



Universidad Nacional

SAN LUIS GONZAGA



Atribución-NoComercial-SinDerivadas 4.0 Internacional

Esta licencia es la más restrictiva de las seis licencias principales Creative Commons, permitiendo a otras solo descargar sus obras y compartirlas con otras siempre y cuando den crédito, pero no pueden cambiarlas de forma alguna ni usarlas de forma comercial.

<http://creativecommons.org/licenses/by-nc-nd/4.0>



UNIVERSIDAD NACIONAL "SAN LUIS GONZAGA"
FACULTAD DE INGENIERÍA DE SISTEMAS
DIRECCIÓN DE INVESTIGACIÓN
EVALUACIÓN DE ORIGINALIDAD



CONSTANCIA

El que suscribe, deja constancia que se ha realizado el análisis con el software de verificación de similitud al documento cuyo título es:

EFICIENCIA DE LA MINERÍA DE DATOS EN EL CONTROL DE JABAS DE LA EMPRESA VANGUARD PERU - ICA PERIODO 2020-2021

Presentado por:

- **GUERRA VERA LUIS PIERRE DE JESUS**

BACHILLER en PREGRADO de la facultad de Ingeniería de Sistemas. El resultado obtenido es (**porcentaje de similitud 0%**) por el cual se otorga el calificativo de:

APROBADO, según el Reglamento de Evaluación de la Originalidad.

Se adjunta al presente el reporte de evaluación con el software de verificación de originalidad.

Ica, 03 de setiembre del 2024


Dr. JAVIER ORLANDO GUTIÉRREZ FERREYRA
Director de la Unidad de Investigación
Facultad de Ingeniería de Sistemas

UNIVERSIDAD NACIONAL “SAN LUIS GONZAGA”

VICERRECTORADO DE INVESTIGACIÓN

FACULTAD INGENIERÍA DE SISTEMAS



EFICIENCIA DE LA MINERÍA DE DATOS EN EL CONTROL DE
JABAS DE LA EMPRESA VANGUARD PERU - ICA PERIODO 2020-
2021

Línea de investigación: Ciencias naturales, ingeniería y tecnologías sostenibles

INFORME FINAL DE TESIS PARA OPTAR EL TITULO
PROFESIONAL DE INGENIERO DE SISTEMAS

Autor: LUIS PIERRE DE JESUS GUERRA VERA

Ica, Perú

2023

DEDICATORIA

A Dios quien fue la luz en mi camino, me guio en el sendero de la vida y me acompaña siempre en las buenas y malas con su inmenso amor.

A mi madre Diana, que nunca me dejo de lado y siempre estuvo con ahínco de verme mejorar y a mi padre Luis, ambos que con su apoyo incondicional y la motivación que me dieron pude lograr todo los retos y metas que me propuse.

De Igual Manera Quiero dedicar esta tesis a mis hermanas, que siempre están cuando las necesito, el cariño y apoyo brindado a mi persona para nunca alejarme de mis metas y objetivos.

AGRADECIMIENTO

Agradecer a Dios, por estar siempre presente sobre todo en los momentos mas complicados de mi vida, llenando de gracia y amor mi camino.

Agradecer a Mi familia, por el apoyo, motivación y esfuerzo en todo este tiempo.

Agradecer a la plana docente y administrativa de la Universidad Nacional San Luis Gonzaga de Ica, sobre todo a la Facultad de Ingeniería de sistemas la cual pertenezco por guiarme y apoyarme en el camino de la obtención de este grado académico, a todos los docentes por su enseñanza y tiempo brindado.

Por último y no menos importante, Agradecer al Dr. Erwin Peña por la asesoría, dirección y conocimientos brindados, aportado para la realización de este Proyecto de Tesis.

INDICE DE CONTENIDOS

Dedicatoria	ii
Agradecimientos	iii
Índice de contenidos	iv
Índice de tablas	v
Índice de figuras	vi
Resumen	vii
Abstract	viii
I. Introducción	1
2.1. Realidad problemática	3
2.2. Antecedentes	5
2.3. Justificación	9
2.4. Objetivos de la investigación	10
II. Estrategia metodológica	11
2.1. Método	12
2.2. Aplicación del método	12
III. Resultados	23
3.1. Aplicación del método	23
3.2. Calcular la transferencia de las jabas	30
3.3. Controlar las jabas	35
IV. Discusión	40
V. Conclusiones	41
VI. Recomendaciones	42
VII. Referencias bibliográficas	43
VIII. Anexos	45
Anexo 1. Certificado de trabajo empresa Vanguard	46
Anexo 2 Fondos grupo Vanguard Perú	47
Anexo 3 Jabas vacías	47

Índice de tablas

TABLA I: METADATOS DEL CONTROL DE JABAS

14

Índice de figuras

Fig. 1. Países donde Vanguard tiene sus sedes con agricultores de las zonas	4
Fig. 2. Variedad de vegetales y frutas que se exportan	4
Fig. 3. Estructura de los datos iniciales de las jabas	13
Fig. 4. Estructura de los datos finales de las jabas	13
Fig. 5. Separación de la Base de Datos	15
Fig. 6. Opción de copiar hoja en nuevo libro	15
Fig.7. Existencia de datos vacíos	16
Fig. 8. Plataforma Google colab	16
Fig. 9. Carga de librerías y carga de los datos	17
Fig. 10. Información de la metadata del archivo	17
Fig. 11. Datos de la Base de datos	18
Fig. 12. Visualización de los primeros y últimos registros	18
Fig. 13. Nuevo dataframe	19
Fig. 14. dataframe resultante	19
Fig. 15. Datos nulos o faltantes	20
Fig. 16. Información del nuevo dataframe Jabas1	20
Fig. 17. dataframe con valores nulos	21
Fig. 18. dataframe con valores nulos	21
Fig. 19. Error de datos detectado por la librería	21
Fig. 20. Errores detectados en la columna JABAS POR COLOR	22
Fig. 21. Estadística descriptiva	23
Fig. 22. Gráfico de traslado de Jabas por Semana	24
Fig. 23. Gráfico de traslado de Jabas desde el Fundo Origen	25
Fig. 24. Gráfico de traslado de Jabas al Fundo de destino	26
Fig. 25. Gráfico de traslado de Jabas según el tipo de Jaba	27
Fig. 26. Gráfico de traslado de Jabas por color	28
Fig. 27. Gráfico de traslado de Jabas según la observación	29
Fig. 28. Reporte generado por la librería	30
Fig. 29. Estructura del reporte generado por la librería	31

Fig. 30. Estadísticas relacionadas con el dataframe (Dataset)	31
Fig. 31. Estadística de la variable Fondo Origen	32
Fig. 32. Frecuencia de las Jabas despachadas	32
Fig. 33. Jabas despachadas	33
Fig. 34. Tipos de Jabas	33
Fig. 35. Jabas según color	34
Fig. 36. Jabas según la observación	34
Fig. 37. Reporte generado por la librería dtale	36
Fig. 38. Selección de la columna Tipo jaba	37
Fig. 40. Opciones de la descripción de la columna Tipo Jaba	37
Fig. 41. Ventana para análisis de las variables o datos	38
Fig. 42. Gráfico de frecuencias y KDE	39
Fig. 43. Distribución de Jabas a cada uno de los Fondos	39

RESUMEN

En el presente trabajo de tesis, cuyo objetivo fue Evaluar el control de Jabas mediante minería de datos en la empresa VANGUARD PERÚ-Ica en el periodo 2020-2021. La metodología empleada se basa en la metodología CRISP-DM Methodology, para el estudio descriptivo del desplazamiento de Jabas entre los Fondos de la empresa. Se facilitó el estudio utilizando herramientas disponibles para la minería de datos como la plataforma del Google Colab en la nube, el lenguaje Python y las librerías para la evaluación de los datos. Librería Pandas para el manejo de los datos con la creación de un DataFrame, librería gráfica interactiva Plotly y librerías para la evaluación y análisis de datos y Data-Profiling, DTable y Sweetviz cuya aplicación requiere de pocas líneas de código. Los resultados han sido satisfactorios se ha evaluado y calculado el movimiento de las Jabas en la empresa, se han creado gráficas de frecuencia en el cálculo, y tabla de evaluación gráfica interactiva con las librerías. El trabajo concluye en la importancia de la minería de datos con una eficiencia muy grande en el control de Jabas de la empresa VANGUARD PERÚ.

Palabras clave: Minería de datos, Jabas, Librerías

ABSTRACT

In this thesis work, whose objective was to evaluate the control of Jabas through data mining in the company VANGUARD PERÚ-Ica in the period 2020-2021. The methodology used is based on the CRISP-DM Methodology, for the descriptive study of the displacement of Jabas between the company's Fundos. The study was facilitated using available tools for data mining such as the Google Colab platform in the cloud, the Python language and libraries for data evaluation: Pandas library for data management with the creation of a DataFrame, plotly interactive graphic library and libraries for data evaluation and analysis and data-profiling, dtale and sweetviz whose application requires only a few lines of code. The results have been satisfactory, the movement of the Jabas in the company has been evaluated and calculated, frequency graphs have been created in the calculation, and interactive graphical evaluation table with the libraries. The work concludes in the importance of data mining with a very great efficiency in the control of Jabas of the company VANGUARD PERU.

Key words: Data mining, Jabas, Libraries.

I. INTRODUCCIÓN

Los adelantos tecnológicos de las últimas décadas han ocasionado que se almacenen grandes cantidades de datos en las industrias de todo tipo, y estas han ido de la mano del desarrollo de software de computación que permiten el análisis de los datos con la finalidad de obtener información útil para la toma de decisiones. Diversas son las investigaciones que se están realizando relacionadas con la Minería de Datos, con la finalidad de analizar patrones y relaciones de las variables para tener información sensible para el control de procesos, diagnóstico de errores y [1]. En ese mismo horizonte el almacenamiento de los datos hoy se concibe como una tarea rutinaria en las empresas apoyados en los sistemas de información, por lo cual la Minería de Datos ha cobrado mayor relevancia en diversas industrias y empresas académicas [2].

Las organizaciones hoy consideran a los datos como el nuevo activo o tesoro por lo valioso que oculta en la información sobre los clientes y sus procesos internos, además se consideran la memoria de la institución, y en la cual la Minería de Datos se convierte en el gestor de nuevos descubrimientos de patrones y tendencias que son obtenidas al examinar grandes volúmenes de información [3].

La inteligencia artificial en la actualidad esta haciendo posible que se puedan generar el análisis y exploración de los datos que se están almacenando en las instituciones del sector público, y que con su poder de computo permita el manejo de enormes cantidades de información que están en las bases de datos de estas. Con esta es importante debido a la disminución de los tiempos de análisis, siendo mas eficientes para este análisis e interpretación, pero además que con esta se puede obtener información que no es posible su visualización a simple vista, porque estas información esta escondida en los datos [4].

La minería de datos según [5], es un campo relativamente nuevo que emplea una serie de técnicas las mismas que se clasifican en dos grandes campos; por un lado el campo de la minería de datos para la predicción de comportamientos y el campo de la minería descriptiva. En ese mismo contexto [5], relacionado con las técnicas descriptivas especifica que no necesitan de modelos, ya que con estas

técnicas solo se describe el comportamiento de estos datos, además de las relaciones que pudieran existir.

En tal sentido [6], la Minería de Datos ha sido desarrollada para descubrir patrones en los datos, de grandes bases de datos y tener información útil que usualmente no esta visible o se encuentra oculta. A medida que avanza el desarrollo tecnológico se van desarrollando nuevas aplicaciones para los diversos ámbitos de las industrias e instituciones. Por lo cual según [7], es de suma importancia que se comprenda como la Minería de Datos puede identificar y reducir las desviaciones y conductas extrañas en los datos materia de estudio.

A pesar del poder de la Minería de Datos en la extracción de datos valiosos de las fuentes de datos, los datos usualmente no se encuentran debidamente preparadas para realizar los procesos de los algoritmos y librerías de la Minería de Datos, motivo por el cual estos datos deben pasar por un procesamiento previo para tener los datos específicos para el análisis, en este sentido según [8], la preparación de los datos incluye la detección y eliminación de los valores atípicos, ya que estos valores presentan comportamientos diferentes al de los datos generales, y que afectan al análisis de los datos mayoritarios.

Las técnicas y algoritmos de la Minería de Datos es muy extensa y se pueden aplicar en todos los ámbitos de la industria y la academia, algunos estudios como el de [9], en el campo de la educación sobre las pruebas SaberPro en la que utilizan la Minería de Datos para analizar los diversos factores relacionados con el desempeño de los estudiantes, y de la que se obtiene conocimiento de las fuentes de datos. similarmente en el mismo campo de la educación en el estudio de [10], en la que la minería de los datos se está empleando para desarrollar estrategias tecnológicas que permitan establecer un seguimiento a los estudiantes que presentan rezago académico y que son proclives a que puedan desertar de las aulas.

En otros campos como la agricultura la Minería de datos se esta utilizando para el análisis de las variables meteorológicas como la temperatura por medio de técnicas de agrupamiento [11].

En nuestra región a pesar del auge y crecimiento que tiene la agroindustria, no se están aplicando estas técnicas de la Minería de Datos, a pesar de que se están generando día a día grandes cantidades de datos del agro como la producción, los despachos, las ventas, la conservación de los productos etc. Pero que no se esta aprovechando esa enorme cantidad de datos generados. Sin embargo, aprovechando la labor desempeñada por la tesista relacionada con el control de las Jabas, es que se planteó el presente trabajo de tesis, como un aporte al conocimiento y a la evaluación de las técnicas de la Minería de Datos para un caso real.

Y en ese entender con la Minería de Datos se puede extraer conocimiento de los datos con los cuales se están produciendo enormes contribuciones en las organizaciones para tomar mejores decisiones [12].

De todo lo realizado en las diversas investigaciones, se puede conceptualizar que el término Minería de Datos es utilizado mayormente para describirlo como un proceso que emplea técnicas y herramientas investigativas, que permiten obtener información de utilidad para las personas, esta información es extraída de las base de datos de la empresa [1].

1.1 Realidad problemática

La gran generación de datos en las empresas son una realidad, los datos son el principal activo de las empresas. Mucha información, pero poco conocimiento siendo este un problema de muchas de las empresas en el mundo, ante esta situación la minería de datos se convierte en una herramienta de suma importancia para poder convertir la información en conocimiento con la finalidad de poder tomar decisiones acertadas en base a conocimientos y no a percepciones. En la Región Ica existen más de 200 empresas agroindustriales, que exportan muchos productos al mundo como la uva, los espárragos, arándanos y muchos más. Para poder gestionar internamente estos productos las empresas agroindustriales cuentan con envase para el transporte y almacenamiento denominadas Jabas. La empresa VANGUARD PERÚ que cuenta con muchas hectáreas dispone de una cantidad muy variada de Jabas que constantemente están transportando los productos entre los diversos campos de cultivo. Tener un control sobre estas jabas es una oportunidad que se presenta para lograr extraer de esa información el conocimiento para el adecuado control de estos recursos.

