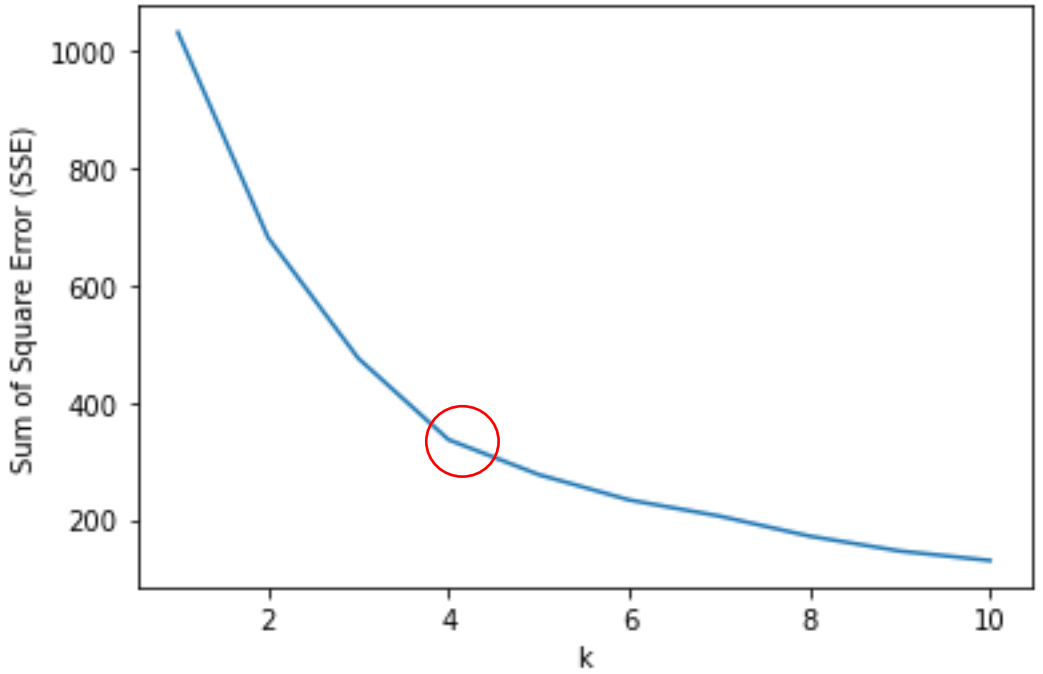
# Week 07 Report

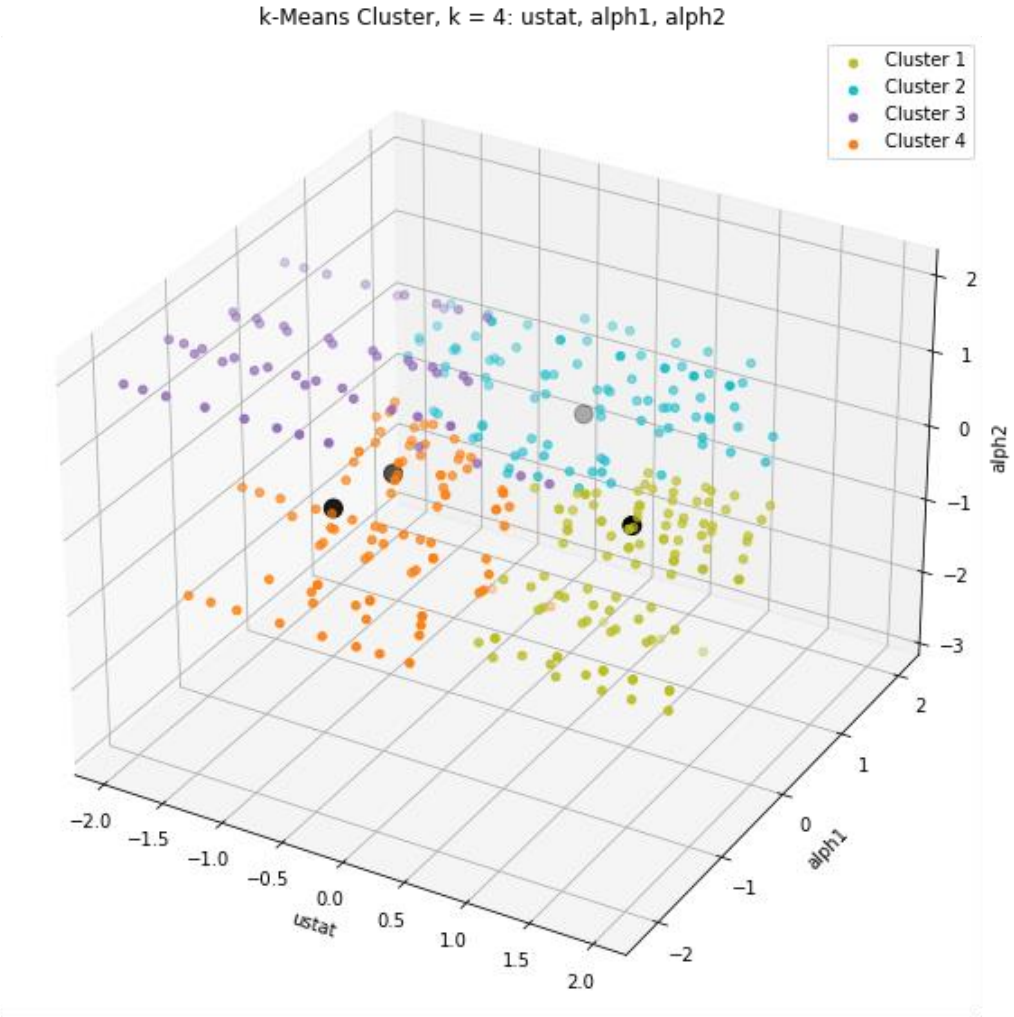
Jeffrey Li

# Previous Week

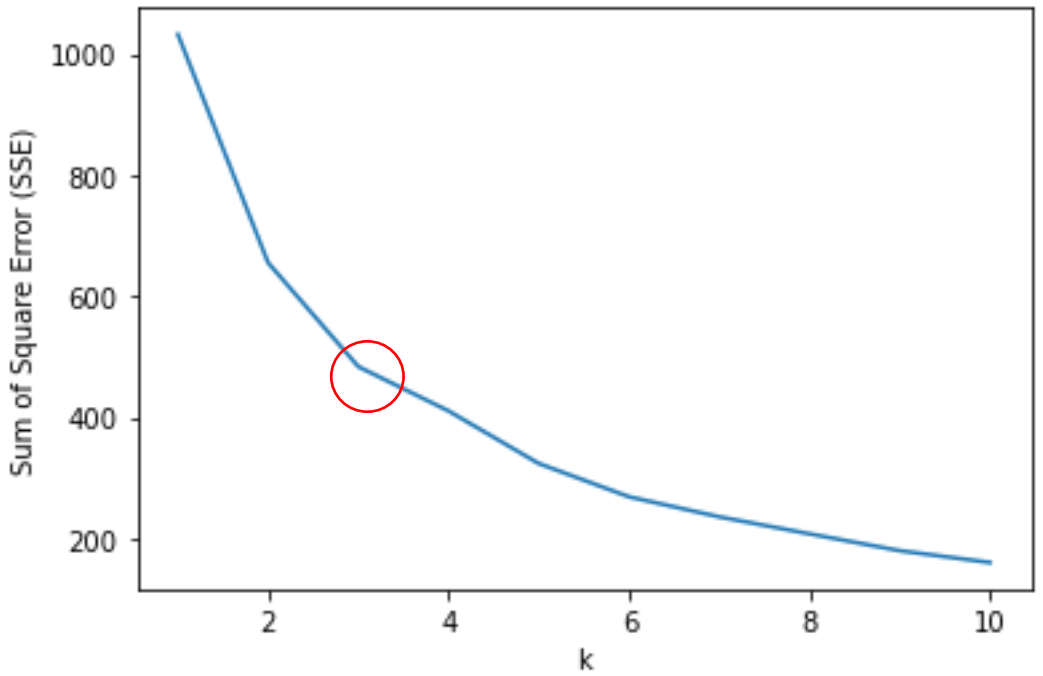
- Used power transformer function on purged data.
- Performed k-Means clustering on newly standardized data.
  - Identified k using elbow method.

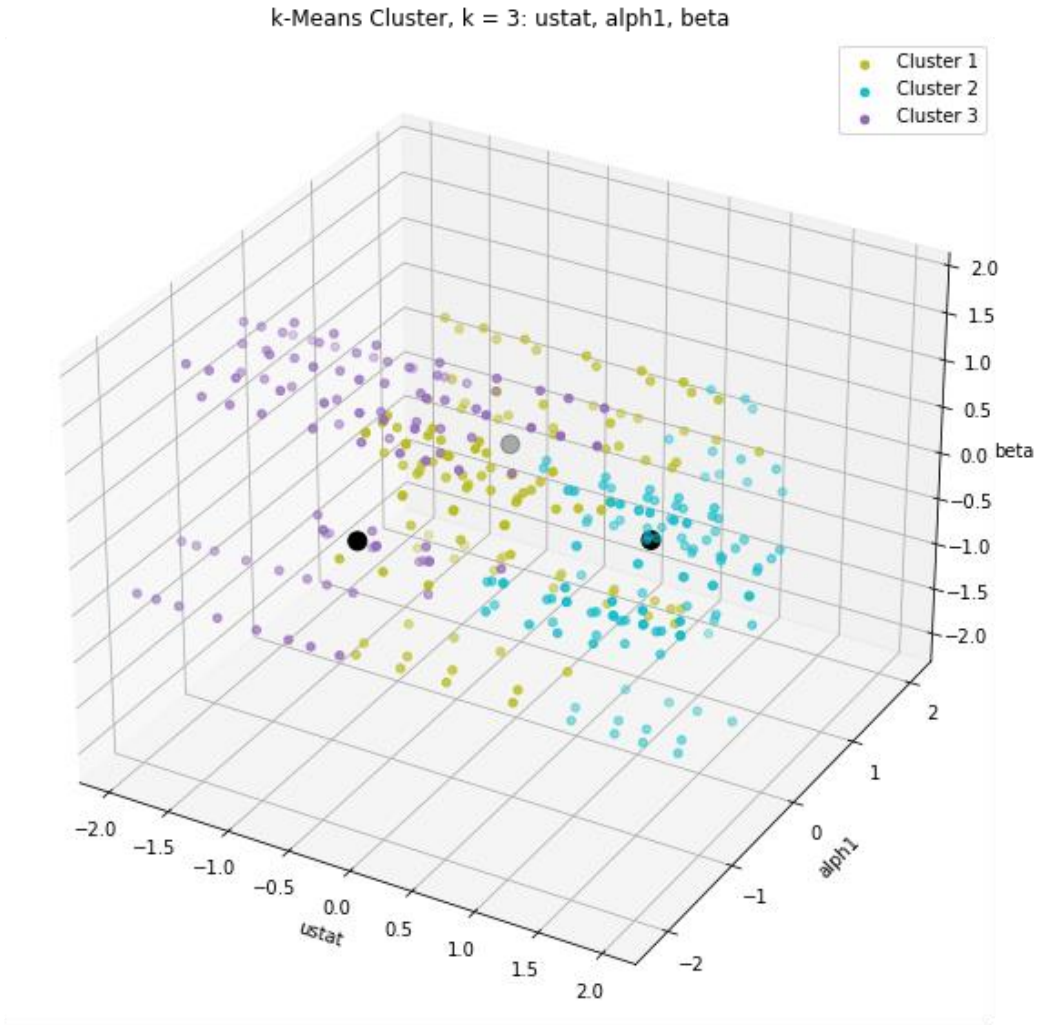
# k = 4: ustat, alph1, alph2



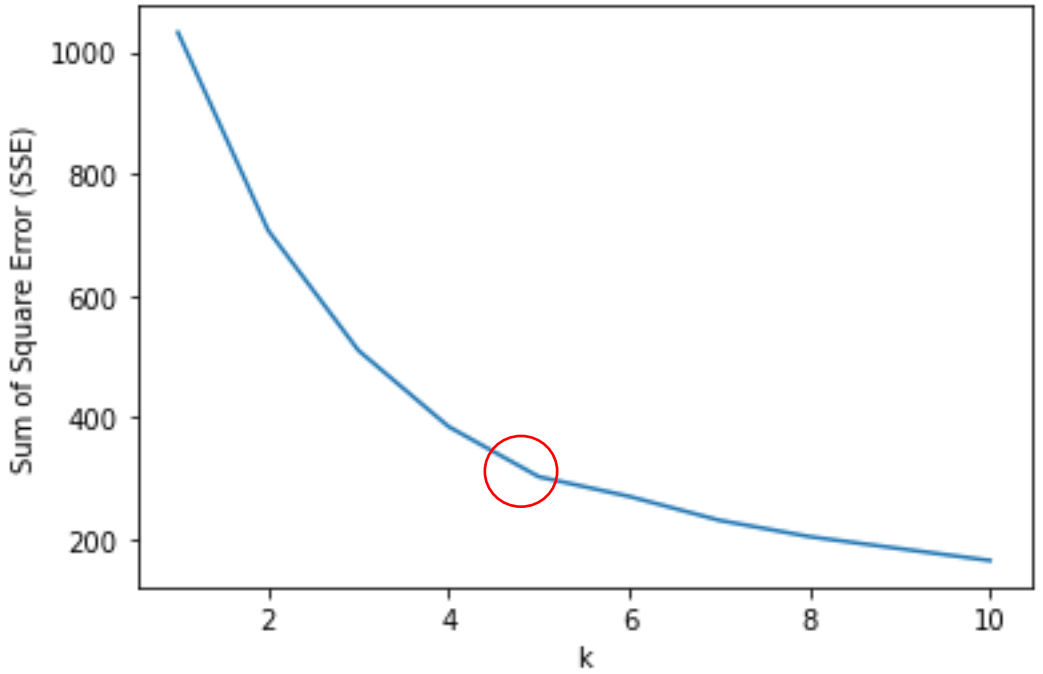



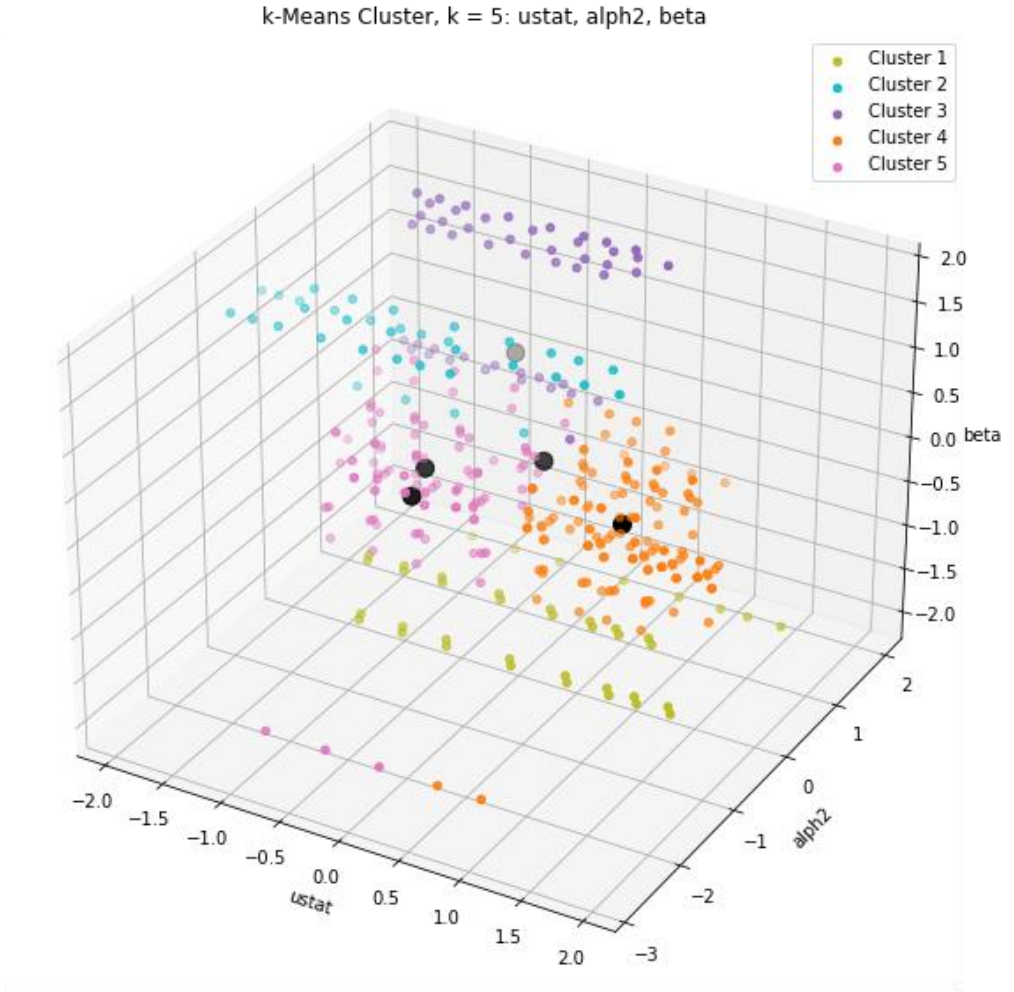
# k = 3: ustat, alph1, beta



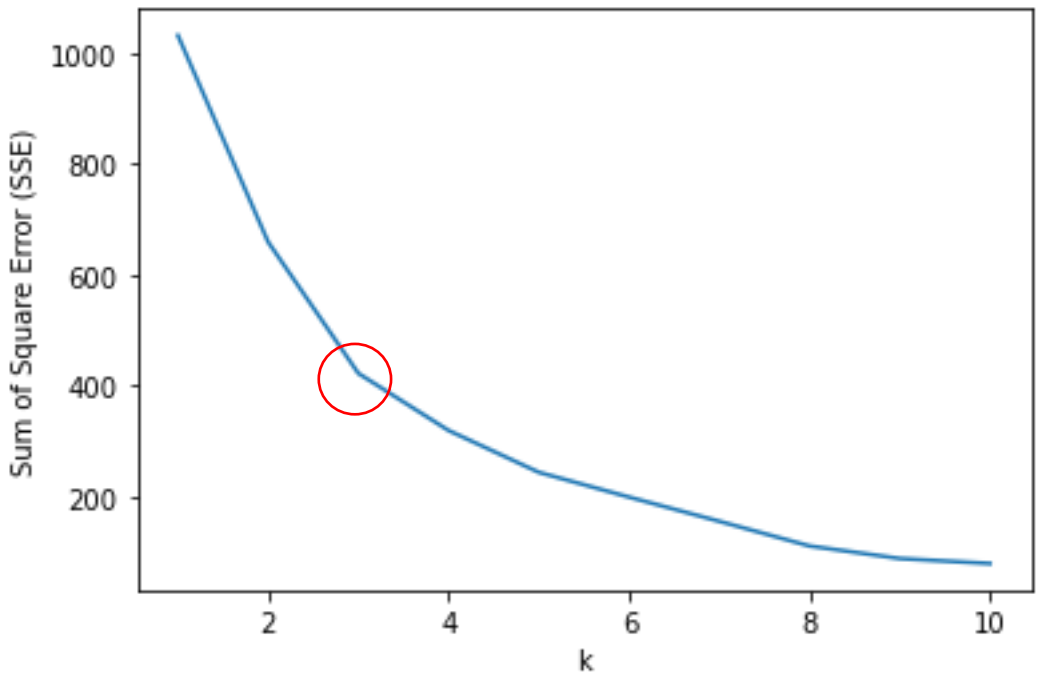



