

## Sprint 1:

### 1 ) Crear el repositorio público y compartirlo con el equipo:

El primer paso del proyecto, fue la creación de un repositorio público en GitHub, que centraliza todas las actividades de desarrollo y colaboración entre los integrantes del equipo. Este repositorio actúa como un espacio de trabajo compartido donde se almacenan todos los recursos relacionados con el proyecto, como scripts, notebooks, documentación, imágenes y otros elementos esenciales para el avance adecuado.

El repositorio fue creado bajo el nombre PremiumDrinks, accesible en el siguiente enlace: <https://github.com/PremiumDrinks/PremiumDrinks>. Desde este punto, todos los miembros del equipo fueron invitados como colaboradores para asegurar el acceso adecuado.

Se crearon dos logos: El de la consultora "The Borcelle" que será la encargada de ayudar a "DrinksPremium" en la toma de decisiones, y el logo de "DrinksPremium" que es una distribuidora de bebidas alcohólicas.

Se crean carpetas de "Dashboard y Presentación","Data sets","Queries SQL",Phyton con el fin de organizar de manera estructurada el proyecto.

### Transformación y Análisis de Datos de Ventas e Inventario para Premium Drinks:

#### 1. Carga y exploración de los datos

Los datos han sido cargados en un DataFrame de pandas. A través del método `df.info()`, se verificaron las columnas, los tipos de datos y los valores no nulos. Se encontraron 9 columnas, que incluyen campos como Brand, Description, Price, PurchasePrice, y VendorName, con diferentes tipos de datos: int64, float64 y object.

#### 2. Transformaciones de datos

Conversión de fechas: Las fechas en la columna SalesDate se han convertido al tipo datetime usando `pd.to_datetime()`. Esto permite un manejo más eficiente de las fechas en análisis futuros, como tendencias de ventas a lo largo del tiempo.

Conversión a numéricos: Varias columnas clave relacionadas con ventas y precios, como SalesQuantity, SalesDollars, SalesPrice, Volume, y ExciseTax, se han convertido a tipo numérico mediante `pd.to_numeric()`. Esto se hizo para corregir valores no válidos y permitir un análisis numérico. Valores no convertibles se convirtieron en NaN usando el parámetro `errors='coerce'`.

#### 3. Limpieza y validación

Se exportaron varios DataFrames a archivos CSV, como:

InvoicePurchases12312016\_Clean.csv

2017PurchasePricesDec\_Clean.csv

PurchasesFINAL12312016\_Clean.csv

EndInvFINAL12312016\_Clean.csv

SalesFINAL12312016\_Clean.csv

BegInvFINAL12312016\_Clean.csv Esto sugiere que los datos fueron procesados y guardados para futuras referencias, asegurando que se mantenga un registro de los datos limpios y listos para su uso.

#### 4. Análisis de precios y gráficos

Se realizaron visualizaciones, como un boxplot de precios y precios de compra (Price y PurchasePrice), para analizar la distribución de estos dos valores clave. Los boxplots permiten identificar visualmente los valores atípicos y la dispersión de los precios en el dataset.

Se generaron gráficos de dispersión (scatter plots) para explorar relaciones entre diversas variables, como precio de ventas vs cantidad de ventas.

#### 5. Conclusiones preliminares

Limpieza exitosa de datos: Se han manejado datos faltantes o errores en las columnas numéricas, garantizando que el análisis posterior sea preciso y no se vea afectado por datos corruptos.

Transformaciones clave: Convertir las fechas y asegurar que las columnas numéricas sean adecuadas ha sido una parte esencial del proceso, preparando el terreno para análisis futuros como la segmentación por tiempo o cálculos financieros.

Exportación: Los DataFrames procesados han sido exportados a archivos CSV, lo que indica que el trabajo está listo para su uso o análisis más detallado en otros contextos.