

Lección 2: Taller de estadística en Python



Tiempo de ejecución: 2 horas

PLANTEAMIENTO DE LA SESIÓN	MATERIALES
<p>Con el entorno de trabajo local instalado, se busca que los estudiantes tengan un contacto más cercano con las herramientas. Esto se consigue con la práctica, la cual, en esta sesión corresponde a hacer algunas operaciones de lecciones anteriores sobre conjuntos de datos conocidos sobre los que se conocen algunos resultados.</p>	<p>Entorno de trabajo instalado (lección 1)</p>

Mediante lenguajes de programación es posible escribir programas que realicen el análisis de datos sobre un conjunto de datos disponible. Algunos de los conjuntos de datos en los que se ha trabajado previamente son el set de datos del titanic y

Utilizando el entorno instalado y en compañía del docente realice las siguientes actividades:

1. Cree una nueva carpeta y allí copie el dataset del titanic. Abra la carpeta utilizando visual studio code.

2. En el programa visual studio code, cree un nuevo archivo llamado estadística.ipynb se encontrará con un entorno similar al de la imagen 1:

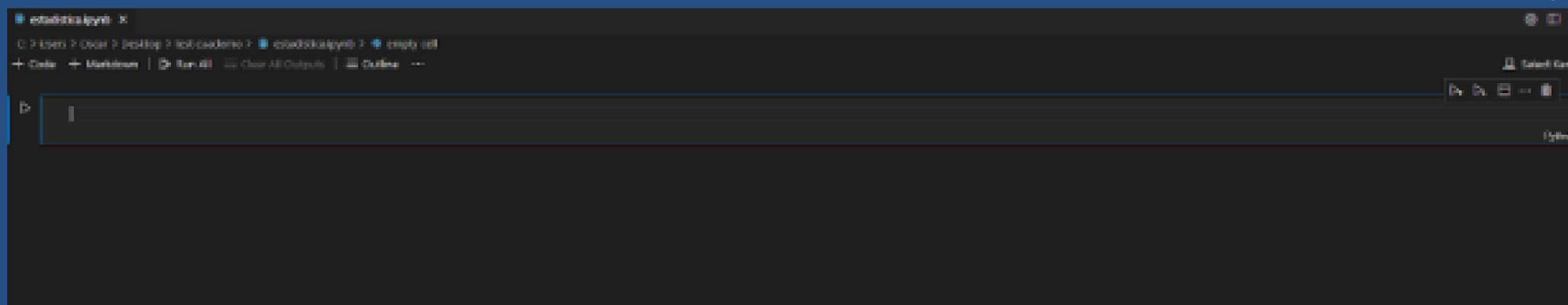


Imagen 1: visualización del cuaderno de Jupyter.



Un cuaderno de Jupyter permite organizar un programa en celdas. Cada celda puede contener código o texto, de esta forma se pueden escribir programas explicativos con texto enriquecido que puede integrar imágenes, ecuaciones y gráficos con el fin de proveer un entorno altamente visual para la ciencia de datos y el desarrollo de experimentos.

Utilizando Excel o Google sheets, copie los valores de la columna sobrevivientes, edad y costo del tiquete desde Excel a una lista de Python. Para esto puede copiarlos entre llaves [] de tal forma que los datos queden almacenados en el código.





3. Utilizando la librería numpy calcule el promedio de edades, repita el proceso separando para sobrevivientes y no sobrevivientes
4. Calcule el promedio del costo del tiquete, también calcule el costo separando por sobrevivientes y no sobrevivientes.
5. Halle el máximo y el mínimo del costo del tiquete y de las edades.
6. Con la librería Matplotlib cree un histograma para ver la distribución de los datos de edad y de costo de tiquete. Repita separando por sobrevivientes y no sobrevivientes.
7. Saque conclusiones al respecto teniendo en cuenta los datos visualizados.

