

Business Cases with Data Science

Case 1: Hotel Customer Segmentation

**Master Degree Program in
Data Science and Advanced Analytics**

March, 2024
Group F

GROUP F

Bernardo
Pinto Leite
Nrº: 20230978

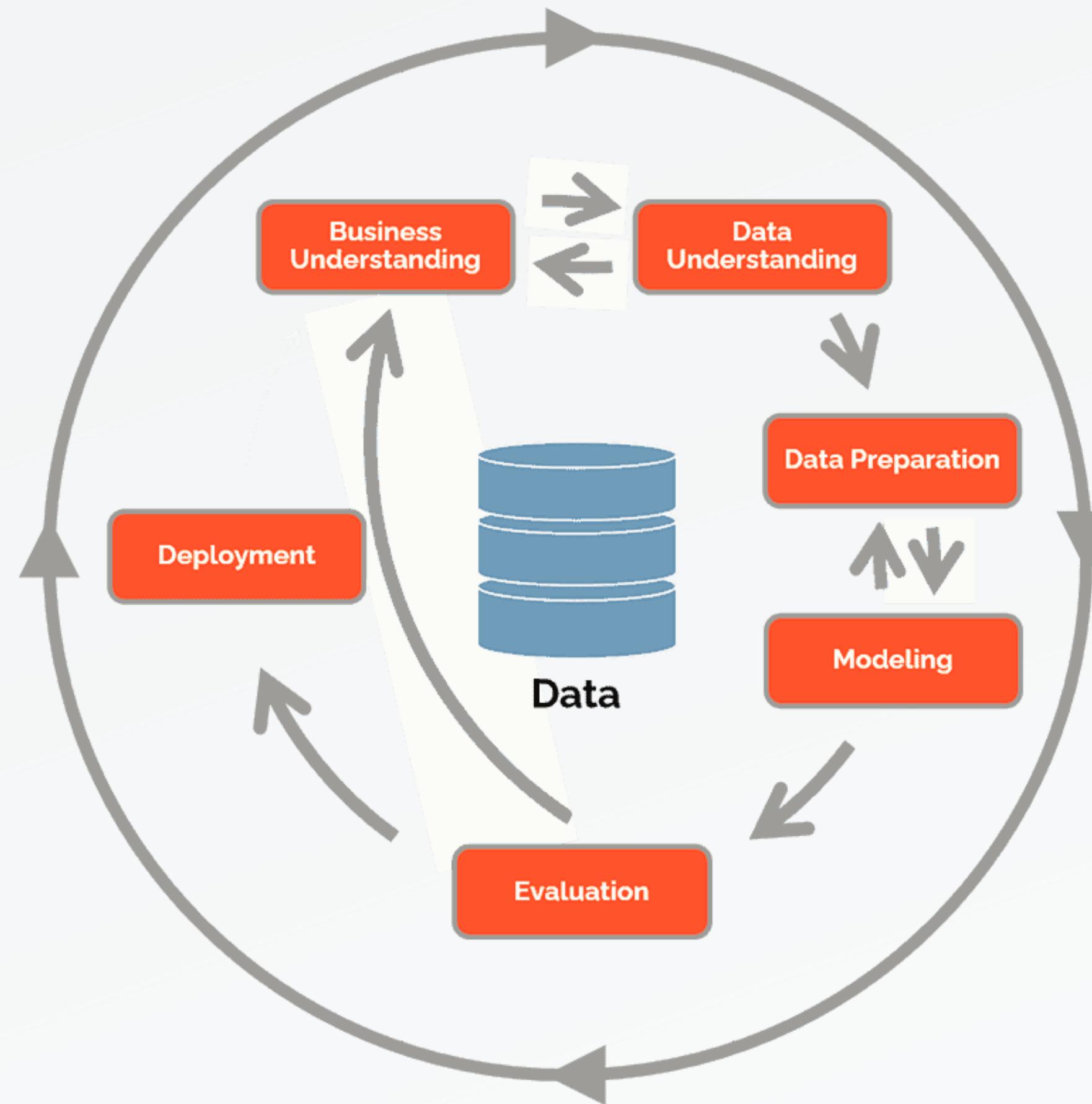
Emília
Santos
Nrº: 20230446

Nicolás
Zerené
Nrº: 20230779

Ricardo
Kayseller
Nrº: 20230450

Steph
Kuznetsov
Nrº: 20231002

METHODOLOGY: CRISP-DM



BUSINESS NEEDS



Customer Segmentation



New Marketing Strategies



BUSINESS OBJECTIVES

Enhance Customer Satisfaction

Optimize Marketing Efforts

Increase Revenue

Define Strategic Decisions

Improve Product/Service Offerings

Drive Operational Efficiency

Leverage Competition Advantage



BUSINESS SUCCESS CRITERIA

Excellent Service

Management

Security

DATA PREPARATION

01

VARIABLE AGE

Under 16 and over 90 excluded
"Age" into 6 groups

02

SERVICE REQUEST VARIABLES

Variables with over 99% zeros were excluded for enhancing analytical accuracy

03

OUTLIERS

Outliers in some variables were removed using IQR and boxplot methods, retaining 99%

