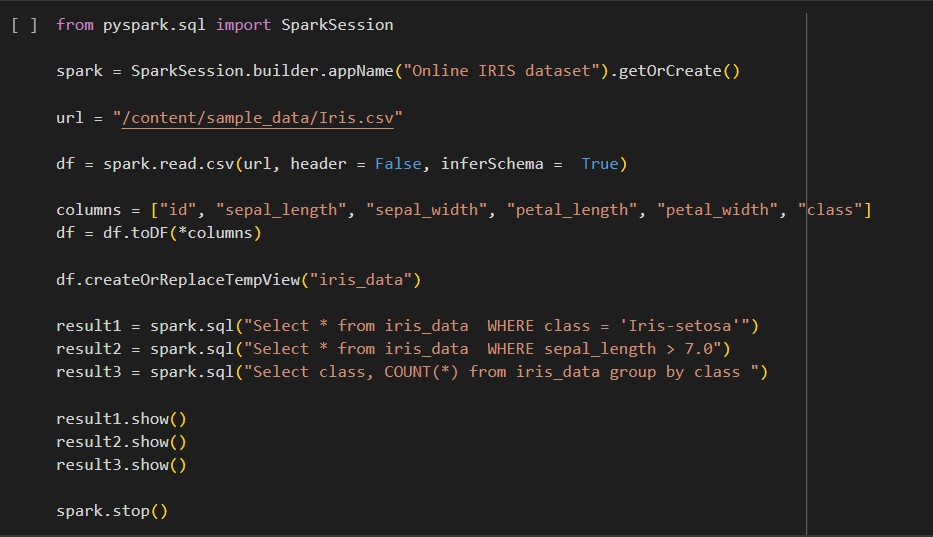
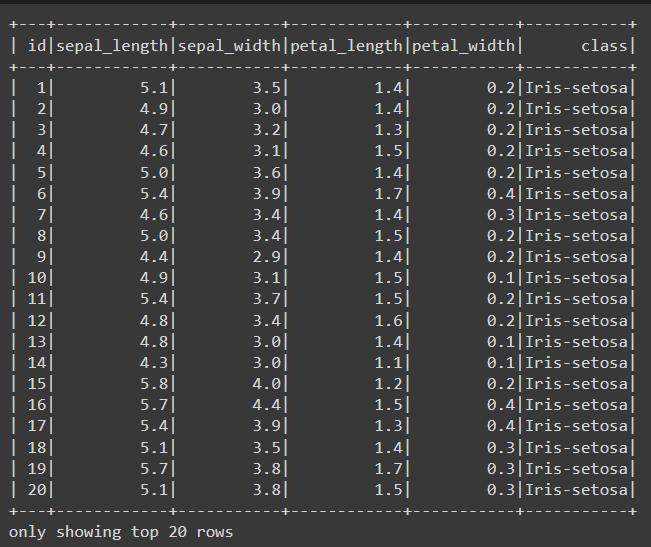
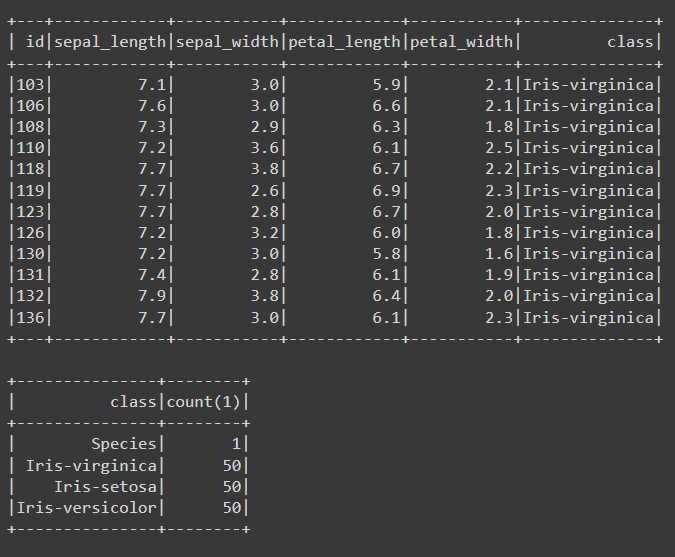
|  |  |
| --- | --- |
| Name: | Prerna Sunil Jadhav |
| Sap Id: | 60004220127 |
| Class: | T. Y. B. Tech (Computer Engineering) |
| Course: | Big Data Infrastructure Laboratory |
| Course Code: | DJ19CEEL6011 |
| Experiment No.: | 08 |

