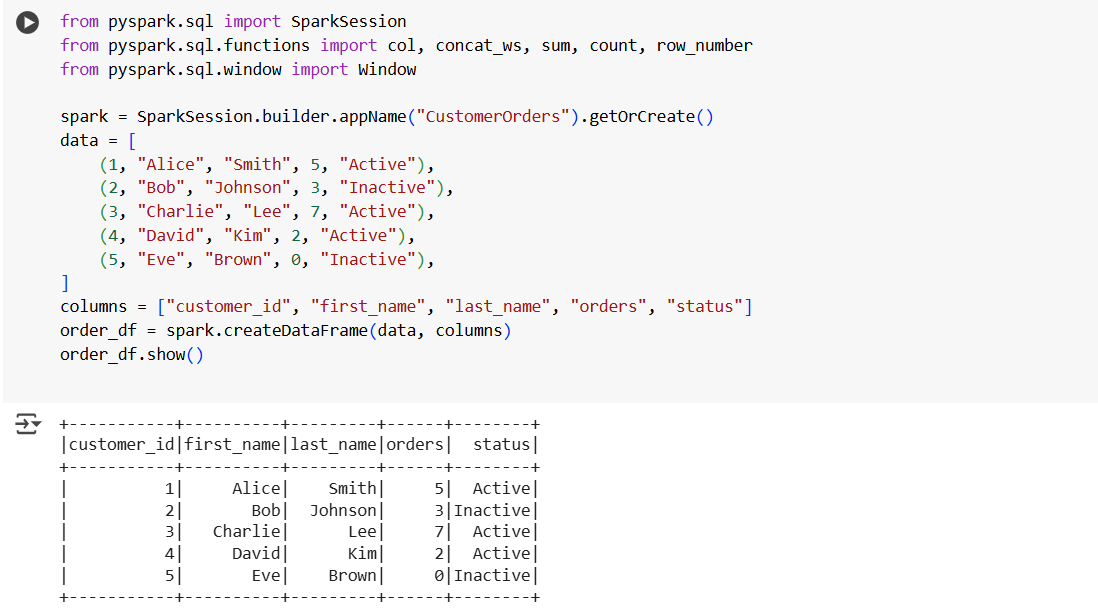
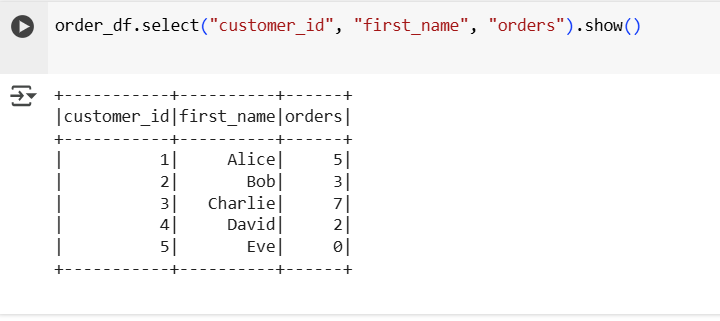
**Coding Assessment – Apply Transformation and actions in Pyspark**