La empresa tiene establecida como misión “Traer a las mesas de todo el mundo los vegetales más frescos, saludables y deliciosos. Es una determinación impulsada por la pasión y que proviene de granjas y campos ubicados estratégicamente en las regiones de cultivo más fértiles de la Tierra”. La misma que se encuentra declarada en su portal web <https://www.vanguardteam.com/es/reach>. Esta empresa tiene presencia en 9 países para la provisión de los vegetales que cultiva con un nivel de calidad superior, como se muestra en la Fig.1

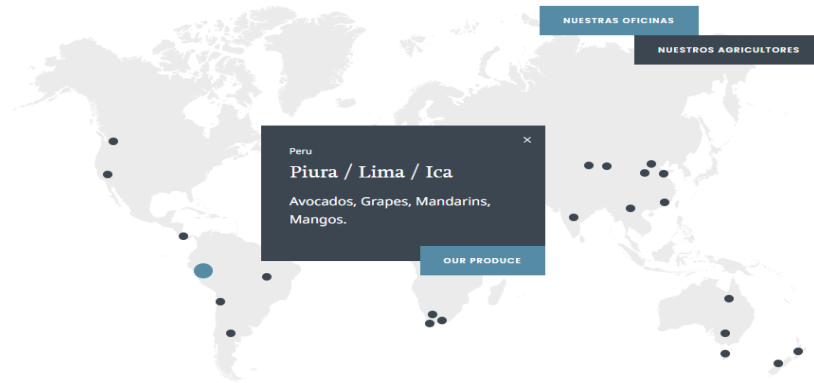


Fig. 1. Países donde Vanguard tiene sus sedes con agricultores de las zonas
 En tener una presencia en diversas partes del mundo le dan a la empresa Vanguard poder cumplir con la entrega precisa de los vegetales que se cultivan. En el caso de nuestro país cuenta con presencia en Piura, Lima e Ica. La empresa exporta a los diversos países del mundo una gran variedad de vegetales que se presentan en la Fig. 2.

	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
APPLES	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ASPARAGUS	✓		✓	✓	✓			✓	✓		✓	✓
AVOCADOS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
BERRIES	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓
BROCCOLI		✓	✓	✓	✓							
CARROTS	✓	✓	✓	✓	✓	✓				✓	✓	✓
CHERRIES	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓
FIGS								✓	✓	✓	✓	
GARLIC	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
GINGER	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
GRAPEFRUITS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
GRAPES	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
KIWIFRUITS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
KUMQUATS	✓	✓	✓							✓	✓	✓
LEMONS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
LETTUCE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
LIMES	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
MANDARINS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
MANGOES	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
MELONS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ONIONS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
ORANGES	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
PEARS	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
PINEAPPLES	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
POMEGRANATES	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓
SQUASHES	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
STONE FRUITS	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓

Fig. 2. Variedad de vegetales y frutas que se exportan

La Fig. 2, muestra la variedad de frutas y vegetales que la empresa produce, a lo largo de todo el año, y que para el traslado o distribución entre los diversos campos de cultivos que tienen la empresa en nuestra ciudad, requiere del uso de jabas para esto.

Como se puede apreciar en las Figuras 3, 4 de los anexos, la empresa Vanguard Perú de la provincia de Ica lleva el control de estos recursos por medio de hojas de cálculo de EXCEL, que a pesar de estar registradas no ofrecen una buena opción para poder obtener conocimiento sobre la distribución, ubicación de esta Jabas.

El manejo de información relacionada con las Jabas se realizan en cuadros Excel, ERP Nisira y Agro +(Que es el sistema que se usa en campo para el control de asistencia, salidas de camiones de fruta, y producción, que es de donde se jala la base de datos proveniente de los tinglados en el campo, La función era ordenar los datos y saber que cantidad de jabas estaban circulando en los diferentes fundos. Habiendo 5 Milagritos, challapampa, los laureles, las casuarinas y el arenal los cuales todos llegan y se procesan en PACKING que es donde llegan los camiones de jabas de Uva y a la vez se devuelven con jabas Vacías para su distribución, se manejan diversos colores de JABAS según el tipo de uva que se solicita, y según la información en vivo que se tiene de los almacenes se distribuye estas jabas, lo cual se trató de gestionar un Excel donde nos muestre que cantidad de jabas se tiene en cada fundo y almacén packing central.

Como Gestor de información del área de post cosecha, Localizada dentro del PACKING de la empresa, siguiendo el la Jerarquía está el jefe de Operaciones de post cosecha y por último el área de gerencia de producción donde se presentaba los cuadros de reporte, tanto diario semanal como mensual de las salidas y entradas diarias.

Como necesidad inmediata y según los conocimientos se manejó y creo un cuadro Excel brindado para dar salida a los pedidos de la empresa.

Esta necesidad en su momento ha generado gran cantidad de datos, motivo por el cual se plantea el estudio de tesis, que como autor veo la oportunidad de poder mejorar el control de estas jabas para la empresa Vanguard.

1.2 Antecedentes

En otra orientación, la investigación igualmente sobre Minería de datos sobre Rendimiento Académico de estudiantes de Ingeniería Agroindustrial en Nicaragua. La investigación se desarrolló con tipo de investigación cualitativo y cuantitativo. Se realizaron diversas pruebas como el análisis estadístico correlacional, univariado y multivariado, modelación estadística avanzada, los datos se obtuvieron de dos fuentes. Por un lado, los datos obtenidos de encuestas

y en segundo lugar los datos académicos registrados en los sistemas informáticos de la universidad. Con los resultados de la Minería de Datos se recomendó la implantación de nuevas estrategias de aprendizaje, la formación continua de los profesores, la reestructuración del currículo y normas académicas, además de nuevas políticas de género y un control del rendimiento académico por medio de esta herramienta [13].

En la investigación sobre Minería de Datos cuyo propósito es extraer información oculta y revelar patrones y conocimientos importantes desde grandes volúmenes de datos. Estos datos pueden provenir de diversas fuentes como BD, Data Warehouse, Portales Web u otras fuentes. En ese sentido las tecnologías como la IoT o Internet de las cosas, es una importante herramienta para obtener datos a gran escala. Como tal existe una relación muy cercana entre estas dos tecnologías, fortalezas que se complementan y se aplican en sectores tan importante como la agricultura inteligente o precisión e investigación agrícola. Estos campos aprovechan estas tecnologías para el estudio de la biodiversidad en artrópodos en los agroecosistemas. La Minería de datos se convierte en una posibilidad para determinar patrones en esta línea y determinar el comportamiento de los artrópodos, con factores abióticos para esta área de estudio [14].

En esta investigación se muestran derivaciones de la investigación cuyo objetivo fue descubrir los factores asociados a la deserción estudiantil utilizando una fuente de datos en la que se procesan aspectos socioeconómicos, académicos, institucionales y otros más de los sistemas de la universidad Nariño sede Pasto, de la carrera de ciencias naturales. Con el tratamiento de la Minería de Datos y el análisis estadístico, utilizando la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining), muy utilizado como un estándar en los proyectos utilizados. La investigación utilizó dos repositorios uno con un periodo de procesamiento de 6 años, y el otro entre el 2006 y 2015 de la información histórica académica. Se aplicó las fases de la metodología, realizando una primera fase de limpieza y transformación o preprocesamiento de los datos. La investigación se realizó con el software weka versión 3.8.2 y los algoritmos de Redes bayesianas, árboles de decisión, J48, y LMT (Logistic Model Tree), los mejores resultados obtenidos con J48 con un promedio de notas bajo, la pérdida de materias de solo 1, fueron los patrones detectados por la minería de datos. Con este conocimiento se tiene la posibilidad de hacer una toma de decisiones más precisa en las directivas de la institución y en la reformulación de los planes académicos para reducir la deserción, además de la mejora en la calidad educativa[15].

Investigaciones como la presente en el empleo de técnicas en la Minería de Datos, la que se determina el papel fundamental de esta para la exploración, análisis, comprensión y uso del conocimiento detectado del enorme volumen de datos, como también la detección de las

tendencias y conductas de la información, con esto se facilita el entendimiento del avance tecnológico. En la investigación se utilizó para el empleo de minería de datos para predecir la producción de espárragos con fines de negocio y agroindustrial del distrito de Jayanca de Lambayeque. Para ello se aplicó los algoritmos para series de tiempo ARIMA, Redes neuronales y Holwinters y hacer un análisis comparativo. Los coeficientes de más alta correlación se dieron con el algoritmo de ARIMA, por lo que se concluye como un buen modelo para la predicción en la producción de espárragos [16].

EL creciente uso de los SIG hace necesario poder evaluar algunas herramientas que se complementen como la minería de datos espaciales. La investigación hace una revisión de estas técnicas para conocer cuál de ellas aporta mejores beneficios, con los resultados de las tendencias de su aplicación en el sector agrícola. La gestión de las variables por medio de estas herramientas propician un mejore proceso de toma de decisiones, y con ello contar con resultados de la evaluación de los datos [17].

En esta investigación en la que el crecimiento de los datos y el desarrollo de nuevas herramientas tecnológicas están apoyando a las organizaciones, se determina lo importante el empleo de la Minería de Datos con la finalidad de procesar y analizar los datos de la organización para poder obtener conocimiento útil. Como tal, en la rama de la ingeniería se está recopilando enormes cantidades de datos de diversos sensores implementados en las industrias, y es aquí donde la Minería de Datos por medio del diseño de modelos y patrones representan un camino a seguir. La investigación se orienta en hacer una revisión de la aplicación de esta herramienta en la agroindustria de la caña de azúcar en 25 años. Esta revisión bibliográfica realizada en la internet y revistas especializadas sobre el tema. Los resultados de esta revisión bibliográfica dan fe del incremento del empleo de la Minería de Datos en la agroindustria de la caña de azúcar, con mayor énfasis en el último quinquenio [18].

En la presente investigación en la que se desea poner en contexto la aplicación de la minería de datos relacionados con el clima, y usando para ella herramientas o técnicas y algoritmos que fueron utilizados en otros proyectos, con metodologías que han permitido encontrar conocimientos que no hubiera sido posible con otras herramientas para las variables climatológicas [19].

En la presente investigación que busca obtener y encontrar patrones de comportamiento de los estudiantes de la universidad nacional del Altiplano, en su plataforma virtual Moodle, empleando BigData de la información estudiantil, considerando para ello los datos del segundo semestre del año 2017. La investigación busca analizar el uso del Moodle y establecer las asociaciones de las variables, encontrar patrones de comportamiento no conocidos. Los resultados muestran de 16

escuelas, una mayor presencia en la virtualidad de Ingeniería de sistemas, Ingeniería de Minas, Ingeniería Mecánica Eléctrica e Ingeniería Económica, los algoritmos de clúster para segmentar o clasificar patrones detectan patrones de comportamiento agrupando en un primer clúster (0): Ingeniería de Minas en el módulo Course y View a las 18 horas; en el clúster (1): Ingeniería económica con mayor actividad en la asignación de tareas a las 22 horas [20].

En el presente estudio de la utilización de la Minería de Datos para el algoritmo de agrupación K-Medias aplicado a un conjunto de imágenes agroindustrial. El algoritmo tradicional agrupa en conjunto de datos con las mismas características denominadas clúster, en la investigación se utiliza el algoritmo de K-medias modificado para manejar una mayor flexibilidad para el caso de datos en el sector agrícola. El procesamiento de esta data en cuanto a la funcionalidad y aplicación del K-medias modificado para la agrupación, evalúa imágenes del nivel infrarrojo que han sido obtenidas de pruebas de la deficiencia hídrica para el cultivo de trigo [21].

En esta investigación, con una orientación de aplicación de la Minería de textos para encontrar patrones y conocimientos relacionadas con las quejas y reclamos del proceso de atención de a clientes de una empresa de galletas. La información obtenida de esta área se extrae la data del periodo de enero 2016 hasta julio del 2019. Los datos fueron tratados en la plataforma del Microsoft Office y para la Minería de textos con el software R de los datos de muestra. Los resultados determinar la frecuencia de las palabras más utilizadas creando una nube de palabras categorizadas sobre productos con mayores quejas, palabras detectadas como húmedas, plástico, quemadas, etc. La evaluación de coocurrencia mostró que la palabra galletas tienen una asociación con gusto, húmeda y menos. En menor proporción se encontraron palabras aisladas como rota, faltan, sin, etc. En tal sentido las quejas del departamento de atención a clientes se dan en relación con el sabor del producto, la humedad, si el producto está roto. Este resultado determina la necesidad de tomar en cuenta para el control y presentación del producto [22].

En la investigación, se resalta la Minería de datos como técnica especializada para diversas áreas con la finalidad de poder encontrar patrones y características ocultas en los repositorios de datos, campos como el marketing y los negocios es donde se ha ido incrementado su aplicación. Un aspecto importante, es que el conocimiento extraído de estas bases de datos siempre está disponible en los tomadores de decisiones. En la investigación se utilizó árboles de decisión, regresión lineal y M5, para el estudio se consideró 13 atributos de cultivo de vinos, con el que se efectuó una discriminación y posterior agrupación en un conjunto nombrado como químicos. Siendo los fenoles los más óptimos de acuerdo con el algoritmo aplicado, siendo este el más adecuado para usar en este cultivo [23].

En la investigación del autor, uno de los retos más importantes de los desarrolladores, es tener una gran cantidad de datos. Este reto obliga al empleo de minería de datos para estos datos con el objetivo de mejorar la cantidad de veces que se procesa estos datos. En la investigación se caracteriza el algoritmo en 3 pasos (S3), K-medias algoritmo para ver la gran cantidad de dimensiones en los datos, la clasificación no supervisada para imágenes. En el estudio para la concurrencia de datos se traza un flujo de datos con instrucción única para datos múltiples. Los resultados confirman la posibilidad de la concurrencia, además de que el modelo S3 no depende de la selección inicial, sino del agrupamiento con K-medias en los grupos iniciales. El algoritmo no supervisado S3 se considera una buena opción para las imágenes médicas con la técnica de minería de datos [24].

En la investigación relacionada con la cadena productiva de uva de mesa en el estado de sonora, donde se cuenta con 320,990.10 toneladas de producción de uva en el 2020, representando el 84.47% de la producción anual con un valor de \$9,296,739.69 (anuario agrícola 2020) estimándose 80% de la producción de exportación a los EEUU, obteniendo ingresos de 731,857 miles de dólares, siendo el producto de mayor valor para el gobierno. La investigación propuso la metodología con técnicas de minería de datos con la finalidad de identificar los patrones del mercado en cuanto a las fechas de inicio, duración y fin, igualmente se caracterizó los patrones de precios con algoritmos de series de tiempo. Toda esta información con la finalidad de poder tomar mejores decisiones, además de la mejor planificación de la producción de la uva de los sonorenses y por correspondencia de México [25]

Basado en la problemática y los antecedentes que lo sustenta, se planteó el problema ¿Cómo será la evaluación con Minería de datos en el control de Jabas de la empresa VANGUARD PERÚ en la provincia de Ica, en el periodo 2020-2021?

