



2303: Internal Meeting

Date/Time 23 March 2018, 5:00 PM – 8:00 PM

Attendees Shubhangi, Tanushree, Arushi

Tasks	Actor	Follow up Action
<p>Selected variables for clustering and prepared them:</p> <ul style="list-style-type: none"> - RFM - Weekday/Weekend - Meal Time - Proportion of Repeat restaurants/cuisines - Discount proportions - Days in Advance Bookings - Tier 	All	Write out report for clustering analysis
<p>Performed clustering for different ranges: Choosing too small a k value can lead to overfitting and over generalization with a cluster comprising of too many observations. At the same time, choosing a very large value for k may lead to clusters that are too small without any defining characteristics.</p> <p>Therefore, to strike a balance between the two, we implemented clustering in ranges, going from a smaller range (3-6) to a larger range (3-15) in increments of 1. Through a process of elimination, we found k = 8 clusters to be the most optimal having the largest 'CCC' score and the fewest number of clusters with less than 5% of total users.</p>	All	Clustering with 8 clusters is selected
<p>Performed other clustering techniques other than K-Means to compare:</p> <ul style="list-style-type: none"> - Normal Mixtures - Latent Class 	All	