



# Webscraping Indeed.com

## Project 3



## Research Objective

1. Determine the industry factors that are most important in predicting the salary amounts for these data.
2. Determine the factors that distinguish job categories and titles from each other. For example, can required skills accurately predict job title?



### Research questions:

1. factors that are most important in predicting the salary amounts for these data?
2. What factors distinguish job categories from each other?



## Tasks:

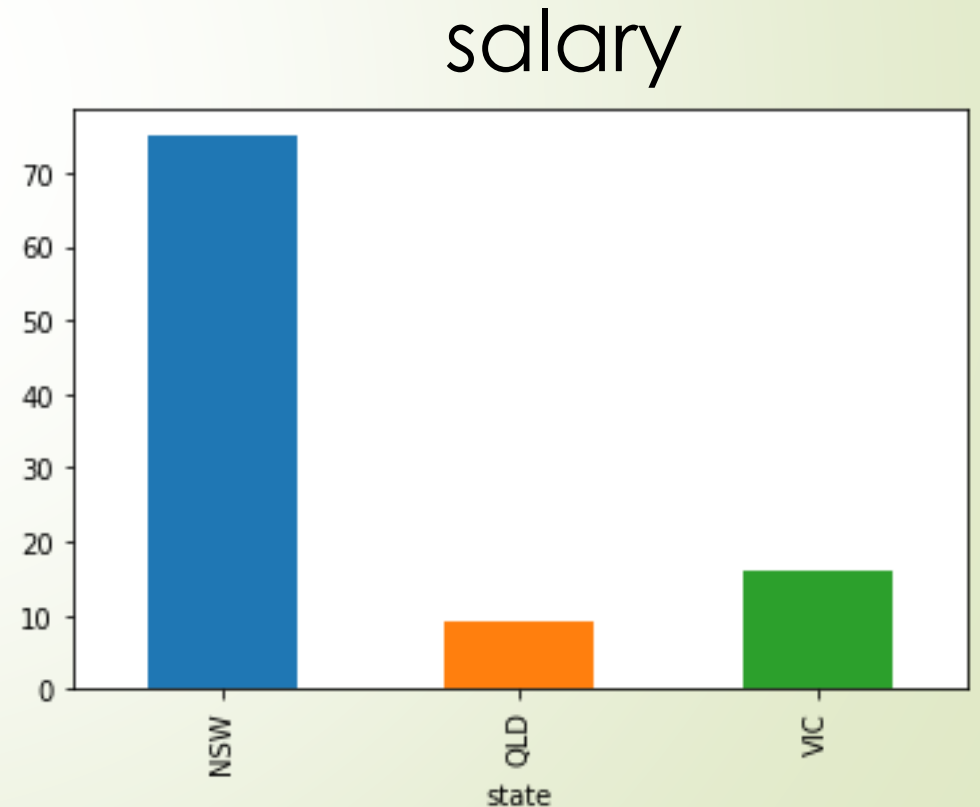
- ➡ 1. Scrape my own data from a job Indeed.com in order to collect the data to best answer these two questions.
- ➡ 2. Clean and Analyse the data

# Analysis

- ➡ scrape\_data.shape:
- ➡ (Rows:1029, Columns: 9)
- ➡ Create an avg salary

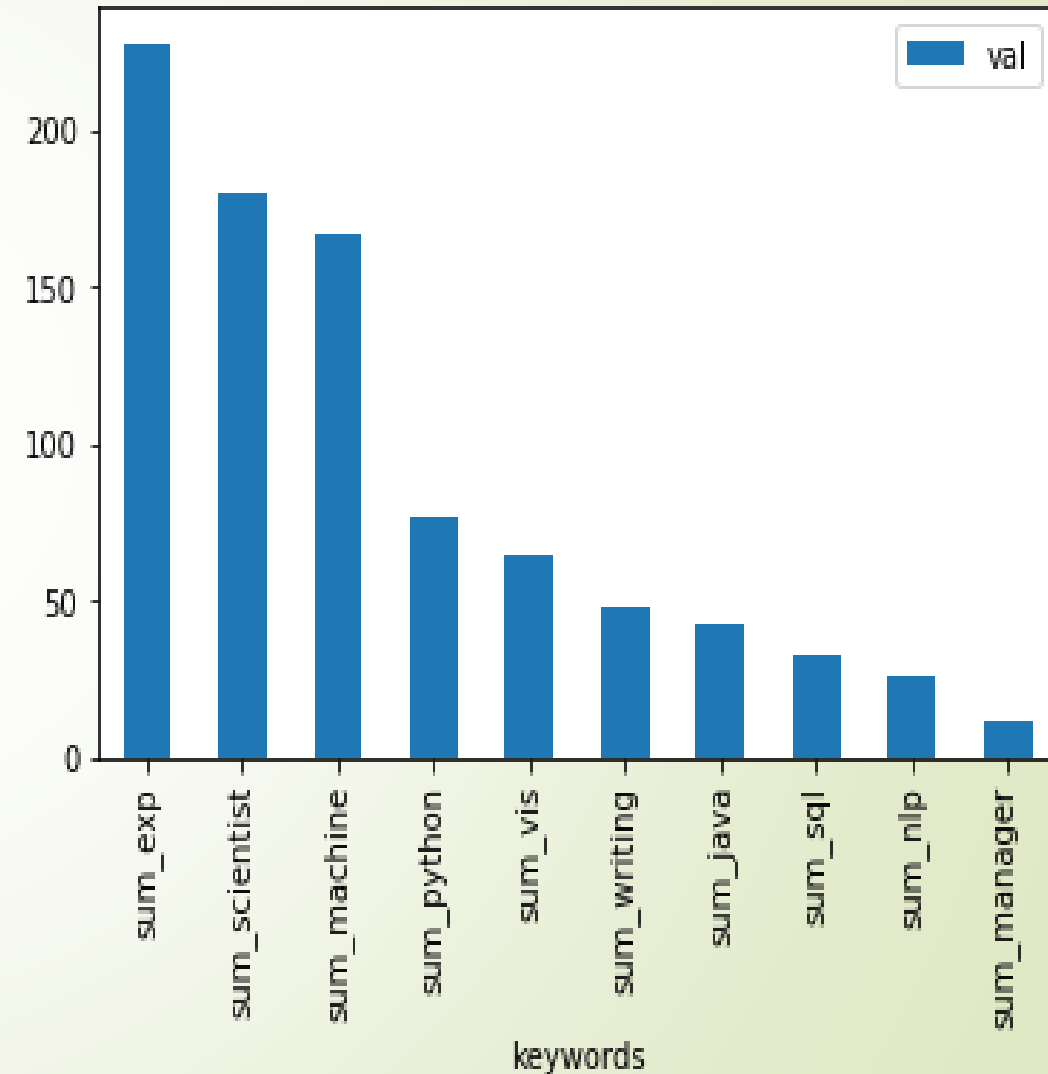
Used : logistic regression and  
RandomForestClassifier

