Targeting customers in hospitality through cluster analysis

Ludovico Gandolfi - 07/2022

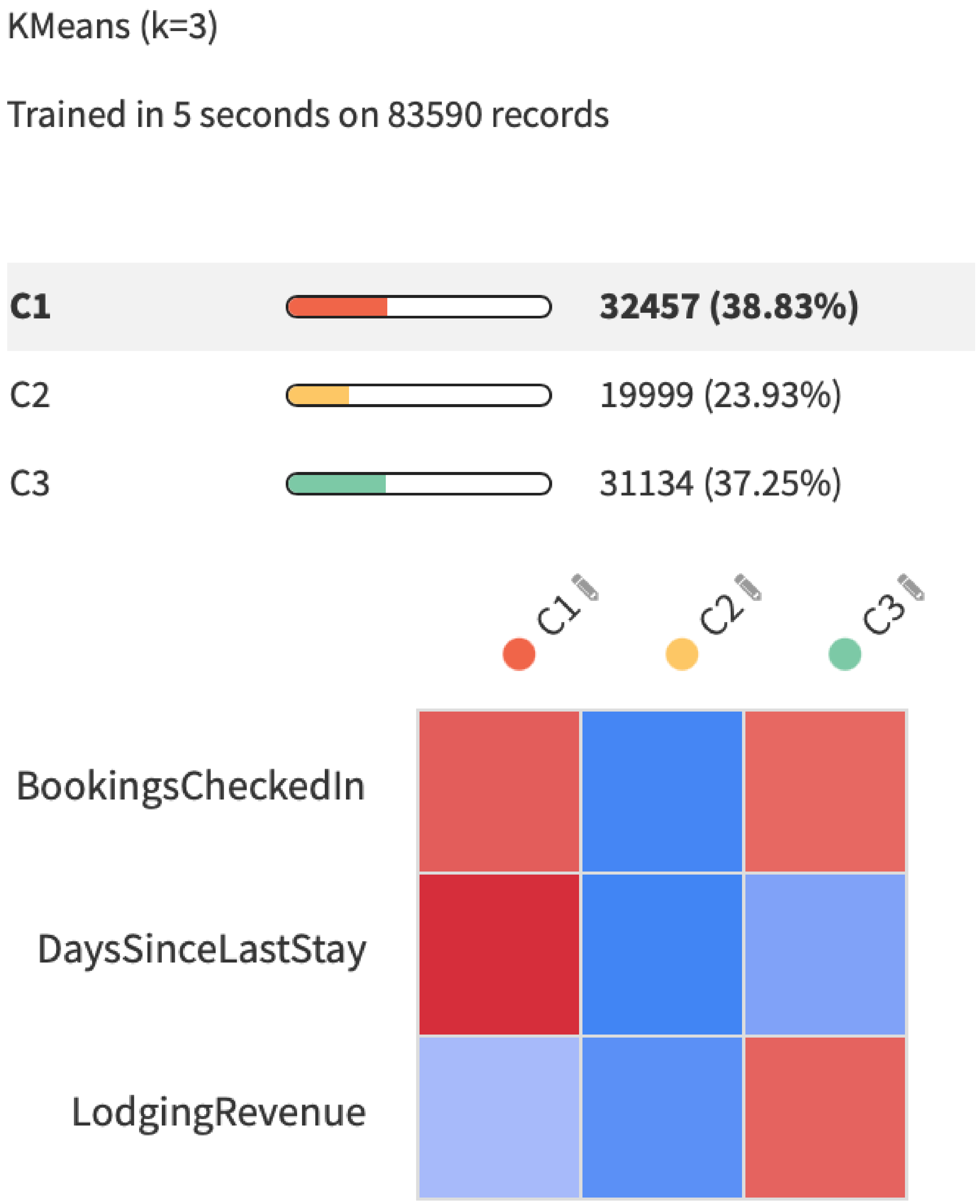
Baseline model

This baseline model brings us three clusters with a similar number of observations in each, which, together with the features evaluated, could allow us to infer on their meaning.

