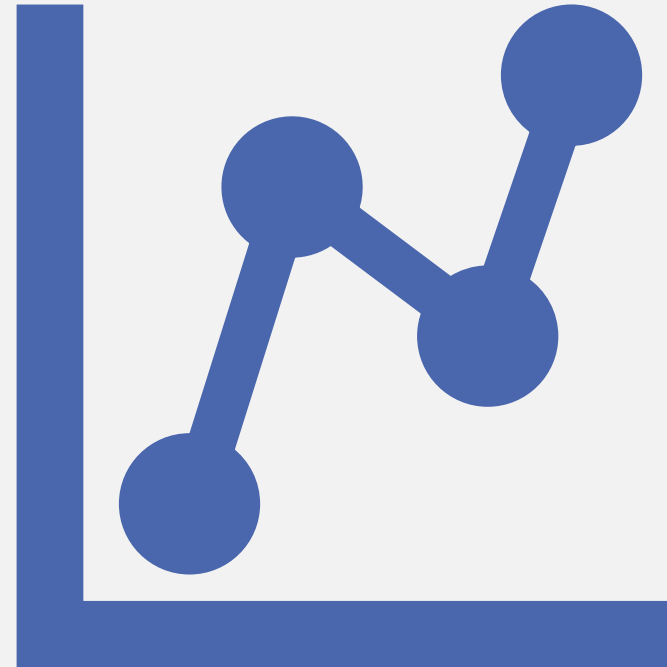# Defining Analytical Occupations

Fernando Garcia

Task

- Provide a working definition of "analytical."

- Explore analytical jobs based on O*NET data.

- Report findings using the Standard Occupational Classification (SOC) system.

Considered Approaches

Assigning Subjective Weights

- Observe database, select variables based on intuition, research, and assign subjective weights to each variable
- Calculate analytical index through summed normalized values for each occupation
- Classify by quantiles of analytical index scores

Feature Selection and Clustering

- Observe database, select variables to create an initial analytical index score
- Apply feature selection using feature selection method to refine variables for calculating analytical index
- Cluster occupations based on updated analytical index (using k-means algorithm) and assign labels to each cluster

Principle Component Analysis

- Apply principal component analysis (PCA) on dataset with all variables from relevant datasets
- Calculate weighted sum of PCA components to create analytical index scores for each job
- Cluster occupations based on analytical score index (using k-means) and assign labels to each cluster

Weighing Approaches

Assigning Subjective Weights

Pros

- Easily interpretable
- Demonstrates strong understanding of topic through required selection of relevant variables and assignment of subjective weights

Cons

- For same reasons above, requires strong understanding of topic and subjective weighting that can lead to inaccurate outcomes
- Does not utilize machine learning techniques that can make classification efficient or select relevant features

Feature Selection and Clustering

Pros

- Utilizes feature selection technique to determine relevant variables
- Utilizes machine learning clustering to label occupations as analytical
- Additionally demonstrates understanding of analytical through initial index

Cons

- Requires understanding of topic to create initial index
- Accuracy of model becomes dependent on feature selection technique

Principle Component Analysis

Pros

- Background understanding of topic not inherently necessary; reliant on all given data
- Determines most important factors from all variables in given dataset when determining index

Cons

- Complex or difficult to interpret
- Linearity of PCA can lead to misleading relationships between variables, hence inaccuracy in model

Decided Approach

- Decided upon feature selection and clustering approach
  - Good balance between leveraging domain knowledge and data driven analysis
  - Minimalizes biases by requiring good understanding of subject to decide upon variables, and then refines index based on importance of each selected variable
  - Uses machine learning techniques to classify data based on calculated index score

Defining Analytical

- Intuitively: In the context of occupations, analytical jobs require strong problem-solving skills, critical thinking, and deductive reasoning to interpret information and identify solutions.

