**Predicting New Store Location.**

**Part 2 – Building the Model**

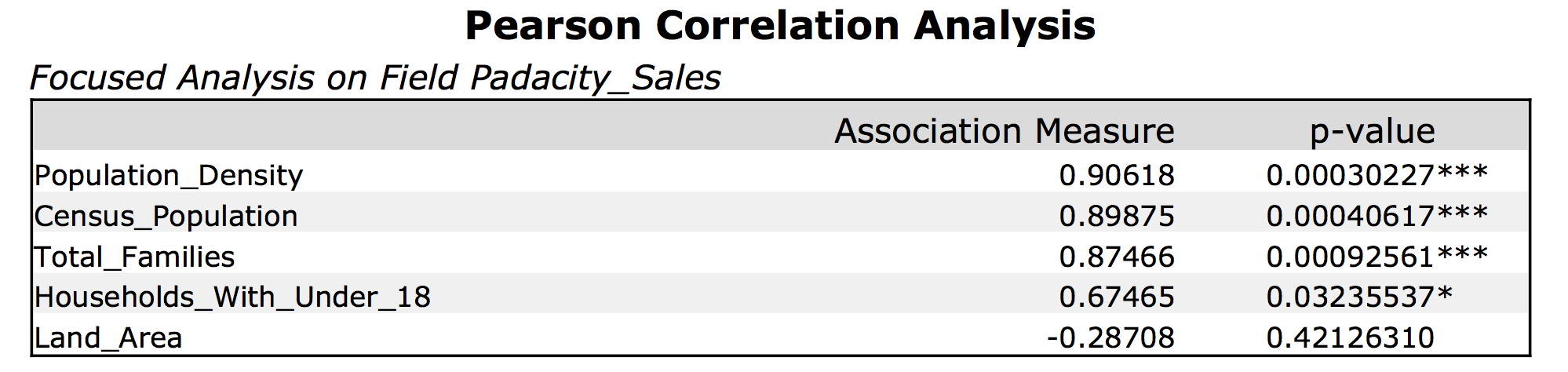
**Creating the model.**

Below is the final dataset used for the regression model.

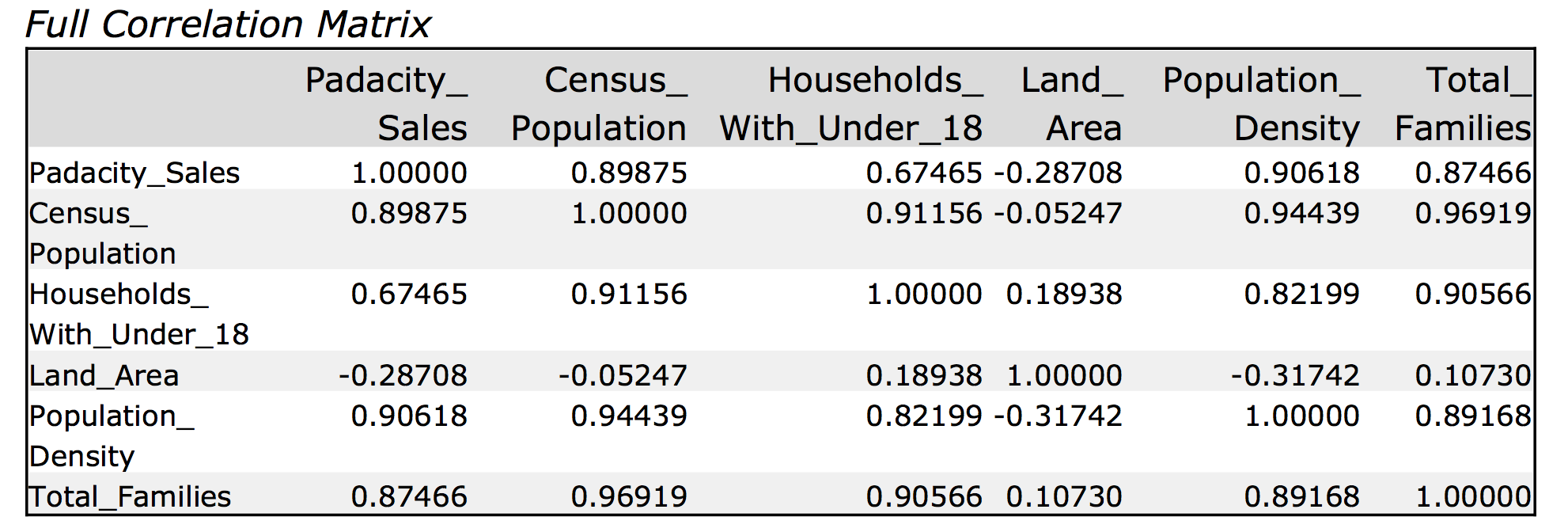
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **City** | **Census\_Population** | **Household\_with\_Under\_18** | **Land\_Area** | **Padacity\_Sales** | **Population\_Density** | **Total\_Families** |
| **Buffalo** | 4585 | 746 | 3115.5075 | 185328 | 1.55 | 1819.5 |
| **Casper** | 35316 | 7788 | 3894.3091 | 317736 | 11.16 | 8756.32 |
| **Cheyenne** | 59466 | 7158 | 1500.1784 | 917892 | 20.34 | 14612.64 |
| **Cody** | 9520 | 1403 | 2998.95696 | 218376 | 1.82 | 3515.62 |
| **Douglas** | 6120 | 832 | 1829.4651 | 208008 | 1.46 | 1744.08 |
| **Evanston** | 12359 | 1486 | 999.4971 | 283824 | 4.95 | 2712.64 |
| **Powell** | 6314 | 1251 | 2673.57455 | 233928 | 1.62 | 3134.18 |
| **Riverton** | 10615 | 2680 | 4796.859815 | 303264 | 2.34 | 5556.49 |
| **Rock Springs** | 23036 | 4022 | 6620.201916 | 253584 | 2.78 | 7572.18 |
| **Sheridan** | 17444 | 2646 | 1893.977048 | 308232 | 8.98 | 6039.71 |