C1 = Disengaged/Occasional *

C2 = Incoming Customers

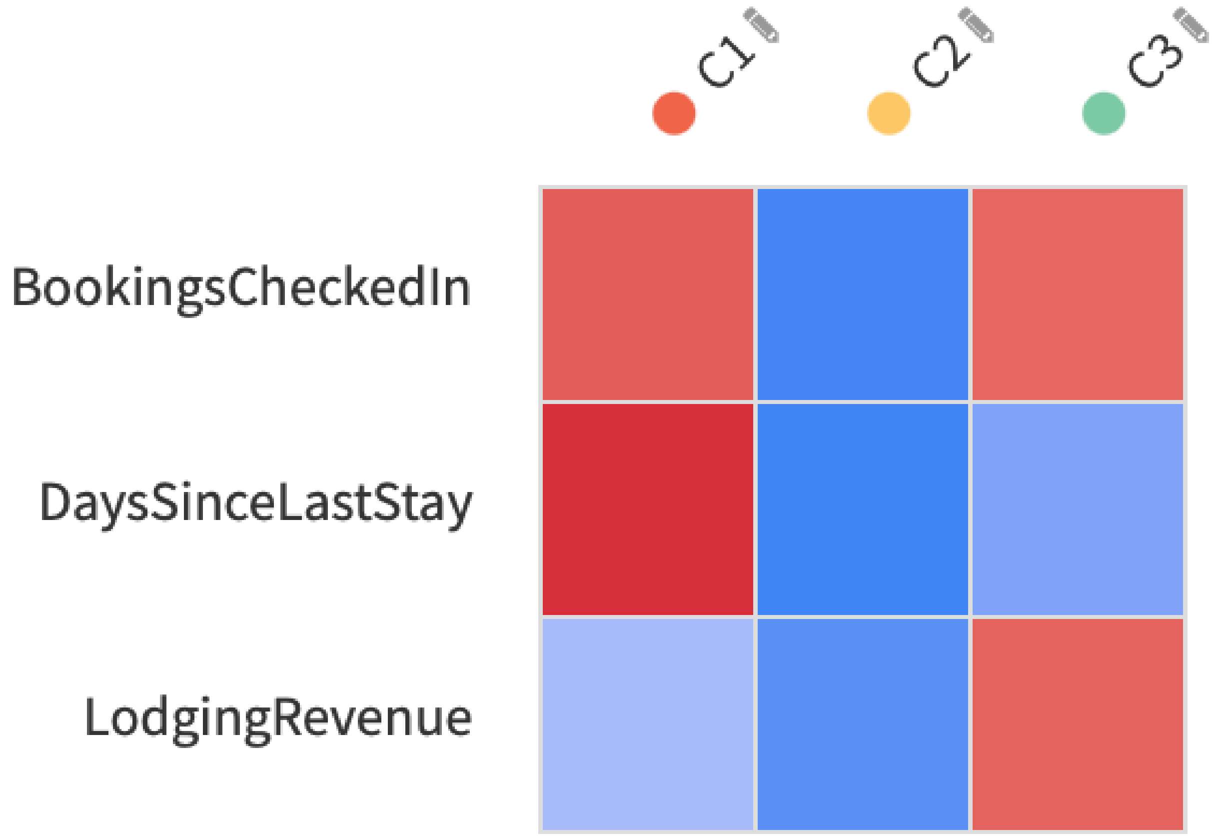
C3 = High-spenders/Occasional **



*profiled as businesspeople **profiled as families

Profiling customers

We defined in the previous page C1, C2 and C3 and imputed some profiles to these clusters.