1.3 Justificación

El proyecto de investigación según [26], existen diversos criterios para determinar la justificación e importancia de un estudio, pero que no necesariamente se debe cumplir con todos, por lo que se presentan los más relevantes:

Conveniencia: el proyecto se considera conveniente porque va a servir para poder tener el control sobre un proceso de una empresa, pero además de hacer más eficiente el proceso que permita tener una distribución más adecuada de las jabas que la empresa utiliza en el transporte de los diversos productos de sus campos de cultivo.

Valor teórico: El estudio tiene un valor teórico importante, debido al auge que viene representando la Minería de Datos en los últimos años hace necesaria que se realicen estudios

sobre esta línea de investigación para contribuir en el conocimiento sobre la aplicación de esta herramienta tecnológica.

Utilidad metodológica: para el proyecto se evalúan diversas metodologías en el campo de la minería de datos, se evaluarán diversos algoritmos de minería de datos, para finalmente aplicar alguno de ellos según la metodología de uso, por lo que se contribuirá en fortalecer el uso de la metodología en bien de la ciencia.

En base a estas justificaciones el proyecto de tesis pretende aportar en el conocimiento de la aplicación de la minería de datos en un proceso específico del sector agroindustrial como es el proceso de control de Jabas.

1.4 Objetivos de la investigación

Para lograr el desarrollo exitoso de la investigación se planteó los siguientes objetivos que guiaron el desarrollo de este.

1.4.1 Objetivo general

Evaluar el control de Jabas mediante minería de datos en la empresa VANGUARD PERÚ-Ica en el periodo 2020-2021.

1.4.2 Objetivos específicos

- ✓ Caracterizar los tipos de Jabas utilizados en la empresa VANGUARD PERÚ-Ica en el periodo 2020-2021.
- ✓ Calcular la transferencia de Jabas desde origen al destino en la empresa VANGUARD PERÚ-Ica en el periodo 2020-2021.
- ✓ Controlar el estado de las Jabas en transferencia del origen al destino en la empresa Vanguard Perú-Ica en el periodo 2020-2021.

II. ESTRATEGIA METODOLÓGICA

La estrategia metodológica que se aplicó en el presente trabajo de tesis consistió en un estudio aplicado tecnológico, cuya línea de investigación se enmarca en la aplicación de la minería de datos, cuyo nivel de investigación es descriptivo, de corte transversal, el estudio retrospectivo cuya data a procesar está en el periodo del 2020-2021. El estudio es observacional no experimental de enfoque cuantitativo. La muestra seleccionada para el estudio esta conformada por las jabs de la empresa como unidad de análisis y la cantidad especificada se corresponden con las jabs del periodo 2020-2021.

Para el procesamiento de los datos se utilizó la plataforma de Google colab, con el lenguaje Python, utilizando las librerías de Pandas para el tratamiento de la base de datos (dataFame), numpy para el manejo numérico, Plotly para el manejo de gráficas. Igualmente, las librerías que permiten la visualización y comunicación de los resultados, librerías especializadas para el Análisis Exploratorio de los Datos (EDA) como ydata-profiling, dtale y sweetviz.

Previo al procesamiento de los datos, se tuvo que instalar las librerías en mención por medio de la sentencia ;pip install.

Instalación de la librería para manejo gráfico interactivos de los resultados con la sentencia: **!pip install plotly==4.14.3**. Para crear informes detallados basados en el dataframe de pandas.

Además para proporcionar información sobre las variables individuales, el análisis de datos faltantes (missing), la correlación de los datos y las interacciones, se instaló la librería: **! pip install ydata-profiling**.

Además, se instaló la librería dtale, la que permite interactúan de manera muy fácil con los datos, con la posibilidad de poder editarlos: **! pip install dtale**.

Adicionalmente se instaló la librería sweetviz. Esta es una librería para la exploración y visualización de datos interactiva, para ello se instala la librería con la sentencia: **! pip install sweetviz**

2.1. **Método:** El método a utilizar estará establecido por el algoritmo que se seleccione para los resultados de la investigación, pudiendo ser estos algoritmos para clasificación de datos numéricos y datos categóricos según lo especificado en los objetivos de la investigación. Igualmente se aplicará el procedimiento si se aplicará para algoritmos supervisados o algoritmos no supervisados. El método que se aplique a la minería de datos estará soportado por el lenguaje de programación Python para data mining, y la herramienta que se utilizará para el procesamiento de los datos se basó en Google colab que dependerá de la evaluación de ellas, es importante conocer que todas las herramientas mencionadas son de libre uso, por lo que no requiere de licencia para su uso.

Se precisa que la plataforma del Google colab, con lenguaje Python, se encuentra disponible en la nube para cualquier persona que tenga una cuenta de Gmail.

Para el proyecto de tesis se aplicará la metodología de ciencia de datos [27] y Minería de datos la misma que sigue las siguientes fases:

1. Business Problem - Problema empresarial
2. Data Acquisition - Adquisición de datos
3. Data preparation - Preparación de los datos
4. Data Modeling - Modelado de datos
5. Visualization and Communication - Visualización y comunicación
6. Deploy and maintains - Despliegue y mantenimiento

Esta metodología denominada CRISP-DM Methodology según [28]

La metodología no utiliza específicamente algún modelo, ya que el estudio llega solo al nivel descriptivo [5], por lo cual no se requiere la fase de la creación de un modelo, como si lo es para otro tipo de estudios.

2.2 Aplicación del método

1. Problema empresarial

El problema que aqueja a la empresa VANGUARD PERÚ-Ica, radica en el poco conocimiento que tiene sobre el movimiento de las jabas que se utilizan para el packing de los productos resultantes de los productos.

2. Adquisición de los datos

Para la adquisición de los datos no presentó ningún inconveniente para obtenerlos, ya que, como trabajador de la empresa, estaba destinado al control de las jabas, por lo cual, para tener

un control básico de estas, fue necesario crear un archivo en formato Excel para registrar el movimiento de estas. El control de estas jabas es responsabilidad de **COMPLETAR**.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	SEM	Fecha	hora partida	Ticket/GR	FUNDO ORIGEN	FUNDO DESTINO	Tipo Jaba	JABAS POR COLOR	n° jaba	PARIHUELA	Observa	LOTE	ESTADO
2	46	11/11/2020	7:10:00	PD001-114	PACKING	EL ARENAL	Jabas Cosecheras AZUL VANGUARD		840		VACIO	EA04	ENVIADO
3	46	11/11/2020	7:15:00	PD001-112	PACKING	EL ARENAL	Jabas Cosecheras AZUL VANGUARD		840		VACIO	EA04	ENVIADO
4	46	11/11/2020	9:22:00	PD001-115	PACKING	EL ARENAL	Jabas Cosecheras AZUL VANGUARD		600		VACIO	EA04	ENVIADO
5	46	11/11/2020	9:25:00	PD001-116	PACKING	EL ARENAL	Jabas Cosecheras AZUL VANGUARD		600		VACIO	EA04	ENVIADO
6	46	11/11/2020	12:00:00	004-008177	ALM CHALLAPAMPA	PACKING	Basales ROJA BASAL		720		VACIO	-	RECIBIDO
7	46	11/11/2020	12:00:00	004-008178	ALM CHALLAPAMPA	PACKING	Basales ROJA BASAL		720		VACIO	-	RECIBIDO
8	46	11/11/2020	13:15:00	PD001-117	PACKING	CHALLAPAMPA	Basales ROJA BASAL		600	62	VACIO	CH14	ENVIADO
9	46	11/11/2020	13:20:00	PD001-118	PACKING	CHALLAPAMPA	Basales ROJA BASAL		600	62	VACIO	CH14	ENVIADO
10	46	11/11/2020	16:00:00	PD001-119	PACKING	CHALLAPAMPA	Jabas Cosecheras CH AZUL		700	14	VACIO	CH14	ENVIADO

Fig. 3. Estructura de los datos iniciales de las jabas

En La Fig. 3, se muestran los datos iniciales del registro de los envíos de las jabas, y cuyos datos son descritos en la Tabla I de los metadatos para tener una comprensión de los datos que se registran en el control de las jabas de la empresa.

	M	N	O	P	Q	R	S	T
1	ESTADO	HORA CORTE	CAMPO CHALLA	CAMPO ARENAL	ALMACEN PACKING	CAMPO CASUARINAS	CAMPO MILAGRITOS	CAMPO LOS LAURELES
2	ENVIADO	7,00	0,00	840,00	-840,00	0,00	0,00	0,00
3	ENVIADO	7,00	0,00	840,00	-840,00	0,00	0,00	0,00
4	ENVIADO	9,00	0,00	600,00	-600,00	0,00	0,00	0,00
5	ENVIADO	9,00	0,00	600,00	-600,00	0,00	0,00	0,00
6	RECIBIDO	12,00	0,00	0,00	720,00	0,00	0,00	0,00
7	RECIBIDO	12,00	0,00	0,00	720,00	0,00	0,00	0,00
8	ENVIADO	13,00	600,00	0,00	-600,00	0,00	0,00	0,00
9	ENVIADO	13,00	600,00	0,00	-600,00	0,00	0,00	0,00
10	ENVIADO	16,00	700,00	0,00	-700,00	0,00	0,00	0,00

Fig. 4. Estructura de los datos finales de las jabas

En la Fig. 4, de los datos finales, se muestra la cantidad de las jabas que se envían y reciben en los campos desde el almacén Packing como el retorno al almacén. En la Fig. x, se puede apreciar que existen 5 campos que tiene la empresa a la que se envían las jabas para el traslado de los cultivos cosechados. Tener en consideración como el proceso se trata de envío y recepción de las jabas se puede comprobar que los valores del campo donde se recibe o envía las jabas van a tener un signo positivo (recepción) o negativo (envío).

TABLA I
METADATOS DEL CONTROL DE JABAS

DATOS	DESCRIPCION	TIPO DATO
SEM	Semana de envío	Categorico
FECHA	Fecha de envío	Fecha-hora
HORA PARTIDA	Hora de envío	Fecha-hora
TICKET/GR	Ticket de despacho	Categorico
FUNDO ORIGEN	Fundo de donde se despacha	Categorico
FUNDO DESTINO	Fundo a donde se despacha	Categorico
TIPO JABA	Tipo de uso de la Jaba	Categorico
JABAS POR COLOR	Color de la Jaba	Categorico
N° JABAS	Cantidad de jabas despachadas	Numérico
PARIHUELA	Cantidad de parihuelas	Numérico
OBSERVACIÓN	Estado en el que se encuentra la Jaba (lleno o vacia)	Categorico
LOTE	Lote asignado a la jaba	Categorico
ESTADO	Estado del envio (Enviado, Recibido)	Categorico
HORA CORTE	Hora en la que se hace corte de envío	Categorico
CAMPO CHALLA	Campo de donde se envía y recibe	Categorico
CAMPO ARENAL	Campo de donde se envía y recibe	Categorico
ALMACEN PACKING	Campo de donde se envía y recibe	Categorico
CAMPO CASUARINAS	Campo de donde se envía y recibe	Categorico
CAMPO MILAGRITOS	Campo de donde se envía y recibe	Categorico
CAMPO LOS LAURELES	Campo de donde se envía y recibe	Categorico

3. Preparación de los datos

En la preparación de los datos, primero se tuvo que separar la hoja que contenía la Base de Datos. para lograr esto simplemente en el Excel se tenía que hacer una copia de la Hoja de la Base de Datos y que esa copia se creara en un archivo nuevo.

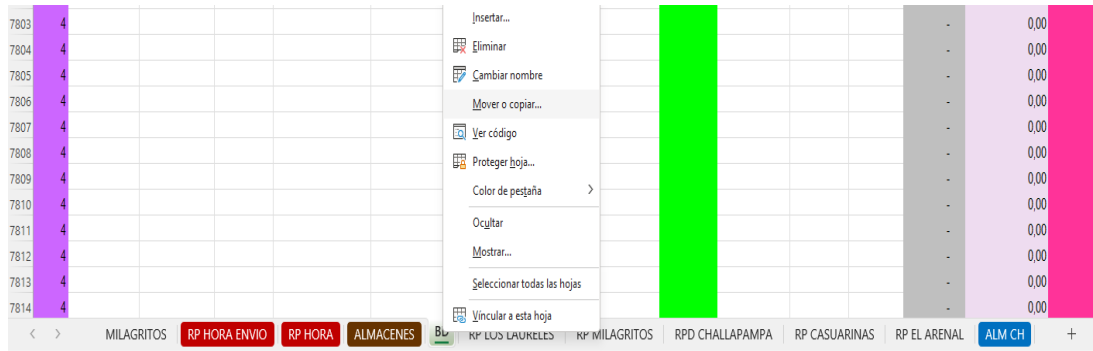


Fig. 5. Separación de la Base de Datos

En la Fig. 5, se puede apreciar la separación de la Base de Datos a un nuevo archivo, solo hay que darle un clic con el botón derecho a la hoja BD y la opción Mover o copiar. Igualmente, la Fig. 5 muestra la necesidad de esta preparación, ya que existen diversas hojas de trabajo que contienen otro tipo de información, pero que, para realizar el procesamiento con la minería de datos, se realiza esta acción.

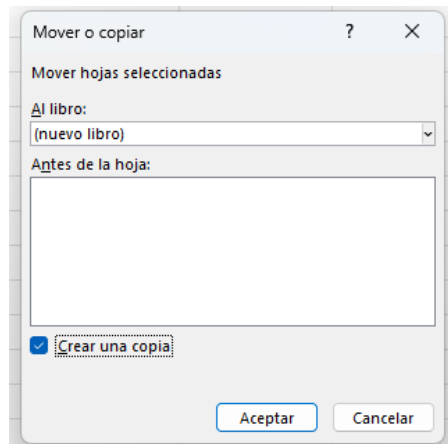


Fig. 6. Opción de copiar hoja en nuevo libro

Posterior a la creación del archivo de Excel, a este se le asignó el nombre **Jabas**. El archivo nuevo creado, al realizar el recorrido por los datos igualmente se puede comprobar que

y la que será examinada y realizar el preprocesamiento previo antes de la aplicación de la minería de datos y lograr los objetivos del proyecto.

Se inicia con la carga de las librerías para realizar la visualización de los datos con las librerías de pandas (para manejo de la base de datos) y la librería numpy para el manejo matemático.

La Fig. 9 muestra dichas acciones.

```
# Carga de librerías
import pandas as pd
import numpy as np

# Carga de los datos
Jabas=pd.read_excel('Jabas.xlsx')
```

Fig. 9. Carga de librerías y carga de los datos

Para conocer los metadatos de los datos del archivo se ejecuto la sentencia para conocer dicha información. `Jabas.info()` cuyos resultados de muestran en la Fig. 10.

```
Jabas.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7546 entries, 0 to 7545
Data columns (total 20 columns):
#   Column                Non-Null Count  Dtype
---  -
0   SEM                    7396 non-null  float64
1   Fecha                  7546 non-null  datetime64[ns]
2   hora partida           7546 non-null  object
3   Ticket/GR              7545 non-null  object
4   FUNDO ORIGEN          7544 non-null  object
5   FUNDO DESTINO         7543 non-null  object
6   Tipo Jaba              7113 non-null  object
7   JABAS POR COLOR       7058 non-null  object
8   nº jabas               7111 non-null  float64
9   PARIHUELA              7041 non-null  float64
10  Observación            7112 non-null  object
11  LOTE                    6714 non-null  object
12  ESTADO                  7546 non-null  object
13  HORA CORTE              7546 non-null  int64
14  CAMPO CHALLA           7546 non-null  int64
15  CAMPO ARENAL           7546 non-null  int64
16  ALMACEN PACKING        7546 non-null  int64
17  CAMPO CASUARINAS       7546 non-null  int64
18  CAMPO MILAGRITOS       7546 non-null  int64
19  CAMPO LOS LAURELES     7546 non-null  int64
dtypes: datetime64[ns](1), float64(3), int64(7), object(9)
memory usage: 1.2+ MB
```

Fig. 10. Información de la metadata del archivo

Como se aprecia en la Fig.10, se tiene que el archivo contiene 7546 registros de datos con un total de 20 columnas (tener en cuenta que la columna inicia en 0). En las líneas finales se

muestran los tipos de datos y que cantidad de cada uno de ellos. Además, se nos informa que el archivo pesa 1.2 MB de almacenamiento.

