**Serie 2 (60 puntos)**

Para este ejercicio, debe de utilizar el dataset final.csv. El set de datos incluye datos demográficos de los 50 estados de Estados Unidos, con las siguientes columnas:

* Population: population estimate as of July 1, 1975.
* Income: per capita income (1974)
* Illiteracy: illiteracy (1970, percent of population)
* Life Exp: life expectancy in years (1969-71)
* Murder: murder and non-negligent manslaughter rate per 100,000 population (1976)
* HS Grad: percent high-school graduates (1970)
* Frost: mean number of days with minimum temperature below freezing (1931-1960) in capital or large city
* Area: land area in square miles

El objetivo de este ejercicio es que usted puede estimar la variable esperanza de vida (life expectancy – life Exp) a partir de las otras variables del set de datos (si aplicaran). Para ello debe de aplicar los conceptos vistos en clase:

* Correlación entre variables
* Validación cruzada (cross validation)
* Feature engineering
* Regresiones lineales

Debe crear un modelo de regresión lineal que nos brinde el menor RMSE de la combinación de variables que usted en base a su percepción y al análisis numérico del set de datos determine pertinentes. Se evaluará el trabajo de ingeniería efectuado (recuerde somos ingenieros) con los datos y como determino usted el mejor modelo. Por temas de tiempo, le recomiendo efectuar un análisis de al menos tres modelos diferentes de regresión y utilizar el que crea usted más conveniente.

CONTADORES QUE NECESITAMOS

Graphical user interface

Description automatically generated with medium confidence

FEATURE ENGINEERING



Entregables:

* PDF detallando las decisiones que tomo al momento de efectuar el proceso de ingeniería de características, el cross validation y las regresiones lineales.
* Código de su solución.