一．直观看出，简单线性关系LRZE= a\*YYSR+b\*YWFY +c\*YYCB +d\*YYSJJFJ +e\*ZCJZSS

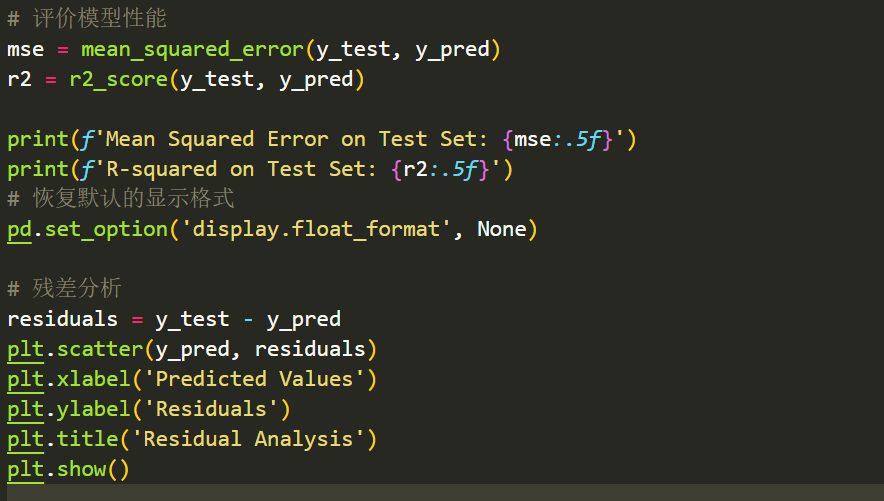
建立线性回归模型，使用线性回归模型的 **coef\_** 属性来获取系数：



结果为Coefficients (a, b, c, d, e): [ 1.70732025e+10 -1.42545409e+09 -1.43747502e+10 -1.95050457e+08

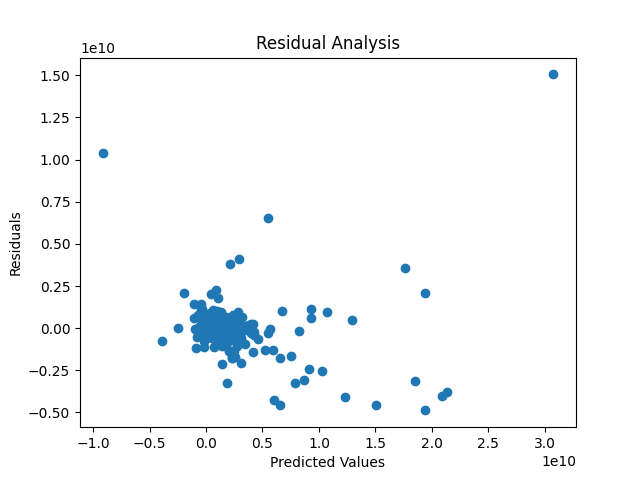
-2.61656664e+08]

二．下面，我们对模型优劣进行评价：



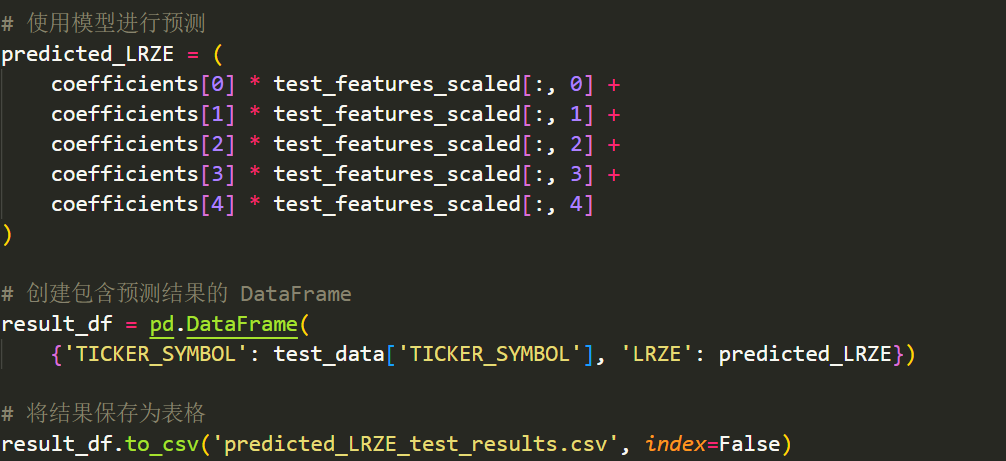
R-squared on Test Set: 0.87390接近1

**R-squared (R²):** R-squared 是一个统计度量，用于衡量模型对因变量变异性的解释程度。R-squared 的取值范围在 0 到 1 之间，越接近 1 表示模型对观测数据的解释越好。在你的情况下，R-squared 值相对较高 (0.87)，表明模型能够解释目标变量的大部分变异



残差分析：残差随着预测值的增加而有规律地变化，呈现锥形，模型拟合较好

三．结果预测：



|  |  |
| --- | --- |
| TICKER\_SYMBOL | LRZE |
| 4953174 | 46313905.43 |
| 4961537 | 4881055.29 |
| 4962538 | -232814286.2 |
| 4968740 | -223146926.8 |
| 4973917 | -302681906.6 |
| 4978589 | 11220602.49 |
| 4978721 | -262503964 |
| 4986535 | -259099985.1 |
| 4990739 | -104342053.9 |
| 4990942 | -301998758.7 |
|  |  |

四：模型改进：1.可能出现过度拟合改进方法：



2. 