

Share your accomplishment!  

Creating Customer Segments

[Student Notes](#) [Project Review](#)

Does Not Meet Specifications

Functionality



SPECIFICATION

All code executes successfully and no errors are produced.

MEETS SPECIFICATION

Reviewer Comments

Be wary of the UserWarning, if you would like to remove warning add

```
import warnings
warnings.filterwarnings('ignore')
```

Responses to Project Questions



SPECIFICATION

At least one idea for what patterns might arise as components in PCA and ICA has been written.

MEETS SPECIFICATION

Reviewer Comments

Awesome idea to display a boxplot of the features to see the 'spread' / variance of the data, as Fresh definitely will arise in the first. If you wanted to go even further, PCA also deals with *correlation* between the features, so a correlation matrix could be appropriate as well.

With ICA, what do you think will be the independent types of vectors here? Hint: what are the 440 rows of this dataset?

SPECIFICATION

The variance explained by each dimension is appropriately reported and explained. A reasonable cutoff point for use has been explained.

MEETS SPECIFICATION

Reviewer Comments

Good reasonable cutoff!

Code Note: I love the biplot for PCA, as being able to visualize PCA is so awesome, if you wanted to make the plot a little bigger, just adjust the x and y limits

```
ax.set_xlim([-2.5, 1])  
ax.set_ylim([-1, 2.5])
```

Code Note 2: Great idea summing up the percent explained as it is a great thing to do to get a clear understanding. Numpy also makes this very easy without a loop

```
np.cumsum(pca.explained_variance_ratio_)
```

SPECIFICATION

Basis vectors for at least two PCA dimensions have been interpreted correctly and their significance has been discussed.

DOES NOT MEET SPECIFICATION

Reviewer Comments

You are off to a good start here. When interpreting these vectors what do the first two mean in terms of features. Therefore what does the component

`[[-0.97653685 -0.12118407 -0.06154039 -0.15236462 0.00705417 -0.06810471]]` represent in terms of the combination of features that represent the first dimension. You can also use your biplot here to confirm this! Do the same with the second component.

Note: Be careful with the comment of "The dimensions seem to represent the customer groups", as this is more of what ICA will show us. PCA in a situation where you have a bunch of independent variables, PCA helps you figure out which ones matter the most.

(<http://stats.stackexchange.com/questions/2691/making-sense-of-principal-component-analysis-eigenvectors-eigenvalues>)

SPECIFICATION

Basis vectors for at least 4 dimensions of ICA have been interpreted correctly and their significance has been discussed.

DOES NOT MEET SPECIFICATION

Reviewer Comments

Awesome job giving an idea of what each component could represent in terms of independent customers, but we also need to interpret these in terms of features. You have done this with some,

but what features are most prevalent in each vector? Also remember that the positive and negative signs do matter here, as they are considered anti-correlated with one another.

Also how could this distributor use this information we just described?

SPECIFICATION

Gaussian Mixtures and K Means have been compared. Student makes a choice which is justified based on the characteristics of the algorithms.

DOES NOT MEET SPECIFICATION

Reviewer Comments

You are off to a good start here by describing clustering and hard vs soft assignments. Just compare these two a bit more and go into the advantages and disadvantages of both. Compare them in terms of speed / complexity / structural information / boundaries. You also need to address the question of "Which algorithm is more suited for this particular dataset?" And justify your answer based on your previous analysis.

If you need a refresher, take a look at the lectures or the sklearn docs

- [Lecture](#)
- [docs](#)

SPECIFICATION

More than one choice of number of clusters has been tried out. Elements from each cluster have been sampled and interpreted.

DOES NOT MEET SPECIFICATION

Reviewer Comments

Your plot looks great, but in order to pass this section you need to experiment with a couple of different cluster numbers here and either choose which would be best in terms of a mathematical derivation with something like a [BIC](#) score or just simply plot at least two different cluster numbers and choose the best looking one for this particular dataset. Also you should sample from the clusters to get a better understanding of the distribution of the clusters, and this also may help you decide on the best cluster number. One technique is to use [PCA.inverse_transform\(\)](#) to take the centroids and use the PCA model to bring the centroids into the original dimensions. Then the centroids can be compared. Also making a bar chart of the centroids in the original dimensions could be very helpful for comparing the centroids to each other and looking for patterns.

SPECIFICATION

PCA has been used to visualize the data in two dimensions. A plot has been created which clearly shows different clusters. If clusters are not clearly visible some discussion has been made of how to improve the visualization.

MEETS SPECIFICATION

SPECIFICATION

One method has been discussed in detail and its usefulness has been discussed.

MEETS SPECIFICATION

Reviewer Comments

Great work as PCA would definitely "help identify the most selling or important product category i.e. Fresh foods"

SPECIFICATION

Some method of improving the ability to get good results from an A/B test has been proposed.

DOES NOT MEET SPECIFICATION

Reviewer Comments

Good idea for an A/B test. But can you describe what an A/B test is? Think control / test group? What will the results from the test show us? How do we implement an A/B test? You are off to a good start, a good idea here would be go through a full example start finish (with product grouping as you have said).

If you need a refresher of what an A/B test is and how to implement one, check out this link. It refers to a website A/B test, but the same terminology can be applied to any situation

- [A/B Test](#)

SPECIFICATION

Some techniques that could be used in a supervised learning analysis have been proposed.

DOES NOT MEET SPECIFICATION

Reviewer Comments

Again another good idea. Can you just go into more detail here and give an idea of how we can use the existing data we have created from one of these techniques and use that data to *predict* new future customers?

Additional Reviewer Comments

Good analysis and I really like the visuals you provided as it definitely makes the project more readable and helps with interpretations. You have good intuition in the techniques, just need to tie them into this particular dataset a bit more.

 [Download project](#)



Best practices for your project resubmission

Ben shares 5 helpful tips to get you through revising and resubmitting your project.

[▶ Watch Video](#) (3:01)



Have a question about your review? Email us at review-support@udacity.com and include the link to this review.

 [RESUBMIT PROJECT](#)

[Front-End Web Developer](#)

[Full Stack Web Developer](#)

[Data Analyst](#)

[iOS Developer](#)

[Android Developer](#)

[Intro to Programming](#)

[Tech Entrepreneur](#)

STUDENT RESOURCES

[Blog](#)
[Help & FAQ](#)
[Catalog](#)
[Veteran Programs](#)

PARTNERS & EMPLOYERS

[Georgia Tech Program](#)
[Udacity for Business](#)