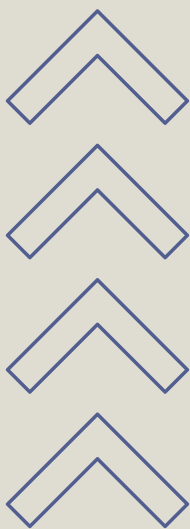




COMISIÓN 71945

# PROYECTO DATA SCIENCE II

ANÁLISIS DE DATOS DE  
CERVEZA ARTESANAL



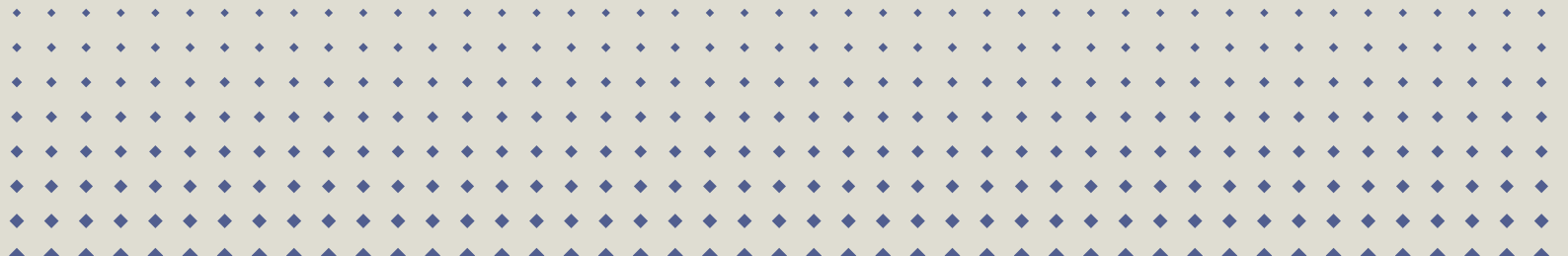
Alumno: Comas Yamil





# ÍNDICE

• Información del dataset	1
• Campos de datos	1
• Desarrollo del objetivo	2
• Hipótesis	2
• Preguntas	2
• Visualizaciones	3-6
• Análisis del resultado	7



## INFORMACIÓN DEL DATASET

Análisis integral de parámetros de elaboración, tendencias de ventas y métricas de calidad en la producción de cerveza artesanal. Período de tiempo que abarca: Enero de 2020 a enero de 2024.

Tamaño del conjunto de datos: 10,000,000 filas y 20 columnas. Asimismo se realiza la reducción de datos para poder analizar una muestra de 25000 filas. **df\_reducido**

Descripción general: Este conjunto de datos representa una recopilación exhaustiva de información de una cervecería artesanal. Incluye una amplia gama de parámetros relacionados con el proceso de elaboración de cerveza, datos de ventas y evaluaciones de calidad. El conjunto de datos proporciona una visión integral de las operaciones de la cervecería y las dinámicas del mercado en el que opera.

El objetivo principal de este proyecto es predecir el Quality\_Score de los lotes de cerveza artesanal en base de los parámetros de elaboración y otras características. Esto permitirá a la cervecería identificar los factores clave que influyen en la calidad del producto y optimizar sus procesos de producción.

