



Inteligencia de negocios

APLICACIÓN DE INTELIGENCIA DE NEGOCIOS AL ANALISIS DE
INVERSION DE UN ESTABLECIMIENTO AGRÍCOLA PARA LA
PRODUCCION DE GRANOS DE SOJA.

Alumnos: Luca Mitas – Lucas Singh

Contenido

Resumen	3
Objetivos generales:.....	3
Objetivos específicos:.....	3
Motivación de este proyecto	3
Orígenes de datos	3
Ubicación Geográfica	4
Distribución de lotes	4
Lote 1	4
Lote 2	4
Lote 3	5
Lote 4	5
Lote 5	5
Lote 6	6
Herramientas elegidas	7
Herramienta para tratamiento de imágenes	7
Herramienta para el proceso ETL.....	7
Herramienta para el almacenamiento (datawarehouse)	7
Herramienta para la generación de informes.	7
Proceso ETL.....	7
Extracción	7
Imágenes:	7
Archivos CSV	7
Transformación	8
Transformaciones de las imágenes:	8
Transformaciones de filtrado:	8
Transformaciones de formato:	8
Carga	8
Carga de información en archivos Excel:.....	8
Carga de datos en base de datos:	8
Carga de datos en Google Data Studio:.....	8
Base de datos relacional	9
Modelo multidimensional.....	10
Dimensiones	10

Tabla provincia:	10
Tabla departamento	10
Hechos	10
Tabla siembracosecha	10
Base de datos relacional para las imágenes	10
Informes	11
Rendimiento por lote	11
Lote 1	11
Lote 2	11
Lote 3	12
Lote 4	12
Lote 5	13
Lote 6	13
Promedio rendimiento lotes	14
Rendimiento promedio del establecimiento	14
Rendimiento República Argentina	15
Rendimiento Departamento General López	15
Rendimiento Provincia de Santa Fe	15
Zonas o departamentos cercanos al establecimiento analizado	15
Conclusiones	16
Conclusiones del modelo planteado:	16
Conclusiones técnicas y sobre la materia:	16

Resumen

El presente trabajo, busca recrear un escenario en el que un emprendedor realiza un análisis de rentabilidad de un campo para la siembra de soja y así poder presentárselo a diversos inversores.

El indicador que servirá para realizar este análisis, es el rendimiento expresado en toneladas (Tn.) por hectárea, y a su vez, se intentará comparar este indicador con el rendimiento de una zona geográfica concreta, en nuestro caso de estudio, el departamento General López ubicado en la provincia de Santa Fe, (zona sojera por excelencia).

Objetivos generales:

Poder determinar, si es viable o no el arrendamiento y posterior siembra con soja de un establecimiento agrícola analizado para la campaña 2021-2022.

Objetivos específicos:

- Obtener el rendimiento de cada área coloreada y asignarle un valor para conseguir el rendimiento promedio del lote, para luego realizar el promedio entre lotes y conseguir el rendimiento promedio del establecimiento en general.
- Realizar el análisis a nivel regional (Departamento General López) con los datos históricos de campañas anteriores, para compararlo con el rendimiento del establecimiento analizado.
- Realizar el análisis a nivel provincial (Provincia de Santa Fe) con los datos históricos de campañas anteriores, para compararlo con el rendimiento del establecimiento analizado.

Motivación de este proyecto

Este proyecto surge de la necesidad de presentar de una manera accesible pero no menos profesional, la información necesaria para poder tomar decisiones en lo que a inversiones agrícolas se refiere, si bien, la inteligencia de negocios es muy utilizada para la agricultura por medio de la agricultura de precisión, los reportes de monitores de maquinarias agrícolas (cosechadoras, sembradoras, pulverizadoras) son demasiado técnicos o se requiere personal calificado para el análisis y la posterior toma de decisiones. Es por eso que este proyecto busca utilizar la misma información, pero que el asesor pueda mostrarla de una manera más amigable a las partes interesadas.

Es importante destacar, que no se busca reemplazar los reportes técnicos, que son esenciales para la parte operativa y que arrojan una gran precisión, si no que este análisis pueda servir como complemento a estos.

Otra motivación importante fue lograr una introducción práctica a la aplicación de inteligencia de negocios sobre una fuente de datos y demanda real, aprendiendo el uso de herramientas informáticas de BI para su correcta manipulación y manejo.

Orígenes de datos

La evaluación y posterior análisis de la información se ha realizado sobre la salida de un monitor de una cosechadora John Deere S700. La salida de este monitor, son imágenes con extensión .jpg que muestran mapas de color mostrando distintos colores para los diferentes rangos de rendimientos de una determinada superficie (lote) expresada en hectáreas.

