

Hotel Customers Analysis

Team 3

Jaya Nagesh, Selma Sentissi El Idrissi, Shamika Kalwe, Shuyi Zhu

Agenda

- Business Problem
- Dataset and Cleaning
- Exploratory Data Analysis
- Unsupervised Learning
 - *Association Rules*
 - *HClust*
 - *DBscan*
 - *KMeans*
- Business Implications
- Conclusions
 - *Key Takeaways*
 - *Limitation and Further Improvement*

Business Problem

The aim is to understand and group customers of a hotel, based on customers' personal, behavioral, demographic, and geographic data.

Dataset and Cleaning

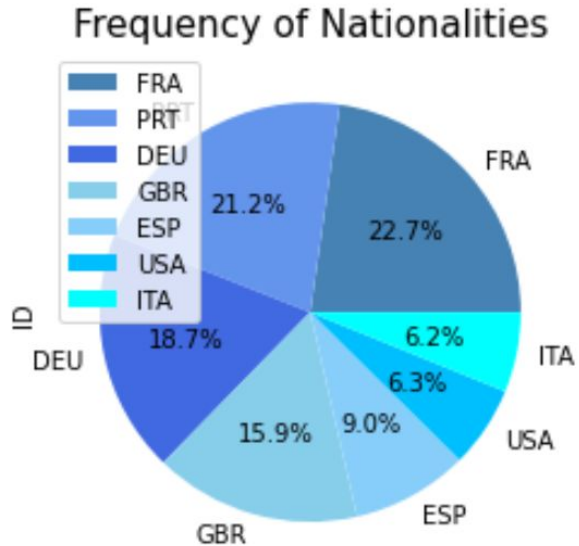
- Data Source: <https://www.kaggle.com/nantonio/a-hotels-customers-dataset>
- Dataset Size:
 - 83,590 rows * 31 columns (before cleaning)
 - 83,590 rows * 29 columns (after cleaning)
- Dropping Columns: NameHash, DocIDHash
- Missing Values: Age
- Outliers: Age, AverageLeadTime
- Sampling the dataset (20%, 16,718 rows) for modeling

| ID | Nationality | Age | DaysSinceCreation | AverageLeadTime | LodgingRevenue | OtherRevenue | BookingsCanceled | BookingsNoShowed | BookingsCheckedIn | PersonsNights | RoomNights | DaysSinceLastStay | DaysSinceFirstStay |
|-------|-------------|-----------|-------------------|-----------------|----------------|--------------|------------------|------------------|-------------------|---------------|------------|-------------------|--------------------|
| 64031 | BEL | 19.000000 | 168 | 149.0 | 548.0 | 150.0 | 0 | 0 | 1 | 10 | 5 | 173.000000 | 173.00000 |
| 78720 | DEU | 56.000000 | 58 | 0.0 | 0.0 | 0.0 | 0 | 0 | 0 | 0 | 0 | 526.859149 | 529.85494 |
| 5429 | PRT | 45.402401 | 976 | 62.0 | 310.7 | 136.0 | 0 | 0 | 1 | 8 | 4 | 980.000000 | 980.00000 |
| 989 | DNK | 62.000000 | 1058 | 13.0 | 1660.0 | 345.5 | 0 | 0 | 1 | 8 | 4 | 1062.000000 | 1062.00000 |
| 43624 | FIN | 30.000000 | 351 | 50.0 | 369.0 | 69.5 | 0 | 0 | 1 | 6 | 3 | 354.000000 | 354.00000 |