- Analytical definition: careful examination of facts and information to discover more about something

- Common analytical skills:
  - Critical thinking
  - Research
  - Data Analysis
  - Problem-solving

Defining Analytical: O*NET Database Observation

- Upon observing O*NET Database, I decided to select variables for occupation skills, knowledge, abilities, and work activities to create my initial analytical index

Skills	Knowledge	Abilities	Work Activities
• Active Listening • Active Learning • Critical Thinking • Complex Problem Solving • Operations Analysis • Quality Control Analysis • Judgment and Decision Making • Systems Analysis • Reading Comprehension	• Mathematics • Computer and Electronics • Engineering and Technology	• Oral Comprehension • Memorization • Written Comprehension • Problem Sensitivity • Mathematical Reasoning • Perceptual Speed • Information Ordering • Deductive Reasoning • Inductive Reasoning • Visualization	• Monitoring • Operations Monitoring • Programming • Systems Analysis • Systems Evaluation • Technology Design • Active Listening • Active Learning • Critical Thinking • Quality Control Analysis • Mathematics

Normalization Approach for Analytical Index

Min Max Normalization

+ Maintains original distribution of data

+ Scales to fixed range

+ Preserves distances between relative points, and thus less affected by outliers

- New values will result in need for recalculation if min or max of data is replaced

- Limiting in nature of fixed range

Z Score Normalization

+ Data is centered around mean and standard deviation, thus classifying outliers is easy

+ Easy to compare different variables

+ Extensible; new data will not likely have huge effect on mean and standard deviation

- No fixed range; can be less intuitive for clients

- Outliers have higher weighting than in min max normalization

Ultimately, I decided to go with Z Score Normalization, as I value extensible tools and understanding how significantly elements deviate from each other relative to the mean.
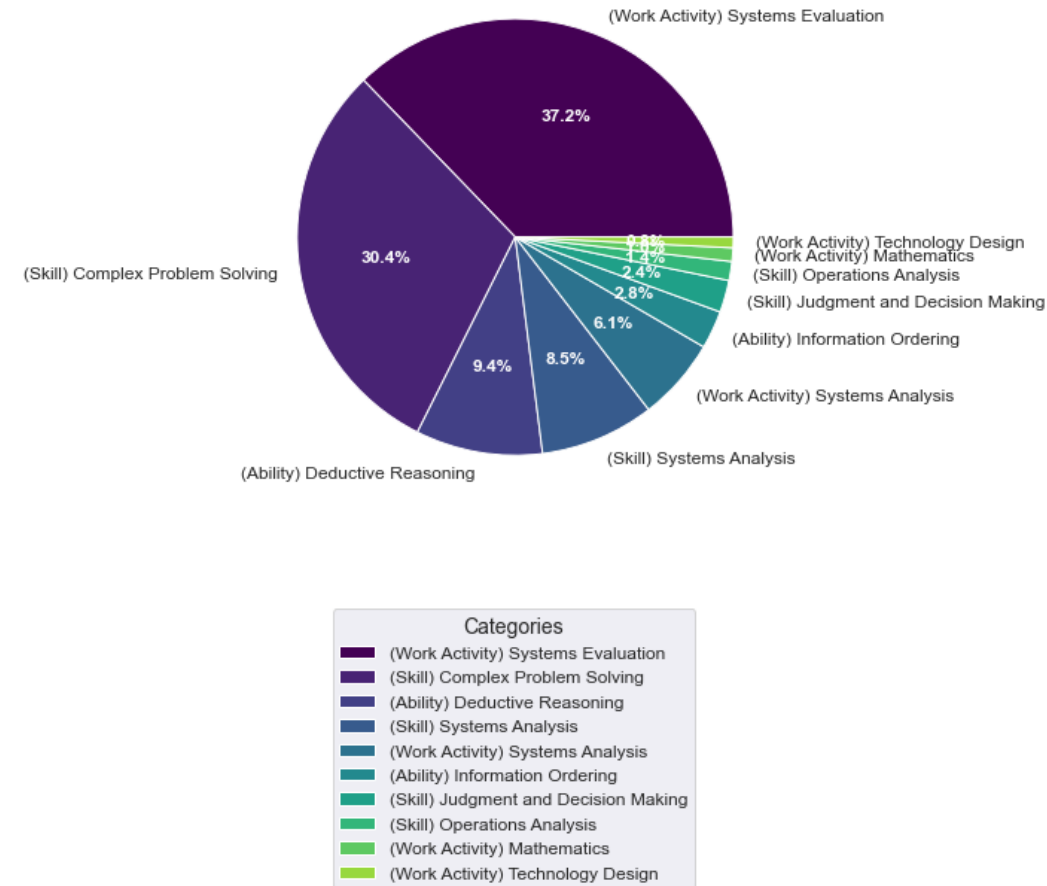
My initial index was created as follows:

Sum of all distinct element z-scores per occupation, where:

z score = ((Element_Level + Element_Importance) – Element_Mean_Combined_Value)/ (Element_Standard_Deviation)

Feature Selection

- Utilized Random Forest Regressor for predicting analytical index scores, and then feature selection based on important features
  - Hyperparameter tuning of model using Randomized Search
  - The R^2 value, used to measure accuracy of model, returned 0.97
  - Was skeptical of omitting originally selected features
    - Complex problem solving, systems evaluation, and deductive reasoning were most important features in predicting the target value
    - Rest of variables were only slightly influential in predicting target value, thus took the top 23 features (importance value > 0.002)
  - Updated analytical index score to account for newly selected features.

Top Feature Importance for Predicting Analytical Jobs

(Work Activity) Systems Evaluation

37.2%

(Work Activity) Technology Design
(Work Activity) Mathematics
(Skill) Operations Analysis

2.4%

(Skill) Judgment and Decision Making

2.8%

(Ability) Information Ordering

6.1%

(Skill) Complex Problem Solving

30.4%

9.4%   8.5%

(Work Activity) Systems Analysis

(Skill) Systems Analysis

(Ability) Deductive Reasoning

Categories
- (Work Activity) Systems Evaluation
- (Skill) Complex Problem Solving
- (Ability) Deductive Reasoning
- (Skill) Systems Analysis
- (Work Activity) Systems Analysis
- (Ability) Information Ordering
- (Skill) Judgment and Decision Making
- (Skill) Operations Analysis
- (Work Activity) Mathematics
- (Work Activity) Technology Design

