Projeto de prática: Recomende uma cidade

**Observe que este projeto é uma continuação do projeto de limpeza de**  **dados.**

## Passo 1: Regressão Linear

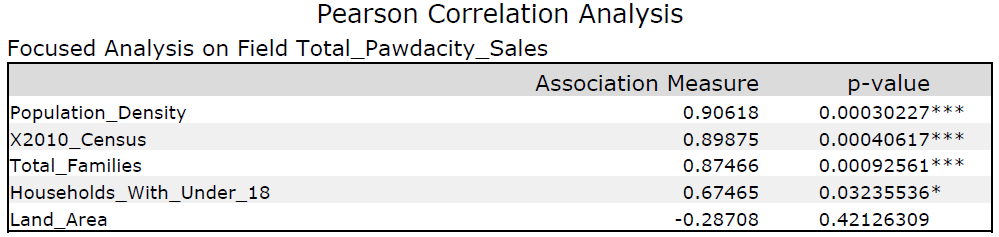
**Criando o Modelo**

Abaixo está o conjunto de dados final usado para modelo de regressão.

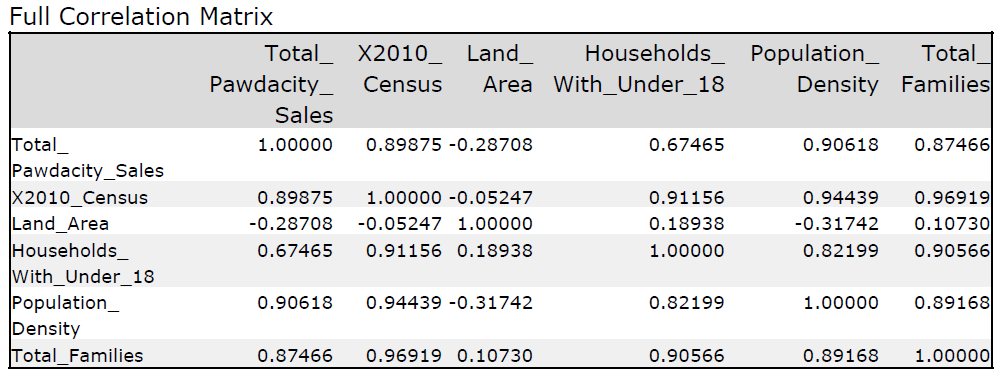
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **City** | **Census\_Population** | **Household\_With\_Under\_18** | **Land\_Area** | **Padacity\_Sales** | **Population\_Density** | **Total\_Families** |
| **Buffalo** | 4585 | 746 | 3115.5075 | 185328 | 1.55 | 1819.5 |
| **Casper** | 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 |

**Selecionando as vereáveis preditoras.**

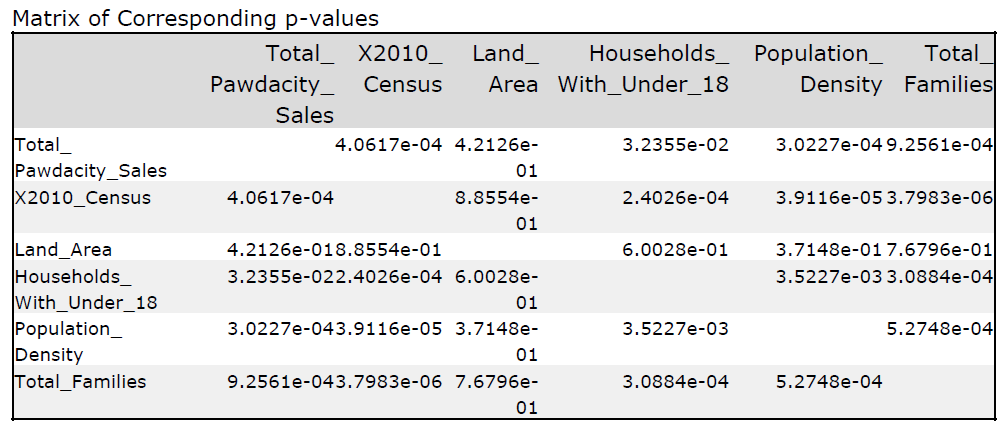
Abaixo está a tabela de correlação Pearson de todas as variáveis.



Matrix completa de correlação.



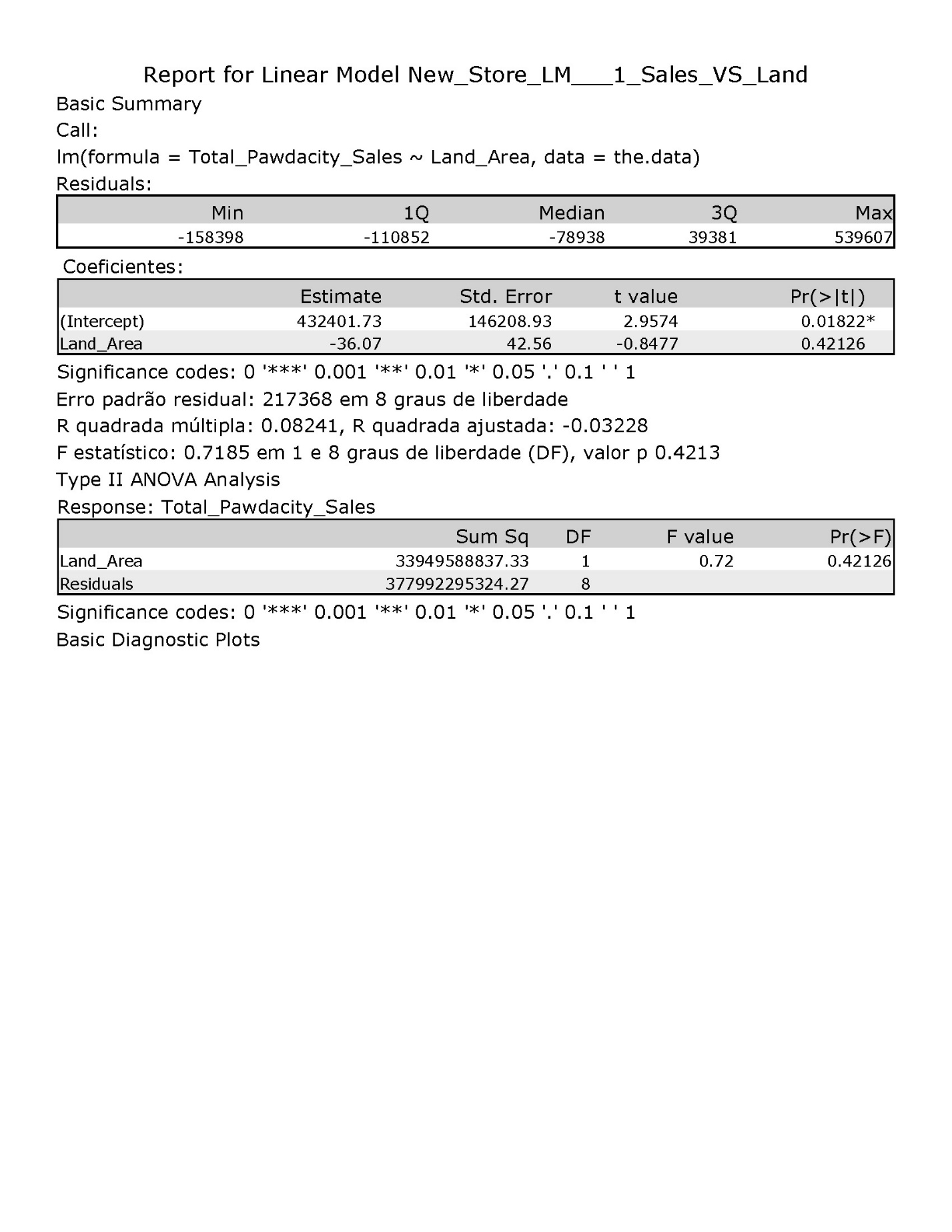
Matrix do p-valores para as variáveis preditoras.



