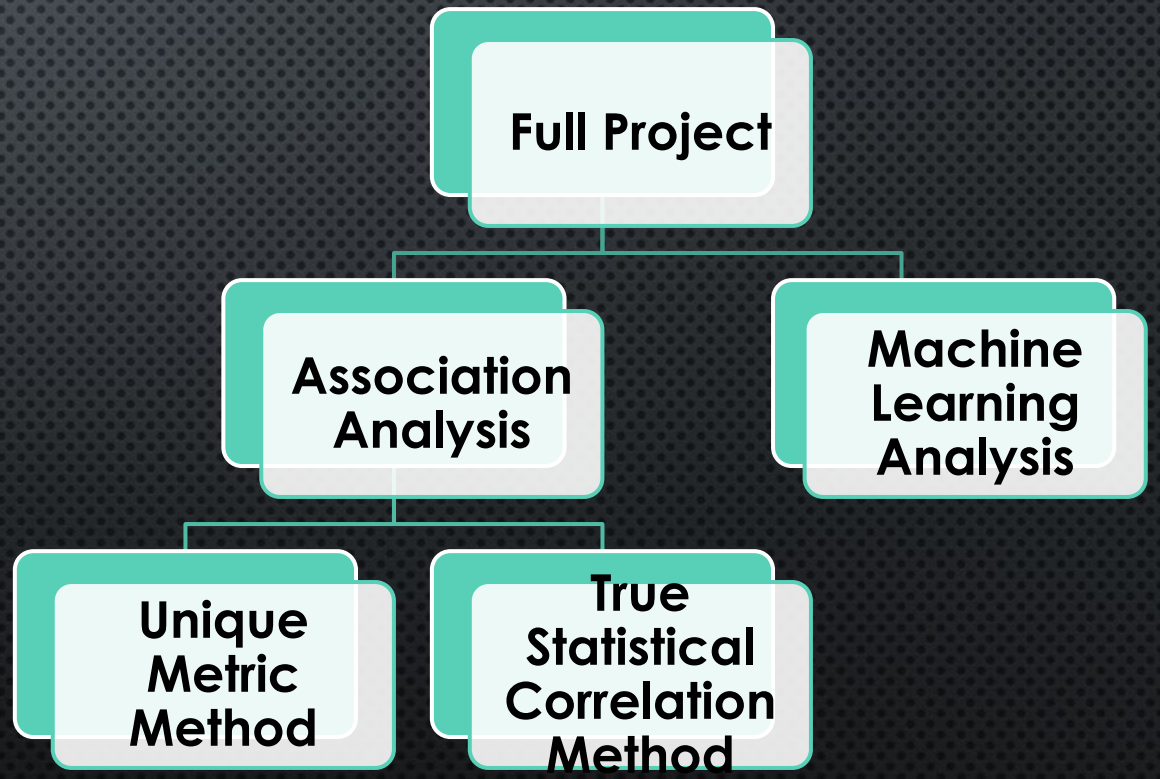
GREENWICH HR PROJECT

JAKE ATLAS | SRIVIDYA GANAPATHI | DANIEL HALTEH | ADI TYAGI

"BRYCE CANYON"

PROJECT OVERVIEW – BRYCE CANYON

1. Correlation between ads, skills in ads and stock performance in next 90 days (by October 25)
2. Predictive models for stock price in the next 90 days based on ads and any other possible external public data set (first model by November 15)
3. Based on ads predictive models for stock outperformance in the next 90 days (end of the quarter)



ASSOCIATION ANALYSIS – UNIQUE METRIC METHOD

$$\overline{\boldsymbol{m}_c} = avg\left(\boldsymbol{m}_{c_j}\right) = avg\left[\left(\overline{\boldsymbol{x}_{j,c}^{*stock}} - \overline{\boldsymbol{x}_{j,c}^{0stock}}\right) - \left(\overline{\boldsymbol{x}_{j,c}^{*market}} - \overline{\boldsymbol{x}_{j,c}^{0market}}\right)\right]$$

$\boldsymbol{x}_{j,c}^{*stock}$ -- Vector of 90-day percent changes in company stock price in the 90 days following the job posting

