

# CUSTOMER PROFILING

## Executive Summary

Customer profiling is crucial for hotel chain to gain a deeper understanding of its guests. This deeper understanding will, in turn, empower Hotels to deliver highly personalized services, enhance guest satisfaction, and optimize revenue generation. This project focuses on using machine learning to segment different customers into groups and build profile of each group. For this purpose, this project has compiled a detailed dataset from a hotel on the Kaggle, including booking details, customer information, and reservation specifics. The project has used DBSCAN for clustering. The results show that 3 main groups of customer are found, each group has its own characteristics. This project provides insights into leveraging machine learning techniques to create customer profiles, thereby helping to develop appropriate strategies for each customer group.

## Introduction

In today's hospitality industry, competition is increasing, as customers have more choices and demand more personalized experiences. With the development of technology and big data, understanding the behavior, preferences and needs of each customer group has become an important factor for hotels to maintain a competitive advantage. However, many hotels still face challenges in analyzing and leveraging customer data to provide appropriate services and improve satisfaction.

Customer profiling not only helps hotels categorize and understand customers better, but also supports in optimizing marketing strategies, resource management and service personalization. This is especially important in building customer loyalty, increasing return rates and improving business efficiency. With these challenges and opportunities, the project focuses on analyzing customer data and using machine learning to create detailed profiles, thereby supporting hotels to provide the best experience and meet the increasingly diverse needs of the market.

# Objectives

## A. Project Objectives

The primary aim of this project is to create comprehensive and accurate customer profiles for a Hotels by harnessing the power of historical data analysis. These profiles will encompass critical aspects such as demographics, booking history, preferences, and feedback, enabling the hotel chain to gain a deeper understanding of its guests. This deeper understanding will, in turn, empower Hotels to deliver highly personalized services, enhance guest satisfaction, and optimize revenue generation.

Major objectives are listed below:

- ***Data collection and analysis:*** This objective involves gathering data from various sources within LuxuryStay Hotels, ensuring data quality and consistency. It includes cleaning, validating, and preparing the data for analysis.
- ***Data modelling with unsupervised machine learning:*** This objective involves applying unsupervised machine learning techniques, such as clustering and dimensionality reduction, to uncover patterns and insights within the data.
- ***Customer profiling and analysis:*** The final objective is to create comprehensive customer profiles that encompass demographics, preferences, behaviors, and feedback.

## B. Key Deliverables

The project will deliver the following key deliverables:

- **Data analysis:** This project will provide a detailed analysis of the data collected, including trends, patterns, and correlations identified.
- **Data visualization:** Visualization of key findings will be provided to enhance understanding and facilitate communication of complex data insights.
- **Customer profile:** The project will provide different groups of customer and their characteristics based on the result of cluster model.

# Methodology

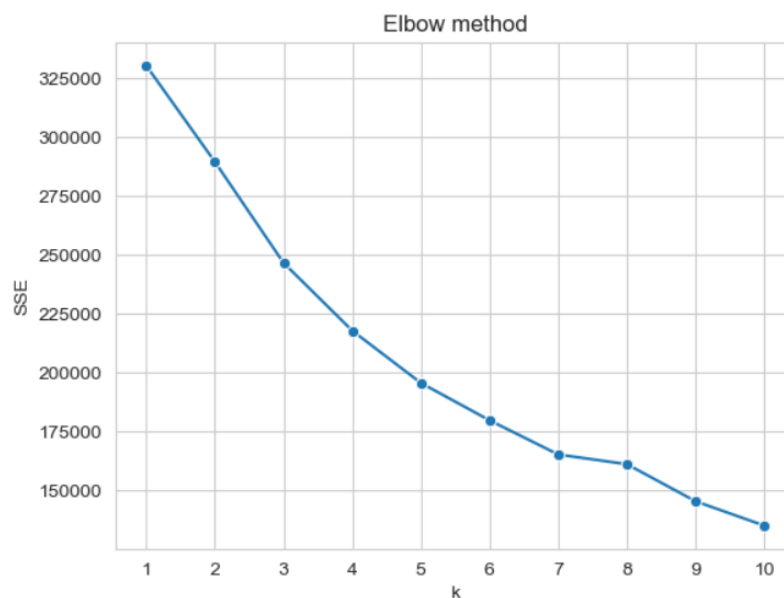
## A. Data Collection

Data was collected from **Kaggle**, ensuring completeness and accuracy. Contains 17 fields in total and 36285 rows, related to booking details, customer information, and reservation specifics. The information has been meticulously gathered from real-world hotel booking scenarios, ensuring authenticity and relevance for building customer profile.