**By Afreen Sultana A**

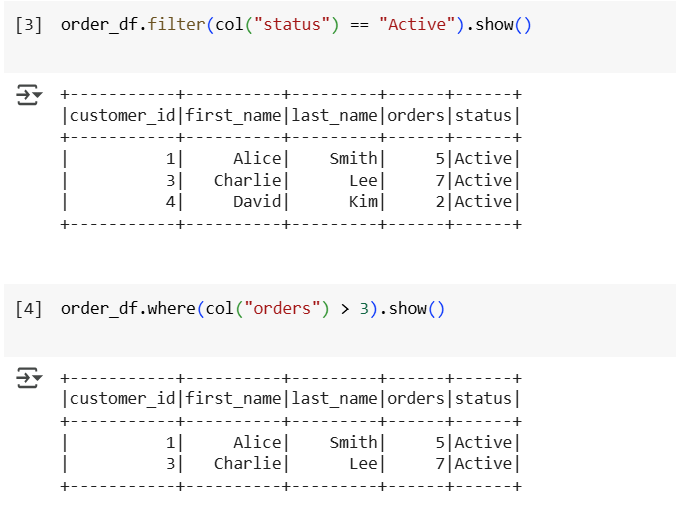
**Transformation**



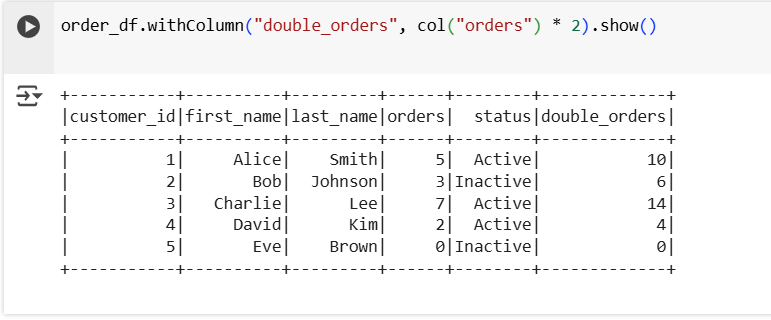
1. **select()** – Select specific columns



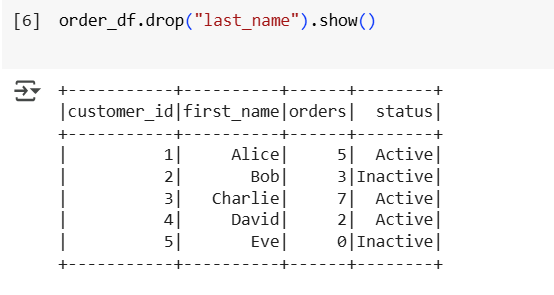
1. **filter() and where()** - Filter rows by condition

****

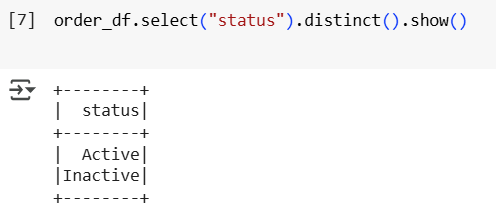
1. **withColumn()** – Add new column or modify existing



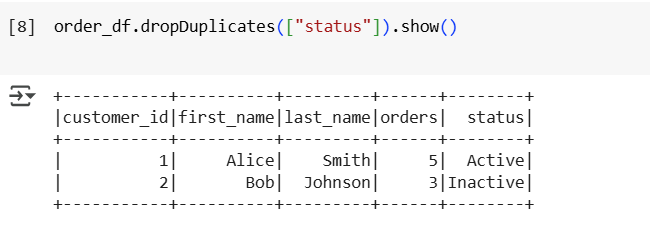
1. **drop()** – Drop a column



1. **distinct()** – Get unique rows



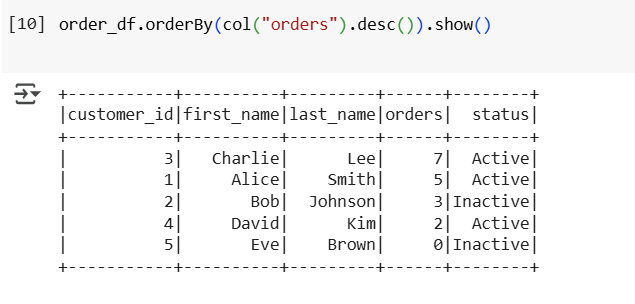
1. **dropDuplicates()** – Drop duplicate rows



1. **groupBy() + agg()** – Group and aggregate



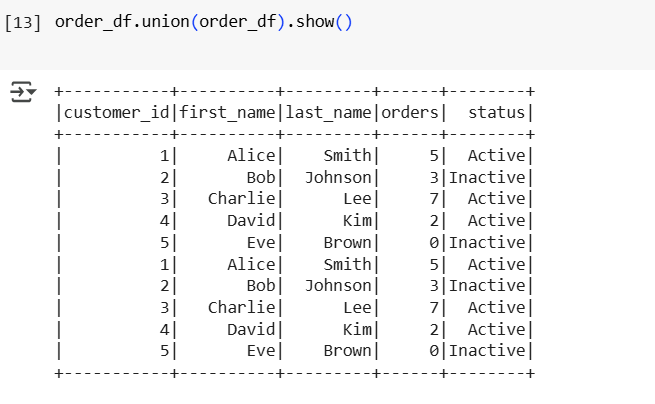
1. **orderBy()**  – Sort rows



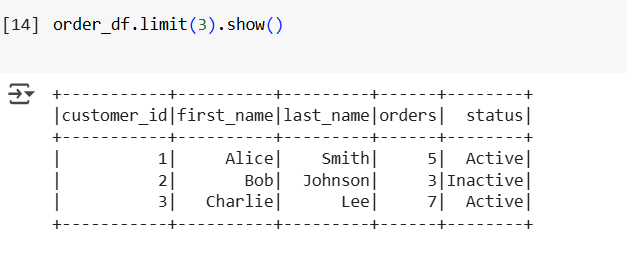
1. **join()** – Join with another DataFrame



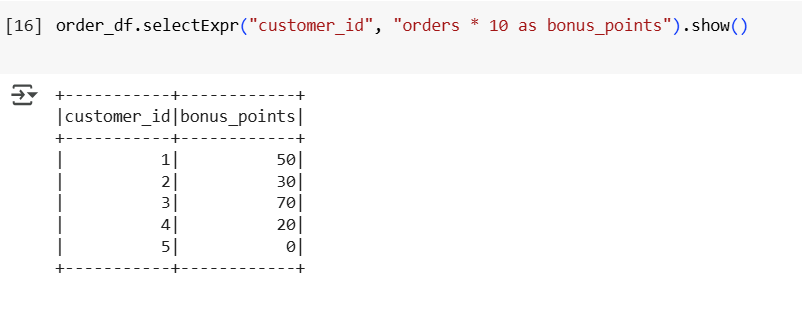
1. **union()** – Combine two DataFrames with same schema



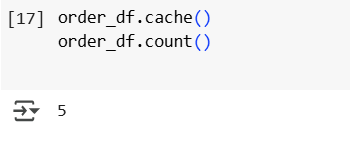
1. **limit()** – Limit rows



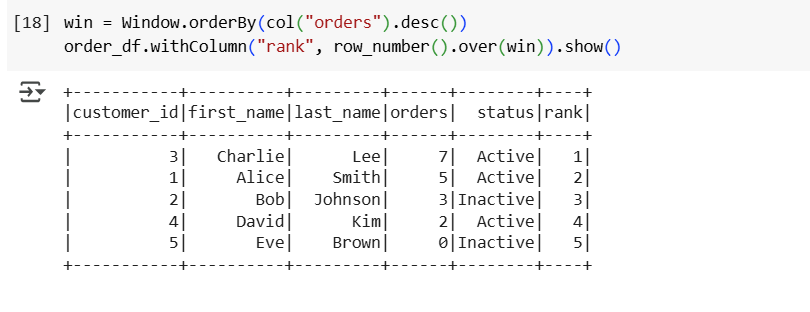
1. **selectExpr()** – SQL expressions inside select



