A model has been built to predict which location should develop a Slacks on Racks. The model will use sales to predict which location will be the most profitable. Data from 68 existing stores was given for consideration for the final model. The variables that were considered were used to predict annual sales for a Slacks on Racks.

Store Variables

A report from IBIS World shows the major market segmentation of Discount Department Stores in 2018.

Major market segmentation (2018) 
10.5% 
aged 25 
14.7% 
Consumers aged 
65 and 
16.3% 
Consumers 
aged 25 to 34 
20.8% 
Consumers aged 35 to 44 
20.6% 
Consumers 
aged 45 to 54 
Total $97.7bn 
17.1% 
Consumers aged 55 to 64 
SOURCE: p.m•ng1SWORLOCOM 

The variables will automatically be considered for the model, along with other variables. The variables in the data that most closely parallels the variables in the diagram are listed below along with the Pearson Correlation with sales.

|  |  |  |  |
| --- | --- | --- | --- |
| AGE\_MIDADLT2534\_8TO | Age: Male Middle Adult (25-34) -0.309 |  | ssdd |
| AGE\_MIDLFE3544\_8TO | Age: Male Midlife Adult (35-44) -0.397 |  |  |
| AGE\_SENIOR65P\_8TO | Age: Senior (65+) -0.030 |  |  |
| AGE\_BABYBOOM5069\_16TO | Age: Baby Boomer (50-69) -0.136 |  |  |
| AGE\_GENX4049\_8TO | Age: GenX (40-49) -0.348 |  |  |
| AGE\_ELEMIDSCHL0514\_8TO | Age: Elementary/Middle School (5-14) -0.366 |  |  |
| AGE\_YNGADLT1824\_16TO | Age: Young Adult (18-24) -0.197 |  |  |

Other Variables

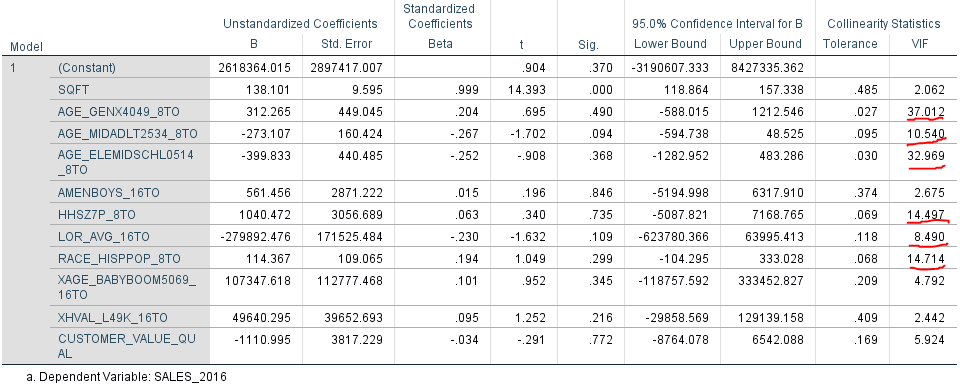
Running a bivariate correlation comparing the sales versus the other variables show the Pearson Correlation between each variable. The variable with the strongest relation with sales is square feet, with a Pearson Correlation of 0.914. Other variables with high Pearson Correlation will be considered for the model are listed below.

|  |  |
| --- | --- |
| Variable | Pearson Correlation with Sales |
| SQFT | 0.914 |
| AGE\_MIDADLT2534\_8TO | -0.385 |
| AGE\_MINOR0417\_8TO | -0.375 |
| AMENBOYS\_16TO | -0.361 |
| HHSZ3P\_8TO | -0.352 |
| HHSZ6\_8TO | -0.421 |
| HHSZ7P\_8TO | -0.440 |
| LOR\_AVG\_16TO | 0.467 |
| RACE\_HISPPOP\_8TO | -0.472 |
| XAGE\_BABYBOOM5069\_16TO | 0.413 |
| XHVAL\_L49K\_16TO | 0.390 |
| CUSTOMER\_VALUE\_QUAL | 0.377 |

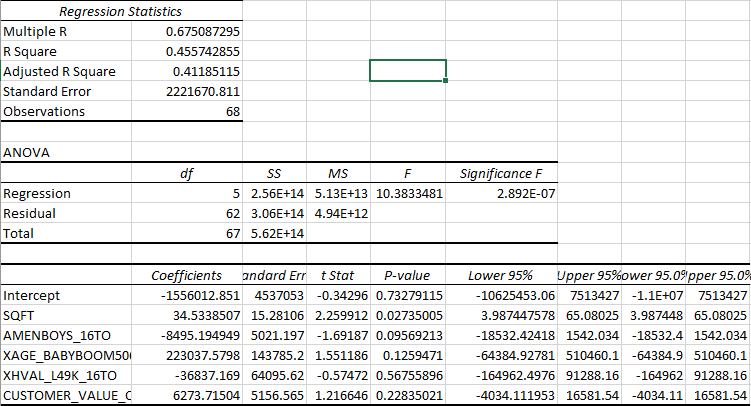
Some of the redundant variables will be removed such as HHSZ6\_8TO: Household Size 6 and HHSZ3P\_8TO: Household Size 3. AGE\_SENIOR65P\_8TO, AGE\_BABYBOOM5069\_16TO, AGE\_YNGADLT1824\_16TO and AGE\_YNGADLT1824\_16TO will be removed as well since they have a low correlation. The rest of the variables are below will be put in a regression to interpret the results

|  |  |
| --- | --- |
| Variable | Pearson Correlation with Sales |
| SQFT | 0.914 |
| AGE\_GENX4049\_8TO | -0.348 |
| AGE\_MIDADLT2534\_8TO | -0.385 |
| AGE\_MINOR0417\_8TO | -0.375 |
| AGE\_MIDADLT2534\_8TO | -0.309 |
| AGE\_MIDLFE3544\_8TO | -0.397 |
| AGE\_ELEMIDSCHL0514\_8TO | -0.366 |
| AMENBOYS\_16TO | -0.361 |
| HHSZ7P\_8TO | -0.440 |
| LOR\_AVG\_16TO | 0.467 |
| RACE\_HISPPOP\_8TO | -0.472 |
| XAGE\_BABYBOOM5069\_16TO | 0.413 |
| XHVAL\_L49K\_16TO | 0.390 |
| CUSTOMER\_VALUE\_QUAL | 0.377 |

Running the regression shows a couple of problems that needs to be addressed before continuing. Some of the independent variables have a high Variance Inflation Factor, which indicates multicollinearity between the independent variables. The variables highlighted in the figure below will be eliminated to improve the accuracy of the model.



Running the regression with the final variables we have these results below.



Based on this model it has shown that store 21266495 will have the most predicted sales.

