Name : Vishal Jha

Batch : Data Engineering

Date : 12-02-2024

Topic : PySpark Coding Challenge

1. Execute Manipulating, Droping, Sorting, Aggregations, Joining, GroupeBy  DataFrames.
2. Creating DataFrames-



1. Join –



1. Group By and Aggregations –



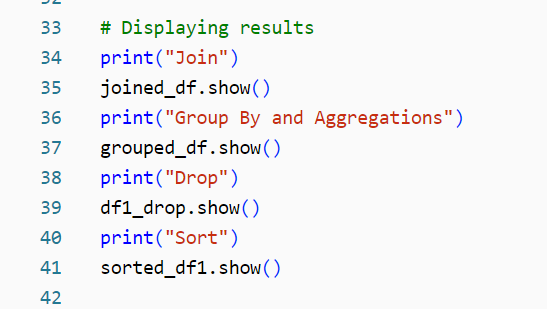
1. Drop –

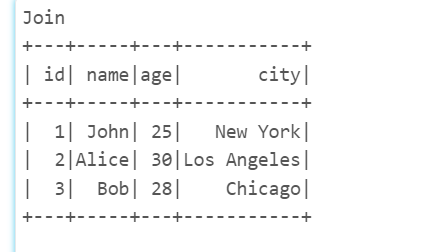


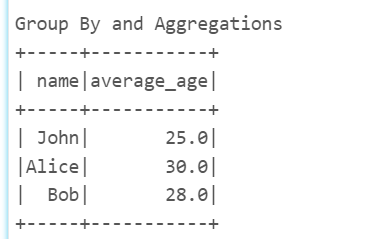
1. Sort –

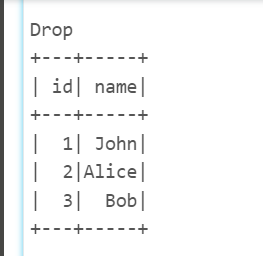


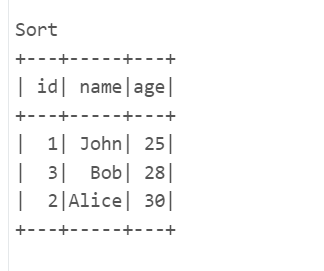
1. Displaying Result –



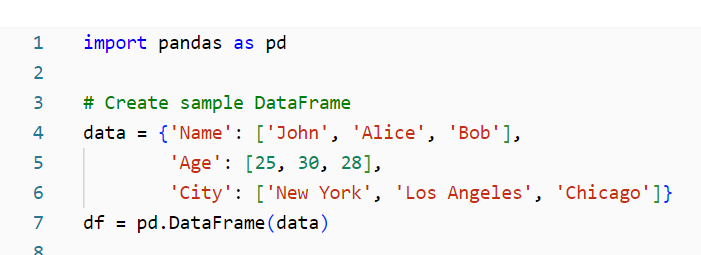




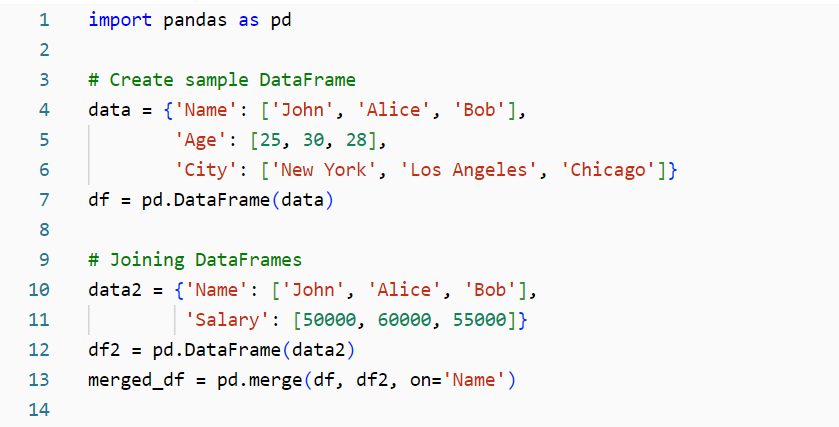




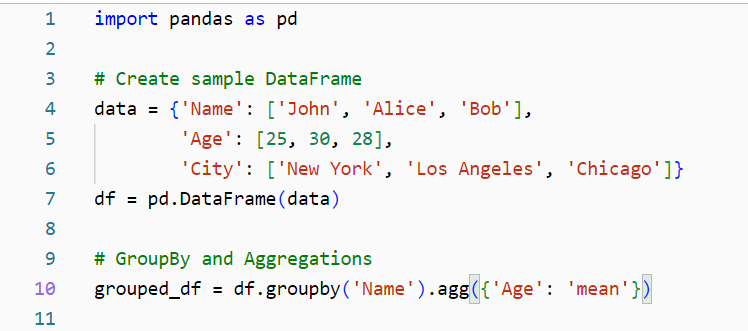
1. Execute Pyspark - SparkSQL joins & Applying Functions in a Pandas DataFrame.
2. Creating Pandas DataFrame –