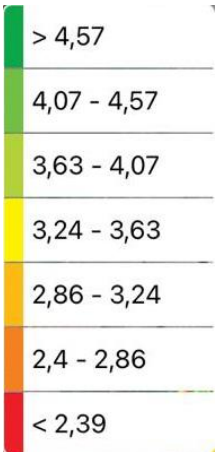
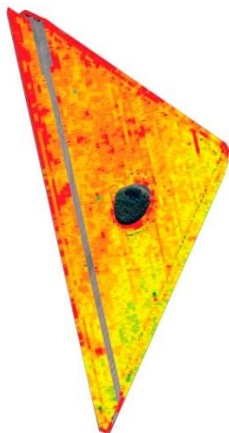
Ubicación Geográfica

Por motivos de confidencialidad se preservaron los datos del nombre del establecimiento, la ubicación precisa y la superficie.

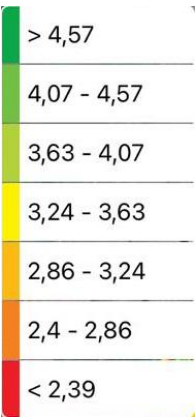
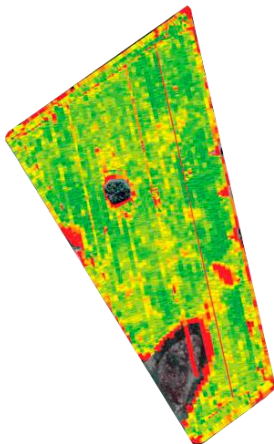
Distribución de lotes

Dicho establecimiento, está dividido en seis lotes que se detallan a continuación.

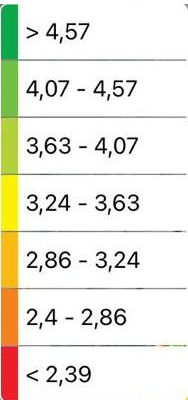
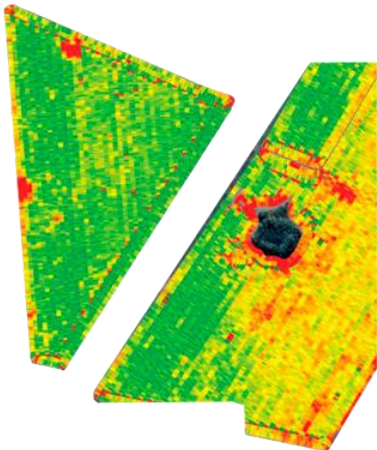
Lote 1



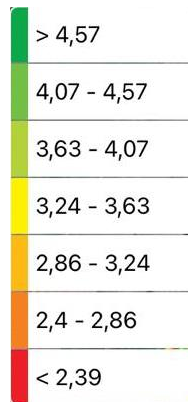
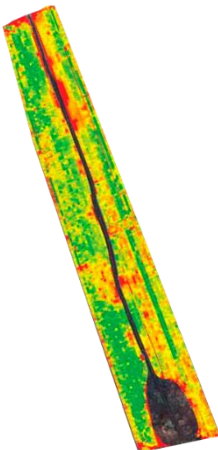
Lote 2



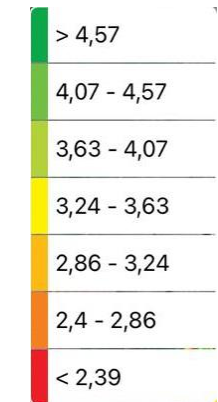
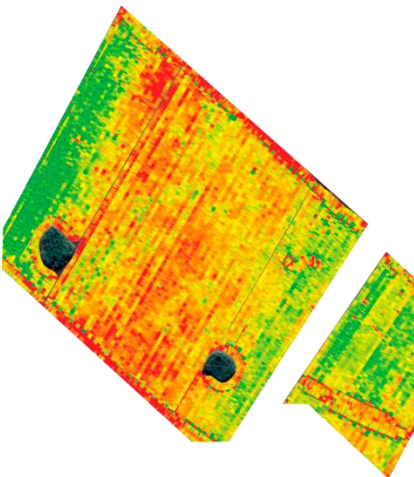
Lote 3



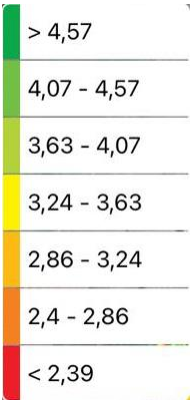
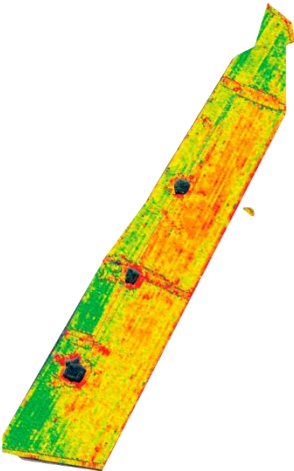
Lote 4



Lote 5



Lote 6



Herramientas elegidas

Herramienta para tratamiento de imágenes

Se desarrolló un script en Python que permite leer los píxeles coloreados totales y los píxeles que están coloreados asociados al color y lote correspondiente.

