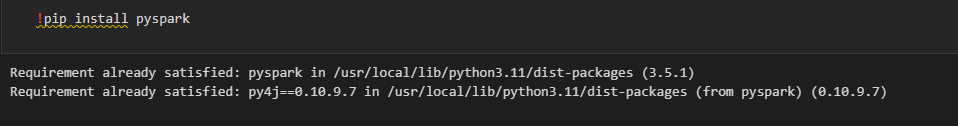
PYSPARK & SPARK CODING CHALLENGE

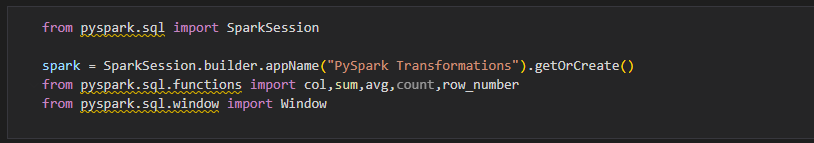
Theophila Murphy S

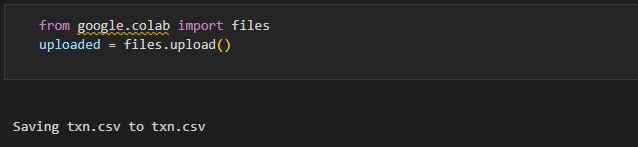
DE-Batch 6

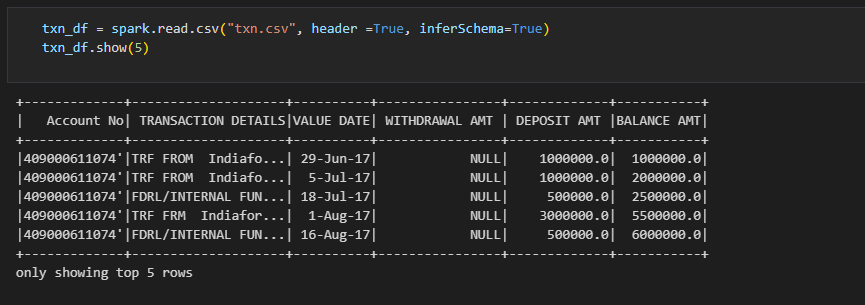
**APPLY TRANSFORMATION & ACTIONS IN PYSPARK**