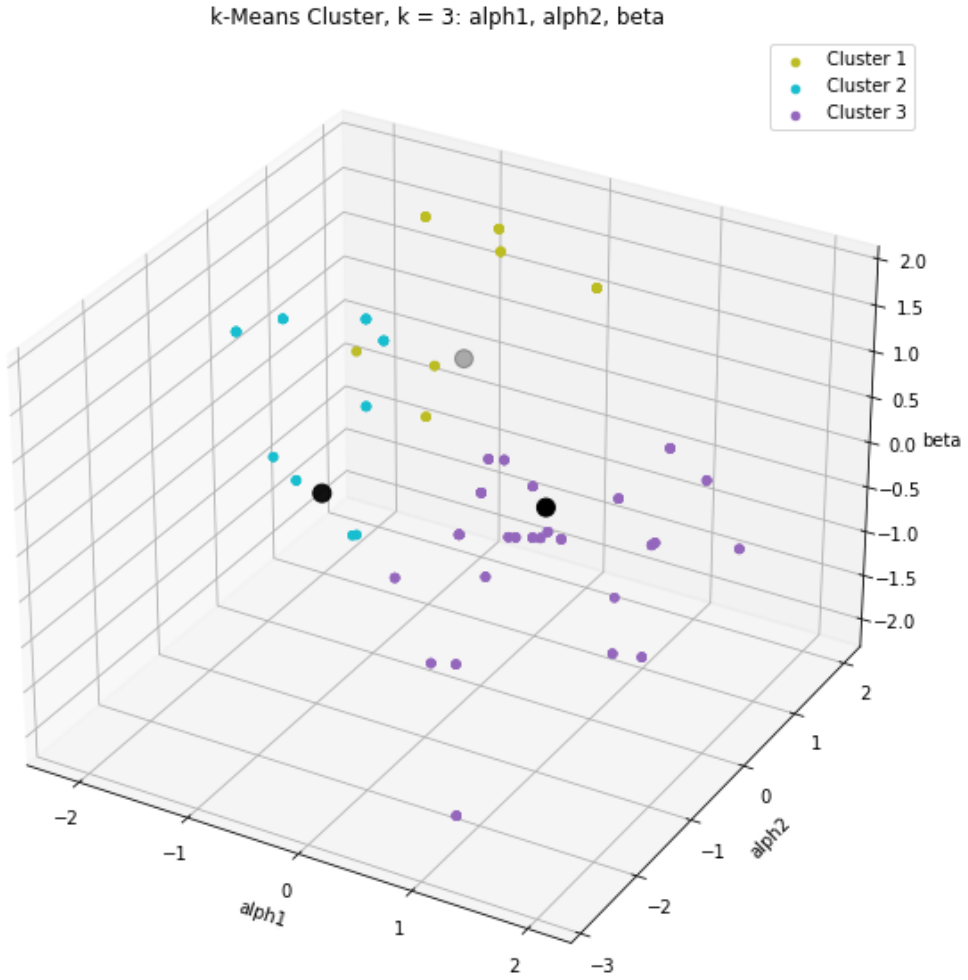
# k = 5: ustat, alph2, beta



# k = 3: alph1, alph2, beta

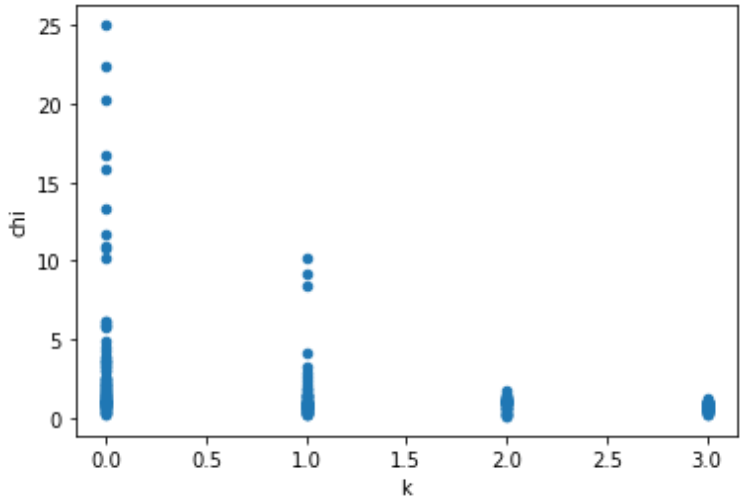


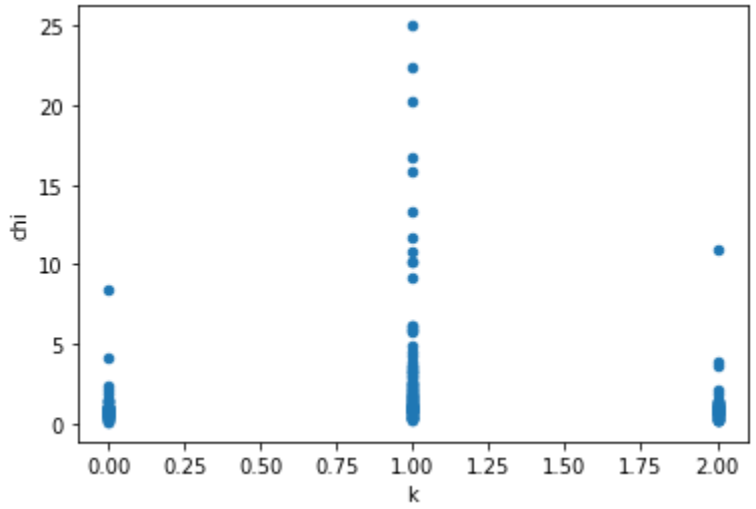
# Updates

- Analyzed clusters.
