



**Stephanie Kui**

GA DSI 21 - Capstone Project



# The Data Science Job Market in SG

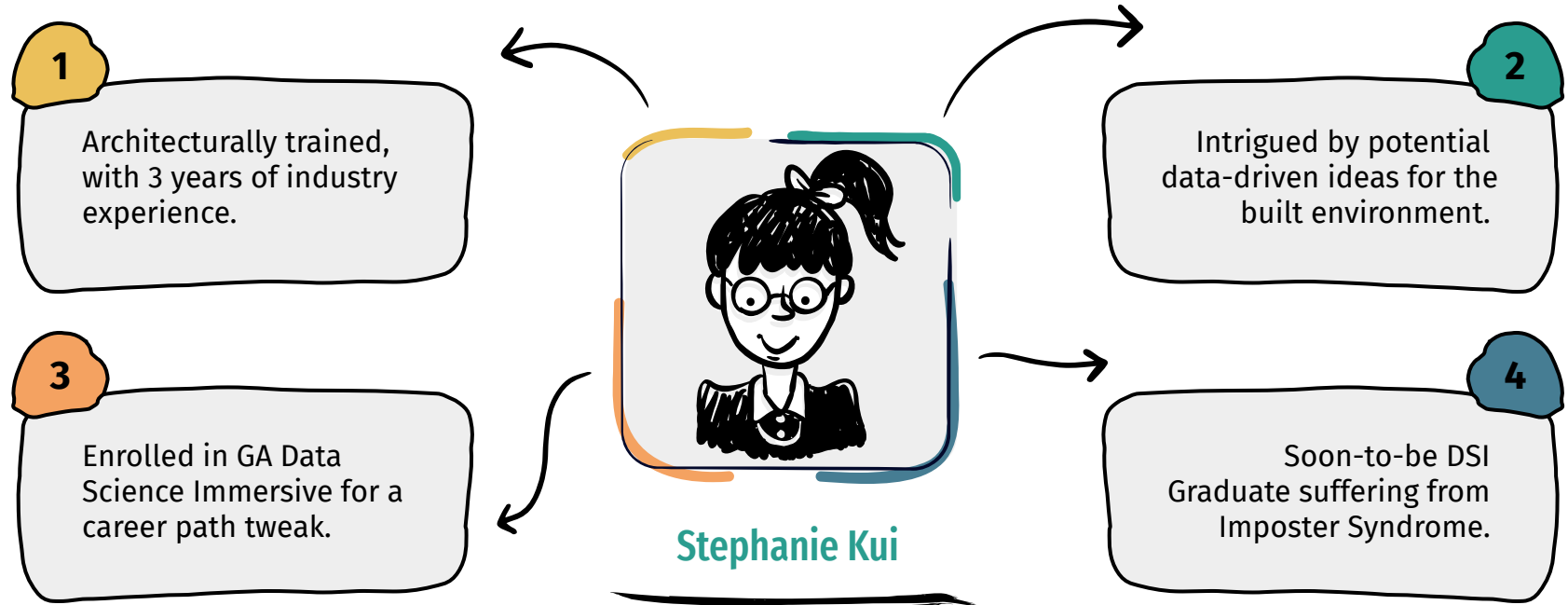
Predicting Your Potential Salary

**Skills**



# Short Background about Me

(& why I chose this capstone topic)



# Why a Salary Prediction Model?

**Salary Negotiations**  
For Job Seekers

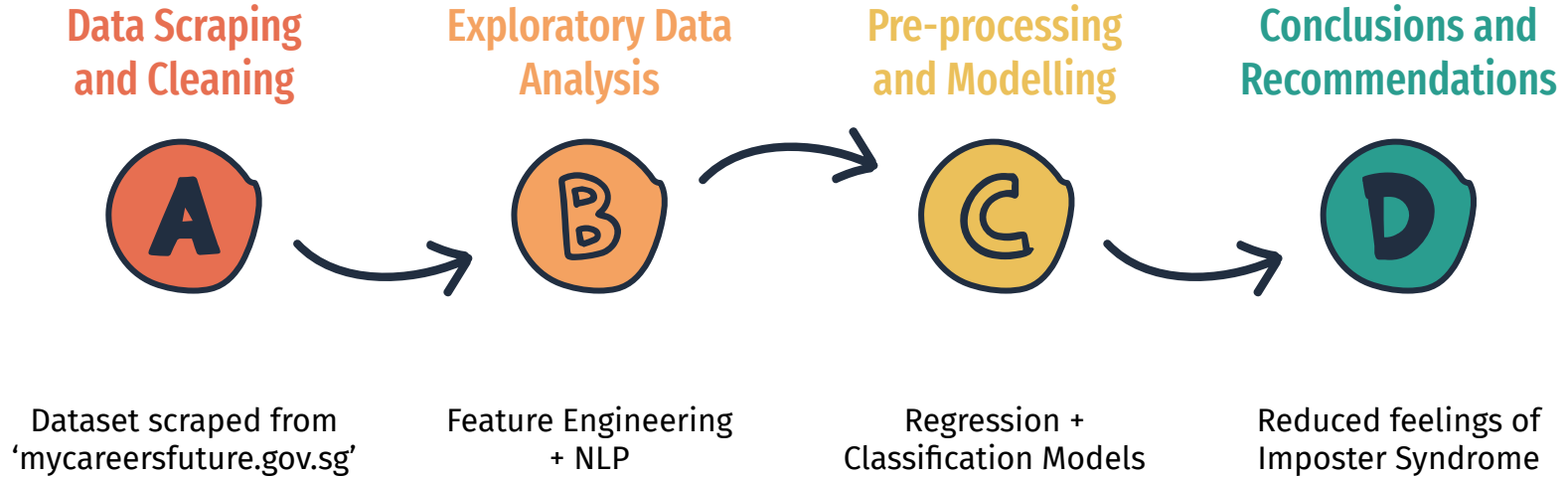


**Hiring Budgets**  
For Companies

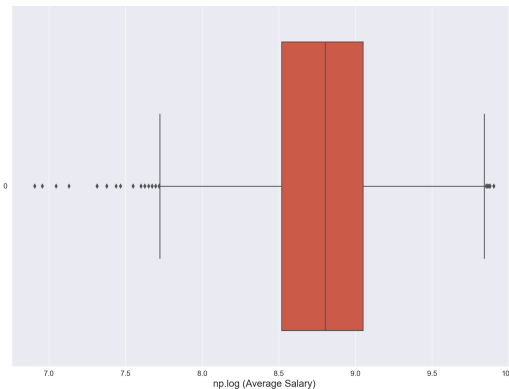
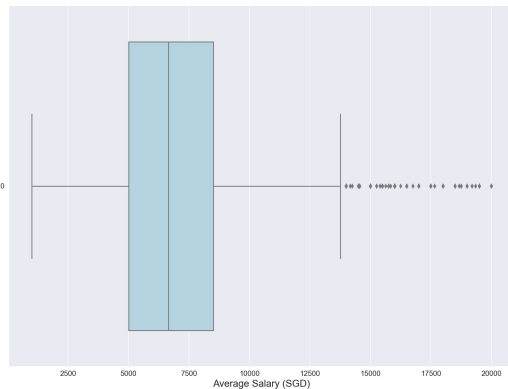
**Improving my Hireability**  
For Job Seekers

“So you want to hire me as a Data Scientist for Intelligent Virtualized Deep Machine Learning Real-time Big Data in the Cloud for Social Networks? Ok, but if you also want Hadoop, increase my salary by 50%.”