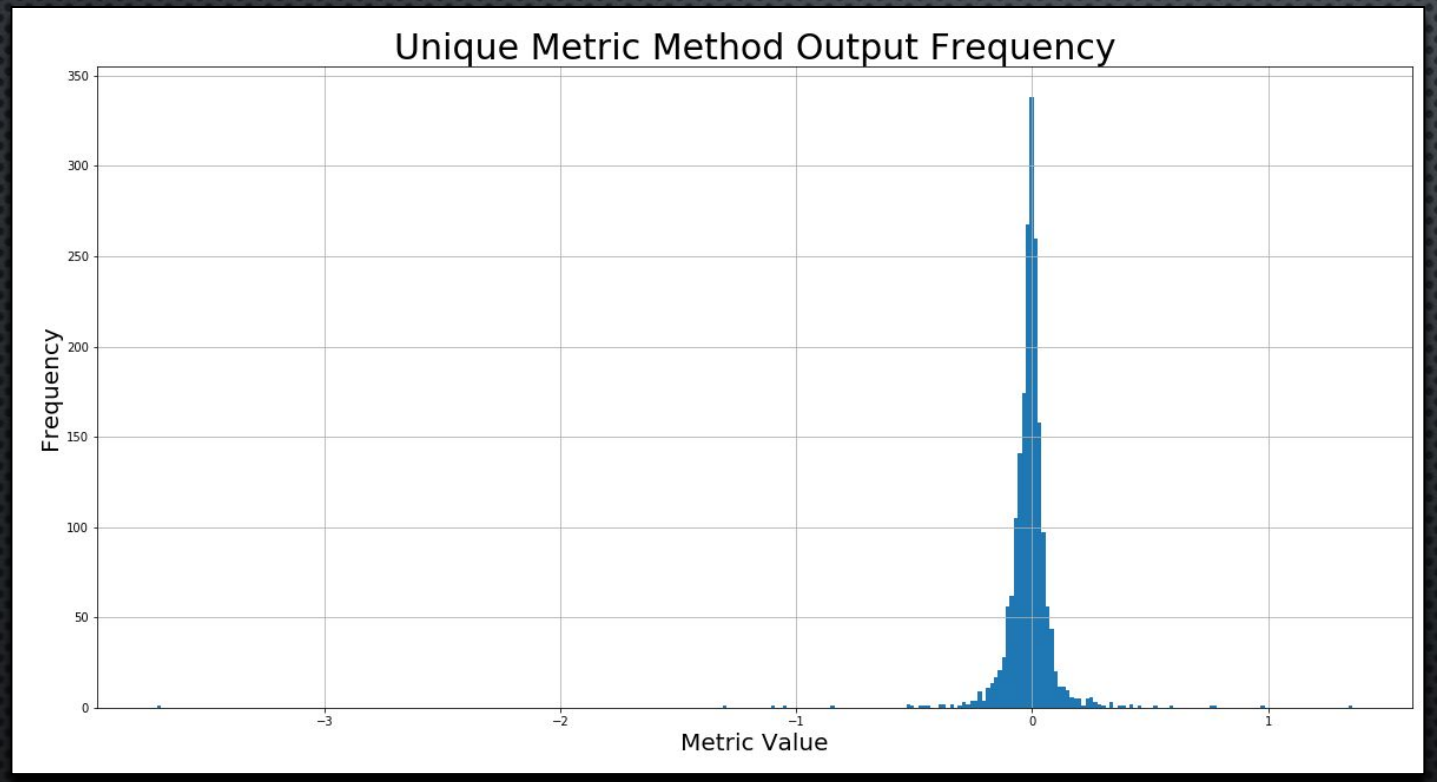
$\boldsymbol{x}_{j,c}^{0stock}$ -- Vector of 90-day percent changes in company stock price in the 90 days leading up to the job posting

$\boldsymbol{x}_{j,c}^{*market}$ -- Vector of 90-day percent changes in market index price in the 90 days following the job posting