  - Reduced k to k =2 and analyzed clusters.
  - Added new features ‘diff’ and dropped ‘alph1’ and ‘alph2.’
- Gaussian Mixture Model.

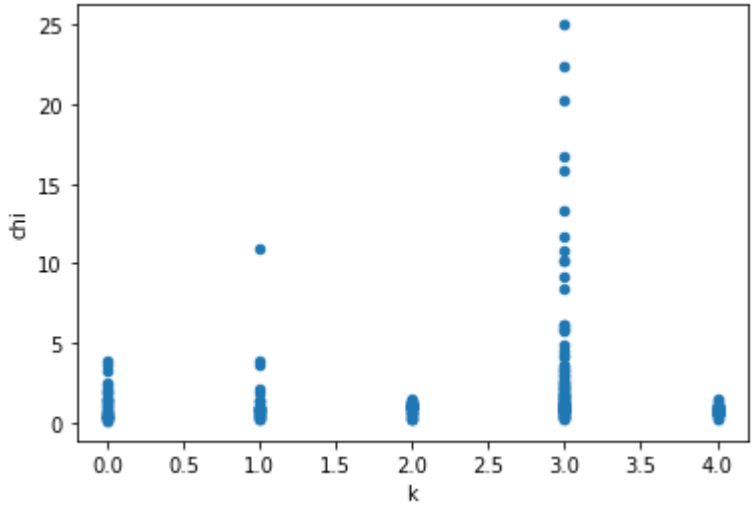
# Clusters (k Determined by Elbow Method)