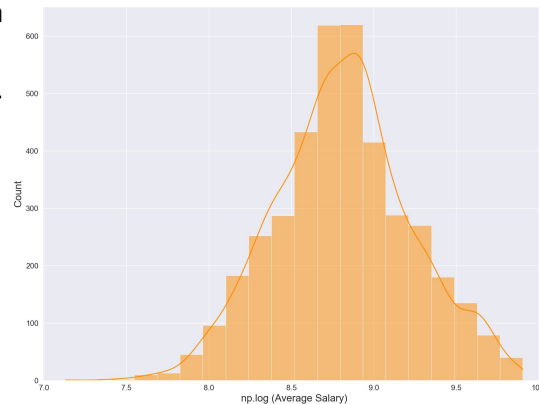
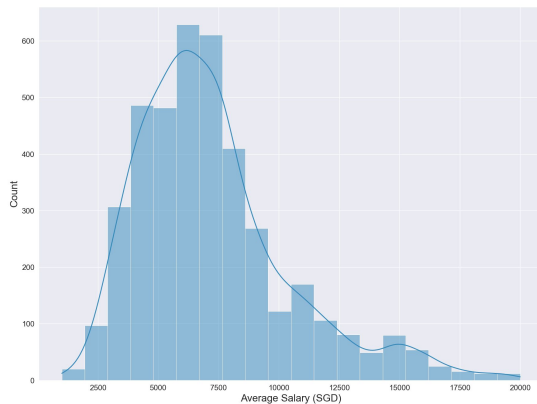
# Project Workflow



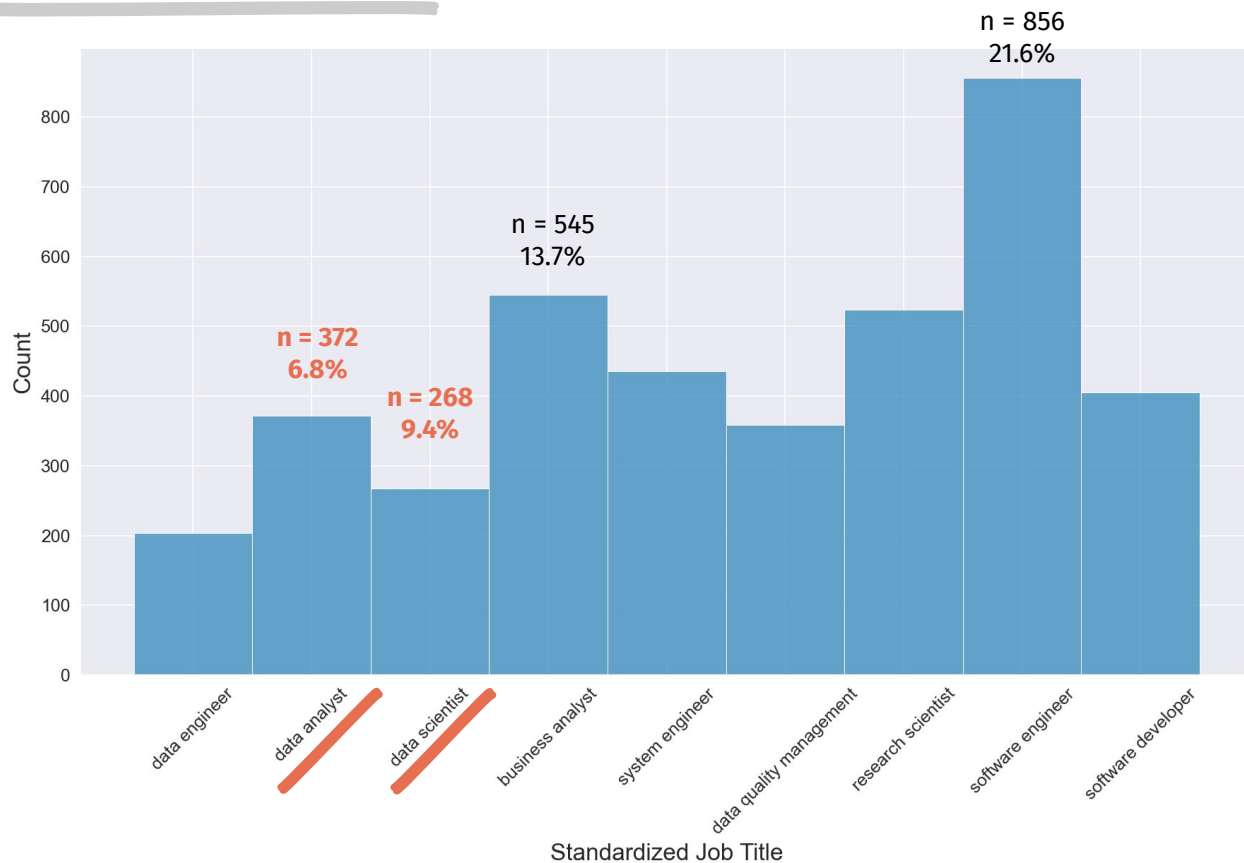
# Distribution of 'Average Salary' in dataset



