Answer to Question 1:

All Variables are Normalized and Using Hartigan-Wong algorithm

Clusters: (in line with Python graph titled "For best K (4), number of points in each cluster")

1: 50% Hired, 50% Not Hired, Highest Amount of Points in Cost\_Estimate\_Cents, Highest number points in Num\_Reviews, Low and same number of points across clusters for Result\_Position

2: 20% Hired, 80% Not Hired, Second to the highest position in Result\_Position, low and same number of points across clusters for Result\_Position

3: Low number of points for Cost\_Estimate\_Cents and Num\_Reviews

4: Second to the highest amount of points in Cost\_Estimate\_Cents, lowest amount of points in Num\_Reviews, Low and same number of points across clusters for Result\_Position

Based on these clusters:

Cluster 1: Those who were hired were correlated with the cost they offered and the highest amount of reviews.

Cluster 2: Those who were mainly not hired were somewhat correlated to result position

Type of pros that customers are interested in:

These groups had the highest amount of data points per cluster (cluster 1 and cluster 2). Pros that were liked were associated within each cluster with those that do not cost that much and are rated high. This is in line with how people approach Yelp/food. This is the suggested way the search result ranking is to be powered: cost and number of reviews are to have highest positive weight when deciding which pro goes to the top of the search result list. This is in line with correlation coefficient results.

Type of pros that customers are NOT interested in:

Those who are not hired though are usually neglected in the search result page. Those that do not get seen as much/ignored will not get hired. In order for them to increase in their ranking, pros may need to lower their cost and/or increase their job quality/customer engagement to increase in their credibility and increase number of reviews they get.