**AIM:** To study and Implement SparkQL using PySpark

**IRIS DATASET:**

**CODE:**

**OUTPUT:**

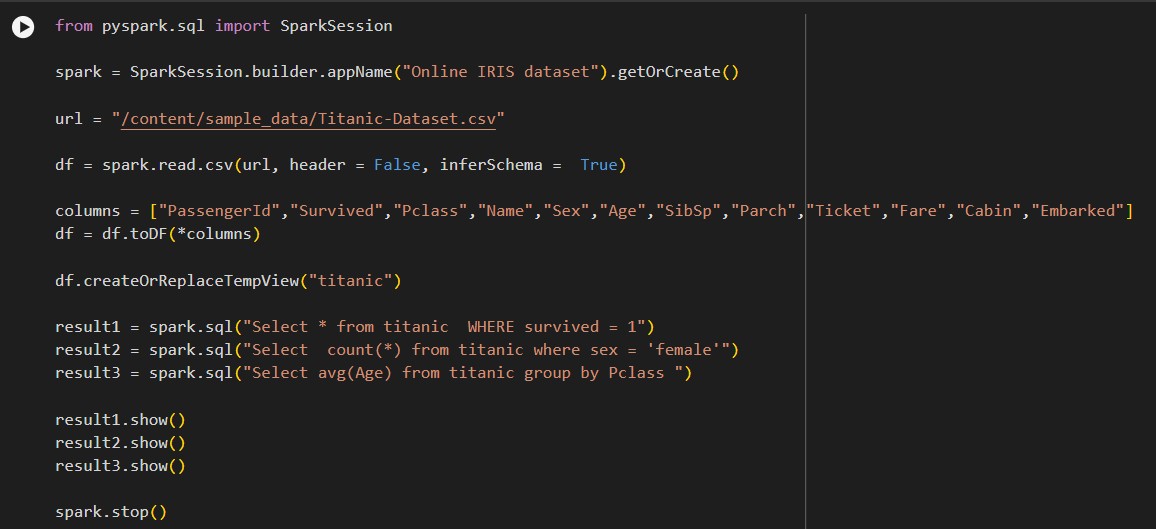


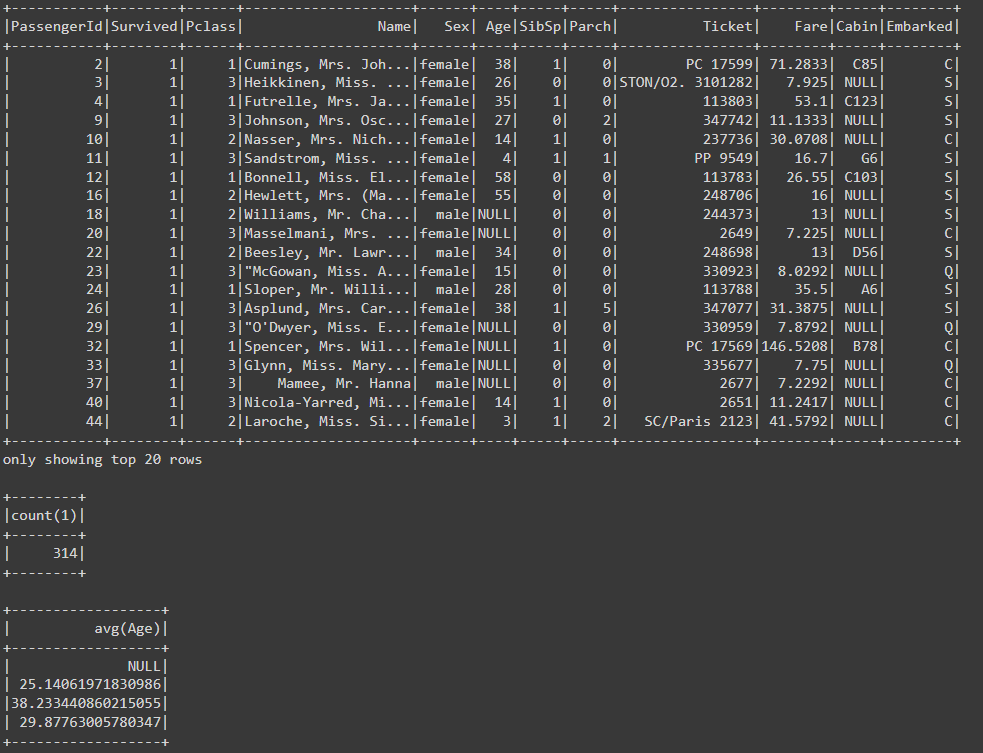
**Titanic Dataset - Perform SQL Queries to find:**

**A. What is the number of passengers who survived the Titanic Disaster?**

**B. How many female passengers were on board the Titanic?**

**C. What is the average age of passengers in each passenger class?**

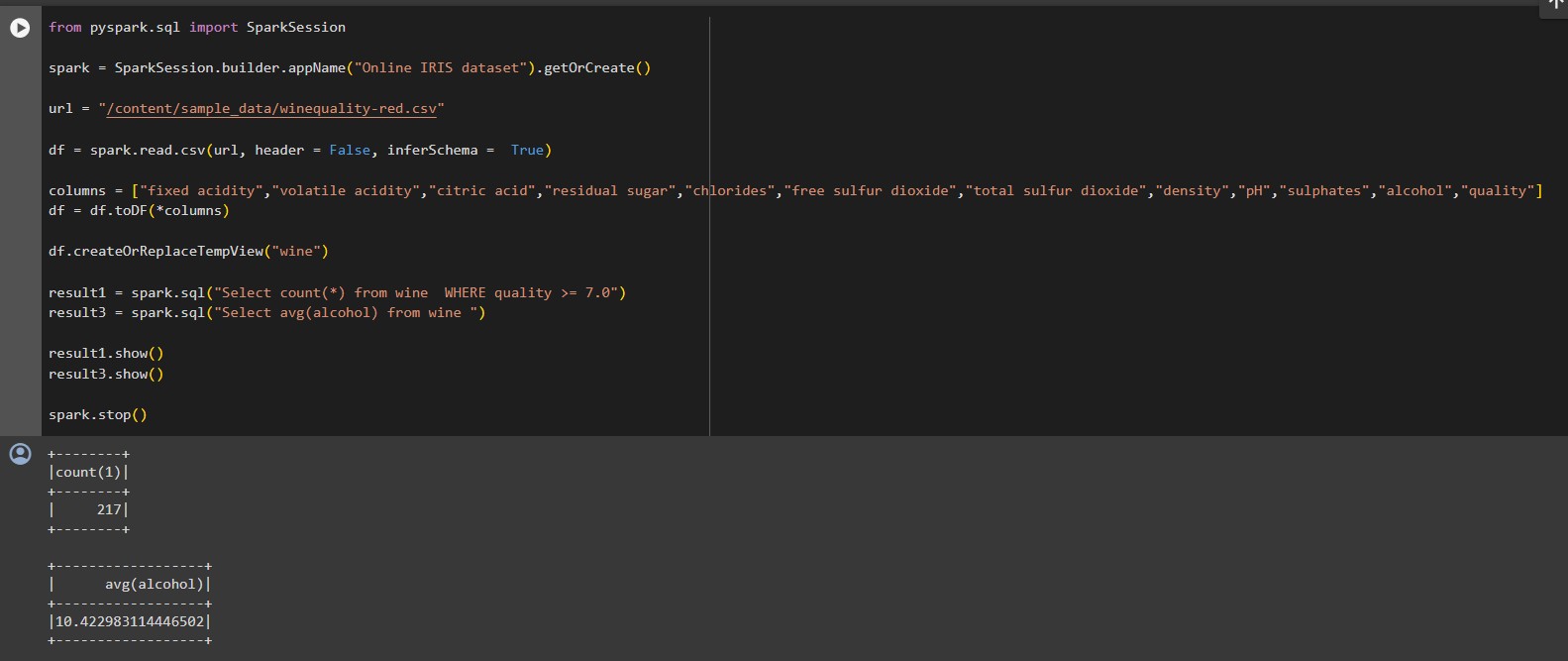
**CODE:**

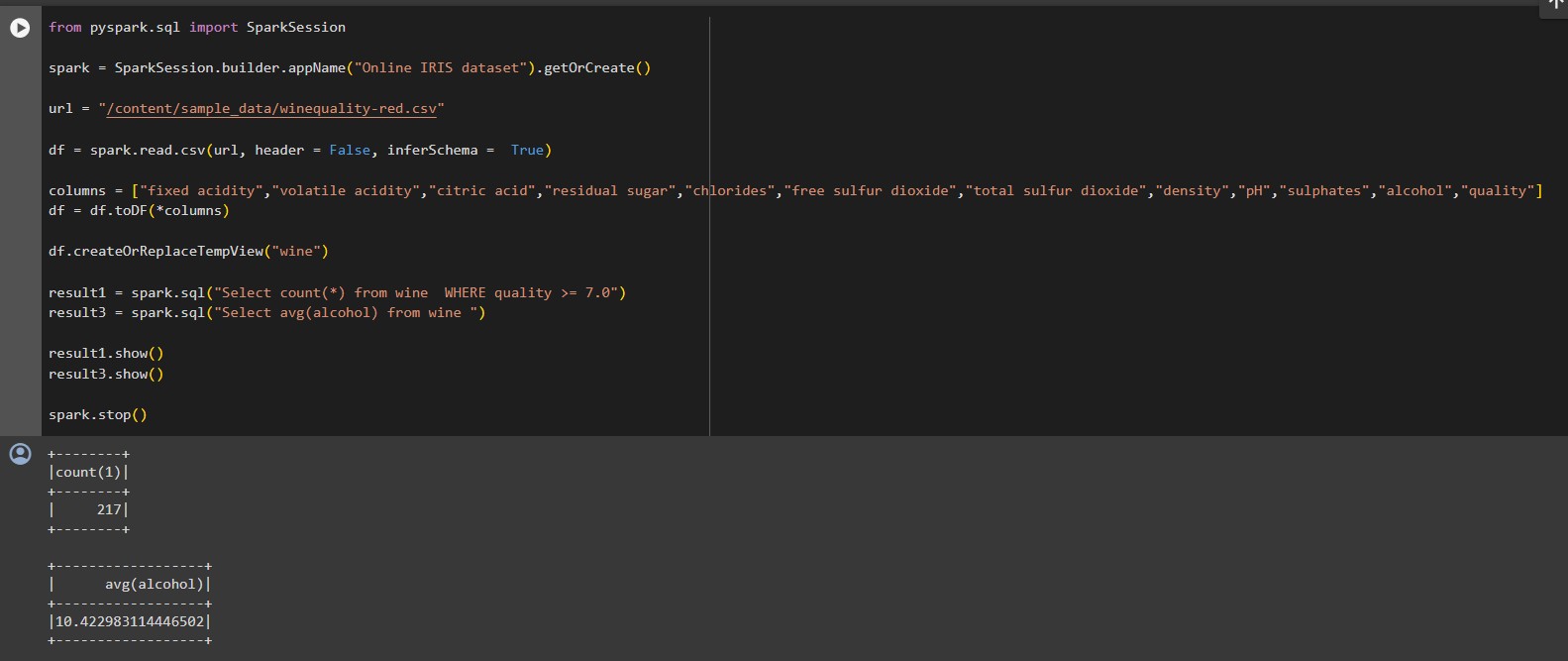
**OUTPUT:**

**Wine Quality Dataset Example**

**A. How many wines are considered high quality (quality score of 7 or higher)**

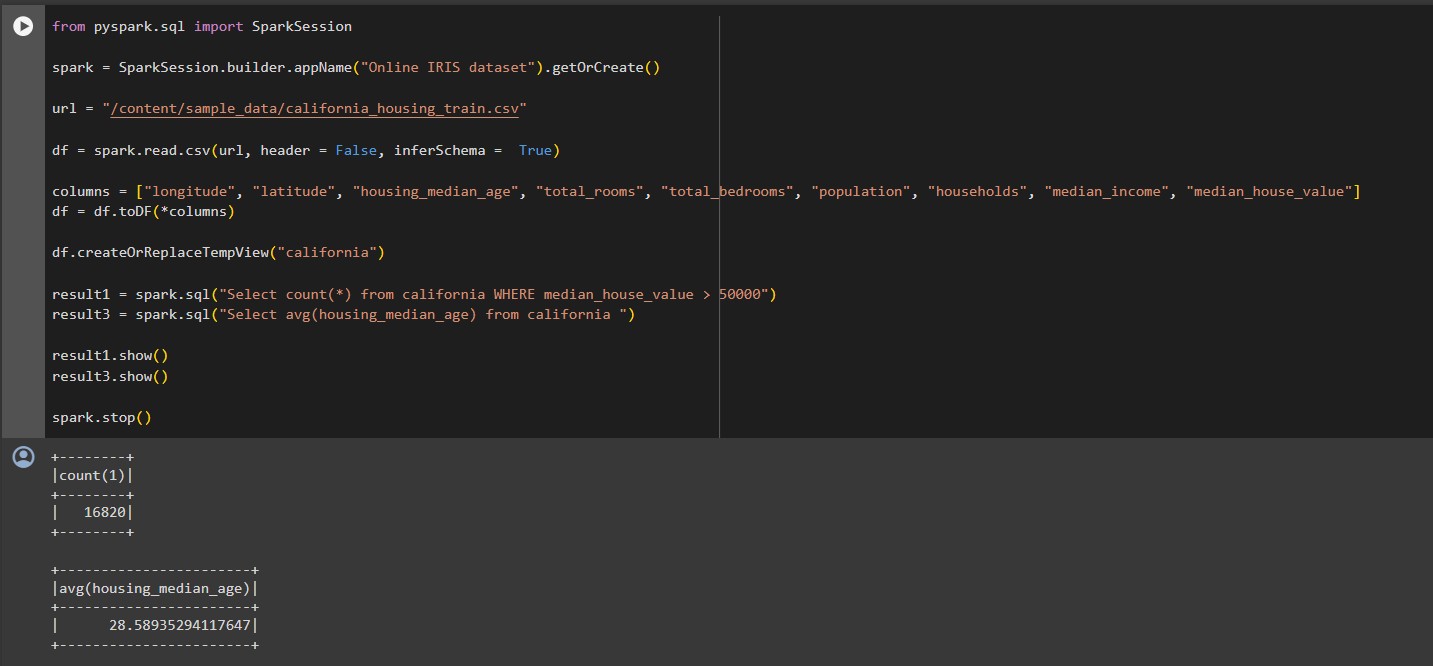
**B. What is the average alcohol content of the wines in the dataset**

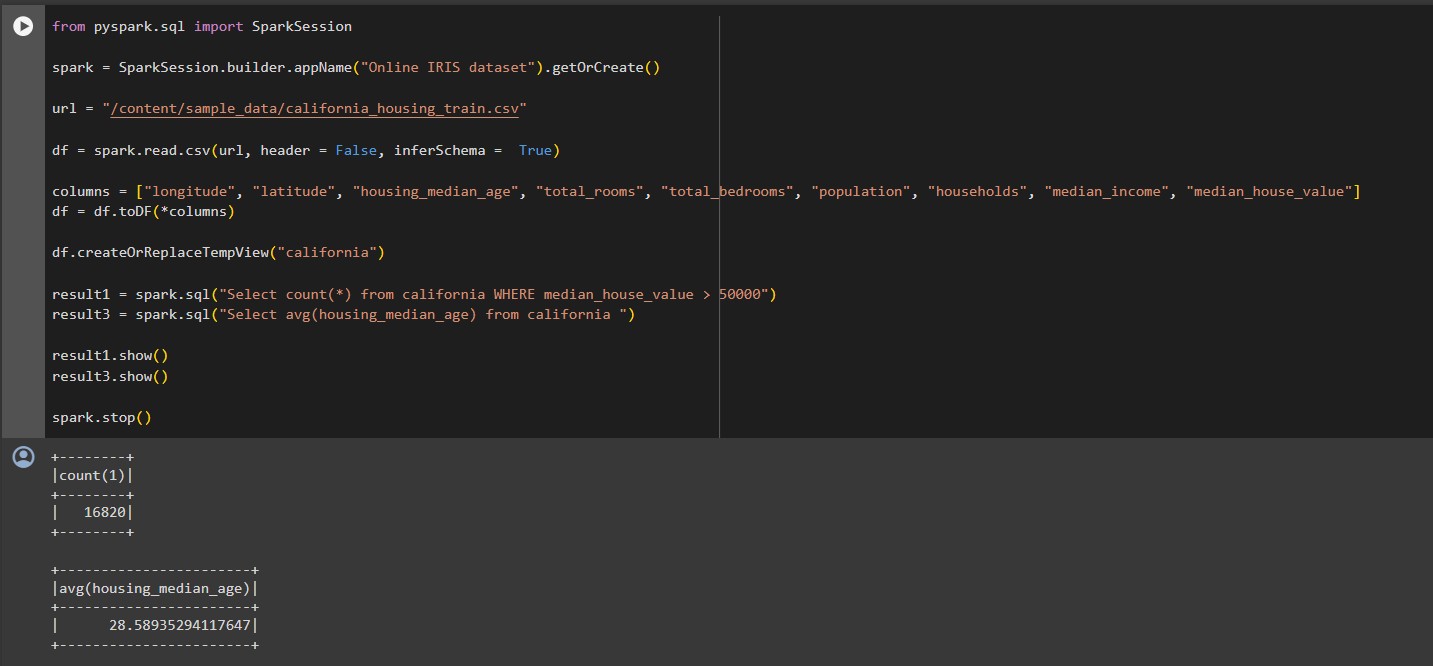
**CODE:**

**OUTPUT:**

**California Housing Dataset Example:**

1. **How many houses have a median value above $5000,000 in California**
2. **What is the average age of the houses in the dataset**

**CODE:**

**OUTPUT:**