Herramienta para el proceso ETL

Para realizar todo el proceso de ETL se utilizó Pentaho Data Integration

<https://www.hitachivantara.com/es-latam/products/data-management-analytics/pentaho-data-integration.html>

Herramienta para el almacenamiento (datawarehouse)

Para realizar el storage de los datos normalizados y procesados se utilizó el gestor de base de datos MySQL (base de datos relacional) con MySQL workbench como herramienta grafica para administrar y gestionar la base de datos. La base de datos se instaló en un hosting privado.

Herramienta para la generación de informes.

Para mostrar los informes se utilizó la herramienta Google Data Studio.

<https://datastudio.google.com/>

Proceso ETL

Extracción

Imágenes:

- **Imágenes de mapas de rendimiento:** Son 6 imágenes .jpg que muestran los lotes pertenecientes a un establecimiento agrícola. Cada imagen es un mapa de color que muestra una carta con rangos de rendimientos.

Archivos CSV

- **Estimaciones-agricolas-2020-08:** Este archivo corresponde a un dataset descargado de la secretaria de Agricultura Ganadería y Pesca de la República Argentina. Cuenta con 143630 registros. Este dataset nuclea información todas las provincias, separado por departamentos, por cultivo, por campaña, la superficie sembrada/cosechada, la producción y el rendimiento.
<https://datos.agroindustria.gob.ar/dataset/estimaciones-agricolas/archivo/95d066e6-8a0f-4a80-b59d-6f28f88eacd5>
- **Análisis de color:** Este archivo CSV contiene la cantidad de píxeles que tiene cada área de color asociadas a cada lote. Es la salida del análisis que realiza el script en Python. Arroja como información, el id del lote, la cantidad de píxeles totales (superficie total) píxeles pintados (cantidad de píxeles de cada color) y el color asociado.

Transformación

Transformaciones de las imágenes:

- **Pasar de CSV a la base de datos la información de cada lote:** Se transformó el CSV obtenido con los datos en crudo de las cantidades de pixeles totales de cada lote y de cada cantidad de pixeles por color para cargarlos en dos tablas en la base de datos, la tabla lote, que contiene el id del lote, el nombre del lote y el tamaño. La otra tabla, datos lote, se conecta por el id lote y tiene la cantidad de pixeles coloreados y el color asociado.
- **Calculo del rendimiento de cada lote:** Se utilizó la transformación cargar desde tabla, para traer la información de los lotes y poder realizar los cálculos de cada área que representa cada color en porcentaje y luego realizar los cálculos utilizando un valor fijo para obtener el rendimiento en Tn.

Transformaciones de filtrado:

- **Filtrado por provincias:** Esta transformación realiza el filtro por provincia (id_provincia).
- **Filtrado por departamentos:** Esta transformación es similar a la anterior, pero en este caso se filtra por el campo departamento (id_departamento).
- **Filtrado por campaña país:** Esta transformación filtra por el id campaña (id_campaña) a nivel país.
- **Filtrado por campaña provincia:** Esta transformación filtra por el id campaña (id_campaña) a nivel provincia.

Transformaciones de formato:

- **Formato de campo campañas:** Esta transformación realiza mediante la opción cutString cambiar el formato del año de campaña de entrada, por ejemplo, en 1998-99 en 1998 quitando el -99
- **Formato de renombrar campo:** Esta transformación permite renombrar un campo, en este caso por ejemplo id_campaña que tiene un carácter especial (ñ) se decidió cambiarlo por idc para evitar conflictos al cargarlos en la base de datos.

Carga

Carga de información en archivos Excel:

- La salida de cada transformación de cada lote, se obtuvo en un archivo Excel que contiene el porcentaje y rendimiento de cada lote para después poder cargarlos en Google data Studio y crear los reportes correspondientes.

Carga de datos en base de datos:

- **Salida a tabla:** La carga de datos se realizó a través de la opción salida a tabla, que contiene la conexión al servidor para la base de datos y que de acuerdo al modelo planteado (ver base de datos y modelo multidimensional) se cargara en la base y tabla correspondiente.