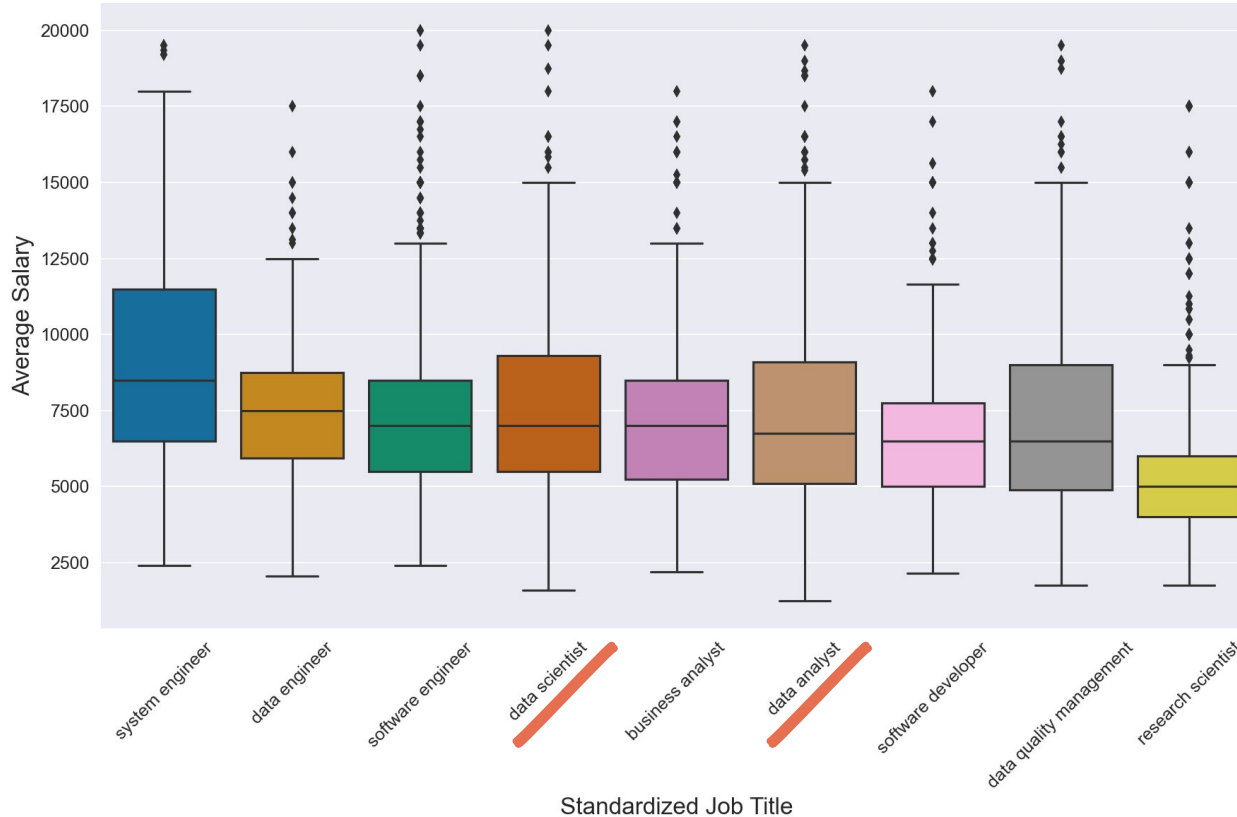
Applying log transformation to achieve a more normalized distribution



# Jobs in demand for the Data Science Industry



# Job Title relative to Average Salary



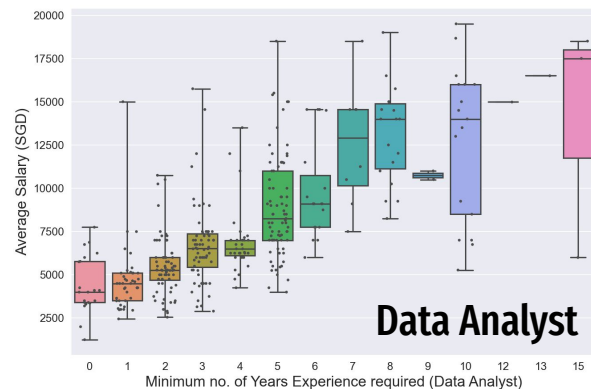
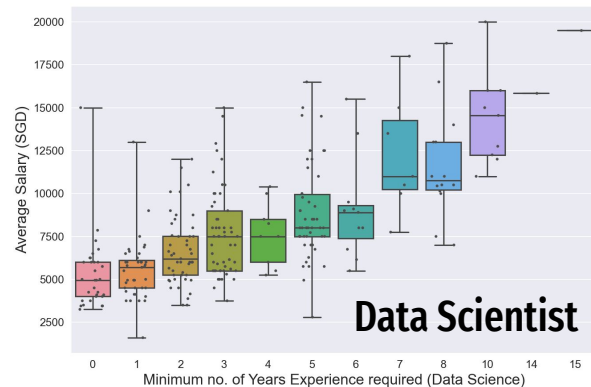
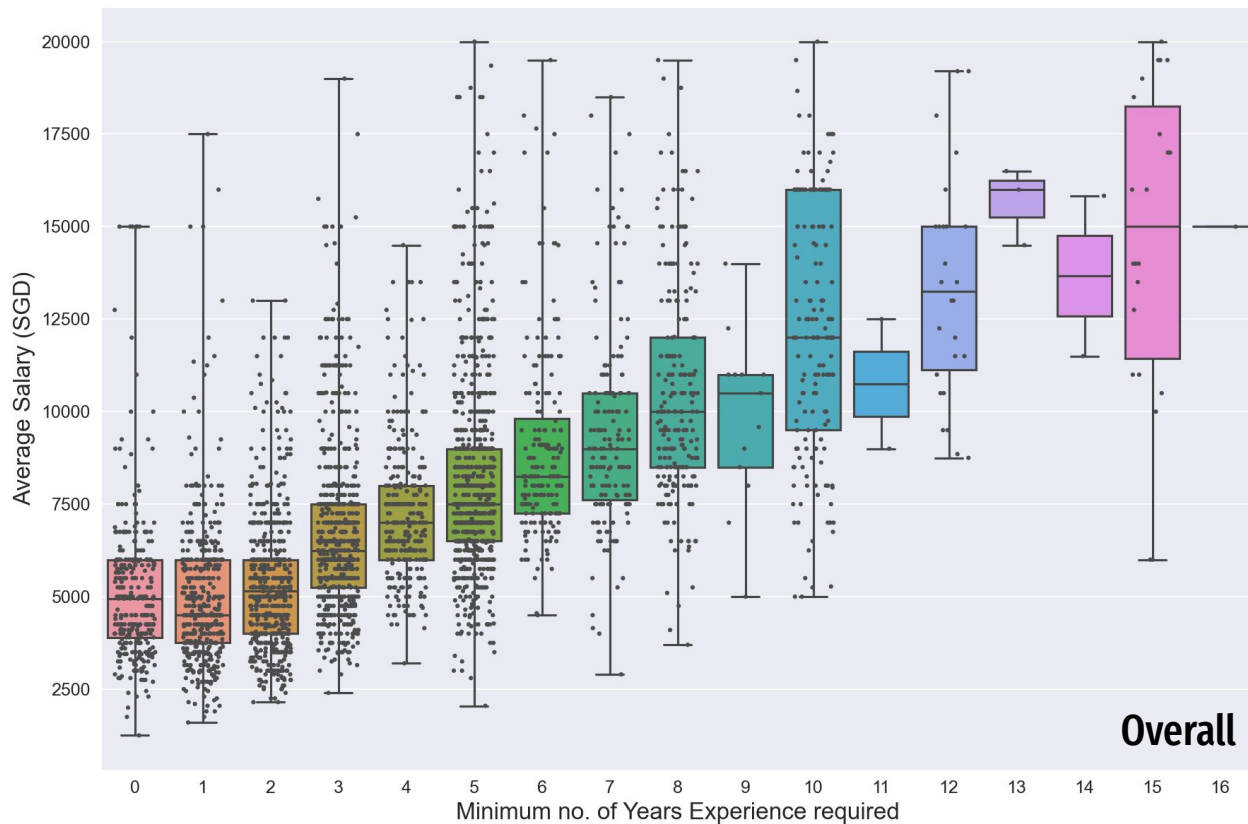
## Mean