Clustering Data

- To classify the data, scaled new analytical index score and fit a K-Means clustering model with three clusters, assigning the following labels:
  - Low analytical
  - Moderately analytical
  - Highly analytical
- Assigned labels to sorted cluster centers to accurately label occupations
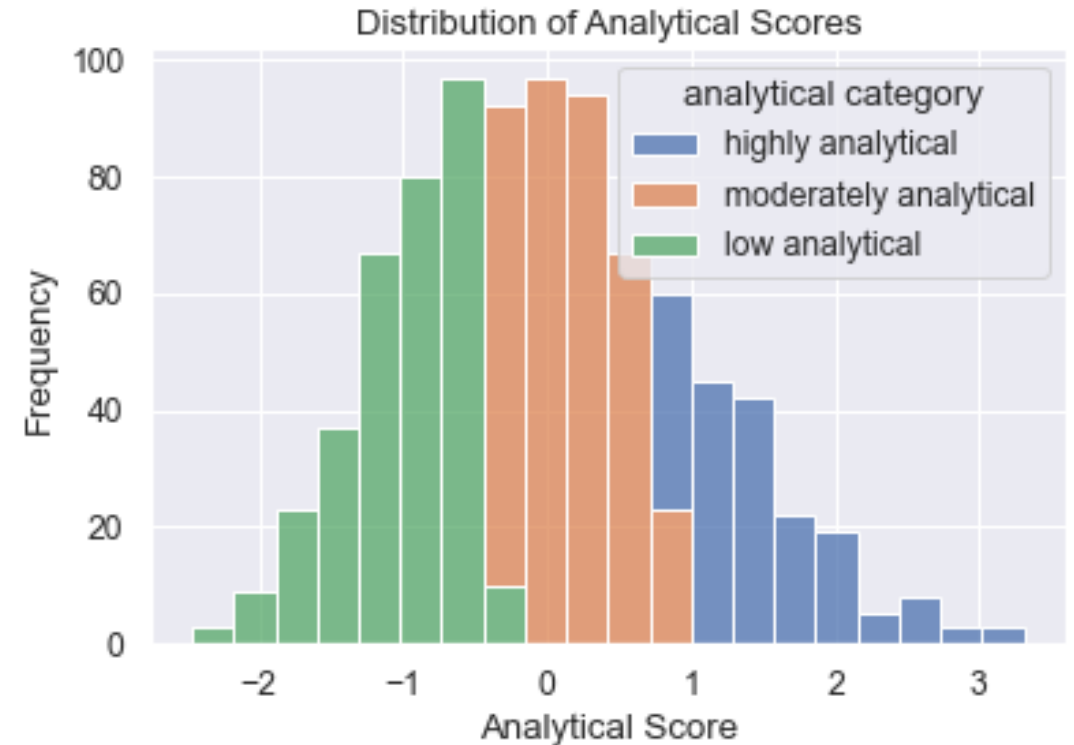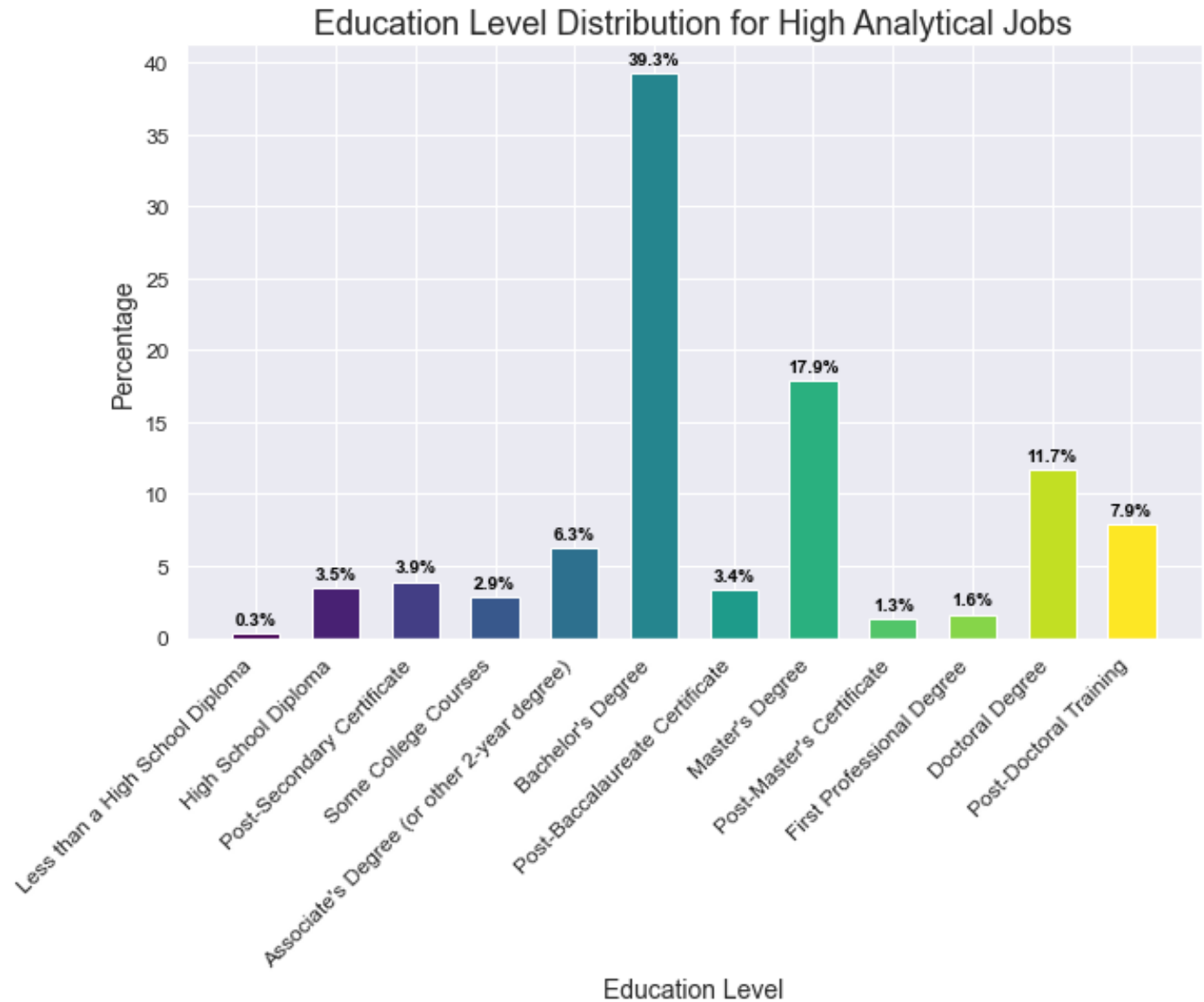
Results