## CAMPOS DE DATOS

Campo	Descripción	Dtype
Batch_ID	Identificador único del lote de producción.	int64
Brew_Date	Fecha y hora de elaboración.	object
Beer_Style	Estilo de cerveza (ej., Ale, Stout, Wheat).	object
SKU	Unidad de mantenimiento de stock.	object
Location	Ubicación de la cervecería.	object
Fermentation_Time	Tiempo de fermentación en horas.	int64
Temperature	Temperatura durante la elaboración.	float64
pH_Level	Nivel de pH de la cerveza.	float64
Gravity	Gravedad específica de la cerveza.	float64
Alcohol_Content	Contenido de alcohol en porcentaje.	float64
Bitterness	Nivel de amargor.	int64
Color	Color de la cerveza.	int64
Ingredient_Ratio	Proporción de ingredientes.	object
Volume_Produced	Volumen producido en litros.	int64
Total_Sales	Ventas totales en unidades monetarias.	float64
Quality_Score	Puntaje de calidad de la cerveza.	float64
Brewhouse_Efficiency	Eficiencia de la sala de cocción.	float64
Loss_During_Brewing	Pérdida durante la elaboración.	float64
Loss_During_Fermentation	Pérdida durante la fermentación.	float64
Loss_During_Bottling_Kegging	Pérdida durante el embotellado/ensado.	float64



# DESARROLLO DEL OBJETIVO

## CONTEXTO

La industria de la cerveza artesanal ha experimentado un crecimiento significativo en los últimos años, con una mayor demanda de cervezas únicas y de alta calidad.

Las cervecerías artesanales necesitan optimizar sus operaciones para mantenerse competitivas, lo que incluye la mejora de la calidad, la eficiencia de la producción y la comprensión de las tendencias del mercado.

Este proyecto se centra en el análisis de un conjunto de datos detallado de una cervecería artesanal, que abarca un período de cuatro años.

## ALCANCE

El análisis de estos datos puede proporcionar información valiosa sobre los factores que influyen en el éxito de la cervecería.

Los hallazgos pueden ayudar a la cervecería a tomar decisiones informadas sobre la producción, las ventas y el marketing.

Además, el proyecto puede servir como un ejemplo de cómo el análisis de datos puede aplicarse a la industria de la cerveza artesanal en general.

## OBJETIVOS DEL ANÁLISIS

- Comprender los patrones de ventas y la demanda de diferentes estilos de cerveza.
- Identificar los parámetros de elaboración que influyen en la calidad de la cerveza.
- Evaluar la eficiencia de la producción y las pérdidas en cada etapa del proceso.
- Determinar los factores que contribuyen a la rentabilidad de la cervecería.

Identificar los parámetros óptimos de producción para cada tipo de cerveza.



