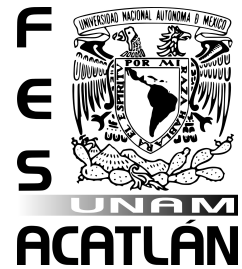




Universidad Nacional Autónoma de México  
Facultad de Estudios Superiores Acatlán  
Diplomado Ciencias de Datos  
Generación 21  
Práctica 2 - Modelos Lineales  
Octubre 2023



## 1.1 SARS-CoV-2

El Gobierno de México nos ha encomendado una tarea crucial: explorar y analizar el Conjunto de Datos Nacional de Salud para identificar los factores determinantes en la hospitalización de pacientes afectados por SARS-CoV-2. Como científicos de datos, nuestro objetivo primordial será discernir los factores clave que inciden en la decisión de hospitalizar a un paciente. Este análisis crucial permitirá optimizar la eficiencia en la atención médica y la asignación de recursos dentro del sistema de salud mexicano.

### Conjunto de Datos

El conjunto de datos está en formato `.pickle`, bajo el nombre de `COVID19MEXICO`. Este conjunto se acompaña de dos archivos adjuntos: `Catalogo` y `Descriptores`, los cuales facilitarán la comprensión e interpretación de los datos contenidos.

### Consideraciones Metodológicas

1. Se deberán excluir del análisis las variables `FECHA_DEF`, `FECHA_INGRESO`, `INTUBADO` y `UCI`.
2. La variable objetivo será `TIPO_PACIENTE`.
3. Se modelará dicha variable objetivo mediante regresión logística.
4. Se identificarán y justificarán las características seleccionadas para el análisis.

### Criterios de Evaluación

- (6 puntos) Se evaluará el código `.ipynb` considerando la limpieza y el orden del mismo. No olvidar guardar modelo en formato `.pkl`.
- (4 puntos) Se requiere un informe conciso (formato diapositiva en extensión `.pdf`) que describa las características influyentes del modelo y conclusión. A su vez se deberá incluir métricas de desempeño que respalden el modelo (No extenderse más allá de 5 diapositivas).

Este análisis debe ser riguroso y meticuloso, ya que los resultados obtenidos influirán directamente en las estrategias y decisiones tomadas por el sistema de salud mexicano en respuesta a la pandemia de SARS-CoV-2. Los hallazgos resultantes del análisis serán cruciales para mejorar los procedimientos y optimizar los recursos, impactando positivamente en la atención y el tratamiento de los pacientes afectados por el virus.

---

## 1.2 La Industria Automotriz

Una importante consultora automotriz los ha contratado para entender los factores que determinan los precios de los autos. Específicamente, desean comprender los factores que afectan los precios de los autos considerando la información disponible con la que cuentan.

### Conjunto de Datos

El conjunto de datos está en formato `.pickle`, bajo el nombre de `Asignación_de_Precio_de_Auto`. Este conjunto se acompaña de un archivos adjunto: `Descriptores`, servirá para la interpretación de los datos contenidos.

### Consideraciones Metodológicas

1. La variable **objetivo** será Precio.
2. Se modelará dicha variable objetivo mediante regresión lineal.
3. Se identificarán y justificarán las características seleccionadas para el análisis.

### Criterios de Evaluación

- (6 puntos) Se evaluará el código `.ipynb` considerando la limpieza y el orden del mismo. No olvidar guardar modelo en formato `.pkl`.
- (4 puntos) Se requiere un informe conciso (formato diapositiva en extensión `.pdf`) que describa las características influyentes del modelo y conclusión. A su vez se deberá incluir métricas de desempeño que respalden el modelo (No extenderse más allá de 5 diapositivas).