

## Laboratorio 1

### Regresión Lineal

**Instrucciones:** Usted usará Python a través de un Jupyter Notebook para llevar a cabo este ejercicio. Recuerde utilizar comentarios para describir lo que está haciendo en cada parte del proceso. Estará usando el dataset proporcionado dentro del portal. Al finalizar recuerde subir al portal un link a su repositorio en el que se pueda correr su notebook.

El dataset proporcionado es parte de la plataforma Kaggle, dentro del cual se muestran los precios de diferentes calabazas en la ciudad de Los Ángeles del 24 de septiembre de 2016 al 30 de septiembre de 2017.

Para este ejercicio se le pide que proporcione un modelo de regresión polinomial que muestre la relación precio y tamaño de la información sobre las calabazas que se muestra en el dataset proporcionado. A continuación se mencionan las generalidades de los pasos sugeridos a realizar.

1. Leer el archivo CSV proporcionado (los-angeles\_9-24-2016\_9-30-2017.csv) y almacenarlo en un **np.array** para ser trabajado en el notebook.
2. Ajustar un modelo polinómico (regresión lineal) en base al dataset cargado de forma matricial que relacione las variables de precio más alto y tamaño (High Price - Item Size).
3. Utilice la implementación vectorial del algoritmo de regresión lineal.
4. Usando *cross-validation* determine el grado del polinomio que mejor describe la nube de puntos (encuentre el mejor balance entre apego a los datos de entrenamiento y generalización para datos previamente no observados).
5. Haga un análisis sobre sus hallazgos.