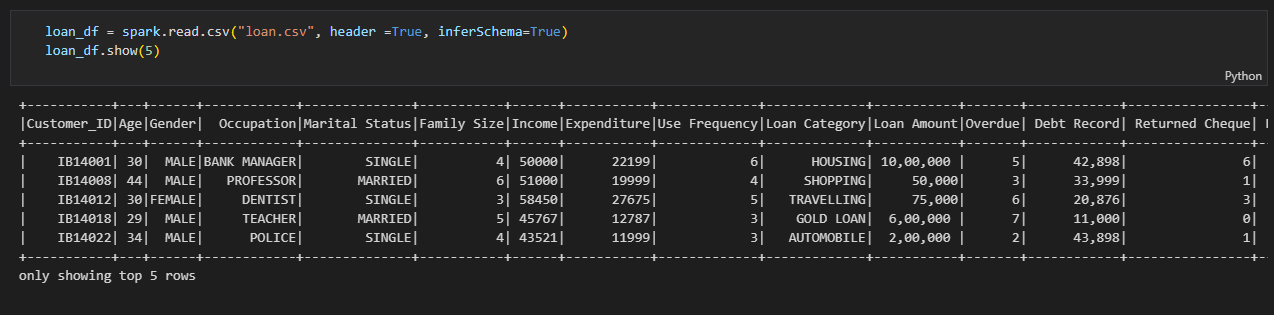
**Initialization and Loading Files**

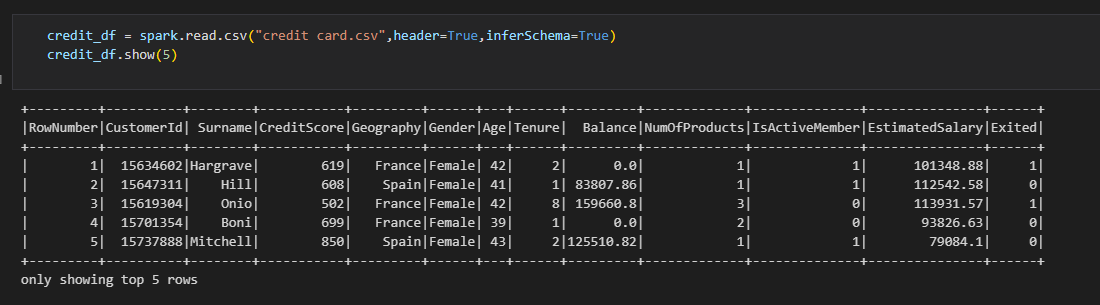






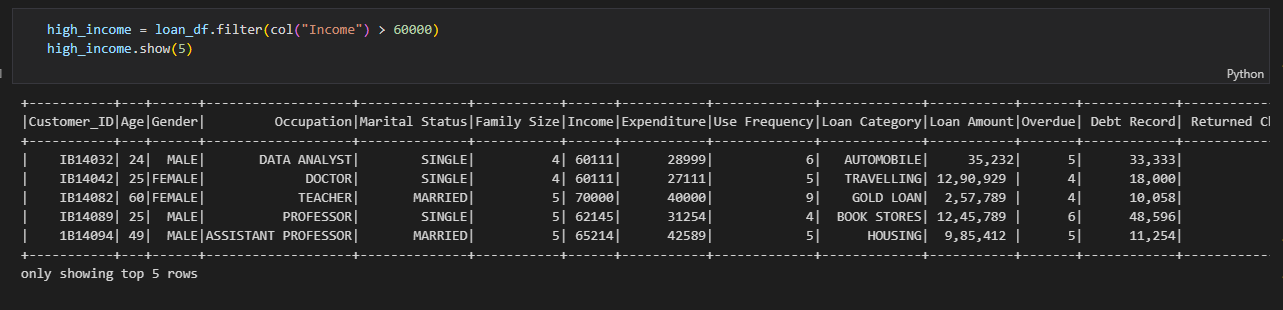




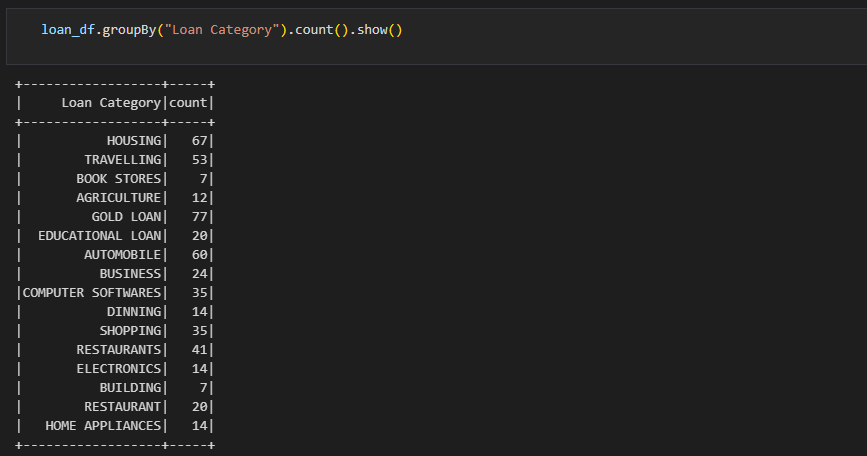


**TRANFORMATIONS**

**Filter()**



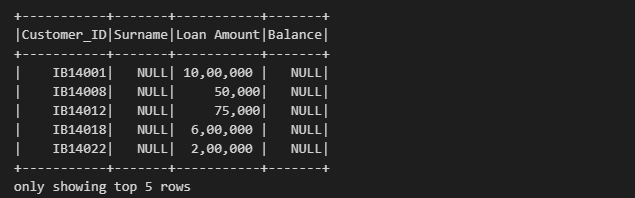