Carga de datos en Google Data Studio:

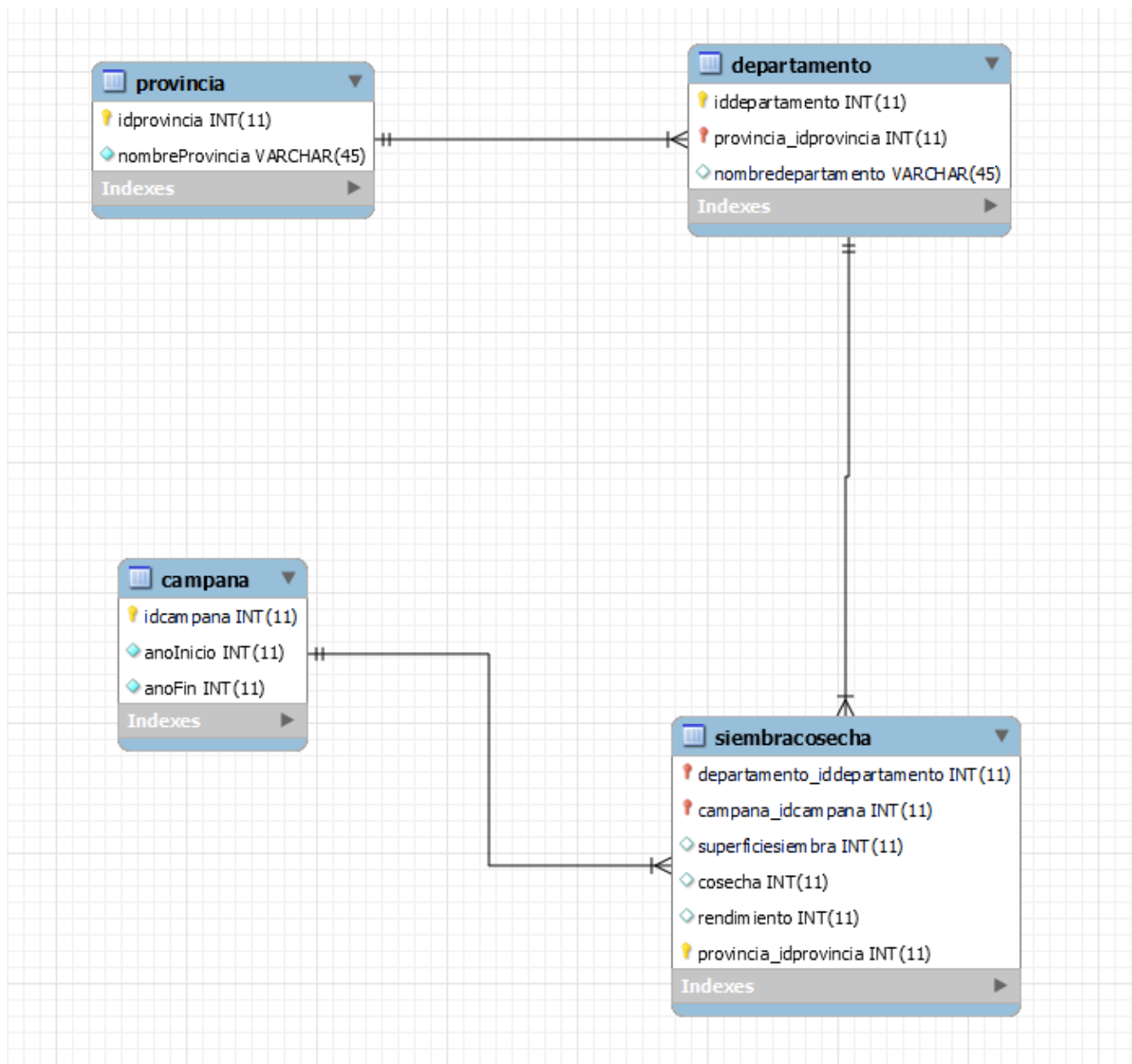
- **Salida a tabla:** Al tener la base de datos alojada en un servicio de hosting pago, este provee un conector MySQL para integrarlo con Google Data Studio y poder utilizar la información en forma de vistas para luego poder crear los informes.

Base de datos relacional

Luego de los filtrados realizados en las transformaciones mencionadas anteriormente, se crearon las tablas correspondientes en donde se cargó la información.

La tabla de provincia, guarda todas las provincias, departamentos los departamentos con la relación de idprovincia, la tabla campana, carga las campañas y tabla siembracosecha genera toda la información que necesitamos relacionadas con las tablas anteriores.