Exploratory Data Analysis

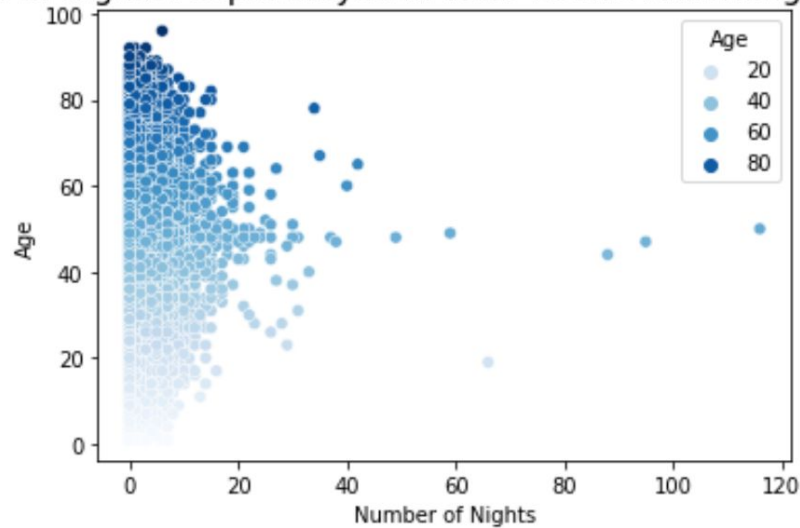
Top 7 Nations by Customers



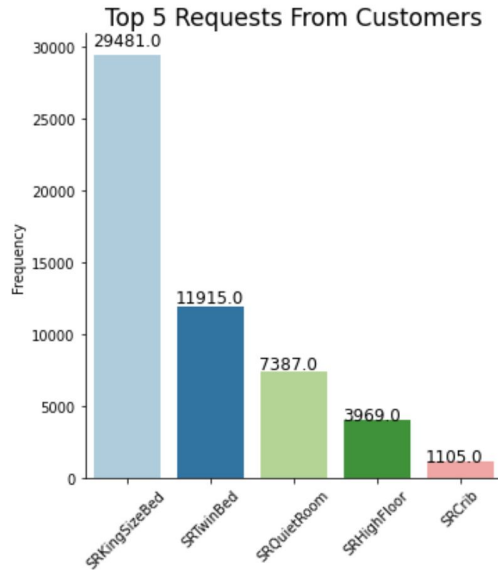
France
Portugal
Germany
United Kingdom
Spain
United States of America
Italy

Number of Nights vs Customers Age

Number of Nights People Stayed in Their Rooms According to Their Age



Top 5 Frequent Requests From Customers



King-size Bed

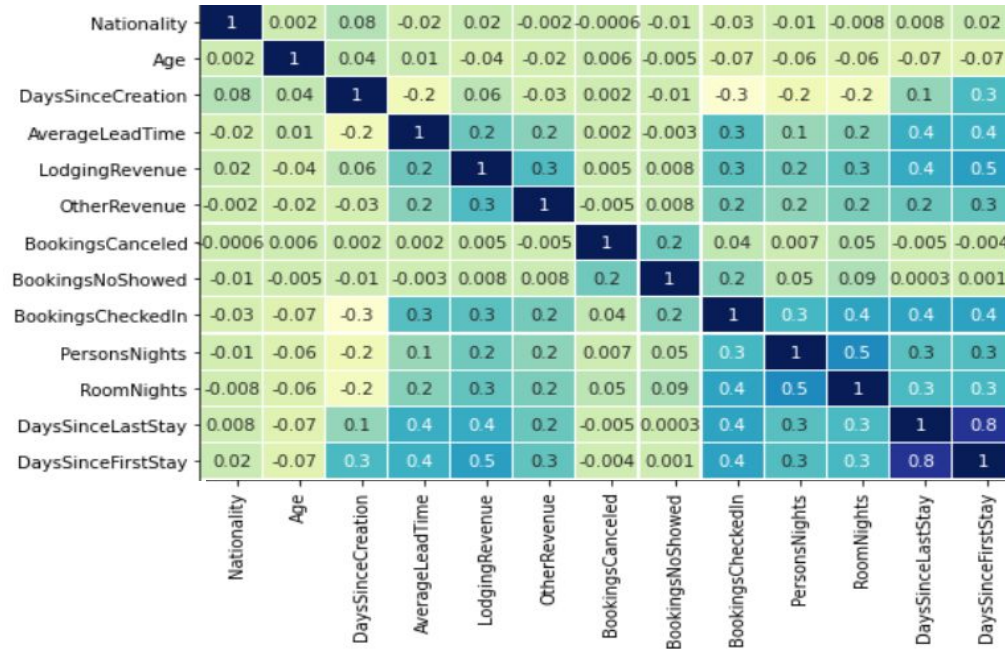
Twin Bed

Quiet Room

High Floor


Crib

Correlation Heatmap (Most Correlated)