A matriz de correlação mostra boa correlação entre as variáveis preditivas, 2010\_Census, Census\_Population, Households\_with\_Under\_18, Population\_Density e Total\_Families. Podendo haver alguma multicolinearidade.

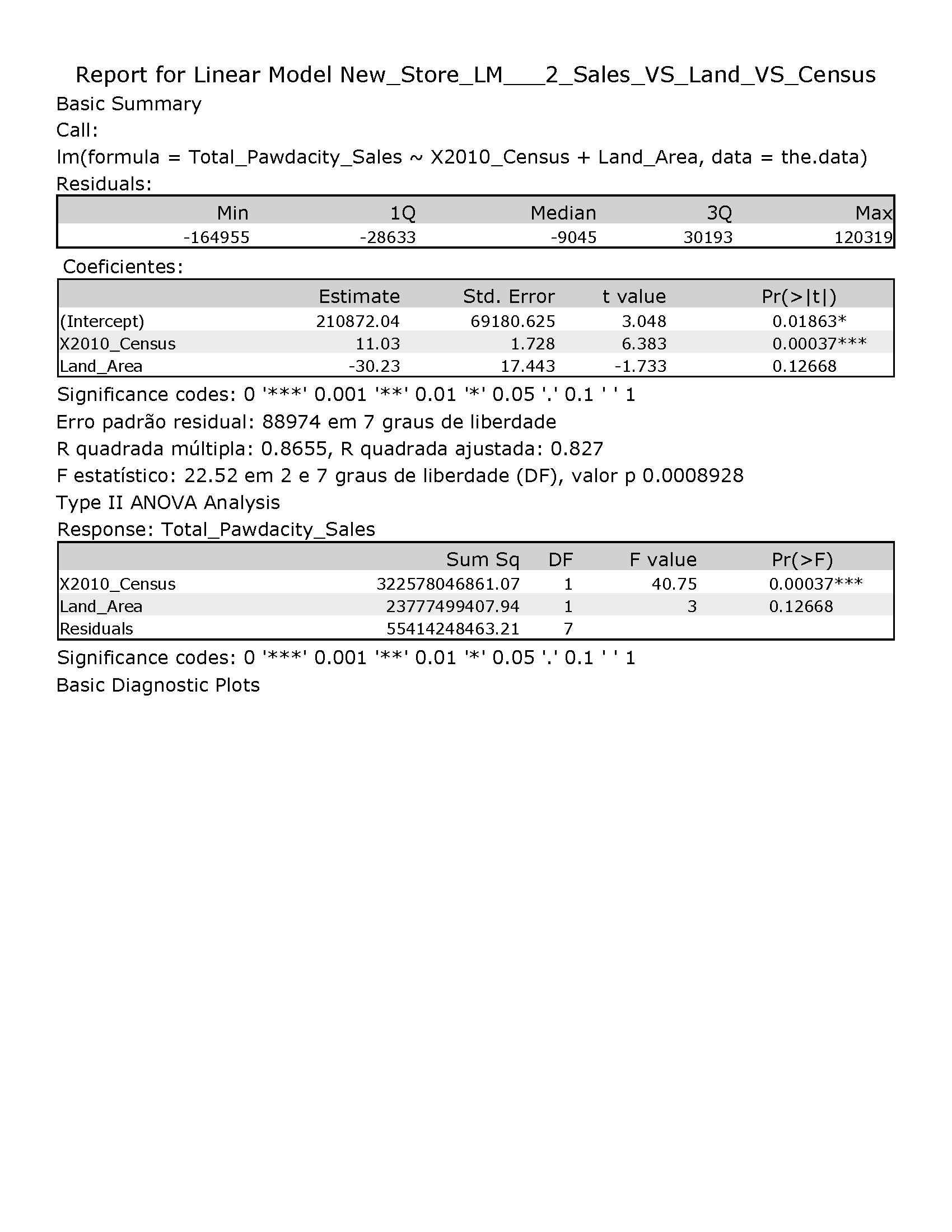
A Land\_Area não mostra grande correlação com as outras variáveis preditivas, portanto, começarei executando uma regressão linear com Land\_Area e ir adicionando as outras variáveis preditoras à regressão.

