

# The Value of an Education

The background of the slide is a deep blue gradient that transitions into a lighter blue and greenish hue towards the bottom right. Overlaid on this background is a complex network of thin white lines connecting small white dots, creating a geometric, crystalline pattern that resembles a molecular structure or a digital network. The lines and dots are more densely packed on the right side of the image, while the left side is more open, providing a clear space for the title text.

# The Problem

- In 2015-2016, ~600 billion was spent on college tuition in the US
- ~ 600 billion spent based on the decisions of 17-18 year old students?
- Are these students making good decisions?
- How should they choose?

<https://nces.ed.gov/fastfacts/display.asp?id=75>

# The Problem

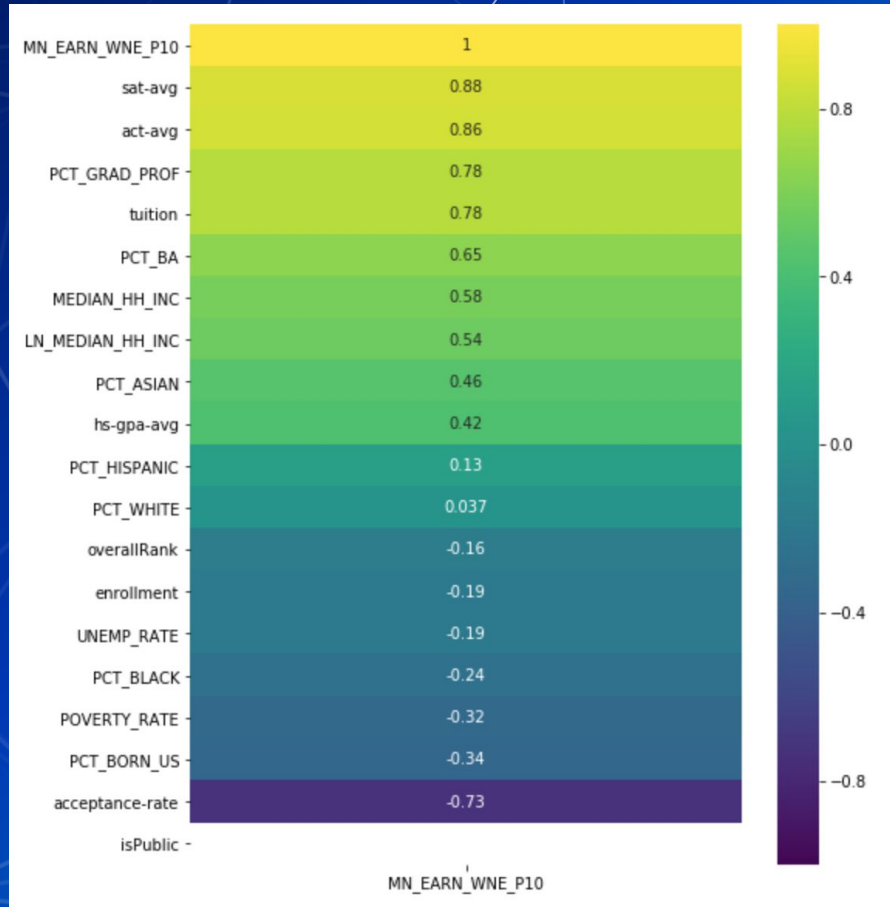
**Using the Kaggle and College Scorecard data, can we use machine learning methods to predict earnings based on the school attended?**

