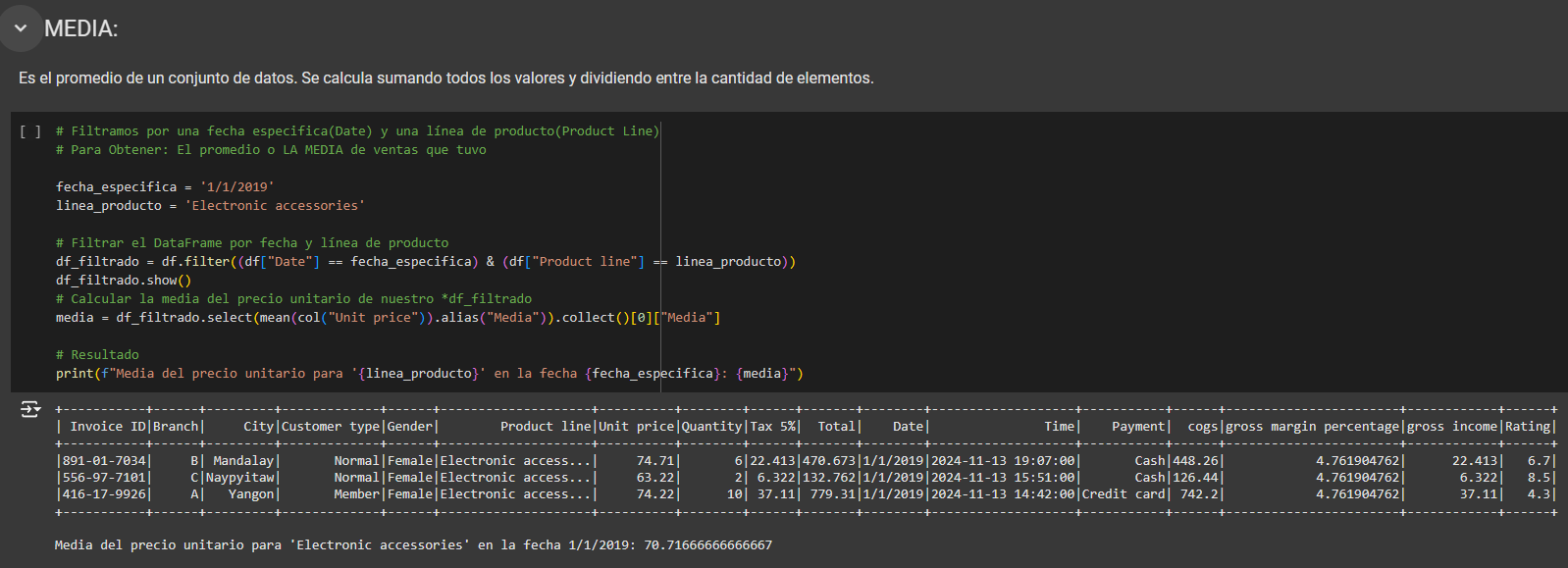
**ACTIVIDAD 14 - 2**

**¿Qué es la media?**

La media, también conocida como promedio, es el valor que se obtiene al dividir la suma de un conjunto de números entre la cantidad de ellos.

Se calcula sumando todos los valores del conjunto de datos y dividiendo la suma por el número total de puntos de datos.

