 3.14 se puede identificar el producto que cierto cliente compró en determinada fecha; sin embargo, como las capacidades del cubo están orientadas a poder analizar la información al agrupar, segmentar o filtrar información, también es posible responder preguntas como el total de las compras de cierto cliente sin importar la fecha, las ventas totales en cierto día, identificar el producto más vendido en un año o a los clientes que más compran cierto producto.

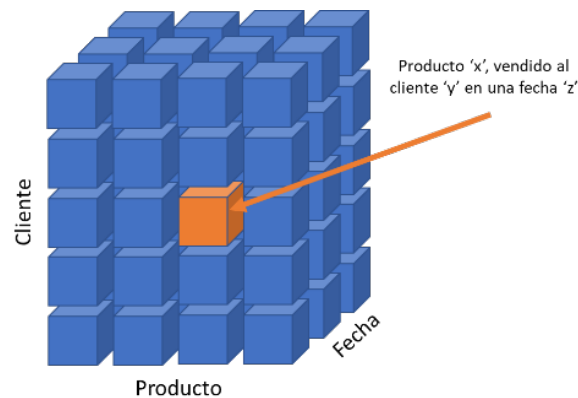


Figura 3.14: Interpretación gráfica de un cubo de datos. Diseño propio.

Una alternativa al uso de cubos de información se da con el modelo tabular, el cual trabaja cargando los datos del data warehouse en memoria para que su consulta sea más rápida [30]. Es un modelo que comparte características tanto de los cubos como de las bases de datos relacionales. Como se puede apreciar en la arquitectura que se usará en este documento, no se hace mención alguna del modelo semántico o de SSAS. Con la finalidad de hacer una demostración completa pero lo más simple posible de un BI, se trabajará bajo un modelo tabular. Esto implica que la conexión entre los el Data Mart y las herramientas de visualización de datos será directa, dando una mayor flexibilidad y una rápida consulta de datos por su disponibilidad en memoria.

2.4. Herramientas de visualización y análisis de datos

Las herramientas para visualizar datos y analizarlos se han popularizado en los últimos años y se tienen distintas alternativas en el mercado. Estas herramientas proveen capacidades relacionadas con la creación de **dashboards** operativos u estratégicos, seguimiento a indicadores clave de desempeño o KPIs, o reportes tanto generales como detallados [30]. Los KPIs, o Indicadores Clave de Desempeño, son valores que demuestran el grado de cumplimiento de uno o más objetivos estratégicos de la organización. Algunos de los KPIs más importantes en el ramo agrícola son el rendimiento, la rentabilidad, el cumplimiento del presupuesto de producción, y los costos de cosecha. Por ejemplo, una empresa puede superar su proyección de ingresos durante la temporada pero quedarse corto en su producción, lo cual se explica por un efecto de un mayor precio de venta. Es para este tipo de enfoque que se desarrollar **dashboards** integrales para analizar el estado de la empresa.

Los **dashboards** proporcionan una visualización de información relevante que es consolidada y organizada en una sola pantalla de manera que pueda ser percibida y entendida rápida y fácilmente [29]. En el 2019, según los Cuadrantes Mágicos de Gartner, los sistemas y proveedores más relevantes en este rubro, fueron Microsoft (no especifica si fue por Power BI, SQL Server Reporting Services o ambos), Tableau y Qlik (igualmente posee diversos servicios como QlikView o QlikSense).



Figura 3.15: Cuadrante Mágico de Gartner para BI y análisis de información. Gartner, 2019.

Todas estas herramientas están enfocadas en el análisis y visualización de datos, inclusive fuera del ámbito de un sistema de BI; es decir, no requieren necesariamente la existencia de un entorno de *Business Intelligence* para operar. Pueden hacerlo mediante conexiones directas a un data warehouse de cualquier tipo, archivos de Excel o Web Services. Como se mencionó anteriormente, herramientas como *Tableau* pueden sustituir el modelo semántico y simularlo bajo sus propias características y lenguajes. Por ejemplo, *Tableau Desktop* tiene la capacidad de conectarse a una base de datos relacional de SQL Server e internamente generar un esquema similar al de los cubos de datos mediante la teoría de conjuntos, la creación de campos calculados y el uso del lenguaje MDX. Sin embargo, como se mencionó anteriormente en las ventajas y desventajas de usar un modelo semántico, la flexibilidad que otorga este tipo de enfoque afecta el mantenimiento de los sistemas, la homogeneidad de la información a través de la organización, y la consistencia de los datos.

Estas tecnologías ofrecen capacidades distintas según sus versiones, pero se destacan

algunas, como son la generación de **dashboards**, las funcionalidades de publicación en línea, las capacidades para trabajo colaborativo, los gráficos dinámicos, y los filtros, por destacar algunos. Dado que la intención de este proyecto es demostrar las capacidades de BI en operaciones agrícolas, y no sugerir ni avalar ninguna tecnología en particular, es recomendable que cada organización realice un estudio de las herramientas disponibles en el momento, tanto gratuitas como de paga, y elija la que más se adecúe a su operación y capacidad. No existe una herramienta de BI que sea perfecta y adaptable para todas las necesidades. Al momento de hacer una elección hay que considerar aspectos como costo, tiempo de implementación, experiencia del personal, flexibilidad, y robustez.

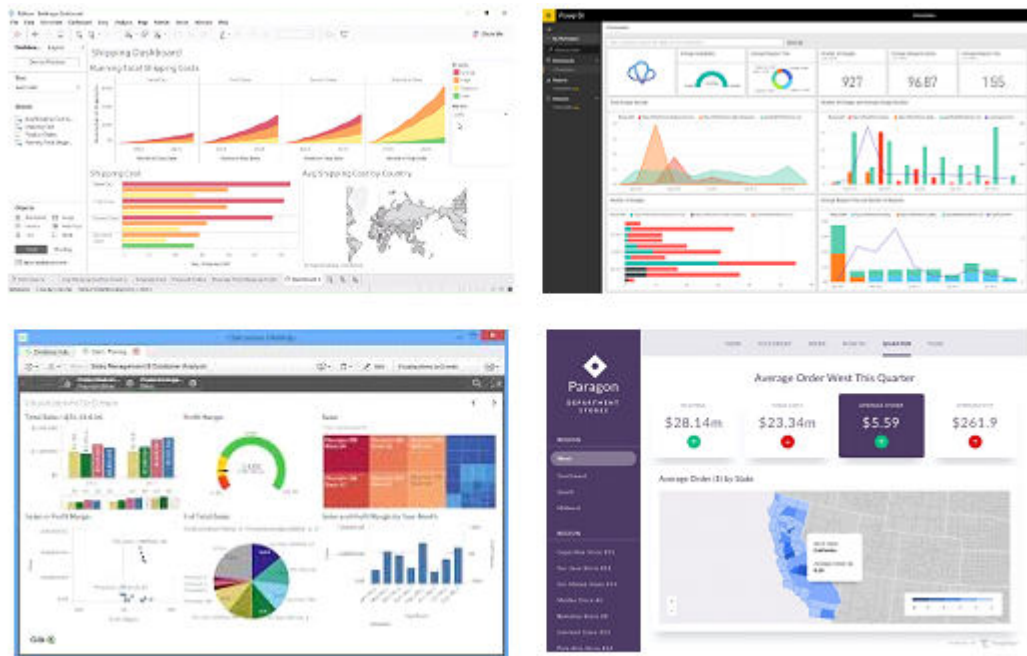


Figura 3.16: Dashboards de plataformas de análisis y visualización de datos; de izquierda a derecha: Tableau, PowerBI, QlikSense y Thoughtspot. Imágenes obtenidas de los sitios de los fabricantes.

Capítulo 4

Desarrollo Metodológico

Duda de los datos hasta que los datos no dejen lugar a dudas
— HENRI POINCARÉ, científico y filósofo francés (1854 – 1912)

Este apartado tiene como propósito describir la metodología a seguir para transformar datos brutos en información de valor para la toma de decisiones. Es importante resaltar que la cantidad y calidad de los datos, así como el grado de tecnificación de los sistemas de información de las unidades agrícolas de producción de berries en México varían de manera importante. Por tanto, el presente documento no pretende estandarizar un mecanismo único para el proceso metodológico, sino demostrar que es factible recabar datos independientemente de las estructuras y fuentes. y así lograr un análisis relevante para la toma de decisiones. Para ello, se utilizarán datos reales de una empresa agrícola de una dimensión de 164 hectáreas que cuenta con operaciones de los cultivos de fresa, frambuesa, zarzamora y arándano.

Con esta metodología se demostrará que es posible hacer una limpieza, procesamiento e interpretación de datos agrícolas que permita mejorar la toma de decisiones basada en datos cuantificables y con potencial de análisis. A su vez, esto permitirá desarrollar las capacidades mínimas necesarias para la creación de modelos de planeación estratégica.

Este capítulo contará con 3 secciones. Las primeras 2 harán uso de Python y Pandas para la preparación de los datos, mientras que en la última se hará uso de Tableau Desktop para ejemplificar el análisis que puede hacerse de estos datos y las preguntas de negocio que pueden ser contestadas.

- En la primera sección, obtención de datos, se ejemplifica el tipo de información agrícola que se puede conseguir, con la intención de mostrar al lector la variabilidad que se puede encontrar en las estructuras de datos.
- En la segunda sección, preparación de datos, se desarrolla un proceso de limpieza y estandarización tanto de la estructura de los datos como de los datos en sí, resaltando algunas de las recomendaciones sobre cómo dar un tratamiento adecuado a éstos para que se encuentren listos para su uso en distintos modelos.

- En la tercera sección, exploración de los datos, se demuestra cómo pueden contestarse algunas de las preguntas de negocio más comunes gracias a dicha estandarización. Para esto, se construyen modelos de análisis y conexión a sistemas de Business Intelligence y se describe cómo conectar los datos limpios a la plataforma de Tableau Desktop para su análisis e interpretación.

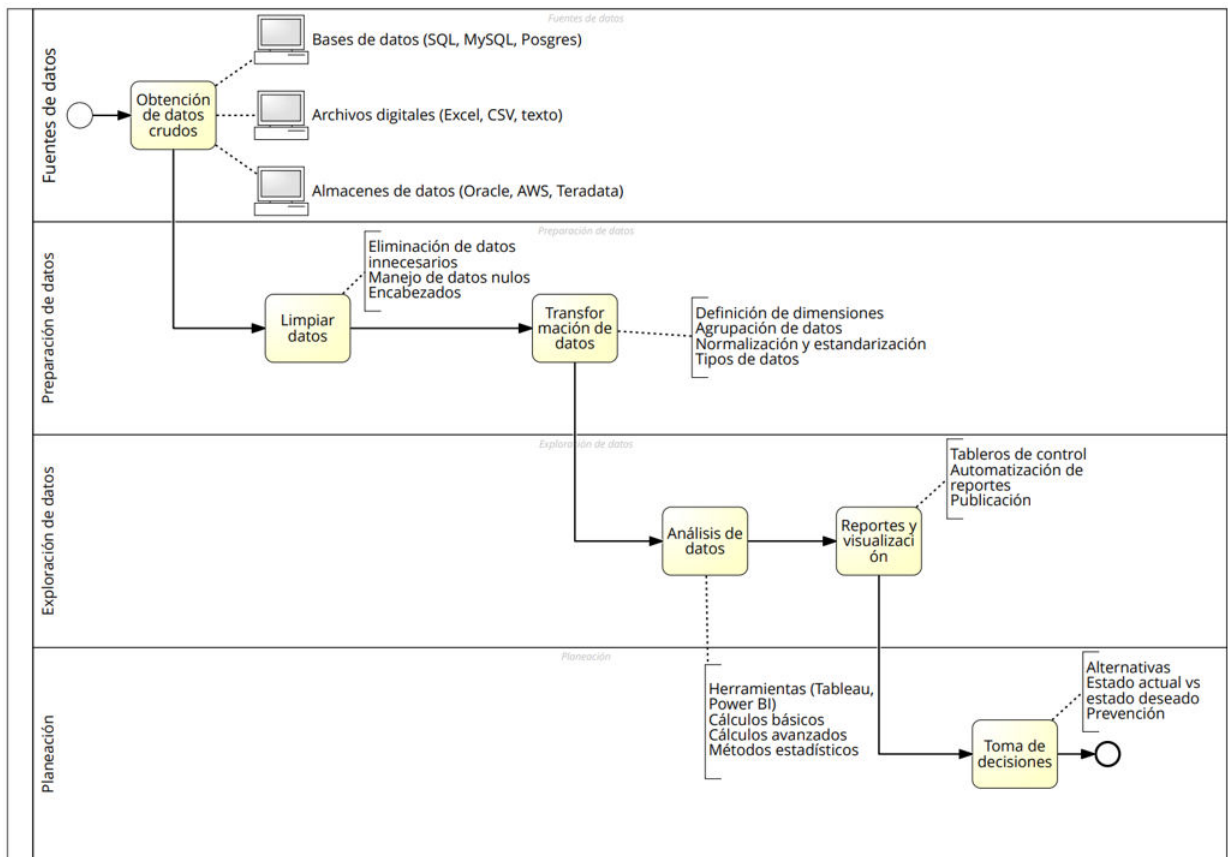


Figura 4.1: Proceso de análisis de datos para la toma de decisiones.