- As I had intuitively assumed, top, highly analytical jobs were all scientist or engineering positions

- Bottom, low analytical jobs were mainly common labor jobs

- I was surprised to not see more Computer Science or Analyst positions make the top of the list

- Additionally surprised to see agricultural sorters score so low

  - Personal experience as a package handler and sorter proved to require memorization, judgment and decision making, as well as active listening
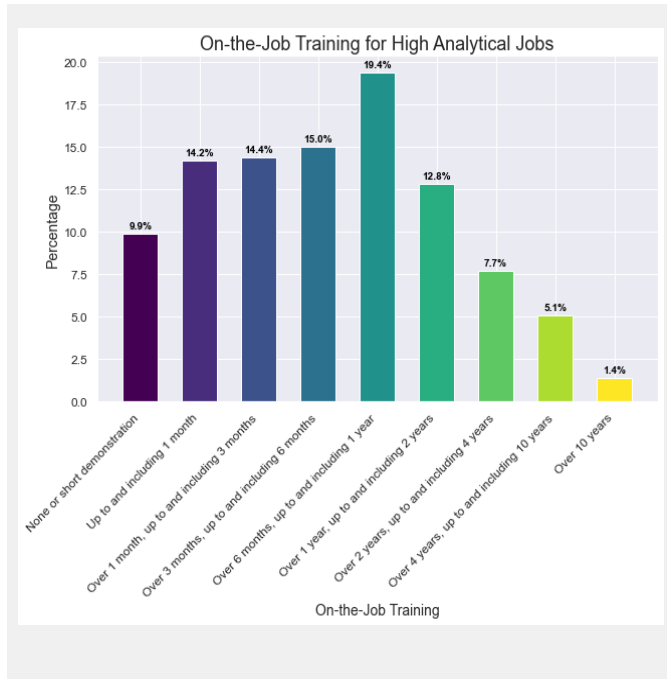