# HIPÓTESIS

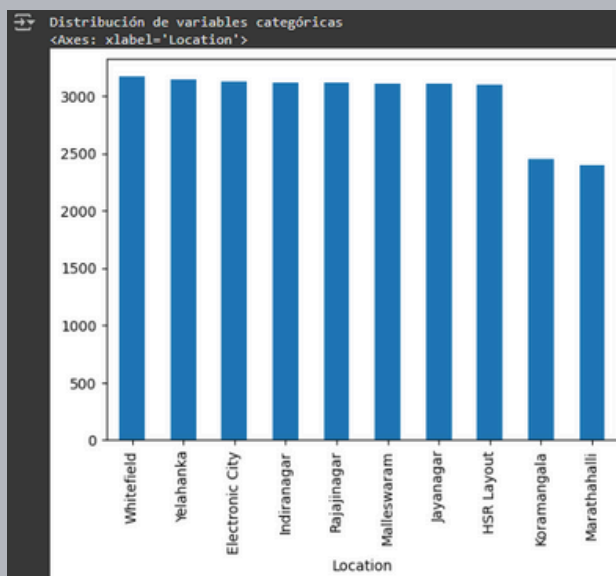
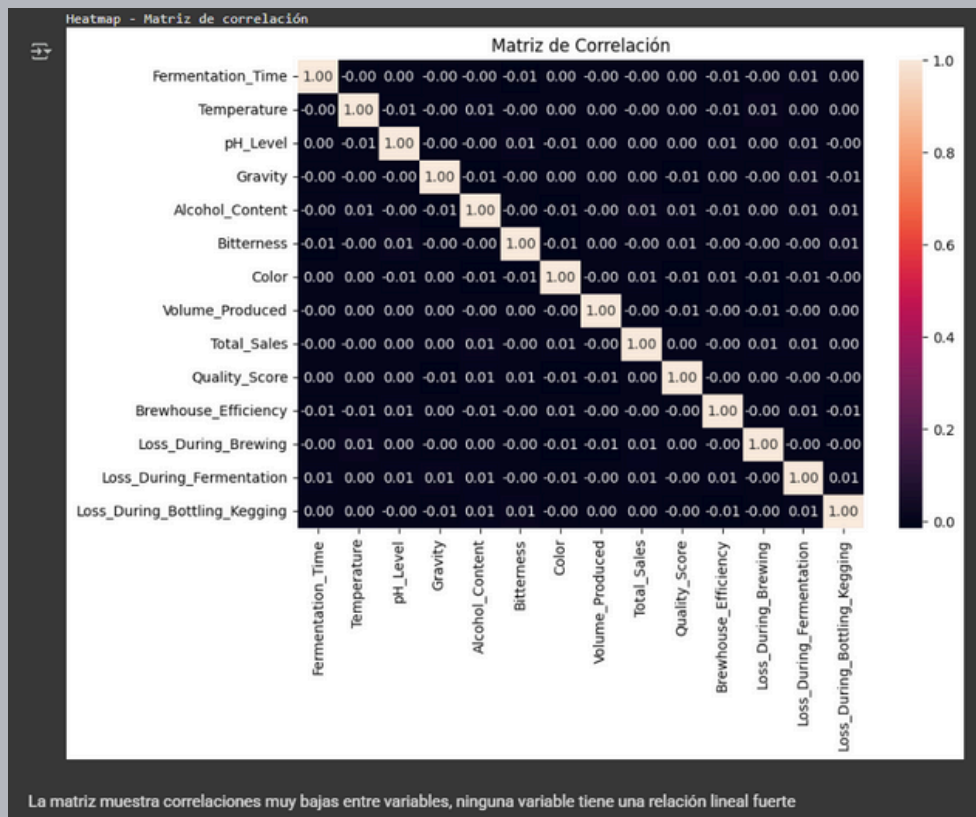
- Los estilos de cerveza más populares varían según la ubicación de la cervecería.
- Los parámetros de elaboración (temperatura, pH, gravedad, etc.) tienen un impacto significativo en la calidad de la cerveza.
- Las pérdidas durante la producción (elaboración, fermentación, embotellado) afectan negativamente la calidad de la cerveza.
- La eficiencia de la cervecería varía según la ubicación.
- Las ventas de cerveza varían según la temporada del año.
- La rentabilidad de la cervecería está influenciada por una combinación de factores, incluyendo las ventas, la calidad y la eficiencia.