1. Obtención de datos

En esta sección se presenta el proceso de recolección de datos necesario para conformar un modelo de análisis de producción. Este proceso cubre 3 necesidades específicas: la determinación de los métricos y dimensiones, la identificación de las fuentes de datos y la selección del tipo de conexión necesaria para adquirirlas.

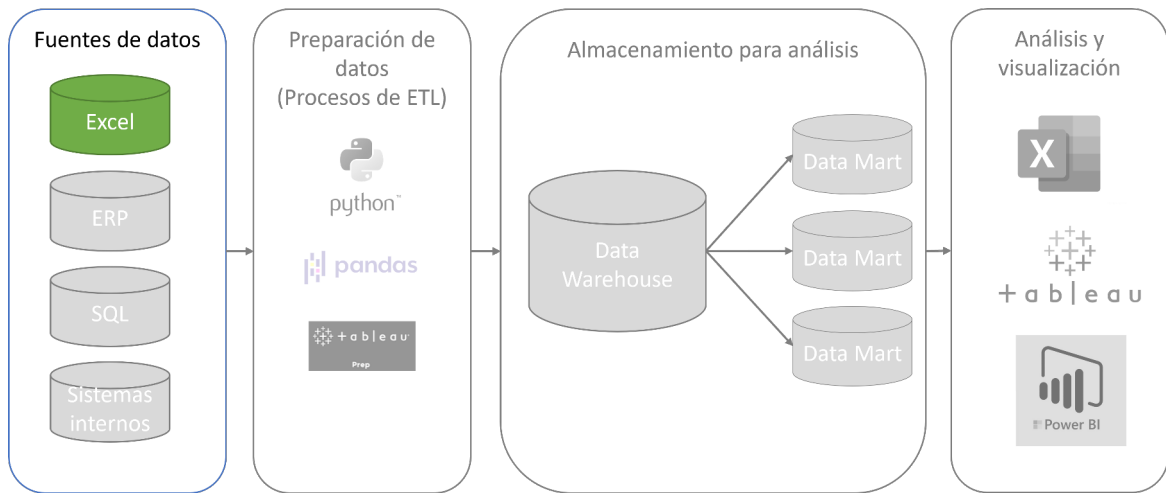


Figura 4.2: Arquitectura. Fase de selección de la fuente de datos.

1.1. Determinación de métricos y dimensiones

La curva de producción y el **rendimiento** obtenido en una producción **agrícola** dependen de una importante cantidad de factores. Para facilitar el entendimiento del modelo, estos factores serán diferenciados como métricos (valores numéricos asociados a una unidad de medida) y dimensiones (contextualización del métrico para darle significado). Es importante recalcar que no necesariamente todos los métricos se encuentran asociados a todas las dimensiones; por ejemplo, la producción está ligada a la semana, mientras que la **superficie** está asociada a toda una temporada y no a una semana específica.

Las dimensiones y métricos que estén disponibles, así como su nivel de detalle y calidad, dependerá de las capacidades tecnológicas actuales de cada unidad de producción **agrícola**. Estas pueden variar drásticamente de una a otra. Para este ejercicio, se usarán las dimensiones y métricos con los que cuenta actualmente la **agrícola** a analizar.

Dimensiones		
Dato	Descripción	Tipo de dato
Temporada	Temporada de producción, considerando como una temporada todo lo que se produce entre la semana 30 de un año y la semana 29 del año subsecuente.	Cadena de 9 caracteres. Ejemplo: 2018-2019
Semana	Semana de producción. Un año tradicional tiene 52 semanas, considerando el inicio de semana en domingo y el cierre en sábado. Cada 4 o 5 años se tiene un año de 53 semanas.	Entero. Ejemplo: 42 ó 50.
Cultivo	Fresa, frambuesa, zarzamora o arándano	Cadena de caracteres
Ciclo	Según el cultivo, se pueden tener 1 solo ciclo productivo (1 ciclo = 1 temporada) o más.	Entero mayor a 1. Un cultivo perenne puede tener una cantidad de ciclos mayor a 5.
Tipo de campo	Existen campos convencionales y orgánicos; estos últimos llevan una producción que prohíbe el uso de químicos.	Cadena de caracteres
Medio	La producción se puede dar en sistemas basados en suelo o en sustrato, es decir, producción dentro de macetas u otros elementos físicos distintos al suelo.	Cadena de caracteres
Zona de producción	Zona donde se produce, ya sea el nombre del municipio, pueblo, cabecera municipal, etc.	Cadena de caracteres
Variedad	Las <i>berries</i> , por ejemplo, la fresa, no son únicas por cultivo. Existe una cantidad significativa de variedades con características particulares. Año con año se desarrollan nuevas variedades.	Cadena de caracteres
Rancho	Nombre del rancho. Esto ayuda a relacionar el rancho con otras variables como puede ser el supervisor de rancho o el tipo de suelo.	Cadena de caracteres
Tipo de suelo	Los suelos tienen distintas propiedades según su tipo. Un suelo pesado es muy distinto a un suelo arenoso.	Cadena de caracteres

Cabe aclarar que existen otras dimensiones que son de gran relevancia para el análisis, especialmente si se pretende crear modelos de pronóstico de la producción de buena precisión. Sin embargo, en el caso de las *agrícolas* en las que se logró acceso a datos, no se cuenta con registros confiables de las mismas. Por ejemplo, la fecha de plantación y/o poda se tiene registrada, pero los métricos como producción en kilogramos no se registran a ese nivel de detalle, por lo que no aportan mayor valor para su análisis.

Métricos		
Dato	Descripción	Tipo de dato
Kilogramos	Kilogramos de <i>fruta</i> en producción	Decimal para los cálculos, aunque suele manejarse como entero sin decimales.
Hectáreas	<i>superficie</i> en producción	Decimal, típicamente con 2 decimales
Rendimiento	Relación kilogramos/hectáreas	Entero

Al igual que con las dimensiones, hay una cantidad importante de variables numéricas que, en opinión del autor de este documento, son fundamentales para averiguar si existe una correlación que permita crear modelos de pronósticos, pero que en las *agrícolas* analizadas no se miden. Algunas de las que podrían destacar son el número de brotes por metro lineal, la longitud de los laterales, los valores nutricionales, la incidencia de las *plagas* y/o las enfermedades, entre otras. Esta situación, en conjunto con la falta de ciertas dimensiones mencionadas anteriormente, demuestran algunos de los retos que tiene la producción de

berries para poder implementar sistemas de pronósticos confiables.

1.2. Identificación de las fuentes de datos

Como se explicó en el marco teórico, las fuentes de datos son variadas y pueden incluir el uso de sistemas de ERP, CRM, SCM, servidores externos, sistemas externos, RFID e incluso datos estructurados y no estructurados. En el caso que se está usando como ejemplo, los datos originales provienen de diversos archivos en formato de hoja de cálculo. Los registros se almacenan en archivos separados por cultivo y temporada. La empresa en cuestión lleva registros de las últimas 7 temporadas de producción.

1.3. Selección del tipo de conexión

Por las razones descritas anteriormente, es necesario seleccionar tecnologías para el análisis y procesamiento de datos. En esa parte del proyecto se trabajará con Python, la librería de Pandas y conexiones a archivos .XLSX.

Aunque ya se ha mencionado, es importante recordar que la obtención de los datos será distinta de compañía en compañía, y no forma parte de los objetivos de este proyecto ni el diseño ni la implementación de una plataforma que permita recabar datos; por el contrario, se pretende demostrar que es posible generar esquemas de análisis para la planeación estratégica que se adapten a distintas estructuras de datos, siempre y cuando estos pasen por un proceso de limpieza y transformación adecuado. Por esta razón, se hará uso de las fuentes de mencionadas a manera de demostración.

A partir de esta sección, se usarán como fuente las hojas de cálculo de una empresa de producción de berries que cuenta con sus archivos de volumen, superficie y precio de 7 temporadas y de 4 cultivos. Estos archivos presentan un mismo formato que tiene las siguientes columnas: compañía, zona de producción, rancho, temporada, cultivo, sistema de producción, medio de producción, tipo de negocio, semana, variedad, hectáreas, kilogramos y precio por kilo.

Las Figuras 4.3 y 4.4 ejemplifican la cantidad de archivos y la estructura que presentan. Cabe destacar que, como se aprecia en la Figura 4.4, no todos los campos tienen necesariamente valores, dado que no hay producción durante las 52 semanas que componen el año.

Una vez que se conocen las fuentes de datos, es necesario elegir una tecnología que permita su obtención y transformación. En este proyecto, y para poder trabajar con estos datos, se utilizará el lenguaje de Python 3.8.3, así como la librería de Pandas. Con este lenguaje podemos leer los 27 archivos e incorporarlos en un solo DataFrame que contenga la información íntegra lista para su procesamiento.

Name	Date modified	Type	Size
Black 13-14.xlsx	10/21/2020 9:56 PM	Microsoft Excel W...	57 KB
Black 14-15.xlsx	10/21/2020 9:56 PM	Microsoft Excel W...	58 KB
Black 15-16.xlsx	10/21/2020 9:56 PM	Microsoft Excel W...	56 KB
Black 16-17.xlsx	10/21/2020 9:56 PM	Microsoft Excel W...	55 KB
Black 17-18.xlsx	10/21/2020 9:56 PM	Microsoft Excel W...	64 KB
Black 18-19.xlsx	10/21/2020 9:56 PM	Microsoft Excel W...	70 KB
Black 19-20.xlsx	10/21/2020 9:56 PM	Microsoft Excel W...	59 KB
Blue 14-15.xlsx	10/21/2020 9:56 PM	Microsoft Excel W...	14 KB
Blue 15-16.xlsx	10/21/2020 9:56 PM	Microsoft Excel W...	15 KB
Blue 16-17.xlsx	10/21/2020 9:56 PM	Microsoft Excel W...	15 KB
Blue 17-18.xlsx	10/21/2020 9:56 PM	Microsoft Excel W...	13 KB
Blue 18-19.xlsx	10/21/2020 9:56 PM	Microsoft Excel W...	18 KB
Blue 19-20.xlsx	10/21/2020 9:56 PM	Microsoft Excel W...	26 KB
Rasp 13-14.xlsx	10/21/2020 9:56 PM	Microsoft Excel W...	25 KB
Rasp 14-15.xlsx	10/21/2020 9:56 PM	Microsoft Excel W...	35 KB
Rasp 15-16.xlsx	10/21/2020 9:56 PM	Microsoft Excel W...	41 KB
Rasp 16-17.xlsx	10/21/2020 9:56 PM	Microsoft Excel W...	33 KB
Rasp 17-18.xlsx	10/21/2020 9:56 PM	Microsoft Excel W...	30 KB
Rasp 18-19.xlsx	10/21/2020 9:56 PM	Microsoft Excel W...	40 KB
Rasp 19-20.xlsx	10/21/2020 9:56 PM	Microsoft Excel W...	37 KB
Straw 13-14.xlsx	10/21/2020 9:56 PM	Microsoft Excel W...	24 KB
Straw 14-15.xlsx	10/21/2020 9:56 PM	Microsoft Excel W...	49 KB
Straw 15-16.xlsx	10/21/2020 9:56 PM	Microsoft Excel W...	39 KB
Straw 16-17.xlsx	10/21/2020 9:56 PM	Microsoft Excel W...	34 KB
Straw 17-18.xlsx	10/21/2020 9:56 PM	Microsoft Excel W...	48 KB
Straw 18-19.xlsx	10/21/2020 9:56 PM	Microsoft Excel W...	76 KB
Straw 19-20.xlsx	10/31/2020 3:13 PM	Microsoft Excel W...	96 KB

Figura 4.3: Archivos fuente de las últimas 7 temporadas.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1											Area	Volume	WA Price
2	Empresa, S.P.R. de R.L. de C.V.	Purepero	Doña Bety	2013-2014	Strawberries	Conventional	Soil High	1330	Variedad 1	1.89			
3									Variedad 2	11.77			
4									Variedad 3	0.59			
5									Variedad 4	0.27			
6									Variedad 5	0.48			
7									1331	Variedad 1	1.89		
8									Variedad 2	11.77			
9									Variedad 3	0.59			
10									Variedad 4	0.27			
11									Variedad 5	0.48			
12									1332	Variedad 1	1.89		

Figura 4.4: Extracto de uno de los archivos con sus datos crudos.

2. Preparación de datos

En esta sección, se presenta el proceso a seguir para la transformación de los datos previo a su procesamiento y análisis. Este proceso incluye transformaciones de distinta índole tanto en la estructura como en los datos mismos. Se hará uso de Python y Pandas para cubrir las siguientes necesidades: rellenar campos vacíos, eliminar columnas innecesarias, definir nombres de los encabezados, eliminar filas con datos nulos, traducir términos del


```
Out[1]:
```

	Unnamed: 0	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6	Unnamed: 7	Unnamed: 8	Unnamed: 9	Area	Volume	WA Price
0	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	2013-2014	Blackberries	Conventional	Soil	Hooped	1330.0	Sleeping Beauty	0.837700	NaN	NaN
1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Sleeping Beauty	1.970821	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Tupi	3.751439	NaN	NaN
3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Vanessa	0.319702	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1331.0	Sleeping Beauty	0.837700	NaN	NaN
...
16392	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2029.0	Anya	0.896025	NaN	3.956212

Figura 4.5: Esquema de datos original cargado en Python y Pandas.

inglés al español y calcular columnas adicionales que permitan, eventualmente, trabajar con series de tiempo.