# The Dataset

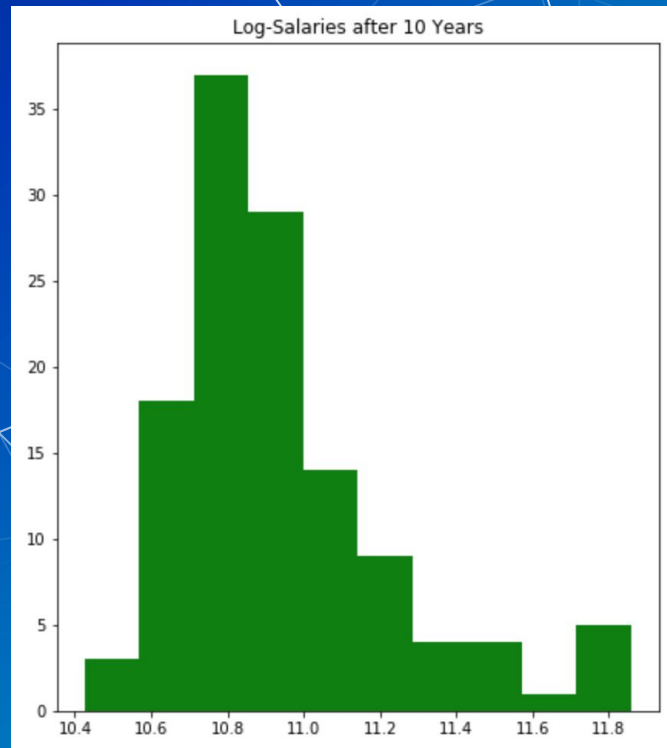
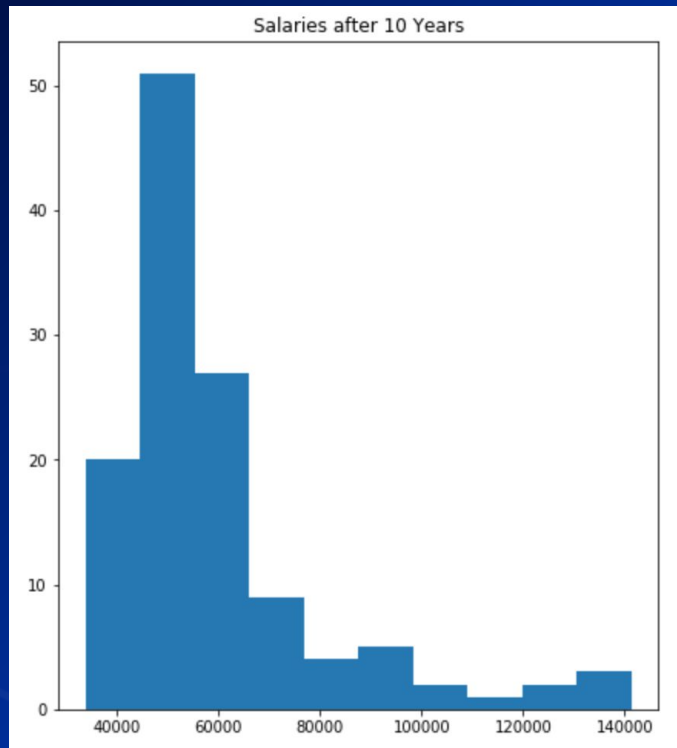
- **Collected from two sources**
  - **Kaggle - collected from [usnews.com](https://www.usnews.com)**
  - **College Scorecard**
- **Result - top ~300 universities**
- **Self-reporting issues**
- **Null imputation**