Se precisa que los datos de tipo object, son considerados como datos de tipo categórico, los demás son o números o fecha. Igualmente, la sentencia no muestra que no existen valores nulos (non-null).

Para visualizar los datos contenidos en la Base de datos se utiliza la instrucción `Jabas.head()`, para visualizar los primeros 5 registros de los datos.

SEM	Fecha	hora partida	Ticket/GR	FUNDO ORIGEN	FUNDO DESTINO	Tipo Jaba	JABAS POR COLOR	n° jabas	PARIHUELA	Observación	LOTE	ESTADO	HORA CORTE	CAMPO CHALLA	CAMPO ARENAL	ALMACEN PACKING	CAMPO CASUARINAS	CAMPO MILAGRITOS	CAMPO LOS LAURELES
0	46.0	2020-11-11	07:10:00	PD001-114	PACKING	EL ARENAL	Jabas Cosecheras	AZUL VANGUARD	840.0	NaN	VACIO EA04	ENVIADO	7	0	840	-840	0	0	0
1	46.0	2020-11-11	07:15:00	PD001-112	PACKING	EL ARENAL	Jabas Cosecheras	AZUL VANGUARD	840.0	NaN	VACIO EA04	ENVIADO	7	0	840	-840	0	0	0
2	46.0	2020-11-11	09:22:00	PD001-115	PACKING	EL ARENAL	Jabas Cosecheras	AZUL VANGUARD	600.0	NaN	VACIO EA04	ENVIADO	9	0	600	-600	0	0	0
3	46.0	2020-11-11	09:25:00	PD001-116	PACKING	EL ARENAL	Jabas Cosecheras	AZUL VANGUARD	600.0	NaN	VACIO EA04	ENVIADO	9	0	600	-600	0	0	0
4	46.0	2020-11-11	12:00:00	004-008177	ALM CHALLAPAMPA	PACKING	Basales	ROJA BASAL	720.0	NaN	VACIO	0 RECIBIDO	12	0	0	720	0	0	0

Fig. 11. Datos de la Base de datos

SEM	Fecha	hora partida	Ticket/GR	FUNDO ORIGEN	FUNDO DESTINO	Tipo Jaba	JABAS POR COLOR	n° jabas	PARIHUELA	Observación	LOTE	ESTADO	HORA CORTE	CAMPO CHALLA	CAMPO ARENAL	ALMACEN PACKING	CAMPO CASUARINAS	CAMPO MILAGRITOS	CAMPO LOS LAURELES
0	46.0	2020-11-11	07:10:00	PD001-114	PACKING	EL ARENAL	Jabas Cosecheras	AZUL VANGUARD	840.0	NaN	VACIO EA04	ENVIADO	7	0	840	-840	0	0	0
1	46.0	2020-11-11	07:15:00	PD001-112	PACKING	EL ARENAL	Jabas Cosecheras	AZUL VANGUARD	840.0	NaN	VACIO EA04	ENVIADO	7	0	840	-840	0	0	0
2	46.0	2020-11-11	09:22:00	PD001-115	PACKING	EL ARENAL	Jabas Cosecheras	AZUL VANGUARD	600.0	NaN	VACIO EA04	ENVIADO	9	0	600	-600	0	0	0
3	46.0	2020-11-11	09:25:00	PD001-116	PACKING	EL ARENAL	Jabas Cosecheras	AZUL VANGUARD	600.0	NaN	VACIO EA04	ENVIADO	9	0	600	-600	0	0	0
4	46.0	2020-11-11	12:00:00	004-008177	ALM CHALLAPAMPA	PACKING	Basales	ROJA BASAL	720.0	NaN	VACIO	0 RECIBIDO	12	0	0	720	0	0	0
...
7541	4.0	2021-01-20	14:55:13	PD04-535	CHALLAPAMPA	PACKING	NaN	NaN	NaN	NaN	NaN	RECIBIDO	14	0	0	0	0	0	0
7542	4.0	2021-01-20	14:55:25	PD01-3497	PACKING	LAS CASUARINAS	Jabas Cosecheras	CH AZUL	350.0	26.0	VACIO LC16	ENVIADO	14	0	0	-350	350	0	0
7543	4.0	2021-01-20	14:56:12	PD02-700	PACKING	LAS CASUARINAS	Jabas Cosecheras	AZUL VANGUARD	650.0	26.0	LLENO LC23	RECIBIDO	14	0	0	650	-650	0	0
7544	4.0	2021-01-20	15:03:50	PD05-629	MILAGRITOS	PACKING	Jabas Cosecheras	ML VERDE	650.0	26.0	LLENO ML08	RECIBIDO	15	0	0	650	0	-650	0
7545	4.0	2021-01-20	15:09:21	PD01-3498	PACKING	CHALLAPAMPA	Jabas Cosecheras	CH AZUL	400.0	8.0	VACIO CH09	ENVIADO	15	400	0	-400	0	0	0

546 rows x 20 columns

Fig. 12. Visualización de los primeros y últimos registros

En la Fig. 12, se puede visualizar los primeros y últimos registros del Dataframe, simplemente invocando al Dataframe Jaba.

Para efectos de las necesidades del análisis que se va a implementar, no se requiere las columnas finales desde la HORA CORTE, hasta la última columna, debido a que la

información de estas columnas es redundantes con las columnas FUNDO ORIGEN, FUNDO DESTINO, n° jabas, por lo cual se eliminarán estas columnas por medio de la sentencia drop.

```
Jabas.drop(columns=['HORA CORTE', 'CAMPO CHALLA', 'CAMPO ARENAL', 'ALMACEN PACKING', 'CAMPO CASUARINAS', 'CAMPO MILAGRITOS', 'CAMPO LOS LAURELES'], inplace=True)
```

En la sentencia se incluyen las columnas que se van a eliminar y el argumento inplace=True, es para que la eliminación modifique el dataframe original. Los resultados se muestran con la siguiente sentencia **Jabas.head()**.

	SEM	Fecha	hora partida	Ticket/GR	FUNDO ORIGEN	FUNDO DESTINO	Tipo Jaba	JABAS POR COLOR	n° jabas	PARIHUELA	Observación	LOTE	ESTADO
0	46.0	2020-11-11	07:10:00	PD001-114	PACKING	EL ARENAL	Jabas Cosecheras	AZUL VANGUARD	840.0	NaN	VACIO	EA04	ENVIADO
1	46.0	2020-11-11	07:15:00	PD001-112	PACKING	EL ARENAL	Jabas Cosecheras	AZUL VANGUARD	840.0	NaN	VACIO	EA04	ENVIADO
2	46.0	2020-11-11	09:22:00	PD001-115	PACKING	EL ARENAL	Jabas Cosecheras	AZUL VANGUARD	600.0	NaN	VACIO	EA04	ENVIADO
3	46.0	2020-11-11	09:25:00	PD001-116	PACKING	EL ARENAL	Jabas Cosecheras	AZUL VANGUARD	600.0	NaN	VACIO	EA04	ENVIADO
4	46.0	2020-11-11	12:00:00	004-008177	ALM CHALLAPAMPA	PACKING	Basales	ROJA BASAL	720.0	NaN	VACIO	0	RECIBIDO

Fig. 13. Nuevo dataframe

En la Fig. 13, se presenta los datos del dataframe después de haber eliminado las columnas de la sentencia drop.

Igualmente visualizando los datos resultantes, se ve la posibilidad de complementar esta eliminación, para lo cual se van a eliminar las columnas que no son representativas para el análisis como las columnas: hora partida, Ticket/GR, PARIHUELA.

Jabas.drop(columns=['hora partida','Ticket/GR','PARIHUELA'], inplace=True) y cuyos resultados se muestran en la Fig. 14.

	SEM	Fecha	FUNDO ORIGEN	FUNDO DESTINO	Tipo Jaba	JABAS POR COLOR	n° jabas	Observación	LOTE	ESTADO
0	46.0	2020-11-11	PACKING	EL ARENAL	Jabas Cosecheras	AZUL VANGUARD	840.0	VACIO	EA04	ENVIADO
1	46.0	2020-11-11	PACKING	EL ARENAL	Jabas Cosecheras	AZUL VANGUARD	840.0	VACIO	EA04	ENVIADO
2	46.0	2020-11-11	PACKING	EL ARENAL	Jabas Cosecheras	AZUL VANGUARD	600.0	VACIO	EA04	ENVIADO
3	46.0	2020-11-11	PACKING	EL ARENAL	Jabas Cosecheras	AZUL VANGUARD	600.0	VACIO	EA04	ENVIADO
4	46.0	2020-11-11	ALM CHALLAPAMPA	PACKING	Basales	ROJA BASAL	720.0	VACIO	0	RECIBIDO

Fig. 14. dataframe resultante

El dataframe resultante cuenta con las características necesarias para poder realizar el análisis de los datos según nuestro estudio y poder evaluar el control de las Jabas planteada en nuestro objetivo principal.

Finalmente, se comprueba si no existen datos faltantes o nulos en el dataframe, para ello utilizamos la sentencia **Jabas.isnull().sum(axis=0)**

```
SEM          150
Fecha        0
FUNDO ORIGEN 2
FUNDO DESTINO 3
Tipo Jaba    433
JABAS POR COLOR 488
nº jabas    435
Observación  434
LOTE        832
ESTADO      0
dtype: int64
```

Fig. 15. Datos nulos o faltantes

En la Fig. 15, se puede comprobar que existen en varias columnas datos faltantes, con excepción de la Fecha y el Estado. ante esta situación, se puede solucionar mediante la imputación de los datos o simplemente eliminando los valores nulos o faltantes. Para nuestro caso tomamos la decisión de eliminar los datos nulos o faltantes, para ello se aplica la sentencia **Jabas1=Jabas.dropna()**, y verificamos los resultados con la sentencia **Jabas1.info()**. Aquí se ha creado un nuevo dataframe Jabas1.

```
SEM          0
Fecha        0
FUNDO ORIGEN 0
FUNDO DESTINO 0
Tipo Jaba    0
JABAS POR COLOR 0
nº jabas    0
Observación  0
LOTE        0
ESTADO      0
dtype: int64
```

Fig. 16. Información del nuevo dataframe Jabas1

En la Fig. 16, se tiene la información del nuevo dataframe en la que se han eliminado los valores nulos o perdidos.

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7546 entries, 0 to 7545
Data columns (total 10 columns):
#   Column                Non-Null Count  Dtype
---  -
0   SEM                    7396 non-null  float64
1   Fecha                  7546 non-null  datetime64[ns]
2   FUNDO ORIGEN          7544 non-null  object
3   FUNDO DESTINO         7543 non-null  object
4   Tipo Jaba              7113 non-null  object
5   JABAS POR COLOR       7058 non-null  object
6   n° jabas               7111 non-null  float64
7   Observación           7112 non-null  object
8   LOTE                   6714 non-null  object
9   ESTADO                 7546 non-null  object
dtypes: datetime64[ns](1), float64(2), object(7)
memory usage: 589.7+ KB

```

Fig. 17. dataFrame con valores nulos

```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 6523 entries, 0 to 7545
Data columns (total 10 columns):
#   Column                Non-Null Count  Dtype
---  -
0   SEM                    6523 non-null  float64
1   Fecha                  6523 non-null  datetime64[ns]
2   FUNDO ORIGEN          6523 non-null  object
3   FUNDO DESTINO         6523 non-null  object
4   Tipo Jaba              6523 non-null  object
5   JABAS POR COLOR       6523 non-null  object
6   n° jabas               6523 non-null  float64
7   Observación           6523 non-null  object
8   LOTE                   6523 non-null  object
9   ESTADO                 6523 non-null  object
dtypes: datetime64[ns](1), float64(2), object(7)
memory usage: 560.6+ KB

```

Fig. 18. dataFrame con valores nulos

Los resultados del dataFrame Jabas de la Fig. 17, muestran que existen 7546 registros con 10 columnas y el archivo pesa 589.7 Kb, y en el dataFrame Jabas1 de la Fig. 18, muestra que existen 6523 registros con las mismas 10 columnas y un peso del archivo de 560.6 Kb notándose la diferencia al haberse eliminado los valores nulos.

Es importante tener en consideración que se pueden presentar datos que no corresponden con los valores de la columna específica, se aplica la librería

```

import sweetviz as sv
report=sv.analyze(Jabas1)
report.show_html()

```

La ejecución de la librería para este análisis se comprueba que existe un error en los datos, en las que se detectó en la columna JABAS POR COLOR, que existen algunos datos que no corresponden con los datos registrados, posiblemente mal registrado.

