

## ACTIVIDAD 1

Realizar la regresión ajustando un modelo de mínimos cuadrados. (nota docente: la guía del taller se encuentra en la sección 1 del cuaderno de jupyter llamado regresión.ipynb).

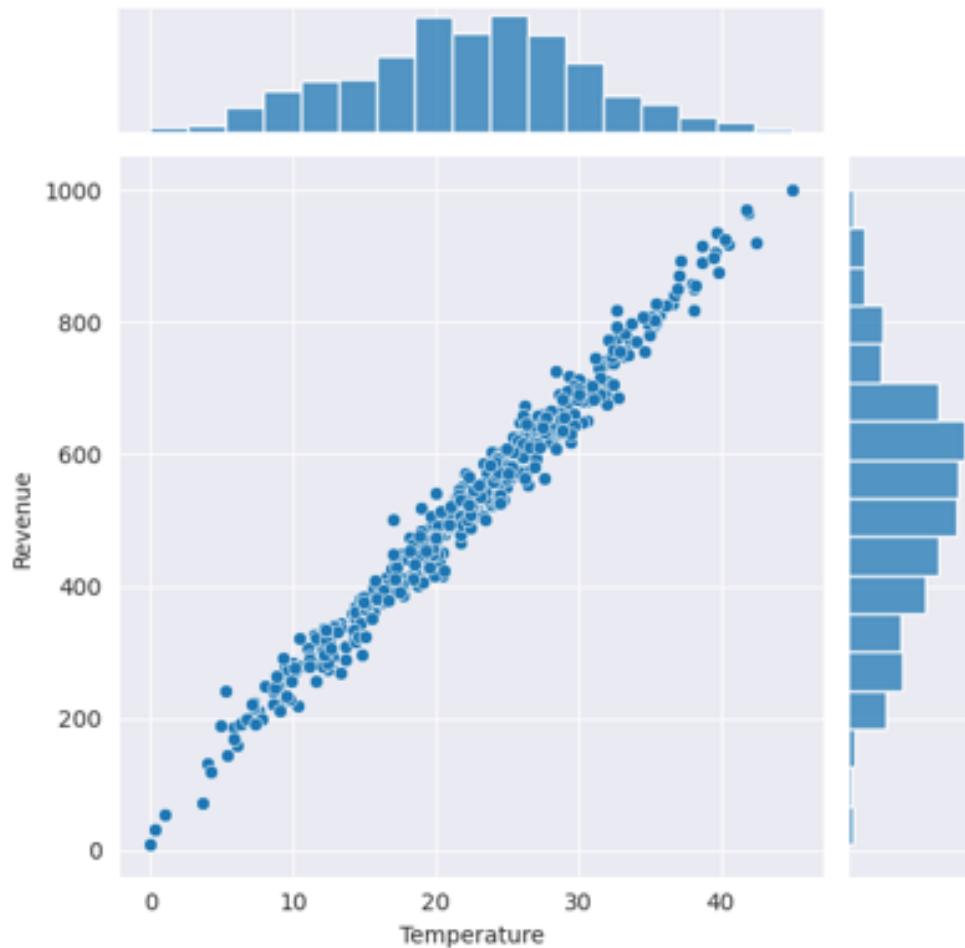
Para explicar mejor el modelo de regresión lineal, se empleará el conjunto de datos “ice cream revenue” que tiene la relación entre ventas de helados y la temperatura ambiente. (disponible en: <https://www.kaggle.com/datasets/vinicio150987/ice-cream-revenue>).

