ABHISHEK DAS

BCGDV MACHINE LEARNING ASSIGNMENT

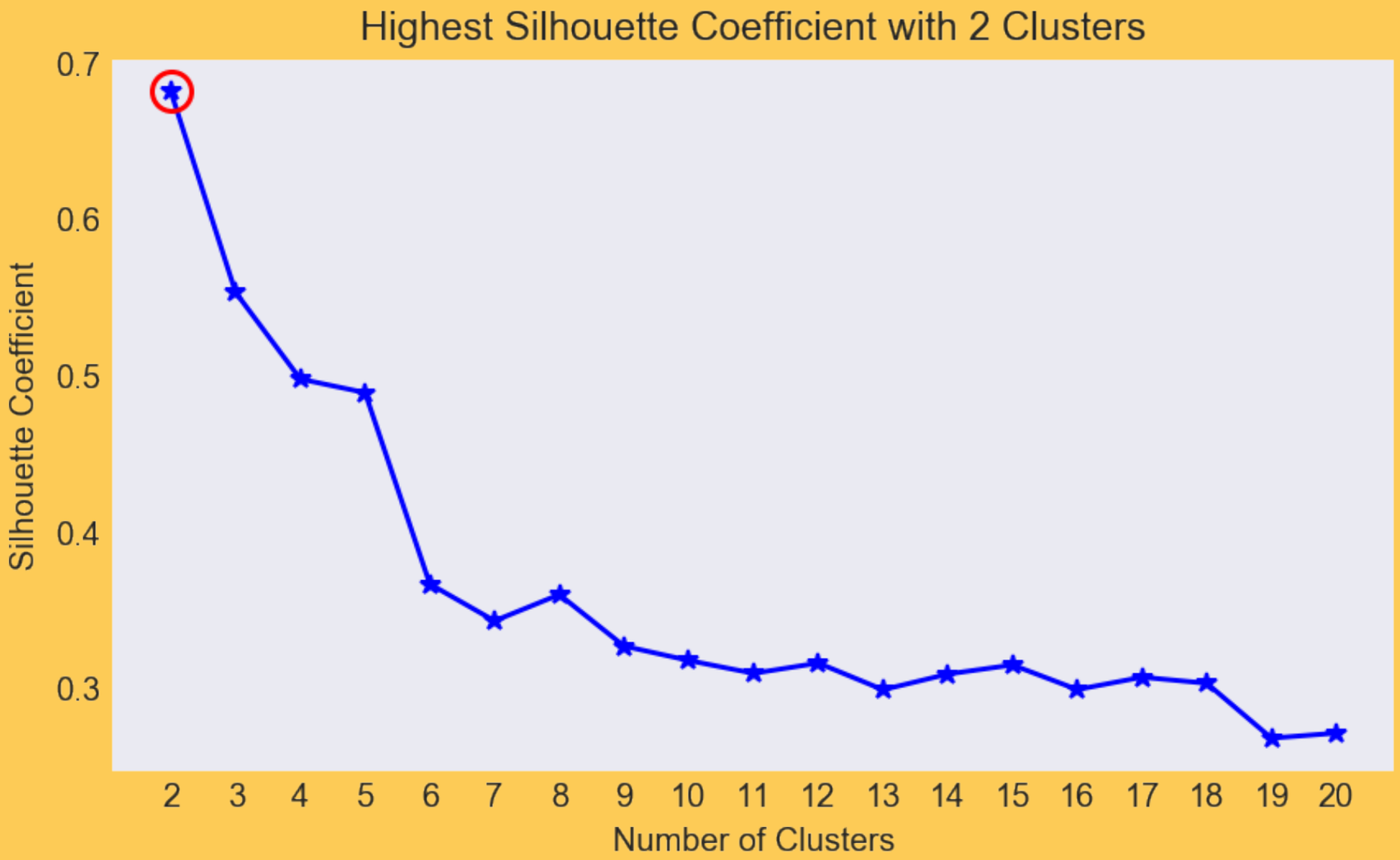
SECTION 1

K-MEANS CLUSTERING

- Intuitive: Let's randomly pick some centroids and move them around till they're the 'center' of our data clusters
- Distance: Minimizes within cluster variance = Euclidean^2
- Cohesion: How far are points from their centroid?
- Separation: How far are clusters from each other?