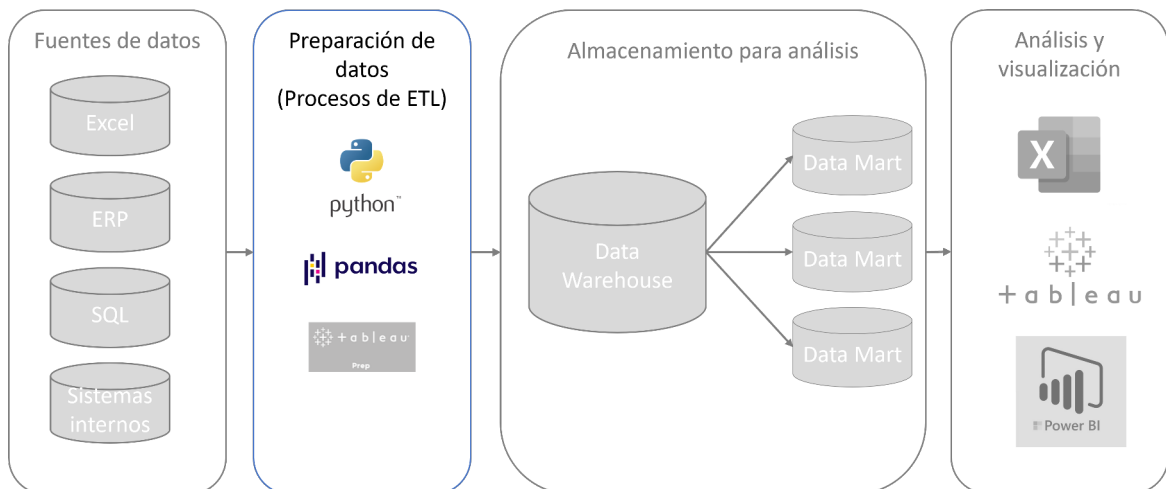


Figura 4.6: Arquitectura. Fase de preparación de datos usando Python y Pandas.

2.1. Proceso de preparación de los datos

En la sección anterior se utilizó Python para consolidar los datos crudos de 27 archivos distintos en uno solo. A continuación, será necesario someter estos datos a un proceso de transformación. Para ello, es fundamental conocer a detalle el significado de los datos que se están trabajando; de lo contrario, se corre el riesgo de incurrir en una malinterpretación de los datos que puede comprometer el resultado de los modelos. Para este `DataFrame` que fue generado en Python será necesario realizar las siguientes transformaciones:

2.1.1. Definir encabezados

Los datos crudos no poseen encabezados en la mayoría de las columnas. Es necesario definir un nombre para cada columna que sea representativo de su contenido.

Out[2]:

	Compañía	Zona	Rancho	Temporada	Cultivo	Sistema	Medio	Ciclo	Semana	Variedad	Hectáreas	Kilogramos	Precio de venta
0	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	2013-2014	Blackberries	Conventional	Soil	Hooped	1330.0	Sleeping Beauty	0.837700	NaN	NaN
1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Sleeping Beauty.	1.970821	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Tupi	3.751439	NaN	NaN
3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Vanessa	0.319702	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	1331.0	Sleeping Beauty	0.837700	NaN	NaN
...
16392	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	2029.0	Anya	0.896025	NaN	3.956212
16393	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	El Dorado	10.850055	NaN	4.253294
16394	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Ivonne	2.383331	NaN	4.080881

Figura 4.7: Agregando encabezados a los datos crudos.

2.1.2. Rellenar celdas nulas de las columnas de tipo dimensión

Los datos se encuentran en forma tabular sin repetición; es decir, los datos como zona o rancho no se repiten en cada renglón. Estas celdas nulas deben ser rellenas con el dato más reciente que se encuentre localizado en el primer renglón superior no vacío.

Out[3]:

	Compañía	Zona	Rancho	Temporada	Cultivo	Sistema	Medio	Ciclo	Semana	Variedad	Hectáreas	Kilogramos	Precio de venta
0	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	2013-2014	Blackberries	Conventional	Soil	Hooped	1330.0	Sleeping Beauty	0.837700	NaN	NaN
1	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	2013-2014	Blackberries	Conventional	Soil	Hooped	1330.0	Sleeping Beauty.	1.970821	NaN	NaN
2	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	2013-2014	Blackberries	Conventional	Soil	Hooped	1330.0	Tupi	3.751439	NaN	NaN
3	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	2013-2014	Blackberries	Conventional	Soil	Hooped	1330.0	Vanessa	0.319702	NaN	NaN
4	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	2013-2014	Blackberries	Conventional	Soil	Hooped	1331.0	Sleeping Beauty	0.837700	NaN	NaN

Figura 4.8: Tabulando los datos.

2.1.3. Traducir términos al español

Aunque no es una tarea crítica, se recomienda traducir la información al español para hacer reportes más adecuados para los usuarios finales.

Out[4]:

	Compañía	Zona	Rancho	Temporada	Cultivo	Sistema	Medio	Ciclo	Semana	Variedad	Hectáreas	Kilogramos	Precio de venta
0	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	2013-2014	Zarzamora	Convencional	Suelo	Perenne	1330.0	Sleeping Beauty	0.837700	NaN	NaN
1	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	2013-2014	Zarzamora	Convencional	Suelo	Perenne	1330.0	Sleeping Beauty	1.970821	NaN	NaN
2	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	2013-2014	Zarzamora	Convencional	Suelo	Perenne	1330.0	Tupi	3.751439	NaN	NaN
3	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	2013-2014	Zarzamora	Convencional	Suelo	Perenne	1330.0	Vanessa	0.319702	NaN	NaN
4	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	2013-2014	Zarzamora	Convencional	Suelo	Perenne	1331.0	Sleeping Beauty	0.837700	NaN	NaN

Figura 4.9: Traduciendo de inglés a español.

2.1.4. Eliminar precio de venta

Por el momento, el precio de venta no es de gran significancia dado que en la obtención de los datos no se encontró un registro adecuado de los costos de inversión, mantenimiento y producción. Es posible omitir estos datos temporalmente y retomarlos en el momento en que sí agreguen valor para el cálculo de rentabilidad.

Out[5]:

	Compañía	Zona	Rancho	Temporada	Cultivo	Sistema	Medio	Ciclo	Semana	Variedad	Hectáreas	Kilogramos
0	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	2013-2014	Zarzamora	Convencional	Suelo	Perenne	1330.0	Sleeping Beauty	0.837700	NaN
1	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	2013-2014	Zarzamora	Convencional	Suelo	Perenne	1330.0	Sleeping Beauty	1.970821	NaN
2	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	2013-2014	Zarzamora	Convencional	Suelo	Perenne	1330.0	Tupi	3.751439	NaN
3	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	2013-2014	Zarzamora	Convencional	Suelo	Perenne	1330.0	Vanessa	0.319702	NaN
4	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	2013-2014	Zarzamora	Convencional	Suelo	Perenne	1331.0	Sleeping Beauty	0.837700	NaN
...
16392	Jukskani, S.P.R. de R.L. de C.V.	Purepero	Bxangaru 19-20	2019-2020	Fresa	Convencional	Suelo	1	2029.0	Anyá	0.896025	NaN
16393	Jukskani, S.P.R. de R.L. de C.V.	Purepero	Bxangaru 19-20	2019-2020	Fresa	Convencional	Suelo	1	2029.0	El Dorado	10.850055	NaN

Figura 4.10: Eliminando precio de venta.

2.1.5. Calcular la fecha de cierre de semana

Para poder hacer uso adecuado de series de tiempo es necesario calcular una fecha basada en el número de semana y el año. Como se puede apreciar en los datos originales, tanto la superficie como los kilogramos se encuentran agrupados por semana.

Out[6]:

	Compañía	Zona	Rancho	Temporada	Cultivo	Sistema	Medio	Ciclo	Semana	Variedad	Hectáreas	Kilogramos	Año	Semana del año	Fecha cierre de semana
0	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	2013-2014	Zarzamora	Convencional	Suelo	Perenne	1330	Sleeping Beauty	0.837700	NaN	2013	30	2013-07-27
1	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	2013-2014	Zarzamora	Convencional	Suelo	Perenne	1330	Sleeping Beauty	1.970821	NaN	2013	30	2013-07-27
2	La Rojeña Don Antonio,	Zamora	1Chirimoyo	2013-2014	Zarzamora	Convencional	Suelo	Perenne	1330	Tupi	3.751439	NaN	2013	30	2013-

Figura 4.11: Calculando la semana.

2.1.6. Separar un DataFrame para volumen y otro para superficie

La superficie se encuentra repetida durante las 52 semanas del año para cada renglón que posea las mismas dimensiones. Esto genera duplicidad de datos, así que es recomendable separar 2 DataFrames: uno para superficie agrupada por temporada y otro de volumen por semana. En el DataFrame de volumen se deben encontrar los valores nulos y sustituirlos por el valor de 0. Los datos crudos incluyen valores nulos donde no hubo producción; sin embargo, es importante cambiar estos valores a 0 para que las series de tiempo identifiquen la existencia de un valor.

Out[7]:

Fecha cierre de semana	Compañía	Zona	Rancho	Cultivo	Temporada	Sistema	Medio	Ciclo	Semana	Variedad	Kilogramos	Año	Semana del año
2013-07-27	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	2013-2014	Convencional	Suelo	Perenne	1330	Sleeping Beauty	0.0	2013	30
2013-07-27	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	2013-2014	Convencional	Suelo	Perenne	1330	Sleeping Beauty	0.0	2013	30
2013-07-27	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	2013-2014	Convencional	Suelo	Perenne	1330	Tupi	0.0	2013	30
2013-07-27	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	2013-2014	Convencional	Suelo	Perenne	1330	Vanessa	0.0	2013	30
2013-08-03	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	2013-2014	Convencional	Suelo	Perenne	1331	Sleeping Beauty	0.0	2013	31
...
2020-07-18	Juiskani, S.P.R. de R.L. de C.V.	Purepero	Bxangaru 19-20	Fresa	2019-2020	Convencional	Suelo	1	2029	Anya	0.0	2020	29
2020-07-18	Juiskani, S.P.R. de R.L. de C.V.	Purepero	Bxangaru 19-20	Fresa	2019-2020	Convencional	Suelo	1	2029	El Dorado	0.0	2020	29
2020-07-18	Juiskani, S.P.R. de R.L. de C.V.	Purepero	Bxangaru 19-20	Fresa	2019-2020	Convencional	Suelo	1	2029	Ivonne	0.0	2020	29
2020-07-18	Juiskani, S.P.R. de R.L. de C.V.	Purepero	Bxangaru 19-20	Fresa	2019-2020	Convencional	Suelo	1	2029	Marquis	0.0	2020	29
2020-07-18	Juiskani, S.P.R. de R.L. de C.V.	Purepero	Bxangaru 19-20	Fresa	2019-2020	Convencional	Suelo	1	2029	Xareni	0.0	2020	29

Figura 4.12: Separando los DataFrames.

2.1.7. Asignación de índices

Para cada `DataFrame` será necesario asignar un índice, el cual será la temporada.

Out[8]:

	Compañía	Zona	Rancho	Cultivo	Sistema	Medio	Ciclo	Semana	Variedad	Hectáreas	Kilogramos	Año	Semana del año	Fecha cierre de semana
Temporada														
2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	Convencional	Suelo	Perenne	1330	Sleeping Beauty	0.837700	NaN	2013	30	2013-07-27
2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	Convencional	Suelo	Perenne	1330	Sleeping Beauty	1.970821	NaN	2013	30	2013-07-27
2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	Convencional	Suelo	Perenne	1330	Tupi	3.751439	NaN	2013	30	2013-07-27
2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	Convencional	Suelo	Perenne	1330	Vanessa	0.319702	NaN	2013	30	2013-07-27
2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	Convencional	Suelo	Perenne	1331	Sleeping Beauty	0.837700	NaN	2013	31	2013-08-03
...
2019-2020	Juiskani, S.P.R. de R.L. de C.V.	Purepero	Bxangaru 19-20	Fresa	Convencional	Suelo	1	2029	Anya	0.896025	NaN	2020	29	2020-07-18
2019-2020	Juiskani, S.P.R. de R.L. de C.V.	Purepero	Bxangaru 19-20	Fresa	Convencional	Suelo	1	2029	El Dorado	10.850055	NaN	2020	29	2020-07-18
2019-2020	Juiskani, S.P.R. de R.L. de C.V.	Purepero	Bxangaru 19-20	Fresa	Convencional	Suelo	1	2029	Ivonne	2.383331	NaN	2020	29	2020-07-18
2019-2020	Juiskani, S.P.R. de R.L. de C.V.	Purepero	Bxangaru 19-20	Fresa	Convencional	Suelo	1	2029	Marquis	1.275861	NaN	2020	29	2020-07-18
2019-2020	Juiskani, S.P.R. de R.L. de C.V.	Purepero	Bxangaru 19-20	Fresa	Convencional	Suelo	1	2029	Xareni	1.811691	NaN	2020	29	2020-07-18

Figura 4.13: Asignación de índices.

