

BUSINESS CASE 1:

HOTEL CUSTOMER SEGMENTATION

André Moreira, 20222132
Beatriz Teixeira, 20211313
Catarina Nunes, 20230083
Pedro Catarro, 20230463

Presentation Structure

1. Business Understanding

2. Data Understanding

3. Data Preparation

4. Modeling

5. Evaluation

6. Deployment

7. Conclusions

1. Business understanding

Background & Context

- **Lisbon, Portugal**
- **Hotel industry**
- **Competitive environment**
- **Growth of potential customers**

Business Problem

- **Outdated customer segmentation**

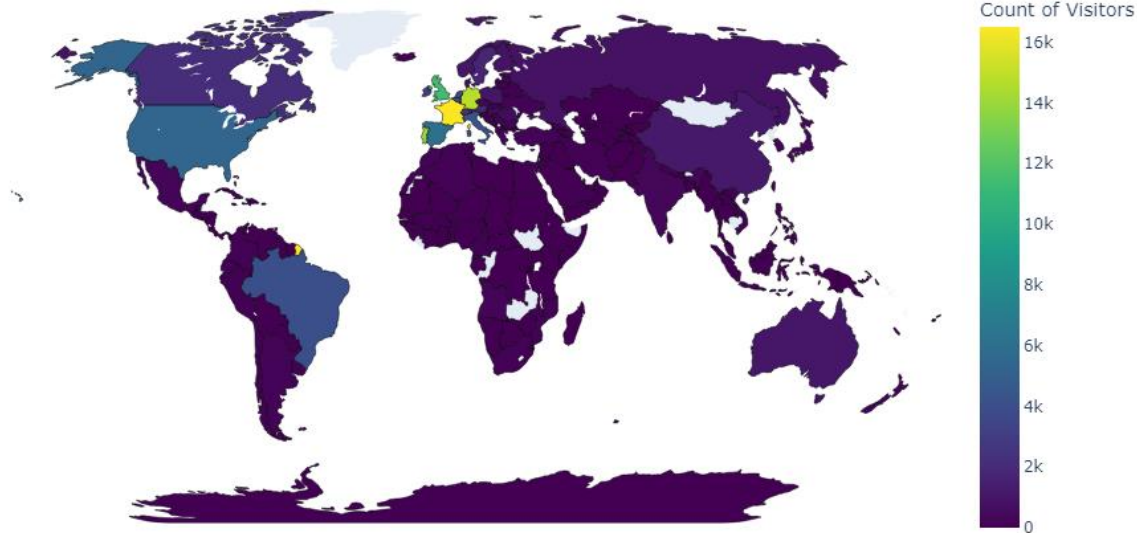


2. Data Understanding:

Univariate EDA:

Our Costumers – Who are they?

Visual Representation of Nationality of Customers



● FRA: 15%

● DEU: 13%

● PRT: 13%

● GBR: 10%

[45, 55] Years: 25%

[35, 45] Years : 20%

[55, 65] Years : 18%

[25, 35] Years : 18%

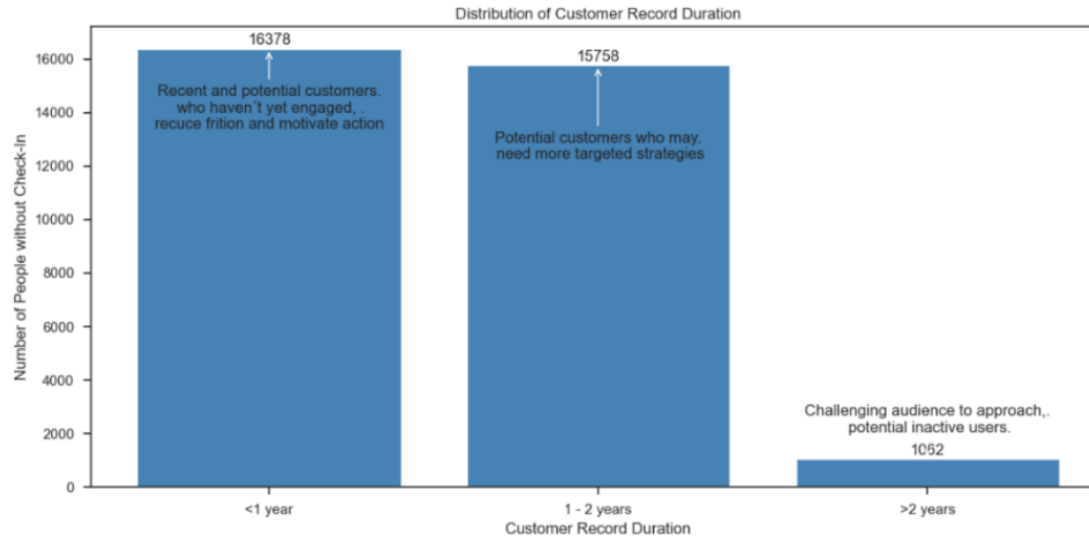
[65, 100] Years : 14%

[18, 25] Years : 6%

2. Data Understanding:

Bivariate EDA:

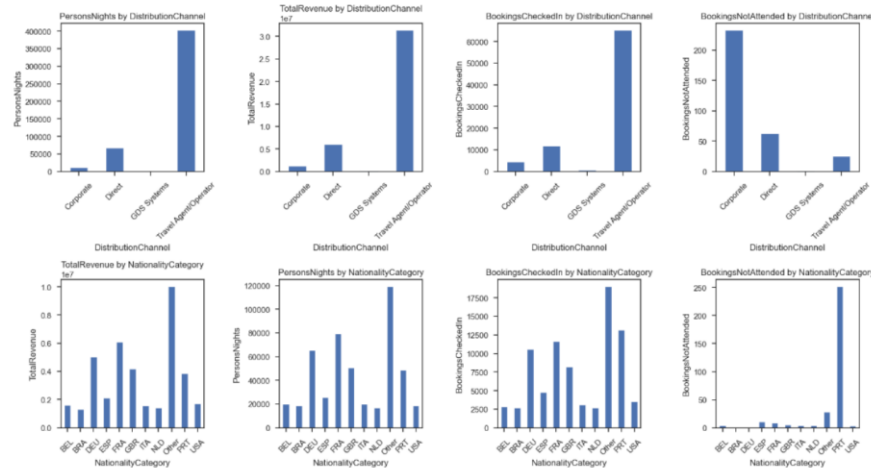
- Potential Customers**



2. Data Understanding:

Bivariate EDA:

- **Distribution Channel vs BookingsNotAttended**
- **Nationality vs BookingsCheckedIn/BookingsNotAttended**



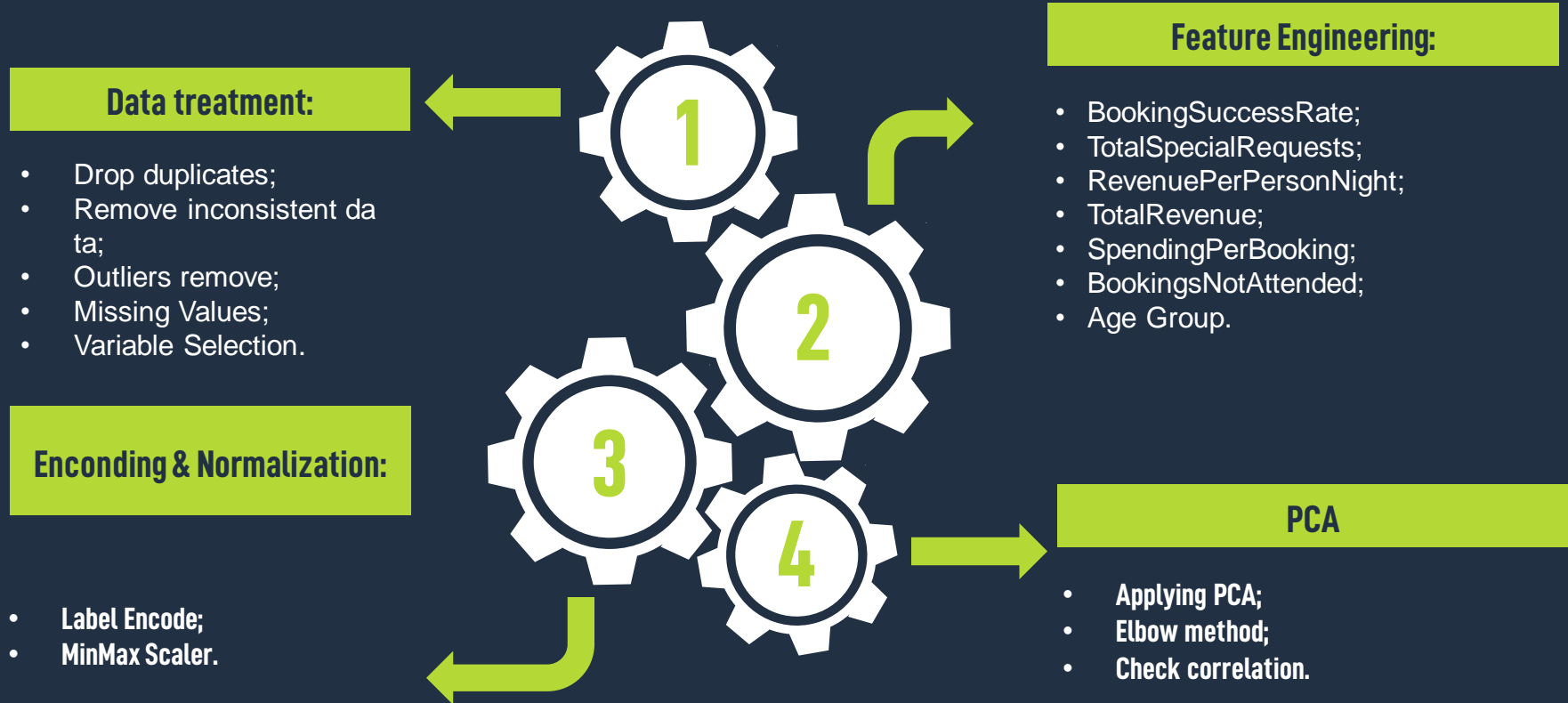
2. Data Understanding:

Bivariate EDA:

Old Segmentation – Why it's no longer serving us?

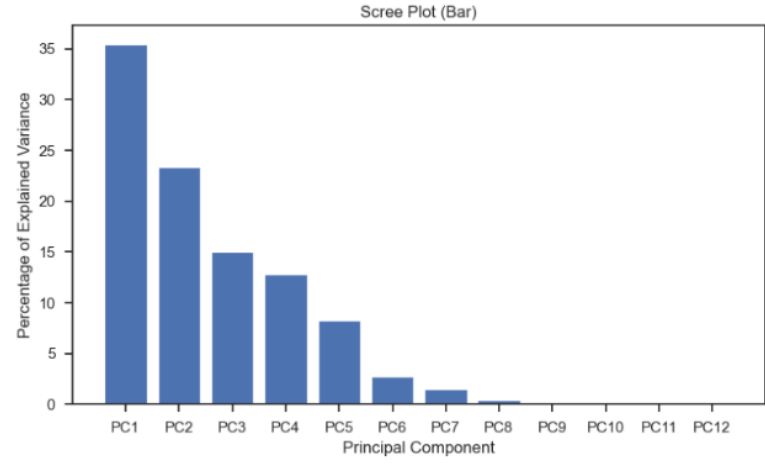
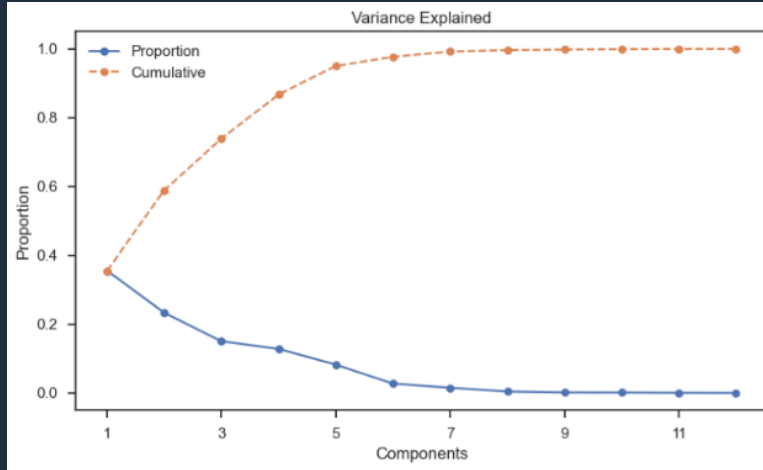
- Competitive landscape demands detailed segment insights beyond sales origin.
- Segments should reveal client and market characteristics to inform strategy alignment.
- Current data shows reliance on Travel Agents and Operators as primary channels.
- Old segmentation is outdated, heavily skewed towards a non-specific 'Other' category.
- "Complimentary" segment.