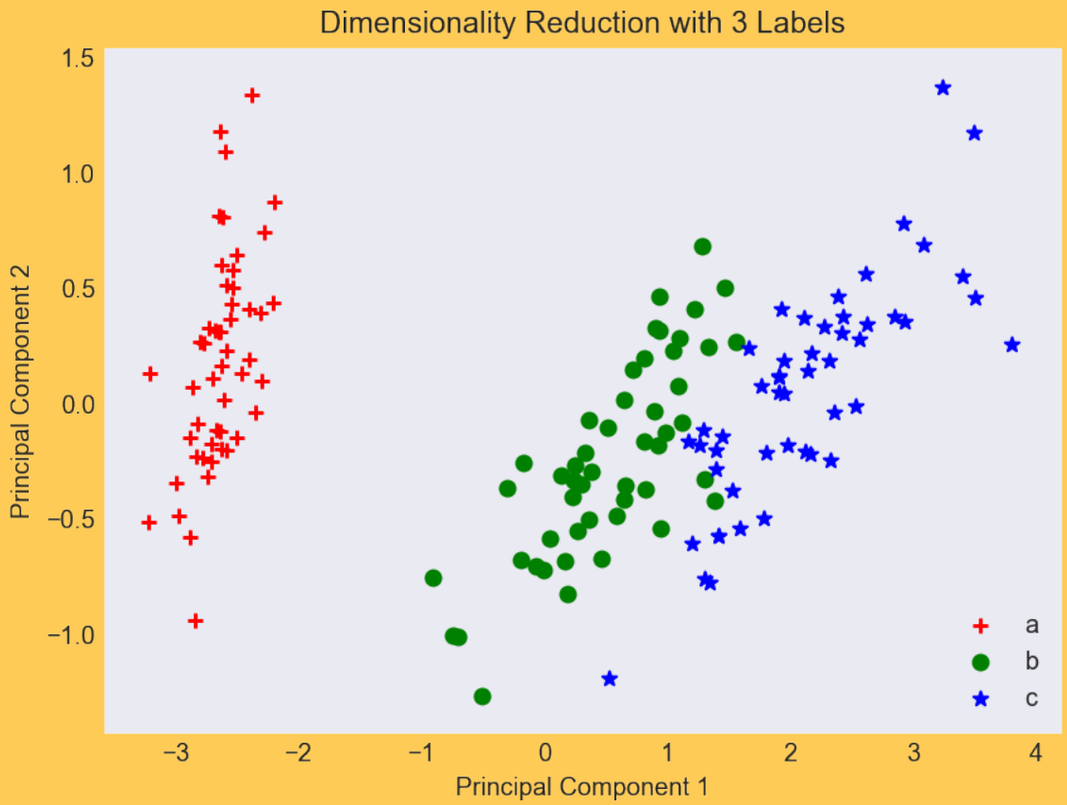
TEXT

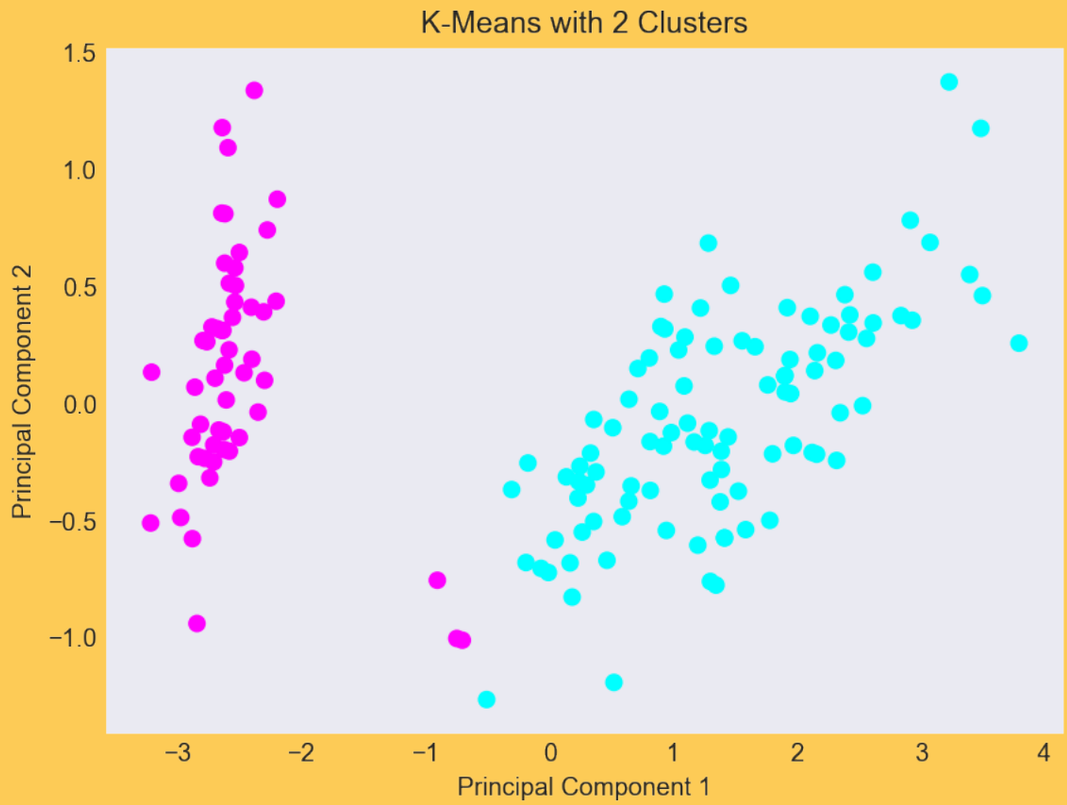
CHOOSING THE K IN K-MEANS