Es un dataset simple, solamente tiene dos columnas que son la temperatura y las ventas totales de helados. (Nota para el docente: Código del análisis disponible en el cuaderno regresión.ipynb, sección 2: Regresión del dataset ICE-CREAM-REVENUE)

Una vez leído el conjunto de datos, se procederá a realizar un rápido análisis exploratorio, empleando las funciones de pandas describe e info. (500 datos, 2 columnas) (el código se extrae del cuaderno para facilitar el seguimiento del ejercicio).

```
df_helados.info  
df_helados.describe()
```

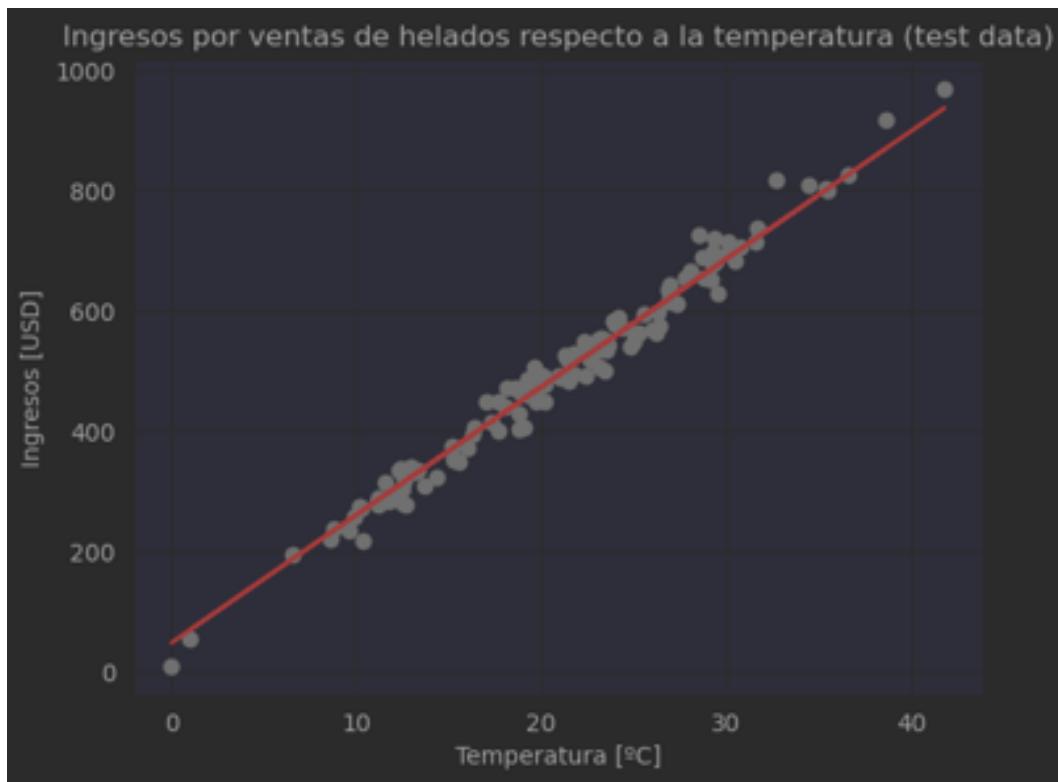
Luego, utilizando seaborn, crearemos un gráfico que permita visualizar todos los datos (en este caso es viable ya que son solo dos columnas). Se generará un gráfico como el de la figura 1 en donde se ve claramente que los datos se pueden modelar mediante una línea recta.



**Figura 1: Gráfico de los datos del dataset Ice cream revenues**

Con el fin de evaluar los modelos de regresión, es necesario separar el conjunto de datos en dos. Un subconjunto de datos de entrenamiento de modelos, que servirá para generar modelos de regresión y un segundo subconjunto de datos de evaluación que ayudará a validar que el modelo tenga un error tolerable y no esté sobre entrenado.

Empleando la clase `LinearRegression` de `scikit-learn` crearemos un modelo y evaluaremos su desempeño (secciones 2.2, 2.4 y 2.4 del cuaderno). Finalmente se evaluará el modelo y se revisarán que los resultados del modelo predigan correctamente (con el menor error posible) los datos de test (ver figura 2)



**Figura 2: Resultados del modelo sobre el conjunto de datos de validación.**