El esquema de la base de datos relacional se puede observar en la siguiente imagen



Modelo multidimensional

Dimensiones

Tabla provincia:

Idprovincia PK	int
nombreProvincia	varchar

Tabla departamento

Iddepartamento PK	int
Provincia_idprovincia FK	Int
nombreddepartamento	varchar

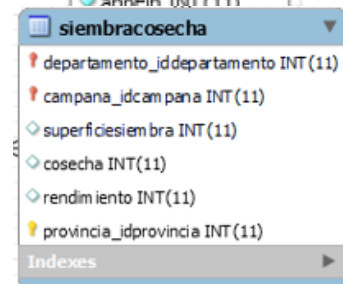
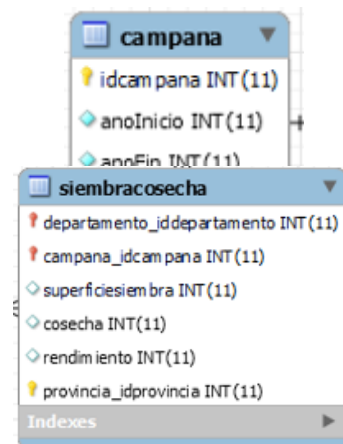
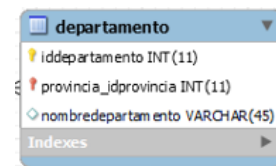
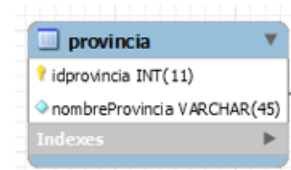
Tabla campaña

Idcampana PK	int
anoInicio	Int
anoFin	varchar

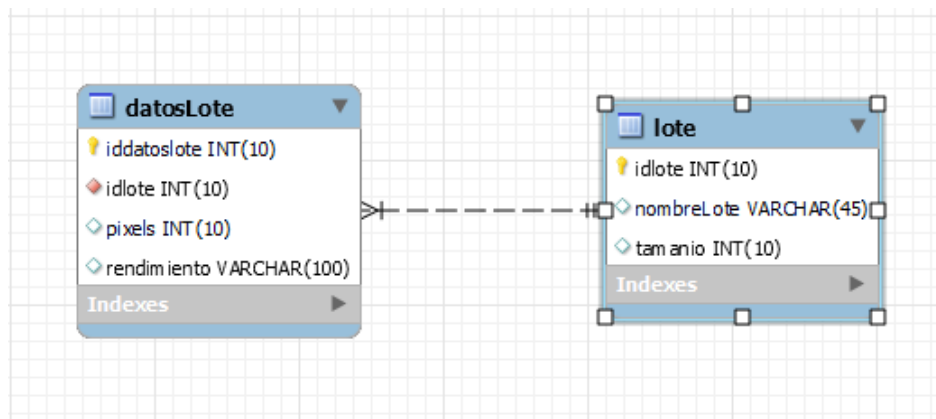
Hechos

Tabla siembracosecha

Departamento_iddepartamento	int
Campana_id_campana	Int
superficiesiembra	varchar
Cosecha	int
rendimiento	int
Provincia_idprovincia	int



Base de datos relacional para las imágenes



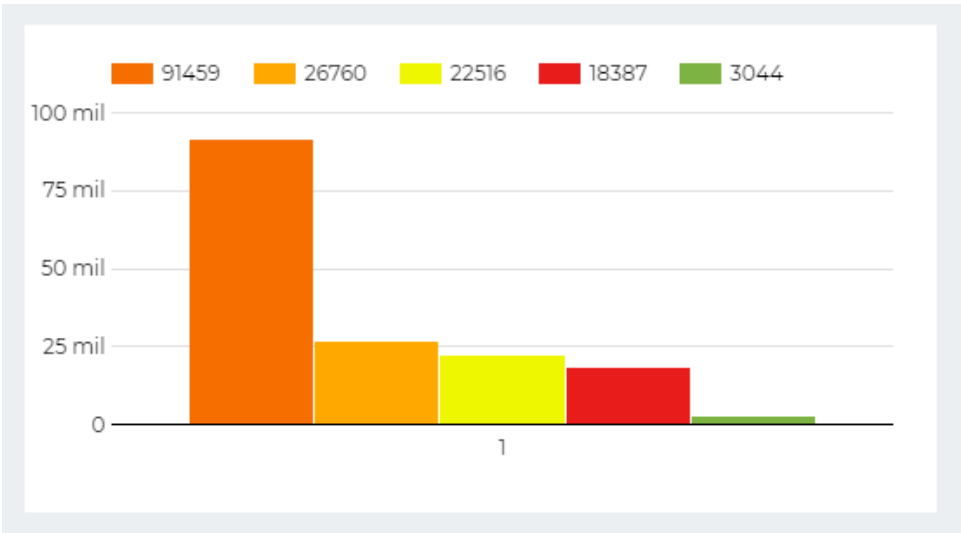
Informes

El informe completo se puede observar en:

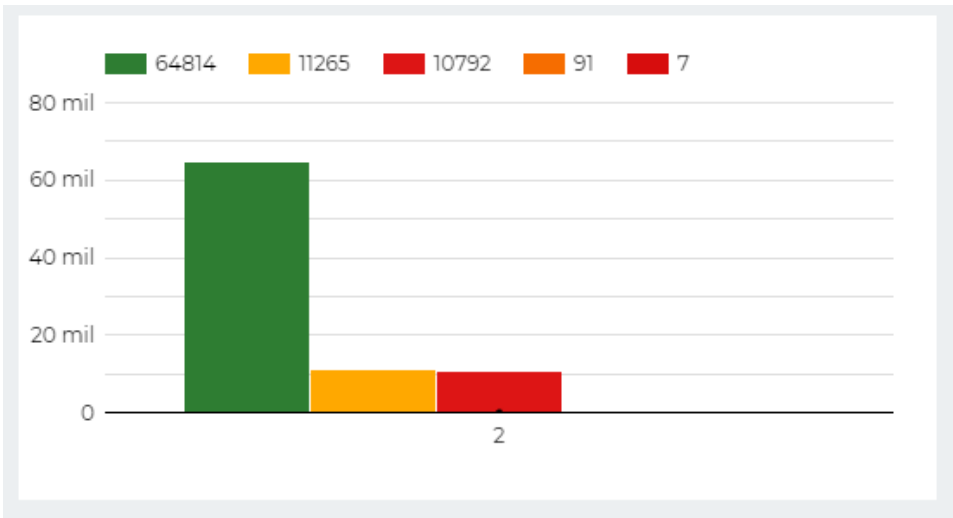
<https://datastudio.google.com/reporting/c44fcd33-eea2-4d78-aede-1917031931af>

