**Model Selection and Variable selection**

**Problem:**

You are a VP of sales and have responsibility for 41 stores.  You have collected data from the stores on advertising costs, store size in square feet, % employee retention, customer satisfaction score, whether a promotion was run or not and sales.  You want to build a model that can predict sales based on these five variables.

Establish a best multiple regression model to predict the sales using forward, backward and stepwise procedure.

The data are shown in Table below..

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Store | Advertising Costs | Size (Sq. Ft) | % Employee Retention | Customer Satisfaction | Promotion | Sales |
| 1 | 124.4 | 22560 | 59 | 32 | 1 | 1581 |
| 2 | 154 | 31181 | 62 | 33 | 1 | 2139 |
| 3 | 123.5 | 16314 | 78 | 28 | 0 | 1043 |
| 4 | 163 | 24205 | 66 | 43 | 0 | 1702 |
| 5 | 107 | 17574 | 82 | 22 | 1 | 1339 |
| 6 | 143.9 | 19584 | 67 | 34 | 0 | 521 |
| 7 | 133.7 | 22682 | 57 | 32 | 1 | 1720 |
| 8 | 121.4 | 23398 | 64 | 23 | 1 | 1197 |
| 9 | 104.6 | 19507 | 88 | 25 | 0 | 950 |
| 10 | 99.2 | 11443 | 87 | 17 | 0 | 266 |
| 11 | 93.8 | 16832 | 82 | 29 | 1 | 1718 |
| 12 | 133.8 | 24326 | 70 | 37 | 1 | 1820 |
| 13 | 131.3 | 18541 | 86 | 27 | 1 | 1805 |
| 14 | 123.2 | 22099 | 67 | 30 | 0 | 1042 |
| 15 | 88.3 | 16928 | 80 | 21 | 0 | 655 |
| 16 | 154.3 | 16237 | 69 | 28 | 1 | 1480 |
| 17 | 112.1 | 15290 | 87 | 19 | 1 | 1057 |
| 18 | 114 | 12947 | 80 | 23 | 0 | 953 |
| 19 | 91.7 | 16326 | 77 | 24 | 0 | 364 |
| 20 | 113.7 | 13024 | 82 | 27 | 0 | 783 |
| 21 | 105.7 | 22054 | 86 | 28 | 1 | 792 |
| 22 | 161.3 | 22637 | 75 | 34 | 1 | 2185 |
| 23 | 143 | 18733 | 64 | 26 | 1 | 1051 |
| 24 | 113.7 | 21126 | 58 | 30 | 0 | 1456 |
| 25 | 69 | 16819 | 57 | 18 | 0 | 146 |
| 26 | 106.7 | 16992 | 74 | 22 | 0 | 899 |
| 27 | 125.8 | 18355 | 88 | 25 | 1 | 1243 |
| 28 | 52.4 | 13958 | 87 | 25 | 0 | 421 |
| 29 | 114.7 | 18298 | 71 | 24 | 0 | 318 |
| 30 | 142.5 | 18016 | 55 | 27 | 0 | 383 |
| 31 | 114.6 | 22317 | 80 | 27 | 1 | 993 |
| 32 | 150.5 | 26221 | 71 | 31 | 1 | 1766 |
| 33 | 74.5 | 19494 | 68 | 24 | 1 | 1123 |
| 34 | 111.4 | 18406 | 85 | 24 | 1 | 1523 |
| 35 | 117.7 | 18880 | 86 | 24 | 1 | 1281 |
| 36 | 108.4 | 21312 | 81 | 30 | 1 | 1899 |
| 37 | 171.7 | 26618 | 72 | 35 | 1 | 2508 |
| 38 | 156.5 | 15981 | 72 | 25 | 0 | 1042 |
| 39 | 156.3 | 20749 | 61 | 32 | 0 | 1416 |
| 40 | 140.6 | 22458 | 77 | 27 | 1 | 1659 |
| 41 | 155.1 | 18579 | 85 | 32 | 0 | 1370 |

The evaluation pattern is as follows:

|  |  |  |
| --- | --- | --- |
| Section | Parameters | Marks |
| A | Objective/Aim | 2 |
| B | Analysis | 3 |
| C | Interpretation | 3 |
| D | Timely submission | 2 |
| Total |  | 10 |