# PREGUNTAS

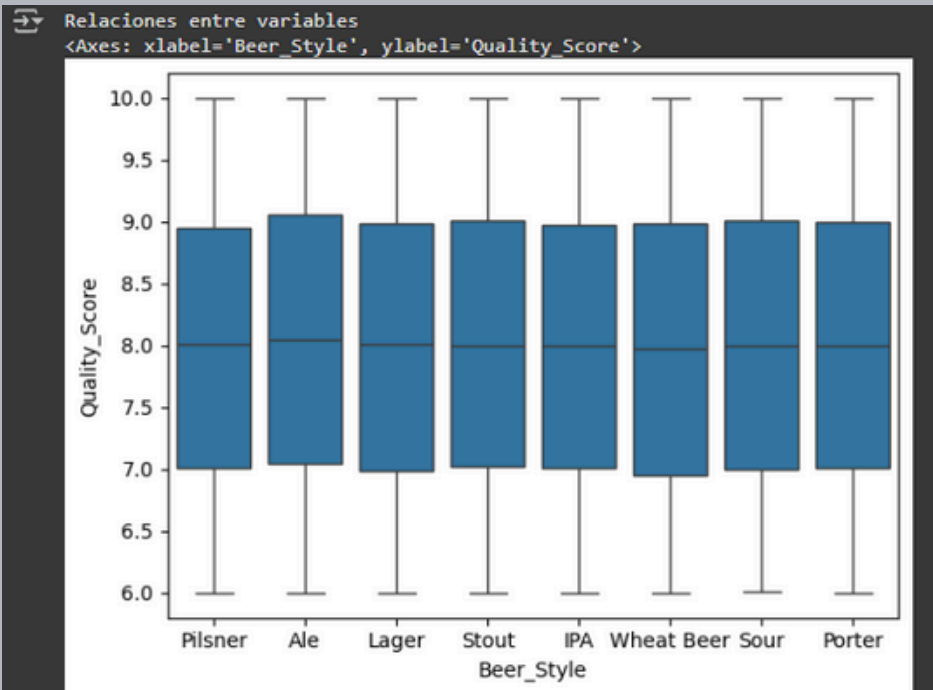
1. ¿Qué estilos de cerveza tienen las mayores ventas en cada ubicación?
2. ¿Existen diferencias significativas en las preferencias de los clientes entre las ubicaciones?
3. ¿Cómo se correlacionan los parámetros de elaboración con el puntaje de calidad de la cerveza?
4. ¿Cuáles son los rangos óptimos de los parámetros de elaboración para producir cervezas de alta calidad?
5. ¿Existe una correlación entre las pérdidas de producción y el puntaje de calidad?
6. ¿Qué etapas del proceso de producción tienen las mayores pérdidas y cómo se pueden reducir?
7. ¿Cuáles ubicaciones presentan los mejores niveles de eficiencia?
8. ¿Qué factores influyen en la eficiencia de la cervecería?
9. ¿Existen patrones estacionales en las ventas de cerveza?
10. ¿Cuáles son los meses o temporadas de mayor y menor demanda?
11. ¿Qué factores tienen el mayor impacto en la rentabilidad de la cervecería?

# VISUALIZACIONES

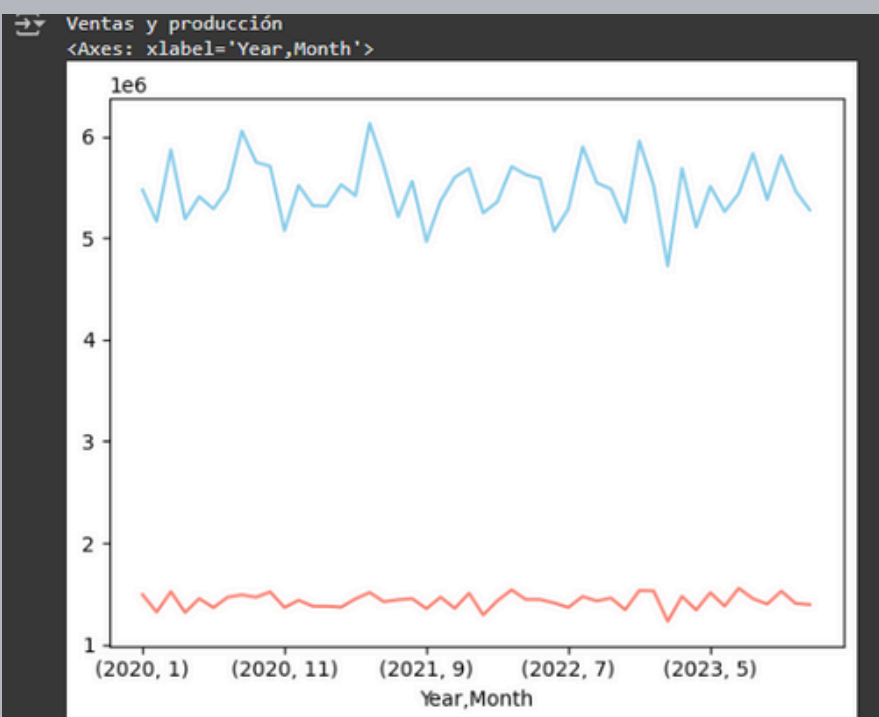


La información del dataset es equilibrada para todas las locaciones, al utilizar una muestra se puede observar que hay dos locaciones que presentan menos muestras.