# Exploring the Data

# Correlations

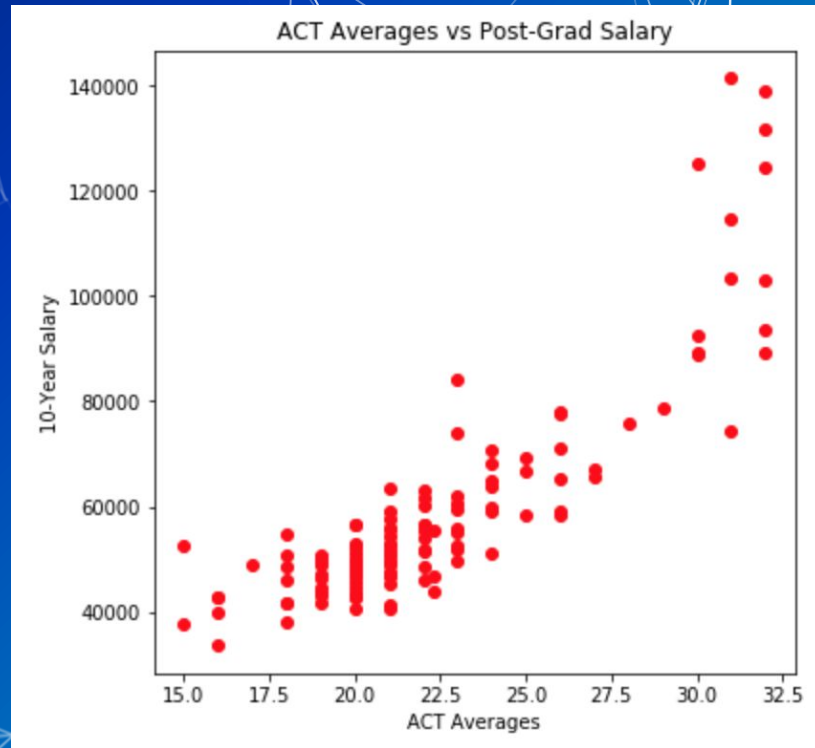
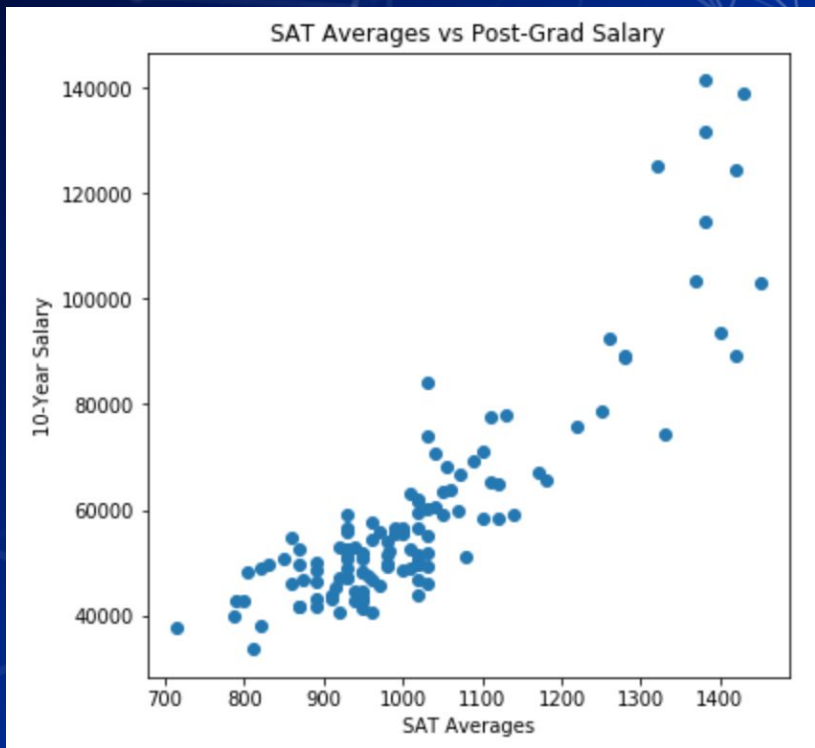