**Ejemplo de Media**

**Explicación del código:**

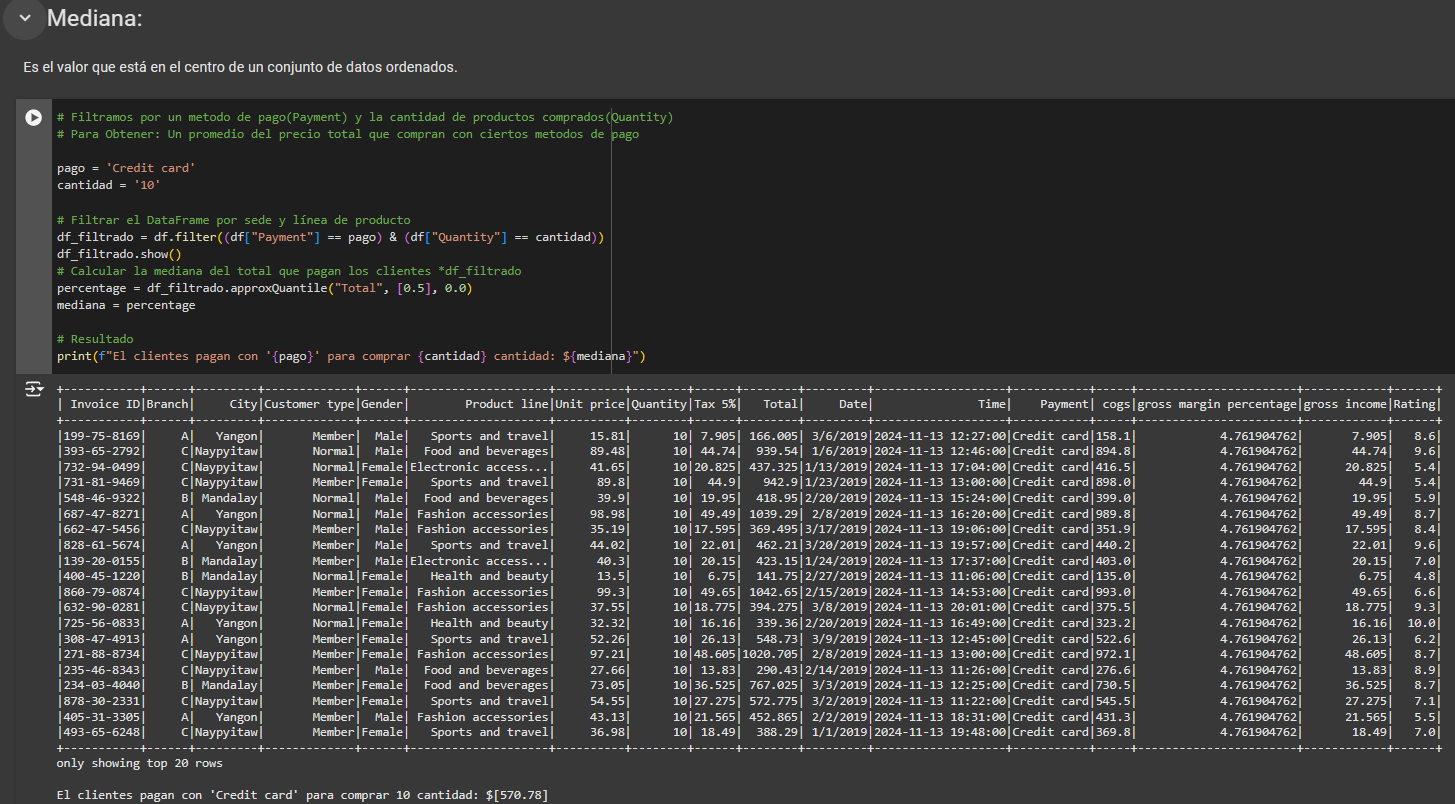
1. **Filtrado del DataFrame:** Filtramos el Dataframe seleccionando solo las filas en las que la fecha (Date) sea igual a '1/1/2019' y la línea de producto (Product line) sea 'Electronic accessories'.
2. **Cálculo de la media:** Usamos la función mean(col("Unit price")) para calcular el promedio del precio unitario de los productos que cumplen con las condiciones de filtro. Luego, con .collect(), obtienes el valor calculado de la media del precio unitario de los productos.

**¿Qué es la mediana?**

La mediana es un conjunto es un valor que se encuentra a la mitad de los otros valores, es decir, que, al ordenar los números de menor a mayor, éste se encuentra justamente en medio.

La media representa el punto de equilibrio y está influida por los valores extremos. Proporciona un valor medio de los datos.

