

Lab 3

En este laboratorio estudiaremos factores que contribuyen o no a la esperanza de vida de una localidad (país o ciudad).

1. (5%) Cree un nuevo Jupyter notebook y cargue las siguientes bases de datos:
<https://www.kaggle.com/datasets/kumarajarshi/life-expectancy-who>
<https://www.kaggle.com/datasets/max-mind/world-cities-database/>
<https://www.kaggle.com/datasets/prasertk/healthy-lifestyle-cities-report-2021>
2. (10%) Para la base de datos de esperanza de vida por país (life-expectancy-who), conserve únicamente el año más reciente. Procure también excluir filas con valores vacíos (que contienen null o NaN).
3. (5%) Para la variable de esperanza de vida, calcule la media, mediana y la desviación estándar de todos los países para el año más reciente. Imprímalos en el notebook.
4. (10%) Realice un histograma con la variable esperanza de vida. Cambie progresivamente el número de barras, desde un número pequeño como 5 hasta un número grande como 100. Reporte qué es lo que cambia en las distribuciones conforme aumenta la cantidad de barras. Escoja qué número de barras le parece óptimo y explique por qué. Investigue.
5. (10%) En la misma gráfica anterior, coloque la media (μ) como una línea vertical y la posición de μ más una desviación estándar (σ) y menos una desviación estándar: $\mu - \sigma$ y $\mu + \sigma$.
6. (20%) Investigue cuáles variables tienen una correlación con la esperanza de vida. Imprima los valores de la correlación y explique cuáles son las tres variables más importantes.
7. (40%) Usando la base de datos world-cities-database, y algún sitio o servicio web (o biblioteca de Python), cree una nueva columna que corresponda a la altitud de cada ciudad. Como resultado, la base de datos contendrá una variable extra, que no se tenía previamente.

Este laboratorio continuará en el lab 6, en el que se estudiará más a fondo los factores que inciden en la esperanza de vida. Guarde su solución, para entregarla junto con el lab 6.