TEXT

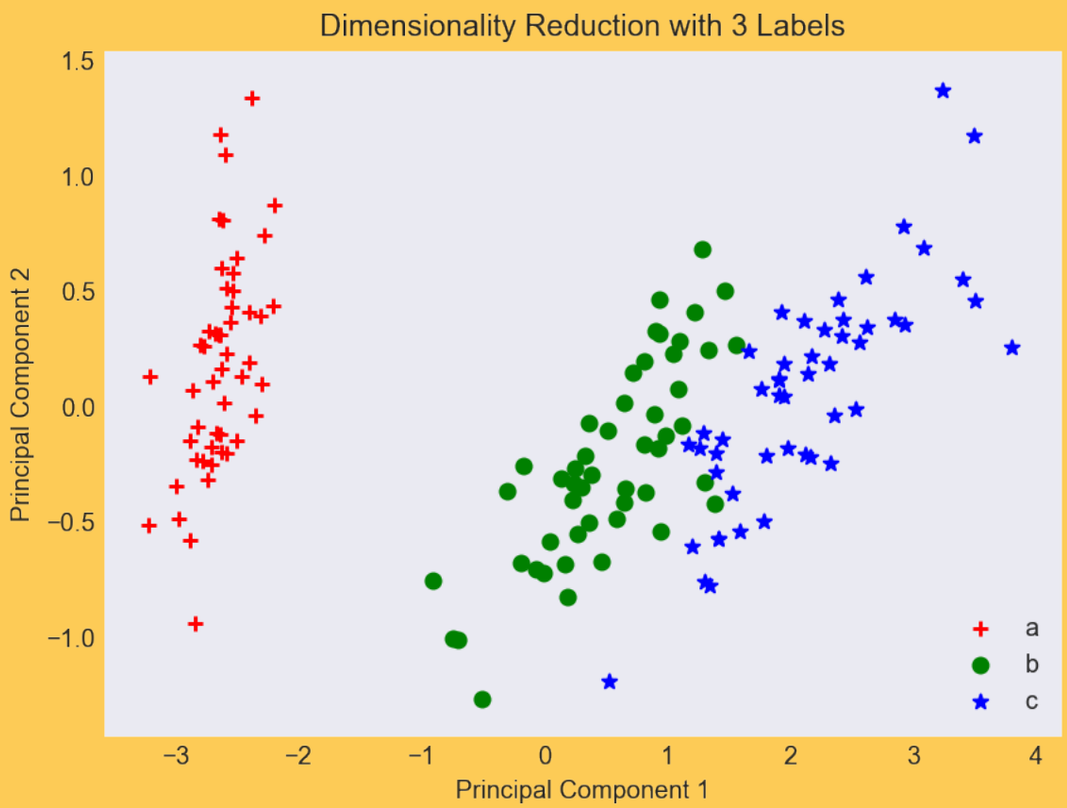
B AND C ARE GROUPED TOGETHER

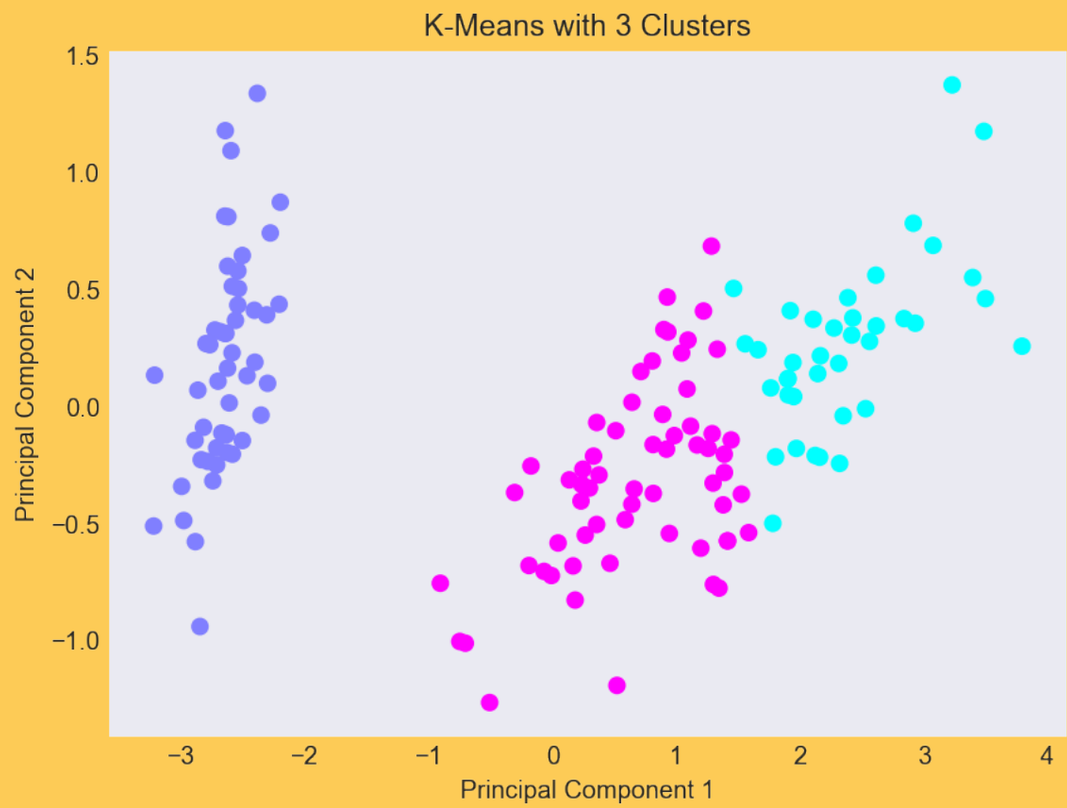