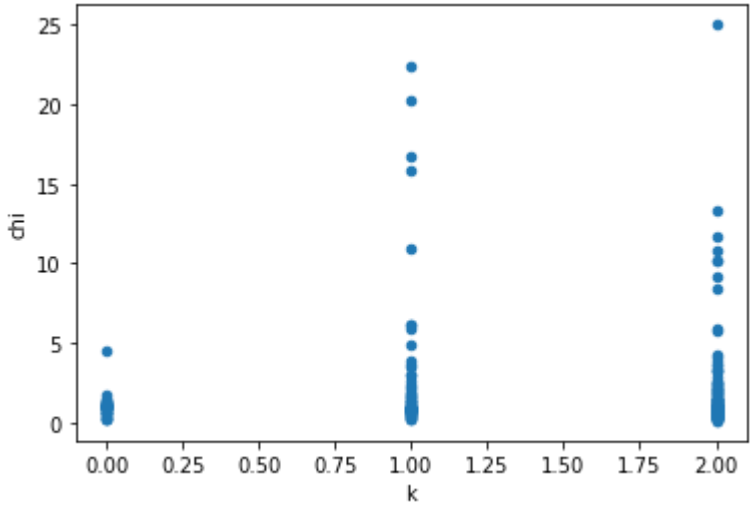

0: 107
3: 93
1: 90
2: 54




1: 125
0: 118
2: 101

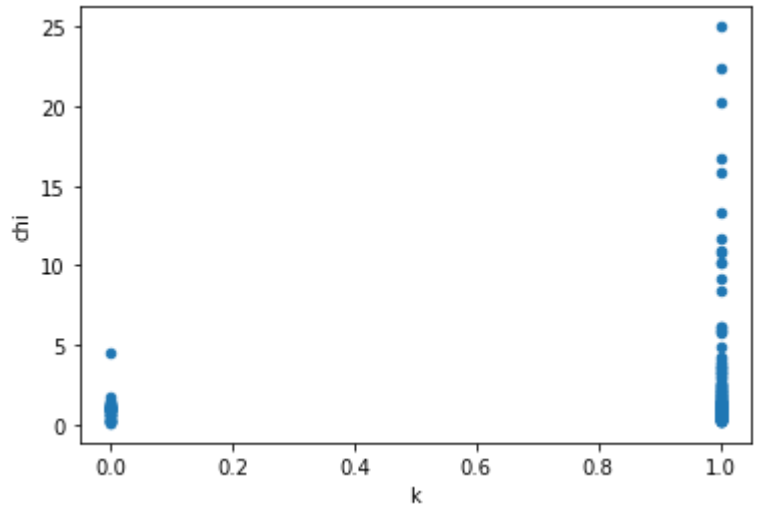



3: 121
4: 98
2: 49
0: 40
1: 36

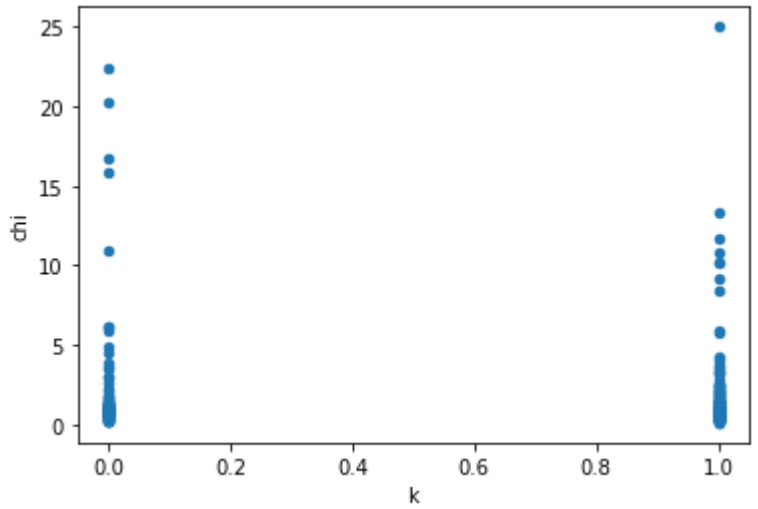



2: 211
1: 80
0: 53

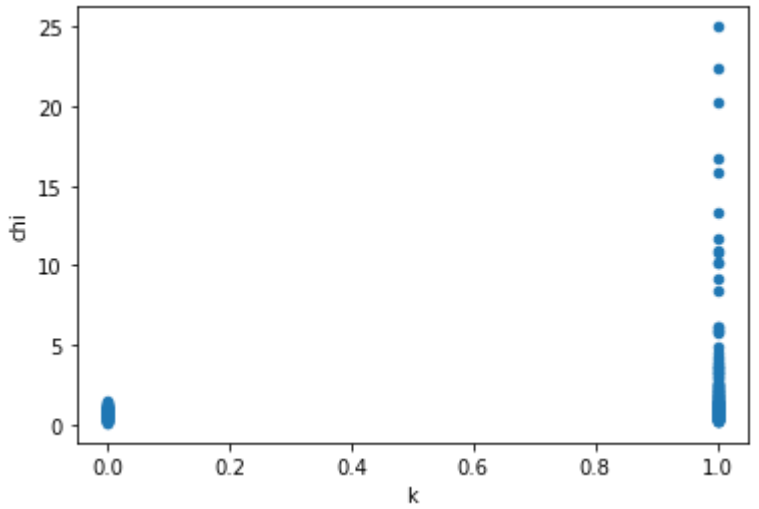
# Clusters (k = 2)



