

Desafío - Estadística descriptiva y probabilidades (parte I)

En este desafío validaremos nuestros conocimientos de estadística descriptiva aprendidos en la sesión. Para lograrlo, necesitarás aplicar lo aprendido, analizando un set de datos asociados a los ingresos obtenidos por la población en base de a la información demográfica entregada en el archivo `ds_salaries.csv`.

Lee todo el documento antes de comenzar el desarrollo individual, para asegurarte de tener el máximo de puntaje y enfocar bien los esfuerzos.

Tiempo asociado: 1 hora cronológica

Descripción

El conjunto de datos contiene un informe salarial fabricado en el que puedes realizar análisis para tratar de estimar el salario basado en las condiciones dadas. Las columnas son las siguientes:

- `work_year`: El año en que se pagó el salario.
- `experience_level`: El nivel de experiencia en el trabajo durante el año, con los siguientes valores posibles: EN (nivel de entrada / junior), MI (nivel medio / intermedio), SE (nivel senior / experto), EX (nivel ejecutivo / director).
- `employment_type`: El tipo de empleo para el puesto: PT (tiempo parcial), FT (tiempo completo), CT (contrato), FL (trabajo freelance).
- `job_title`: El puesto de trabajo desempeñado durante el año.
- `salary`: El monto total del salario bruto pagado.
- `salary_currency`: La moneda en la que se pagó el salario, utilizando un código de moneda ISO 4217.
- `salary_in_usd`: El salario en USD (tipo de cambio dividido por la tasa promedio de USD para el año correspondiente a través de fxdata.foorilla.com).
- `employee_residence`: El país de residencia principal del empleado durante el año laboral, utilizando un código de país ISO 3166.
- `remote_ratio`: La cantidad total de trabajo realizado de forma remota, con los siguientes valores posibles: 0 (sin trabajo remoto, menos del 20%), 50 (parcialmente remoto), 100 (totalmente remoto, más del 80%).
- `company_location`: El país de la oficina principal del empleador o sucursal contratante, utilizando un código de país ISO 3166.

- `company_size`: El número promedio de personas que trabajaron para la empresa durante el año: S (menos de 50 empleados, pequeña), M (50 a 250 empleados, mediana), L (más de 250 empleados, grande).

Se busca realizar un análisis estadístico de estos salarios, considerando:

1. El promedio general de los salarios, desviación estándar, quintiles y rango. (hint: escoge la columna adecuada para comparar)
2. Una comparación entre los salarios agrupados por 3 diferentes categorías a tu elección. ¿Para cuál de ellas las medidas de tendencia central son más representativas? ¿Para cuál son menos representativas? Explica.
3. ¿Qué cargos reciben mejores sueldos en las empresas con sede en Estados Unidos?

Requerimientos

Dentro del archivo de Jupyter Notebook debes ir ejecutando las siguientes acciones y explicar lo que estás haciendo. Debes considerar los siguientes aspectos.

1. Calcular los indicadores estadísticos solicitados, sobre columnas específicas. **(3 Puntos)**
2. Calcular indicadores estadísticos, agrupando y filtrando los valores **(3 Puntos)**
3. Interpretar los indicadores estadísticos en cada caso, considerando medidas de tendencia central y dispersión. **(4 Puntos)**



¡Mucho éxito!

Consideraciones y recomendaciones

- Debes entregar tu trabajo en un archivo de Jupyter Notebook, con todo el código y las explicaciones respectivas para desarrollar tu trabajo.