$\boldsymbol{x}_{j,c}^{0market}$ -- Vector of 90-day percent changes in market index price in the 90 days leading up to the job posting

j is the set of jobs posted companies are denoted by c

ASSOCIATION ANALYSIS – UNIQUE METRIC METHOD



Relatively 0-centered normal distribution with small tails

Standard deviation is 0.13

Outliers exist and are consistently penny-stocks

ASSOCIATION ANALYSIS – CORRELATION METHOD

DEFINE A VECTOR FOR EACH COMPANY, CONTAINING ENTRIES FOR EACH JOB POSTING:

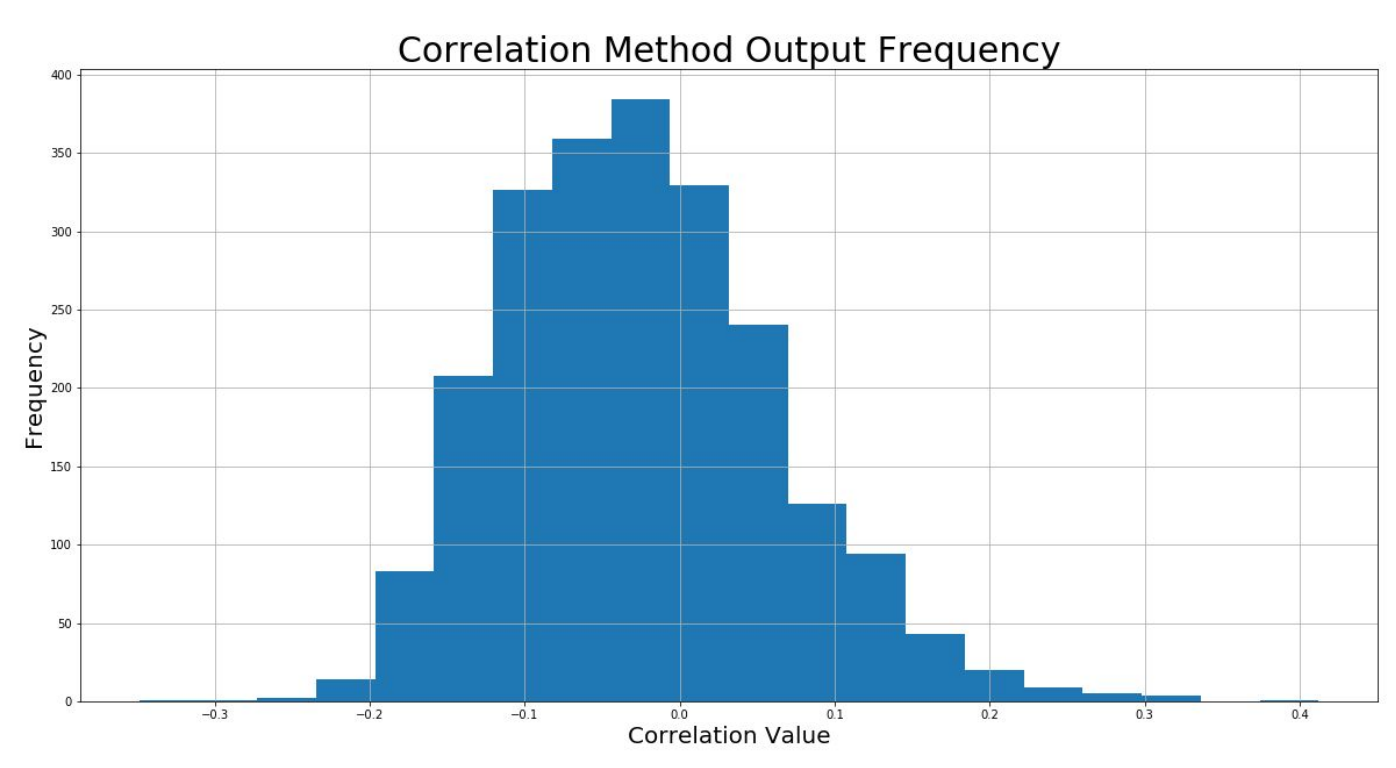
$$\boldsymbol{L_{1_c}} = vector\ of\ \#\ of\ jobs\ posted\ each\ day$$

