

# Analysis of Salary Data

Data Scientist: Greg Hein



# Agenda

## **Findings of linear regression modeling with tech salary data**

- Data Description
- Regression Results
- Interpretation of Model Results

# Data Description

There are a total of 375 employees in the original dataset.

For each employee, the original dataset has entries for: Age, Gender, Education Level, Job Title, Years of Experience, and Salary.

To make use of Job Title column, it was analyzed to see whether certain words were contained in the Job Title

List of Job Title words used: Director, Junior, Senior, Manager, Analyst, Engineer

# Data Description

After inspection, it was found 2 of the employees had null data so those two employees were dropped from the dataset leaving 373 employees.

The Salaries in the dataset range from \$350 to \$250,000

# Data Description

## Relationship between employee characteristics and Salary

### Average Salaries for each Gender

Female 97011.173184

Male 103867.783505

---

### Average Salaries for each Education Level

Bachelor's 74756.026786

Master's 129795.918367

PhD 157843.137255

The average values for Salaries for Gender and Education Level here show differences that suggest an influence of both on Salary

# Data Description

## Relationships between employee characteristics and Salary (cont'd)

Years of Experience plotted  
Against Salary suggests  
A linear relationship



# Regression Results

After an initial regression run it was determined that the Job Title fields in the dataset associated with the words: junior, analyst, and engineer were not statistically significant.

Those fields were removed from the final linear regression model.

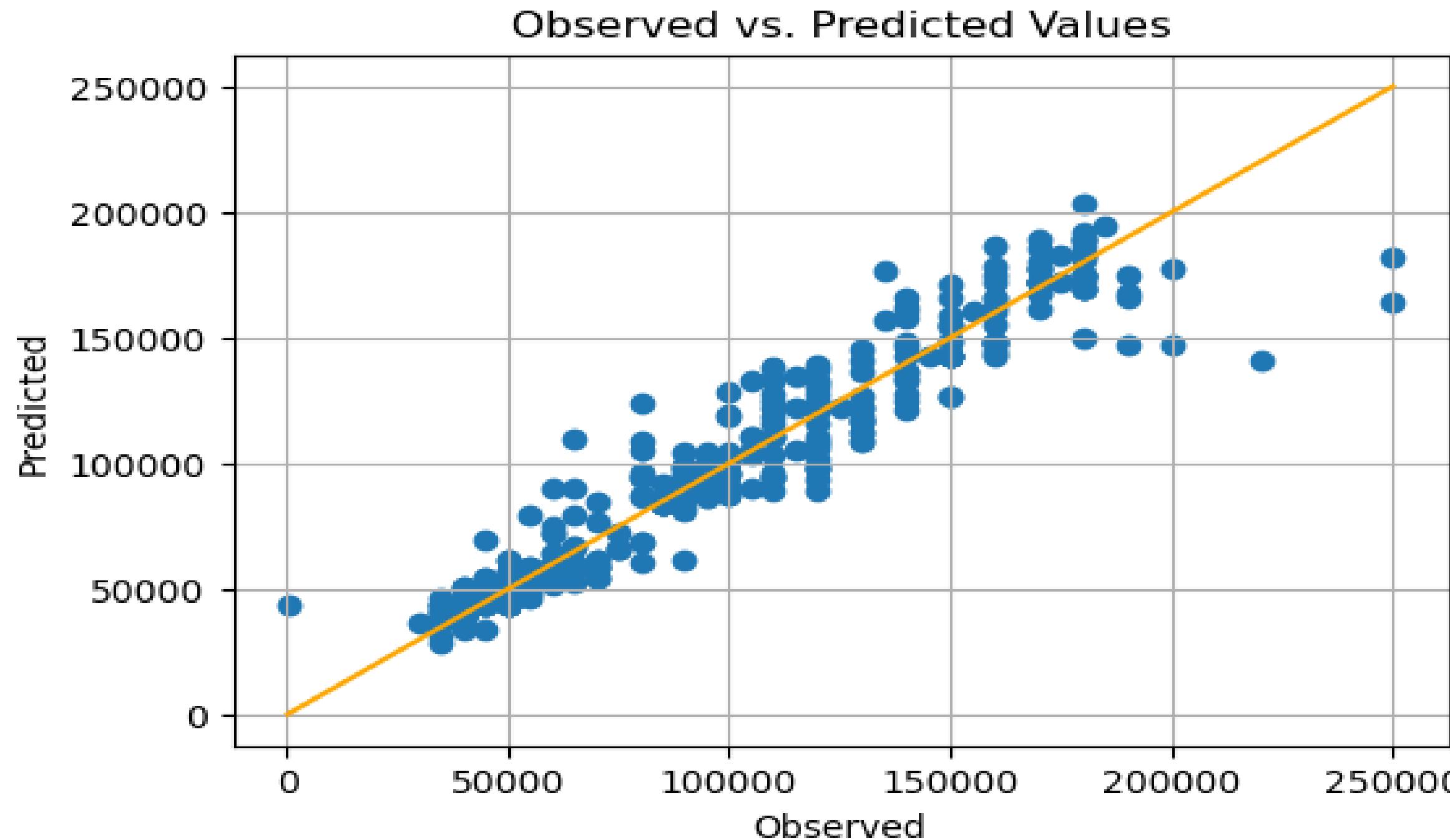
After their removal, all remaining fields were statistically significant.

# Regression Results

## OLS Regression Results

Dep. Variable:	Salary	R-squared:	0.914			
Model:	OLS	Adj. R-squared:	0.913			
Method:	Least Squares	F-statistic:	557.4			
Date:	Wed, 10 Dec 2025	Prob (F-statistic):	1.46e-190			
Time:	21:53:52	Log-Likelihood:	-4092.6			
No. Observations:	373	AIC:	8201.			
Df Residuals:	365	BIC:	8233.			
Df Model:	7					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
intercept	2.876e+04	1677.411	17.146	0.000	2.55e+04	3.21e+04
Gender_Male	7629.6521	1481.834	5.149	0.000	4715.649	1.05e+04
is_director	2.498e+04	3503.028	7.131	0.000	1.81e+04	3.19e+04
is_senior	1.445e+04	1820.427	7.936	0.000	1.09e+04	1.8e+04
is_manager	4284.6595	1822.582	2.351	0.019	700.579	7868.740
Years of Experience	5104.4388	177.226	28.802	0.000	4755.927	5452.950
Education Level_Master's	1.404e+04	2032.078	6.908	0.000	1e+04	1.8e+04
Education Level_PhD	2.308e+04	2760.740	8.360	0.000	1.77e+04	2.85e+04
Omnibus:	126.635	Durbin-Watson:		1.790		
Prob(Omnibus):	0.000	Jarque-Bera (JB):		1010.895		
Skew:	1.196	Prob(JB):		3.07e-220		
Kurtosis:	10.702	Cond. No.		60.8		

# Regression Results



# Interpretation of Model Results

## Years of Experience

With each additional year of experience, the expected change in salary is an increase of \$5,104.43.

We can be 95% confident that salary will increase between 4755.93 and 5452.95 dollars with each additional year of experience

# Interpretation of Model Results

## Senior Job Title

If the employee has a senior job title, the expected change in salary is an increase of \$14,450.

We can be 95% confident that salary will increase between 10,900 and 18,000 dollars with a senior job title.