# The Target



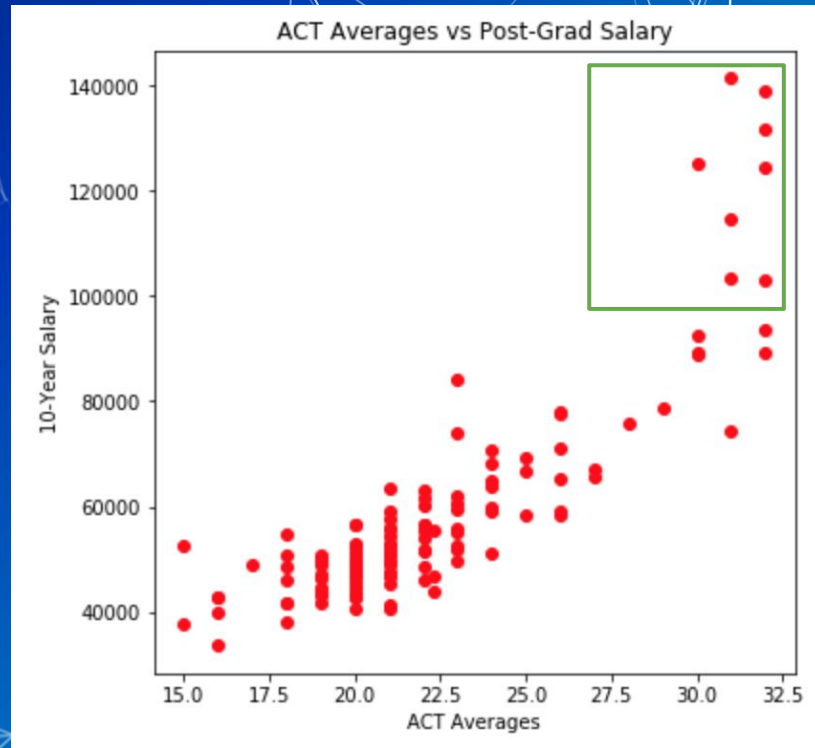
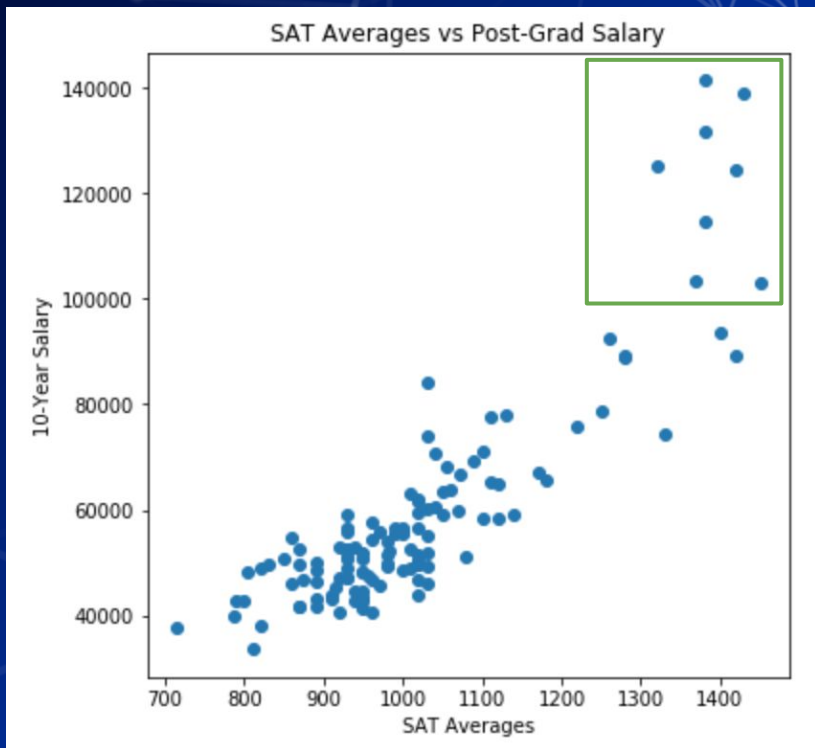