- $$\boldsymbol{L_{2_c}} = \begin{pmatrix} \%\ change\ in\ the\ stock's\ closing\ price \\ 90\ days\ after\ the\ job\ posting \end{pmatrix} - \begin{pmatrix} average\ \%\ change\ in\ the\ market\ indices' \\ closing\ prices\ 90\ days\ after\ the\ job\ posting \end{pmatrix}$$

THEN:

$$\mathrm{COR}(\boldsymbol{L_{1_c}}, \boldsymbol{L_{2_c}})\ provides\ the\ final\ metric\ and\ p-value\ of\ the\ association$$

ASSOCIATION ANALYSIS – CORRELATION METHOD

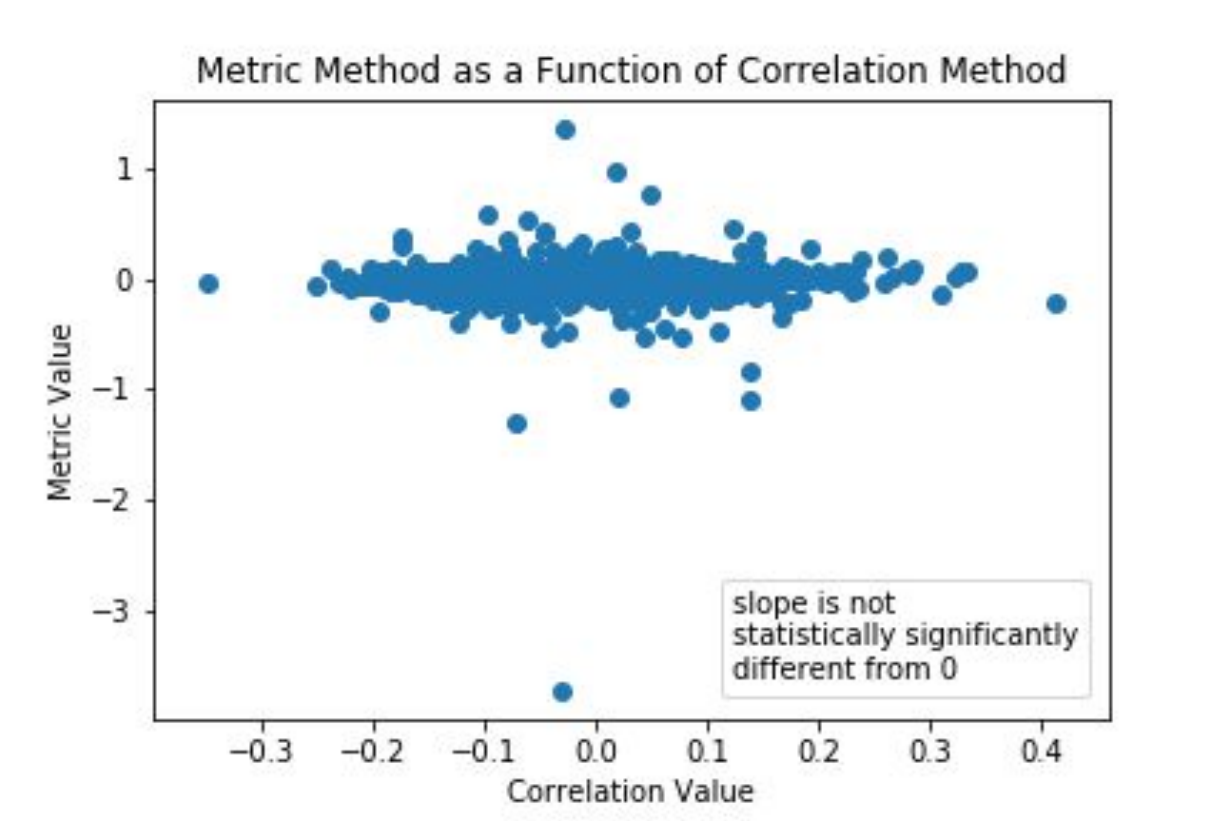


Relatively 0-centered but skewed distribution with minimal tails

Standard deviation is 0.09

Mean and median are approximately equal despite apparent skew

ASSOCIATION ANALYSIS – RECOMMENDATIONS



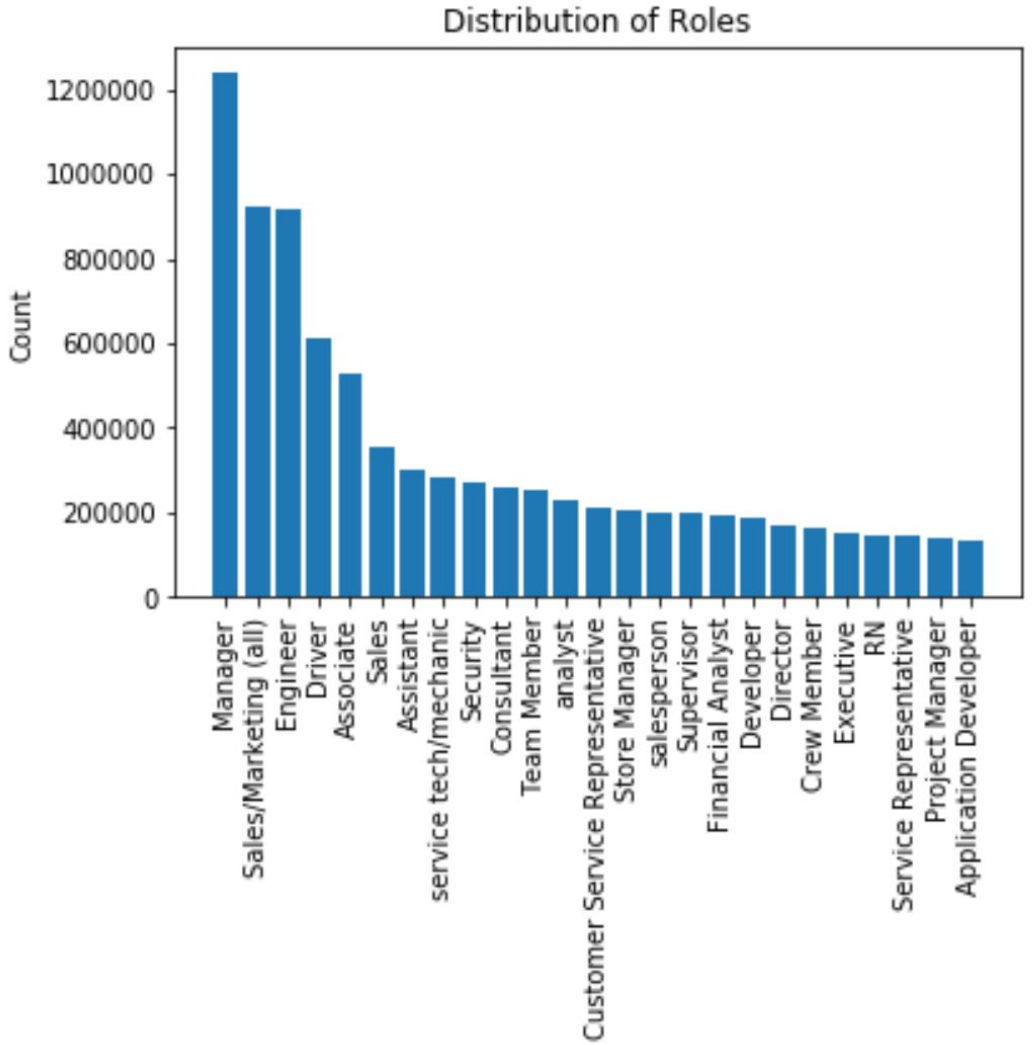
Beware of trying to pick stocks by using single numbers. This is likely not effective. Use these lists sparingly, and when using them, use both