Unsupervised Learning

- Association Rules
 - HClust
 - DBscan
 - KMeans
- 

Association Rules

- Loyalty Check of Customers: Repeated Bookings \leq Age Group, Nationality, Length of Stay
- Length of Stay \leq Age Group, Distribution Channel, Avg Daily Price Group, Nationality

| | Age Group | Distribution Channel | Avg Daily Price | Nationality | Repeated Bookings | Length of Stay |
|-------|-------------|-----------------------|-----------------|------------------|------------------------|----------------|
| 23938 | Adult 18-39 | Travel Agent/Operator | Low | GBR | Book 2 Times | Holiday |
| 66410 | Senior 60+ | Travel Agent/Operator | Low | Others Countries | Book 2 Times | Weekend Trip |
| 10527 | Adult 18-39 | Travel Agent/Operator | Medium | Others Countries | Book 2 Times | Holiday |
| 29667 | Adult 40-59 | Travel Agent/Operator | Low | Others Countries | Book 3 Times and above | Weekend Trip |
| 11813 | Adult 40-59 | Travel Agent/Operator | Low | Others Countries | Book 3 Times and above | Weekend Trip |

Association Rules

- Customer Loyalty Check by Age, Nationality, and Length of Stay

| | antecedents | consequents | antecedent support | consequent support | support | confidence | lift | leverage | conviction |
|----|---------------|--------------------------|--------------------|--------------------|----------|------------|----------|----------|------------|
| 11 | (Adult 40-59) | (Book 3 Times and above) | 0.510257 | 0.126470 | 0.084384 | 0.165375 | 1.307626 | 0.019852 | 1.046614 |
| 19 | (Senior 60+) | (Book 2 Times) | 0.196695 | 0.629983 | 0.155068 | 0.788368 | 1.251413 | 0.031154 | 1.748402 |

| | antecedents | consequents | antecedent support | consequent support | support | confidence | lift | leverage | conviction |
|----|-------------|--------------------------|--------------------|--------------------|----------|------------|----------|----------|------------|
| 10 | (ESP) | (Book 3 Times and above) | 0.030470 | 0.126470 | 0.005932 | 0.194670 | 1.539261 | 0.002078 | 1.084686 |
| 15 | (FRA) | (Book 2 Times) | 0.065565 | 0.629983 | 0.046429 | 0.708131 | 1.124049 | 0.005124 | 1.267753 |
| 21 | (GBR) | (Book 2 Times) | 0.045475 | 0.629983 | 0.032482 | 0.714286 | 1.133818 | 0.003834 | 1.295061 |

| | antecedents | consequents | antecedent support | consequent support | support | confidence | lift | leverage | conviction |
|----|----------------|--------------------------|--------------------|--------------------|----------|------------|----------|----------|------------|
| 12 | (Holiday) | (Book 3 Times and above) | 0.319211 | 0.126470 | 0.049112 | 0.153855 | 1.216534 | 0.008742 | 1.032364 |
| 6 | (Holiday) | (Book 2 Times) | 0.319211 | 0.629983 | 0.220527 | 0.690853 | 1.096622 | 0.019430 | 1.196897 |
| 10 | (Weekend Trip) | (Book 2 Times) | 0.492285 | 0.629983 | 0.322423 | 0.654952 | 1.039635 | 0.012292 | 1.072366 |