1. New\_Store\_LM - 1\_Sales\_VS\_Land



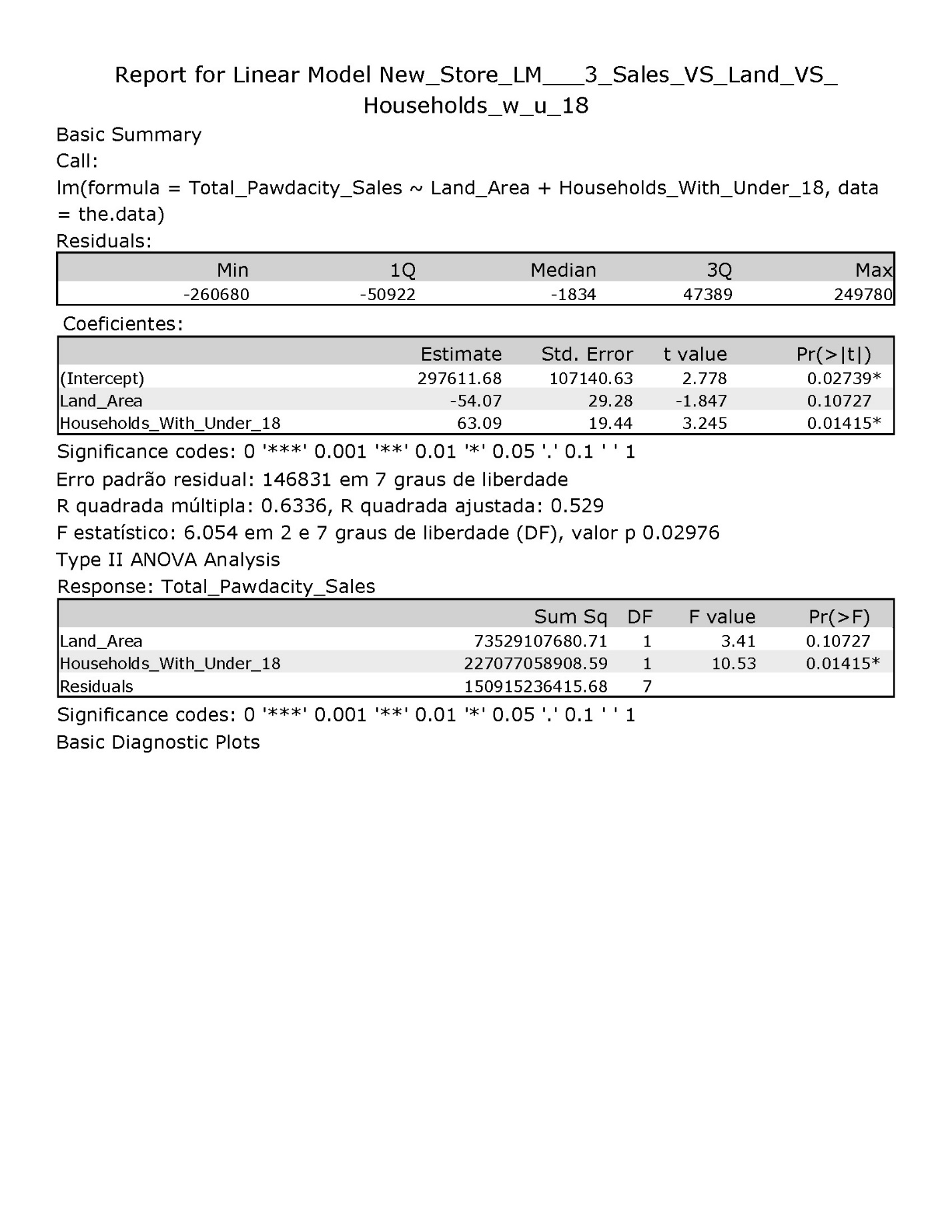
R quadrada do modelo linear para Total\_Pawdacity\_Sales X Land\_Area = **0.08241**

1. New\_Store\_LM - 2\_Sales\_VS\_Land\_VS\_Census



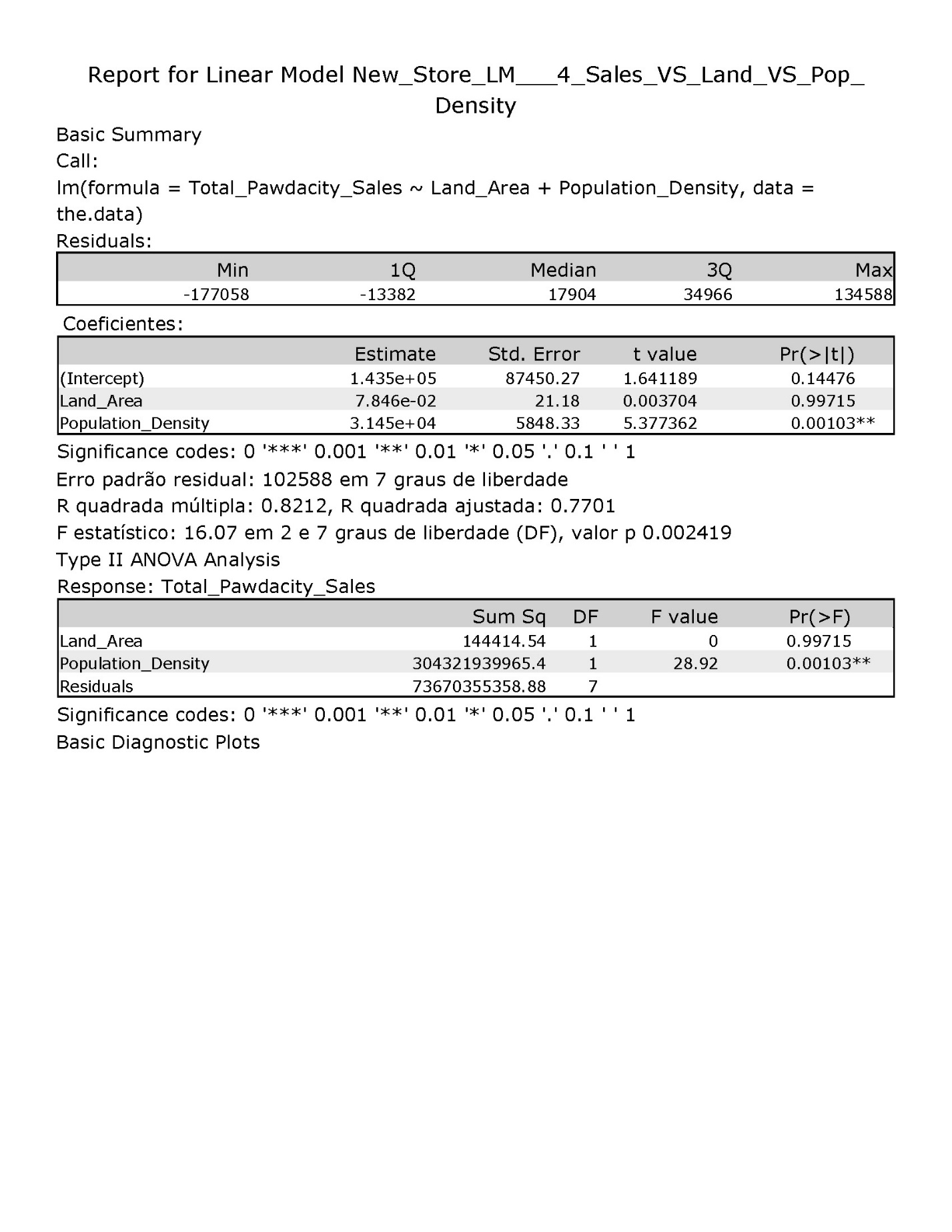
R quadrada ajustada do modelo linear para Total\_Pawdacity\_Sales X Land\_Area X 2010\_Census = **0.827**