# Do the SAT and ACT Matter?

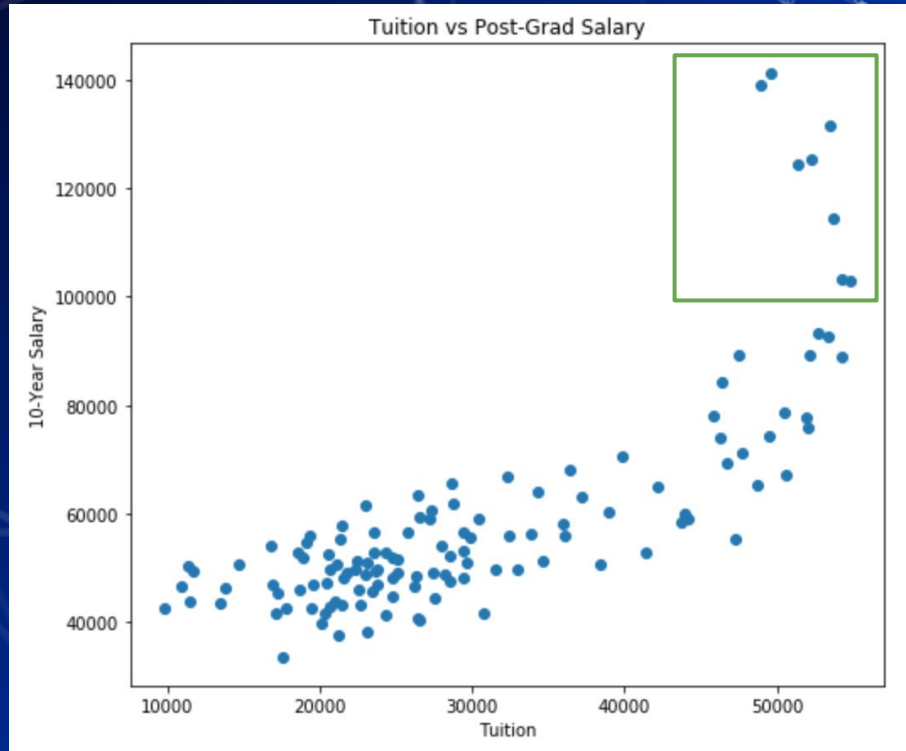




# Do the SAT and ACT Matter?



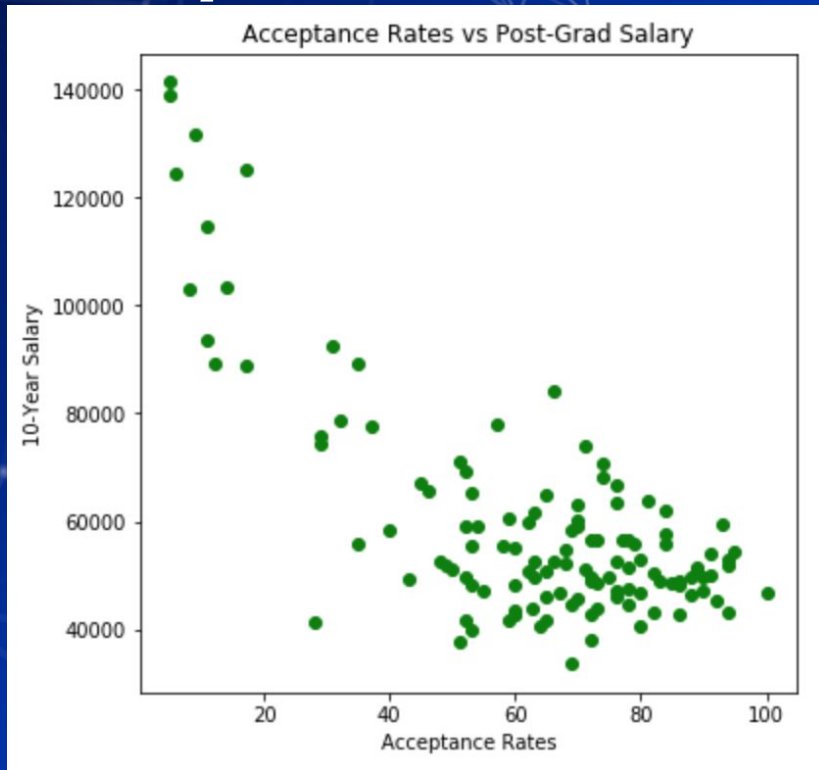
# Does Tuition Matter?



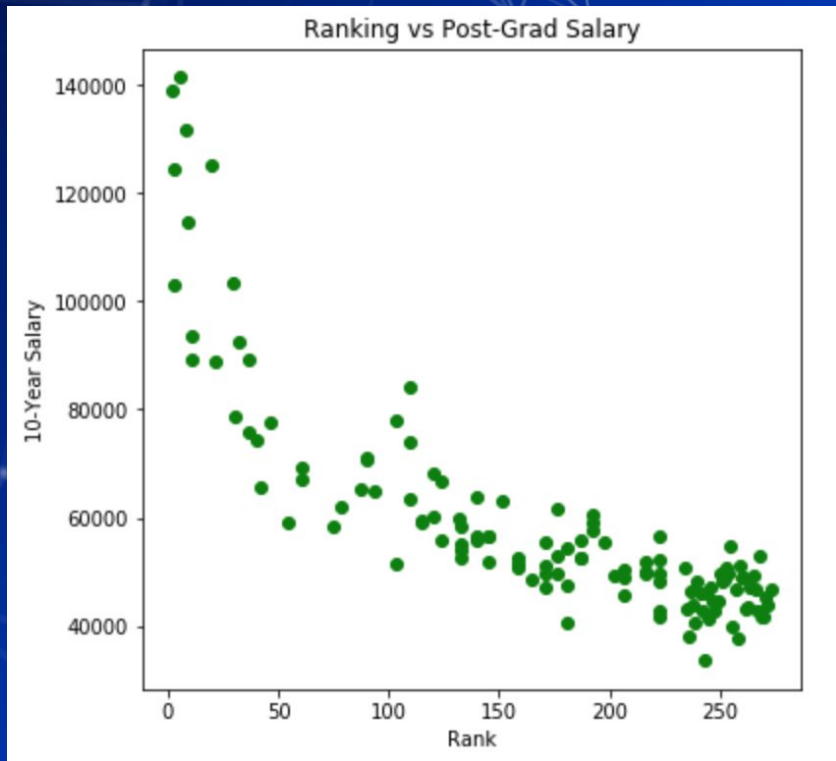
- Just like the SAT and ACT, Tuition rates are a strong predictor of future salaries

- Who are these outliers though?

# Exclusivity and Demand



# Does Rank matter?



# The Outliers

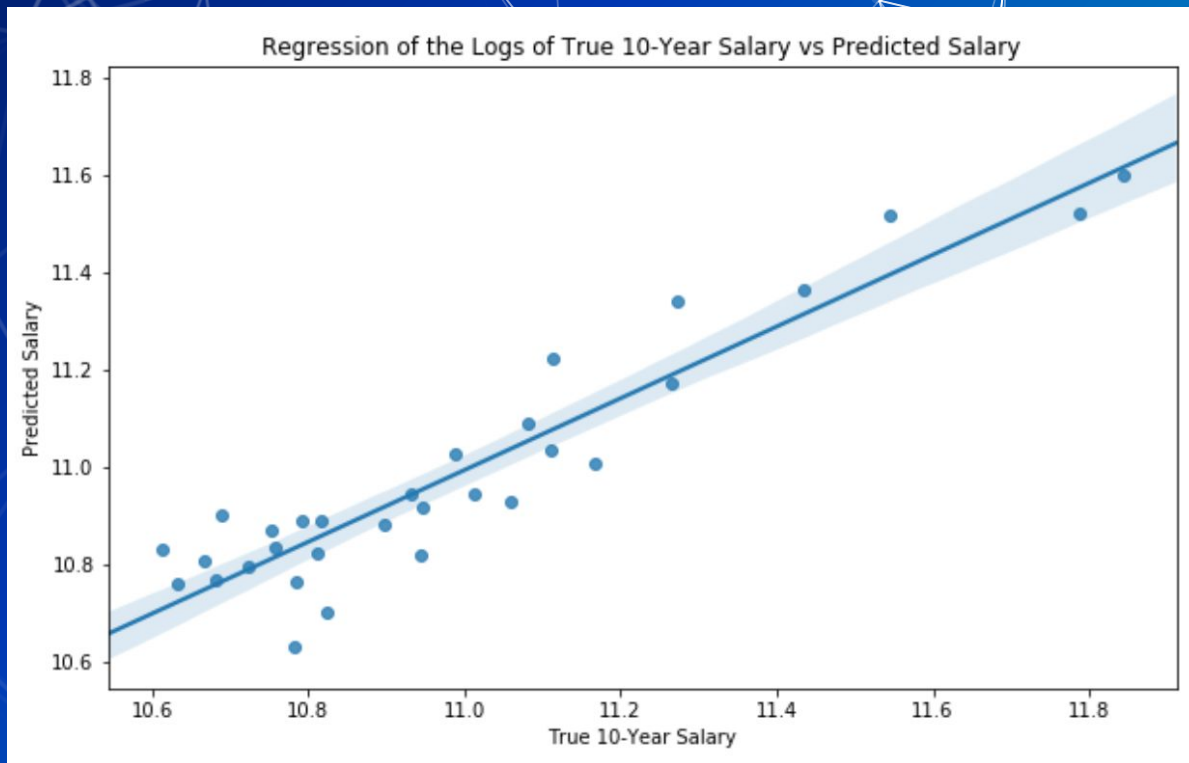
Turns out these are constantly the same “elite” universities

University	Tuition	Acceptance Rate	10-Yr Salary
Harvard U.	49k	5.0%	140k
U. Chicago	55k	8.0%	103k
Yale U.	51k	6.0%	124k
Stanford U.	49k	5.0%	141k
U. Penn	53k	9.0%	131k
Duke U.	54k	11.0%	114k
Georgetown	52k	17.0%	125k
Tufts U.	54k	14.0%	103k