Ir.Score=~0.9

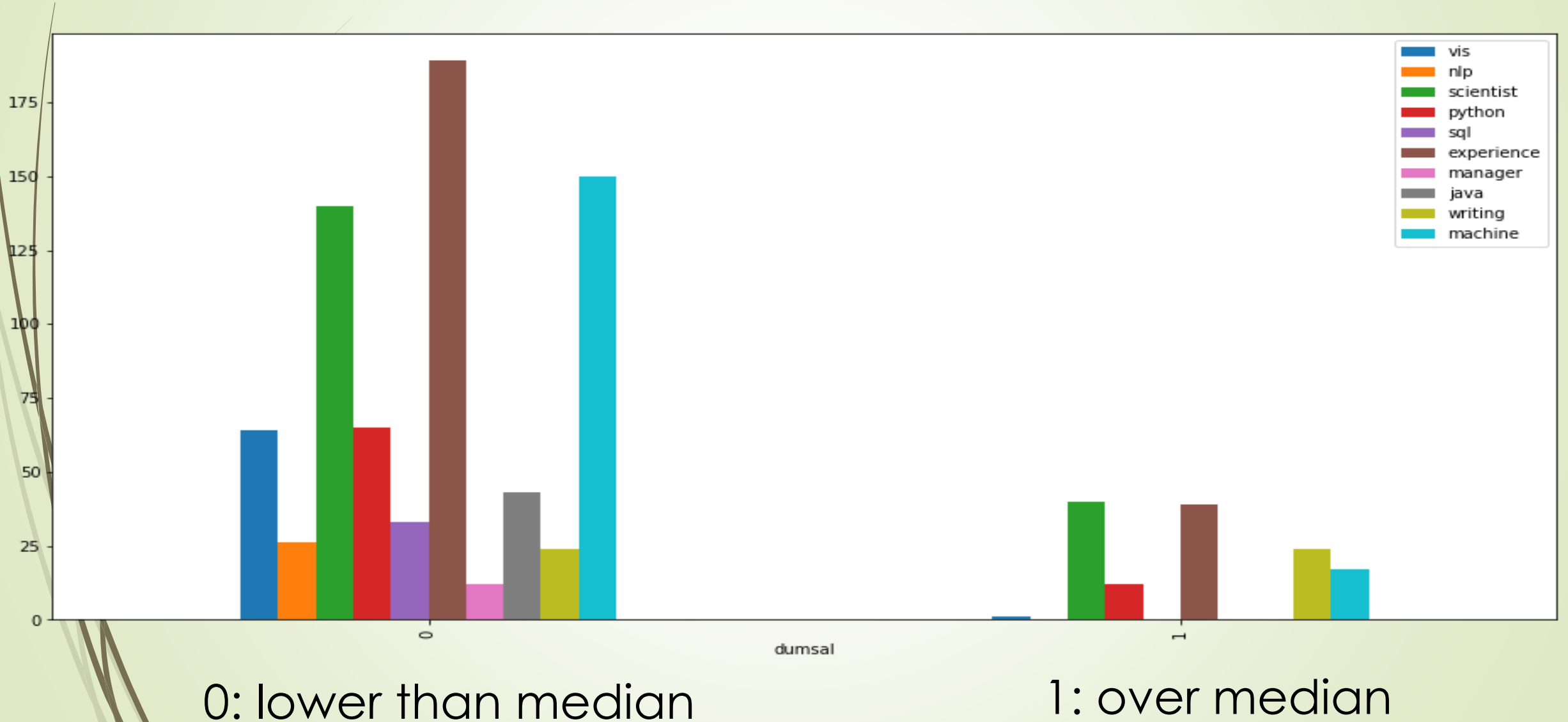


# Job description: Import job description words

<b>writing</b>	0.222347
<b>experience</b>	0.093193
<b>intelligence</b>	0.063077
<b>linux</b>	0.054975
<b>artificial</b>	0.051900
<b>programming</b>	0.045941
<b>solving</b>	0.027982
<b>machine</b>	0.027056
<b>scientist</b>	0.020335