**Ejemplo de Mediana**

****

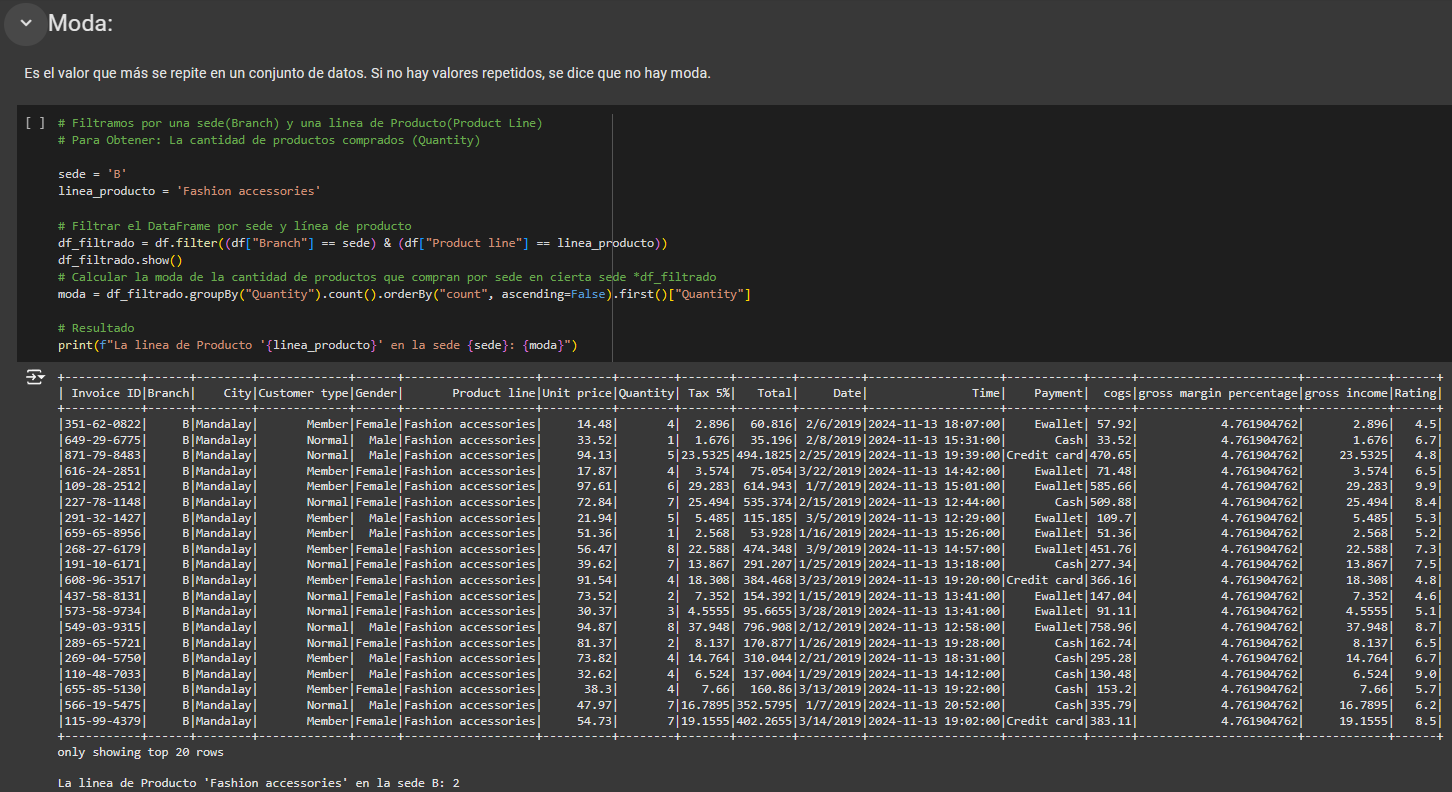
**Explicación del código:**

1. **Filtrado del DataFrame:** Filtramos el DataFrame para seleccionar solo las filas donde el método de pago es "Credit card" y la cantidad de productos comprados es igual a 10.
2. **Cálculo de la mediana:** Usas approxQuantile para calcular la mediana del total de las compras. La función approxQuantile nos permite calcular percentiles aproximados, y en este caso, calculas el percentil 50 (la mediana), y el parámetro 0.0 en approxQuantile indica una aproximación precisa.

**¿Qué es la moda?**

La moda es el valor que aparece más dentro de un conjunto de datos. A diferencia de la media y la mediana, la moda no requiere valores numéricos.

Se llama *amodal* cuando en un conglomerado no se repiten los valores.

**Ejemplo de Moda**

**Explicación del código:**

1. **Filtrado del DataFrame:** Filtramos el Dataframe para seleccionar solo las filas que corresponden a una sede específica (Branch = 'B') y una línea de producto específica (Product line = 'Fashion accessories').
2. **Cálculo de la moda:** Después, agrupar los datos por la columna Quantity y contar cuántas veces aparece cada cantidad de productos comprados. Luego, ordenas estos grupos y seleccionas la cantidad de productos más frecuente, que es la moda.