# The Model

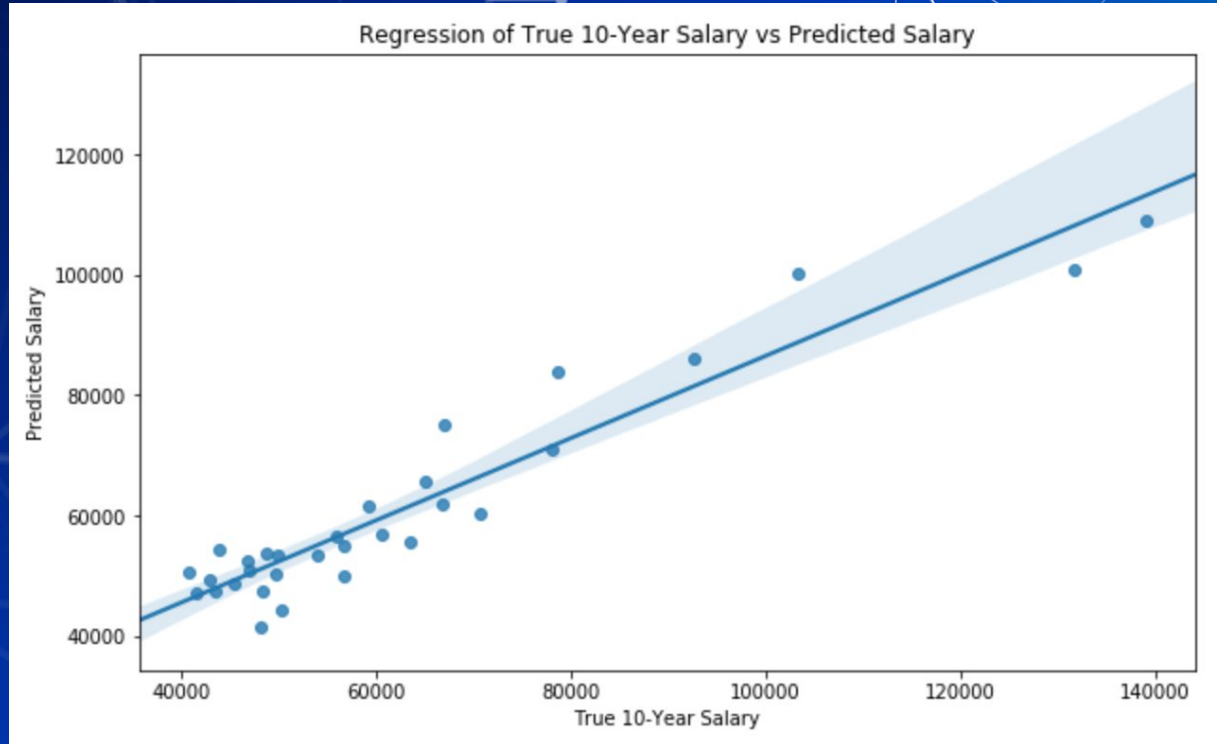
# The Model

**Linear Regression  
proved to be the most  
effective model to use**





# The Model



**0.8705 -> 0.8543**

**Train Accuracy**

**Test Accuracy**

**9,466.51 \$**

**Root Mean Squared Error**

**6,540.56 \$**

**Mean Absolute Error**

# Important Features

## Percent Graduate Professors:

- For every 1% increase in Graduate Professors, we see a 1.01 increase in 10-year salaries for the university

PCT_HISPANIC	1.005026
act-avg	1.005413
PCT_WHITE	1.005626
PCT_GRAD_PROF	1.012784
PCT_ASIAN	1.026791

# Conclusions

- Generally, higher values in each category result in higher salaries
- Rank as it turns out is a decent rule of thumb
- Outliers exist at the top of the ranks where the “name” carries extra value

# Next Steps

- Expanding scope of data to more years and universities
- Building a multi-feature recommender model with more comprehensive data
- Cost vs. Value assessment based on cost of attendance and employment/salary outcomes
- Compare against salaries of non-college-goers

# **CREDITS**

**Special thanks to my amazing and talented  
classmates and instructors for a wonderful  
12 weeks**

**And also to General Assembly**





**Questions?**



# Sources

- <https://collegescorecard.ed.gov/data/>
- <https://collegescorecard.ed.gov/data/>