<b>statistics</b>	0.019148
<b>sql</b>	0.016297

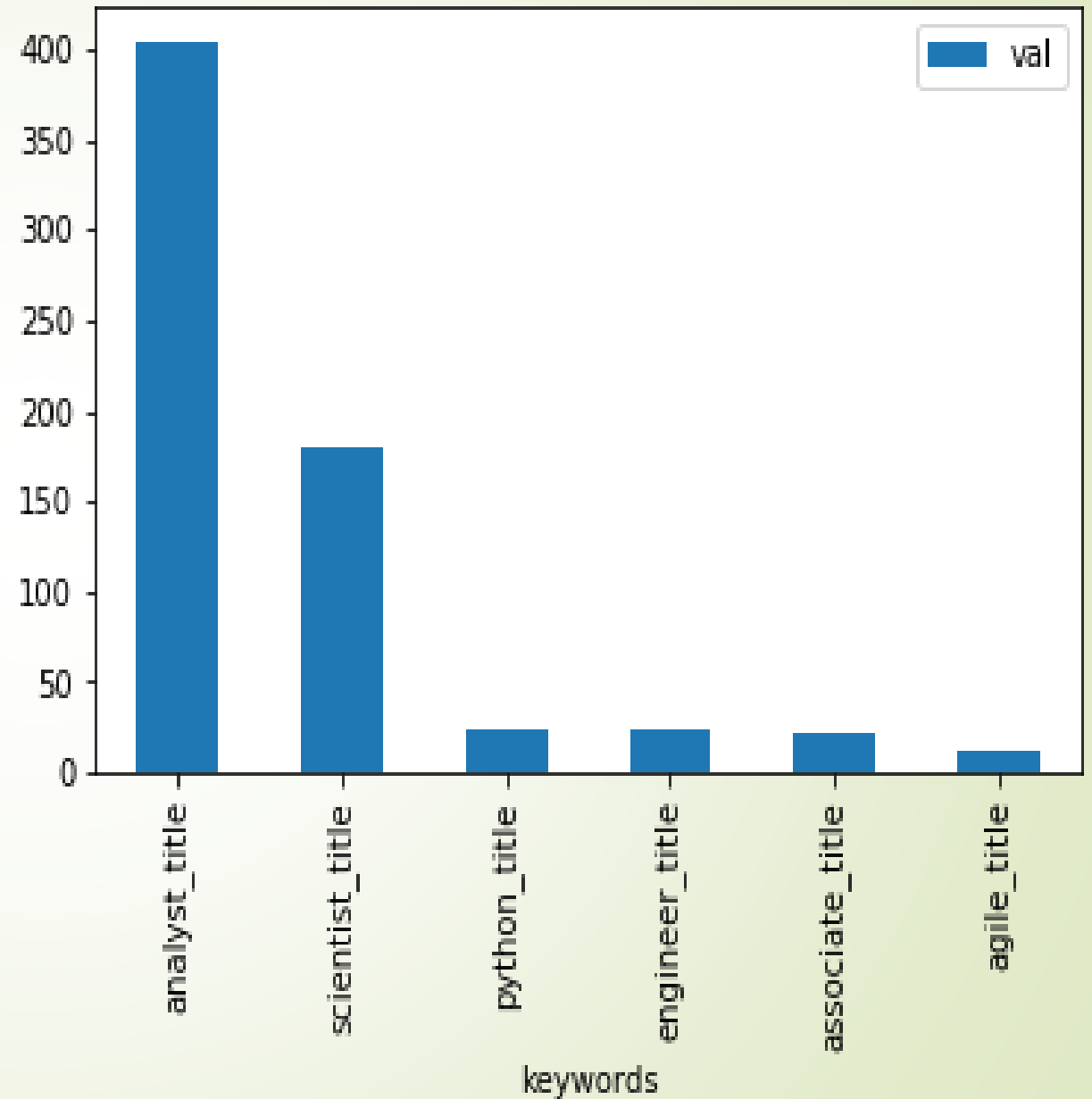


# Job description: Important keywords for two jobs payments



# JOB Title

<b>scientist</b>	0.088349
<b>analyst</b>	0.078580
<b>associate</b>	0.076165
<b>engineer</b>	0.069471
<b>python</b>	0.067455
<b>agile</b>	0.056153
<b>business</b>	0.055829



# Important keywords:

