1. Conduct exploratory data analysis on the fraud\_data.csv data set and explain whether any customer attributes are associated with fraud. The file contains a list of 50,000 sales and includes a customer ID column, four customer attribute columns (X1-X4), and one column indicating whether the customer committed fraud (fraud occurred when Y=1).
2. Write a SQL select statement that returns a list of customers who enrolled and did not pay a deposit. You have access to two tables: “enrollments” and “deposit\_payments”. Specify which SQL syntax you used.

|  |  |
| --- | --- |
| **5 sample rows from "enrollments" table** | |
| **customer\_id** | **enrollment\_date** |
| 1 | 1/5/2021 |
| 2 | 1/5/2021 |
| 3 | 1/6/2021 |
| 4 | 1/7/2021 |
| 5 | 1/8/2021 |

|  |  |
| --- | --- |
| **5 sample rows from "deposit\_payments" table** | |
| **customer\_id** | **payment\_amount** |
| 4 | 16 |
| 6 | 6 |
| 10 | 2 |
| 11 | 9 |
| 22 | 13 |

1. Write a SQL select statement that calculates the 90th percentile height of Baby Boomer men from the table named “height\_data”. Assume the measurements were recorded today. Specify which SQL syntax you used.

|  |  |  |  |
| --- | --- | --- | --- |
| **5 sample rows from "height\_data" table** | | | |
| **height** | **weight** | **age** | **male** |
| 151.765 | 47.8256065 | 63 | 1 |
| 139.7 | 36.4858065 | 63 | 0 |
| 136.525 | 31.864838 | 65 | 0 |
| 156.845 | 53.0419145 | 41 | 1 |
| 145.415 | 41.276872 | 51 | 0 |