



Customer Segmentation

Github repo:
<https://github.com/Trent-Kindvall/Customer-Segmentation>



Problem

- In marketing finding and being able to target different customer segments is an important optimisation
- Especially with digital marketing you can target very niche populations and influence the click through rate and conversion rate



Problem

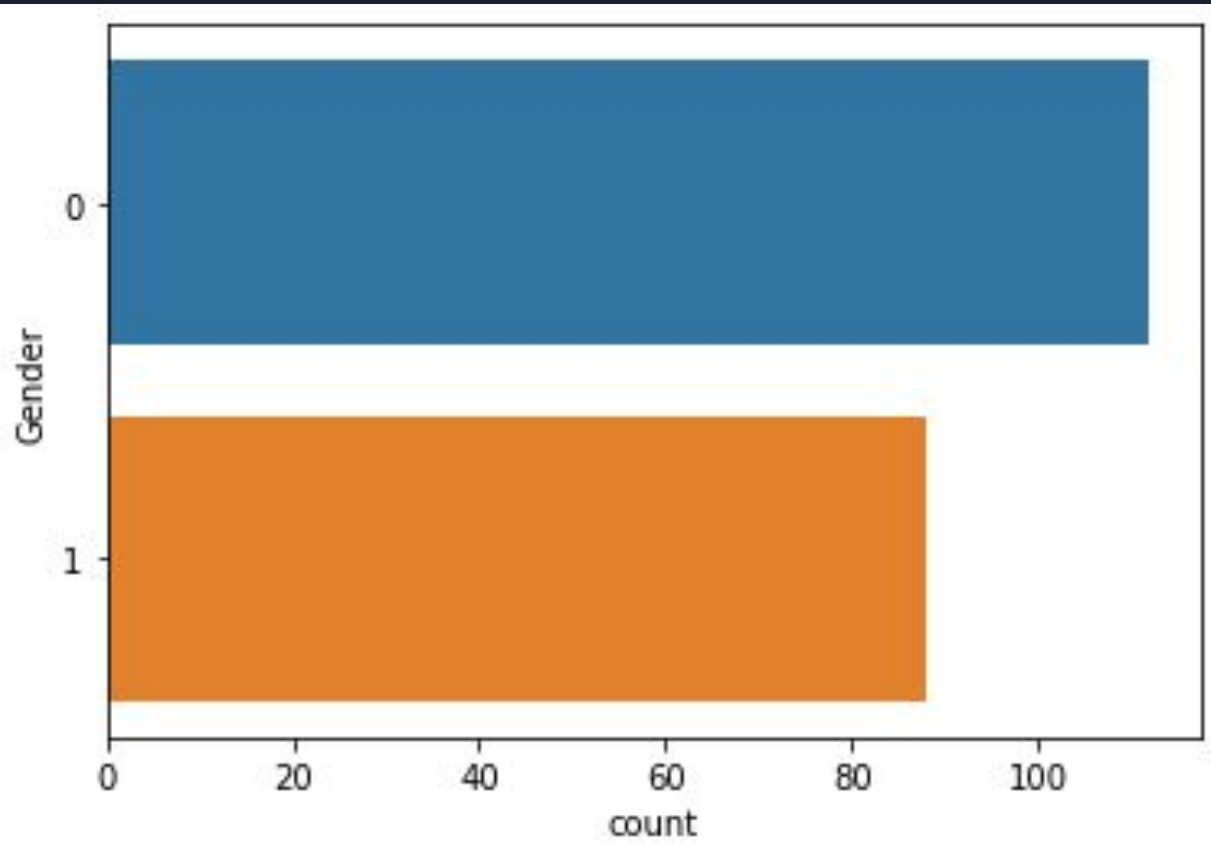
- This can be difficult because there is no way to check the accuracy. Which makes it a great candidate for an unsupervised model like K-means clustering