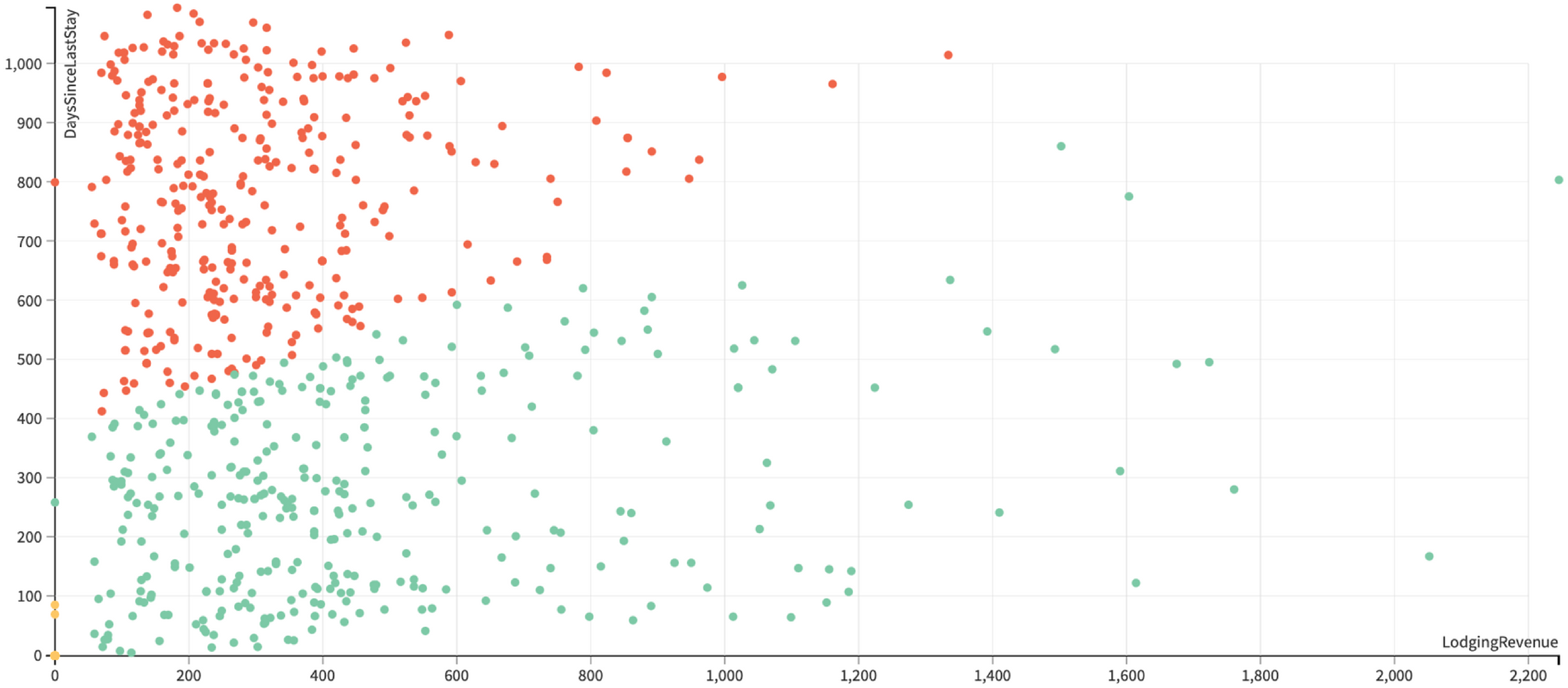


In fact, being the customers in C1 more prone to use seldomly the hotel, for very brief periods and with low spending, these could be profiled as **businesspeople**.

Customers in C3 display a higher frequency and spending (also linked to higher days of stay), therefore they could be regarded as **families**.

Lastly, customers in C2 have negative *DaysSinceLastStay* and, on average, 0 *BookingsCheckedIn*, signaling they are **incoming customers**.

Visual analysis



It could be particularly interesting to look at the scatterplot of *LodgingRevenue* on *DaysSinceLastStay* , which signals how recurrent customer are also the ones who bring the highest revenue per-stay.

Suggestions per-group on the baseline model

C1 - Businesspeople

This segment could be targeted by offering business-oriented deals to corporations, in order to increase the flow of managers in the hotel.

C2 - Incoming Customers

No marketing campaign needed, try to improve the customer experience based on the previous customers' reviews.

C3 - Families

An e-mail re-targeting campaign, offering the appropriate rewards and discounts could appeal to these occasional customers.

*continued in the next page

*continued

The exploitation of analytical techniques could allow us to assign values to our customers, as the CLV, and running a supervised model would allow us to predict with a certain accuracy which specific customers we should be targeting.

In fact, it would not be optimal to assign customers to a cluster based the characteristics and labeling we assigned to each cluster, as this could provide us with non-significant results.