```

/usr/local/lib/python3.10/dist-packages/sweetviz/dataframe_report.py:74: FutureWarning: iteritems is deprecated and will be removed in a future version. Use .items instead.
  all_source_names = [cur_name for cur_name, cur_series in source_df.iteritems()]
/usr/local/lib/python3.10/dist-packages/sweetviz/dataframe_report.py:109: FutureWarning: iteritems is deprecated and will be removed in a future version. Use .items instead.
  filtered_series_names_in_source = [cur_name for cur_name, cur_series in source_df.iteritems()]
Feature: JABAS POR COLOR [redacted] [55%] 00.03 -> (00.0
/usr/local/lib/python3.10/dist-packages/sweetviz/series_analyzer_numeric.py:25: FutureWarning: The 'mad' method is deprecated and will be removed in a future version. To compute the
  stats["mad"] = series.mad()
/usr/local/lib/python3.10/dist-packages/sweetviz/series_analyzer_cat.py:28: FutureWarning: iteritems is deprecated and will be removed in a future version. Use .items instead.
  for item in category_counts.iteritems():
/usr/local/lib/python3.10/dist-packages/sweetviz/series_analyzer_cat.py:28: FutureWarning: iteritems is deprecated and will be removed in a future version. Use .items instead.
  for item in category_counts.iteritems():
/usr/local/lib/python3.10/dist-packages/sweetviz/series_analyzer_cat.py:28: FutureWarning: iteritems is deprecated and will be removed in a future version. Use .items instead.
  for item in category_counts.iteritems():
/usr/local/lib/python3.10/dist-packages/sweetviz/series_analyzer_cat.py:28: FutureWarning: iteritems is deprecated and will be removed in a future version. Use .items instead.
  for item in category_counts.iteritems():
-----
TypeError                                 Traceback (most recent call last)
<jupyter-input-23-3ac8a7ad9b68> in <cell line: 1>()
----> 1 report=sv.analyze(Jabas1)
      2 report.show_html()

```

Fig. 19. Error de datos detectado por la librería

En la Fig. 19, la aplicación de la librería **sweetviz** detectó un error en los datos de la columna **JABAS POR COLOR**, dicho error se debe que se encontró tres registros con valor 0 en lugar del color de la Jaba, pero además esta información correspondía a las parihuelas, las cuales son los contenedores y no las Jabas, los mismos que se presentan en la Fig. 20.

Count	JABAS POR COLOR									
	0	AMARI...	AZUL VANG...	AZULES	CH AZUL	MADERA	ML VERDE	NARAN...	NEGRAS	NEGRAS AZULES
1	0.0	112.0	92.0	1.0	526.0	1.0	597.0	0.0	38.0	0.0
2	0.0	3.0	0.0	0.0	37.0	0.0	16.0	0.0	0.0	0.0
3	0.0	105.0	27.0	3.0	463.0	0.0	450.0	0.0	20.0	5.0
4	0.0	39.0	57.0	0.0	301.0	0.0	220.0	0.0	17.0	0.0
46	0.0	5.0	11.0	0.0	20.0	0.0	0.0	0.0	0.0	0.0
47	0.0	6.0	12.0	0.0	16.0	0.0	0.0	0.0	0.0	0.0
48	0.0	37.0	84.0	0.0	130.0	0.0	34.0	0.0	0.0	0.0
50	3.0	53.0	53.0	0.0	325.0	0.0	181.0	4.0	0.0	0.0
51	0.0	82.0	219.0	0.0	303.0	0.0	177.0	5.0	0.0	0.0
52	0.0	35.0	64.0	0.0	98.0	0.0	99.0	2.0	5.0	0.0
53	0.0	34.0	125.0	0.0	170.0	0.0	165.0	0.0	14.0	0.0
Total	3.0	511.0	744.0	4.0	2389.0	1.0	1939.0	11.0	94.0	5.0

Fig. 20. Errores detectados en la columna JABAS POR COLOR

En la Semana 50 se detecta que se ha registrado tres jabas con valor cero (0), que no es coherente con el tipo de dato, con la eliminación de estos datos se cuenta con 6520 registros finales.

Con los realizado en esta tercera fase de la metodología, se tiene los datos preparados y listos para realizar el análisis respectivo. Estando ya los datos preparados, procedemos a guardar los datos por precaución con la sentencia: **Jabas1.to_csv('/content/Jabas.csv')** el archivo original utilizado estaba en formato xlsx, y al guardar los datos finales estos se están guardando en formato csv, que son los mas utilizados por ser menos pesados que los archivos Excel.

La visualización y comunicación de los resultados, así como el despliegue y mantenimiento se presentan en el capítulo de los resultados.

III. RESULTADOS

3.1. Aplicación del método

Un primer acercamiento a los datos se logra con la sentencia para ver los estadísticos descriptivo para los valores numéricos. **Jabas1.describe()**

	SEM	nº jabas
count	6523.000000	6523.000000
mean	24.994328	535.650008
std	24.193849	224.440204
min	1.000000	0.000000
25%	3.000000	400.000000
50%	4.000000	600.000000
75%	51.000000	700.000000
max	53.000000	6540.000000

Fig. 21. Estadística descriptiva

En la Fig. 21, la estadística descriptiva de los datos arroja para la cantidad de jabas que se trasladan, en promedio se trasladan 535 jabas con un mínimo de 224 y un máximo de 6540 jabas.

```
# Carga de librerias
import pandas as pd
import numpy as np
```

```
# Carga de los datos
Jabas=pd.read_csv('Jabas.csv')
```

```
# cargando librerias para análisis gráfico con plotly
```

```
import plotly.express as px
import plotly.graph_objects as go
```

```
# Creación de gráfico del número de Jabas por Semana
```

```
tabla=pd.pivot_table(Jabas, values='n° jabas', index='SEM', aggfunc=np.sum)
```

```
fig=px.bar(tabla, y='n° jabas')
```

```
fig.show()
```

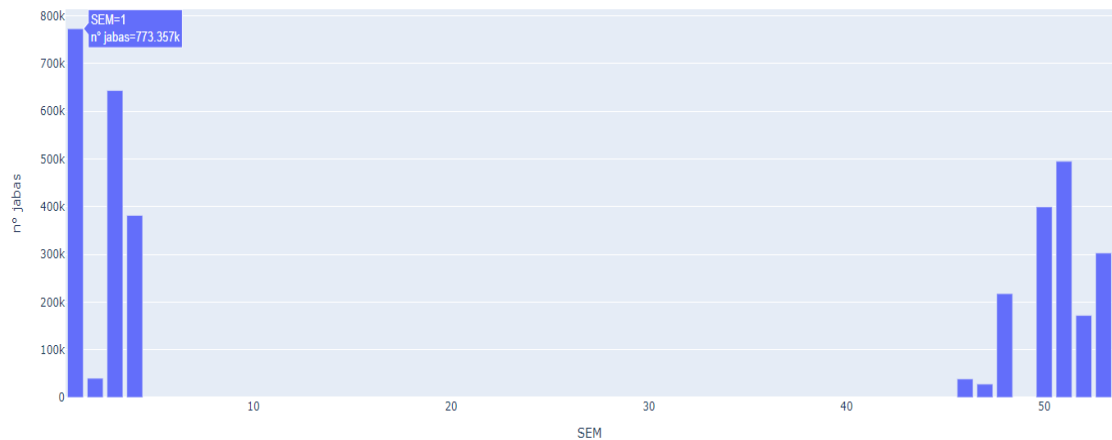


Fig. 22. Gráfico de traslado de Jabas por Semana

En la Fig. 22, donde se muestra las jabas trasladadas entre las semanas de la primeras a la semana 50, la gráfica muestra como en las primeras semanas, la primera semana presenta un alto volumen de traslado de las Jabas, y en las últimas semanas como la 50, 51 igualmente. La gráfica muestra que no ha habido desplazamiento de las Jabas entre las semanas intermedias a estas semanas.

Se precisa igualmente que cuando se emplea la librería del plotly para la creación de los gráficos, estos gráficos son interactivos, vale decir que la información de cada columna de las gráficas se muestra cuando se pasa el puntero por encima de la columna de los datos, por ello es que el sistema no requiere tener una visualización fija de los valores en las columnas, como usualmente se muestra en gráficas realizadas por otros softwares como Excel.

Creación de gráfico del número según el Fundo de donde se despachan

```
tabla=pd.pivot_table(Jabas, values='n° jabas', index='FUNDO ORIGEN', aggfunc=np.sum)
```

```
fig=px.bar(tabla, y='n° jabas')
```

```
fig.show()
```

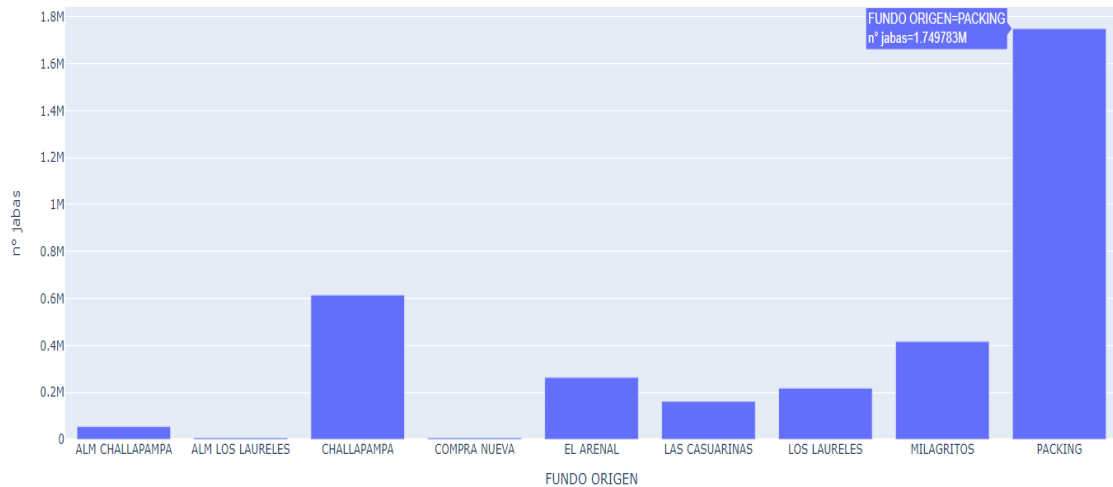


Fig. 23. Gráfico de traslado de Jabas desde el Fundo Origen

La Fig. 23, muestra, que el mayor volumen de desplazamiento de Jabas según el origen corresponde la Zona de Packing, con cerca de dos millones de Jabas.

La gráfica también refleja que las Jabas que se desplazan como origen, en realidad es de donde se envía por lo que las Jabas que han sido enviadas a los Fundos, posteriormente retornas la Zona de Packing, o son enviadas a otros Fundos donde se necesiten.

La gráfica muestra que el fundo que mayor despacho de jabas realiza, es del Fundo Challapampa, seguido del Fundo Milagritos y el Fundo el Arenal.

Creación de gráfico del número según el Fondo a donde se despachan

```
tabla=pd.pivot_table(Jabas, values='n° jabas', index='FUNDO DESTINO', aggfunc=np.sum)
```

```
fig=px.bar(tabla, y='n° jabas')
```

```
fig.show()
```

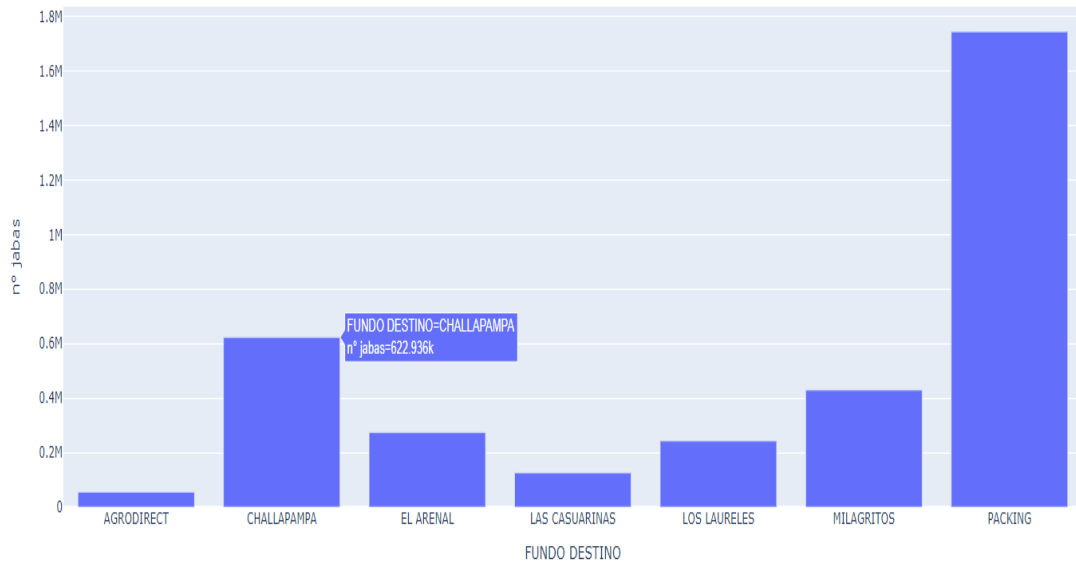


Fig. 24. Gráfico de traslado de Jabas al Fondo de destino

La Fig. 24, muestra el desplazamiento de las Jabas a los fondos de destino, la gráfica es muy específica en mostrar que el lugar de destino mayormente es la zona de Packing y en cuanto a los destinos de las Jabas, mayormente se dan a los Fondos de Challapampa, el Fondo Milagritos y el Fondo el Arenal.

Este comportamiento es muy similar a la gráfica de la Fig. 23, desde donde se despachan las Jabas.

Creación de gráfico del número según el tipo de Jaba

```
tabla=pd.pivot_table(Jabas, values='n° jabas', index='Tipo Jaba', aggfunc=np.sum)
```

```
fig=px.bar(tabla, y='n° jabas')
```

```
fig.show()
```

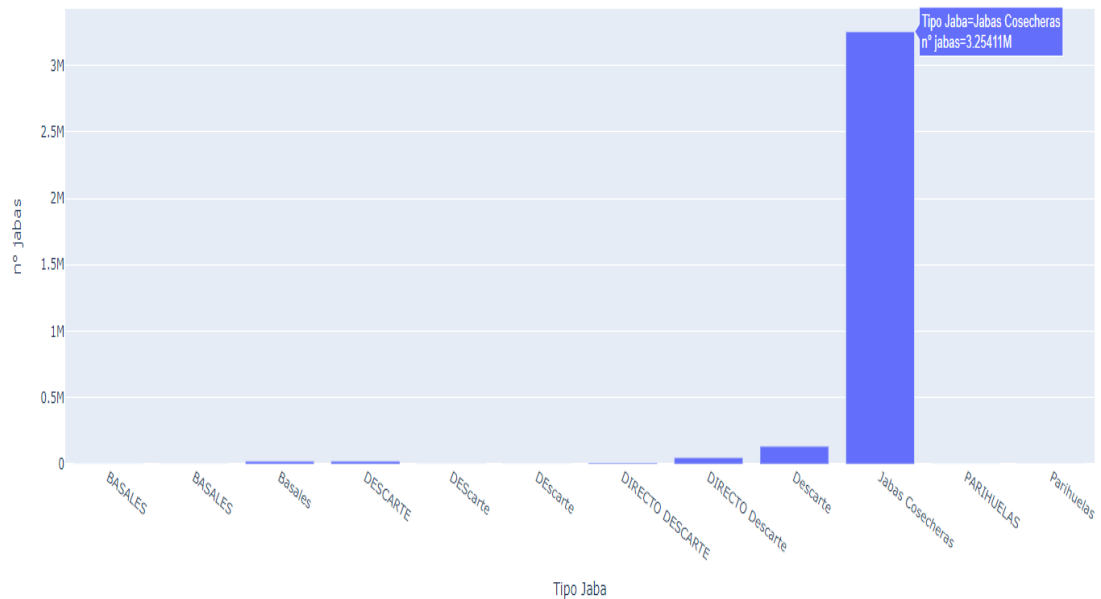


Fig. 25. Gráfico de traslado de Jabas según el tipo de Jaba

En esta gráfica de la Fig. 25, se puede comprobar, que el registro de los datos en esta variable no está bien estructurada, ya que se tiene datos que no están bien establecidos, se escriben datos todo en mayúsculas, en otro caso mayúscula y minúscula. En el análisis de datos el valor que contiene una columna no debe variar, debe ser exacto el dato.

Además, en cuanto a la distribución de las Jabas, según el tipo de Jabas, se muestra que son las Jabas cosecheras las que mayor movimiento tienen. También la gráfica demuestra que existe una pequeña cantidad de Jabas que pasan a descarte.