Data

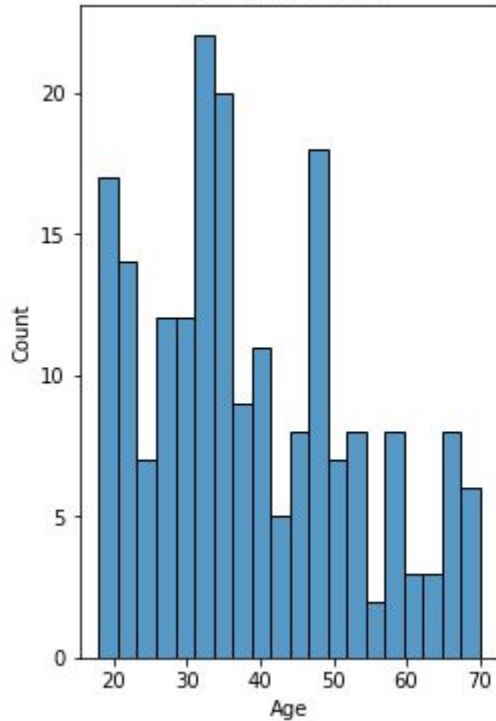
	Gender	Age	Annual_Income	Spending_Score
count	200.000000	200.000000	200.000000	200.000000
mean	0.440000	38.850000	60.560000	50.200000
std	0.497633	13.969007	26.264721	25.823522
min	0.000000	18.000000	15.000000	1.000000
25%	0.000000	28.750000	41.500000	34.750000
50%	0.000000	36.000000	61.500000	50.000000
75%	1.000000	49.000000	78.000000	73.000000
max	1.000000	70.000000	137.000000	99.000000