1. New\_Store\_LM - 3\_Sales\_VS\_Land\_VS\_Households\_w/u\_18



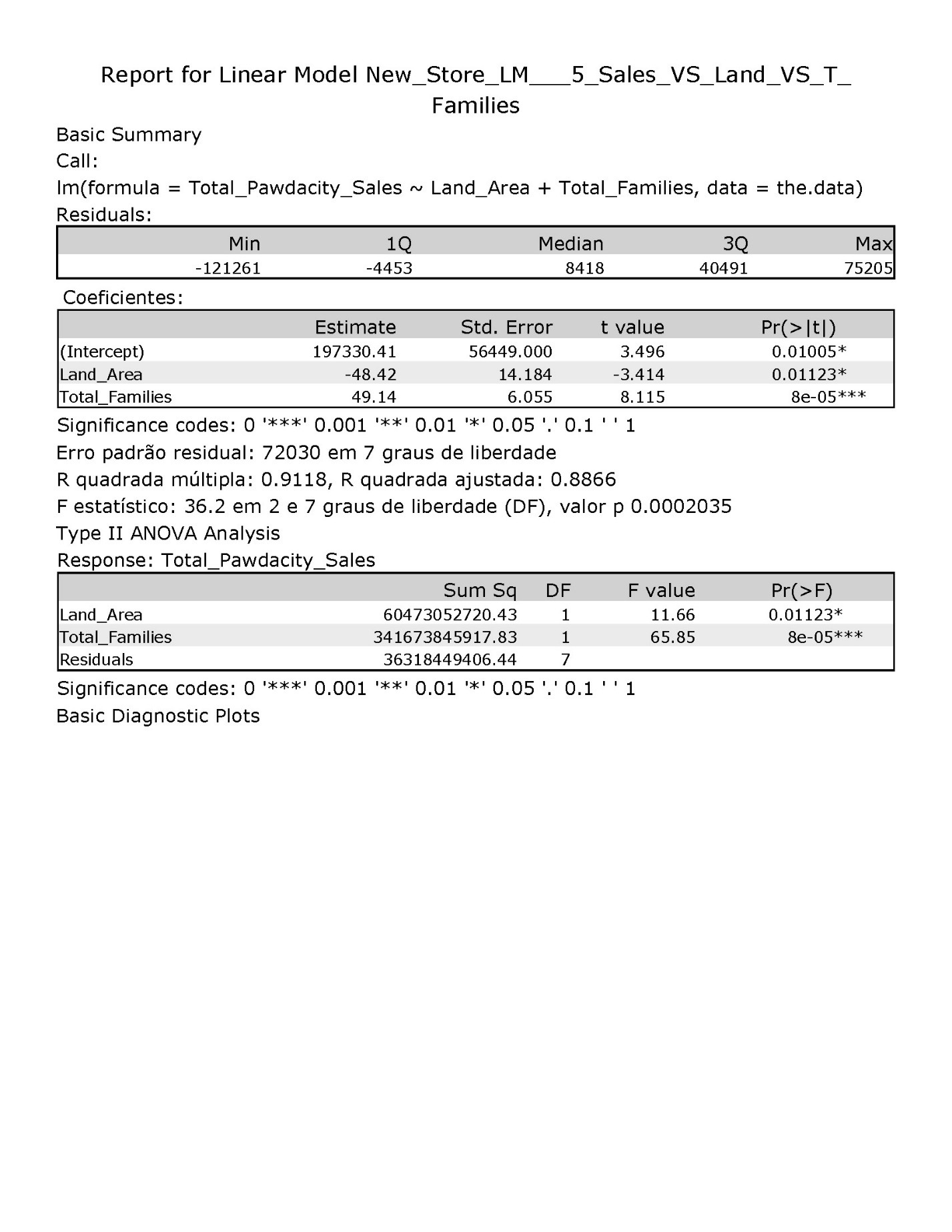
R quadrada ajustada do modelo linear para Total\_Pawdacity\_Sales X Land\_Area X Housesholdes\_W/u\_18 = **0.529**

