MACHINE LEARNING PROJECT 3

CLUSTERING ARTWORKS

LEE KUCZEWSKI

METADATA

Transforming the Datatype

pl	given location on the planar dimensions
si	variations in height, width, area
va	the various degrees between white and black
te	variation in the fineness or coarseness of an area having a given value; includes blur
co	hue, using the repertoire of colored sensations which can be produced at equal value
or	various orientations, ranging from the vertical to the horizontal in a distinct direction
sh	a mark with a constant size can nonetheless have an infinite number of different shapes
reflection	indicates the work contains an image given back by a reflecting surface, or an image seen in a mirror or shiny surface
po	A POINT represents a location on the plane that has no theoretical length or area. This signification is independent of the size and character of the mar
li	A LINE signifies a phenomenon on the plane which has measurable length but no area. This signification is independent of the width and characteristic
ar	An AREA signifies something on the plane that has a measurable size. This signification applies to the entire area covered by the visible mark.
notes	notes, description

In order to read in the metadata, we will need to transform the datatype.

Out[30]:

	artist	country_of_origin	date	reflection	notes
0	Giorgio de Chirico	Italy	1913	False	distorted perspective, shadow, signification o...
1	Giovanni Anselmo	Italy	1967	False	hard to understand the viewpoint, sense of for...
2	Milton Avery	America	1958	False	flatish, textured shapes & specific colors-lin...
3	Gillian Avery	UK	1957	False	shapes, layers, paint handlng/texture, orienta...
4	Joseph (Jef) Banc	France	1956	False	ambiguity through abstraction, odd shape, v te...

TRANSFORM

.ASTYPE(INT) > FROM BOOLEAN TO INTEGER

Part 2 Set (True, False) values to (1, 0)

Transform Data type for columns ‘reflection’ & ‘has_text’

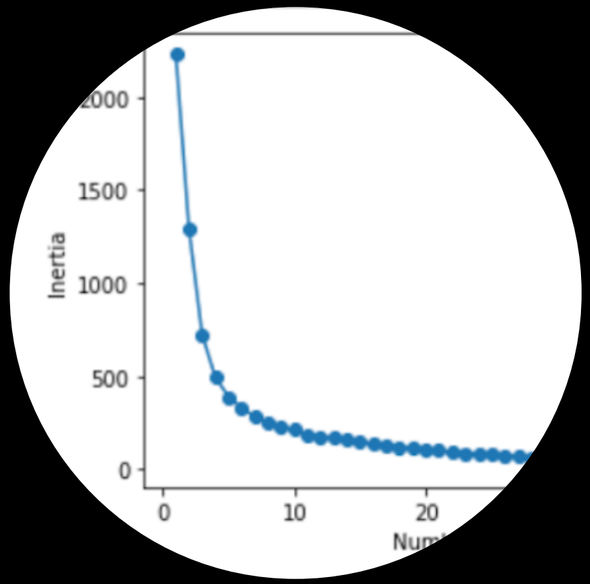
Convert data type in Featured Columns

```
In [147]: # Convert the "reflection" and "has_text" colums from True/False to 0,1
          data['reflection'] = data['reflection'].astype(int)
          data['has_text'] = data['has_text'].astype(int)
```

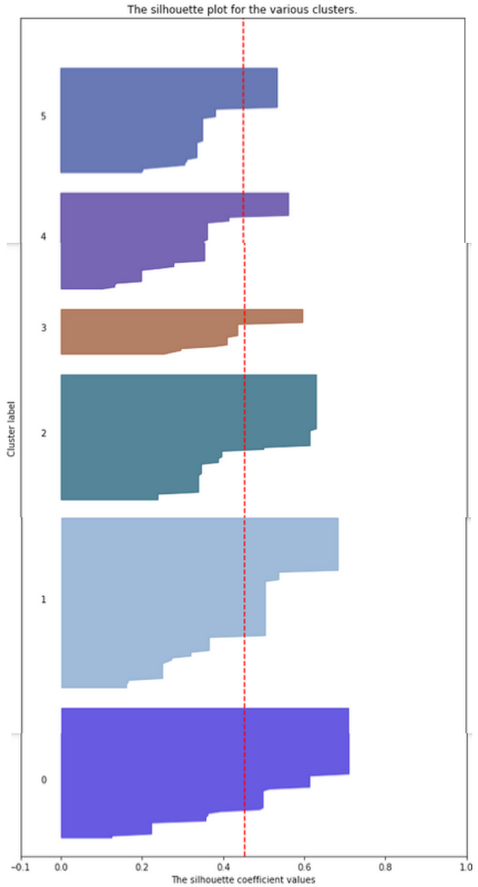
has_text	va	te	co	or	sh	reflection	po	li	ar
0	3	0	0	4	3	0	0	0	5
0	1	3	0	3	2	0	0	0	5
0	2	4	4	2	4	0	0	0	5
0	1	2	0	2	4	0	0	0	5
0	1	4	1	0	4	0	0	0	5
1	1	2	0	0	0	0	0	5	0
0	4	1	0	4	1	0	0	0	5

TRUE/FALSE → **1, 0**

Number of Clusters



(5, 6, 7) - Decided on 6 Clusters Using Features: "va", "or", "reflection", and "has_text"



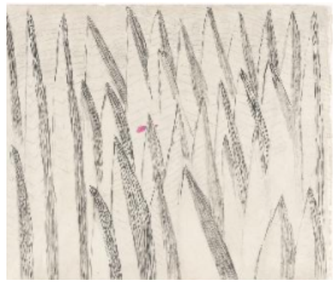

```
For n_clusters = 6 The average silhouette_score is : 0.4521922986577857
```

Silhouette analysis for KMeans clustering on sample data with n_clusters = 6

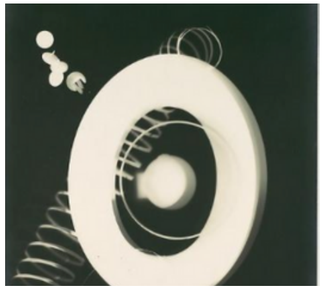
- **"VA" - Various Degrees between white and black**
- **"OR" - Various Orientations from Horizontal to Vertical**
- **"Reflection" - Image has reflective surface**
- **"Has_Text" - Image has Text**

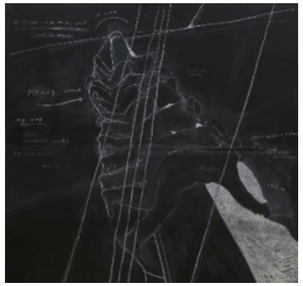
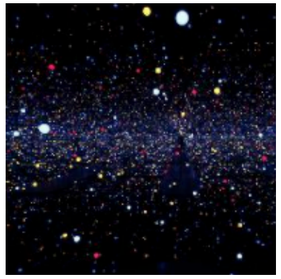
Clusters

Cluster 1




Cluster 2



Cluster 3




Cluster 4




Cluster 5




Cluster 6