Another interesting approach could be gathering data on a large variety of characteristics that appeal more to the customers, in order to explore these data through a dimensionality reduction with unsupervised algorithms such as a Principal Component Analysis, which could allow us to define the factors (ensemble of features) which affect the most our customers' choices, in order to improve the hotel on the characteristics where it lacks the most.

Implementing RFM

RFM models combine three different customer attributes to rank customers. The components are **Recency** (days since last purchase), **Frequency** (flow of transactions) and **Monetary** (total money spent). RFM analysis would be useful in finding our best customers, understanding their behavior and then running targeted marketing campaigns.

In our case, we evaluated Recency on the basis of the *DaysSinceLastStay*, Frequency on *BookingsCheckedIn* and Monetary on *LodgingRevenue* .

Tailoring a recipe

Creating values for Recency

If (DaysSinceLastStay < 30)

Then (Recency= 5)

Else if (DaysSinceLastStay < 50) AND (DaysSince
LastStay > 30)

Then (Recency= 4)

Else if (DaysSinceLastStay < 100) AND (DaysSinc
eLastStay > 50)

Then (Recency= 3)

Else if (DaysSinceLastStay < 500) AND (DaysSinc
eLastStay > 100)

Then (Recency= 2)

Else if (DaysSinceLastStay > 500)

Then (Recency= 1)

Creating values for Frequency

If (BookingsCheckedIn >= 50)

Then (Frequency= 5)

Else if (BookingsCheckedIn < 50) AND (BookingsC
heckedIn >= 30)

Then (Frequency= 4)

Else if (BookingsCheckedIn < 30) AND (BookingsC
heckedIn >= 10)

Then (Frequency= 3)

Else if (BookingsCheckedIn < 10) AND (BookingsC
heckedIn >= 2)

Then (Frequency= 2)

Else if (BookingsCheckedIn == 1)

Then (Frequency= 1)

Else if (BookingsCheckedIn == 0)

Then (Frequency= 0)

Creating values for Monetary

If (LodgingRevenue >= 500)

Then (Monetary= 5)

Else if (LodgingRevenue < 500) AND (LodgingRev
enue >= 300)

Then (Monetary= 4)

Else if (LodgingRevenue < 300) AND (LodgingRev
enue >= 200)

Then (Monetary= 3)

Else if (LodgingRevenue < 200) AND (LodgingRev
enue >= 100)

Then (Monetary= 2)

Else if (LodgingRevenue < 100) AND (LodgingRev
enue >= 1)

Then (Monetary= 1)

Else if (LodgingRevenue < 1)

Then (Monetary= 0)