Top and Bottom 10 Analytical Jobs
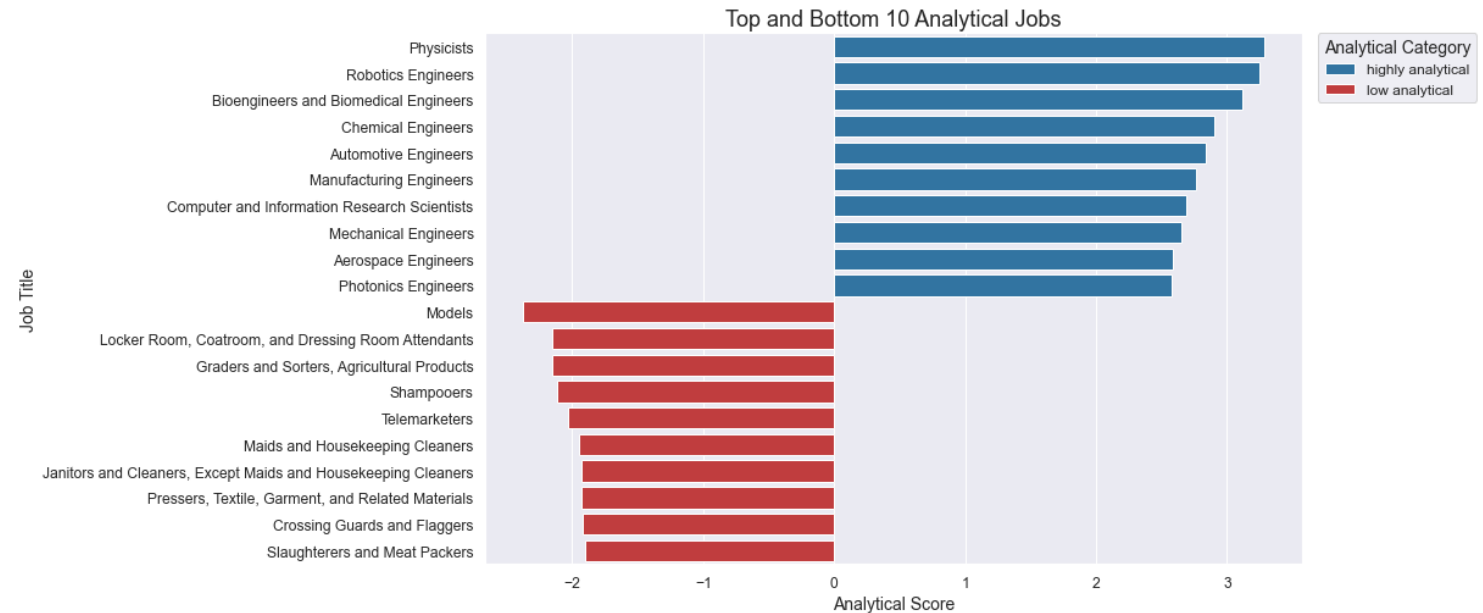
Highly Analytical Jobs: Education Level

Most highly analytical jobs require an education of a bachelor's degree or higher.

Education Level Distribution for High Analytical Jobs

Education Level	Percentage
Less than a High School Diploma	0.3%
High School Diploma	3.5%
Post-Secondary Certificate	3.9%
Some College Courses	2.9%
Associate's Degree (or other 2-year degree)	6.3%
Bachelor's Degree	39.3%
Post-Baccalaureate Certificate	3.4%
Master's Degree	17.9%
Post-Master's Certificate	1.3%
First Professional Degree	1.6%
Doctoral Degree	11.7%
Post-Doctoral Training	7.9%

Additional Insights

6m-1 Year on the job training **19.4%**

On-the-Job Training for High Analytical Jobs

Most highly analytical jobs include some form of on-the-job training for up to a year.