2.1.8. Eliminar superficies nulas y duplicadas del `DataFrame` de superficie

Una vez que se crea el `DataFrame` de superficie, se deben eliminar todos aquellos registros nulos; es decir, donde no hay superficie, dado que no aportan ningún valor al análisis. Además, al crear el `DataFrame` sin número de semana, se vuelve necesario eliminar todos los registros duplicados para que queden solamente los registros únicos y con datos relevantes.

Out[9]:

Temporada	Compañía	Zona	Rancho	Cultivo	Sistema	Medio	Ciclo	Variedad	Hectáreas
2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	Convencional	Suelo	Perenne	Sleeping Beauty	0.837700
2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	Convencional	Suelo	Perenne	Sleeping Beauty	1.970821
2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	Convencional	Suelo	Perenne	Tupi	3.751439
2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	Convencional	Suelo	Perenne	Vanessa	0.319702
2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	Milagro Org	Zarzamora	Orgánico	Suelo	Perenne	Dasha	0.348030
...
2019-2020	Juiskani, S.P.R. de R.L. de C.V.	Purepero	Bxangaru 19-20	Fresa	Convencional	Suelo	1	Anya	0.896025
2019-2020	Juiskani, S.P.R. de R.L. de C.V.	Purepero	Bxangaru 19-20	Fresa	Convencional	Suelo	1	El Dorado	10.850055
2019-2020	Juiskani, S.P.R. de R.L. de C.V.	Purepero	Bxangaru 19-20	Fresa	Convencional	Suelo	1	Ivonne	2.383331
2019-2020	Juiskani, S.P.R. de R.L. de C.V.	Purepero	Bxangaru 19-20	Fresa	Convencional	Suelo	1	Marquis	1.275861
2019-2020	Juiskani, S.P.R. de R.L. de C.V.	Purepero	Bxangaru 19-20	Fresa	Convencional	Suelo	1	Xareni	1.811691

Figura 4.14: Eliminando superficies repetidas.

2.1.9. Guardar los DataFrames para su posterior uso en los modelos

Se guardarán 3 DataFrames: uno para volumen, otro para superficie y un tercero previo a la separación de estos conceptos como referencia. Así tendremos estos DataFrames disponibles tanto para Python como para otras tecnologías. Será de utilidad guardarlos en Excel para usos posteriores. Aunque lo ideal es usar un Data Warehouse formal, tal y como se muestra en la Figura 4.15, este ejercicio se mantendrá simple usando como salida un archivo de Excel que ya se encuentra estandarizado.

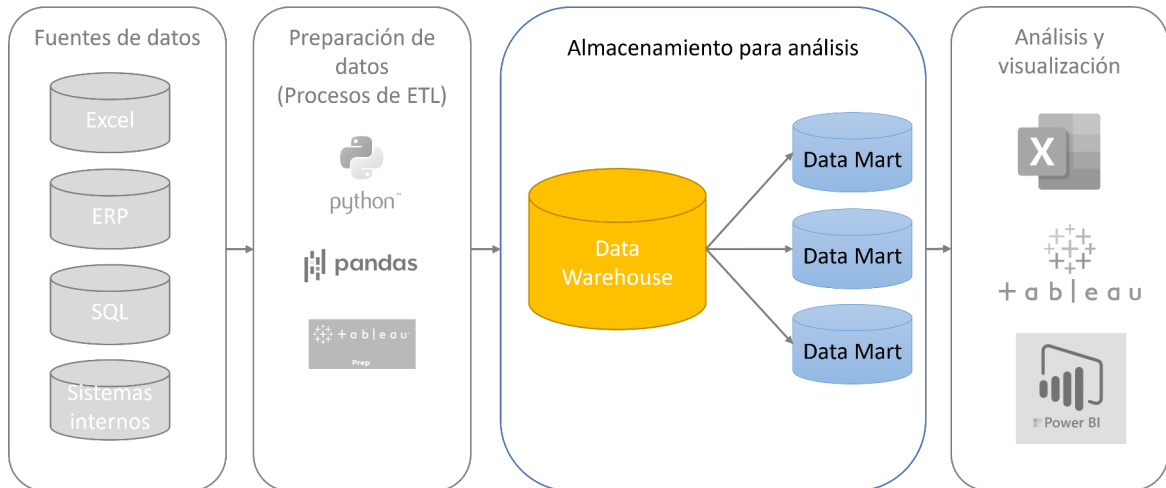


Figura 4.15: Arquitectura. Fase de almacenamiento.

2. PREPARACIÓN DE DATOS CAPÍTULO 4. DESARROLLO METODOLÓGICO

	A	B	C	D	E	F	G	H	I	J
1	Temporada	Compañía	Zona	Rancho	Cultivo	Sistema	Medio	Ciclo	Variedad	Hectáreas
2	2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	Convencional	Suelo	Perenne	Sleeping Beauty	0.83770002
3	2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	Convencional	Suelo	Perenne	Sleeping Beauty.	1.97082082
4	2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	Convencional	Suelo	Perenne	Tupi	3.75143922
5	2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	Convencional	Suelo	Perenne	Vanessa	0.31970194
6	2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	Milagro Org	Zarzamora	Orgánico	Suelo	Perenne	Dasha	0.34802996
7	2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	Milagro Org	Zarzamora	Orgánico	Suelo	Perenne	Vanessa	1.42854158
8	2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	Ocales	Zarzamora	Convencional	Suelo	Perenne	Dasha	2.51714692
9	2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	Ocales	Zarzamora	Convencional	Suelo	Perenne	Tupi	5.22854312
10	2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	Ocales	Zarzamora	Convencional	Suelo	Perenne	Vanessa	1.0926522
11	2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	Rojeña	Zarzamora	Convencional	Suelo	Perenne	BQ706.7	0.06474976
12	2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	Rojeña	Zarzamora	Convencional	Suelo	Perenne	BQ706-3	0.06474976
13	2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	Rojeña	Zarzamora	Convencional	Suelo	Perenne	BQ727.3	0.06474976

Figura 4.16: DataFrame de superficie después de su procesamiento.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Fecha cierre de semana	Compañía	Zona	Rancho	Cultivo	Temporada	Sistema	Medio	Ciclo	Semana	Variedad	Kilogramos	Año	Semana del año
2	2013-07-27 00:00:00	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	2013-2014	Convencional	Suelo	Perenne	1330	Sleeping Beauty	0	2013	30
3	2013-07-27 00:00:00	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	2013-2014	Convencional	Suelo	Perenne	1330	Sleeping Beauty.	0	2013	30
4	2013-07-27 00:00:00	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	2013-2014	Convencional	Suelo	Perenne	1330	Tupi	0	2013	30
5	2013-07-27 00:00:00	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	2013-2014	Convencional	Suelo	Perenne	1330	Vanessa	0	2013	30
6	2013-08-03 00:00:00	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	2013-2014	Convencional	Suelo	Perenne	1331	Sleeping Beauty	0	2013	31
7	2013-08-03 00:00:00	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	2013-2014	Convencional	Suelo	Perenne	1331	Sleeping Beauty.	0	2013	31
8	2013-08-03 00:00:00	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	2013-2014	Convencional	Suelo	Perenne	1331	Tupi	0	2013	31
9	2013-08-03 00:00:00	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	2013-2014	Convencional	Suelo	Perenne	1331	Vanessa	0	2013	31
10	2013-08-10 00:00:00	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	2013-2014	Convencional	Suelo	Perenne	1332	Sleeping Beauty	0	2013	32
11	2013-08-10 00:00:00	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	2013-2014	Convencional	Suelo	Perenne	1332	Sleeping Beauty.	0	2013	32
12	2013-08-10 00:00:00	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	2013-2014	Convencional	Suelo	Perenne	1332	Tupi	0	2013	32
13	2013-08-10 00:00:00	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	2013-2014	Convencional	Suelo	Perenne	1332	Vanessa	0	2013	32
14	2013-08-17 00:00:00	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	2013-2014	Convencional	Suelo	Perenne	1333	Sleeping Beauty	0	2013	33

Figura 4.17: DataFrame de volumen después de su procesamiento.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Temporada	Compañía	Zona	Rancho	Cultivo	Sistema	Medio	Ciclo	Semana	Variedad	Hectáreas	Kilogramos	Año	Semana del año	Fecha cierre de semana
2	2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	Convencional	Suelo	Perenne	1330	Sleeping Beauty	0.83770002	2013	30	2013-07-27 00:00:00	
3	2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	Convencional	Suelo	Perenne	1330	Sleeping Beauty.	1.97082082	2013	30	2013-07-27 00:00:00	
4	2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	Convencional	Suelo	Perenne	1330	Tupi	3.75143922	2013	30	2013-07-27 00:00:00	
5	2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	Convencional	Suelo	Perenne	1330	Vanessa	0.31970194	2013	30	2013-07-27 00:00:00	
6	2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	Convencional	Suelo	Perenne	1331	Sleeping Beauty	0.83770002	2013	31	2013-08-03 00:00:00	
7	2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	Convencional	Suelo	Perenne	1331	Sleeping Beauty.	1.97082082	2013	31	2013-08-03 00:00:00	
8	2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	Convencional	Suelo	Perenne	1331	Tupi	3.75143922	2013	31	2013-08-03 00:00:00	
9	2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	Convencional	Suelo	Perenne	1331	Vanessa	0.31970194	2013	31	2013-08-03 00:00:00	
10	2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	Convencional	Suelo	Perenne	1332	Sleeping Beauty	0.83770002	2013	32	2013-08-10 00:00:00	
11	2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	Convencional	Suelo	Perenne	1332	Sleeping Beauty.	1.97082082	2013	32	2013-08-10 00:00:00	
12	2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	Convencional	Suelo	Perenne	1332	Tupi	3.75143922	2013	32	2013-08-10 00:00:00	
13	2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	Convencional	Suelo	Perenne	1332	Vanessa	0.31970194	2013	32	2013-08-10 00:00:00	
14	2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	Convencional	Suelo	Perenne	1333	Sleeping Beauty	0.83770002	2013	33	2013-08-17 00:00:00	
15	2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	Convencional	Suelo	Perenne	1333	Sleeping Beauty.	1.97082082	2013	33	2013-08-17 00:00:00	
16	2013-2014	La Rojeña Don Antonio, S.P.R. de R.L. de C.V.	Zamora	1Chirimoyo	Zarzamora	Convencional	Suelo	Perenne	1333	Tupi	3.75143922	2013	33	2013-08-17 00:00:00	

Figura 4.18: DataFrame general después de su procesamiento.

Este procesamiento ha permitido generar 3 estructuras de datos basados en los mismos datos crudos, y que pueden ser exploradas y analizadas para extraer información desde distintas aristas que ayuden a la toma de decisiones. En las siguientes secciones se evaluarán formas de análisis tanto simples como algunas más elaboradas para dimensionar el alcance de estos conjuntos de datos.

3. Exploración de datos

En esta sección se presentan algunos ejemplos básicos de cómo los datos que ya fueron procesados pueden analizarse tanto gráfica como numéricamente. Existen diversas compañías en el mercado que han desarrollado aplicaciones con este propósito, destacando Tableau y Power BI como fuertes competidores en los Cuadrantes Mágicos de Gartner.

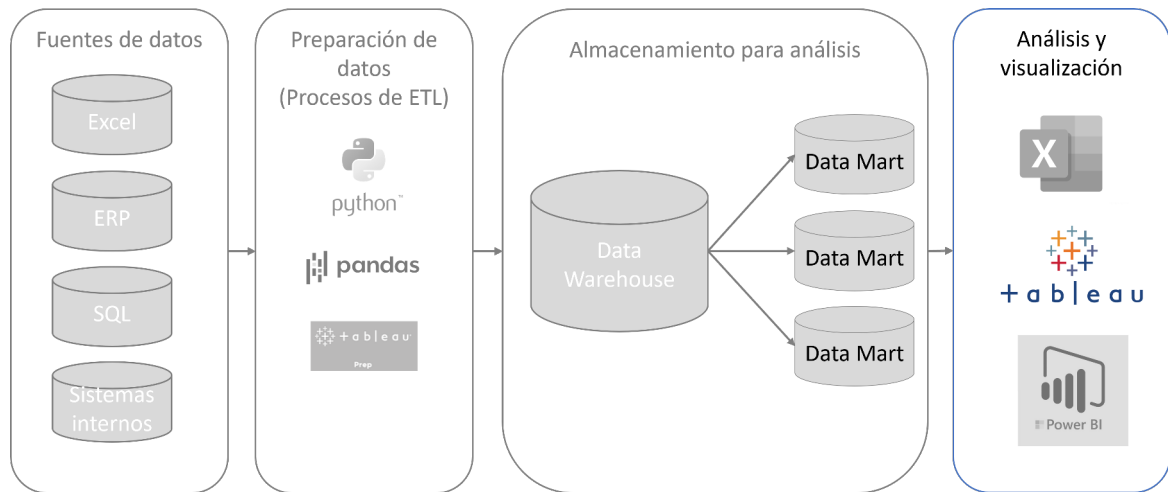


Figura 4.19: Arquitectura. Fase de análisis y visualización.

