

# Universidad Externado

## Taller de Ciencia de Datos

### Predicción: Uber y City Bike Share

#### Modelos Supervisados

Junio 2024

Departamento de Matemáticas

# Análisis de Datos: Uber y City Bike Share

June 16, 2024

## Introducción

En este taller, los estudiantes tendrán la oportunidad de analizar datos de los casos de Uber y City Bike Share utilizando técnicas de regresión lineal y logística. Además, se explorarán diferentes métodos de validación cruzada en Python y se validarán los resultados con todas las métricas de regresión o clasificación correspondientes.

## Descripción del Taller

Los estudiantes podrán elegir uno de los dos casos para su análisis:

- **Uber:** Análisis de los datos de viajes para predecir la demanda y otros factores relevantes.
- **City Bike Share:** Análisis de los datos de uso de bicicletas compartidas para entender los patrones de uso y otros factores relevantes.

## Metodología

El taller se desarrollará en los siguientes pasos:

1. **Selección del Caso:** Escoger entre el caso de Uber o City Bike Share.
2. **Exploración de Datos:** Realizar un análisis exploratorio de datos (EDA) para entender la estructura y características principales del conjunto de datos seleccionado.
3. **Preprocesamiento de Datos:** Limpiar y preparar los datos para el modelado. Esto incluye manejo de valores faltantes, normalización, y codificación de variables categóricas.
4. **Selección del Modelo:** Escoger entre regresión lineal o regresión logística, dependiendo del objetivo del análisis.
5. **Implementación del Modelo:** Implementar el modelo en Python utilizando bibliotecas como scikit-learn.
6. **Validación Cruzada:** Aplicar diferentes métodos de cross-validation (K-Fold, Leave-One-Out, etc.) para evaluar la robustez del modelo.

7. **Evaluación del Modelo:** Validar los resultados utilizando todas las métricas de evaluación pertinentes:
  - Para regresión: MAE, MSE, RMSE,  $R^2$ .
  - Para clasificación: Accuracy, Precision, Recall, F1-Score, AUC-ROC.
8. **Informe de Resultados:** Documentar los hallazgos, incluyendo gráficos y tablas que resuman el desempeño de los modelos.

## Recursos

Los siguientes recursos estarán disponibles para los estudiantes:

- Conjunto de Datos de Uber
- Conjunto de Datos de City Bike Share
- Documentación de scikit-learn: [https://scikit-learn.org/stable/user\\_guide.html](https://scikit-learn.org/stable/user_guide.html)
- Tutoriales y guías de Python para Ciencia de Datos

## Entregables

Al finalizar el taller, los estudiantes deberán entregar:

- Código fuente en Python con comentarios explicativos.
- Debe tener un informe detallado en formato markdown que incluya el análisis exploratorio, descripción del modelo, resultados de la validación cruzada y la evaluación del modelo.
- Gráficos y tablas que respalden los hallazgos y conclusiones.