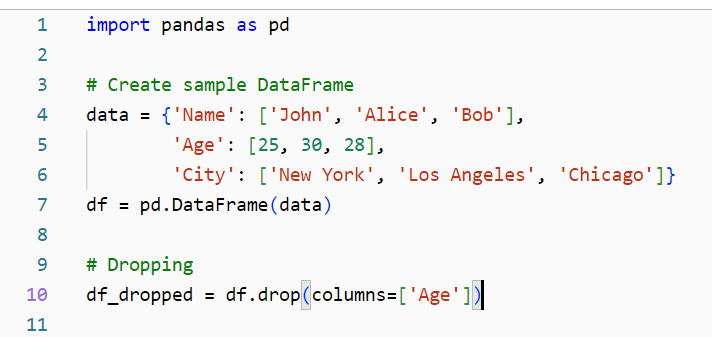
1. Join Function –



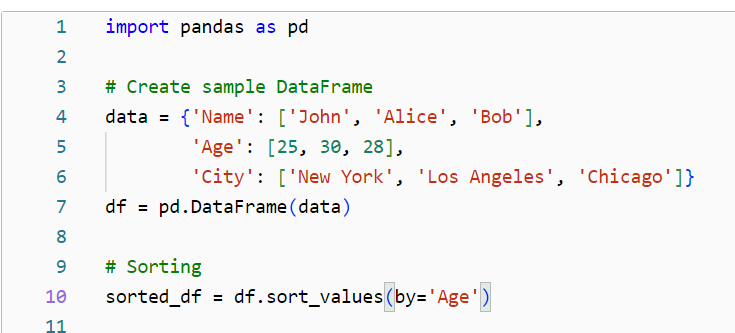
1. Group By and Aggregation Function –



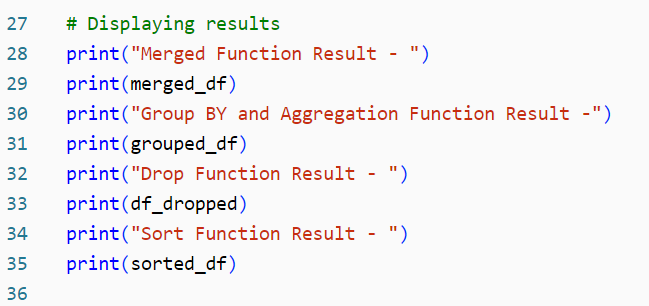
1. Drop Function –

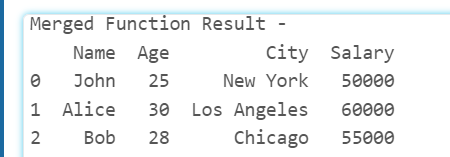


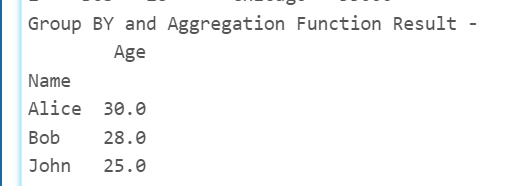
1. Sort Function –

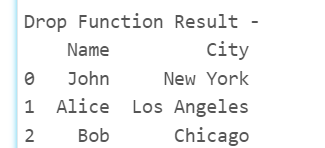


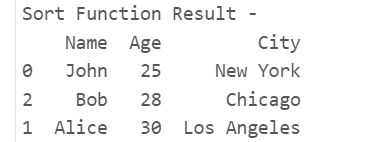
1. Displaying Result of Functions –







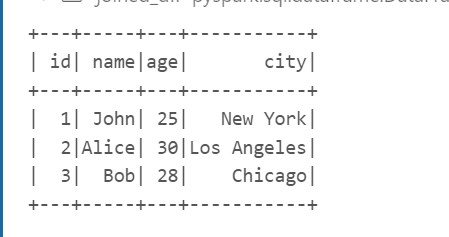




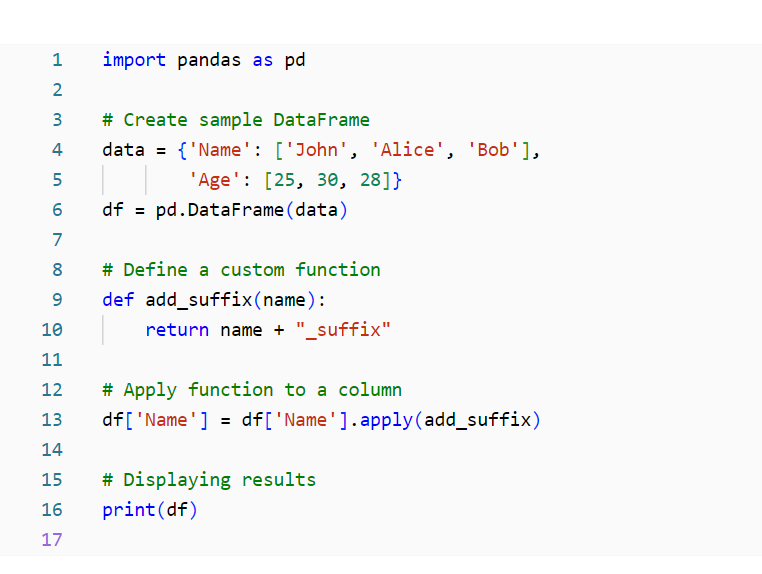
1. PySpark – Spark SQL Join –



Output –



1. Pandas – Applying Function –



Output :