System Engineer	<b>\$9,200</b>
Data Scientist	<b>\$7,800</b>
Business Analyst	<b>\$7,200</b>
Data Analyst	<b>\$7,600</b>
Research Scientist	<b>\$5,500</b>

## Median

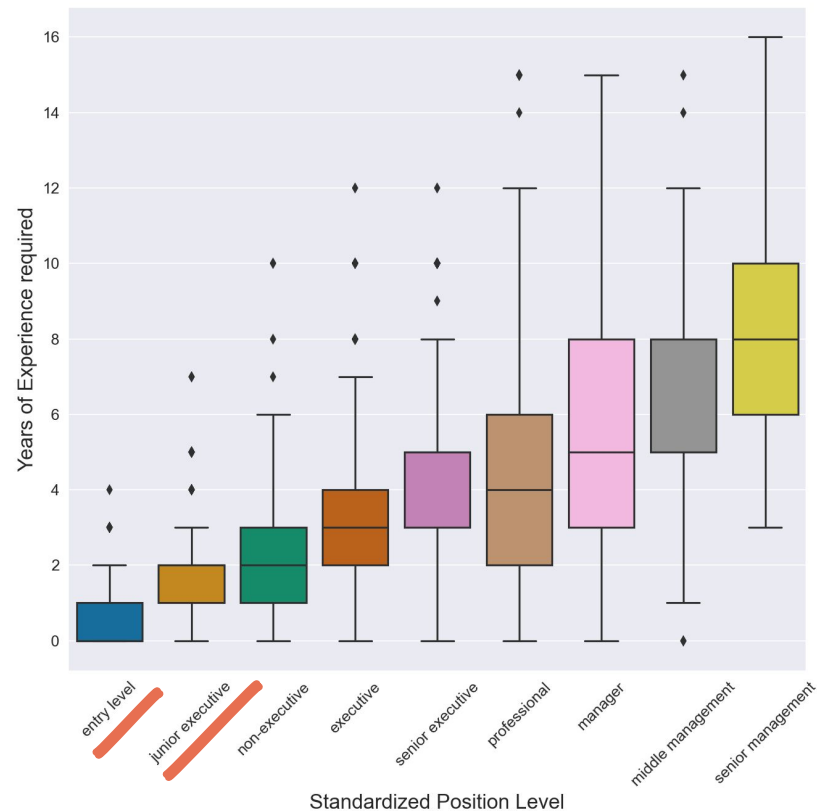
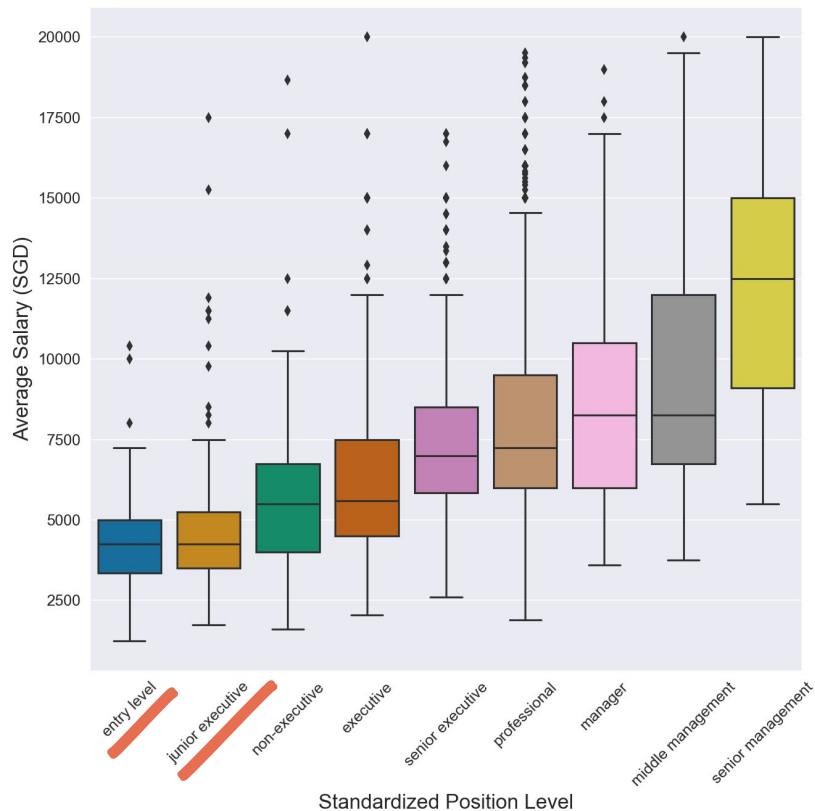
System Engineer	<b>\$8,500</b>
Data Scientist	<b>\$7,000</b>
Business Analyst	<b>\$7,000</b>
Data Analyst	<b>\$6,750</b>
Research Scientist	<b>\$5,000</b>