2-4 Years of Related Work **23.4%**

Related Work Experience Distribution for High Analytical Jobs
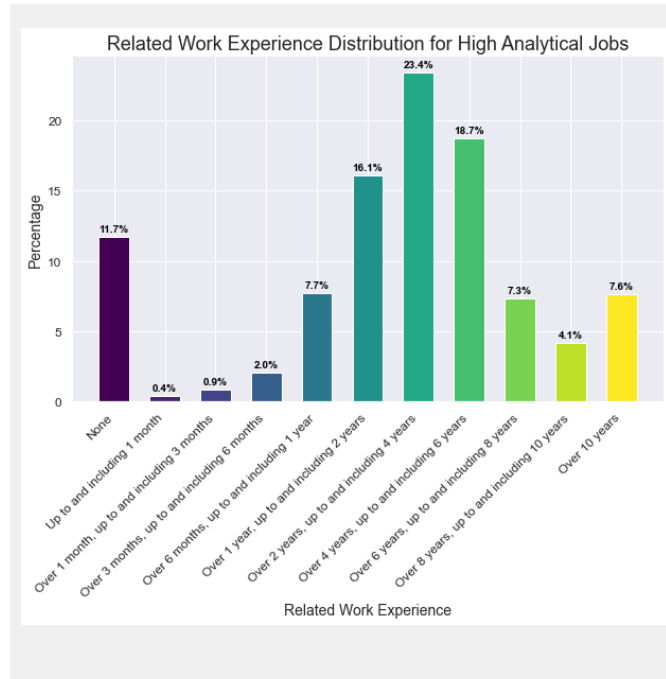
Most highly analytical occupations require over 2 years of related work experience.

No on-site or in-plant training **20.7%**

On-Site or In-Plant Training for High Analytical Jobs
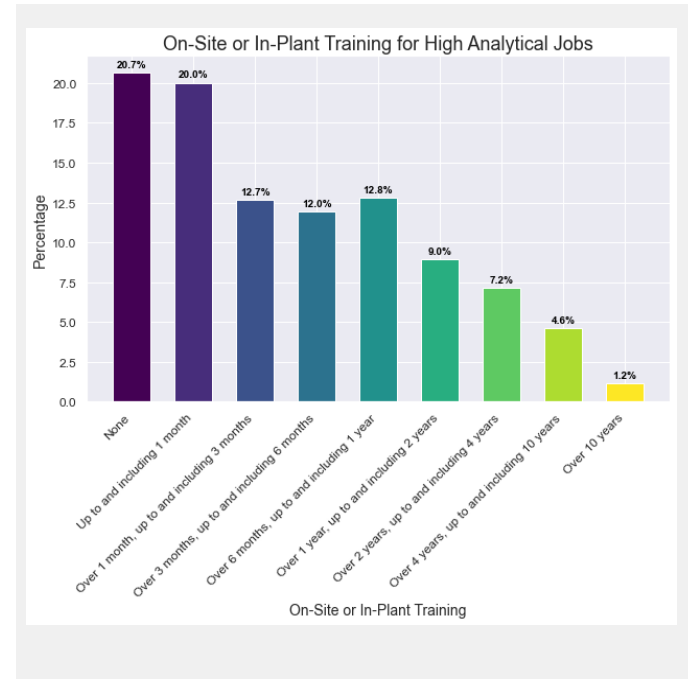
Most highly analytical occupations do not include on-site or in-plant training.

Insights (Cont.)

- Clients can better understand the necessary skills, abilities, knowledge, job responsibilities, and qualifications for the highest analytical STEM jobs in the labor market.

- Rising concerns:
  - Are there any relevant features that were not considered?
  - How subjective is defining "analytical"?

- Recommended Action:
  - Test for more input features that could potentially help better define analytical, as well as test other machine learning techniques to classify data
  - Increase sample size for more accurate results

Project Overview

Summary:

Created analytical index using selected variables to classify occupations in the O*NET database

Utilized feature selection as well as clustering through machine learning techniques

Strengths:

Comprehensive balance between data-driven techniques and domain knowledge of "analytical" to complete task

Accurate both technically and intuitively

Improvements:

More research for variable selection is necessary

Consider analyzing public job posting data (or web scraping) for more informed variable selection

Consider testing a variety of different models to optimize target label prediction