**Selecting the predictor variables.**

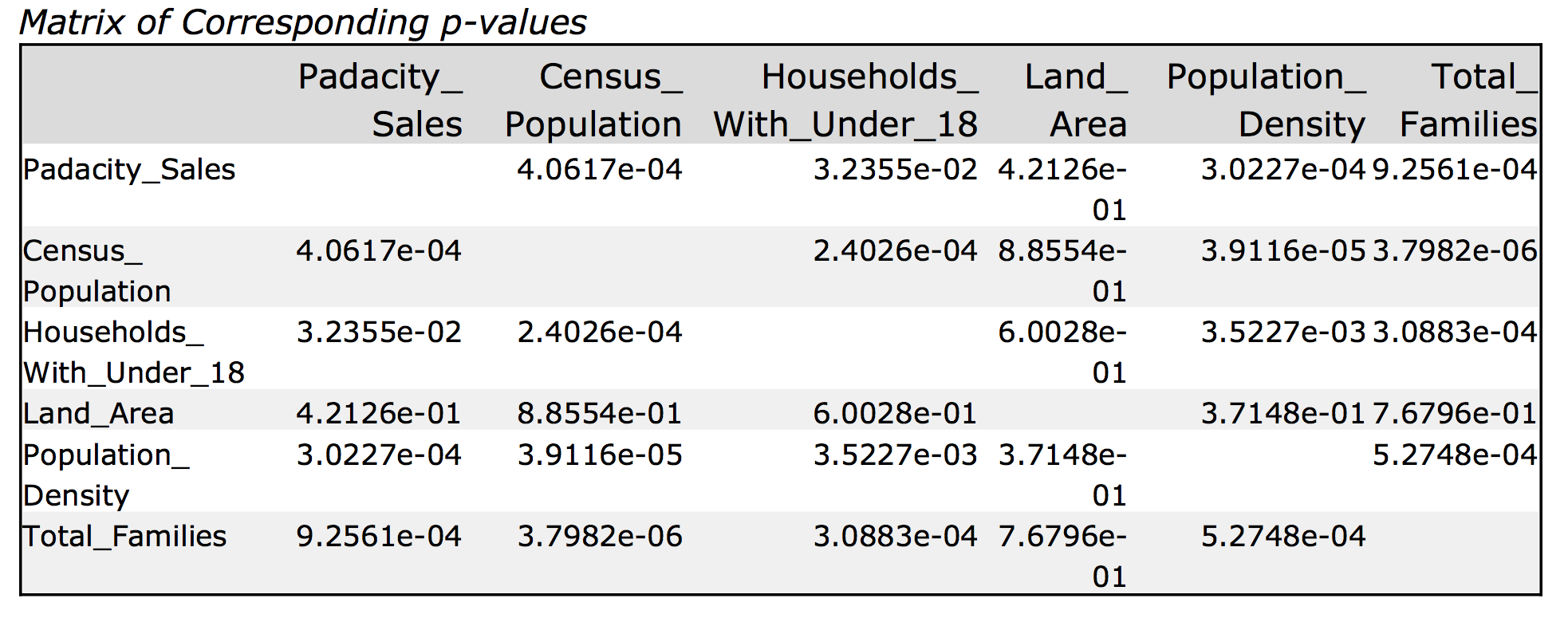
Below is a table of all the variables and their Pearson correlation.



Full correlation matrix.

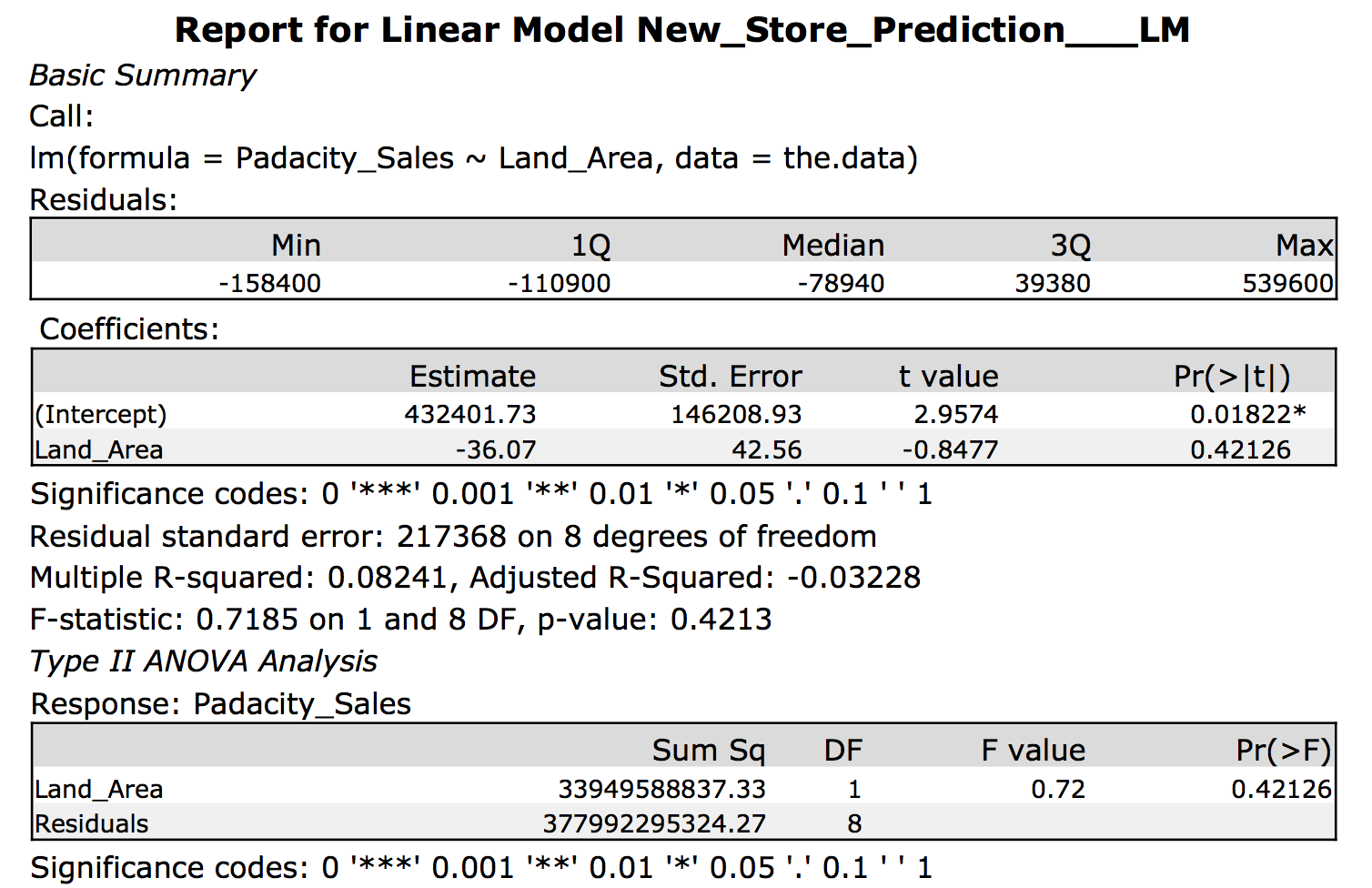


Matrix of p-values for Predictor Variables.