Association Rules

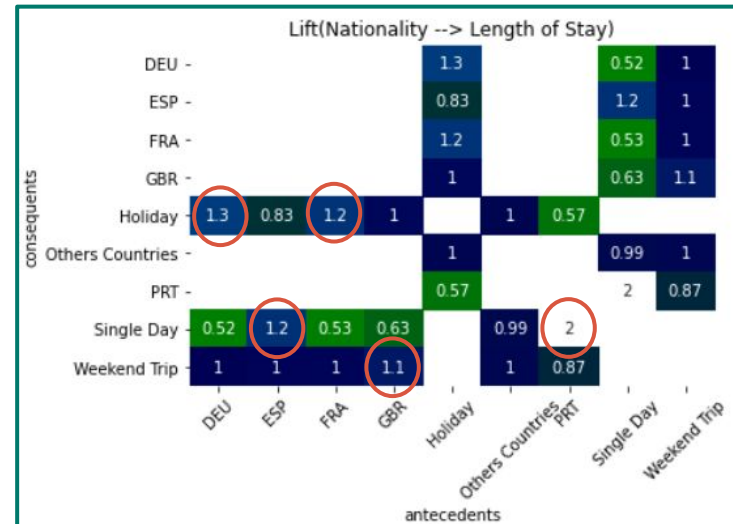
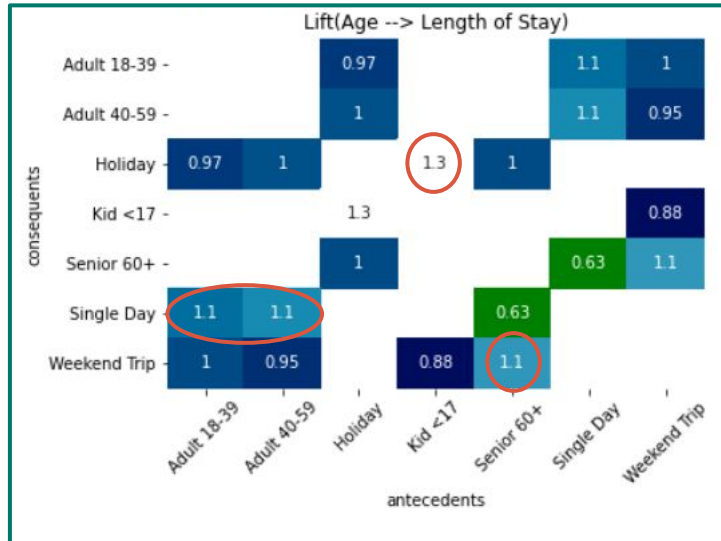
- Customer Loyalty Check by Age, Nationality, and Length of Stay

| | antecedents | consequents | antecedent support | consequent support | support | confidence | lift | leverage | conviction |
|-----|----------------------------------------------|----------------|--------------------|--------------------|----------|------------|----------|----------|------------|
| 183 | (Holiday, Senior 60+) | (Book 2 Times) | 0.065128 | 0.641301 | 0.054603 | 0.838392 | 1.307330 | 0.012836 | 2.219558 |
| 189 | (Weekend Trip, Senior 60+) | (Book 2 Times) | 0.112813 | 0.641301 | 0.089687 | 0.795007 | 1.239679 | 0.017340 | 1.749809 |
| 243 | (Weekend Trip, Others Countries, Senior 60+) | (Book 2 Times) | 0.071157 | 0.641301 | 0.056530 | 0.794444 | 1.238802 | 0.010897 | 1.745024 |
| 43 | (Senior 60+) | (Book 2 Times) | 0.200524 | 0.641301 | 0.158719 | 0.791523 | 1.234246 | 0.030123 | 1.720570 |
| 159 | (Others Countries, Senior 60+) | (Book 2 Times) | 0.128280 | 0.641301 | 0.101102 | 0.788136 | 1.228964 | 0.018836 | 1.693061 |

| | antecedents | consequents | antecedent support | consequent support | support | confidence | lift | leverage | conviction |
|-----|---------------------------------|--------------------------|--------------------|--------------------|----------|------------|----------|----------|------------|
| 192 | (Adult 40-59, Holiday) | (Book 3 Times and above) | 0.164797 | 0.131047 | 0.034837 | 0.211394 | 1.613117 | 0.013241 | 1.101885 |
| 186 | (Adult 40-59, Others Countries) | (Book 3 Times and above) | 0.309730 | 0.131047 | 0.054751 | 0.176771 | 1.348911 | 0.014162 | 1.055542 |
| 18 | (Adult 40-59) | (Book 3 Times and above) | 0.501557 | 0.131047 | 0.086574 | 0.172611 | 1.317166 | 0.020847 | 1.050235 |
| 198 | (Adult 40-59, Weekend Trip) | (Book 3 Times and above) | 0.234669 | 0.131047 | 0.039235 | 0.167193 | 1.275825 | 0.008482 | 1.043403 |