TEXT

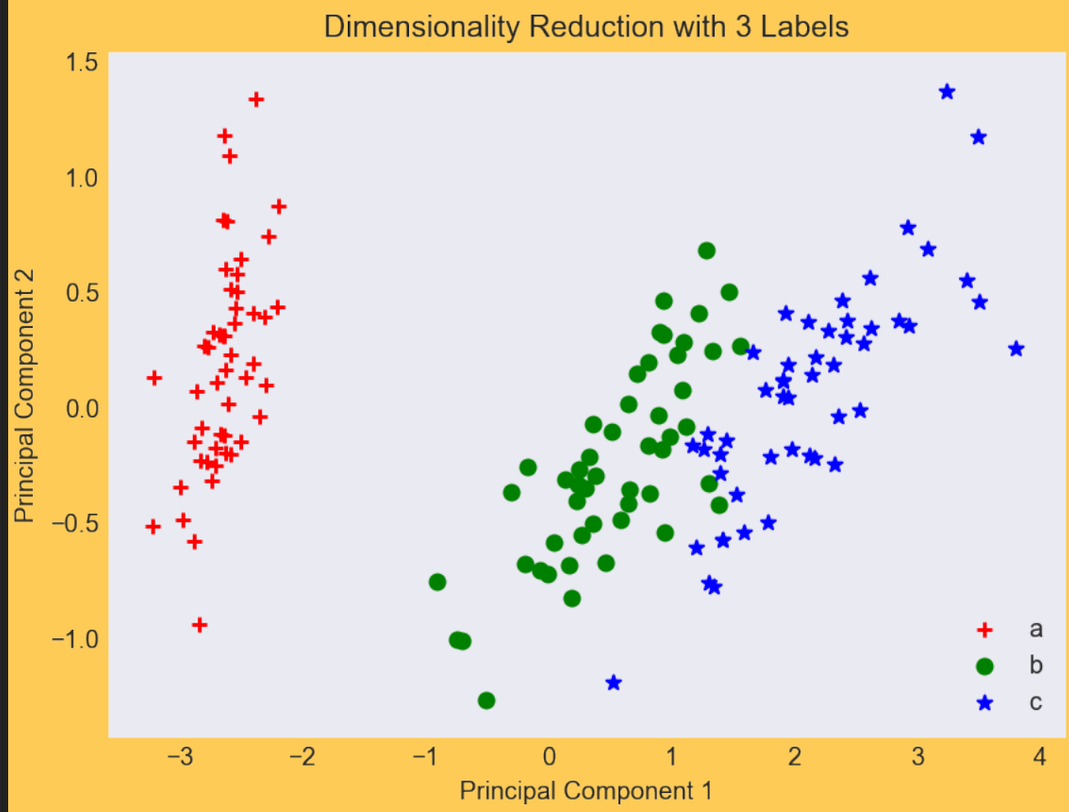
3 CENTROIDS IS NOT A LOT BETTER...