```
# Creación de gráfico del número según el color de la Jaba
```

```
tabla=pd.pivot_table(Jabas, values='n° jabas', index='JABAS POR COLOR', aggfunc=np.sum)
```

```
fig=px.bar(tabla, y='n° jabas')
```

```
fig.show()
```

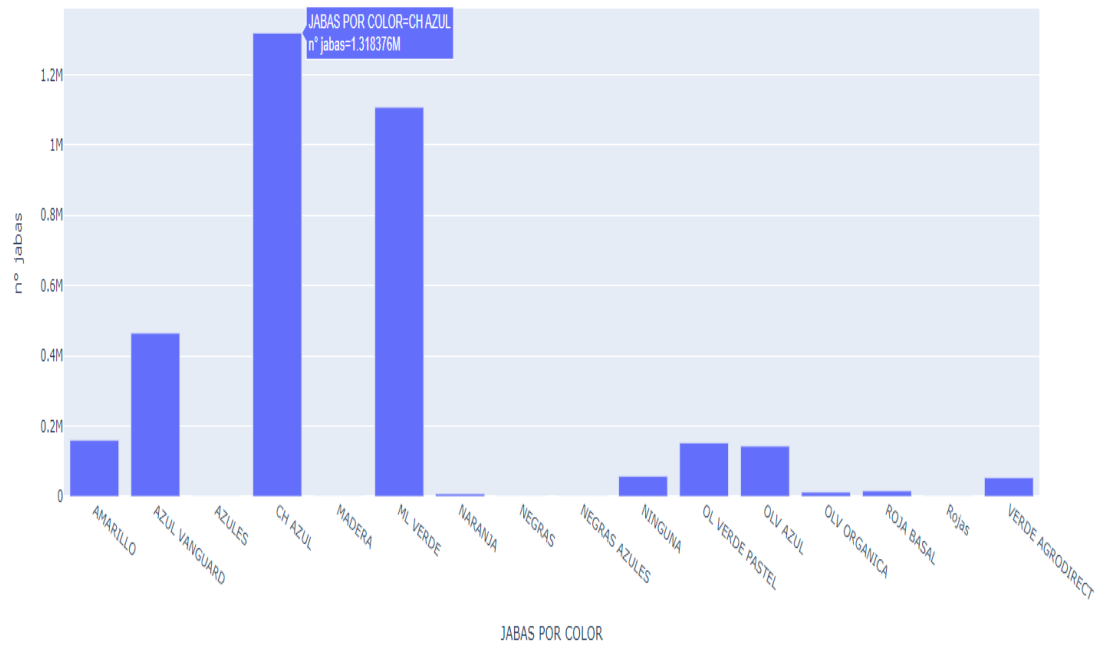


Fig. 26. Gráfico de traslado de Jabas por color

La fig. 26 muestra el desplazamiento de las Jabas según su color, y como se puede comprobar, las Jabas del color CH AZUL, son las que mayormente de trasladan con más de 1 millón de ellas, seguido de las Jabas de color ML Verde y las Jabas de color AZUL VANGUARD.

```
# Creación de gráfico del número según la Observación
```

```
tabla=pd.pivot_table(Jabas, values='n° jabas', index='Observación', aggfunc=np.sum)
```

```
fig=px.bar(tabla, y='n° jabas')
```

```
fig.show()
```

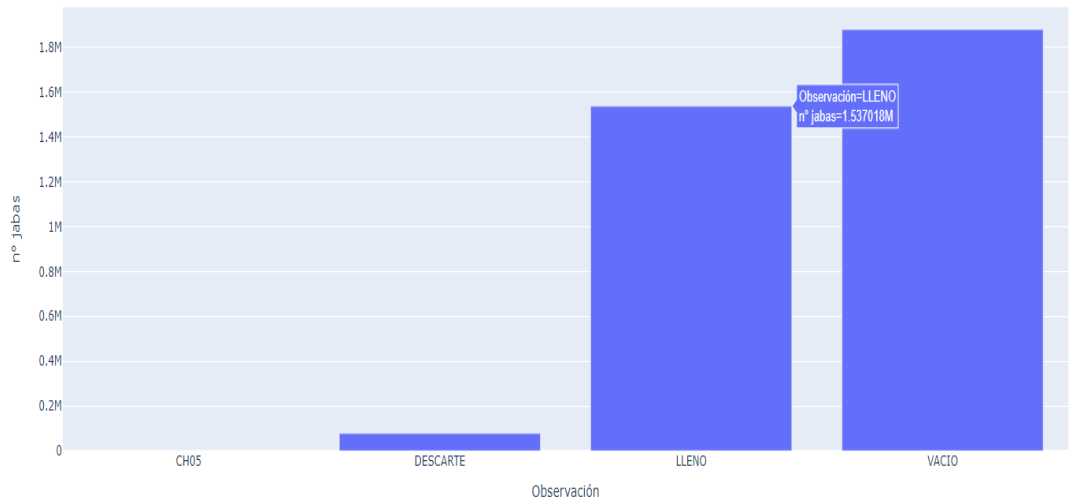


Fig. 27. Gráfico de traslado de Jabas según la observación

Las graficas generadas con la herramienta de plotly, son de muy buena utilidad desde el punto de vista de la interacción que puede realizar el investigador, pero que el investigador debe estar recorriendo las gráficas para evaluar la información investigada y conocer los valores respectivos.

A continuación, se utilizará algunas librerías especializadas para obtener resultados sin necesidad de tener que realizar la programación de diversas sentencias.

Instalar la librería

```
# libreria para EDA
```

```
!pip install ydata-profiling
```

Creación de reportes

```
# cargando libreria y generando reporte
```

```
from ydata_profiling import ProfileReport
```

```
ProfileReport(Jabas)
```

3.2. Calcular la transferencia de las jabs

A continuación, se utilizará algunas librerías especializadas para obtener resultados sin necesidad de tener que realizar la programación de diversas sentencias.

Instalar la librería

```
# libreria para EDA  
!pip install ydata-profiling
```

Creación de reportes

```
# cargando libreria y generando reporte  
from ydata_profiling import ProfileReport  
ProfileReport(Jabas)
```

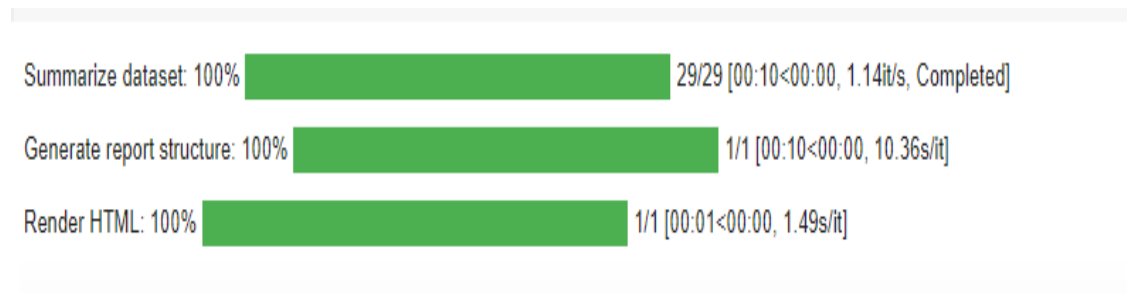


Fig. 28. Reporte generado por la librería

La información generada presenta una serie de opciones para poder evaluar y analizar los datos de los reportes generados, los que han sido generados según se muestra en la Fig. 28.

Overview

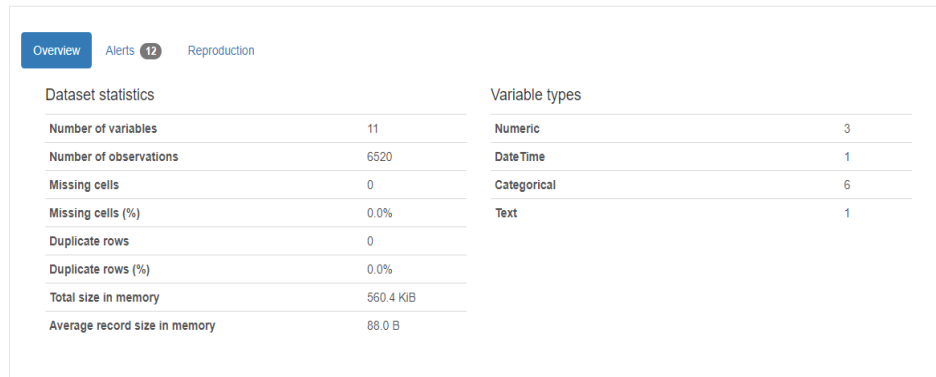


Fig. 29. Estructura del reporte generado por la librería

La fig. 29 muestra la estructura del reporte en la parte superior derecha se destaca en las pestañas los datos a evaluar: Overview, Variables, Interactions, Correlations, Missing values y Sample, los mismos que analizaremos a continuación. Para los efectos del estudio solo se evalúan las dos primeras pestañas.

1. Overview

Overview

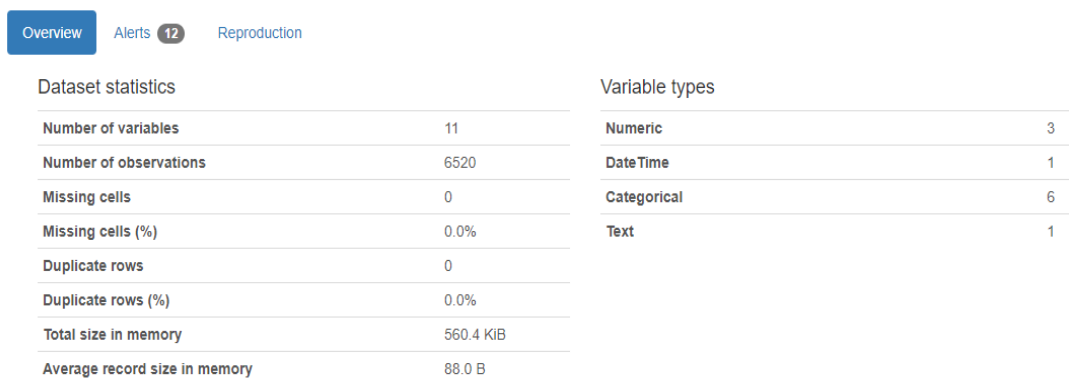


Fig. 30. Estadísticas relacionadas con el dataframe (Dataset)

En el primer reporte se muestra la información sobre el dataframe, relacionados con las variables, cantidad de registros si existen celdas vacías y los tipos de variables e información complementaria.

2. Variables

En esta segunda pestaña, que nos permite hacer la evaluación del traslado de las Jabas, aquí se puede realizar un análisis univariado de cada una de las variables como se presenta en las siguientes gráficas.

Variables

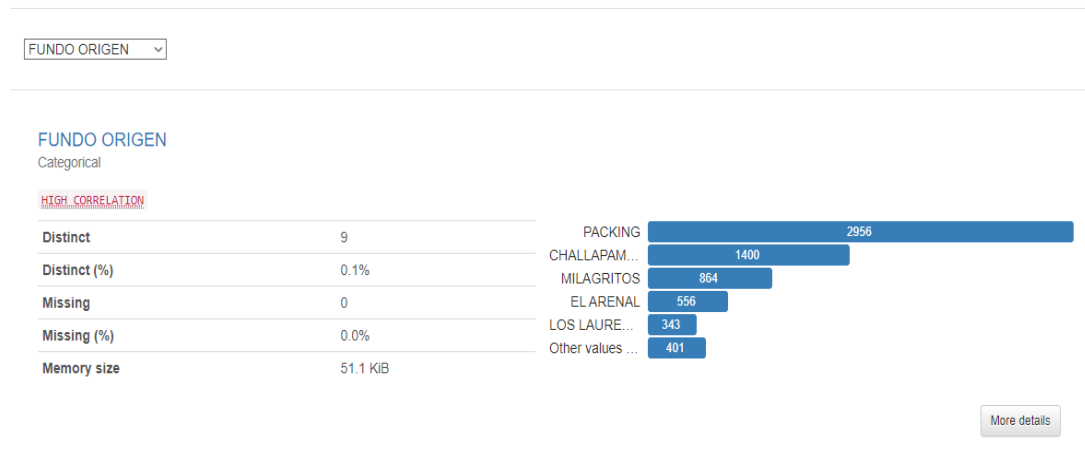


Fig. 31. Estadística de la variable Fondo Origen

En esta pestaña, se puede seleccionar la variable que se desea tener el análisis gráfico, donde se aprecia que existen 9 orígenes que se analizan donde se destaca por la cantidad de jabas despachadas los 5 primeros Fundos, y los otros 4 orígenes por ser menor los valores se agrupan como otros. Para obtener mayor detalle sobre los resultados en las gráficas, se puede visualizar con el botón más detalle. La gráfica muestra las frecuencias de estos despachos y que se muestra en la Fig. 32.



Fig. 32. Frecuencia de las Jabas despachadas

Como se muestra en la Fig. 32, la ubicación desde donde se envían las Jabas esta en el origen de Packing con más del 45% y el Fundo Challapampa con 21.5% y el Fundo Milagritos con 13.3%.

Value	Count	Frequency (%)
PACKING	3564	54.7%
CHALLAPAMPA	1063	16.3%
MILAGRITOS	791	12.1%
EL ARENAL	423	6.5%
LOS LAURELES	390	6.0%
LAS CASUARINAS	194	3.0%
AGRODIRECT	95	1.5%

Fig. 33. Jabas despachadas

En cuanto a los destinos del envío de las Jabas, se tienen según la Fig. 33 que se destina mas del 54% a la zona de Packing, al Fundo Challapampa con 16.3% y al Fundo Milagritos 12.1%.

Value	Count	Frequency (%)
Jabas Cosecheras	5726	87.8%
Descarte	436	6.7%
DIRECTO Descarte	117	1.8%
Parihuelas	79	1.2%
DESCARTE	72	1.1%
Basales	37	0.6%
PARIHUELAS	26	0.4%
DIRECTO DESCARTE	23	0.4%
BASALES	1	< 0.1%
DEscarte	1	< 0.1%
Other values (2)	2	< 0.1%

Fig. 34. Tipos de Jabas

Con relación a los tipos de jabas, en la Fig. 34, se puede apreciar que la mayor proporción de estos traslados son de las Jabas Cosecheras con el 87.8%, y un porcentaje muy bajo de Jabas de descarte.

En la Fig. 34, también se puede apreciar que existe una deficiente estructuración de los datos que deben de estar definidos en los tipos de Jabas. Existe redundancia en los tipos de Jabas como Descarte, DESCARTE, Parihuelas, PARIHUELAS, Directo Descarte, DIRECTO DESCARTE, lo que ocasiona confusión al momento de la evaluación de los traslados.

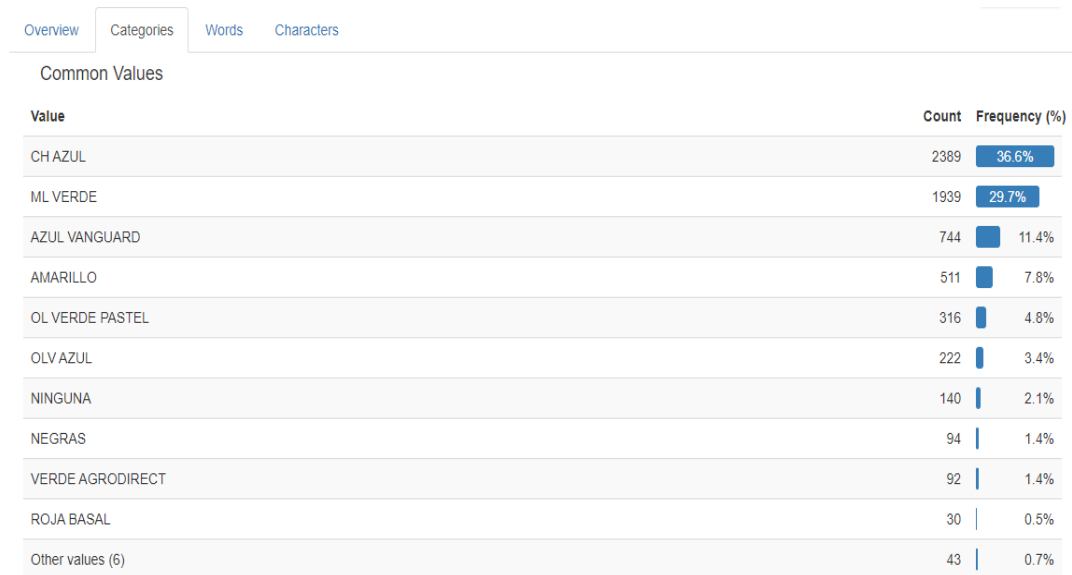


Fig. 35. Jabas según color

En la Fig. 35 se muestra la distribución de las Jabas con relación al color, siendo las Jabas CH AZUL las de mayor distribución con 36.6% y las ML VERDE con 29.7% y en tercer lugar las Jabas AZUL VANGUARD con 11.4%. las mismas que representan aproximadamente el 78% del total.

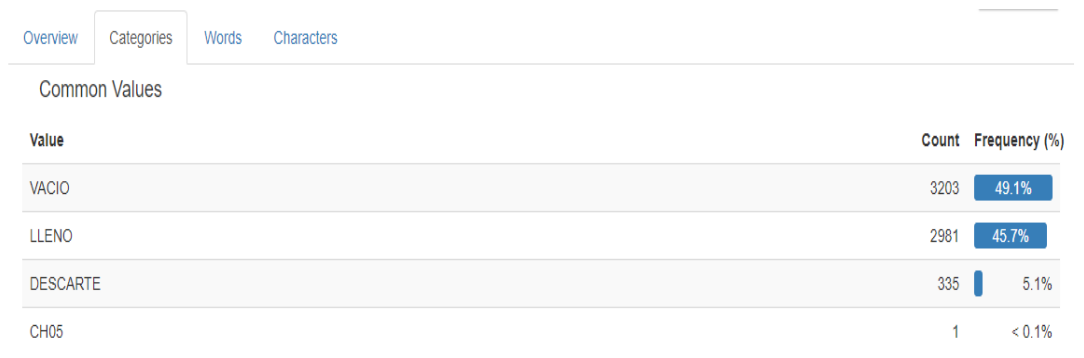


Fig. 36. Jabas según la observación

La variable observación, muestra si las Jabas trasladadas están vacías o llenas. El reporte muestra que las Jabas vacías representan el 49.1%, y las llenas un 45.7%, sin embargo, es bueno tener en consideración que existe un 5.1% de Jabas que pasan a descarte.

3.3. Controlar las jabas

Para controlar la distribución de las Jabas y las diversas variables se tiene una librería que ayuda mucho de manera interactiva a conocer sobre los traslados de estas a los Fondos o a la zona de Packing. Para lograr esto se sigue las siguientes sentencias.