## B. Data Analysis Techniques

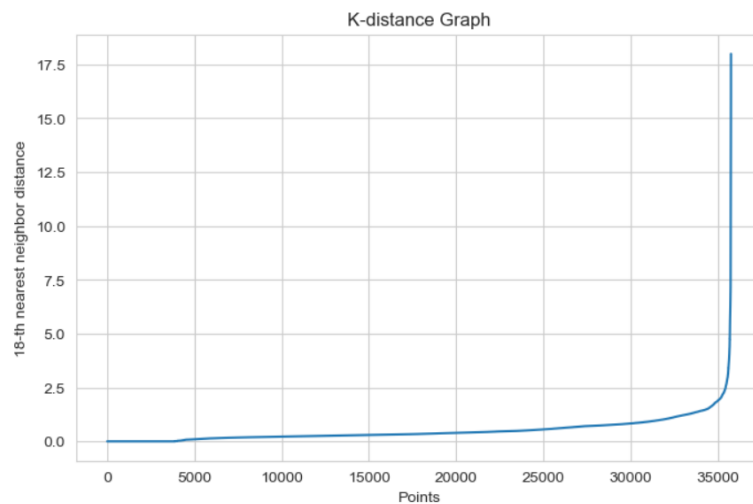
Our project employed advanced data analysis techniques, including **Descriptive Analysis**, **Cluster Analysis**,... to uncover patterns, trends, and correlations within the data. These techniques were chosen based on the nature of the data and the objectives of the analysis. By leveraging these sophisticated methods, we aimed to extract valuable insights and actionable recommendations to drive informed decision-making and strategic planning.

## Key Findings



It seems that SSE starts to move sideways from 7 to 8 clusters, which is a fairly large number of clusters. We hope to have only about 2-4 clusters so that we can easily come up with strategies suitable for each group. Having so many clusters can lead to having clusters with very few numbers, while other clusters have very many numbers.

In part, this can be affected by outliers as K-Means is not a good model for handling outliers. Therefore, we will use another model, DBSCAN.



With DBSCAN, we need to determine 2 main parameters: eps and min\_samples. With min\_samples, we usually choose larger than the number of data dimensions:  $\text{min\_sample} = D + 1$  where  $D$  is the number of data dimensions, ideally choose  $2 * D$ . In this case, we have  $D = 9$ , so  $\text{min\_sample} = 18$ .

To choose eps, we will use the k-distance chart. The break point of the graph will be chosen as the eps of the model. In the chart above, we see the break point is around 2.4.

## Results

### cluster 1:

***This group is mainly short-term tourists or business travelers:***

- Most guests are adults, traveling alone or in pairs (few children are accompanied).
- The length of stay is short (2-3 nights) and mainly on weekdays, indicating that guests may be on business trips.
- The average spend per room is reasonable (around 100-120 units), indicating that this group tends to choose standard rooms rather than premium rooms.
- The early booking period (average 3 months) indicates that customers plan in advance, not spontaneously.

- This is a new group of guests, with a low return rate (only 1.39%). They may not have had many special experiences or impressions of the service to return.
- The number of special requests is low (less than 1 request) and very few guests request parking.

**cluster 2:**

***This is a group of families with children and short-term stays of 2-3 nights.***

- They tend to plan ahead, booking rooms about 2 months before the trip.
- Although they are new guests, they choose rooms with higher prices, suitable for families who need more space or good service.
- Special requests from this group (nearly 1 request per guest) may be related to services for children or additional amenities.

**cluster 3:**

***This is a loyal group of customers:***

- This is a group of repeat guests (100% regular guests), often traveling alone or in pairs and without children.
- They tend to stay short-term (1-2 nights), focusing on weekdays rather than weekends.
- This group often books rooms close to the departure date (about 3 weeks before the trip).
- This is a group of guests with high savings, prioritizing lower-priced rooms (average ~80).
- They have a history of cancellations, but are generally still inclined to return to book.

**cluster noise:**

- This is a very diverse group of guests, including individuals, families, and large groups. Their staying behavior and spending power vary significantly, from cheap short-term stays to expensive long-term stays.
- This group has a very high cancellation history, and is unstable in implementing plans.
- Although the rate of returning customers is quite high (66%), inconsistent and unpredictable behavior reduces trust.

# Conclusion

In summary, this project successfully leveraged various statistical and machine learning techniques to solve a real-world problem in the context of hotels. The goal was to create comprehensive and accurate customer profiles for a Hotels by harnessing the power of historical data analysis. These profiles will encompass critical aspects such as demographics, booking history, preferences, and feedback, enabling the hotel chain to gain a deeper understanding of its guests. This deeper understanding will, in turn, empower Hotels to deliver highly personalized services, enhance guest satisfaction, and optimize revenue generation..

The project followed a well-defined workflow, including data collection, cleaning and pre-processing, exploratory data analysis, model building, and evaluation. Each stage was approached systematically, using tools such as Python and visualization libraries to facilitate the analysis and modeling process.

As the project came to an end, there were many opportunities for further optimization and refinement. By incorporating additional data and exploring advanced techniques, model performance can be improved, leading to better results in customer profiling.

In summary, this project successfully addresses the problem of customer profiling, highlights the value of data science methods, and demonstrates the benefits of a modular architecture for implementation. The insights gained from this project can inform decision making and contribute to improving the operations of hotels.