3.1. Conectando Tableau Desktop a la fuente de datos

Como se mencionó en el marco teórico, existen distintas aplicaciones en el mercado enfocadas a proveer capacidades de *Business Intelligence*, análisis y/o reporte. Para ejemplificar el alcance que se puede lograr en negocios agrícolas una vez que se tienen datos procesados y estandarizados, se usará la plataforma de Tableau Desktop en su versión 2020.3.3 para hacer exploración de datos. Por el momento, se hará la conexión directamente a los *DataFrames* que fueron guardados en hojas de cálculo para demostrar la calidad de análisis que se puede lograr aun, cuando no se esté usando un Data Warehouse formal.

Para esto, el primer paso es configurar la conexión de Tableau hacia los *DataFrames* estandarizados (ver Figura 4.20). Por practicidad, se han colocado los *DataFrames* de volumen y superficie en una sola hoja de cálculo pero en hojas de trabajo separadas. Es a este archivo al que Tableau debe conectarse para comenzar la exploración.

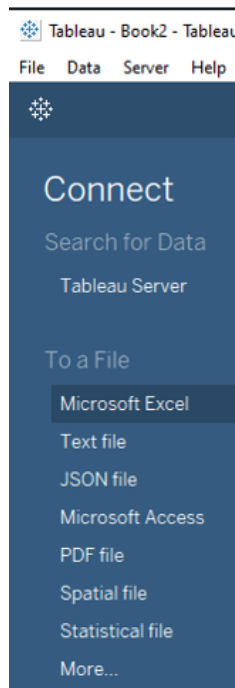


Figura 4.20: Conexión a Excel desde Tableau.

Una vez conectados a la fuente de datos, es necesario definir las entidades que formarán parte de nuestro análisis. De no establecer la conexión adecuadamente, se corre el riesgo de que **Tableau** no interprete los datos correctamente. En este caso, las hojas de volumen y superficie representan las entidades relacionadas (ver Figura 4.21).

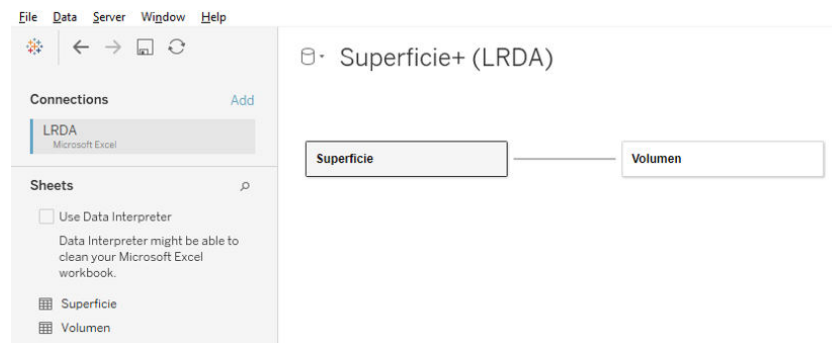


Figura 4.21: Estableciendo la fuente de datos.

Por último, se debe especificar cómo es la naturaleza del comportamiento de esta relación. Para ello, es necesario especificar cuáles son los campos de cada hoja que deben coincidir para asegurar que los métricos estén haciendo referencia a las dimensiones correctas (ver Figura 4.22).

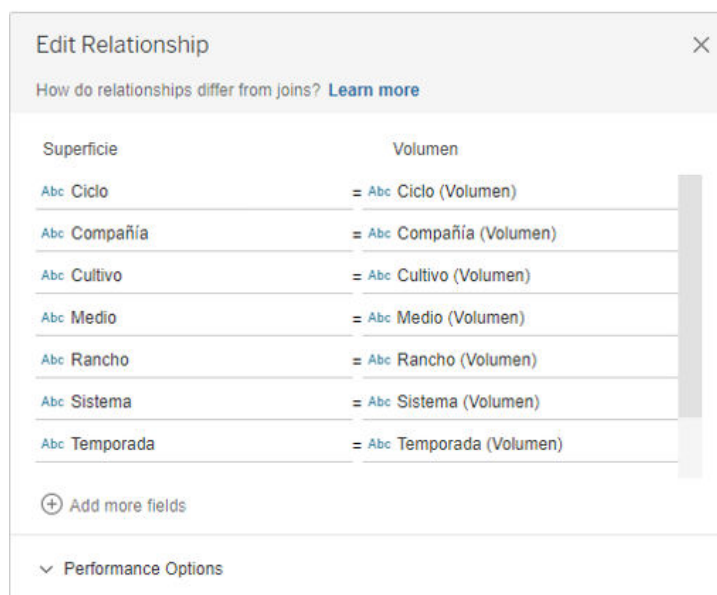


Figura 4.22: Estableciendo relaciones en Tableau.

3.2. Exploración simple de datos

Una vez que se ha especificado la fuente de datos y se han configurado las relaciones que existen entre las dimensiones, es posible comenzar a usar **Tableau** para crear hojas de análisis, reportes y/o tableros de control. En esta sección, se demostrará el potencial de esta forma de trabajo para hacer exploraciones simples de los datos que permitan responder preguntas que estén directamente relacionadas con los datos originales; es decir, todo lo relacionado con el volumen de producción y la **superficie**.

3.2.1. ¿Cuál ha sido la evolución de la agrícola a través de los años?

En las Figuras 4.23 y 4.24 se grafica la evolución en **superficie** y volumen que la agrícola ha tenido temporada tras temporada. Desde una apreciación visual simple, se puede apreciar que el ritmo de crecimiento en **superficie** ha sido, en general, más acelerado que el crecimiento en kilogramos, lo cual nos lleva a pensar que el **rendimiento** por hectárea en las temporadas 2015-16, 2016-17 y 2018-19 fue menor al de otros años. Esto podrá ser evaluado más adelante cuando se realice una exploración de datos más avanzada que haga uso de cálculos en **Tableau** a través de la creación de lo que se conoce como campos calculados.

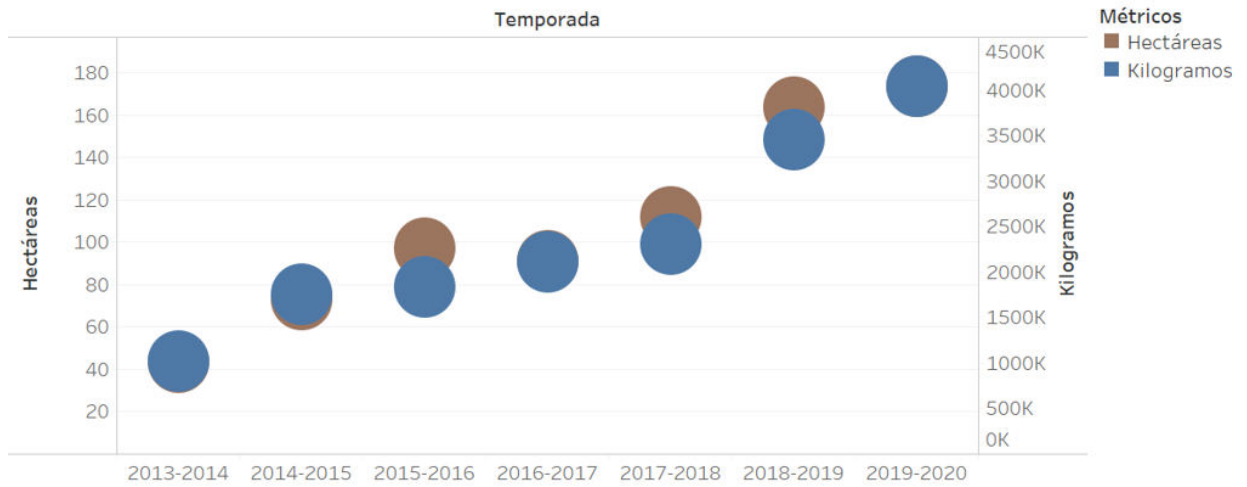


Figura 4.23: Evolución de la agrícola en superficie y volumen.

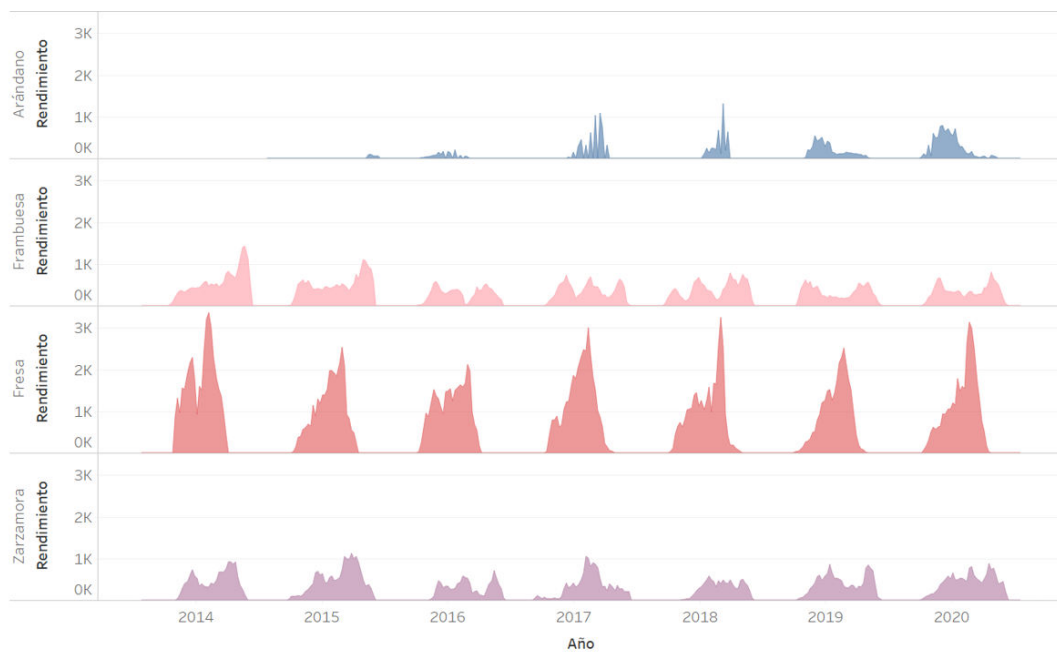


Figura 4.24: Evolución de la agrícola en volumen.

3.2.2. ¿Será que el efecto anterior se presenta en todos los cultivos?

Es posible que el efecto general de la agrícola no se vea reflejado en todos los cultivos. Cabe la posibilidad que algunos cultivos tengan una mejor tendencia. Para ello, la Figura 4.25 permite, nuevamente por apreciación visual, identificar las tendencias por cada cultivo. Evidentemente, en algún momento será necesario hacer una evaluación estadística de estos valores para tener la seguridad de la tendencia que presenta cada cultivo; sin embargo,

las referencias visuales son muy útiles para determinar en qué casos es necesario hacer un análisis de mayor profundidad.

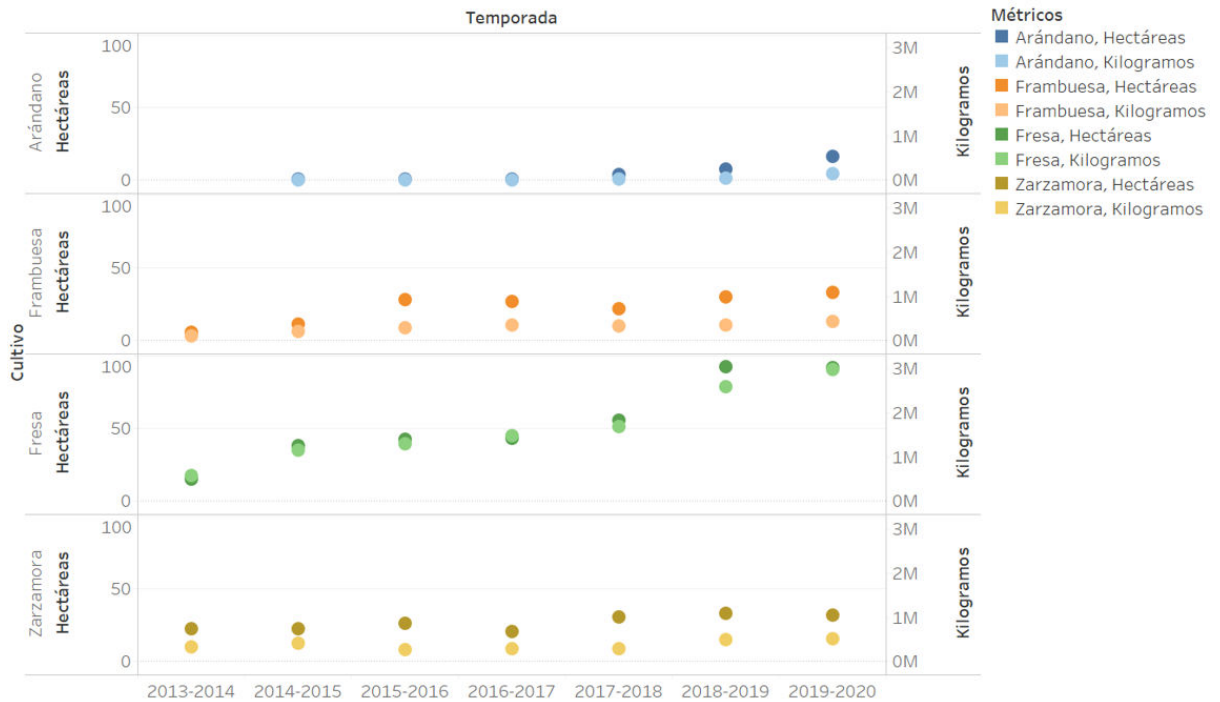


Figura 4.25: Desglose por cultivo de superficie y volumen.