Concatenating and creating RFM

Columns to concatenate

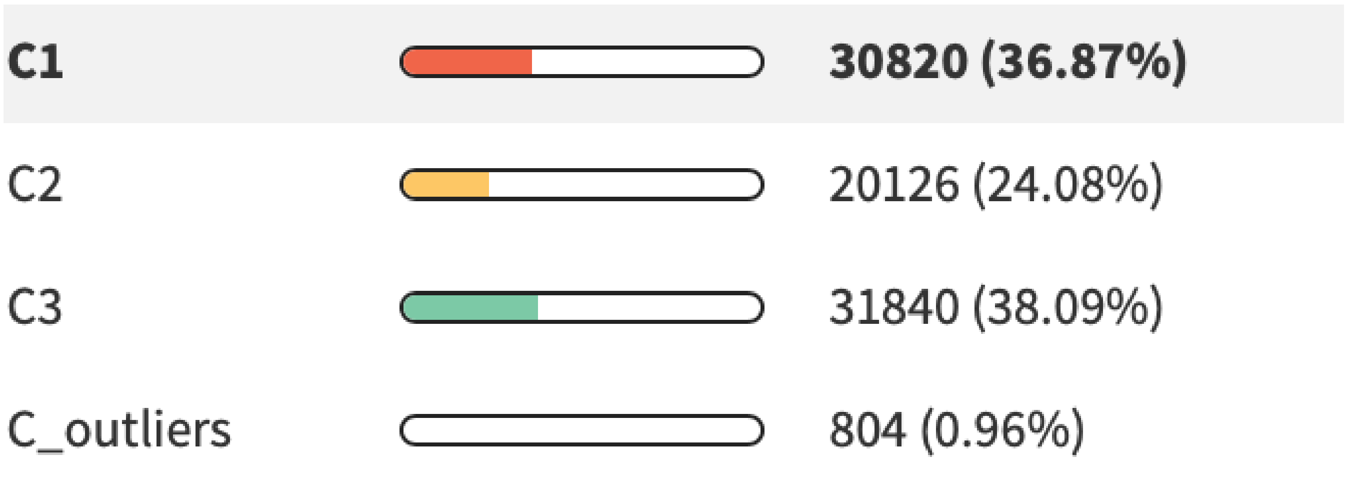
- Recency
- Frequency
- Monetary

↓

RFM Value
bigint
Integer
224
113
500
113

Improving the model – 1

Silhouette = 0.5907 K = 3

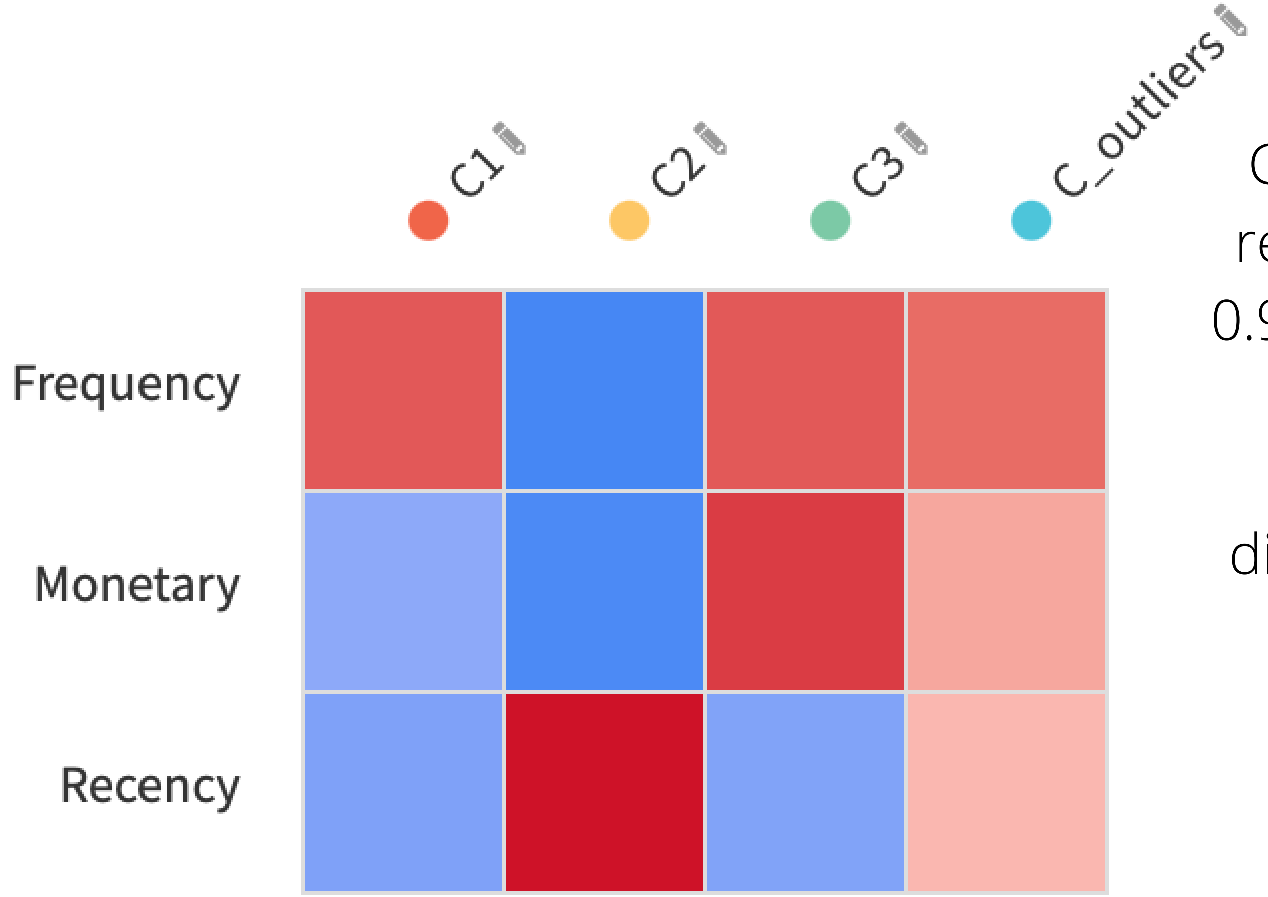


This first clustering model ran on the prepared dataset gave us some intuitions that did not differ that much from the baseline model.