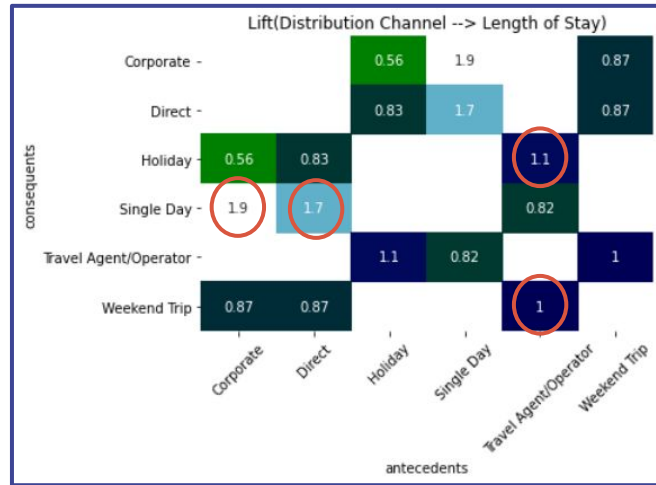
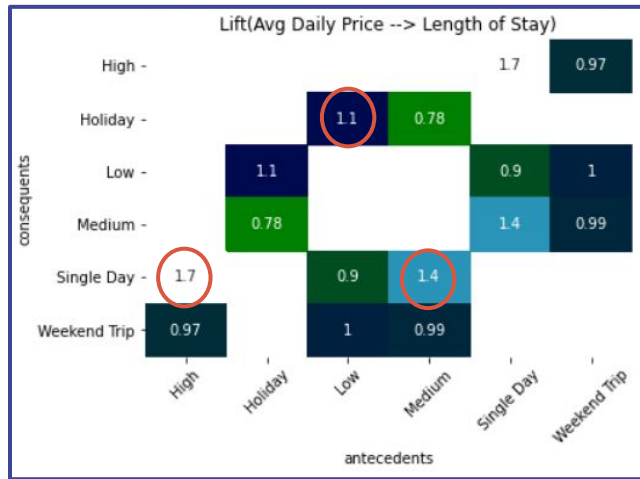
Association Rules

- Length of Stay by Age, Nationality, Average Daily Price Group, Distribution Channel

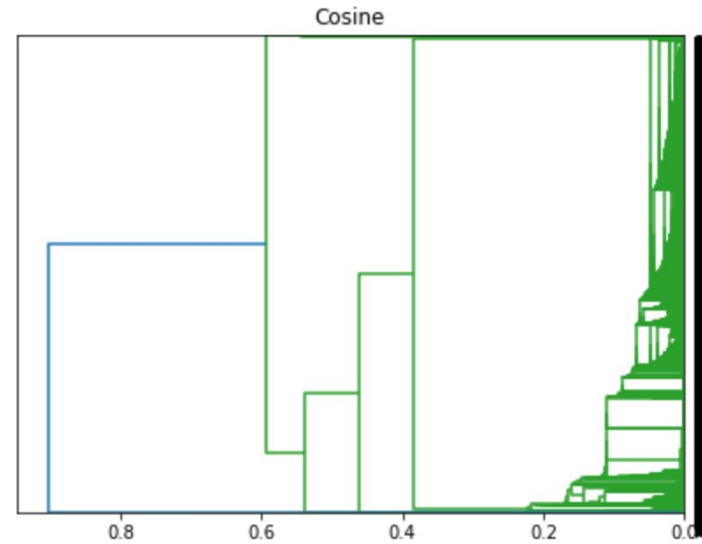
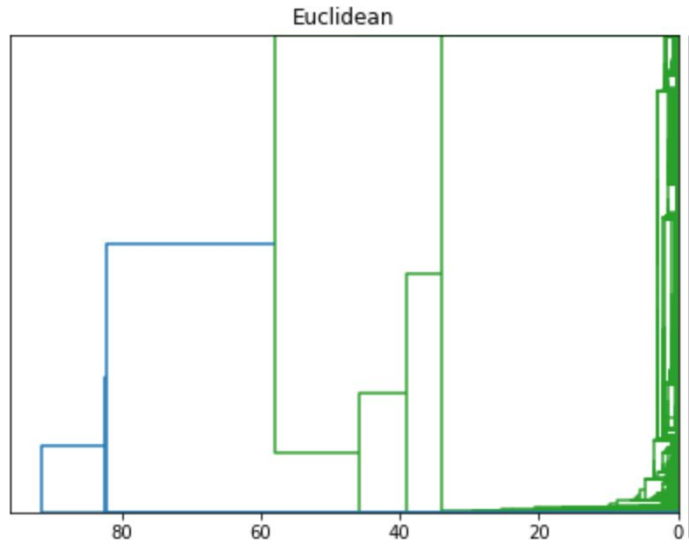