# Years of Experience relative to Average Salary

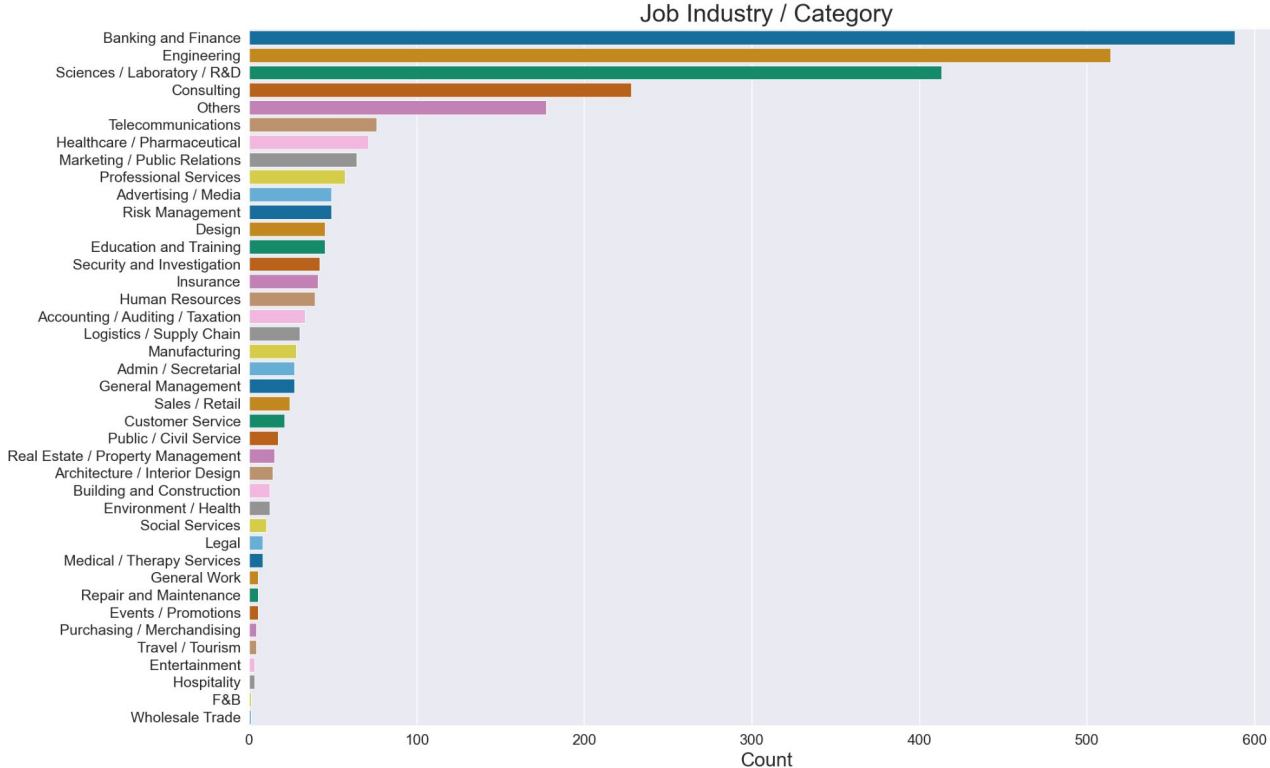




# Career Progression in the Data Science Industry



# Transferable Industry Knowledge



## Top Industries

1

Banking & Finance

2

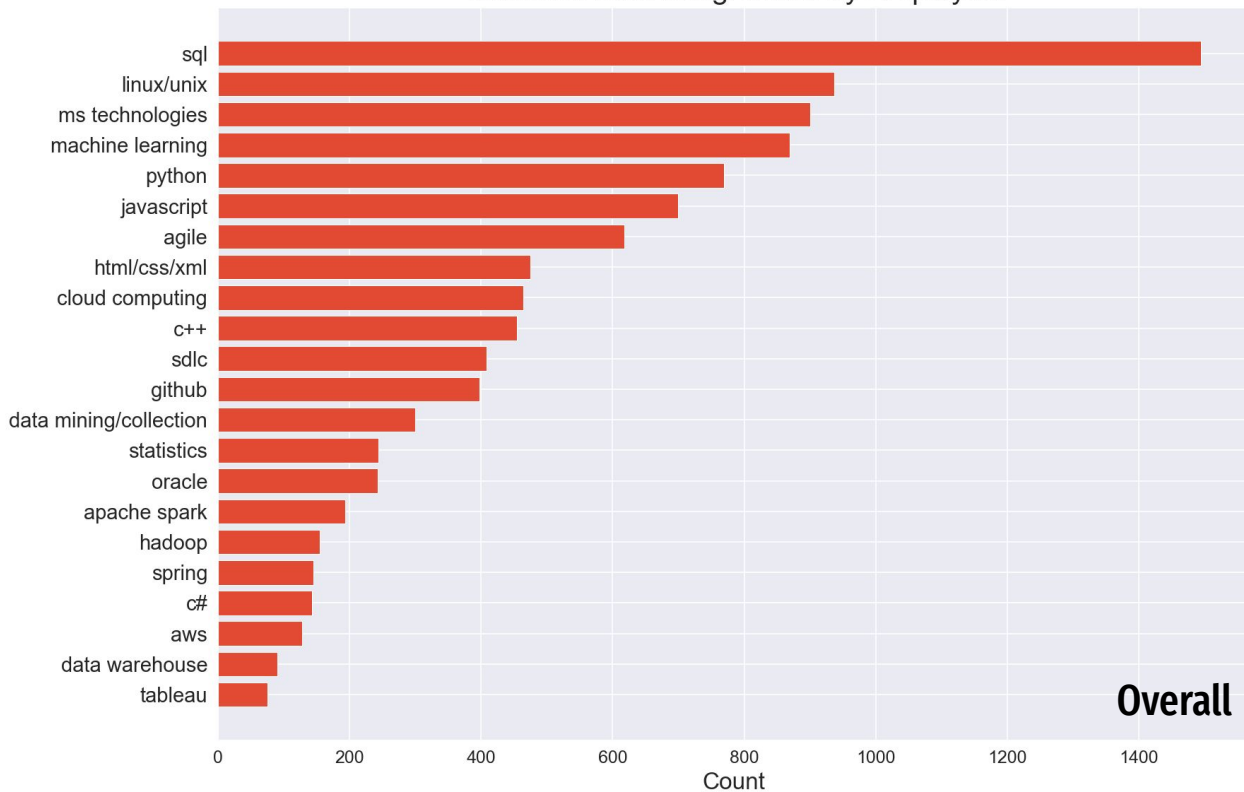
Engineering

3

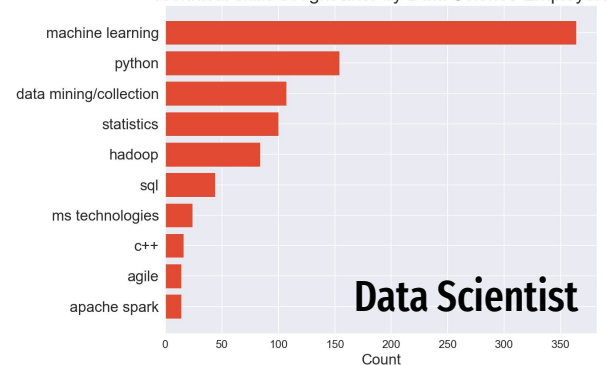
Sciences/Lab/R&D

# Improve your Hireability - Top skills to acquire

Technical skills sought after by Employers

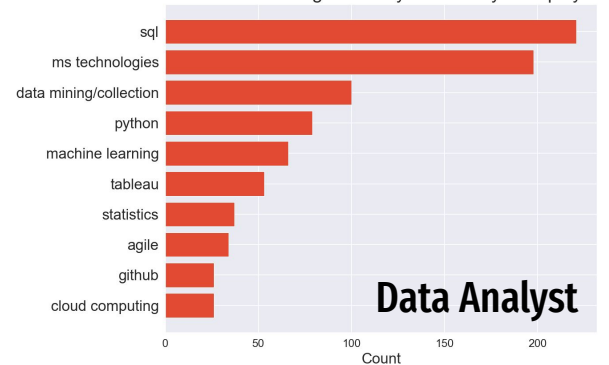


Technical skills sought after by Data Science Employers



**Data Scientist**

Technical skills sought after by Data Analyst Employers



**Data Analyst**

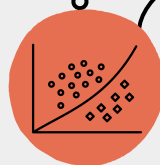
# Pre-processing & Modelling Workflow

## Input

## Best Model

1

Numerical  
& OHE Variables



Regression

Predicting 'Average Salary'

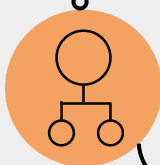
**Gradient Boosting  
Regressor**

Train Score / Test Score

**0.638 / 0.576**

2

Numerical  
& OHE Variables



Classification

Predicting 'Salary above Median'