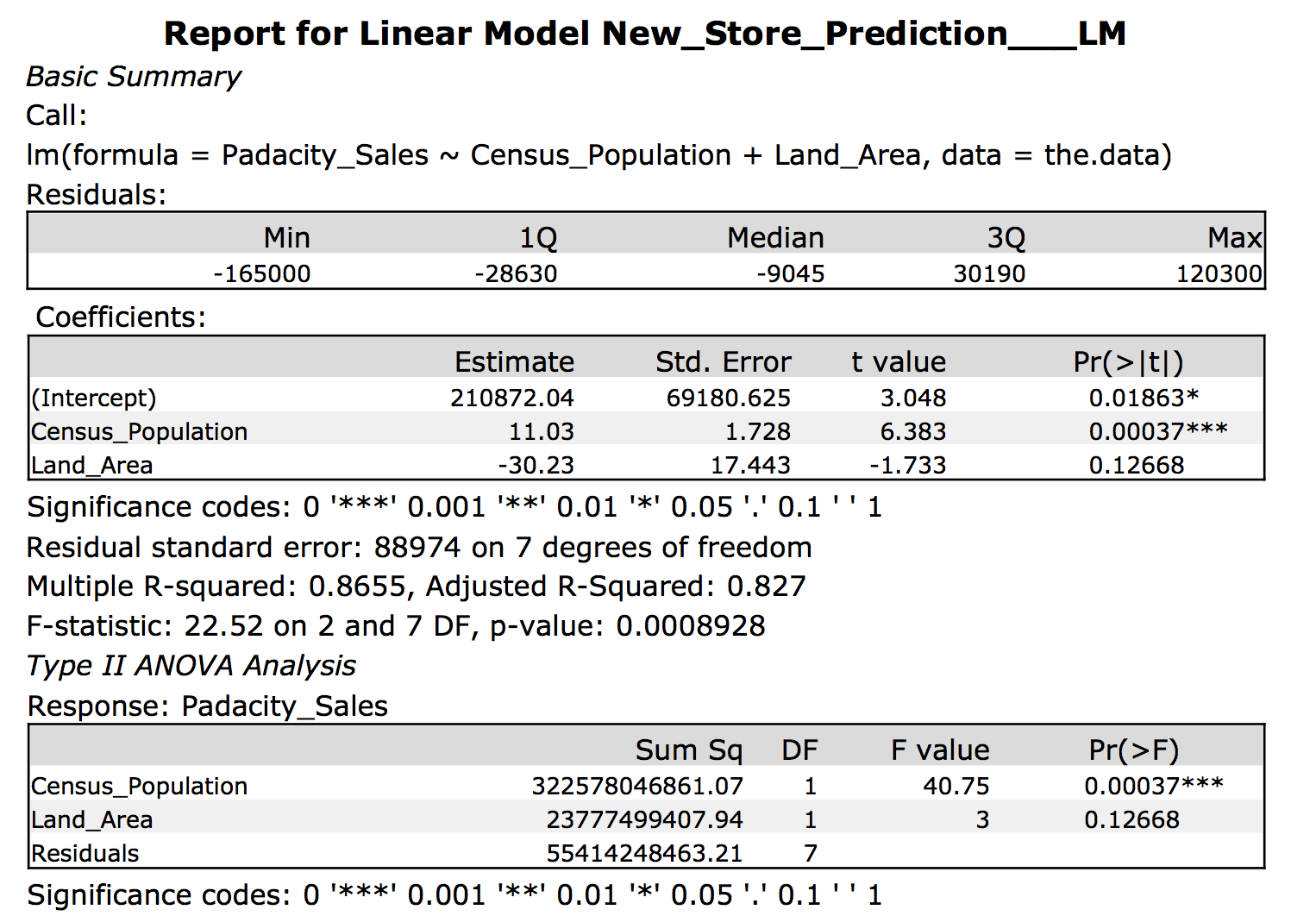


The full correlation matrix shows good correlation between predictor variables, Census\_Population, Households\_with\_Under\_18, Population\_Density and Total\_Families. There may be some multicollinearity here.

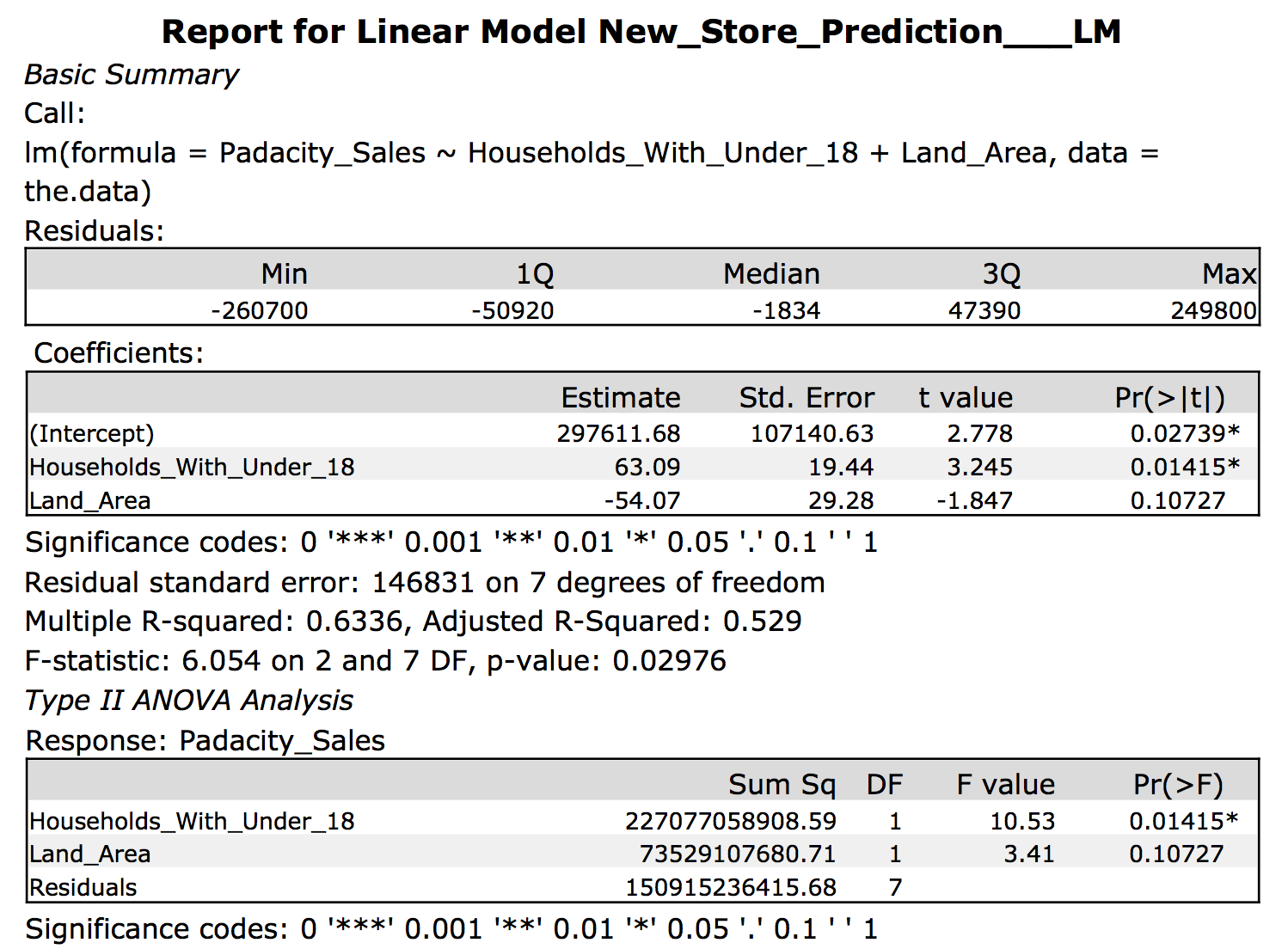
Land\_Area does not show great correlation with the other predictor variables so I will start by running a regression with Land\_Area and add other predictor variables to the regression.