Association Rules

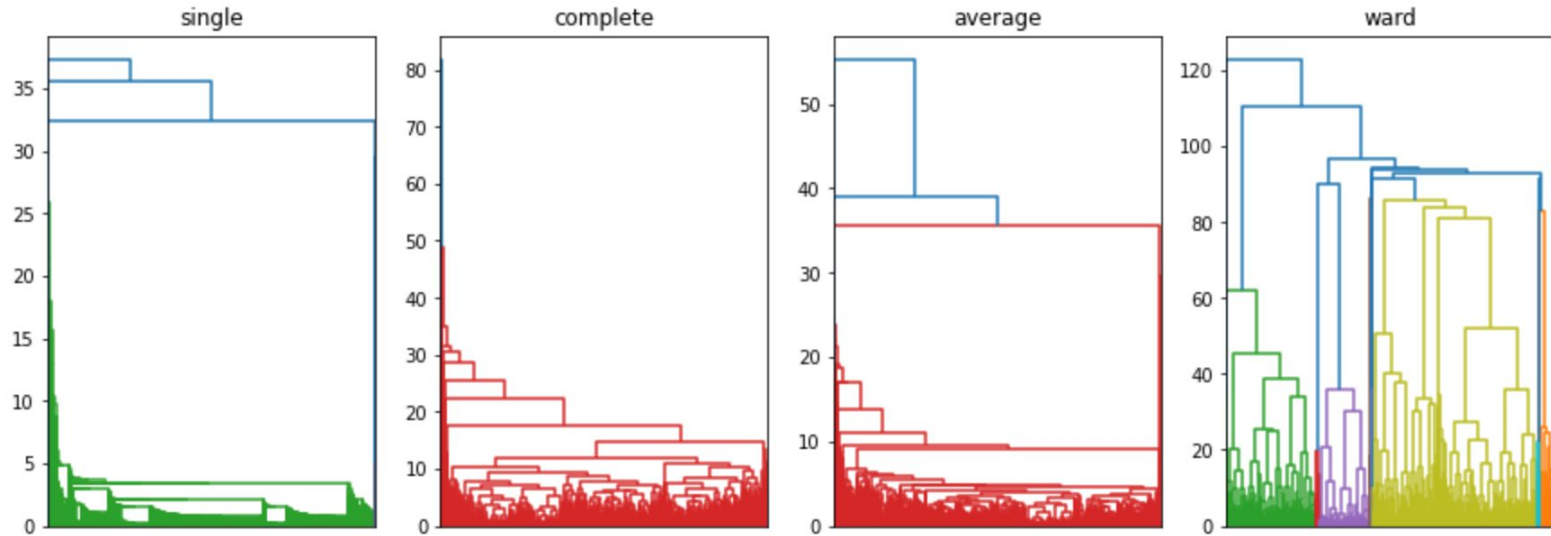
- Length of Stay by Age, Nationality, Average Daily Price Group, Distribution Channel



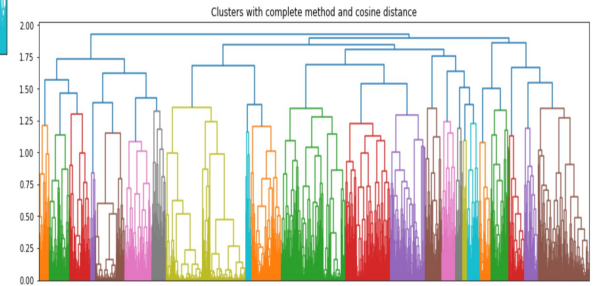
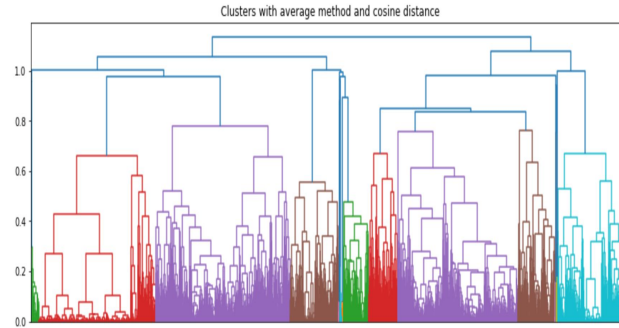
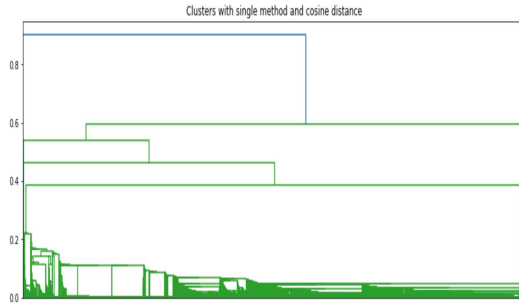
Hierarchical Clustering - Distances