```
# Instalacion de libreria
```

```
!pip install dtale
```

```
# Utilizacion de la libreria
```

```
import dtale
```

```
import dtale.app as dtale_app
```

```
dtale.app.USE_COLAB=True
```

```
# Creación del reporte
```

```
dtale.show(Jabas)
```

La aplicación de las sentencias genera en el sistema un enlace (link) a la información generada:

<https://pd5dl23tjmo-496ff2e9c6d22116-40000-colab.googleusercontent.com/dtale/main/1>, al

acceder al link nos presenta la siguiente información.

D-TALE												
Actions		Visualize	Highlight	Settings								
6520	11	Unnamed: 0	SEM	Fecha	FUNDO ORIGEN	FUNDO DESTINO	Tipo Jaba	JABAS POR COLOR	n° jabas	Observación	LOTE	ESTADO
0	0	46.00	2020-11-11	PACKING	EL ARENAL	Jabas Cosecheras	AZUL VANGUARD	840.00	VACIO	EA04	ENVIADO	
1	1	46.00	2020-11-11	PACKING	EL ARENAL	Jabas Cosecheras	AZUL VANGUARD	840.00	VACIO	EA04	ENVIADO	
2	2	46.00	2020-11-11	PACKING	EL ARENAL	Jabas Cosecheras	AZUL VANGUARD	600.00	VACIO	EA04	ENVIADO	
3	3	46.00	2020-11-11	PACKING	EL ARENAL	Jabas Cosecheras	AZUL VANGUARD	600.00	VACIO	EA04	ENVIADO	
4	4	46.00	2020-11-11	ALM CHALLAPAMPA	PACKING	Basales	ROJA BASAL	720.00	VACIO	0	RECIBIDO	
5	5	46.00	2020-11-11	ALM CHALLAPAMPA	PACKING	Basales	ROJA BASAL	720.00	VACIO	0	RECIBIDO	
6	6	46.00	2020-11-11	PACKING	CHALLAPAMPA	Basales	ROJA BASAL	600.00	VACIO	CH14	ENVIADO	
7	7	46.00	2020-11-11	PACKING	CHALLAPAMPA	Basales	ROJA BASAL	600.00	VACIO	CH14	ENVIADO	
8	8	46.00	2020-11-11	PACKING	CHALLAPAMPA	Jabas Cosecheras	CH AZUL	700.00	VACIO	CH14	ENVIADO	
9	9	46.00	2020-11-11	ALM CHALLAPAMPA	PACKING	Jabas Cosecheras	CH AZUL	700.00	VACIO	0	RECIBIDO	
10	10	46.00	2020-11-11	ALM CHALLAPAMPA	PACKING	Jabas Cosecheras	CH AZUL	700.00	VACIO	0	RECIBIDO	
11	11	46.00	2020-11-11	PACKING	CHALLAPAMPA	Jabas Cosecheras	CH AZUL	700.00	VACIO	CH14	ENVIADO	
12	12	46.00	2020-11-11	PACKING	EL ARENAL	Descarte	AMARILLO	700.00	VACIO	EA04	ENVIADO	
13	13	46.00	2020-11-12	ALM CHALLAPAMPA	PACKING	Jabas Cosecheras	OL VERDE PASTEL	700.00	VACIO	0	RECIBIDO	
14	14	46.00	2020-11-12	PACKING	CHALLAPAMPA	Descarte	AMARILLO	450.00	VACIO	CH14	ENVIADO	
15	15	46.00	2020-11-12	PACKING	EL ARENAL	Basales	ROJA BASAL	240.00	VACIO	EA04	ENVIADO	
16	16	46.00	2020-11-12	ALM CHALLAPAMPA	PACKING	Jabas Cosecheras	OL VERDE PASTEL	700.00	VACIO	0	RECIBIDO	
17	17	46.00	2020-11-12	EL ARENAL	PACKING	Jabas Cosecheras	AZUL VANGUARD	500.00	LLENO	EA04	RECIBIDO	
18	18	46.00	2020-11-12	PACKING	EL ARENAL	Jabas Cosecheras	OL VERDE PASTEL	700.00	VACIO	EA04	ENVIADO	
19	19	46.00	2020-11-12	ALM CHALLAPAMPA	PACKING	Jabas Cosecheras	OL VERDE PASTEL	900.00	VACIO	0	RECIBIDO	
20	20	46.00	2020-11-12	PACKING	EL ARENAL	Jabas Cosecheras	OL VERDE PASTEL	700.00	VACIO	EA04	ENVIADO	

Fig. 37. Reporte generado por la librería dtale

En la Fig. 37 se muestra el resultado de la información generada por la librería dtale, y que, para explorar los datos, se debe dar un clic en la columna que, para realizar el análisis de cualquier columna numérica o categórica, se debe dar un clic en la columna para obtener la información de análisis. La prueba de esta funcionalidad se muestra en la Fig. 38.

SEM	Fecha	FUNDO ORIGEN	FUNDO DESTINO	Tipo Jaba	JABAS POR COLOF
46.00	2020-11-11	PACKING	EL ARENAL		ANGUARI
46.00	2020-11-11	PACKING	EL ARENAL		ANGUARI
46.00	2020-11-11	PACKING	EL ARENAL		ANGUARI
46.00	2020-11-11	PACKING	EL ARENAL		ANGUARI
46.00	2020-11-11	ALM CHALLAPAMPA	PACKING		JA BASA
46.00	2020-11-11	ALM CHALLAPAMPA	PACKING		JA BASA
46.00	2020-11-11	PACKING	CHALLAPAMPA		JA BASA
46.00	2020-11-11	PACKING	CHALLAPAMPA		JA BASA
46.00	2020-11-11	PACKING	CHALLAPAMPA		CH AZU
46.00	2020-11-11	ALM CHALLAPAMPA	PACKING		CH AZU
46.00	2020-11-11	ALM CHALLAPAMPA	PACKING		CH AZU
46.00	2020-11-11	PACKING	CHALLAPAMPA		CH AZU
46.00	2020-11-11	PACKING	EL ARENAL		MARILL
46.00	2020-11-12	ALM CHALLAPAMPA	PACKING		E PASTE
46.00	2020-11-12	PACKING	CHALLAPAMPA		MARILL
46.00	2020-11-12	PACKING	EL ARENAL		JA BASA
46.00	2020-11-12	ALM CHALLAPAMPA	PACKING		E PASTE
46.00	2020-11-12	EL ARENAL	PACKING		ANGUARI
46.00	2020-11-12	PACKING	EL ARENAL		E PASTE
46.00	2020-11-12	ALM CHALLAPAMPA	PACKING		E PASTE
46.00	2020-11-12	PACKING	FI ARFNAI	Jabas Cosecheras	OI VFRDF PASTF

Column "Tipo Jaba"
• Data Type: string

Asc Desc None

Lock

Hide

Delete

Rename

Replacements

Type Conversion

Clean Column

Duplicates

Describe (Column Analysis)

Formats

≠ Aa equals

Select...

0

Fig. 38. Selección de la columna Tipo jaba

Cuando se selecciona la columna para analizar los datos, se despliega una ventana en la que podemos seleccionar la opción **Describe** para realizar el análisis de los datos.

Tipo Jaba (string)
(Use ↑ ↓ buttons to switch columns)

Describe Word Value Counts Value Counts

(Use ← → buttons to switch charts)

- Total Rows: 6,520
 - Count (non-nan): 6,520
 - Count (missing): 0
 - % Missing: 0
- Most Frequent: Jabas Cosecheras
 - Frequency: 5726
- Unique: 12
- Characters
 - Min # Chars: 7
 - Average # Chars: 15.23
 - Max # Chars: 16
 - STD # Chars: 2.31
 - Rows w/ Spaces: 5939
 - Rows w/ Accent Chars: 0
 - Rows w/ Numeric Chars: 0
 - Rows w/ Uppercase Chars: 6520
 - Rows w/ Lowercase Chars: 6397
 - Rows w/ Punctuation: 0
 - Rows Starting w/ Space: 0
 - Rows Ending w/ Space: 73
 - Rows w/ Multi Spacing: 0
 - Rows w/ Hidden Chars: 0
- Words
 - Min # Words: 1
 - Average # Words: 1.91
 - Max # Words: 2
 - STD # Words: 0.28

Unique Row Values (top 100 most common):
 Jabas Cosecheras (5726), Descarte (436), DIRECTO Descarte (117), Parihuelas (79), DESCARTE (72), Basales (37), PARIHUELAS (26), DIRECTO DESCARTE (23), BASALES (1), BASALES (1), DESCarte (1), Descarte (1)

Fig. 40. Opciones de la descripción de la columna Tipo Jaba

La opción describe, presenta tres pestañas para poder analizar los datos, la primera una descripción de los datos de la columna seleccionada, que se muestra en la Fig. 40. Y que para un mejor análisis de las variables se puede acceder a la opción **Open in New Tab** de la parte superior derecha. La misma que los lleva al análisis de la Fig. 41.

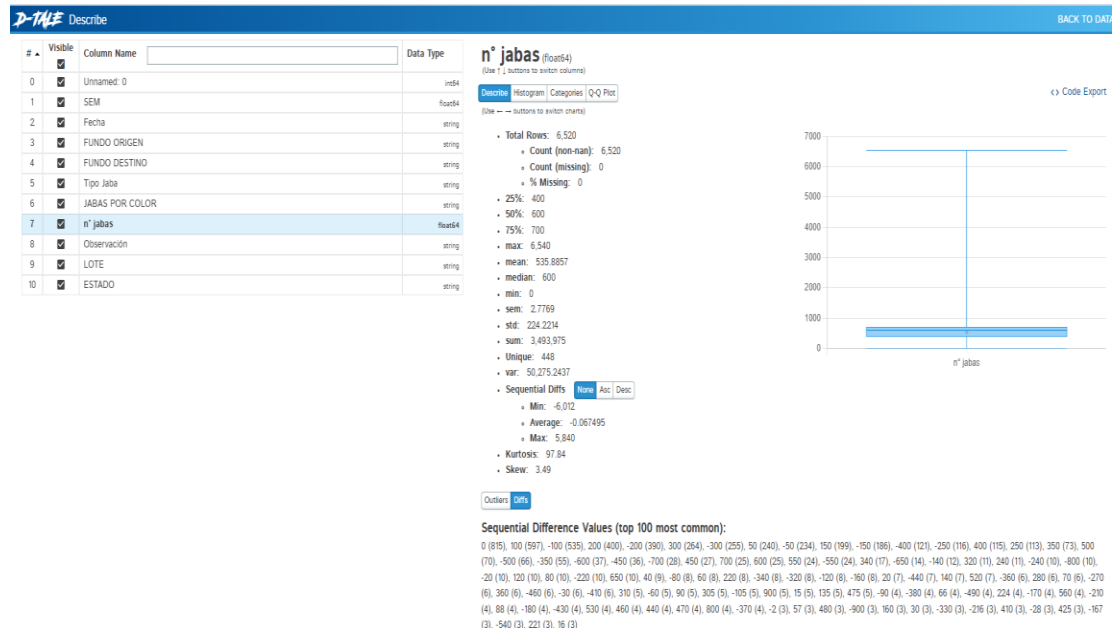


Fig. 41. Ventana para análisis de las variables o datos

En la Fig. 41, en la ventana nueva, se tiene en la parte del lado derecho las variables a analizar y en la parte derecha los gráficos para poder realizar el análisis por el analista de datos. En el caso de variables numérica pueden crear un box plot o cambiar a un histograma como se muestra en la Fig. 42. Asimismo, se puede jugar con las categorías para más gráficos como se muestra en la Fig.43.

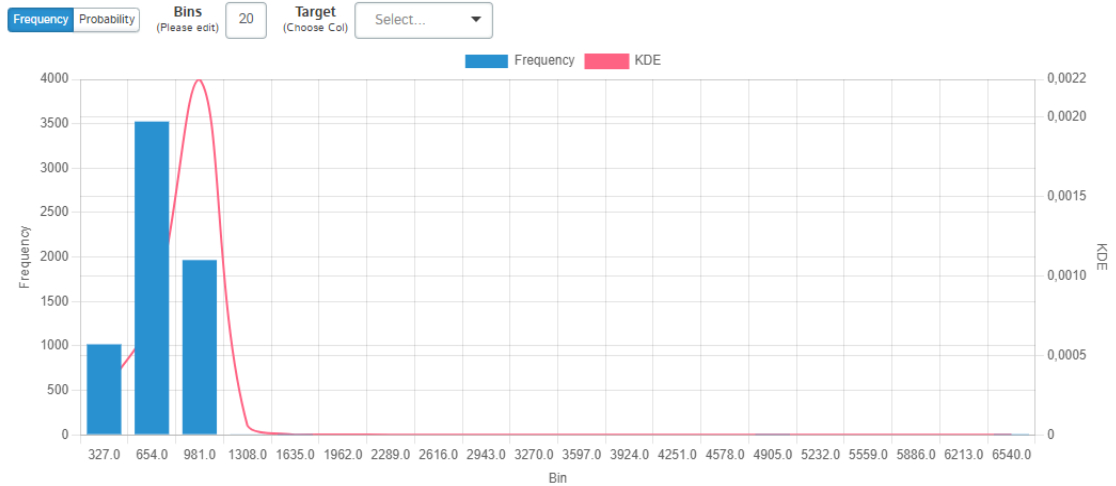


Fig. 42. Gráfico de frecuencias y KDE



Fig. 43. Distribución de Jabas a cada uno de los Fondos.

En la Fig. 43 en el análisis de las Jabas trasladadas a los Fondos, se puede jugar con la operación (para este caso de suma), pero que puede utilizarse otros tipos de operaciones (promedio, desviación estándar, etc.)

IV. DISCUSIÓN

De los resultados obtenidos en el presente trabajo de investigación, en la que se ha logrado obtener resultados de análisis de datos como es el control de Jabas de la empresa Vanguard, Similarmente el estudio de [1] en la que se analiza como la Minería de datos juega un rol importante para la creación de modelos y patrones en el campo de la agroindustria azucarera. En el mismo campo de la agricultura, en la que se evalúan diversas técnicas existentes y cuáles proporcionan mayores beneficios [5]. De igual forma estudios como el de [6] en la que se describe, analiza y aplican Minería de Datos en una base de datos para la agricultura de precisión, para encontrar patrones en el comportamiento de los artrópodos.

El campo de la Minería de datos está creciendo notablemente, sin embargo, para tener un adecuado análisis es necesario seleccionar adecuadamente los algoritmos y librerías adecuadas para tener los resultados esperados, a pesar de analizar el mismo campo de estudio como es el campo agrícola, pero en el que existen una diversidad muy grande de posibilidades; en nuestro caso relacionados con el control de la Jabas de la empresa Vanguard, otros la agricultura de precisión, el control de plagas, etc.

V. CONCLUSIONES

1. En el desarrollo de la tesis, se ha podido comprobar la importancia de la minería de datos, para la evaluación del control de Jabas, en principio se ha evaluado que el registro de la información del archivo en Excel no es el adecuado, existen datos en algunas columnas que no son coherentes con lo que deben registrar.
2. La caracterización de las jabas por medio de estadísticas descriptivas, para los valores numéricos por medio de una sola sentencia facilita la comprensión del traslado de la Jabas, y en cuanto a los valores categóricos la librería para gráficos interactivos con plotly, permite una mayor comprensión de los traslados de las Jabas.