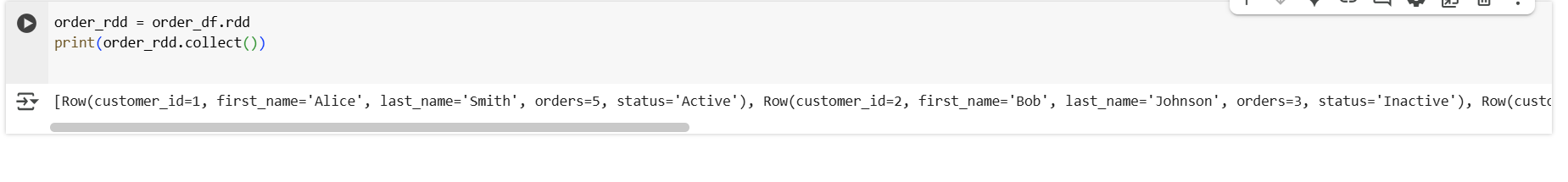
1. **cache()** – Store result in memory



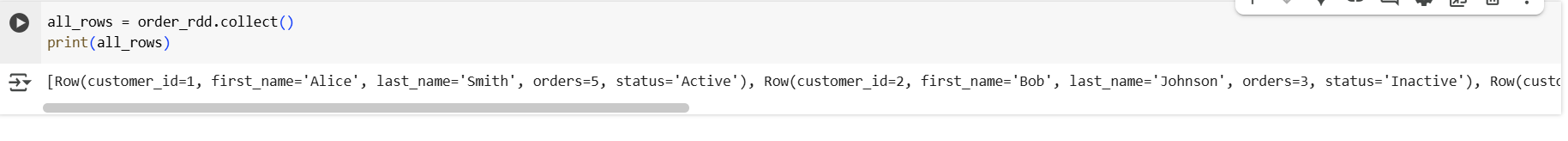
1. **Window Functions** – Ranking, row numbers



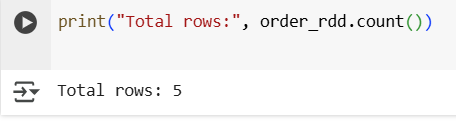
**Actions**

****

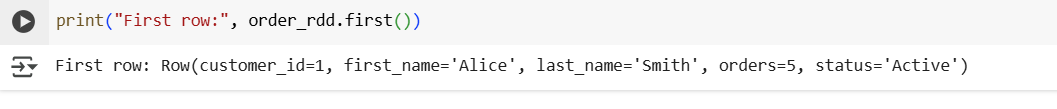
1. **.collect() Action**

****

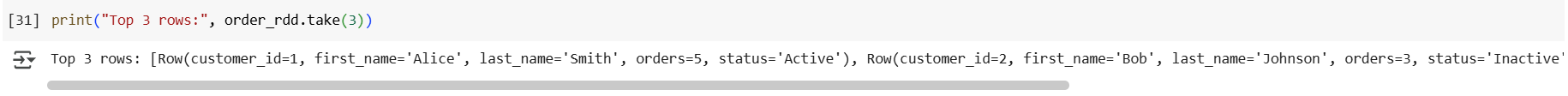
1. **.count() Action**

****

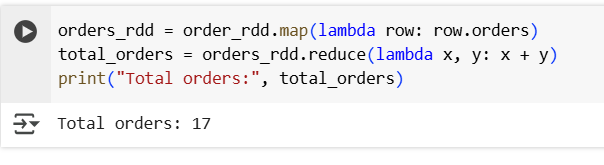
1. **.first() Action**

****

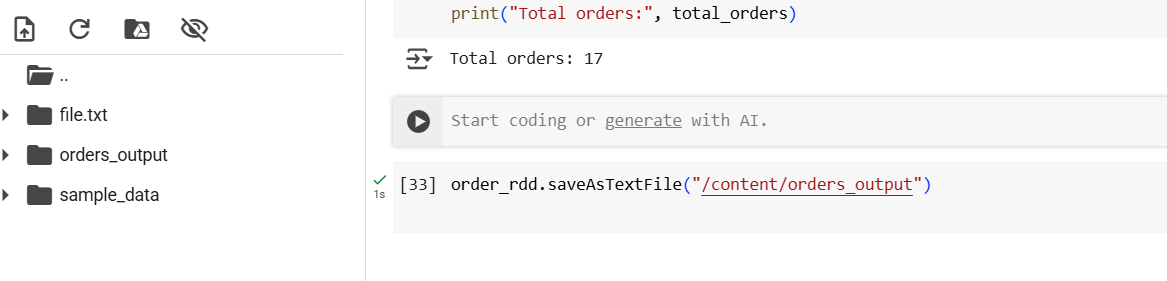
1. **.take(n) Action**

****

1. **.reduce() Action**

****

1. **.saveAsTextFile() Action**

****