# VISUALIZACIONES



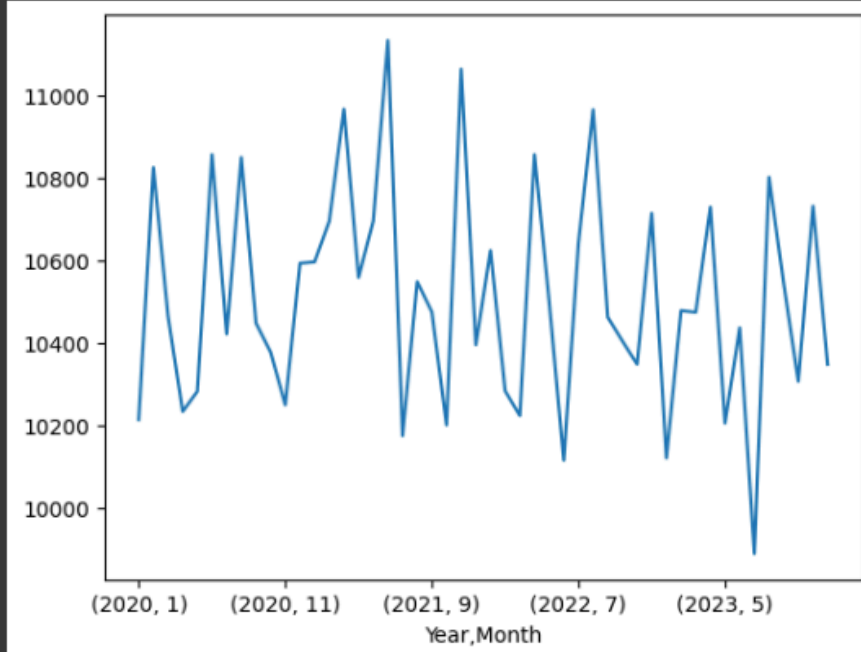
No se observa outliers, presentan valores equitativos entre variables.



Presentan una curva similar entre Total\_Sales y Volume\_Produced indicando una sincronización entre las ventas y la producción.

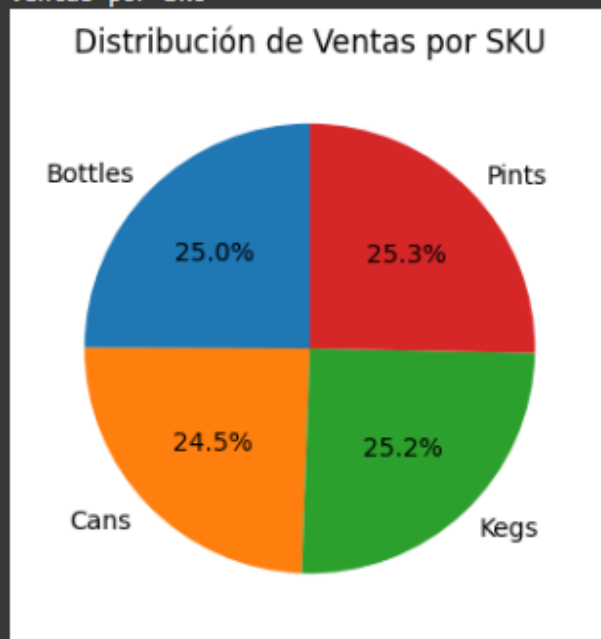
## VISUALIZACIONES

➔ Promedio de ventas mensuales a lo largo del tiempo  
<Axes: xlabel='Year,Month'>



Las ventas mensuales presentan estabilidad general, con una ligera volatilidad estacional.