**Group By()**



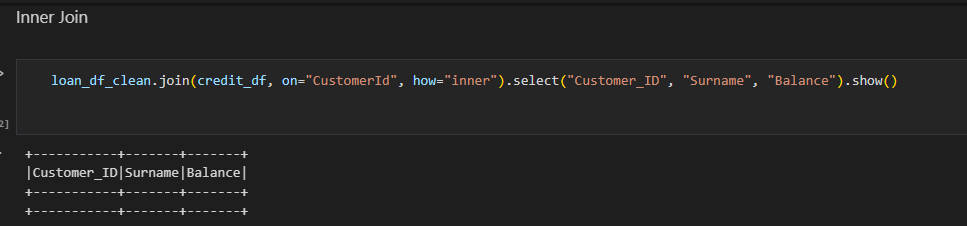
**JOINS**

**Left Join**

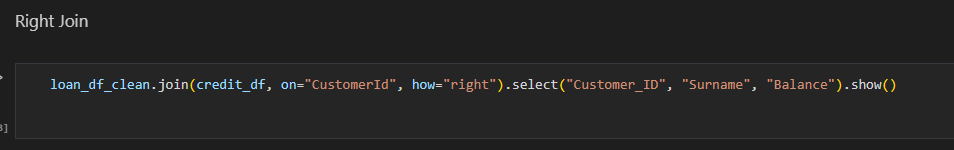


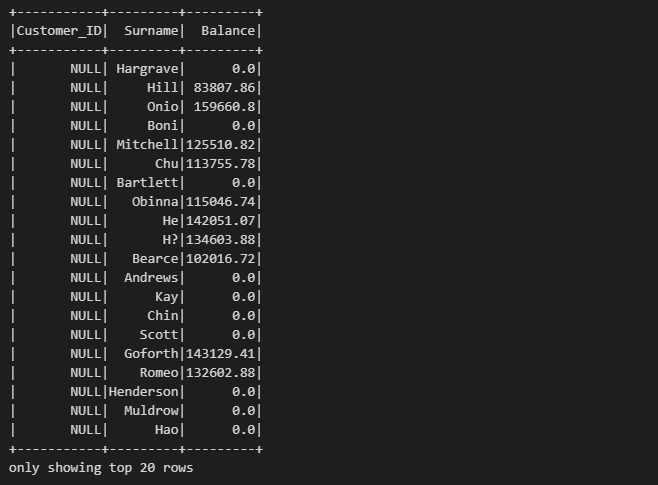


**Inner Join**

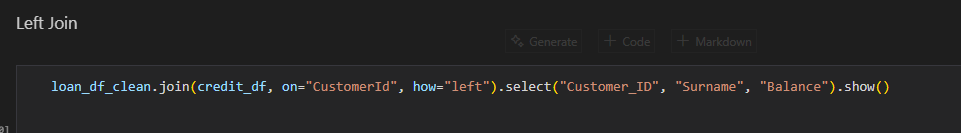


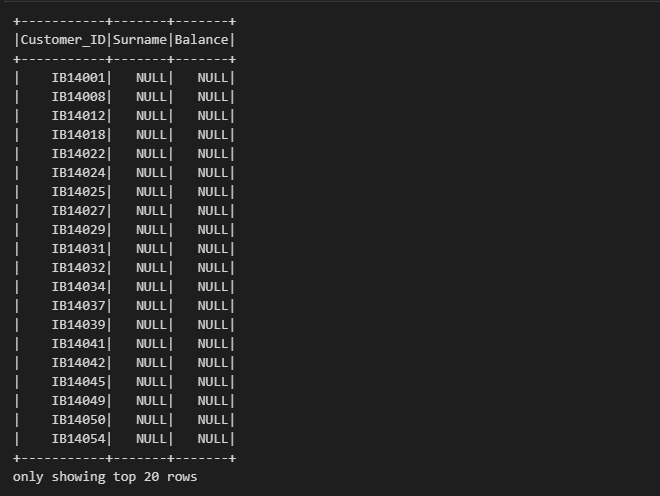
**Right Join**





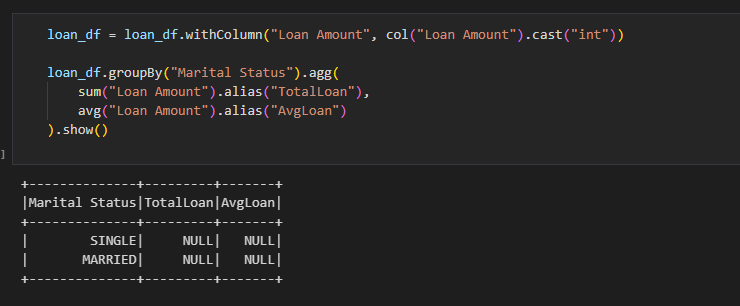