1. New\_Store\_LM - 4\_Sales\_VS\_Land\_VS\_Pop\_Density



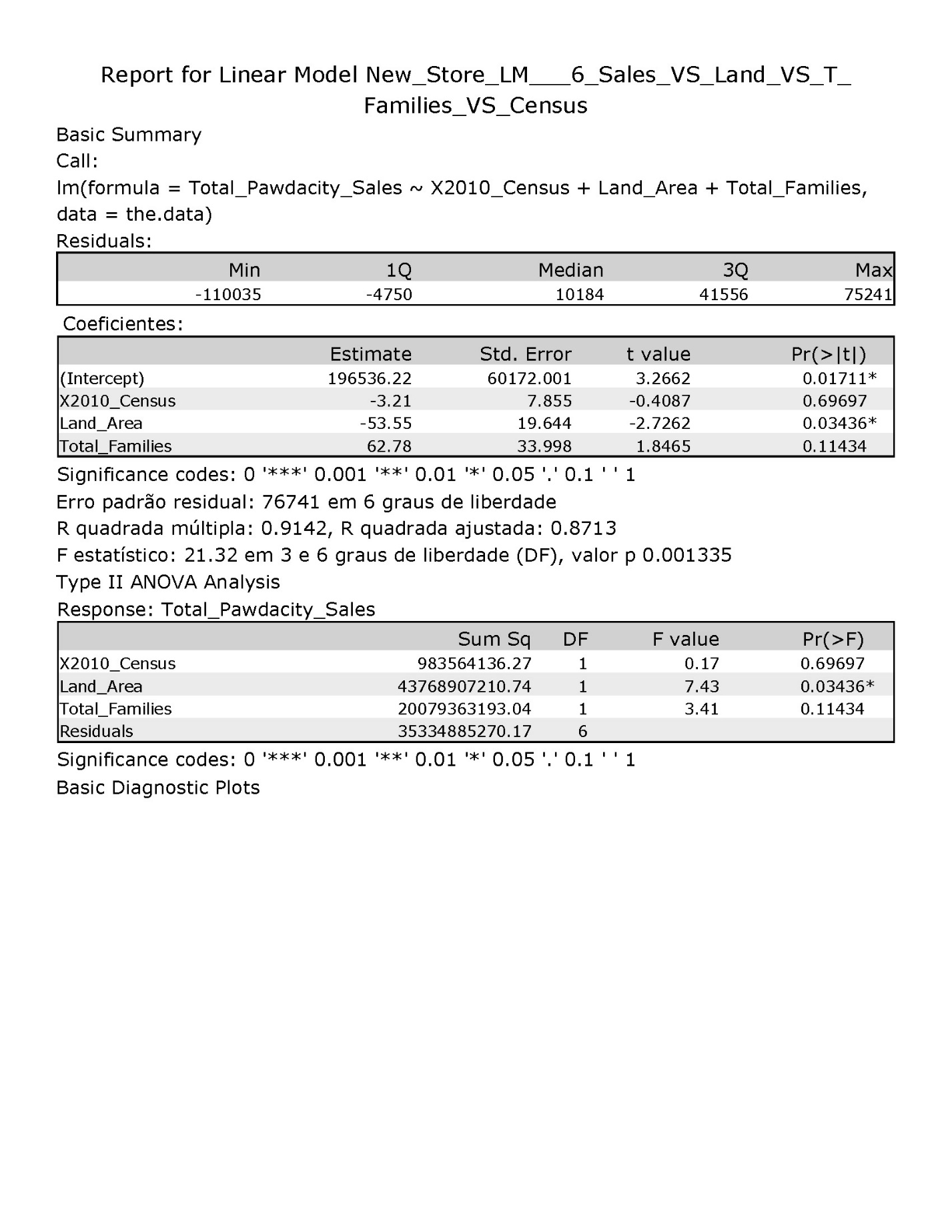
R quadrada ajustado do modelo linear para Total\_Pawdacity\_Sales X Land\_Area X Population\_Density = **0.7701**

1. New\_Store\_LM - 5\_Sales\_VS\_Land\_VS\_T\_Families



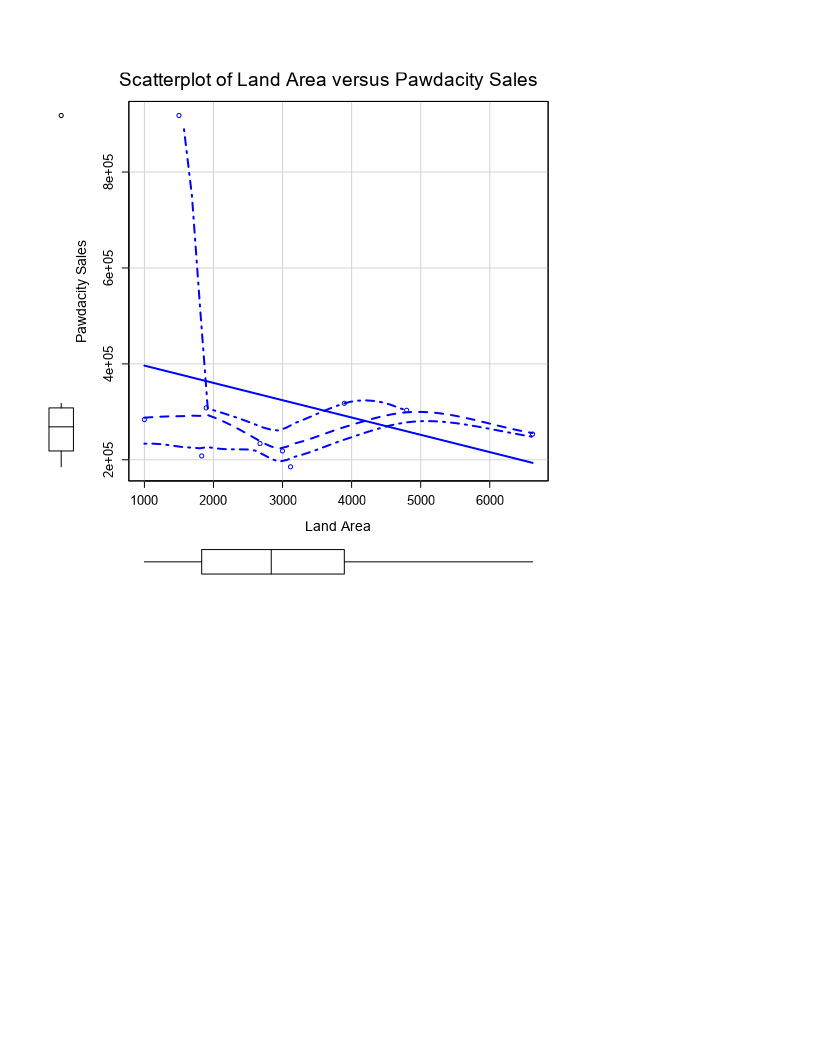
R quadrada ajustado do modelo linear para Total\_Pawdacity\_Sales X Land\_Area X Total\_Families = **0.8866**