Rendimiento por lote

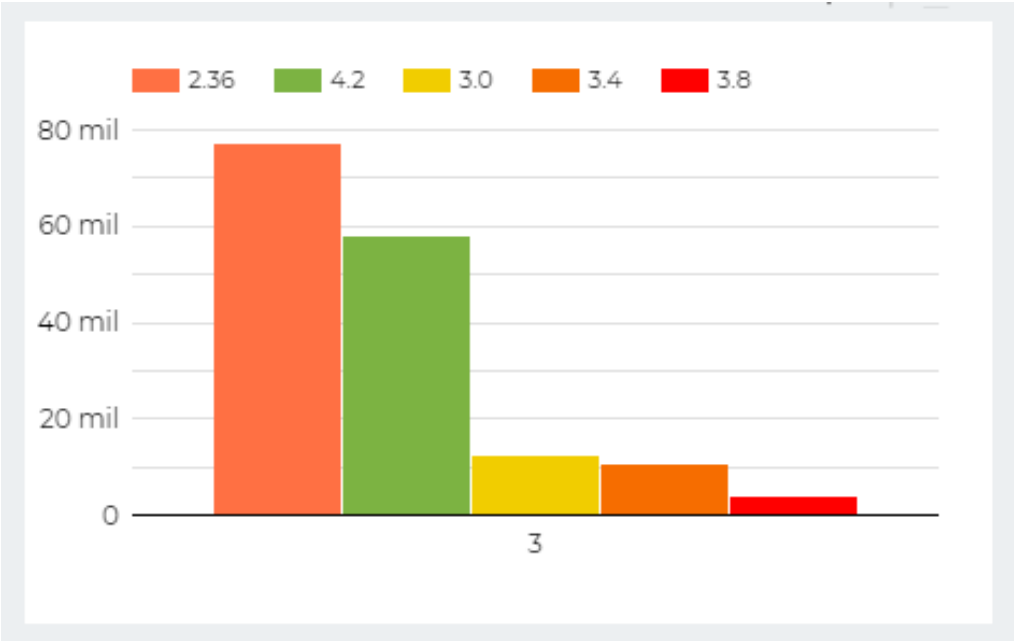
Lote 1



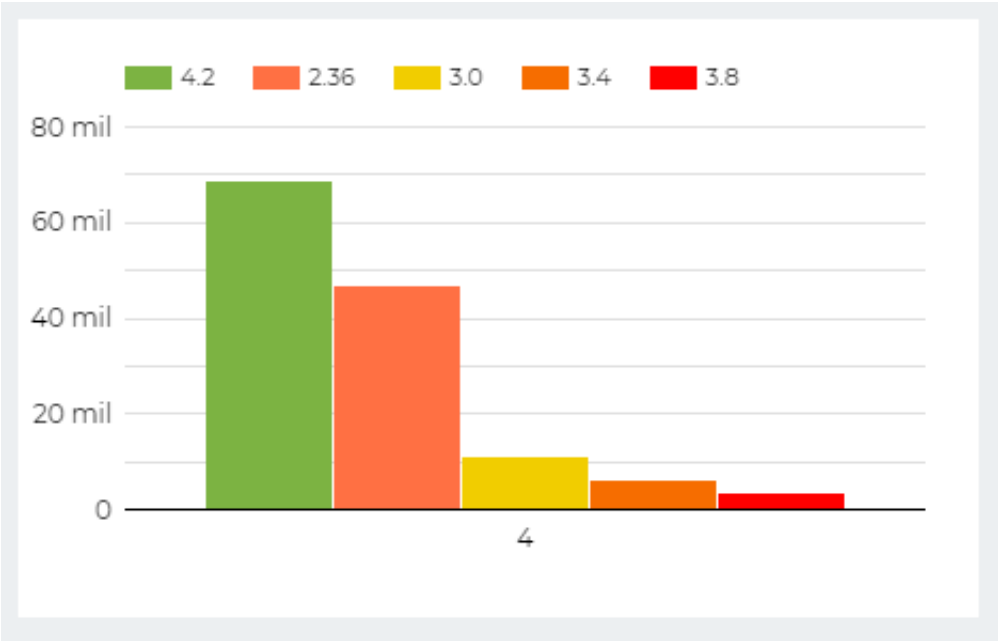
Lote 2



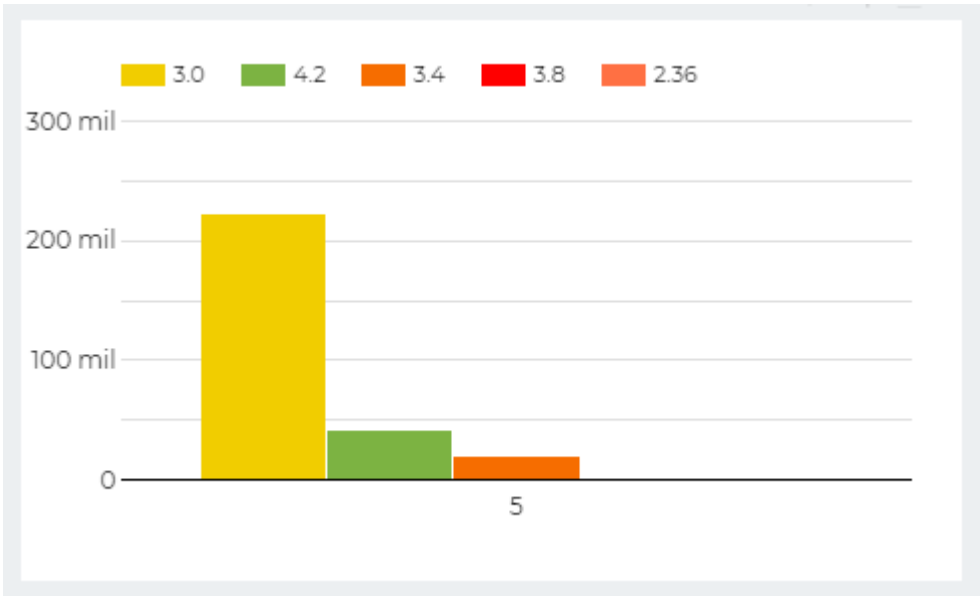
Lote 3



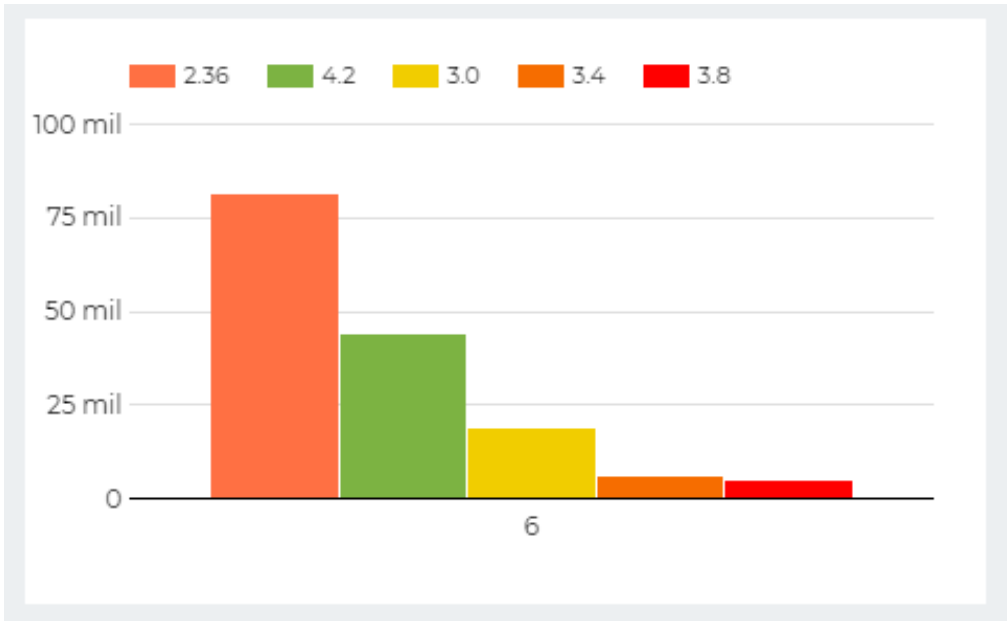
Lote 4



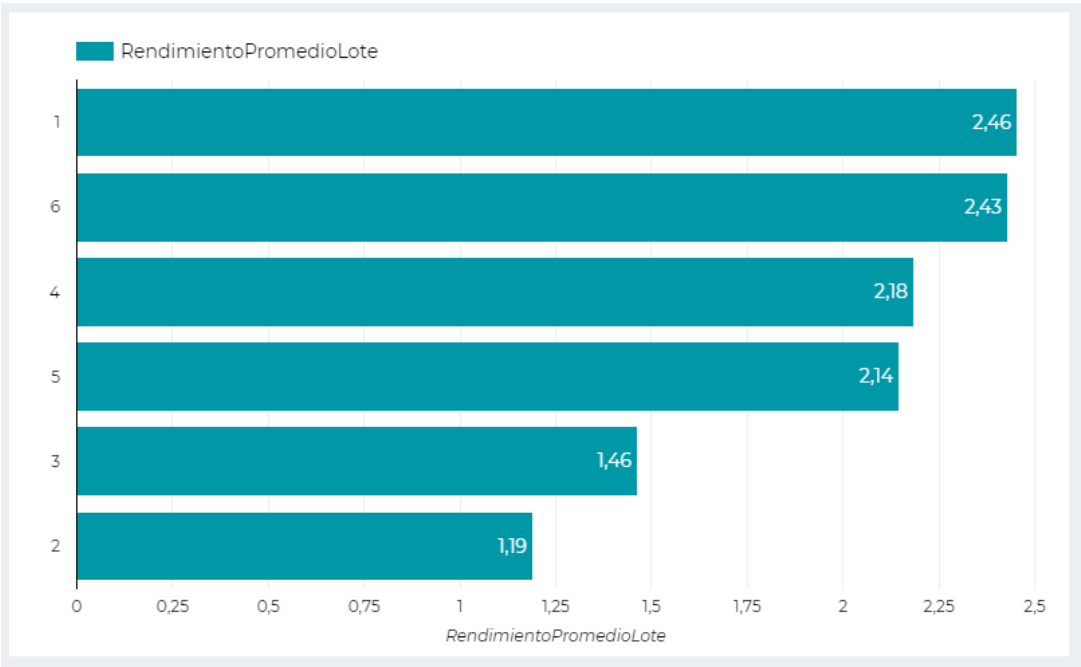
Lote 5



Lote 6



Promedio rendimiento lotes



Rendimiento promedio del establecimiento



Rendimiento República Argentina



Rendimiento Departamento General López



Rendimiento Provincia de Santa Fe



Zonas o departamentos cercanos al establecimiento analizado.

nombredepartamento ▾	
Tres Arroyos	<div>Rendimiento entre:</div> <div>1,5 - 2,1</div> <div>Toneladas por Ha.</div>
San Salvador	
Fray Justo Santa Maria de Oro	
Coronel Suarez	
2 de Abril	
12 de Octubre	

Conclusiones

Conclusiones del modelo planteado:

Según los resultados obtenidos podemos concluir que el rendimiento resultante en la extracción de las imágenes, es un rendimiento relativamente bajo en comparación con la zona de mayor rendimiento a nivel nacional.

A simple vista, se observa que la inversión no es viable ya que los costos de implantación del cultivo de soja más el arrendamiento del establecimiento igualan o superan el rendimiento de tonelada por hectárea obtenida

Conclusiones técnicas y sobre la materia:

Consideramos que fue una materia de mucho provecho para nuestra formación académica, ya que nos permitió entender la importancia de la inteligencia de negocios en el proceso de toma de decisiones. Algunos de los puntos que queremos destacar son los siguientes:

- Poder transformar archivos de diversa índole en datos, en este caso para el análisis de imágenes png.
- Poder conectar distintas plataformas entre sí para la emisión y recepción de datos, logrando un flujo bidireccional de la información.