Hierarchical Clustering with Euclidean distance



Hierarchical Clustering with Cosine Distance

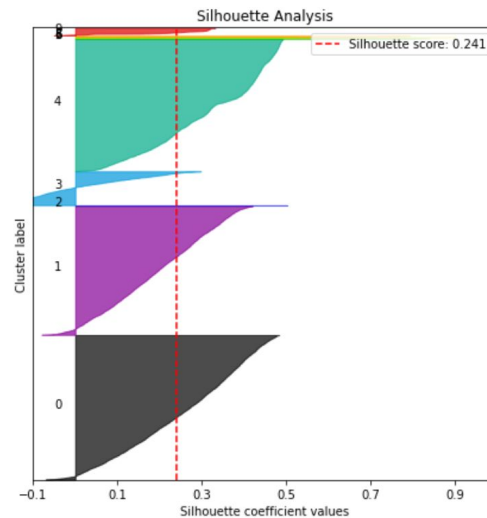
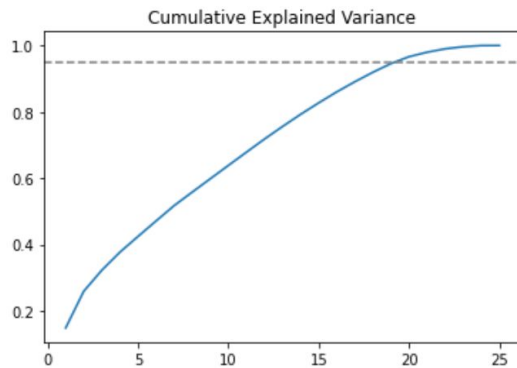
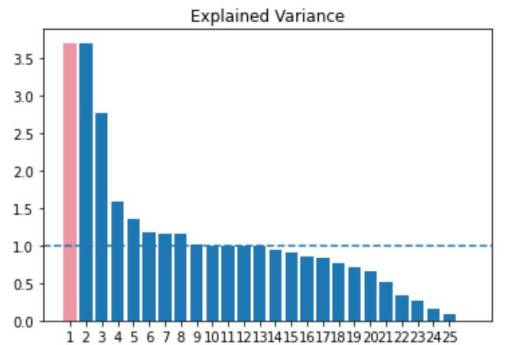


KMeans Baseline

| | |
|----|------|
| 0 | 1067 |
| 1 | 4537 |
| 2 | 1872 |
| 3 | 5 |
| 4 | 4478 |
| 5 | 8 |
| 6 | 222 |
| 7 | 59 |
| 8 | 2 |
| 9 | 11 |
| 10 | 26 |
| 11 | 30 |
| 12 | 4401 |

- No dimensionality techniques utilized
- 13 clusters seemed optimal
- Clusters imbalanced

