For this challenge we had to demonstrate use of SparkSQL. We created a repository from home sales data that was provided. We read in the data and then created a temporary table to perform some search functions with SQL. The output to answer each question is below. For the last question, we partitioned the data by date built, created a temporary table for the parquet data. Finally, we uncached the temporary table and verified that it was uncached.

* What is the average price for a four-bedroom house sold for each year? Round off your answer to two decimal places.

A close-up of a date

Description automatically generated

* What is the average price of a home for each year the home was built, that has three bedrooms and three bathrooms? Round off your answer to two decimal places.

A number of numbers on a white background

Description automatically generated with medium confidence

* What is the average price of a home for each year the home was built, that has three bedrooms, three bathrooms, two floors, and is greater than or equal to 2,000 square feet? Round off your answer to two decimal places.

A close-up of a number

Description automatically generated

* What is the average price of a home per "view" rating having an average home price greater than or equal to $350,000? Determine the run time for this query and round off your answer to two decimal places.

A screenshot of a computer

Description automatically generated

* Verified data uncached

A screenshot of a computer program

Description automatically generated