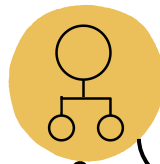
**Logistic Regression**

Train Score / Test Score

**0.668 / 0.634**

3

Text / String  
Variables  
(NLP)



Classification

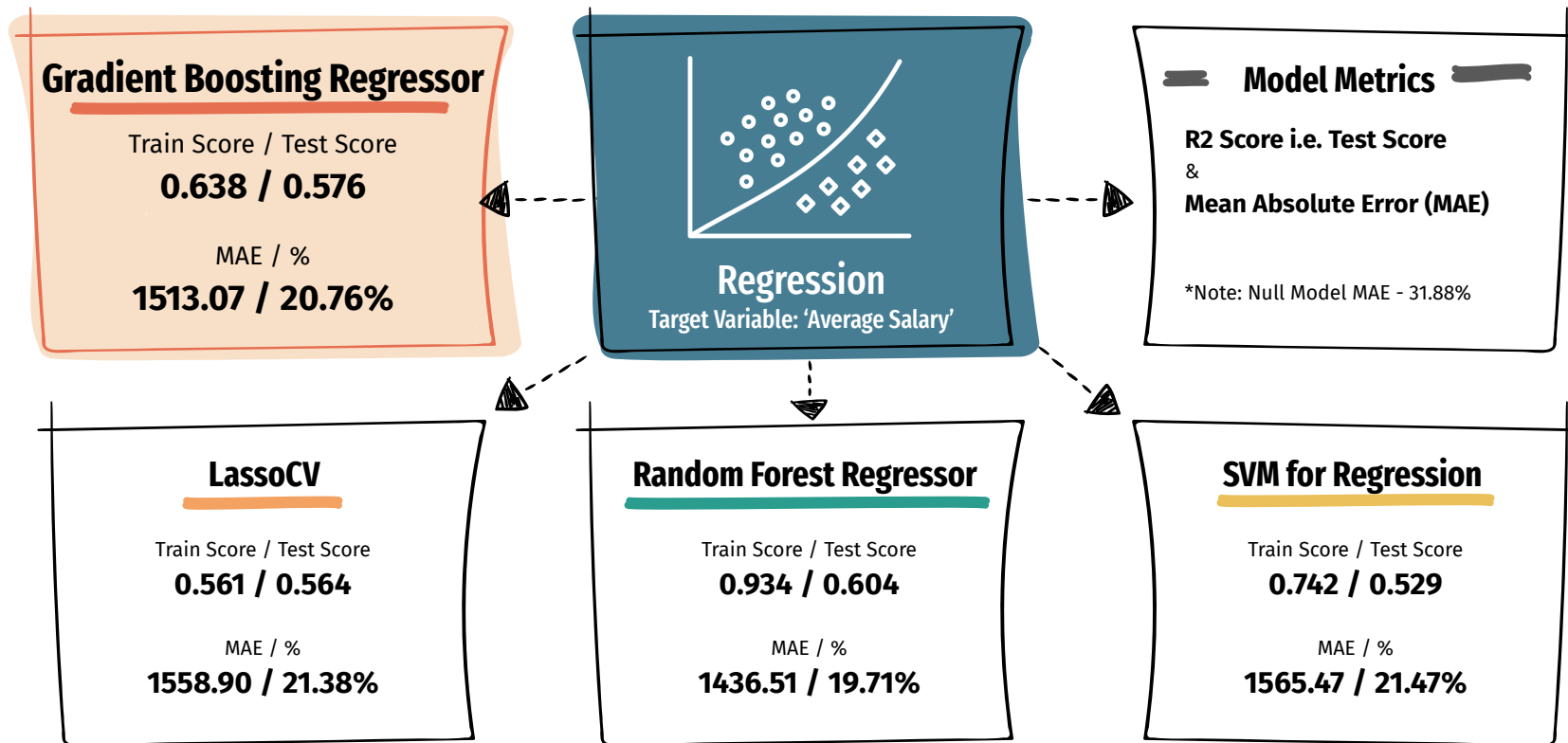
Predicting 'Salary above Median'

