# Statsmodels OLS Salary Prediction Analysis

Results Analysis (generated 2025-08-21 11:15:31.509616)

Dataset: salary\_data.csv

1. Parameter Trends:  
 - Slopes range: 9162.7155 to 11263.3333  
 - Intercepts range: 21048.67 to 28344.34

2. Performance Metrics:  
 - Mean Train RSS (min/max): 22786589.02 / 56420523.56  
 - Mean Test RSS (min/max): 36156721.84 / 83506466.27  
 - Train R2 (min/max): 0.771 / 0.968  
 - Test R2 (min/max): 0.893 / 0.953

3. Correlations Across Splits:  
 - Training % vs slope correlation: -0.6666  
 - Slope vs Pearson(r) correlation: -0.9814

4. Interpretation:  
 - Slopes stabilize as more training data is used (variance in slope decreases).  
 - Train RSS generally decreases with more data; test RSS may plateau or slightly increase if overfitting at low data.  
 - R2 improves then levels off, indicating sufficient data coverage.  
 - Pearson correlation derived from slope remains consistent, reflecting stable linear relationship.

5. Recommendations:  
 - Use >= 60% training split for stable parameter estimation.  
 - Examine residual plots (not produced here) for heteroscedasticity.