Exploratory Data Analysis

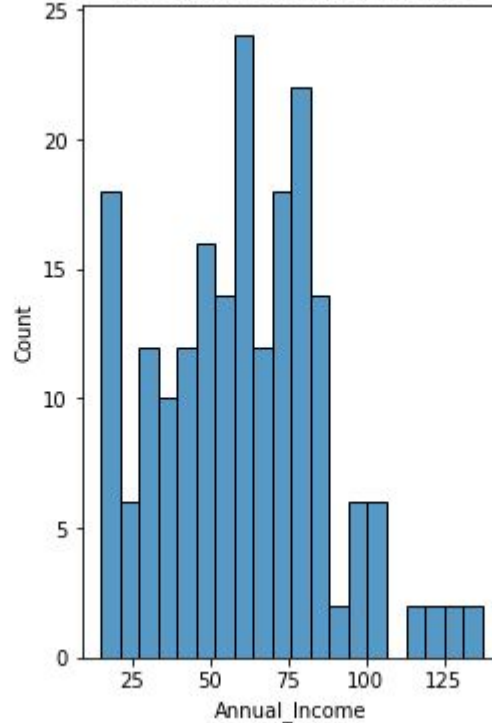


Exploratory Data Analysis

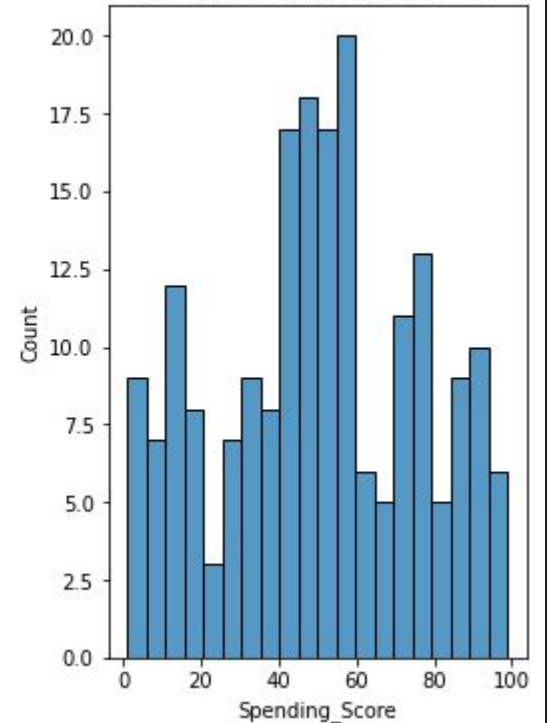
Histogram of Age



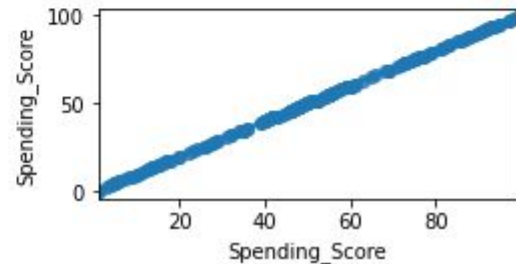
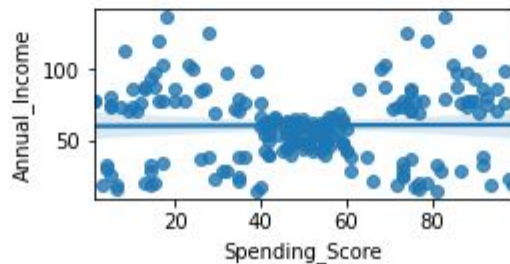
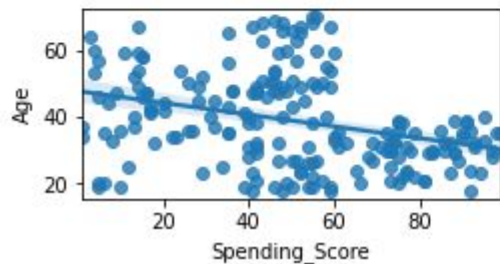
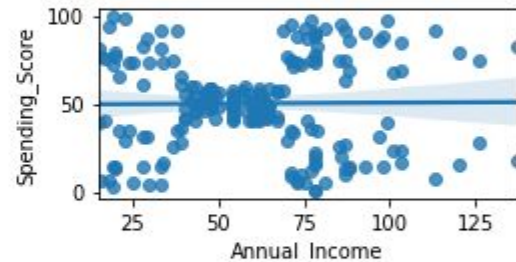
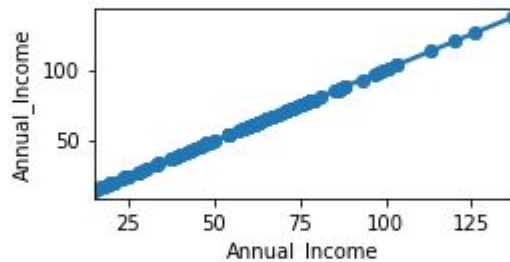
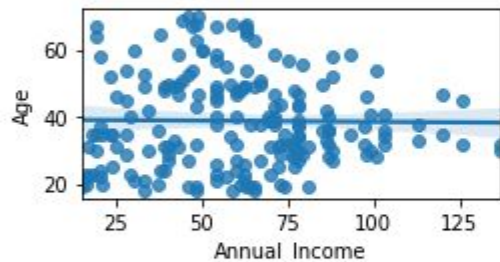
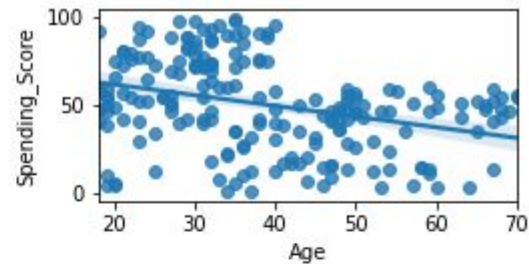
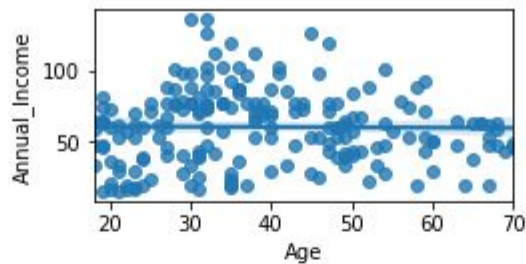
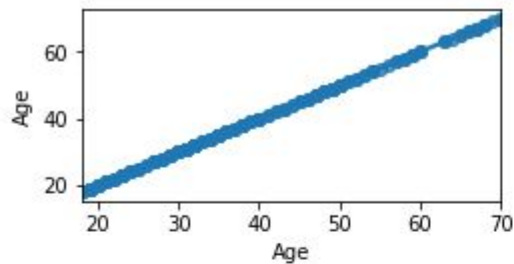
Histogram of Annual_Income