**Left Join**





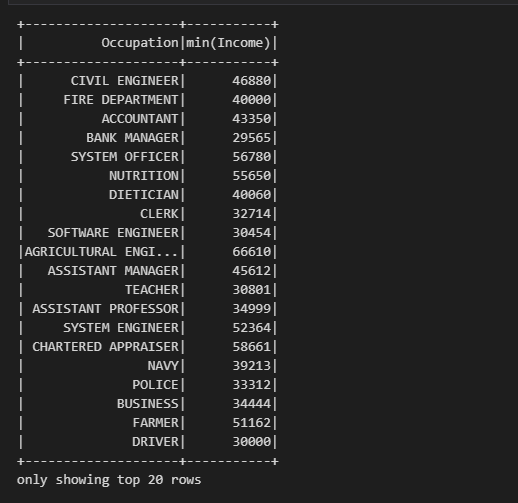
**AGGREGATIONS**

**Sum & Avg**

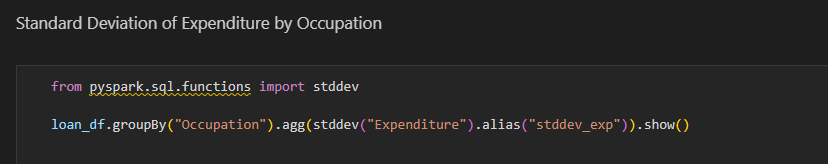


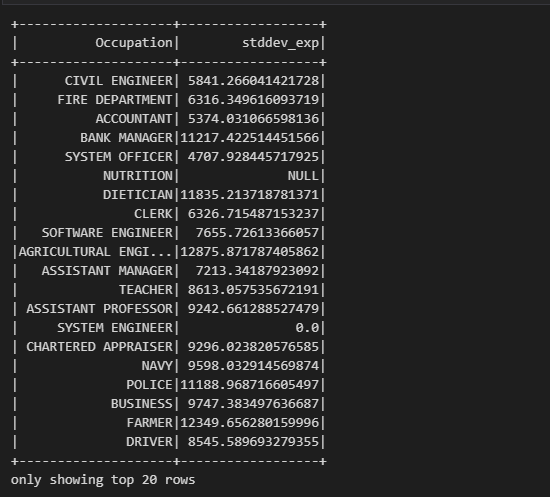
**Max & Min**



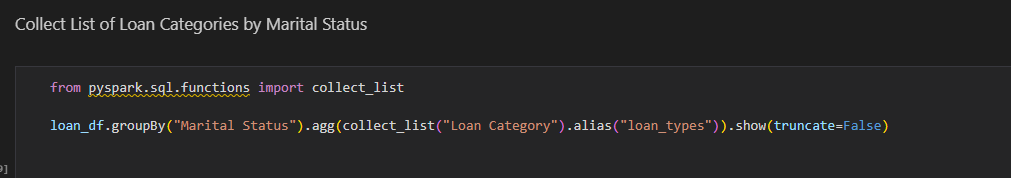


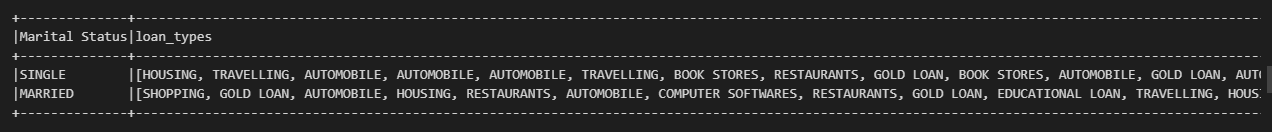