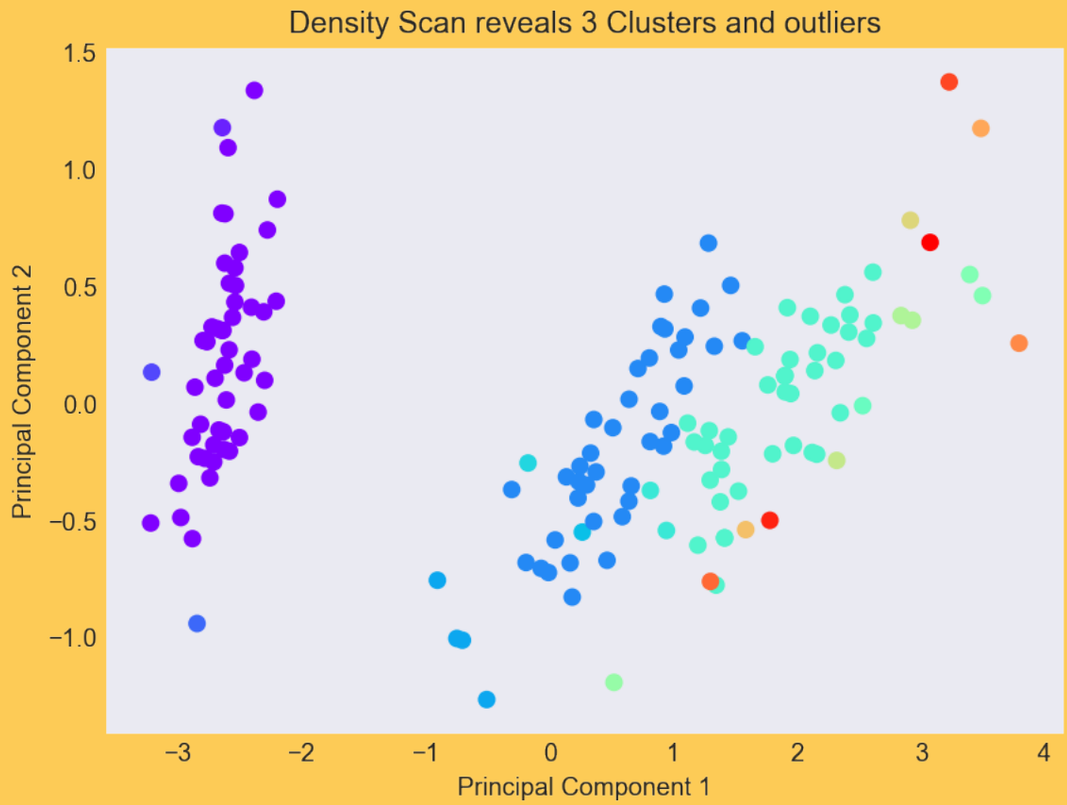




TEXT

DBSCAN BETTER BUT HAS OUTLIERS





SECTION 2

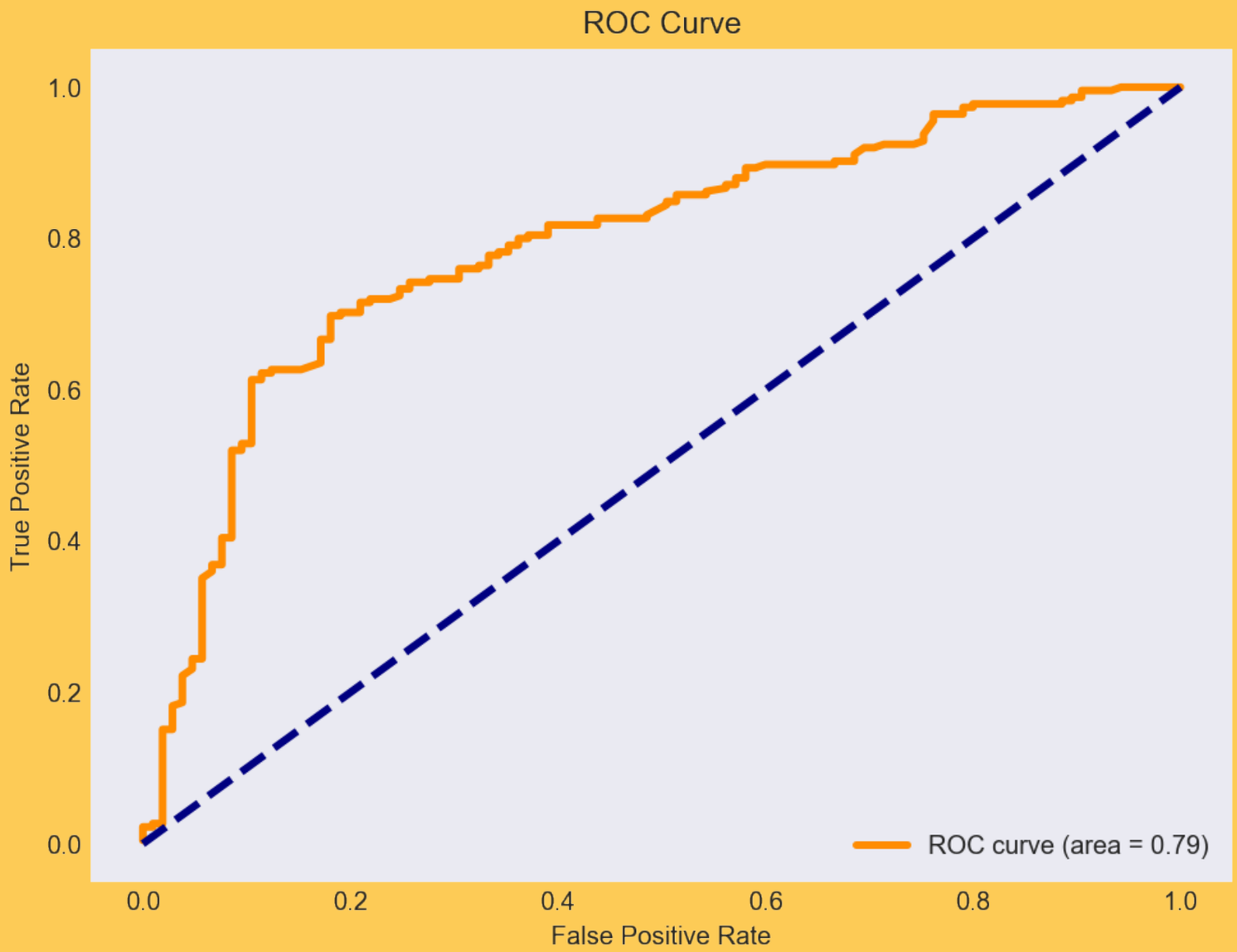
LOGISTIC CLASSIFIER

- The more young children you are looking after and the more hungry you are, but the less alternative eating options in your area means a higher probability of you eating dinner at McDonalds tonight.

- More comprehensible model

TEXT

79% ASSIGNED TO CORRECT CLASS



KNN CLASSIFIER

- Residents of Glebe like residents of Newtown, Marrickville, Erskineville and Surry Hills are more likely to take public transport to work

- Group into a class based on neighbors
- Also intuitive

TEXT

AUC LOGISTIC > AUC KNN