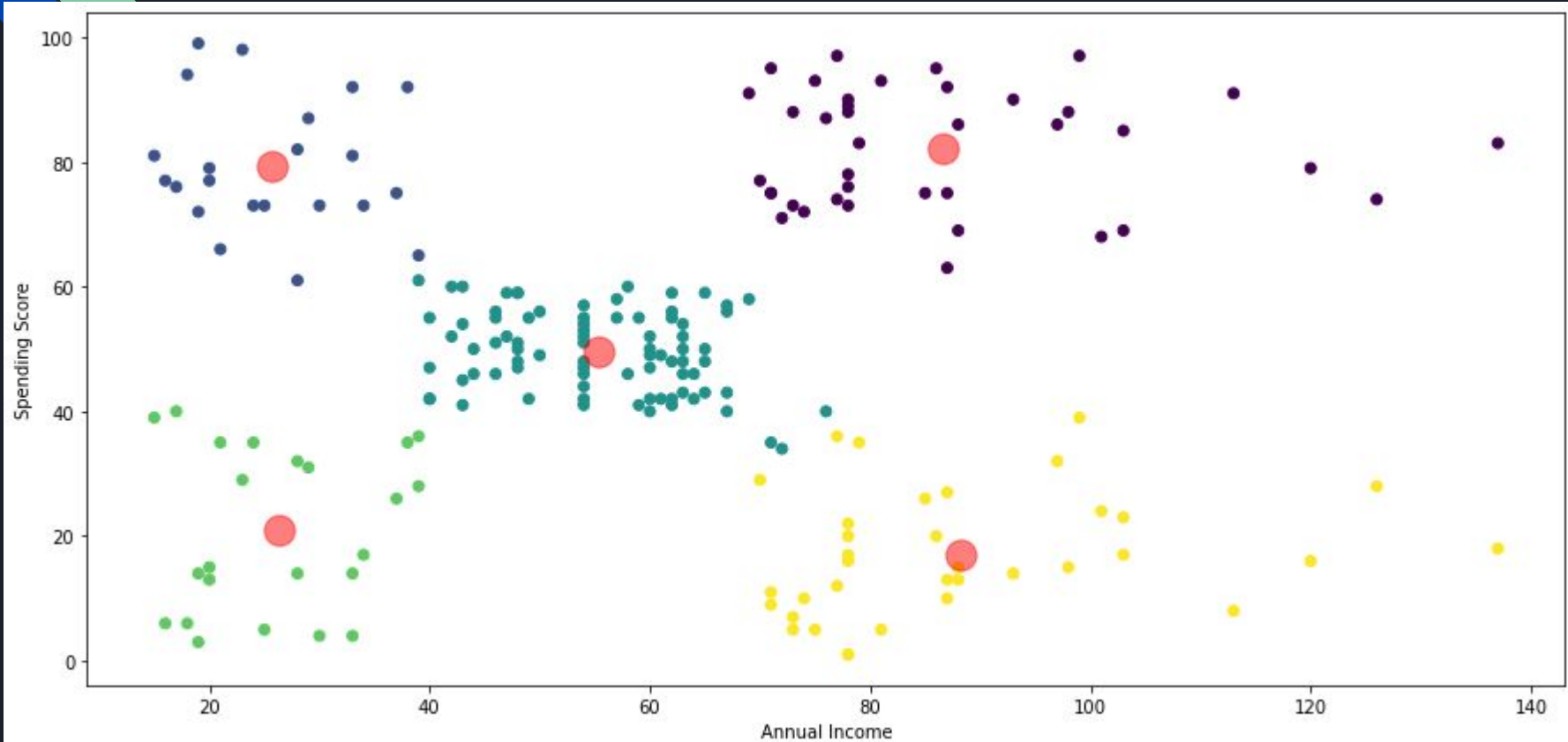
Histogram of Spending_Score



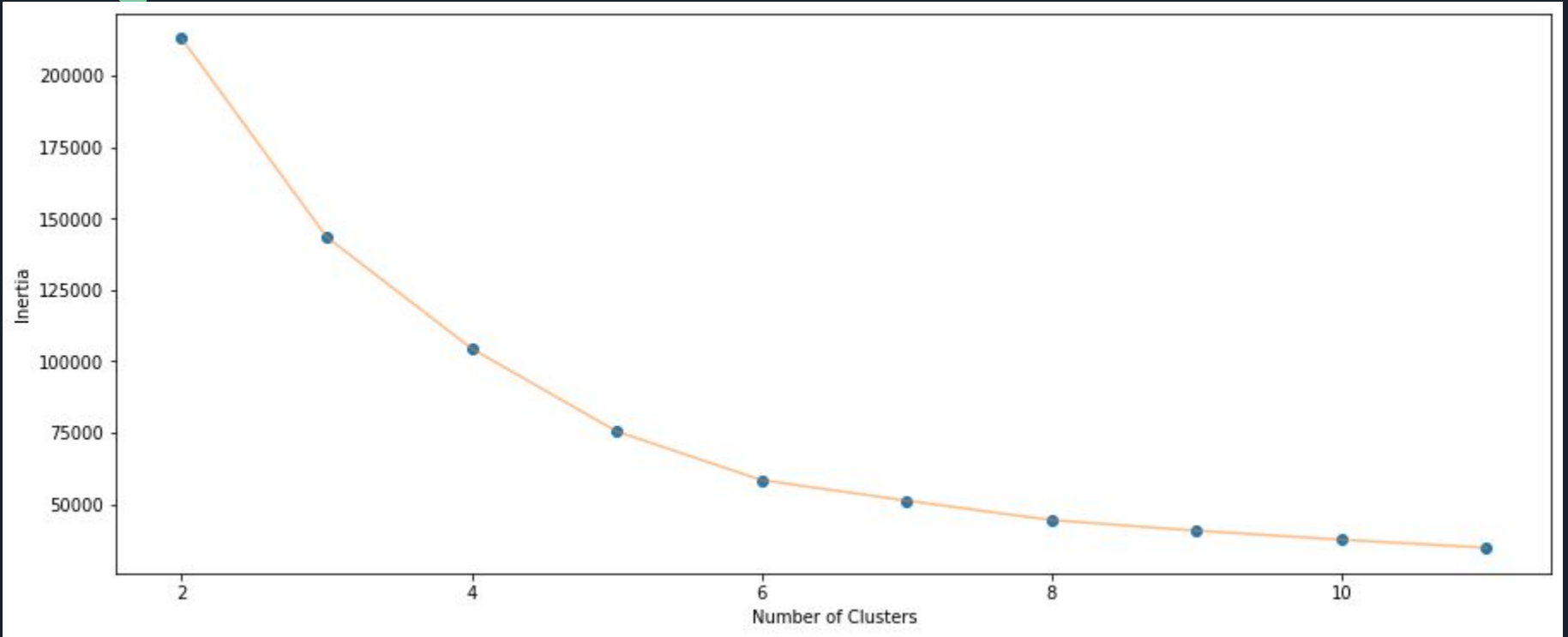
Exploratory Data Analysis



2D Example



Evaluation (Inertia)





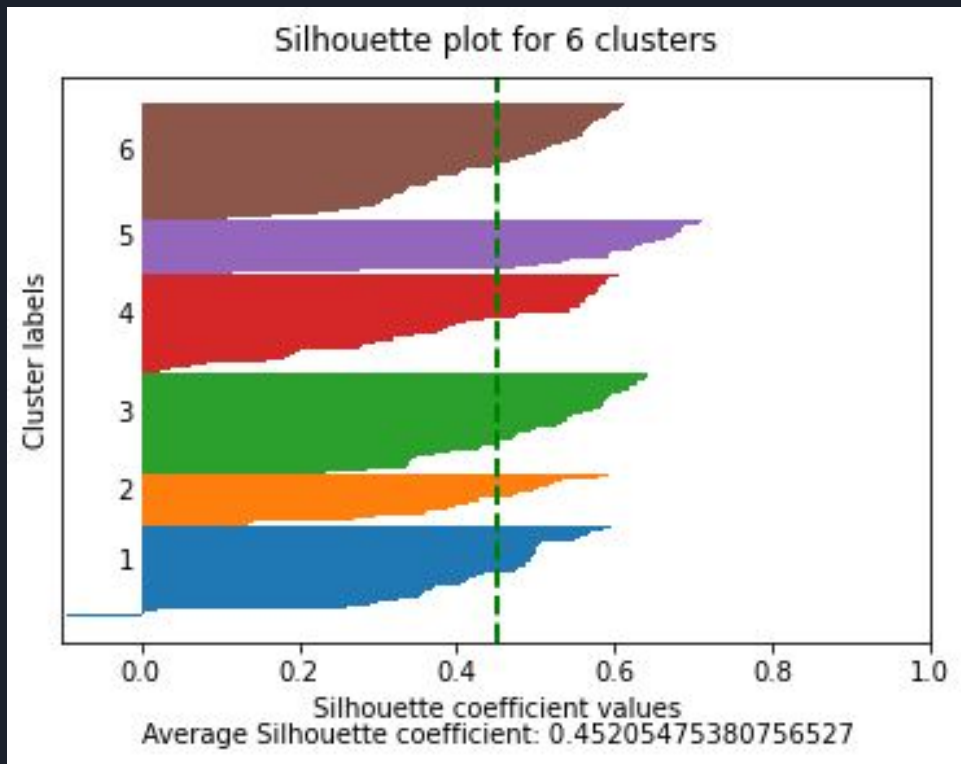
Evaluation (Silhouette Coefficient)

a_i : the average distance from all data points in the same cluster

b_i : the average distance from all data points in the closest cluster

$$\frac{b_i - a_i}{\max(a_i, b_i)}$$

Evaluation (Silhouette Coefficient)





Final Model

The best model I created was a K-means cluster with 6 centroids.

The .45 silhouette coefficient is a good