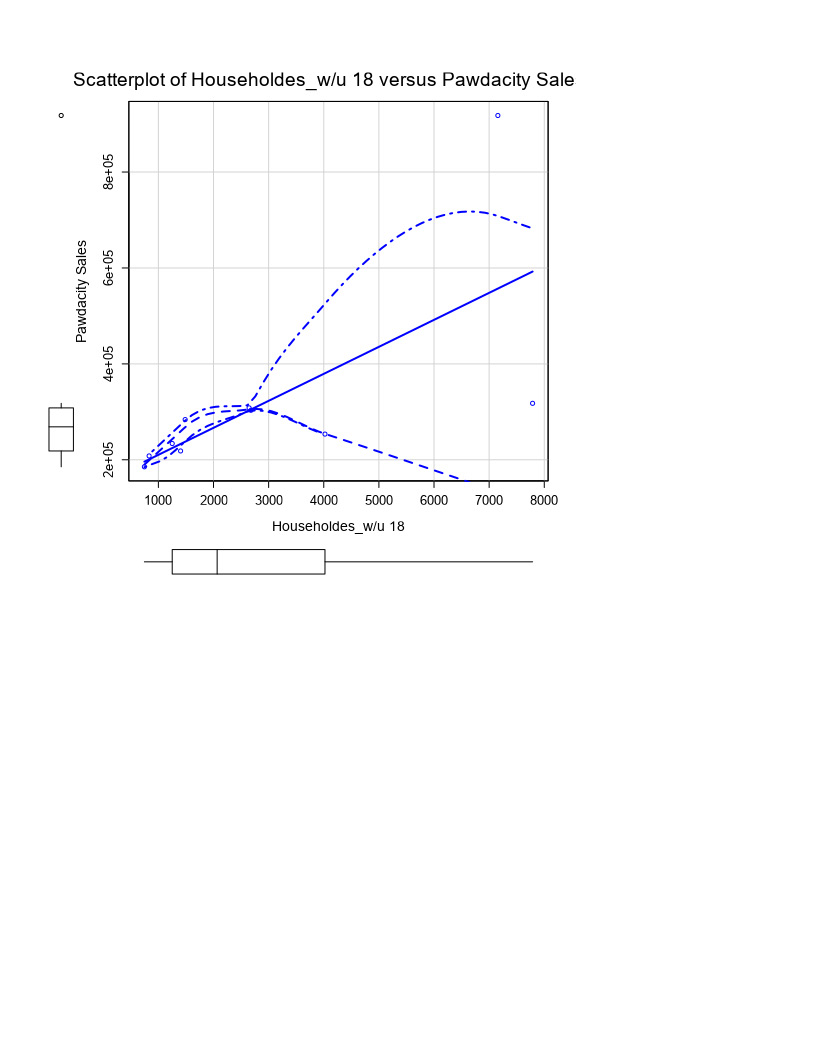
1. New\_Store\_LM - 6\_Sales\_VS\_Land\_VS\_T\_Families\_VS\_Census.

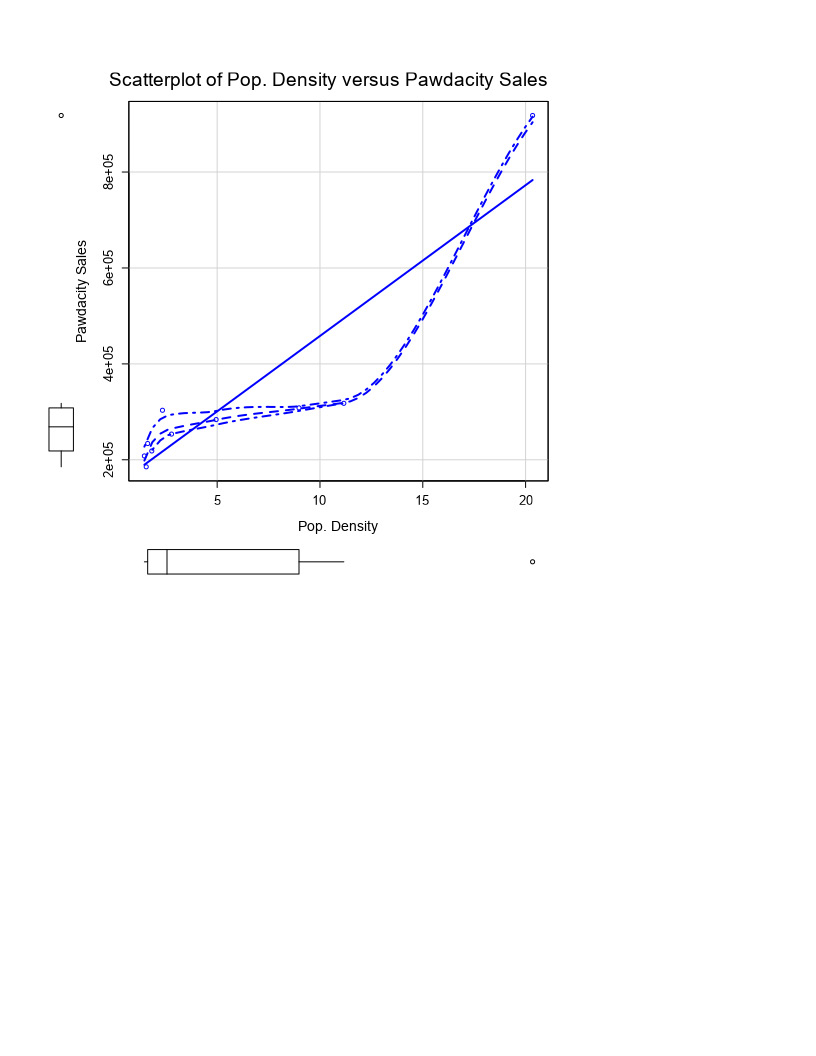


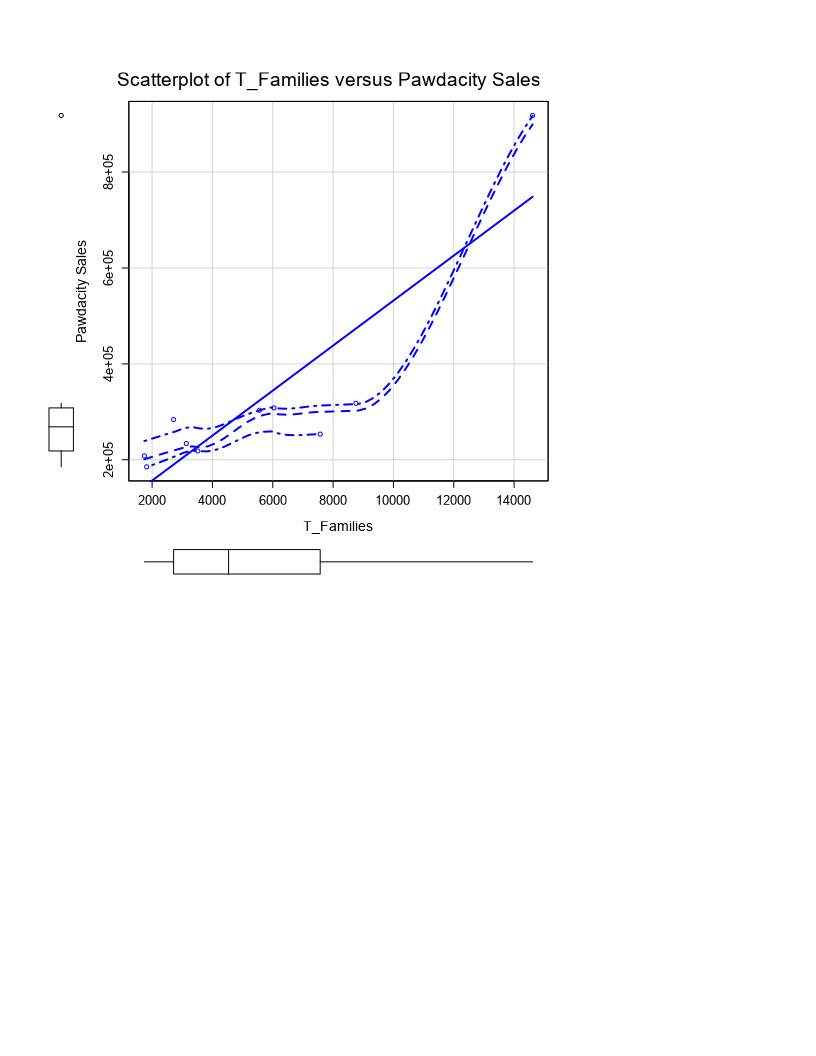
R quadrada ajustado do modelo linear para Total\_Pawdacity\_Sales X Land\_Area X Total\_Families X 2010 Census = **0.8713**