**Standard Deviation**





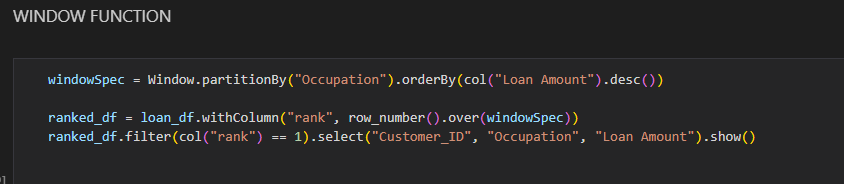
**Agg & Alias**

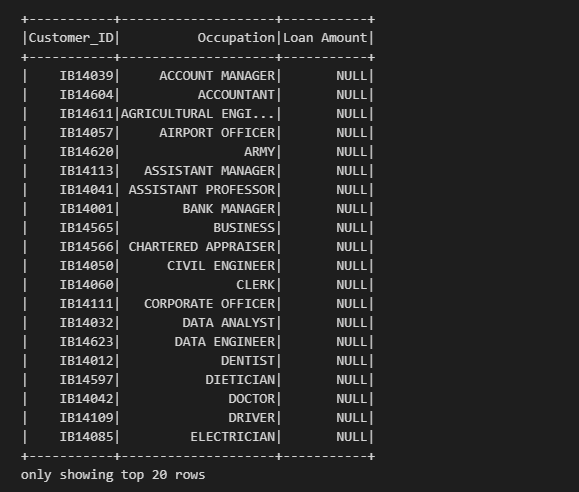




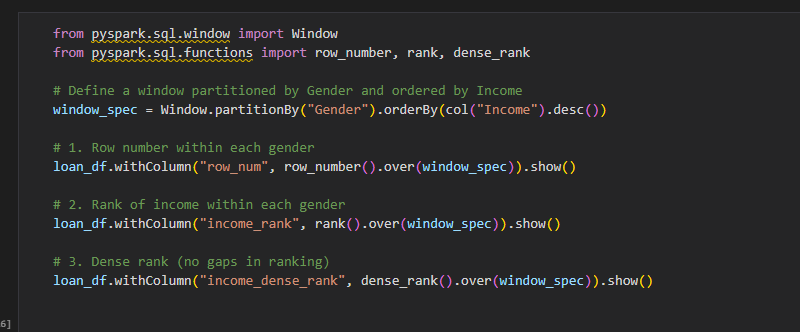
**WINDOW FUNCTIONS**

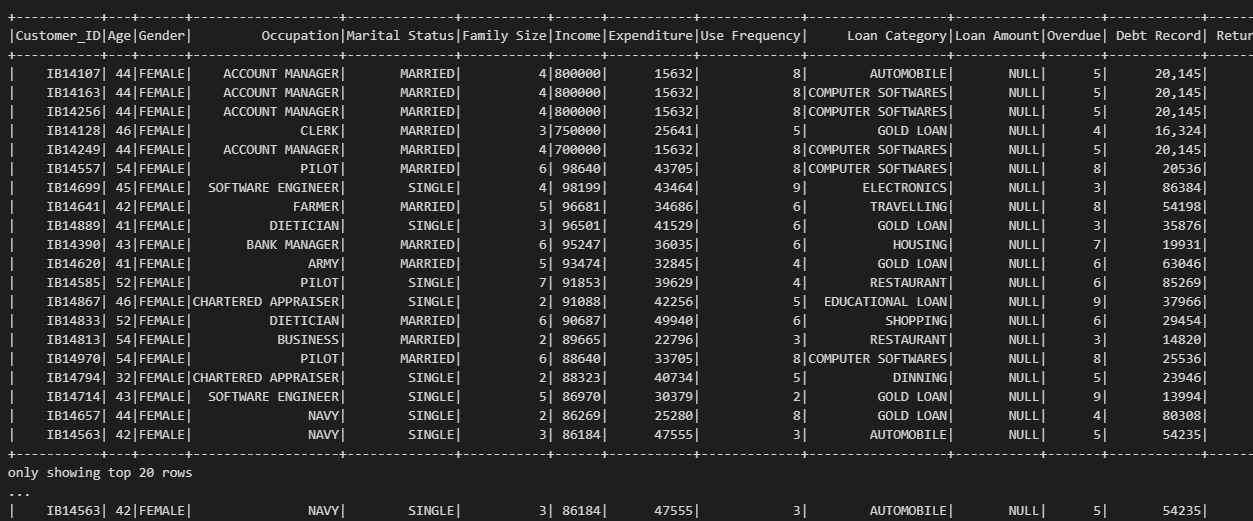
**windowSpec()**





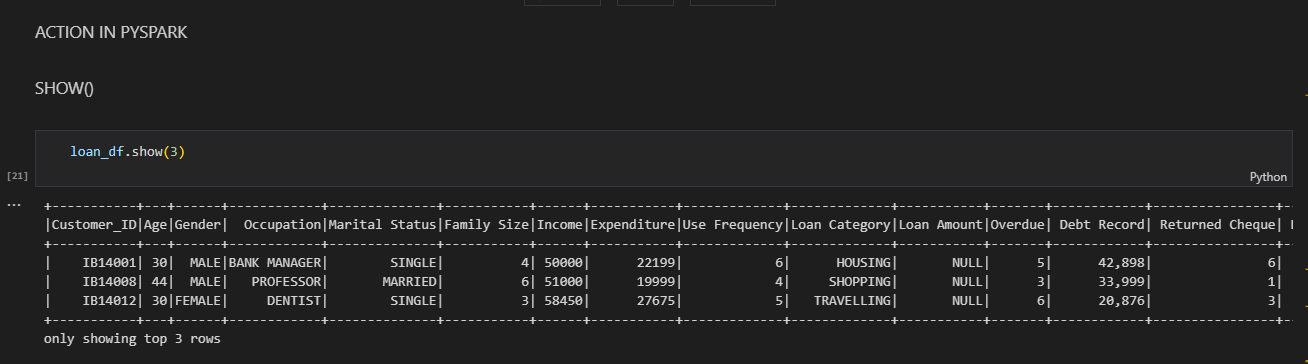