➔ Ventas por SKU



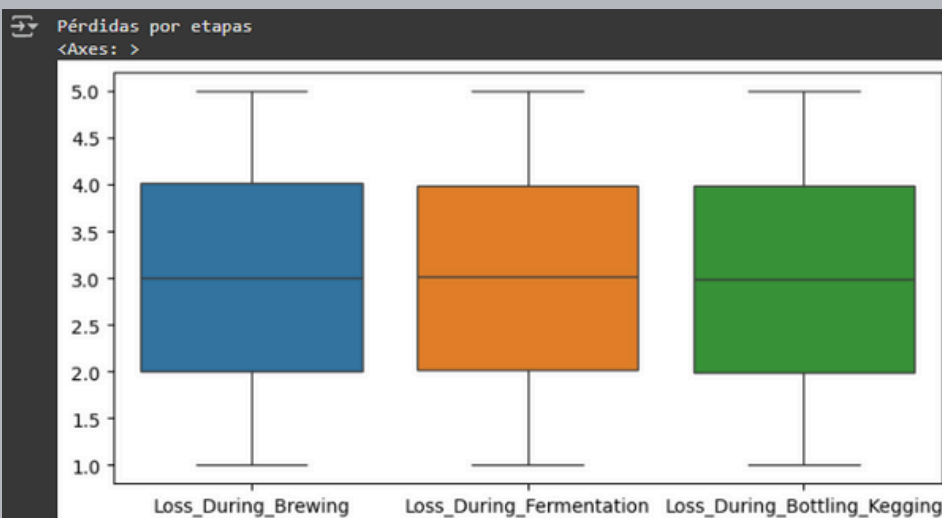
Se puede observar el que dataset es sintético teniendo datos equitativos .

# VISUALIZACIONES

## EFICIENCIA POR UBICACIÓN

```
Location
Electronic City    79.994885
HSR Layout         80.005645
Indiranagar        80.134104
Jayanagar          80.024274
Koramangala        79.978101
Malleswaram        79.810873
Marathahalli       80.050499
Rajajinagar        80.297440
Whitefield         80.047214
Yelahanka          79.776445
Name: Brewhouse_Efficiency, dtype: float64
```

La eficiencia por localizacion esta equilibrada



Al ser un dataset sintético, presenta una óptima y equitativa pérdida por eficiencia.





# ANÁLISIS DE RESULTADOS

## CONCLUSIONES GENERALES

Luego de analizar en profundidad los datos sobre la producción y ventas de cerveza artesanal entre los años 2020 y 2024, se pudieron identificar ciertos patrones relevantes.

### DISTRIBUCIÓN POR UBICACIÓN Y TIPO DE CERVEZA:

Las ventas se reparten de manera bastante pareja entre las distintas ubicaciones.

### RELACIÓN ENTRE ESTILO Y CALIDAD:

Los gráficos del Quality\_Score por estilo muestran una gran variación interna, pero no se observan diferencias significativas entre estilos. Esto podría deberse a inconsistencias en la forma en que se mide la calidad, o bien a que faltan variables técnicas importantes para entenderla mejor.

### PRODUCCIÓN VS. VENTAS:

A lo largo del tiempo se mantuvo una tendencia estable tanto en producción como en ventas mensuales, con variaciones normales según la época del año.

### COMPARACIÓN DE MODELOS:

Se probaron tres modelos (Regresión Lineal Múltiple, Random Forest Regressor y Gradient Boosting Regressor) para predecir la calidad, pero ninguno logró buenos resultados.

En todos los casos, el valor de  $R^2$  fue negativo, lo que indica que las variables disponibles no permiten explicar la calidad del producto de manera efectiva.

Esto evidencia que faltan datos clave en la base actual para realizar este tipo de predicciones.

```
MSE:  
RLM: 1.34  
RFR: 1.34  
GBR: 1.34  
R2:  
RLM: -0.00  
RFR: -0.00  
GBR: -0.00
```