3. En cuanto al calculo del traslado de la Jabas entre los Fondos y la zona de Packing fue de mucha ayuda la librería de ydata-profiling para generar la tabla de frecuencias de los movimientos de estas Jabas.
4. La librería dtale para el control de los traslados de la Jabas es de suma importancia además de ser muy fácil de obtener, solo requiere de pocas líneas de sentencias del lenguaje y el generador de reportes interactivo permite que el usuario sin experiencia pueda tener un control de las variables del movimiento de las Jabas entre los diversos fondos y la zona de Packing.
5. Finalmente se concluye que las librerías utilizadas como ydata-profiling y dtale permiten ahorrar enorme cantidad de tiempo, al generar los reportes para análisis del control de las Jabas con pocas líneas de código, pero con resultados muy grandes.

VI. RECOMENDACIONES (SOLO SI SON PROCEDENTES)

El análisis descriptivo por medio de la Minería de datos en el control de Jabas, con los resultados gráficos correspondientes, nos permite recomendar que para tener un adecuado control de cualquier tipo de procesos, se debe planificar rigurosamente la estructura de los datos que se deben registrar, ya que en el estudio se demuestra que no hay una buena estructuración de los datos. de otro lado e importante es aprovechar las librerías de mayor eficiencia para este tipo de análisis, y evitar el uso de muchas líneas de código, cuando se pueden utilizar estas librerías que simplifican la labor de evaluación y análisis.

VII. REFERENCIAS BIBLIOGRÁFICAS

- [1] O. Gozá-León, «Aplicaciones de la Minería de datos en la agroindustria azucarera de caña», *ResearchGate*, vol. 55, n.º 3, p. 13, 2022.
- [2] J. I. Salazar Torres y E. Girón Cardenas, «Análisis y aplicación de algoritmos de minería de datos», *Rev. Perspect.*, vol. 6, n.º 21, pp. 71-88, dic. 2021, doi: 10.26620/uniminuto.perspectivas.6.21.2021.71-88.
- [3] M. I. Uvidía Fassler, A. S. Cisneros Barahona, P. M. Méndez Naranjo, y H. M. Villa Yáñez, «Minería de datos para la toma de decisiones en la unidad de nivelación y admisión universitaria ecuatoriana», *Cumbres*, vol. 4, n.º 2, pp. 55-67, feb. 2019, doi: 10.48190/cumbres.v4n2a5.
- [4] Y. Torres-Quezada, «Minería de datos para determinar los factores más influyentes en la ocurrencia de siniestros de tránsito en Ecuador en el año 2020», *CEDAMAZ*, vol. 11, n.º 2, pp. 124-132, dic. 2021, doi: 10.54753/cedamaz.v11i2.1181.
- [5] A. F. Gil-Torres, A. L. Monroy-García, y J. S. González-Sanabria, «Minería de datos espacial en la agricultura en Latinoamérica - Una aproximación conceptual», *Pensam. Acción*, n.º 28, pp. 19-33, dic. 2019, doi: 10.19053/01201190.n28.2020.10976.
- [6] M. Droz, C. Alves, G. Etchart, y B. Diaz, «Aplicación de minería de datos e Internet de las Cosas (IoT) para investigaciones agropecuarias sobre artrópodos», *SEDICE*, vol. 6, n.º 1, pp. 383-397, 2019.
- [7] A. Gallo, F. Pérez, y D. Salinas, «Minería de Datos y Proyección a Corto Plazo de la Demanda de Potencia en el Sistema Eléctrico Ecuatoriano», *Rev. Téc. Energ.*, vol. 18, n.º 1, pp. 72-85, jul. 2021, doi: 10.37116/revistaenergia.v18.n1.2021.461.
- [8] M. P. Orellana Cordero y P. Cedillo, «Detección de valores atípicos con técnicas de minería de datos y métodos estadísticos», *Enfoque UTE*, vol. 11, n.º 1, pp. 56-67, ene. 2020, doi: 10.29019/enfoque.v11n1.584.
- [9] A. I. Oviedo Carrascal y J. Jiménez Giraldo, «Minería de datos educativos: Análisis del desempeño de estudiantes de ingeniería en las pruebas SABER-PRO», *Rev. Politécnica*, vol. 15, n.º 29, pp. 128-140, jul. 2019, doi: 10.33571/rpolitec.v15n29a10.
- [10] C. C. Russo, «Minería de datos aplicada a estrategias para minimizar la deserción universitaria en carreras de Informática de la UNNOBA», *Rev. Iberoam. Tecnol. En Educ. Educ. En Tecnol.*, n.º 24, p. e12, dic. 2019, doi: 10.24215/18509959.24.e12.
- [11] T. T. Pham, G. A. Lobos, y C. L. Vidal-Silva, «Innovación en Minería de Datos para el Tratamiento de Imágenes: Agrupamiento K-media para Conjuntos de Datos de Forma

- Alargada y su Aplicación en la Agroindustria», *Inf. Tecnológica*, vol. 30, n.º 2, pp. 135-142, mar. 2019, doi: 10.4067/S0718-07642019000200135.
- [12] B. Marín y L. López, «APLICACIÓN DE LA TÉCNICA DE MINERÍA DE DATOS PARA LA EXPLORACIÓN DE CONOCIMIENTO IMPLÍCITO Y PREDICCIÓN DEL CLIMA», *Sapientia*, vol. 9, n.º 17, pp. 38-43, 2017, doi: <https://doi.org/10.54278/sapientia.v9i17.27>.
- [13] L. M. Dicovskiy Riobóo y M. E. Pedroza, «CONSTRUCTOS QUE HAN INCIDIDO EN EL RENDIMIENTO ACADÉMICO EN LA CARRERA DE INGENIERÍA AGROINDUSTRIAL EN LA SEDE REGIONAL UNI NORTE», *Rev. Científica Cienc. Tecnol. El Higo*, vol. 10, n.º 1, p. 11, 2020.
- [14] M. Droz, C. Alves, G. Etchart, y B. Diaz, «Aplicación de minería de datos e Internet de las Cosas (IoT) para investigaciones agropecuarias sobre artrópodos», *SEDICE*, vol. 6, n.º 1, pp. 383-397, 2019.
- [15] E. Erazo Ortiz y S. Mora Portilla, «APLICACIÓN DE LA MINERÍA DE DATOS PARA LA DETECCIÓN DE FACTORES ASOCIADOS A LA DESERCIÓN ESTUDIANTIL EN LOS PROGRAMAS PROFESIONALES DE LAS CIENCIAS NATURALES Y CIENCIAS SOCIALES DE LA UNIVERSIDAD DE NARIÑO SEDE PASTO», UNIVERSIDAD TECNOLÓGICA DE PEREIRA, 2018.
- [16] L. H. Fernández Rojas, «APLICACIÓN DE TÉCNICAS DE MINERÍA DE DATOS PARA PRONÓSTICO DE PRODUCCIÓN DE ESPÁRRAGOS», Universidad Señor de Sipan, 2019.
- [17] A. F. Gil-Torres, A. L. Monroy-García, y J. S. González-Sanabria, «Minería de datos espacial en la agricultura en Latinoamérica - Una aproximación conceptual», *Pensam. Acción*, n.º 28, pp. 19-33, dic. 2019, doi: 10.19053/01201190.n28.2020.10976.
- [18] O. Gozá-León, «Aplicaciones de la Minería de datos en la agroindustria azucarera de caña», *ResearchGate*, vol. 55, n.º 3, p. 13, 2022.
- [19] B. Marín y L. López, «APLICACIÓN DE LA TÉCNICA DE MINERÍA DE DATOS PARA LA EXPLORACIÓN DE CONOCIMIENTO IMPLÍCITO Y PREDICCIÓN DEL CLIMA», *Sapientia*, vol. 9, n.º 17, pp. 38-43, 2011, doi: <https://doi.org/10.54278/sapientia.v9i17.27>.
- [20] T. Paola y A. Rozas, «PATRONES DE COMPORTAMIENTO EN EL USO DE LAS AULAS VIRTUALES DE LA UNIVERSIDAD NACIONAL DEL ALTIPLANO ÁREA DE INGENIERÍAS UTILIZANDO TÉCNICAS DE MINERÍA DE DATOS», *Rev. Investig. Esc. Posgrado*, vol. 9, n.º 4, pp. 229-243, 2020.
- [21] T. T. Pham, G. A. Lobos, y C. L. Vidal-Silva, «Innovación en Minería de Datos para el Tratamiento de Imágenes: Agrupamiento K-media para Conjuntos de Datos de Forma

- Alargada y su Aplicación en la Agroindustria», *Inf. Tecnológica*, vol. 30, n.º 2, pp. 135-142, mar. 2019, doi: 10.4067/S0718-07642019000200135.
- [22] I. Poma, A. Eccoña, y R. Silva, «Análisis de quejas y reclamos reportados en el servicio de atención al cliente de una empresa elaboradora de galletas», *Agroindustrial Sci.*, vol. 2, n.º 1, pp. 147-154, 2021, doi: <http://dx.doi.org/10.17268/agroind.sci.2021.01.03>.
- [23] J. I. Salazar Torres y E. Girón Cardenas, «Análisis y aplicación de algoritmos de minería de datos», *Rev. Perspect.*, vol. 6, n.º 21, pp. 71-88, dic. 2021, doi: 10.26620/uniminuto.perspectivas.6.21.2021.71-88.
- [24] R. Sanchez, «Clasificación no supervisada de imágenes médicas y minería de datos. Algoritmo S3 vs. K-medias», *Rev. Cuba. Investig. Bioméd.*, vol. 40, n.º supl, p. 19, 2021.
- [25] V. Solis Sandoval, «Análisis del comportamiento del precio de la uva de mesa sonoreña en Estados Unidos mediante estrategias de minería de datos», Unidad de Sonora-México, 2022.
- [26] R. Hernández, C. Fernández, y P. Baptista, *Metodología de la Investigación*. México, 2014.
- [27] Simplilearn, «Data Science In 5 Minutes | Data Science For Beginners | What Is Data Science? | Simplilearn», *Youtube*, 2018.
<https://www.youtube.com/watch?v=X3paOmcrTjQ&t=21s> (accedido 4 de noviembre de 2022).
- [28] Tuinsti, «Introducción a Minería de Datos», *Youtube*, 2020.
<https://www.youtube.com/watch?v=fteVv1wrztE> (accedido 4 de noviembre de 2022).

VIII. ANEXOS

ANEXO 1. Certificado De trabajo Empresa Vanguard



CERTIFICADO DE TRABAJO

LOS OLIVOS DE VILLACURI S.AC. hace constar:

Que el señor, **GUERRA VERA LUIS PIERRE DE JESUS**, identificado con DNI **70314770**, ha laborado para nuestra empresa desde el 09 de noviembre del 2020 hasta el 24 de febrero del 2021, desempeñándose como **GESTOR DE INFORMACION Y DISTRIBUCION DE JABAS .**

Se expide la presente a solicitud del interesado, para los fines que crea conveniente.

Ica, 09 de marzo del 2021

LOS OLIVOS DE VILLACURI S.A.C.

Soledad Vargas Vera
JEFE DE RECURSOS HUMANOS

ANEXO 2. Fondos grupo Vanguard Perú

+



ANEXO 3. Imagen jabs vacías