**Multinomial  
Naive Bayes**

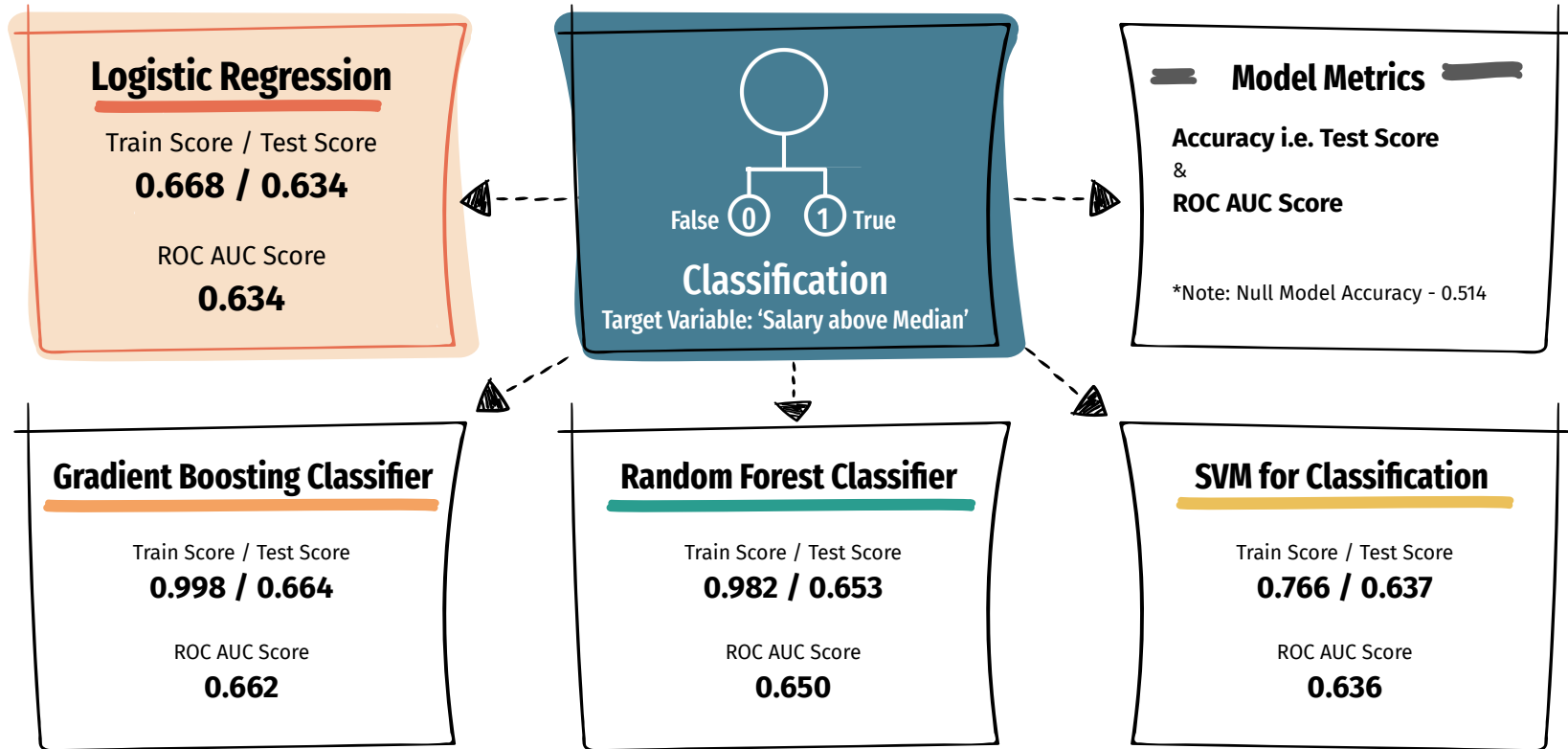
Train Score / Test Score

**0.813 / 0.694**

# Model 1 - Regression Modelling Results

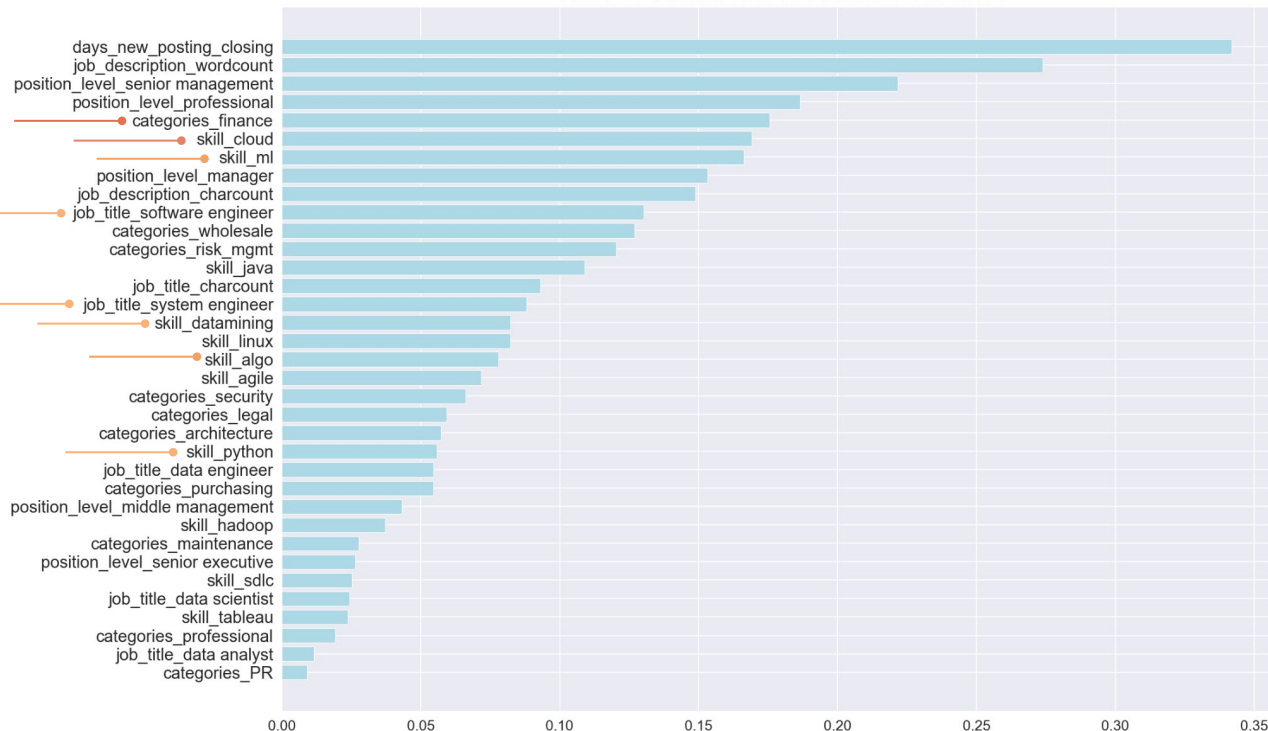


# Model 2 - Classification Modelling Results



# What can you do to increase your Potential Earnings?

Value of Coefficients to Predictor Variables



1

Break into the Fintech industry

2

Pick up Cloud Computing skills