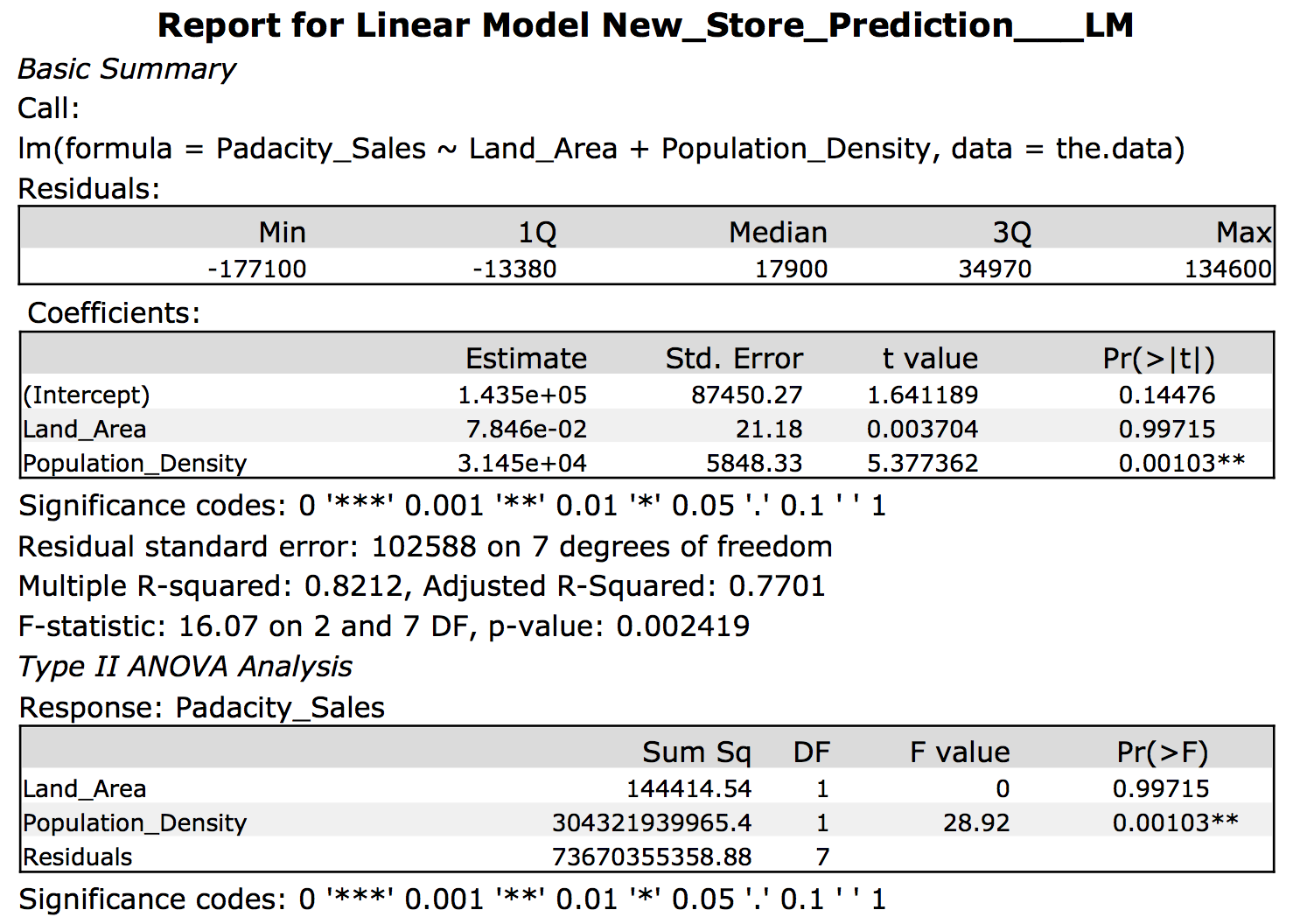
R-Squared for linear model between Sales vs Land\_Area = **0.08241**



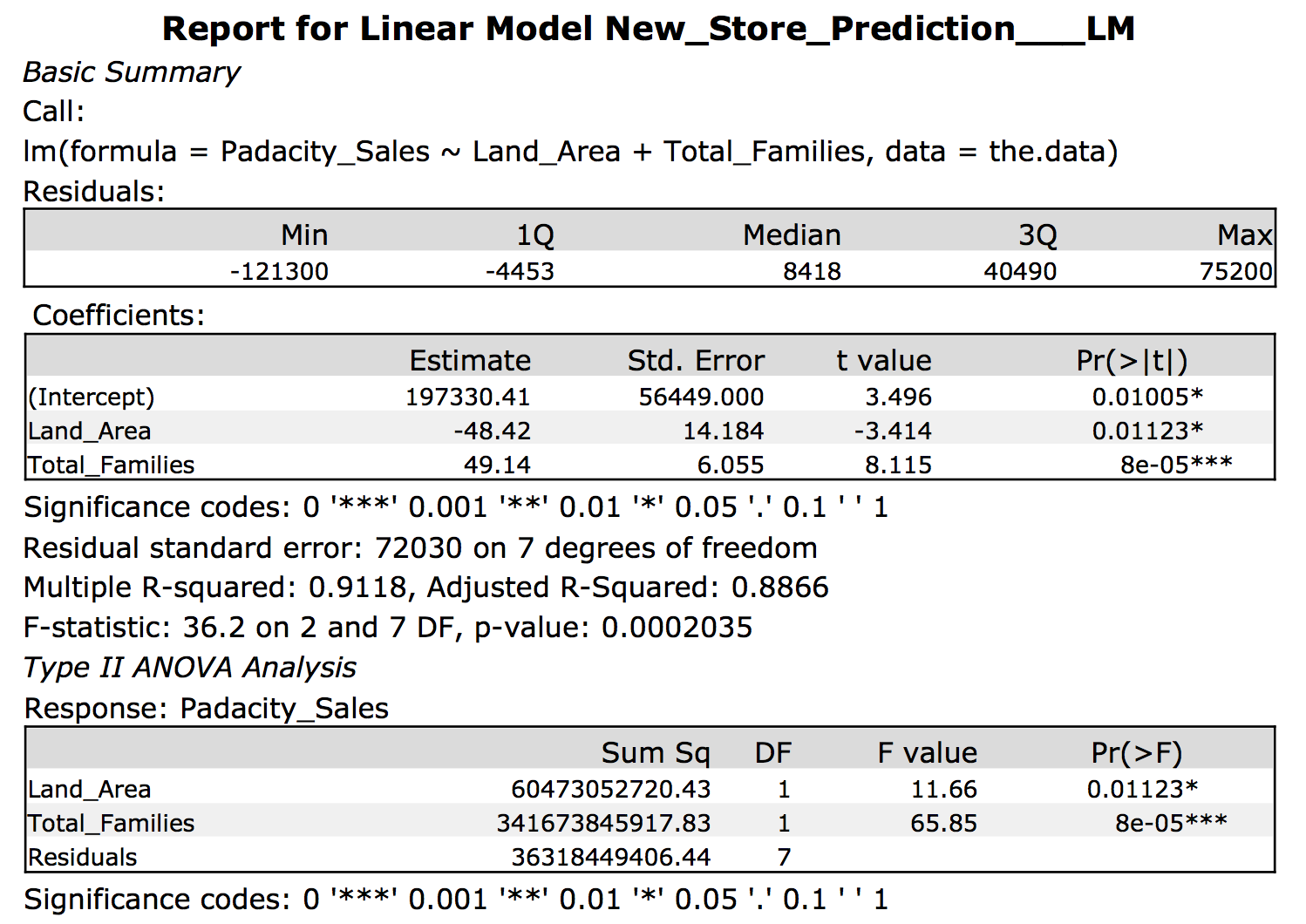
Adjusted R-Squared for linear model between Sales vs Land\_Area vs Census\_Population = **0.827**