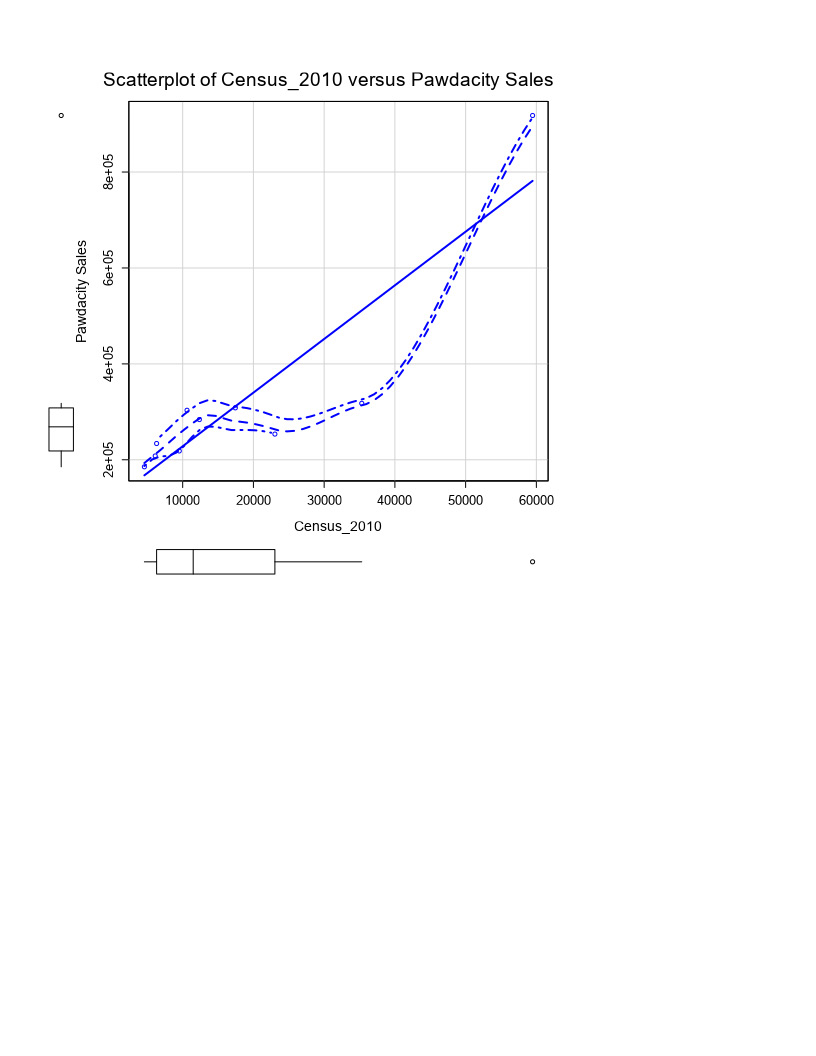
Abaixo estão os Scatterplot para todas as variáveis preditoras X variável alvo (Total\_Pawdacity\_Sales).