3.2.3. Uso de tableros de control

Uno de los usos más comunes en las tecnologías de análisis de datos es la incorporación de *dashboards*, también conocidos como tableros de control, los cuales permiten agrupar distintas vistas de métricos y dimensiones, de tal manera que permita al lector tener un panorama más amplio de la información en una sola hoja. Por ejemplo, la Figura 4.26 muestra, al mismo tiempo, distintas aristas y maneras de presentar la información de la superficie de la agrícola solamente de la temporada 2019-20. Algunas de éstas incluyen su distribución por cultivo, las variedades plantadas y los ranchos en producción.

Hasta el momento, el enfoque ha sido utilizar Tableau para representar gráficamente los datos ya procesados, de manera que permita al lector lograr un primer nivel de comprensión de los datos que es difícil lograr desde un listado meramente numérico. A continuación, la intención es realizar algunos análisis más avanzados haciendo uso de campos calculados. No hay que confundir los análisis avanzados con la implementación de métodos para análisis estadísticos ni con la implementación de modelos para la toma de decisiones, los cuáles quedan fuera del alcance de este capítulo.

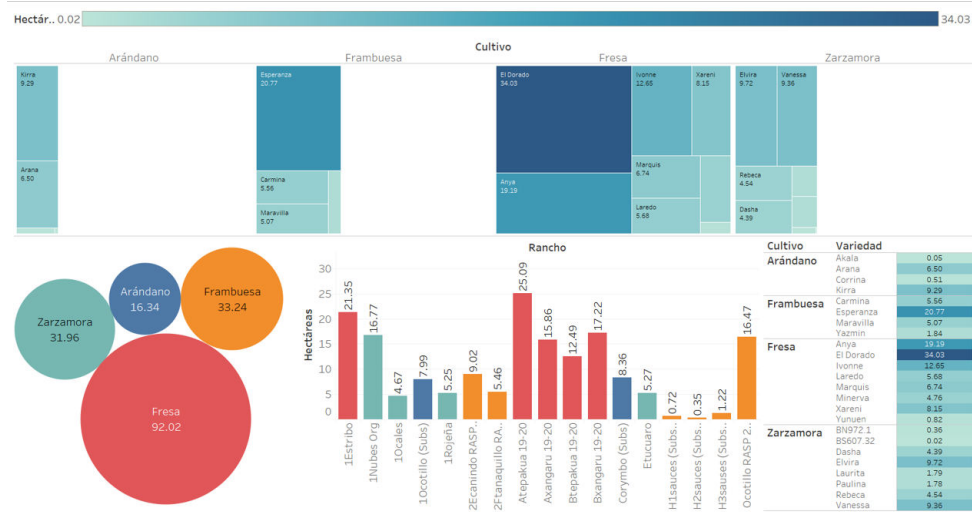


Figura 4.26: Ejemplo de un tablero de control.

3.3. Exploración avanzada de datos

Para ejemplificar un análisis más avanzado de los datos, incorporaremos un nuevo métrico llamado **rendimiento** que no se encuentra en los datos originales pero que puede calcularse al dividir el volumen entre la superficie. Dado que el volumen viene dado por los kilogramos producidos y la superficie se mide en hectáreas, el **rendimiento** usará como unidad de medida los kilogramos por hectárea ó kg/ha.

Es importante recalcar que este paso no se realiza en el ETL al ser un cálculo dinámico y no estático. Esto significa que el cálculo de rendimiento se debe recalcular según el nivel al que se busque analizar la información. Por ejemplo, este cálculo puede hacerse a nivel de zona de producción, de productor, de rancho, o de variedad.

3.3.1. ¿Cuáles han sido los rendimientos históricos de cada cultivo?

La Figura 4.27 resume la **productividad** de los cultivos en cada temporada en kg/ha. Debido a la naturaleza de cada cultivo, su potencial de producción es distinto entre ellos; en otras palabras, el **rendimiento** de un cultivo no es comparable con el de otro. Por ejemplo, mientras que un **rendimiento** de 14,000 kg/ha en frambuesa es un resultado muy bueno, tener ese mismo **rendimiento** en fresa sería lamentable.

Además de calcular el **rendimiento**, se ha agregado un cálculo del promedio de **rendimiento**. Aunque lo ideal sería utilizar un **rendimiento** ponderado que tome en cuenta la superficie en producción, por fines prácticos se está usando un **rendimiento** aritmético tradicional. Por ejemplo, la fresa muestra un **rendimiento** promedio de 31,873 kg/ha.

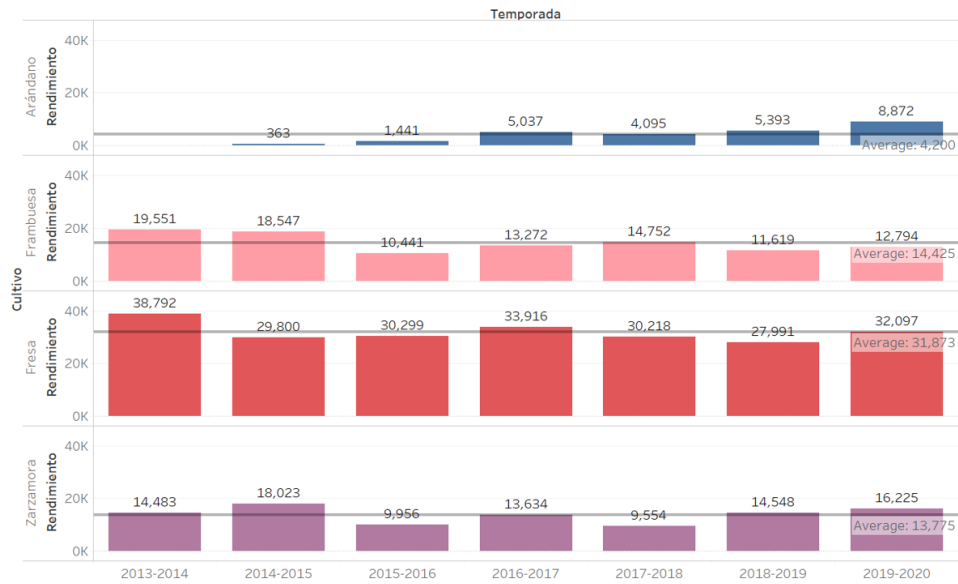


Figura 4.27: Rendimiento histórico por cultivo en kg/ha.

3.3.2. ¿Cuáles variedades han impactado el rendimiento promedio en fresa?

En el ejercicio anterior se pudo apreciar que 4 de las últimas 7 temporadas están por debajo del promedio en fresa. Resultaría importante entender aristas de esta situación; una de ellas es saber si es un efecto general de todas las variedades o si han sido una o más las que causan este efecto y se deba considerar retirarlas de programas futuros.

La Figura 4.28 permite identificar varios aspectos relevantes. Para facilitar la interpretación, se ha usado una gama de colores que permite identificar la combinación de variedad/temporada que han excedido el promedio (verde), las que están cercanas al promedio (gris) y las que han presentado un pobre desempeño (rojo). También se puede apreciar la superficie, la cual está ligada al tamaño del recuadro: entre más grande el recuadro, mayor superficie. Por cuestiones de confidencialidad, los nombres reales de variedades y ranchos en este documento serán sustituidos por otros términos. Algunos ejemplos de lo que se puede identificar visualmente:

- La variedad Var1 ha sido inconsistente, con temporadas malas, regulares y buenas.
- El comportamiento de Var3 ha sido bastante estable a través del tiempo.
- Variedades con un comportamiento pobre y constante como Var8 y Var10 se han retirado ya de los programas.
- En la temporada 2019-20 se introduce una nueva variedad Var12 con resultados por encima del promedio.

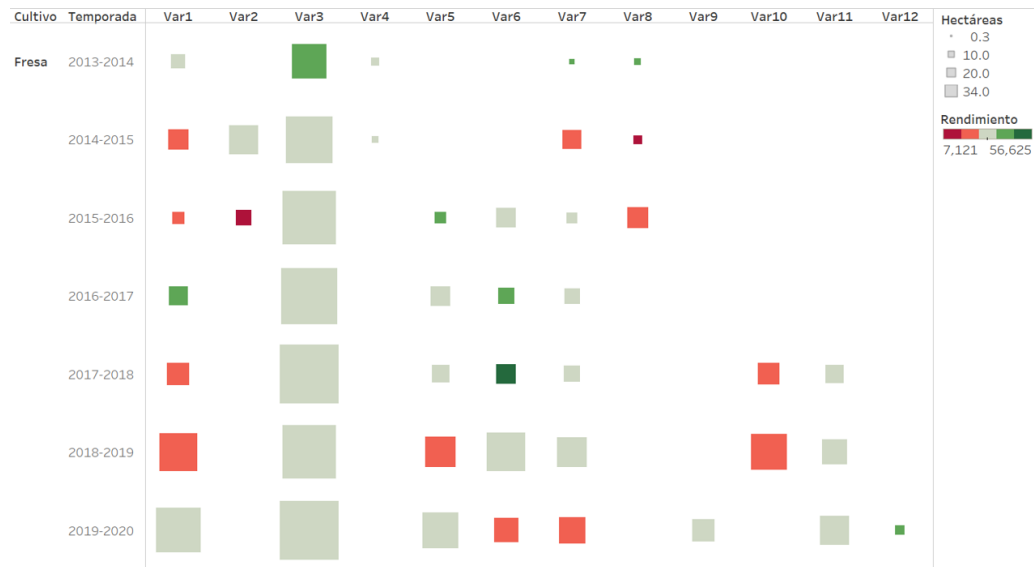


Figura 4.28: Rendimiento histórico por variedad de fresa en kg/ha.

3.3.3. ¿Cuáles ranchos han impactado el rendimiento promedio en zarzamora?

Otro enfoque válido es entender si los distintos ranchos de la [agrícola](#) han presentado resultados similares, o si hay una variabilidad importante entre ellos. La Figura 4.29 muestra los 5 ranchos de zarzamora que posee la [agrícola](#). Usando el [rendimiento](#) promedio de la zarzamora para la evaluación de los [rendimientos](#), se puede identificar lo siguiente:

- El Rancho1 tuvo 3 temporadas seguidas con resultados entre mediocres y malos, llevando a su eliminación al final de la temporada 2017-18.
- El Rancho2 es el más pequeño y con malos resultados consistentes, llevando a su eliminación al final de la temporada 2018-19. Se retomará este rancho más adelante para entenderlo a más detalle.
- El resto de los ranchos han mostrado una tendencia a la mejoría durante las últimas 3 temporadas.

3.3.4. ¿El caso del Rancho2 de zarzamora se debe a todas las variedades o a alguna en específico?

Cabe la posibilidad de que los malos resultados de un rancho se deban a que una de sus variedades tenga un mal desempeño. Hacer un análisis más a profundidad del rancho permite responder esta pregunta. En el caso específico del Rancho2, en la Figura 4.30 se observa que todas las variedades presentaron un mal desempeño, lo cual respalda la decisión de darlo de baja. ¿Cuáles son las razones del mal desempeño general? Las



Figura 4.29: Rendimiento histórico por rancho de zarzamora en kg/ha.

fuentes de datos disponibles no ayudan a dar respuesta, ya que se carece de datos que pudieran ayudar a identificar si se debe a uno o más de los siguientes factores: mala selección de la ubicación del campo, suelo inadecuado, mal manejo nutricional, o falta de experiencia de supervisión, entre otras. Este tipo de cuestiones evidencian la necesidad de complementar los DataFrames usados con otro tipo de registros y métricos que ayuden a dar una explicación y a encontrar propuestas de mejora.



Figura 4.30: Rendimiento histórico del Rancho2 en kg/ha.

3.3.5. Comentarios de cierre

Como se puede apreciar en los pasados ejemplos, las posibilidades de exploración y análisis de datos son ilimitadas y se podrían dedicar decenas de páginas a visualizar dis-

tintos enfoques de un mismo problema. Sin embargo, el propósito de este proyecto no es ahondar en todas las capacidades de tecnologías como Python ni Tableau, tampoco lo es adentrarse profundamente en la situación pasada y actual de la agrícola que amablemente proporcionó sus datos. La gran diversidad de productores de berries que existe en México tiene como consecuencia que también exista una gran gama de:

- Medios de captura de datos, desde manuales hasta uso de ERPs.
- Tipo de datos que se capturan. Superficie y volumen es lo mínimo, pero pueden existir otros datos como valores nutricionales, clima, precios, costos, entre otros.
- Calidad de datos. Hay quienes capturan datos solamente a nivel general del rancho, cuando también hay quien puede llegar a registrar datos de cada planta, variedad, sector o algún otro tipo de unidad.

Es debido a esta realidad que este capítulo ha sido dedicado a demostrar que es posible hacer análisis simples y avanzados en unidades de producción agrícola independientemente de su tamaño, medios de captura y demás variables mencionadas con anterioridad. En la siguiente sección de análisis de resultados, el enfoque será evaluar el potencial inmediato y futuro de estas tecnologías en el sector agrícola, basándonos en los ejemplos provistos en este capítulo.

Capítulo 5

Resultados

La información es el aceite del Siglo XXI, y la analítica es el motor de combustión.