04

CORRELATED VARIABLES

High correlation:
"PersonsNights",
"MarketSegment" dropped

DATA PREPARATION

05

06

07

REDUCING NUMBER OF CATEGORIES

"Nationality" encoded to the top 6, with others grouped into one

"DistributionChannel" to 2 main categories

NEW VARIABLES

"CancellationRate", "TotalSpecialRequests" and "TotalRevenue"

DUPLICATED VALUES

We enhanced data integrity by removing 111 duplicates and fix inconsistencies in "DocIDHash," ensuring accuracy

MODELING



PCA

Variability/Dimensionality



Kmeans --> Distortion Score

**Importance of Silhouette

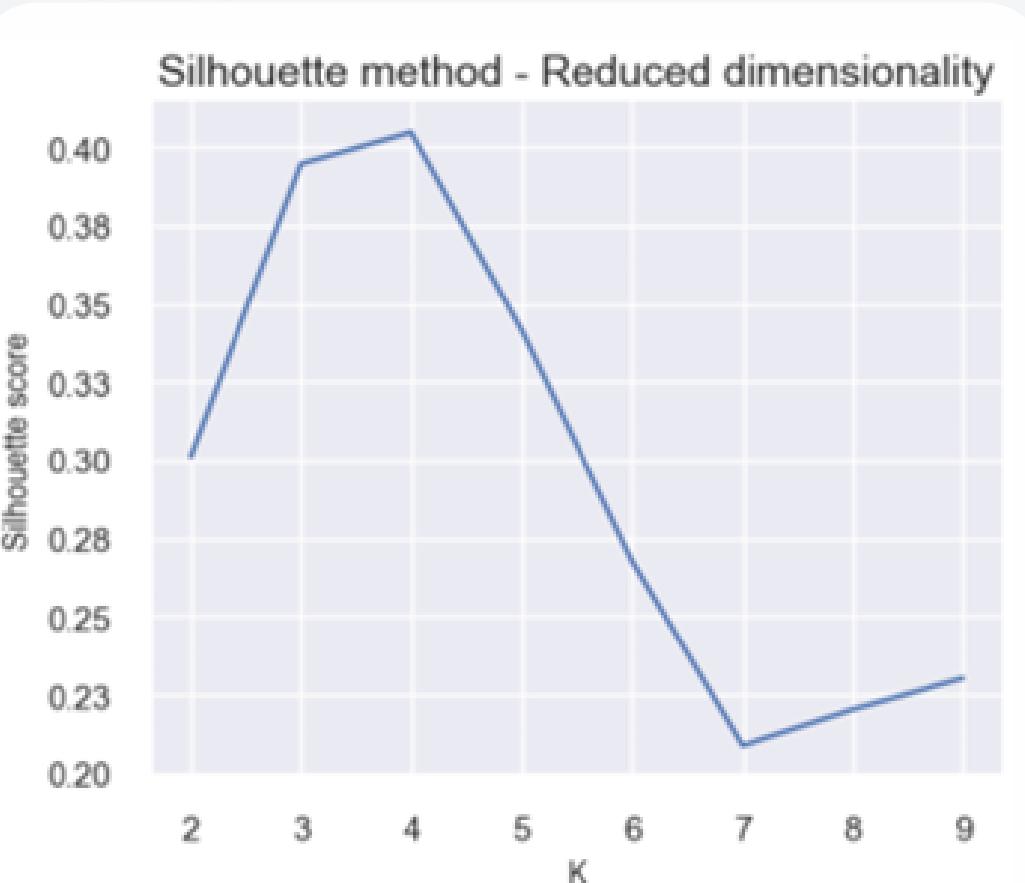


EVALUATION

PCA

36 ---> 18 dimensions

97.63% of the cumulative variance explained



18 dimensions

4 clusters

Cardinality and Magnitude correlated

K-MEANS

Segmentation by 3 perspectives:

5 clusters

Demographic,
Behavioral and
Geographic

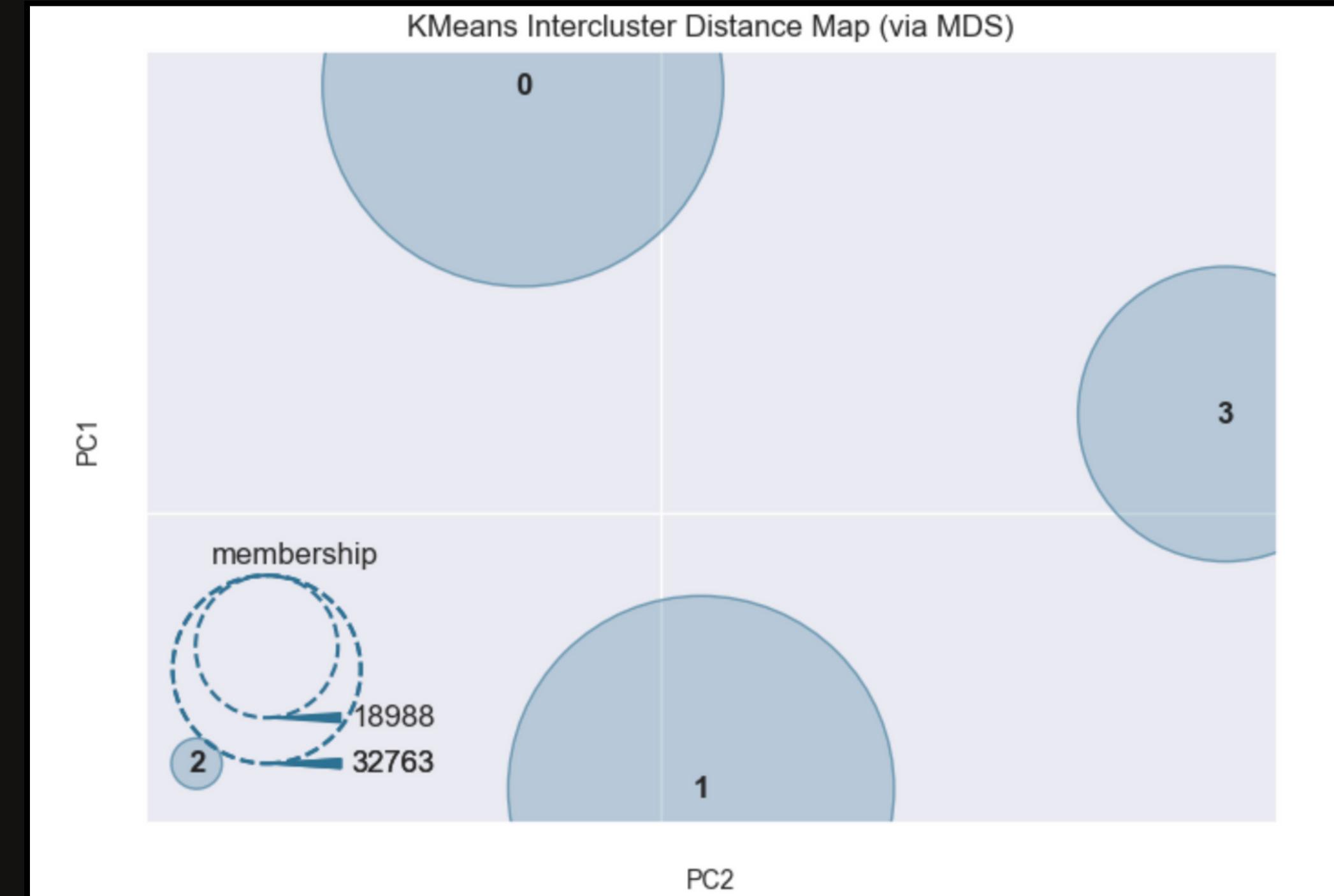
Silhouette score ~ 0.2

K-PROTOTYPES

CLUSTER ANALYSIS

4 clusters

4 CLUSTERS



CLUSTER 0

36174
Individuals

- Young workers, between ages 24 and 35
- Specific Special Requests
- Describes majority of the customers



CLUSTER 1

32763

Individuals

- Mostly French and German Nationality
- Spends the least money: ~342,88€
- Book through Travel Agencies/Operators



CLUSTER 2

12124

Individuals

- Mainly Portuguese people
- Ages between 56 and 65
- Book directly or other method
- Spends the most money (~372,76€)
- Highest number of Check-ins



CLUSTER 3

18988

Individuals

- Mix Nationality
- Ages <25 and >65
- The least number of Check-ins
- Requests: Twin beds, High Floors and 1 specific request



DEPLOYMENT PLAN

3 WEEKS

Preparation

BUSINESS
ALIGNMENT

DATA
ANALYSIS

STRATEGY
DEVELOPMENT

1 MONTH

Marketing
Goals

SURVEYS

SOCIAL
MEDIA
ENGAGEMENT

5 MONTHS

Deployment

PERFORMANCE
EVALUATION

SEO, SEM

FULL SCALE
LAUNCH

DAILY CYCLE

Maintenance

MONITORING

REGULAR
FEEDBACK

IMPROVEMENTS



MAINTENANCE PLANS



**International
middle-aged
business travelers:**

Special business
amenities and
discounts
Strategy to partner
with travel agencies

CLUSTER 1



**Budget-Conscious
Travelers:**

Target with budget-
friendly packages
Preferences for
quieter
accomodations

CLUSTER 2



**Senior Luxury
Regular Customers:**

Provide direct
booking incentives
Premium services
and exclusive offers

CLUSTER 3



**Younger and Senior
Customers:**

Staff training for
proactive
preparations and
customized services
Advance Booking

CLUSTER 4

FUTURE MODEL IMPROVEMENT

Collection of New Types of Information

- Customer Feedback
- Socioeconomic Data
- Travel Preferences
- Online Behavioral Data

Improvements in Data Quality

- Data Standardization
- Real-time Data Validation
- Regular Data Audits
- Collaborative Data Sharing Initiatives

Business Cases with Data Science

Case 1: Hotel Customer Segmentation

Master Degree Program in
Data Science and Advanced Analytics

March, 2024
Group F