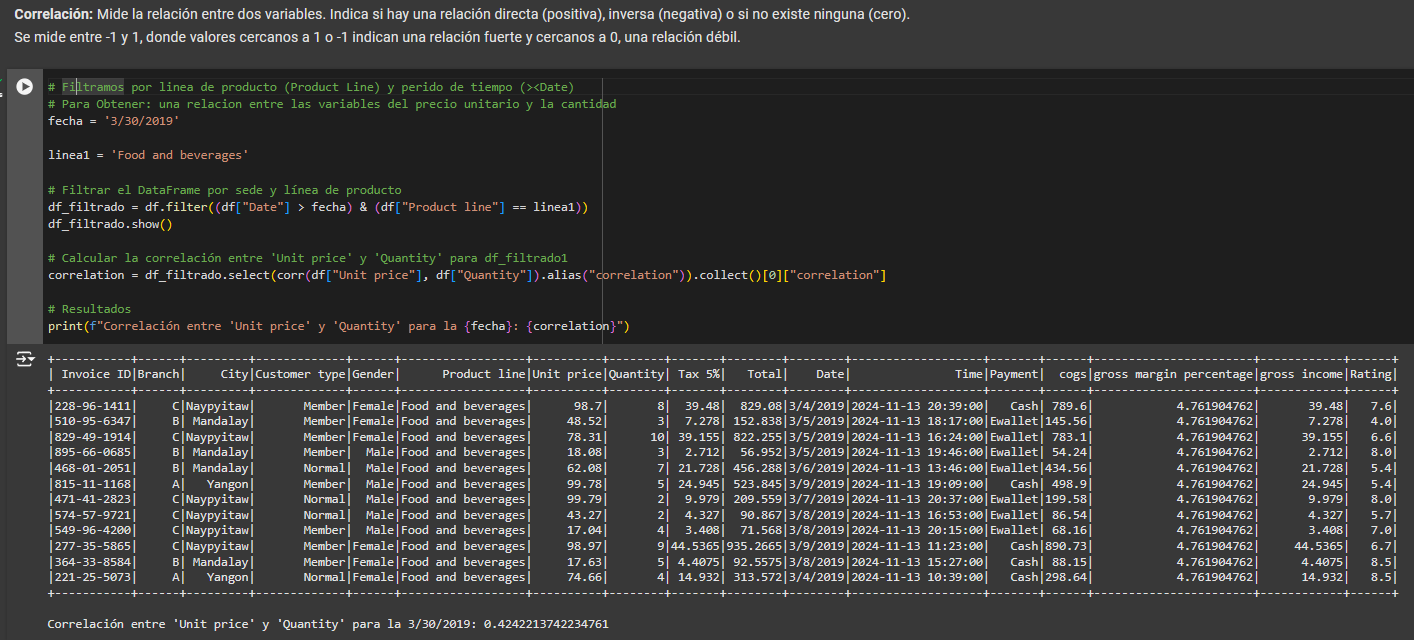
**¿Qué es la correlación?**

La correlación es un tipo de asociación entre dos variables numéricas, específicamente evalúa la tendencia (creciente o decreciente) en los datos.

La correlación nos permite medir el signo y magnitud de la tendencia entre dos variables.

* El **signo** nos indica la dirección de la relación:
  + El valor **positivo** nos indica una relación directa.
  + El valor **negativo** nos indica una relación indirecta.
  + El valor nulo nos indica que no existe tendencia en ambas variables, es decir no tiene relación o que es muy complejo para la tendencia.
* La **magnitud** indica la fuerza de la relación, se mide entre el -1 a 1. Mientras más cercano sean los intervalos a los extremos, más fuerte será la tendencia entre las variables, mientras que más cerca del cero esté la relación, más débil será la tendencia.
  + Si la correlación vale en 1 o -1 diremos que tiene una correlación perfecta.
  + Si está en 0 diremos que no existe correlación o no están correlacionadas.

**Ejemplo de Correlación**



La correlación entre 'Unit price' y 'Quantity' para la línea de producto 'Food and beverages' después del 30 de marzo de 2019 es 0.424, lo que indica una relación moderada entre estas dos variables en este periodo de tiempo. Esto sugiere que, para este subconjunto de datos, a medida que aumenta el precio unitario, también tiende a aumentar la cantidad comprada, aunque la relación no es completamente fuerte.

**Referencias:**

**Media-Mediana-Moda:**

<https://www.questionpro.com/blog/es/la-media-la-mediana-y-la-moda/>

**Correlación:**

<https://www.maximaformacion.es/blog-dat/que-es-la-correlacion-estadistica-y-como-interpretarla/>