**windowSpec() with rank(), row\_number(), dense\_rank()**



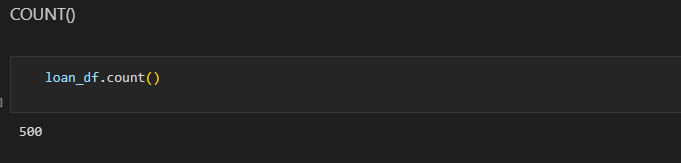


**ACTIONS IN PYSPARK**

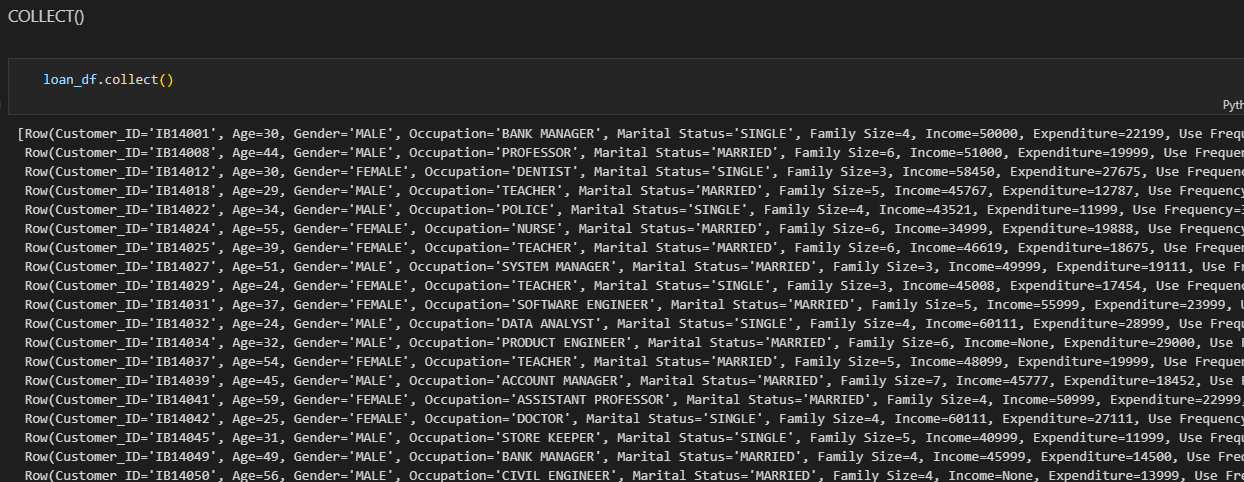
**Show()**



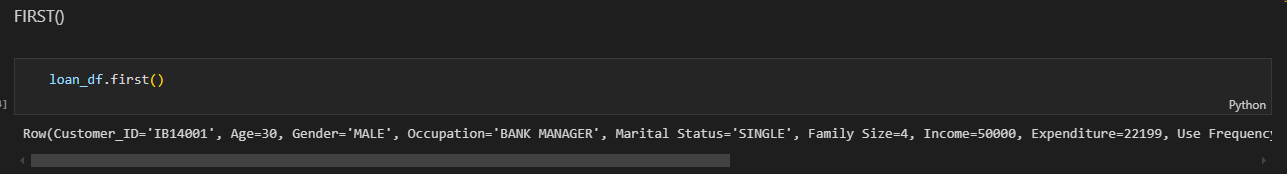
**Count()**



**Collect()**



**First()**



**Take()**