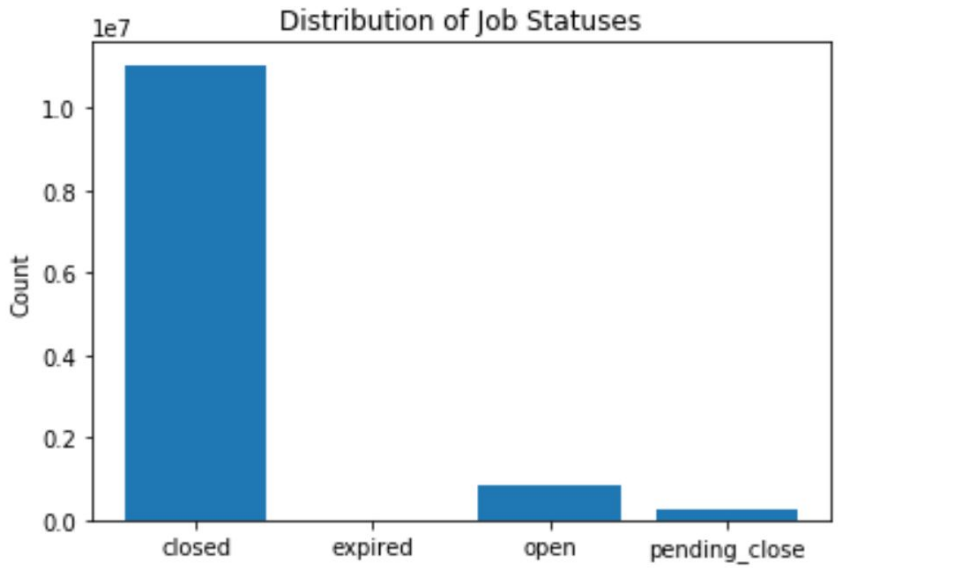
Exercise extreme caution when considering investment in penny stocks

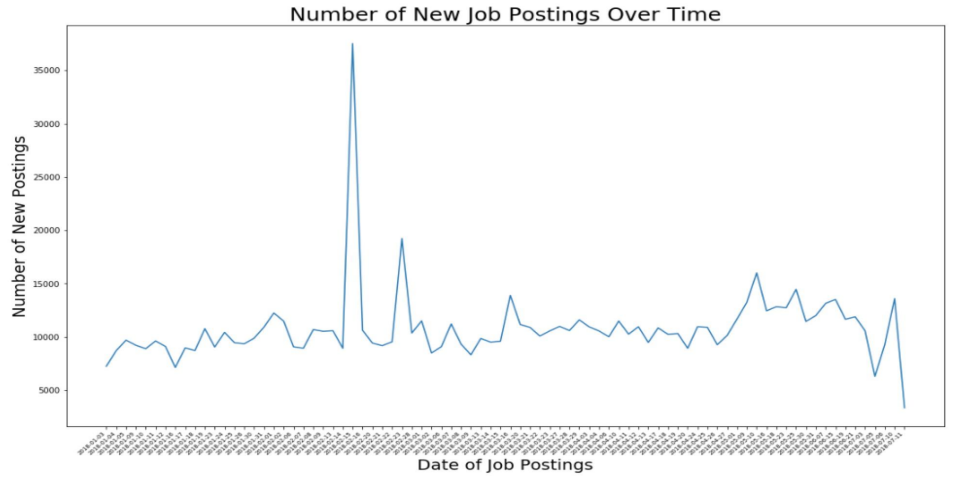
Create simulated portfolios to provide statistical significance for all values of interest, with regards to stock performance

Consider statistical testing to determine optimal window length

MACHINE LEARNING MODELING – DATA EXPLORATION







MACHINE LEARNING MODELING – INSIGHTS

Linear models allowed characterization of the ways in which a job posted by each company had different effects on the predicted stock price.
The linear regression coefficients provide a great discriminant between various types of stocks and their job posting activity, hence they were used

The first model built was multiple linear regression, while using the "pct_change_90_stock" (or percent change in stock closing price over the last 90 days) as our dependent variable . This model resulted in an R-squared value of approximately 0.58292.

The next two models implement regularization in the form of LASSO and Ridge regressions.

	LASSO	RIDGE
alpha	0.001	0.01
R-squared	0.52668	0.58292
MSE	0.01481	0.013029

Lastly, for the final validated Multiple Regression Model, the test set generated an MSE of approximately 0.013029.