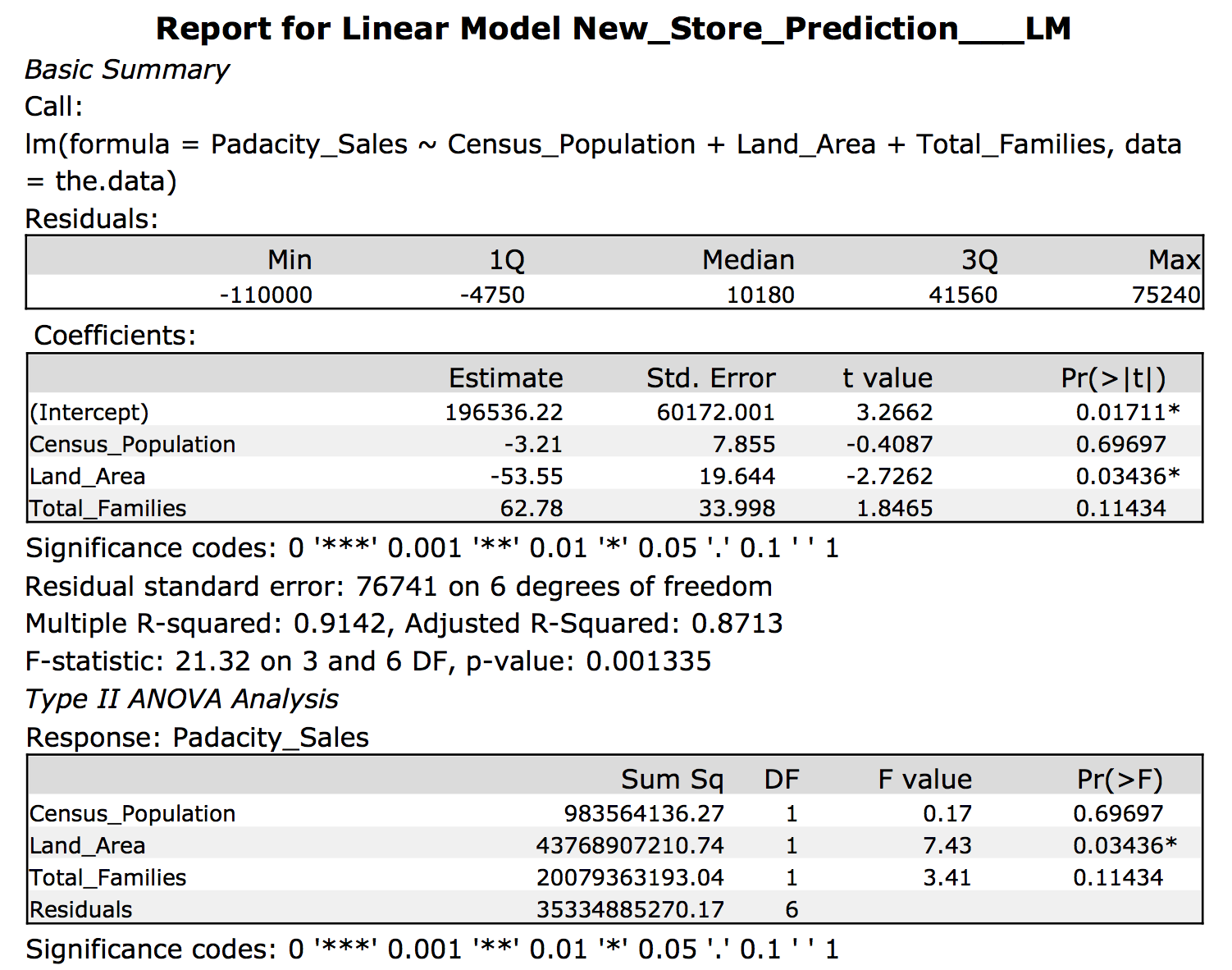
Adjusted R-Squared for linear model between Sales vs Land\_Area vs Households\_With\_Under\_18 = **0.529**



Adjusted R-Squared for linear model between Sales vs Land\_Area vs Population\_Density = **0.7701**

