Ejercicio 4: Medidas de Tendencia Central.

instituto tecnológico de morelia

Ejercicios de práctica

M.I.A. WILFRIDO CORTES OROZCO

2023

|  |  |  |
| --- | --- | --- |
| Nombre del Alumno | Número de control | Número de actividad |
|  |  |  |

# Descripción y objetivos de la actividad.

Objetivos:

* Que el estudiante ponga en práctica las herramientas teóricas aprendidas hasta el momento.
* Usar las herramientas de Pandas para sacar muestras y visualizarlas.

Descripción de la actividad:

La actividad consiste en crear un archivo en Jupyter-Notebook, dentro debe cargar un set de datos a través de la librería de pandas, para ello, debe usar los dataframe.

Una vez que se haya cargado el dataframe, el estudiante debe completar las siguientes tareas:

1. Mostrar información del dataframe, contabilizar cuántas columnas son numéricas y cuántas son de tipo objeto y contabilizar nulos.
2. Mostrar cuál es el mayor puntaje (“total\_cup\_points”) del dataframe.
3. Mostrar el histograma de la columna “species”, tanto con Pandas como con Seaborn.
4. Mostrar el histograma de la columna “total\_cup\_points” con distintos bins:
   1. Mostrar con 10 bins.
   2. Mostrar con 50 bins.
   3. Mostrar con 100 bins.
   4. Mostrar con bins = número de valores únicos que puede tomar la variable/columna.
5. Del punto anterior, hacer un análisis, mencione qué impacto tiene el número de bins durante el análisis y cuál sería la medida ideal.
6. Mencione que tipo de distribución tiene esa columna (si es necesario puede usar el KDE en el histograma para ayudar a la visualización).
7. Haga un muestreo simple de 100 datos y guárdelo en la variable “dfMuestreo100”.
8. Haga una selección de los 100 primeros datos del dataframe y guárdelos en la variable “dfPrimeros100”.
9. Muestre la media, la mediana y la moda del dataset completo.
10. Muestre la media, la mediana y la moda de la variable “dfMuestreo100”.
11. Muestre la media, la mediana y la moda de la variable “dfPrimeros100”.
12. Compare los resultados de los tres pasos anteriores y haga sus conclusiones sobre los muestreos y qué tanta afectación hay con respecto a las medidas de tendencia central.
13. Muestre el histograma de “total\_cup\_points” con la mejor selección de bins para el dataframe original, para la variable “dfMuestreo100” y para la variable “dfPrimeros100”, mencione si hubo algún cambio importante entre cada gráfica.

El set de datos que se va a usar es “Calificaciones\_Cafe.csv”.