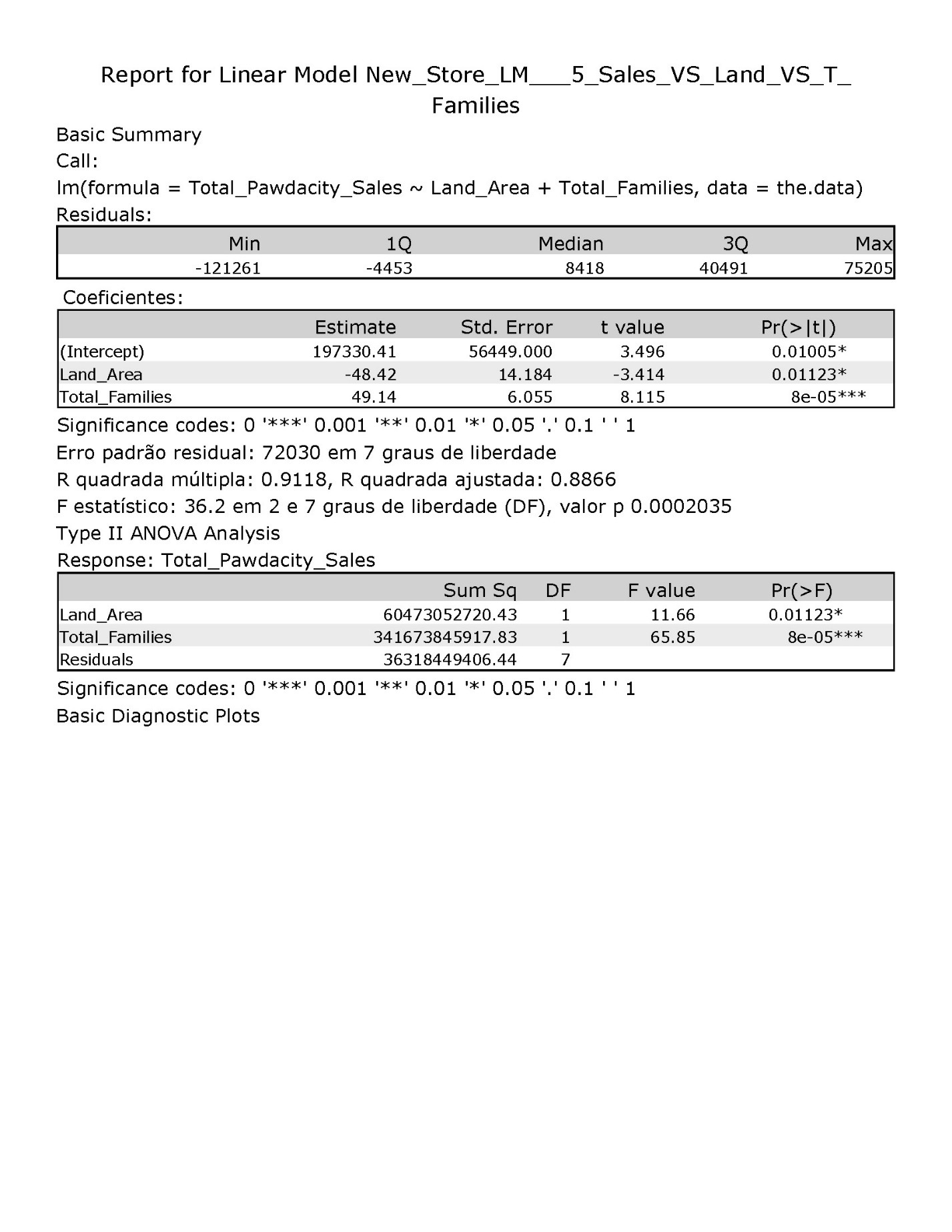


Os gráficos de dispersão acima fornecem uma boa representação da linearidade entre a variável alvo (Total\_Pawdacity\_sales) e sua respectiva variável preditora.

Começando com Land\_Area como uma variável preditora (R-quadrado = 0,08241) e adicionando outras variáveis, possível ver uma maior diferença no R-quadrado quando estão sendo usadas as variáveis Land\_Area e Total\_Families (R-quadrado ajustado = 0,8866).

Usarei Land\_Area e Total\_Families como minhas variáveis preditoras para o meu modelo linear.

Abaixo o resumo do modelo de regressão multilinear.



A equação do modelo de regressão linear é:

**Y (Total\_Pawdacity\_Sales) = 1973330.41 – 48.42 (Land\_area) + 49.12 (Total\_Families).**

## Análise – Recomendação.

Esses são os critérios para a escolha da nova cidade:

1. A nova loja deve estar localizada em uma nova cidade. Isso significa que não deve haver lojas existentes na nova cidade.

2. O total de vendas para toda a competição na nova cidade deve ser inferior a US $ 500.000

3. A nova cidade onde você deseja construir sua nova loja deve ter uma população superior a 4.000 pessoas (com base na estimativa do Censo dos EUA em 2014).

4. As vendas anuais previstas devem ser superiores a US $ 200.000.

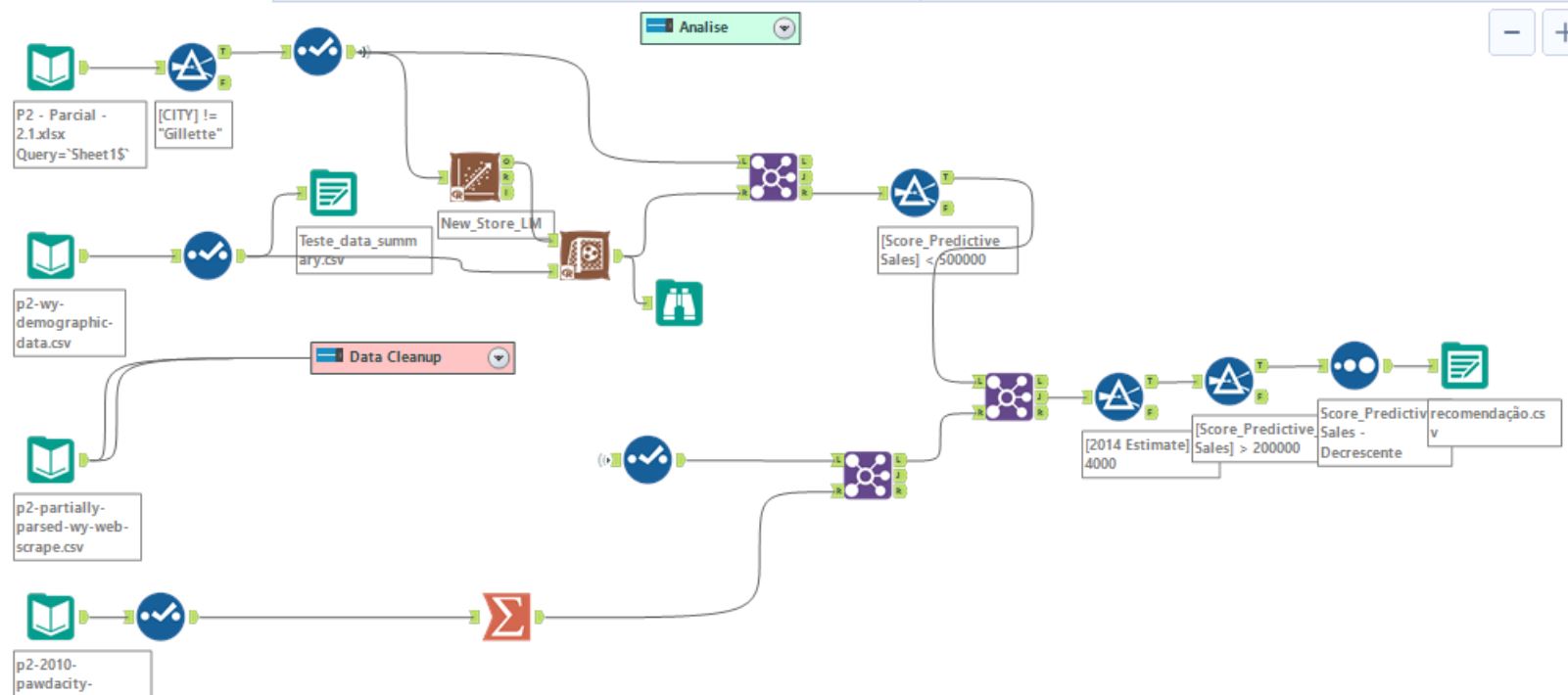
5. A cidade escolhida tem as vendas previstas mais altas do conjunto previsto.

Com os critérios acima, recomendo a Laramie City, que atualmente não contém uma loja e tem uma população estimada em 2014 de 32.081 habitantes e previsão de vendas é de US $ 305.013,88.

Abaixo esta um resumo das 6 possíveis cidades para abertura da 14° loja, em destaque está minha recomendação.

|  |  |  |  |
| --- | --- | --- | --- |
| **City** | **2014\_Census** | **Total\_Families** | **Score\_Pawdacity\_sales** |
| **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 |

**Alterxy Workflow**

****