3

Familiarize yourself with  
Machine Learning Algorithms

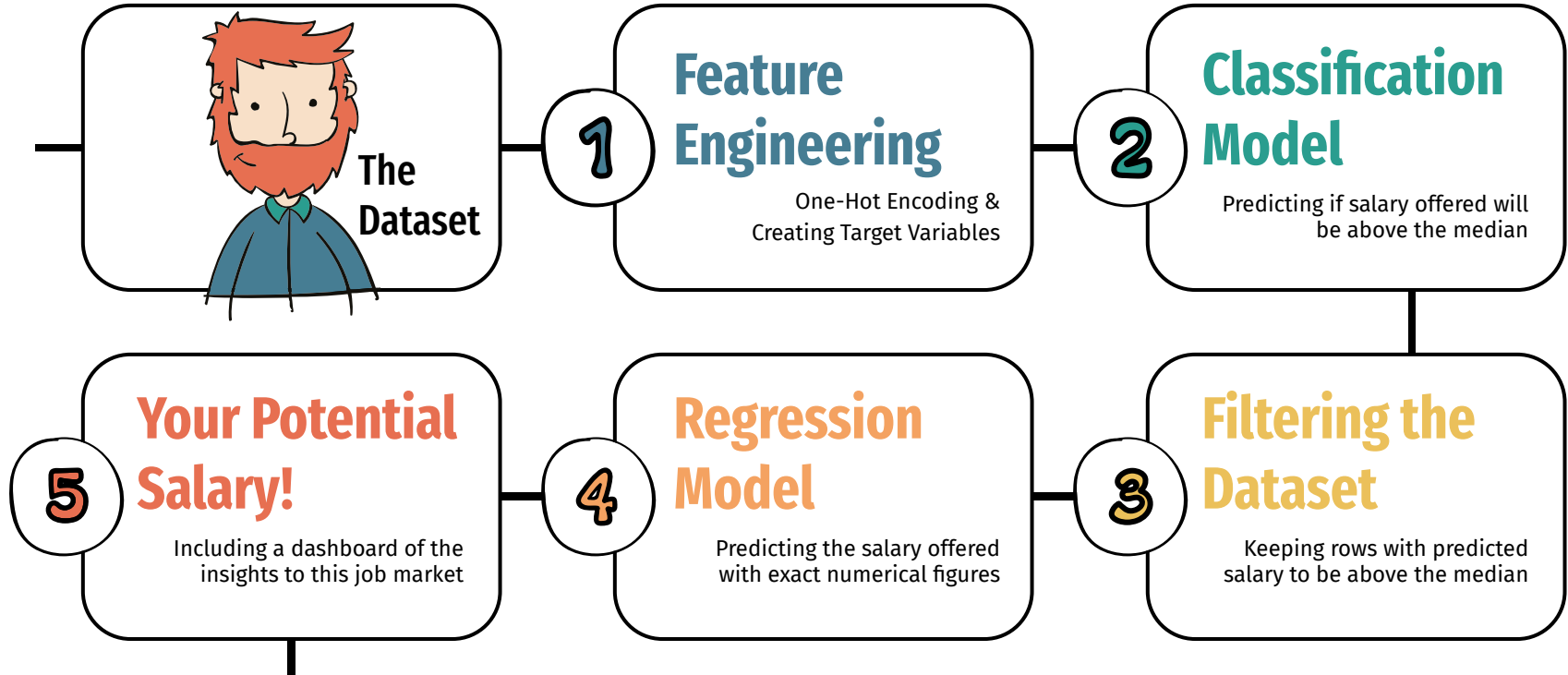
4

Look out for Software Engineer /  
System Engineer roles

5

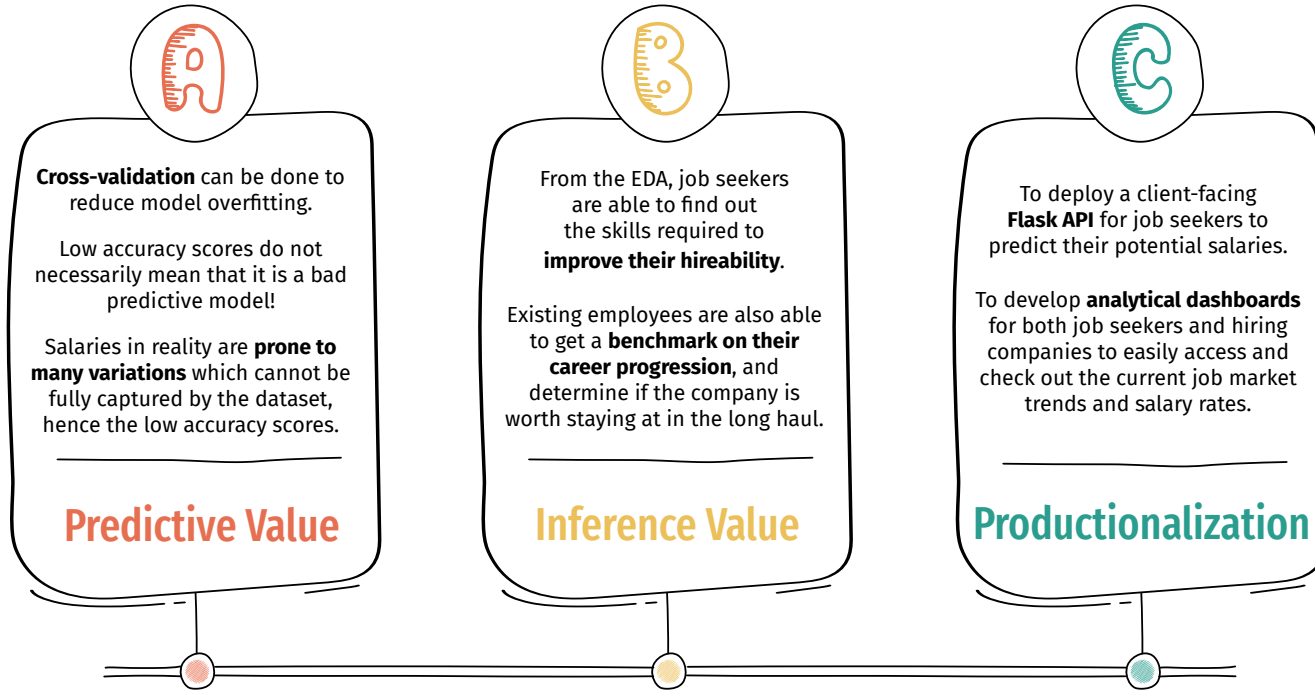
Improve your data mining skills  
e.g. SQL, Python, Scraping

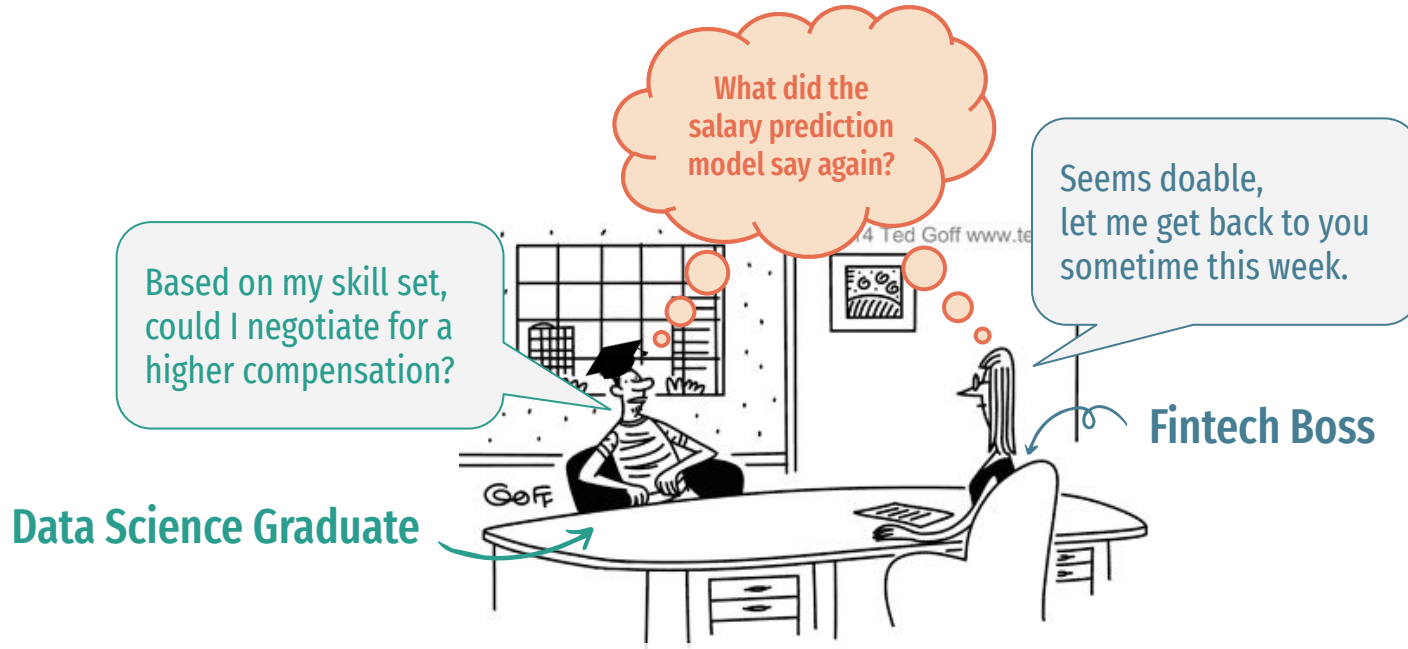
# Recommended Modelling Process





# Conclusion & Future Development





**END**

Thank you for your time!