3. Data Preparation



3. Principal Components Analysis

Principal Components = 5



Threshold of 0.95 (cumulative explained variance of at least 95% of the data variance)

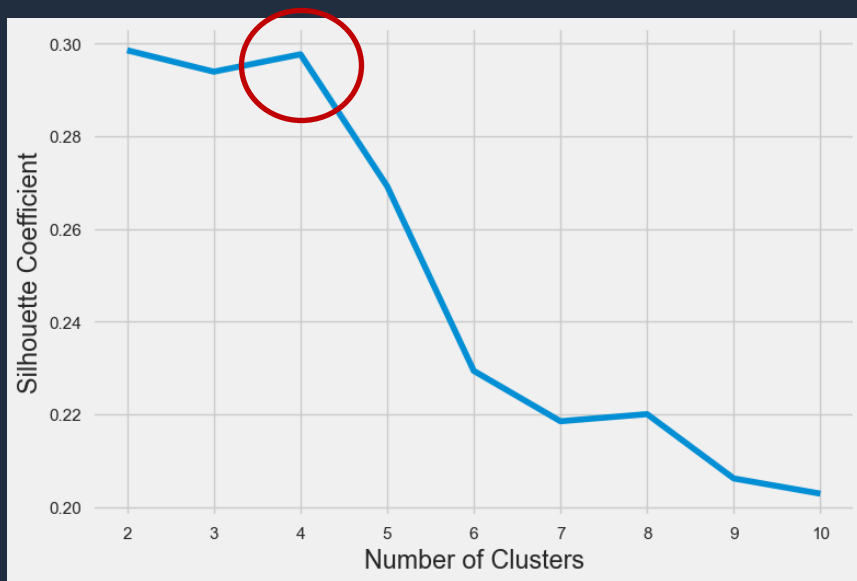
Achieved better results with four components (PCA_Component_1, PCA_Component_2, PCA_Component_3, and PCA_Component_5), as PCA_Component_4 overlaps in role with PCA_Component_3, primarily explaining age variance.

4. Modeling: K-Means

K = 4

For K = 4 The average silhouette_score is : 0.30

Optimal Clusters: Silhouette coefficient suggests 4 clusters, though score is suboptimal.

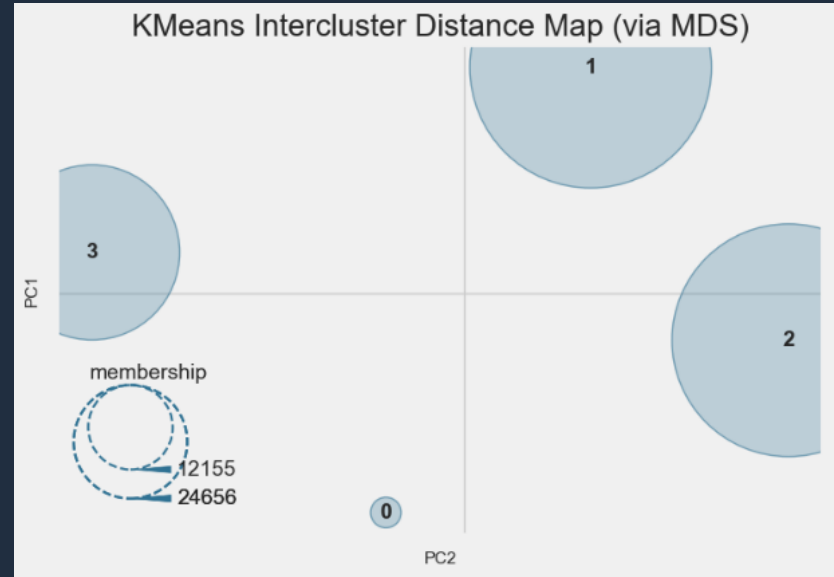


4. Modeling: K-Means

K = 4

Cluster Sizes: Varied cluster sizes are expected, but imbalances like a smaller cluster (e.g., Cluster 0) may require attention for potential **niche segments**.

Average Centroid Distances: Lower distances imply tighter clusters, indicating high similarity among data points. Cluster 1 and 2 show the lowest distances, indicating cohesive clusters.



