

## Key problems (1/3)

Until 2015 hotel chain C operated 4 hotels, however, with the acquisition of new hotels, the hotel chain board decided to invest more in marketing.

However, it was not until 2018 that the hotel chain created a marketing department and hired a new marketing manager, A.

A realized that the current customer segmentation was not adequate, as it **only reflected one only customer characteristic**, its sales origin. It **did not reflect** geographic characteristics, such as the country of origin, demographic characteristics, such as age, or behavioral characteristics, such as the number of stays.

## Key problems (2/3)

Without proper customer segmentation, it is difficult for A to define a strategy to reach new customers and to continue to captivate the current customers.

Taking into consideration the multiple distribution channels that hotels operate nowadays (travel agencies, travel operators, online travel agencies – OTA, brand websites, meta searchers websites, among others). For example, corporate customers tend to make reservations very near the arrival date, book directly with the hotel, and be willing to pay more for a better-equipped room, while a customer on holiday tends to make reservations more distant from the arrival date, book with a travel operator or OTA, and to look for better price opportunities.

## Key problems (3/3)

Therefore, products “creation”, pricing definitions, and other marketing tasks, such as advertising, must take into consideration the targets of its efforts according to the different channels and groups of customers.



## Data (1/6)

The provided dataset is composed of the following columns:

- **ID:** Customer ID
- **Nationality:** Nationality of the customer in ISO 3166-1 (Alpha 3) format
- **Age:** The age of the customer
- **DaysSinceCreation:** Number of elapsed days since the customer was created
- **NameHash:** Hash of the customer's name
- **DocIDHash:** Hash of the customer's personal document identification number (usually a passport or ID card)
- **AverageLeadTime:** Average number of days before arrival date the customer makes bookings

## Data (2/6)

- **LodgingRevenue:** Total amount of lodging revenue paid by the customer so far
- **OtherRevenue:** Total amount of other revenue (e.g., food & beverage, spa, etc.) paid by the customer so far
- **BookingsCanceled:** Number of bookings the customer made but subsequently canceled
- **BookingsNoShowed:** Number of bookings the customer made but subsequently made a "no-show"
- **BookingsCheckedin:** Number of bookings the customer made, which actually ended up staying



## Data (3/6)

- **PersonNights:** Total person/nights the customer has stayed at the hotel so far. Persons/Nights are the sum of *Adults* and *Children* in each booking, multiplied by the number of *Nights* (Length-of-stay) of the booking
- **RoomNights:** Total of room/nights the customer has stayed at the hotel so far. Room/Nights are the multiplication of the number of rooms of each booking by the the number of *Nights* (Length-of-stay) of the booking
- **DistributionChannel:** Distribution channel normally used by the customer to make bookings at the hotel
- **MarketSegment:** Current market segment of the customer
- **SRHighFloor:** Indication if the customer usually asks for a room in a higher floor (0: No, 1: Yes)

## Data (4/6)

- **SRLowFloor:** Indication if the customer usually asks for a room in a lower floor (0: No, 1: Yes)
- **SRAccessibleRoom:** Indication if the customer usually asks for an accessible room (0: No, 1: Yes)
- **SRMediumFloor:** Indication if the customer usually asks for a room in a middle floor (0: No, 1: Yes)
- **SRBathtub** Indication if the customer usually asks for a room with a bathtub (0: No, 1: Yes)
- **SRShower:** Indication if the customer usually asks for a room with a shower (0: No, 1: Yes)
- **SRCrib:** Indication if the customer usually asks for a crib (0: No, 1: Yes)



## Data (5/6)

- **SRKingSizeBed:** Indication if the customer usually asks for a room with a king size bed (0: No, 1: Yes)
- **SRTwinBed** Indication if the customer usually asks for a room with a twin bed (0: No, 1: Yes)
- **SRNearElevator:** Indication if the customer usually asks for a room near the elevator (0: No, 1: Yes)
- **SRAwayFromElevator:** Indication if the customer usually asks for a room away from the elevator (0: No, 1: Yes)
- **SRNoAlcoholInMiniBar:** Indication if the customer usually asks for a room with no alcohol in the mini bar (0: No, 1: Yes)



## Data (6/6)

**SRQuietRoom:** Indication if the customer usually asks for a room away from the noise (0: No, 1: Yes)

NOTE: All time-based columns (e.g., *Age* or *DaysSinceCreation*) were calculated at the dataset extraction date.

- 
1. Explore the data and identify the variables that should be used to segment customers
  2. Use K-Means clustering to identify customers segments
    1. Justify your selection of K (taking into consideration the business use)
    2. Use PCA to reduce dimensionality and speed-up model development
  3. Suggest business applications for the findings