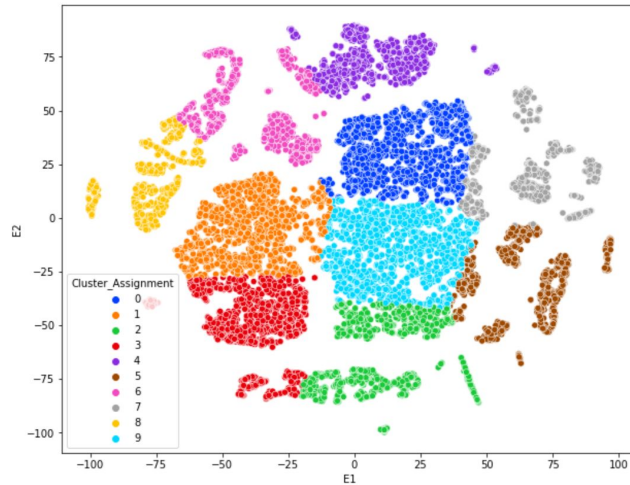
PCA → KMeans



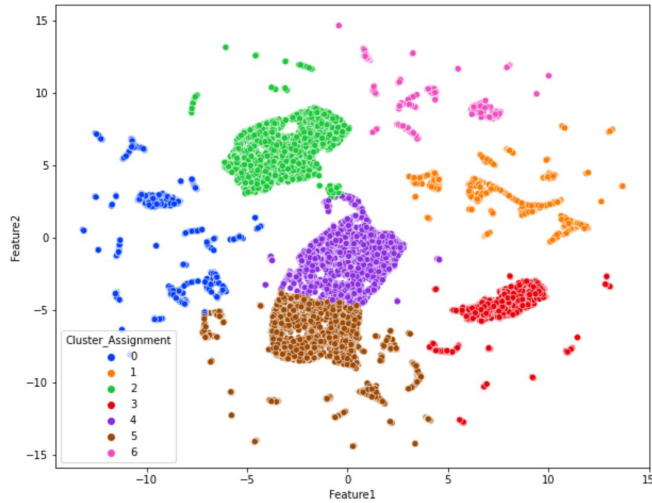
10 Clusters

TSNE vs. UMAP

TSNE



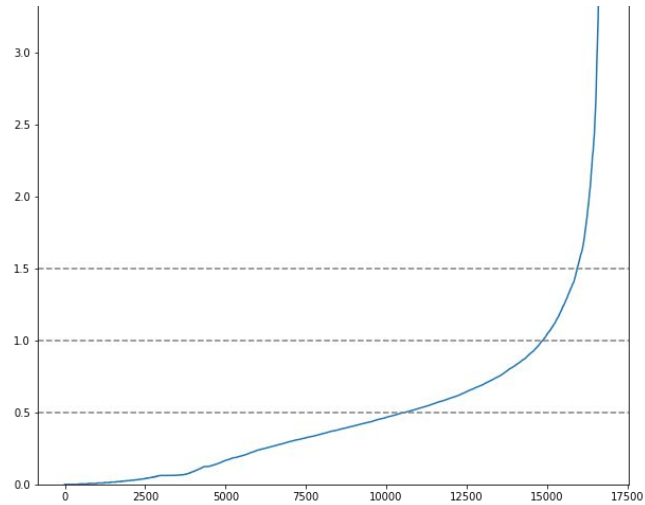
UMAP



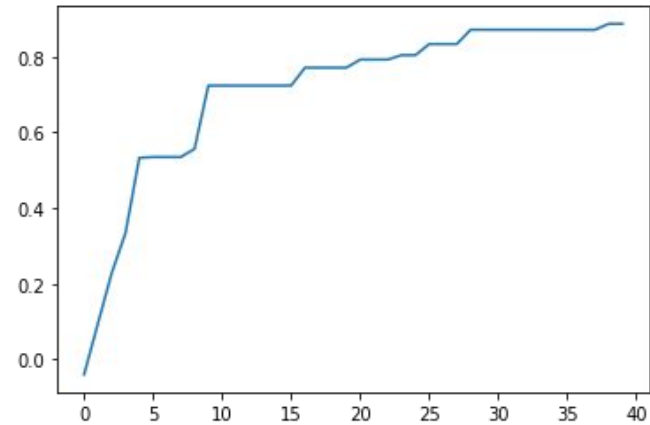
DBSCAN

Parameters: `eps`, `min_samples`

NearestNeighbors distance plot



Silhouette score (varying `eps`)



Business Implications

Overall Findings

- Among the 10 clusters, the average age of overall customers is around 48.
- More than 3 clusters have customers who created their account more than 2 years ago.
- Among the 10 clusters profile, no cluster shows the value 1 in Cancelling/ Not Showed for the reservation.
- With our analysis of lodging revenues and room nights, we realized that the people who spent the least were the people who booked their stay the closest to the actual date.
- People who booked their stay the closest to the actual date were also the most likely to not check in with their reservation afterward.
- The most frequent requests: asking for a king size or a twin size bed, as well as for a quiet room. All the other requests were not used a lot on our sample.

Recommendations

- Since the average customer is middle aged, the hotel should target these people especially (Adults 40-59 are most likely to rebook).
- The hotel is greatly dependent on loyal customers; it is reasonable to suggest the hotel to provide more rewards to attract and maintain those old customers.
- Since people did not show a pattern of canceling or not showing their reservation, this means that the hotel does not have to adopt a credit card mandatory fill form at the moment of the reservation.
- Adjust the price/night based on association rules, provide a higher daily price for customers booking only for 1 night, but a lower price for customers booking for holiday trips (Could add some coupon for customers booking from Travel Agent/Operator).



Thank you!
Questions?