— PETER SONDERGAARD, Vicepresidente y Líder Global de Investigación de Gartner (1965 – a la fecha)

1. Resultados

La metodología descrita en el capítulo anterior fue implementada en una [agrícola](#) de producción de [berries](#) de 160 hectáreas localizada en el municipio de Zamora, Michoacán y denominada como LRDA. Con el fin de demostrar que dicha metodología puede ser aplicada exitosamente en distintas aristas del negocio, se seleccionaron dos temas para su implementación: cosecha y mano de obra. Otro punto importante fue decidir no hacer una implementación directa con personal de experiencia en sistemas de información; en su lugar, se dio una capacitación a una persona que funje como Auxiliar de [Agronomía](#) con 1 año de experiencia profesional y que actualmente es responsable de la recolección de datos y generación de reportes de forma manual. Dicha capacitación tuvo una duración de 10 horas donde se revisaron ejemplos relacionados con la recolección de datos, así como su limpieza, procesamiento, análisis y [reporteo](#).

Actualmente, la auxiliar de producción de LRDA publica diferentes reportes de producción con el objetivo de controlar la calidad del producto, lo que la obliga a administrar diferentes tablas de datos. En sus palabras:

El tener que registrar diariamente datos que se convierten en información que debería ser certera y precisa resulta difícil, puesto que, al reportar la información de los diferentes huertos al Auxiliar de [Agronomía](#) suele ser tardada, desordenada en los diferentes formatos y hace difícil el identificar los hallazgos negativos que se están presentando en los huertos, esto conlleva a que la calidad del producto se vea afectada por la detección inoportuna del problema. Muchas veces el analizar la información tan importante que estos reportes

representan se vuelve imposible, ya que se pierde mucho tiempo en la alimentación de dichas bases de datos, a consecuencia del retardo de información, la falta de comunicación, el desconocimiento del grado de importancia que tienen estos datos no solo para el área de producción sino a las diferentes áreas que componen la organización. La desorganización del área de producción es evidente.

Previo a la capacitación, se midió el tiempo que llevaba el proceso existente para la generación de reportes estáticos, el cual tuvo una duración de 2 horas por tema, es decir, 4 horas diarias en total considerando lo relacionado a mano de obra y a cosecha. Tras la capacitación, se logró reducir el tiempo de captura de datos de manera considerable. A continuación se enlistan los principales resultados:

- Se redujo el tiempo dedicado a la generación de reportes de 4 horas diarias a 1 minuto con 30 segundos.
- Los reportes estáticos se sustituyeron con reportes dinámicos con capacidades de filtrado y minería de datos.
- Se generaron tableros de control para combinar distintas fuentes de información en un solo reporte interactivo.
- Se eliminaron las inconsistencias de información al hacer uso del proceso de limpieza y estandarización. Una de las situaciones a destacar durante este proceso es que dos de los ranchos de la agrícola habían estado presentando información errónea durante toda la temporada debido a errores humanos, por lo que los resultados de rendimiento no reflejaban la realidad. Al pasar por la limpieza y estandarización, el nuevo proceso identificó y corrigió los errores.
- La persona responsable de los reportes ahora cuenta con más tiempo disponible para recorrer campos y dedicarse a analizar los datos procesados para proponer acciones correctivas en tiempo.

Resultados			
Aspecto	Antes	Después	Puntos clave
Tiempo en generación de reportes	96 horas al mes	1 hora al mes	La limpieza de datos facilita el análisis gracias a la homologación de nomenclatura, se automatiza el proceso de transformación de datos, y los reportes gráficos se actualizan automáticamente con la información más reciente
Flexibilidad de los reportes	Reportes estáticos, limitados, mostrando una sola perspectiva de los datos	Reportes dinámicos con capacidades de filtración de datos y análisis más detallados cuando se desea entender un nivel más profundo de los datos	Reportes homologados, facilidad para evaluar un mismo reporte desde distintos filtros y/o dimensiones, e incorporación de tecnologías de análisis y visualización de datos
Tableros de control	Múltiples reportes, cada uno con un propósito específico	Reportes integrales mediante tableros de control	Unificación de múltiples aristas del negocio en un solo tablero de control, diseño de reportes integrales de múltiple propósito en lugar de reportes estáticos desechables
Integridad de la información	Datos erróneos, poca confiabilidad en la información, falta de procesos de comprobación	Datos homogéneos y revisados con alta confiabilidad	Los procesos de ETL estandarizan y homologuean los datos, y también identifican errores en la lógica de los datos, permitiendo su corrección previo a la publicación de reportes
Optimización de recursos	Una persona enfocando al menos el 50 % del tiempo de su jornada laboral en tareas repetitivas de recopilación de datos y publicación de reportes	La misma persona dedica ahora menos del 5 % de su tiempo a las mismas tareas de recopilación de datos y publicación de reportes, permitiéndole pasar más tiempo en análisis, corroboración, y generación de propuestas	Se realizó un análisis de los procesos que permita su automatización, desde la recopilación de datos hasta la publicación de reportes

2. Discusión

Retomando los objetivos de este proyecto, el enfoque desde un principio ha sido demostrar que las tecnologías de BI pueden ser aplicadas al campo de manera exitosa para llevar una mejor administración de los datos. Esto quedó demostrado al lograr capacitar de manera satisfactoria a una persona sin conocimientos previos de BI, administración de datos o herramientas de [reporteo](#). Sin embargo, una correcta implementación de capacidades de BI en una organización es un proceso que lleva tiempo y que es evolutivo, como se observa en la [Figura 5.1](#).

Futuros trabajos deben enfocarse en identificar la factibilidad de llevar el BI a otro nivel. Se recomienda no considerar a la empresa como un todo en el sentido de llevar a todos las áreas los procesos al mismo nivel, ya que la calidad de los datos y los procesos no tienen necesariamente la misma madurez. Por ejemplo, es posible que el departamento de [Agronomía](#) esté listo para reportes dinámicos, mientras que Finanzas solamente esté listo para reportes estáticos. Cada proceso debe evolucionar a su propio ritmo.

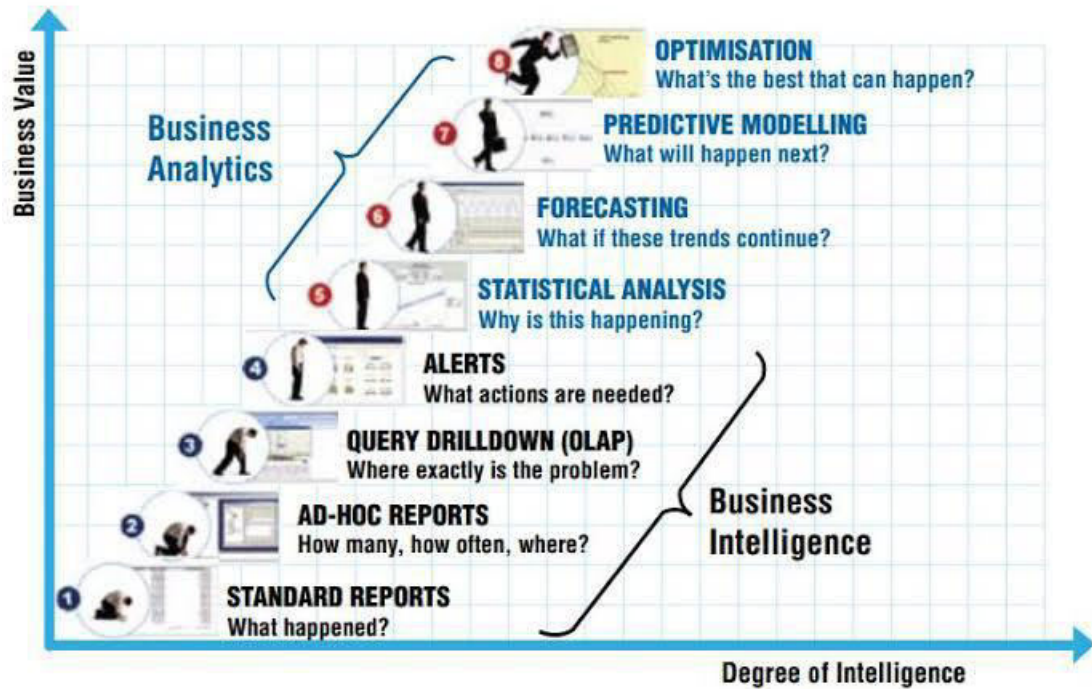


Figura 5.1: Evolución del BI.

Por último, aunque ha quedado demostrado que es posible capacitar personal para mejorar la calidad de la información y reducir los tiempos de procesamiento, se debe reconocer el valor que tiene para estas organizaciones contar con perfiles especializados. Personal con estudios de Informática o de Ciencia de Datos pueden acelerar la evolución de la adopción del BI, logrando llevarla a un nivel de evolución mayor, como la identificación de patrones ocultos, o desarrollo de modelos predictivos.

Capítulo 6

Conclusiones

Los datos se están convirtiendo en la nueva materia prima de los negocios.

— CRAIG MUNDIE, Asesor Senior del CEO de Microsoft y su ex Director de Investigación y Estrategia (1949 – a la fecha)

1. Conclusiones

Se ha demostrado con este proyecto que el **Business Intelligence** es aplicable para empresas agrícolas de distintas dimensiones. Por una parte, se logró exponer de manera teórica los beneficios que una empresa agrícola puede obtener mediante la implementación de paradigmas de análisis de datos ya existentes, pero que suelen ser usados en industrias con procesos más maduros y controlables. Por otro lado, se logró demostrar, a través de un prototipo y en escenarios reales, los beneficios que otorgan los procesos de obtención y limpieza de datos, el análisis e interpretación, y la publicación de información de valor a través de tecnologías de visualización de datos.

A través de los resultados preliminares obtenidos con la prueba que ya fue comentada en la sección de Resultados, se comprueba que las unidades de producción agrícola de **berries** en México pueden implementar de manera satisfactoria las metodologías de **Business Intelligence**. Esto queda demostrado al haber podido reducir drásticamente los tiempos de obtención de información, incrementado la calidad de los datos, y volviendo más confiable la toma de decisiones.

Por último, el análisis de datos que se puede apreciar a través de los reportes y tableros de control demuestra que es posible analizar datos históricos y usar estos mismos para definir el rumbo estratégico de la empresa dependiendo de los factores que se estén midiendo. Aunque en el prototipo se hayan usado variables básicas e indispensables como **superficie**, **volumen** y **rendimiento** a distintos niveles (zonas, variedades, entre otras), éstas no son las únicas variables existentes ni necesarias para la creación de modelos; sin embargo, incorporar cualquier variable adicional debe seguir el mismo proceso aquí descrito.

2. Trabajo futuro

A través de este proyecto se ha demostrado que es posible para las empresas agrícolas el incorporar procesos de **Business Intelligence** comenzando desde la obtención y limpieza de datos, y terminando en la publicación de reportes y tableros de control para la toma de decisiones. Esta información es de alto valor corporativo; sin embargo, la mejora constante que ofrecen otras tecnologías como el aprendizaje máquina, internet de las cosas, y aprendizaje profundo, deberán ser validados en posteriores estudios para comprender su impacto en el campo agrícola.

El proceso de desarrollo genético de nuevos materiales vegetales toma más de 7 años. En otras palabras, definir cuáles son las variedades de fresa que se deberán ofrecer en el mercado en el año 2028 es un trabajo que comienza hoy. Aunque esto se debe en parte a un tema de propagación, también es una realidad que este tiempo podría disminuir si se identifican patrones ocultos que ayuden a determinar si un nuevo material vegetal tiene una alta probabilidad de éxito al analizar parámetros como tamaño de raíz, longitud de los peciolo, o características específicas de la **fruta**. Esto ya se realiza a cierto nivel; sin embargo, el beneficio que pueden agregar las tecnologías ya mencionadas podría revelar información que no es perceptible para el ojo humano.

Pongamos como ejemplo un dato conocido en el cultivo de la fresa. Existe una correlación directa entre el grosor de los peciolo y el tamaño de la **fruta** que arroja; es decir, a mayor grosor del peciolo, mayor tamaño de **fruta**. Esto ayuda a poder hacer proyecciones de producción hasta 6 semanas antes de que la **fruta** sea visible en la planta. Si técnicas como el aprendizaje máquina pueden identificar relaciones similares, pero desde etapas tempranas del desarrollo del material vegetal, la industria sería capaz de acelerar el proceso de liberación de nuevas variedades que sean más productivas, nutritivas, rentables, y resistentes a **plagas** y enfermedades, dando así el soporte necesario para cumplir con los Objetivos de Desarrollo Sostenible planteados por la ONU.