5. Evaluation:

Cluster 0: Young Elite

**Young;
Mostly European tourists;
Group travelers (or more
nights);
Big spenders.**

Cluster 2: New Corporate Clients

**Middle-aged;
Portuguese /locals;
Recent customers;
Corporate-oriented
Short lead time.**

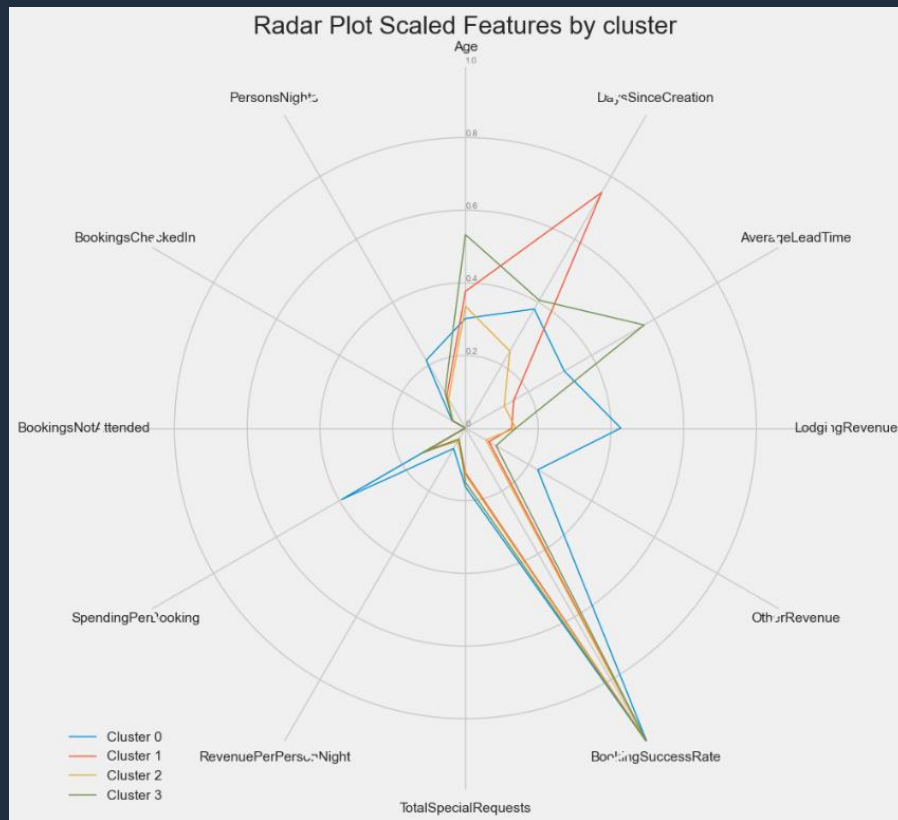
Cluster 1: Established Guests

**Middle-aged;
Long-term customers;
Portuguese /locals;
Standardized preferences.**

Cluster 3: Experienced Planners

**Elderly;
Tourists (predominantly from Germany);
Reliable;
Advanced planners.**

5. Evaluation:



4 Clusters

**Results based on the mean
of each variable per cluster**

6. Deployment

Young Elite

- Web approach
- Create partnerships
- Booking price

New Corporate Clients

- Aim for a second booking
- Partnership with companies

Established Guests

- Discount on bookings in advance
- Outside advertisement
- Room adaptation

Experienced Planners

- Traditional advertisement
- Personalized customer service
- Loyalty benefits

6. Deployment

Smart Helper

Steps to help

1. Input customer information
2. Assign customers to clusters
3. Tailored marketing strategies

```
Enter value for Age: 35
Enter value for DaysSinceCreation: 500
Enter value for AverageLeadTime: 60
Enter value for LodgingRevenue: 800
Enter value for OtherRevenue: 100
Enter value for BookingsCheckedIn: 2
Enter value for PersonsNights: 8
Enter value for SRHighFloor: 0
Enter value for SRCrib: 0
Enter value for SRKingSizeBed: 1
Enter value for SRTwinBed: 0
Enter value for SRQuietRoom: 1
Enter value for BookingSuccessRate: 0.9
Enter value for TotalSpecialRequests: 2
Enter value for RevenuePerPersonNight: 100
Enter value for SpendingPerBooking: 900
Enter value for BookingsNotAttended: 0
Enter value for Origin_North_America: (0/1) 0
Enter value for Origin_Other_Europe: (0/1) 1
Enter value for Origin_Others: (0/1) 0
Enter value for Origin_Portugal: (0/1) 0
Enter value for Origin_South_America: (0/1) 0
Enter value for DistributionChannel_Corporate: 0
Enter value for DistributionChannel_Direct: 1
Enter value for DistributionChannel_GDS Systems: 0
Enter value for DistributionChannel_Travel Agent/Operator: 0
Cliente atribuído ao Cluster: 2
Marketing Strategies:
- Offer discounts on second bookings to encourage recurrence
- Establish direct deals with corporate clients' companies
- Highlight cultural diversity in marketing materials
- Offer multilingual customer support
```

6. Conclusion

Limitations and Potential Improvements

- Implement a policy requiring the collection of age documentation and nationality for all customers – Help with missing values and Duplicates.
- Incorporate additional information: **number of nights spent** or the **average duration of stay**.
- Introduce **time variables**, such as the month of stays, to enable the analysis of seasonality trends.