C1 = Disengaged/Occasional *

C2 = Incoming Customers

C3 = High-spenders/Occasional **



C_outliers represents 0.96% of our data, so we disregard it

*profiled as businesspeople **profiled as

Improving the model – 2

However, in our second model, our performance metrics increased exponentially and gave us interesting insights to develop based on the RFM.

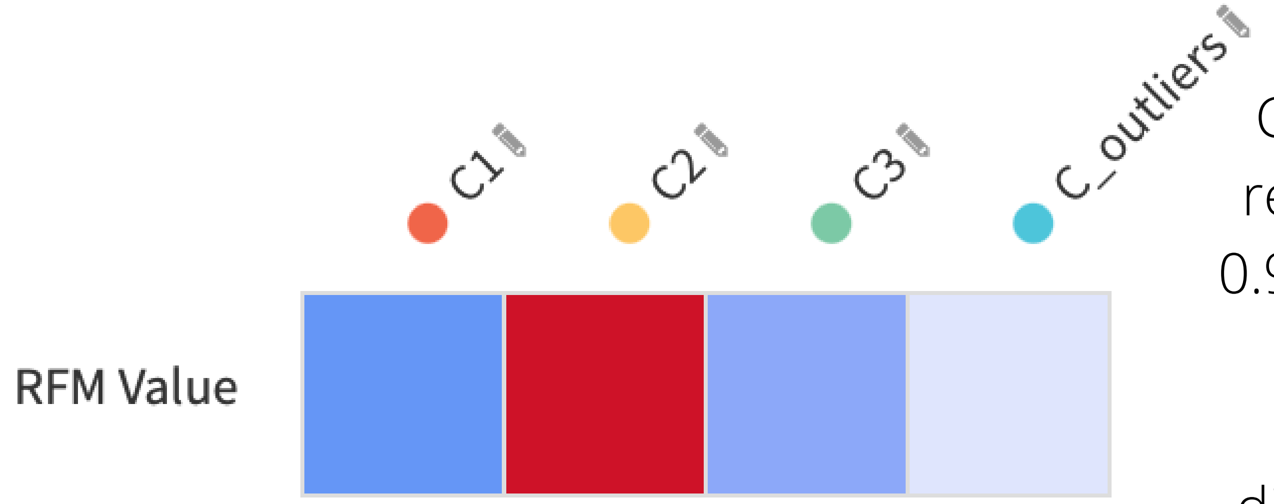
C1 = Low value Customers (low on RFM)

C2 = High value Customers (high on RFM)

C3 = Mid value Customers (mixed RFM)

Silhouette = 0.929 K = 3

C1	**33585 (40.18%)**
C2	21768 (26.04%)
C3	27447 (32.84%)
C_outliers	790 (0.95%)



C_outliers represents 0.95% of our data, so we disregard it

Suggestions per-group on the improved model

C1 - Low-value customers

No marketing campaign should be implemented for these customers if we base our decision on the RFM, as this group scores low on all metrics.

C2 - High-value customers

This segment must be rewarded and targeted on the basis of their loyalty, signaled by their high RFA scores. Loyalty programs and discounts offer an option.

C3 - Mid-value customers

An e-mail retention campaign, offering discounts and "customized deals" could appeal to this group of mixed customers.

*continued in the next page

*continued

With regards to those customers we defined as "Mid-Value Customers", we understand that they could score a mixed RFM in a variety of different ways.

For this reason, we defined the option of "customized solutions".

In fact, they could score high in Monetary and Recency, therefore we should target their Frequency, i.e. the number of times these customers turn to our hotel by offering loyalty plans.

Differently, the customers who score high in Monetary and Frequency should be targeted on Recency, by applying, for instance, a real retention plan or a retargeting campaign in order to push them again towards choosing our accommodation.