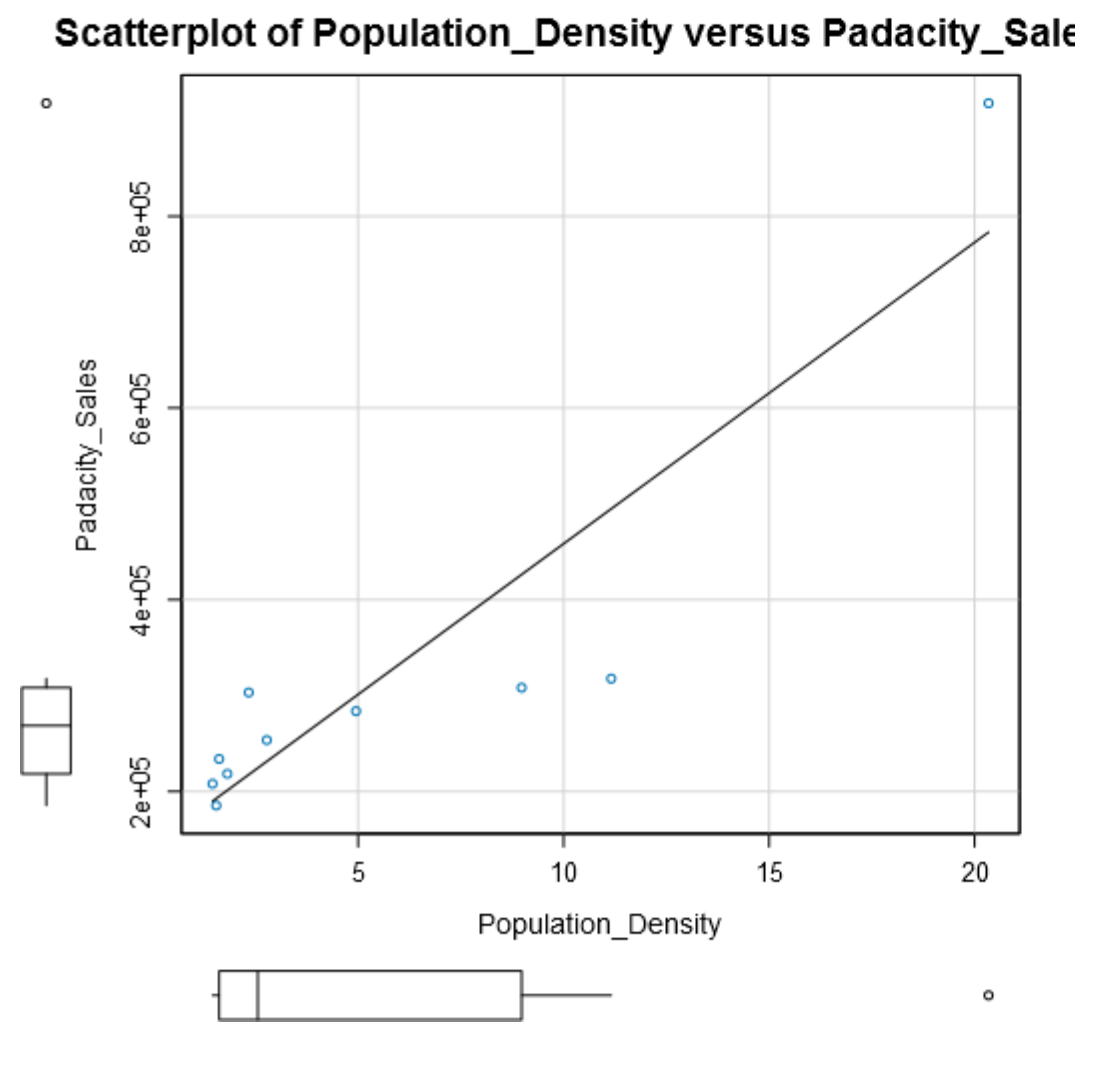
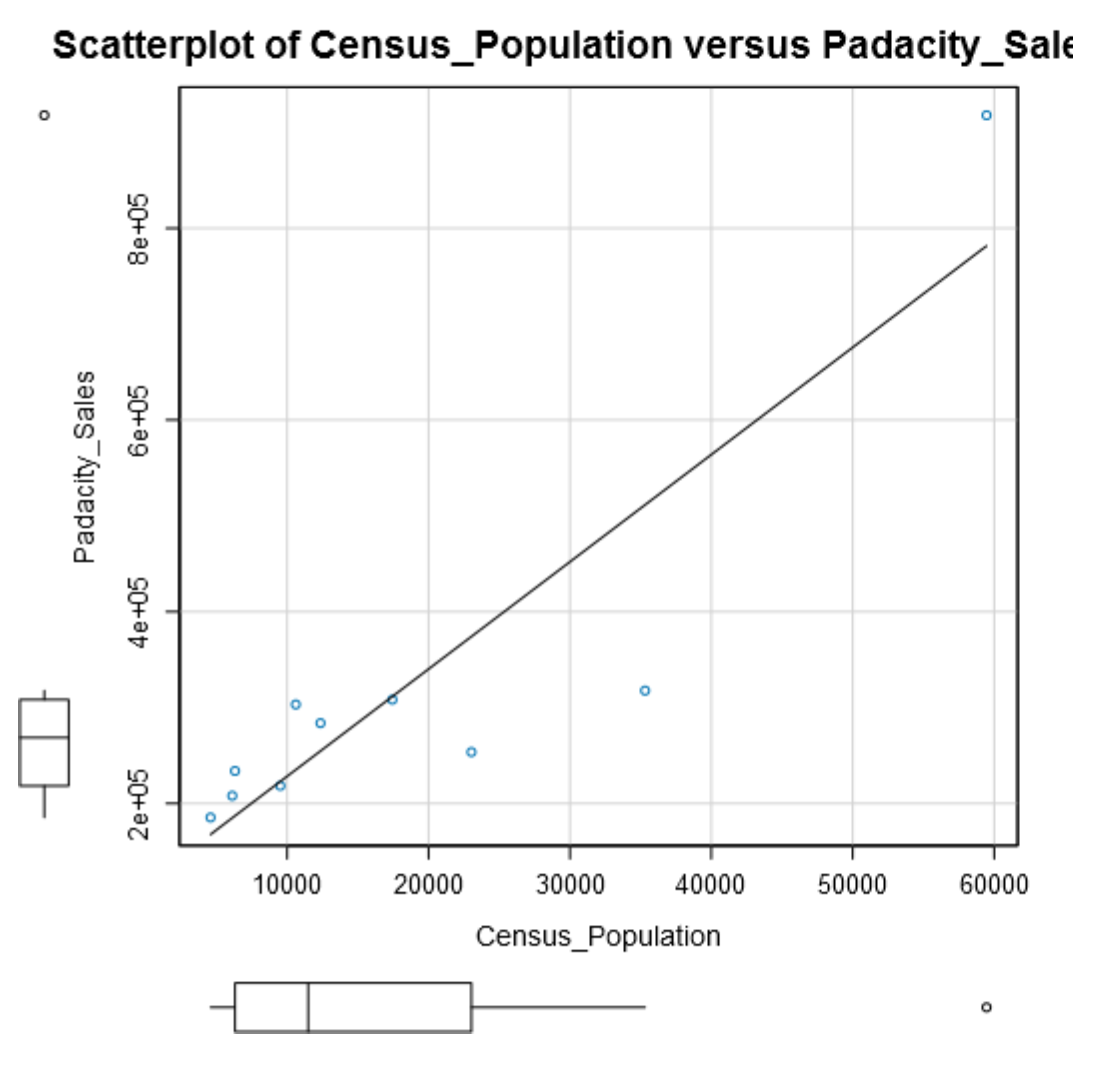


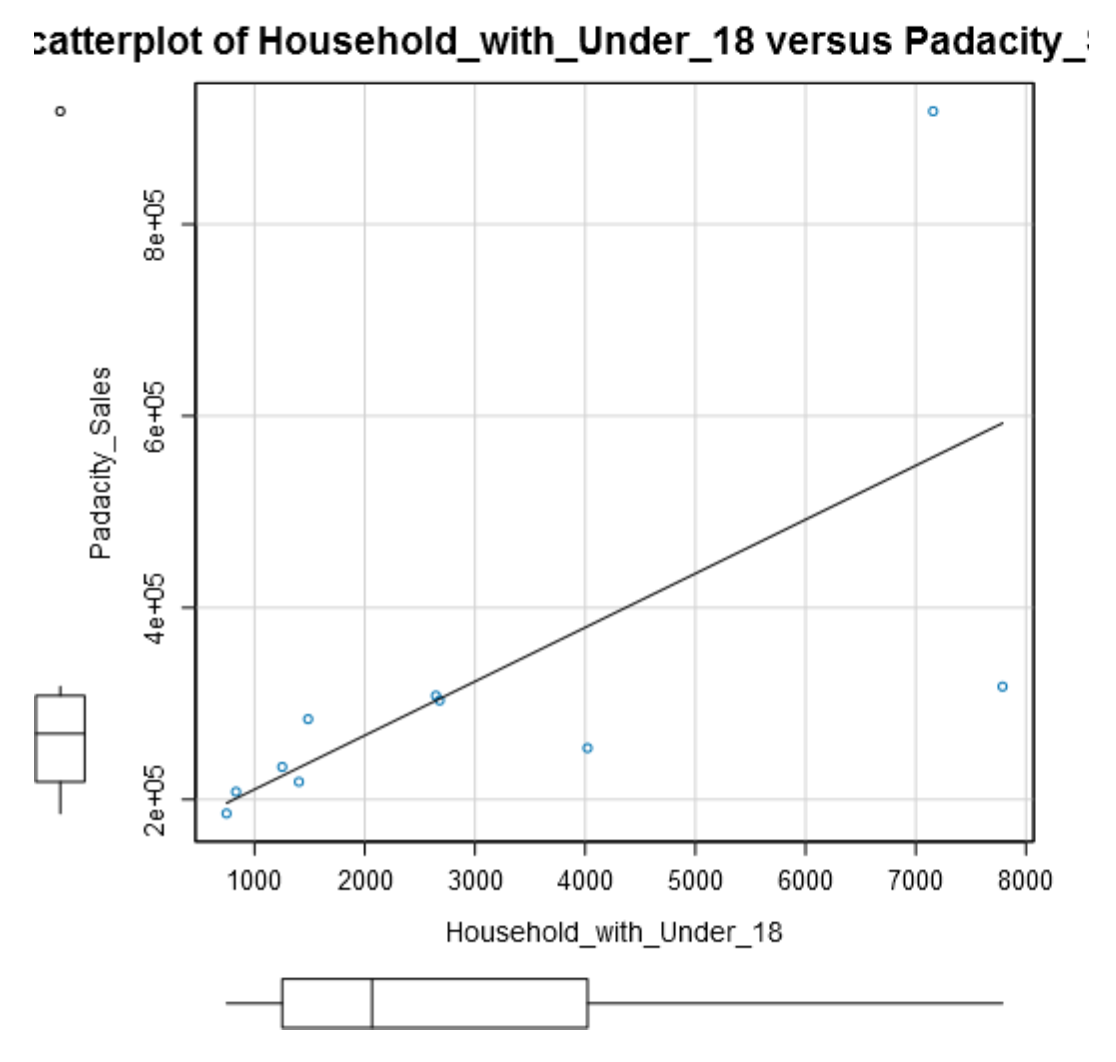
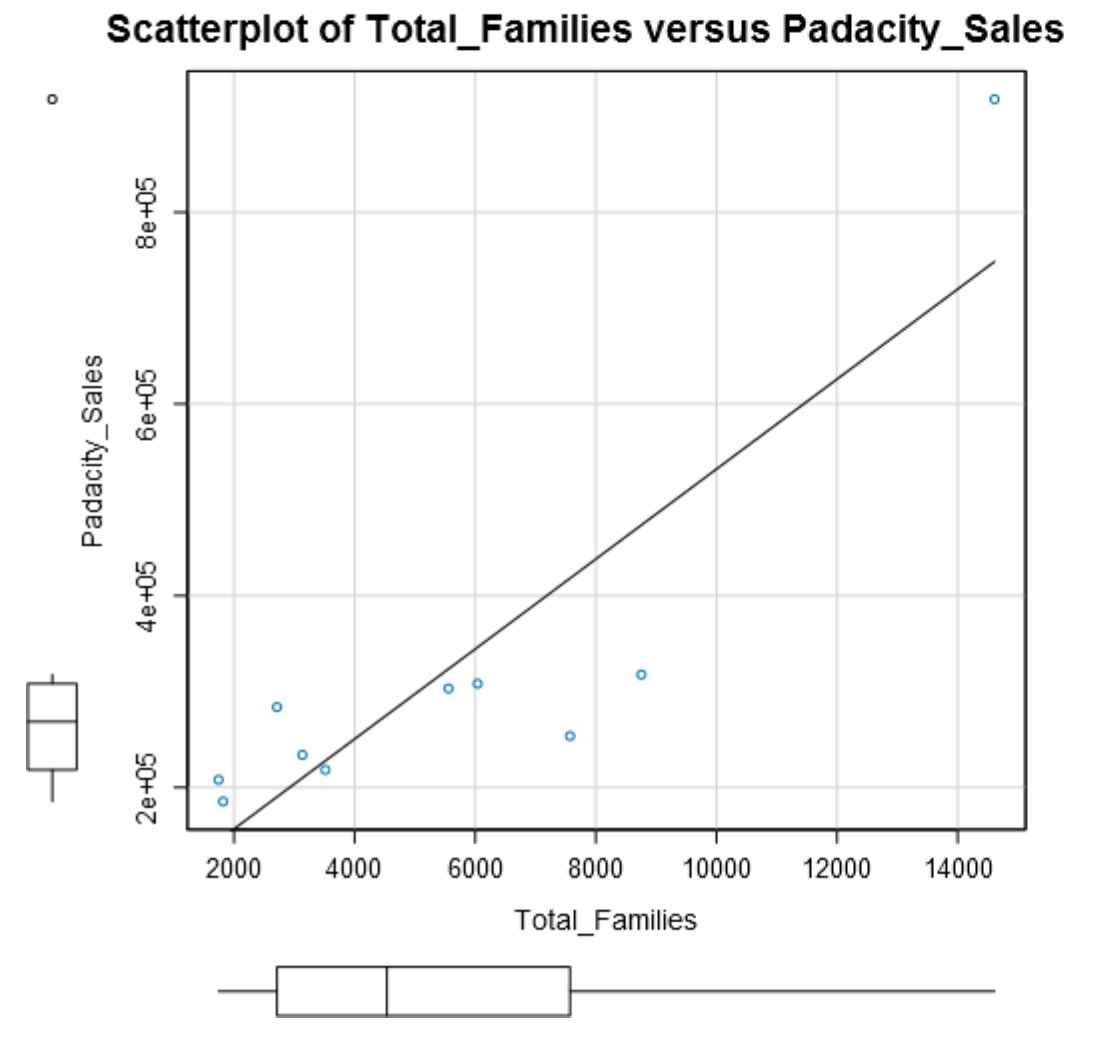
Adjusted R-Squared for linear model between Sales vs Land\_Area vs Total\_Families = **0.8866**

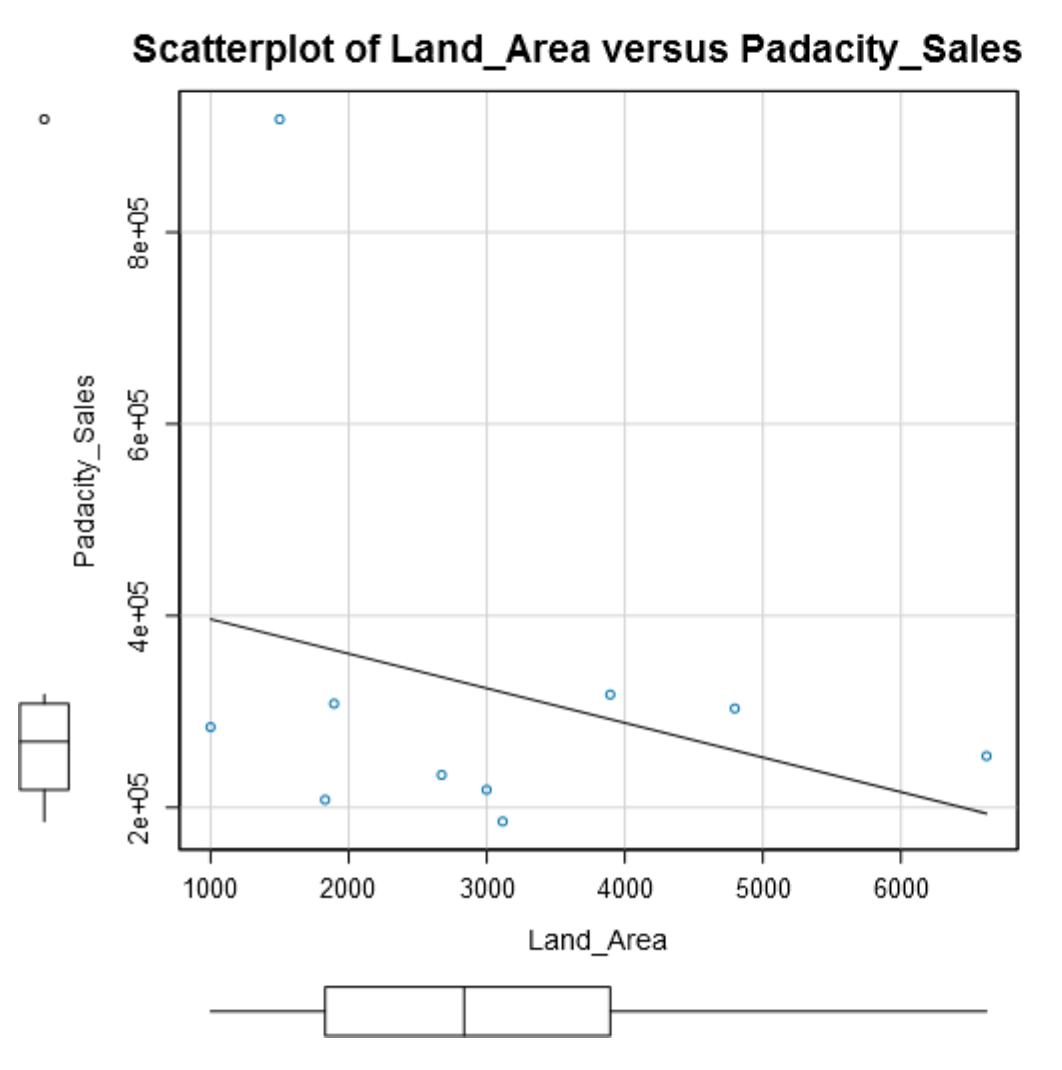


Adjusted R-Squared for linear model between Sales vs Land\_Area vs Total\_Families vs Census\_Population = **0.8713**

Below are the scatter plots of the predictor variables vs our target variable (Pawdacity\_Sales).





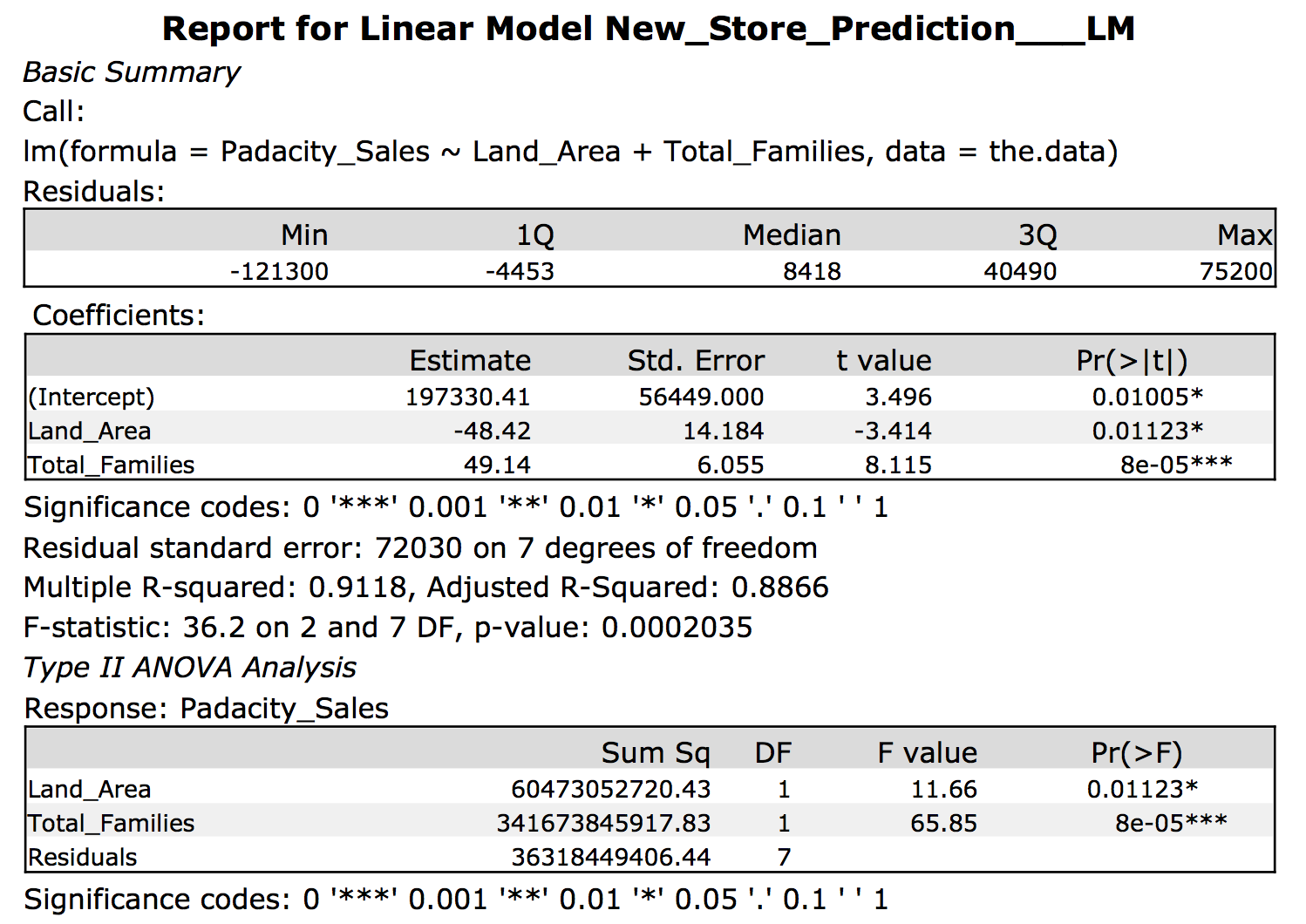


The scatter plots above give a good representation of the linearity between the target variable and its respective predictor variable.

Starting with Land\_Area as a predictor variable (R-Squared = 0.08241) and adding the other variables, I can see that the largest jump in R-Squared comes from Land\_Area and Total\_Families (adjusted r-squared = 0.8866)

I will use Land\_Area and Total\_Families as my predictor variables for my linear model.

**Below is the summary of the multilinear regression model.**



From the summary, the equation for the linear regression model is:

*Y (Pawdacity\_Sales) = 197330.41 – 48.42(Land\_Area) + 49.14(Total\_Families)*

**Final recommendation.**

Here are the criteria’s given to you in choosing the right city:

1. The new store should be located in a new city. That means there should be no existing stores in the new city.
2. The total sales for the entire competition in the new city should be less than $500,000
3. The new city where you want to build your new store must have a population over 4,000 people (based upon the 2014 US Census estimate).
4. The predicted yearly sales must be over $200,000.
5. The city chosen has the highest predicted sales from the predicted set.

With the required criteria, I would recommend Laramie City. Laramie City does not currently contain a store, has an estimated census population for 2014 of 32,081 and predicted sales of **$305,013.88**.

Below is a summary of the final possibilities for a new store with the highlighted row as the recommendation.

|  |  |  |  |
| --- | --- | --- | --- |
| **City** | **2014\_Census\_Pop\_Est** | **Total\_Families** | **Score** |
| **Laramie** | 32081.00 | 4668.93 | 305013.88 |
| **Torrington** | 6736.00 | 2548.50 | 245081.79 |
| **Jackson** | 10449.00 | 2313.08 | 225870.82 |
| **Lander** | 7642.00 | 3876.81 | 225751.40 |
| **Green River** | 12630.00 | 3977.40 | 224372.00 |
| **Worland** | 5366.00 | 1364.32 | 201700.33 |