Glosario

- agricultura** Conjunto de técnicas y conocimientos relativos al cultivo de la tierra. 1, 5, 7–9
- agricultura de precisión** Conjunto de tecnologías que se aplican al trabajo en el campo como satélites, sensores, imágenes y datos geográficos, que reúnen la información necesaria para entender las variaciones del suelo y los cultivos.. 11
- agronomía** Conjunto de conocimientos aplicables al cultivo de la tierra, derivados de las ciencias exactas, físicas y económicas.. 50, 52
- agrícola** Empresa cuyo negocio consiste en la producción de alimentos del campo.. 3, 5, 8, 27, 29, 30, 40, 42–44, 47, 49–51, 54, 55
- aronías** Género de plantas que pertenece a la familia de las rosáceas.. 9
- berries** Conocidas en español como bayas, son un tipo de fruto carnoso con semillas rodeadas de pulpa. En este documento hace referencia a la fresa, la frambuesa, la zarzamora, y el arándano azul.. 1–3, 6–11, 13, 27, 30, 31, 49, 50, 54
- Business Intelligence** Inteligencia de Negocios, o BI, es un término general que cubre los procesos y métodos de recopilación, almacenamiento y análisis de datos de operaciones o actividades comerciales para optimizar el rendimiento.. 2, 5, 13, 25, 28, 40, 54, 55
- cambio climático** Variación global del clima de la Tierra y que amenaza el medioambiente.. 2, 5
- CRM** Customer Relationship Management o Gestión de Relaciones con Clientes. Es un programa informático que permite organizar y gestionar los parámetros relacionados con los clientes y que reúne información de cada proceso de venta individual, desde la captación del cliente hasta el análisis de satisfacción.. 17, 31
- dashboard** Herramienta de gestión de la información que se permite monitorear, analizar y mostrar visualmente los indicadores clave de desempeño (KPI), métricos y datos fundamentales de una empresa, departamento, o proceso específico.. 24, 26, 44

- DataFrame** Estructura de datos con dos dimensiones en la cual se puede guardar datos de distintos tipos (como caracteres, enteros, valores de punto flotante, factores y más) en columnas.. 31, 33, 36–38, 40, 48
- Deep Learning** Tipo de machine learning que entrena a una computadora para que realice tareas como las hacemos los seres humanos, como el reconocimiento del habla, la identificación de imágenes o hacer predicciones.. 8
- ERP** Enterprise Resource Planning (ERP) o Planeación de Recursos Empresariales. Se refiere al tipo de software que usan las empresas para administrar sus actividades diarias, como la contabilidad, el suministro, la administración de proyectos, el cumplimiento de órdenes, y las operaciones de la cadena de suministro.. 3, 17, 31, 49
- ETL** Extract, Transform, and Load o Extraer, Transformar y Cargar. Son tres funciones de base de datos que se combinan en una herramienta para extraer datos de una fuente de datos, modificarlos y colocarlos en otra fuente de datos.. 17, 21, 22
- fruta** Fruto comestible de ciertas plantas cultivadas; p. ej., la pera, la manzana, la fresa, etc.. 2, 10, 12, 30, 55
- grosella** Fruto del grosellero, que es una uva o baya globosa de color rojo, blanco o negro, jugosa y de sabor agrídulce muy grato. Su jugo es medicinal, y suele usarse en bebidas y en jalea.. 9
- herbáceo** Que tiene la naturaleza o cualidades de la hierba.. 12
- Machine Learning** Disciplina científica del ámbito de la Inteligencia Artificial que crea sistemas que aprenden automáticamente.. 5, 8
- macrotúnel** Estructuras construidas para la protección de las plantas a fin de generar mejores condiciones para el desarrollo de éstas.. 8, 11
- MDX** MultiDimensional eXpressions o Expresiones Multidimensionales. Son un lenguaje de consulta para bases de datos multidimensionales sobre cubos OLAP.. 25
- Odoo** Conjunto de aplicaciones de código abierto que cubren diversas necesidades de las organizaciones como CRM, comercio electrónico, contabilidad, inventario, punto de venta, gestión de proyectos, entre otros.. 8
- plaga** Aparición masiva y repentina de seres vivos de la misma especie que causan graves daños a poblaciones animales o vegetales, como, respectivamente, la peste bubónica y la filoxera.. 2, 7, 8, 10, 30, 55
- plataforma** Entorno informático determinado, que utiliza sistemas compatibles entre sí.. 2, 3, 7, 8, 28, 31, 40

- plug** Conocido también como cepellón, es una pella de tierra que se deja adherida a las raíces de los vegetales para trasplantarlos.. 12, 13
- productividad** Capacidad o grado de producción por unidad de trabajo, superficie de tierra cultivada, equipo industrial, etc.. 5, 45
- Redes Neuronales** Modelo simplificado que emula el modo en que el cerebro humano procesa la información.. 8
- rendimiento** Proporción entre la producción y la superficie. Típicamente medida en kilogramos por hectárea, cajas por hectárea, o libras por acre.. 2, 3, 5, 7–9, 12, 13, 29, 42, 45, 47, 51, 54
- rentabilidad** Capacidad que tiene una empresa para aprovechar sus recursos y generar ganancias o utilidades. En las operaciones agrícolas, esta rentabilidad se basa en el costo de inversión y el costo de cosecha.. 1–3, 9, 35
- reporteo** Conjugación de reportar. Trata de las acciones orientadas a generar información de valor en forma de reporte.. 18, 40, 50, 52
- RFID** Radio Frequency Identification o Identificación por Radiofrecuencia. Es un sistema de almacenamiento y recuperación de datos remotos que permiten transmitir la identidad de un objeto por medio de ondas de radio.. 17, 31
- SaaS** Software como Servicio (SaaS) es un modelo de entrega de software basado en la nube en el que el proveedor desarrolla y mantiene el software, proporciona actualizaciones automáticas y lo pone a disposición de sus clientes a través de Internet con un sistema de pago por uso.. 8
- saskatun** Especie originaria de América del Norte, donde también es conocida como fresa de junio occidental, que se aprovecha en jardinería como arbusto o árbol ornamental.. 9
- saúco** Arbusto o arbolillo de la familia de las caprifoliáceas, con tronco de dos a cinco metros de altura, lleno de ramas, de corteza parda y rugosa y médula blanca abundante, hojas compuestas de cinco a siete hojuelas ovales, de punta aguda, aserradas por el margen, de color verde oscuro, de olor desagradable y sabor acre, flores blancas y fruto en bayas negruzcas.. 9
- SCM** Supply Chain Management o Administración de la Cadena de Suministro. Es el proceso por el cual se planifica, controlan y ponen en funcionamiento las operaciones de una red de suministro y almacenaje como proveedores o vendedores.. 17, 31
- SQL** Structured Query Language o Lenguaje de Consulta Estructurado. Es un lenguaje de computación para trabajar con conjuntos de datos y las relaciones entre ellos.. 22

- suite** Conjunto de varias aplicaciones de software diseñadas como herramientas de una misma paquetería.. 8
- superficie** Extensión de tierra, típicamente medida en hectáreas o en acres según el país.. 5-7, 29-31, 35-38, 40-42, 44-46, 49, 54
- Tableau** Herramienta de Inteligencia de Negocios que permite analizar, visualizar y compartir grandes volúmenes de información en forma rápida, flexible y amigable.. 22-28, 40-42, 44, 49
- vida de anaquel** Periodo de tiempo en el cual el alimento conserva los atributos esperados por el consumidor y es el momento adecuado para comercializarlo.. 10, 13
- XML** Extensible Markup Language o Lenguaje de Marcado. Define un conjunto de reglas para la codificación de documentos.. 22

Referencias

- [1] SIAP, “Atlas agroalimentario 2019.pdf.”
- [2] J. C. Huerta, “Las berries ya son el tercer producto agrícola de México más exportado,” library Catalog: www.elfinanciero.com.mx. [Online]. Available: <https://elfinanciero.com.mx/economia/las-berries-se-colocan-como-el-tercer-producto-agricola-mexicano-mas-exportado-al-extranjero>
- [3] FAO, “La agricultura mundial en la perspectiva del año 2050,” p. 28.
- [4] —, “FAO y los ODS.” [Online]. Available: <http://www.fao.org/3/a-i6919s.pdf>
- [5] —, “TRANSFORMAR LA ALIMENTACIÓN Y LA AGRICULTURA PARA ALCANZAR LOS ODS.” [Online]. Available: <http://www.fao.org/3/i9900es/I9900ES.PDF>
- [6] La asamblea general adopta la agenda 2030 para el desarrollo sostenible. Library Catalog: www.un.org Section: Destacadas - Cumbre. [Online]. Available: <https://www.un.org/sustainabledevelopment/es/2015/09/la-asamblea-general-adopta-la-agenda-2030-para-el-desarrollo-sostenible/>
- [7] M. Moran. Infraestructura. Library Catalog: www.un.org. [Online]. Available: <https://www.un.org/sustainabledevelopment/es/infrastructure/>
- [8] A. S. Baldivia, “La disponibilidad de alimentos en México: un análisis de la producción agrícola de 35 años y su proyección para 2050,” vol. 23, no. 93. [Online]. Available: <https://www.redalyc.org/jatsRepo/112/11252977008/html/index.html>
- [9] V. Lakshmi, J. Corbett, and J. Corbett, “How artificial intelligence improves agricultural productivity and sustainability: A global thematic analysis,” p. 10.
- [10] W. E. Forum. Big data, big impact: New possibilities for international development. [Online]. Available: http://www3.weforum.org/docs/WEF_TC_MFS_BigDataBigImpact_Briefing_2012.pdf
- [11] T. Guo and Y. Wang, “Big data application issues in the agricultural modernization of China,” vol. 28, no. 107, pp. 3677–3688.
- [12] J. Tyrychtr, M. Ulman, and V. Vostrovský, “Evaluation of the state of the business intelligence among small Czech farms,” vol. 61, pp. 63–71. [Online]. Available: <http://www.agriculturejournals.cz/web/agricecon.htm?volume=61&firstPage=63&type=publishedArticle>
- [13] C. Parra Montalvo and F. Vázquez Torres, “Las pymes y business intelligence en México,” no. 69. [Online]. Available: <http://www.boletin.upiita.ipn.mx/index.php/ciencia/791-cyt-numero-69/1604-las-pymes-y-business-intelligence-en-mexico>
- [14] FIRA, “Panorama agroalimentario: Berries 2016.” [Online]. Available: https://www.gob.mx/cms/uploads/attachment/file/200633/Panorama_Agroalimentario_Berries_2016.pdf
- [15] Infoagro. Las berries mexicanas y el secreto del éxito, en entrevista con Salvador Álvarez Barboza. | revista infoagro México. Library Catalog: mexico.infoagro.com. [Online]. Available: <https://mexico.infoagro.com/las-berries-mexicanas-y-el-secreto-del-exito-en-entrevista-con-salvador-alvarez-barboza/>
- [16] SAGARPA, “Planeación agrícola nacional 2017-2030: Fresa mexicana.”
- [17] Agriculture decision support system, agricultural DSS, precision farming in India. [Online]. Available: <http://sxagro.com/home/what-is-SXagro.aspx#container>
- [18] Prospera technologies. Library Catalog: home.prospera.ag. [Online]. Available: <https://home.prospera.ag/>
- [19] rNet. rNet | soluciones informáticas. [Online]. Available: <https://rnet.mx/>

- [20] F. d. J. G. Razo, S. R. Rebollar, J. H. Martínez, J. L. M. Hernández, and O. R. Abarca, "SITUACION ACTUAL y PERSPECTIVAS DE LA PRODUCCIÓN DE BERRIES EN MÉXICO," vol. 44, pp. 260–272. [Online]. Available: <https://www.redalyc.org/jatsRepo/141/14161295012/html/index.html>
- [21] Y. Zhao, *Berry Fruit: Value-Added Products for Health Promotion*. Taylor & Francis Ltd., OCLC: 476084443. [Online]. Available: <http://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=305945>
- [22] INCyTU, "Agricultura de precisión," no. 15.
- [23] M. Liotta, "Manual de capacitación: riego por goteo."
- [24] C. Pernuzzi, M. Sordo, M. Travadelo, M. Maina, and A. Pato, "Evaluación de la conveniencia de los macrotúneles en comparación con microtúneles para el cultivo de frutilla en corona," vol. 15, no. 51.
- [25] SAGARPA, "Planeación agrícola nacional 2017-2030: Frutas del bosque mexicanas." [Online]. Available: https://www.gob.mx/cms/uploads/attachment/file/257076/Potencial-Frutas_del_Bosque.pdf
- [26] R. Silva Murillo, "Los sistemas de información como arma estratégica en la gestión empresarial," *ñolPERSPECTIVAS*, 2006. [Online]. Available: <https://www.redalyc.org/articulo.oa?id=425942516006>
- [27] J. Curto Díaz and J. Conesa Caralt, *Introducción al Business Intelligence*. Editorial UOC, OCLC: 688377734.
- [28] B. Evelson, "The forrester wave: Enterprise business intelligence platforms, q3 2008."
- [29] J. L. R. Salgueiro, G. C. Carrión, and J. L. G. González, "LOS SISTEMAS DE INTELIGENCIA DE NEGOCIO COMO SOPORTE a LOS PROCESOS DE TOMA DE DECISIONES EN LAS ORGANIZACIONES," p. 22.
- [30] R. Rad, *Microsoft SQL Server 2014 Business Intelligence development beginner's Guide: get to grips with Microsoft Business Intelligence and data warehousing technologies using this practical guide*. Packt Publ. [u.a.], OCLC: 908612027.
- [31] J. L. Cano, "BUSINESS INTELLIGENCE: COMPETIR CON INFORMACIÓN."
- [32] S. Ramos, "Microsoft business intelligence: vea el cubo medio lleno," p. 129.
- [33] R. Kimball and M. Ross, *The data warehouse toolkit: the definitive guide to dimensional modeling*, third edition ed. John Wiley & Sons, Inc, OCLC: ocn840431951.