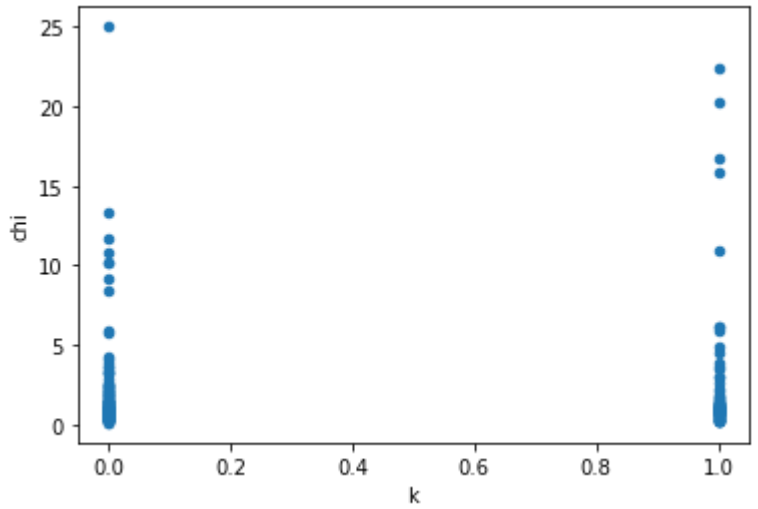

1: 287
0: 57




1: 215
0: 129

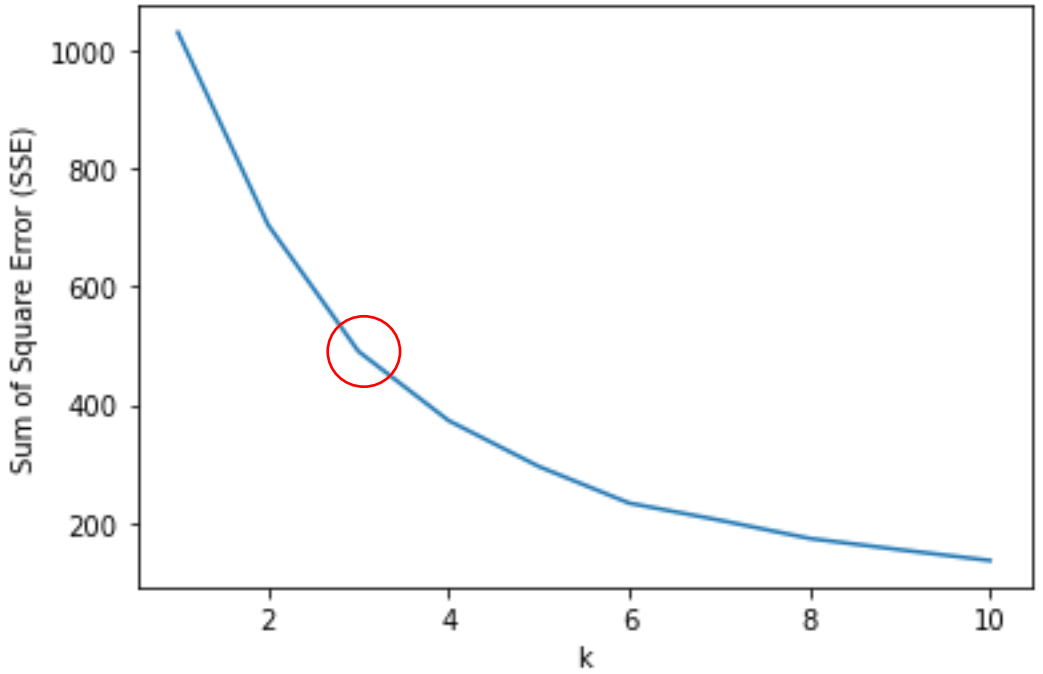


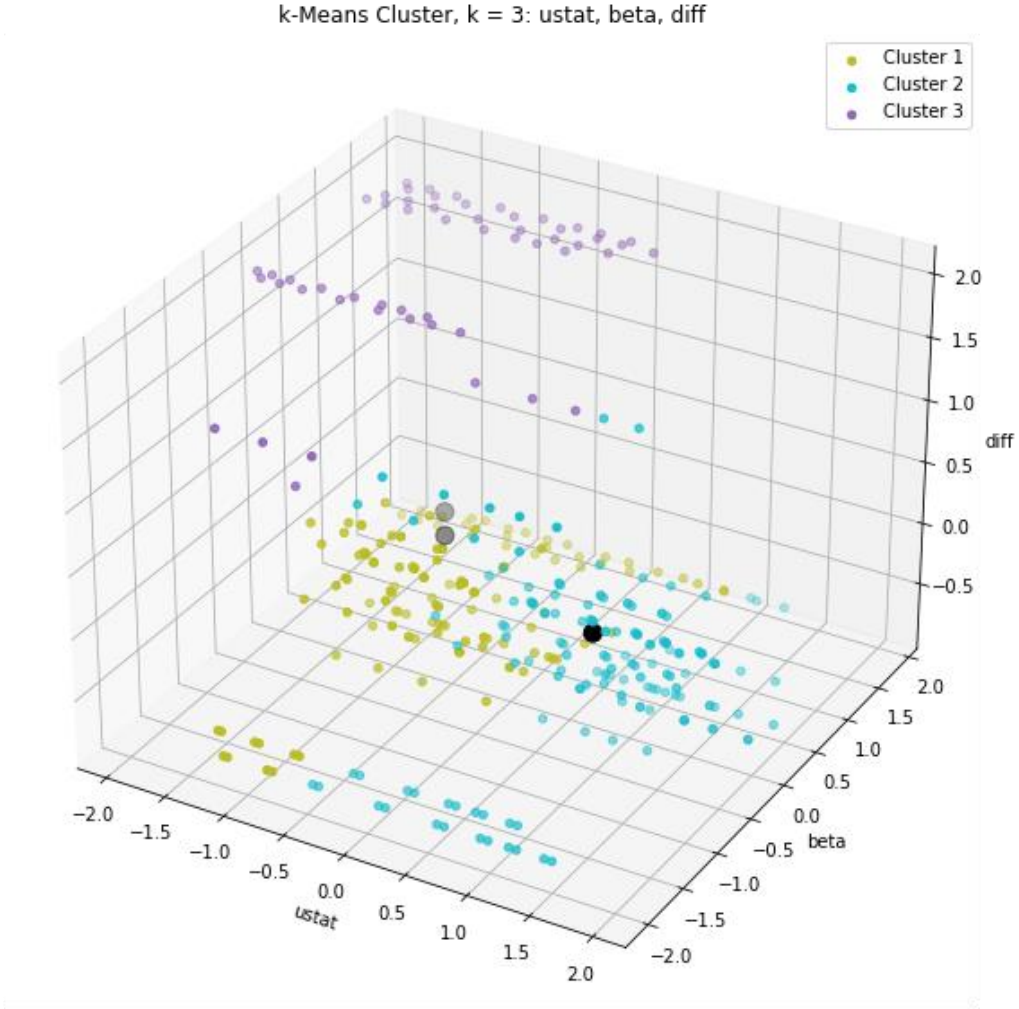

1: 201
1: 143




1: 227
1: 117

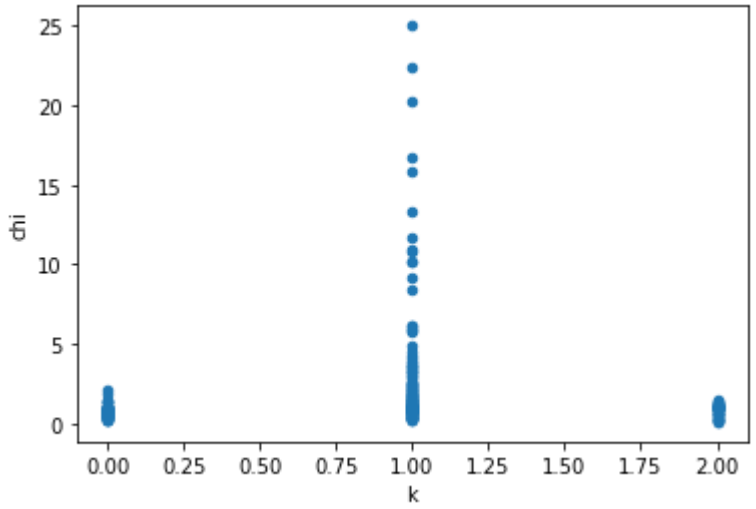
# Adding ‘Diff’ as Feature



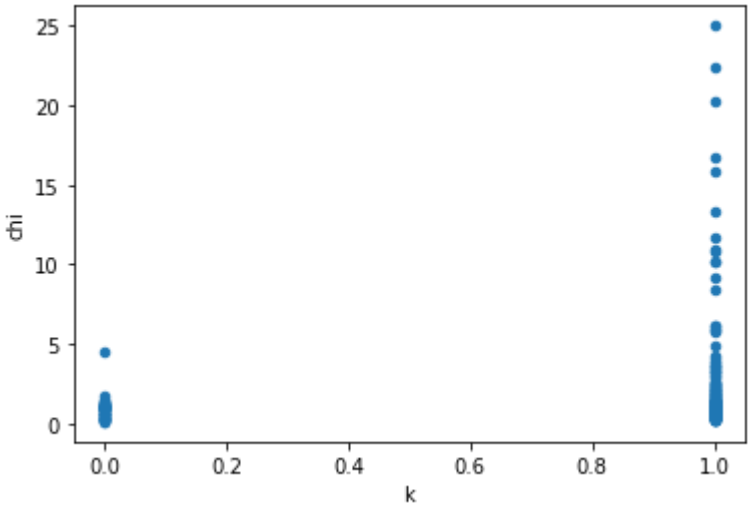
# Adding ‘Diff’ as Feature (cont.)

k = 3




1: 156
0: 133
2: 55

k = 2




1: 279
0: 65

# Gaussian Mixture Model (GMM)

- Assumes data is generated from a Gaussian distribution.
- The resulting fit is not a clustering model, but a generative probabilistic model describing the distribution of the data.
- In the simplest case, GMMs can be used for finding clusters in the same manner as k-means.
  - k-Means is a form of  hard clustering, resulting in a partition.
  - GMM is a form of soft clustering, resulting in a probability.