PS4 Smith

February 18, 2024

I would be interested in scraping the data from the US Bureau of Labor Statistic, focusing on women in the HVAC industry. I read an article from an HVAC publication and website

https://www.achrnews.com/articles/153719-2023-top-women-in-hvac-list

that said only two percent of HVAC employees are women. I would love to dig deeper into that statistic to see more about locations in the US, income and education levels.

1   Question 6.7 Answer

The df1 data frame type is a "tbl-df". The df data frame type is a "tbl-spark".

2   Question 6.8 Answer

The df1 data frame column titles include the ".", such as "Petal.Length". However, in the spark df, they're renamed to replace the period with an underscore, such "Petal-Length". This is done to make sure the names match SQL Spark naming convention.