Prediction of Hotel H2 booking cancellations

Data Vision Analytics

Problems

High cancellation rate impacts revenue

Overbooking leads to reallocation costs

Restrictive cancellation policies decrease bookings

Deliverables

GOALS

SHORT TERM

1

Develop a predictive model for cancellations

2

Implement a revised over booking policy

LONG TERM

3

Optimize demand, minimize cancellations, maximize revenue

RESULTS

Reduce cancellation rate to **20%**

Minimize revenue loss

Summary

01

Data Understanding

Gained insights from the dataset.

Identified relevant features and examined their relationships.

02

Data Preparation

Cleaned data and engineered features

Ensure data quality and improve model performance.

03

Modeling

Selected, trained, evaluated, and fine-tuned models for predicting cancellations.

04

Deployment

Integrated & implemented the model into the hotel chain's processes, monitored its performance.

Data preprocessing

01

Duplicates and Coherence Check

02

Missing Values

Children, Country, Agent, Company

03

Feature Engineering

Length of Stay, Total Bookings, Time in the System, Room Change Status, Booking Date

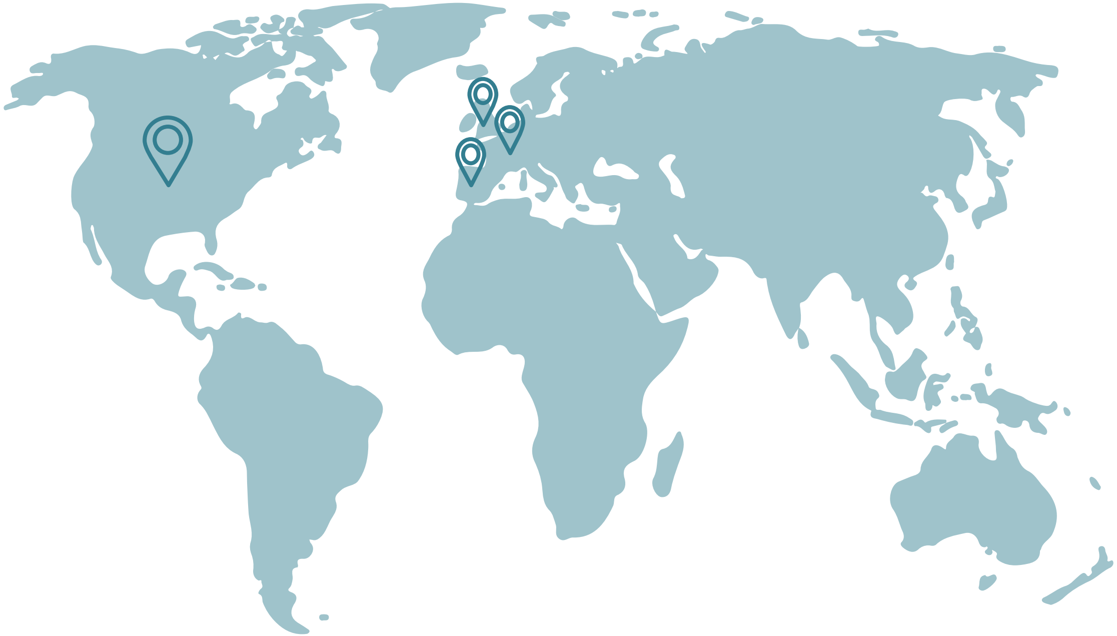
04

Feature Selection

Methods: Pearson, 2Chi-Square, L1-Regularization, RFE.

Removed deterministic features.

Where the guests come from?



Countries

%23.7 Portugal

%15.3 France

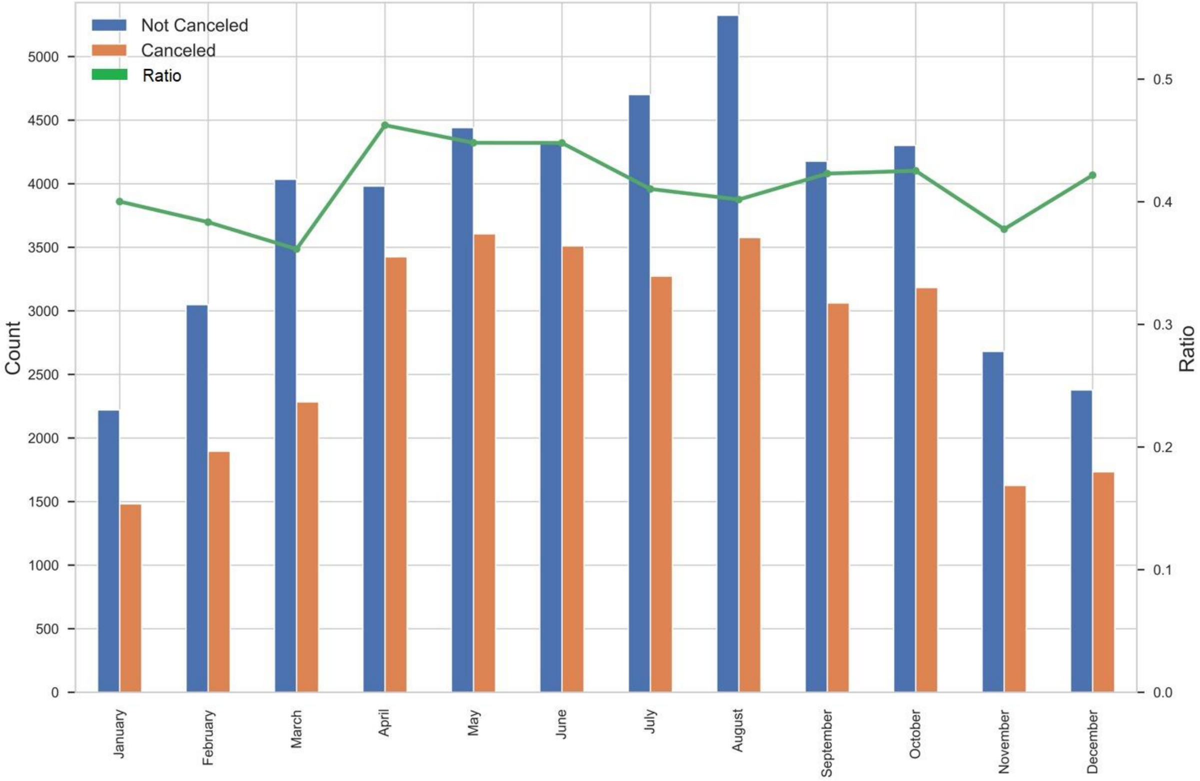
%10.93 Germany

%8.2 UK
…

%2.6 USA

Number of guests per month

Top Months: May - August



Guests: Customer Type

% Distribution of Customer Type

Transient

75%

Transient-Party

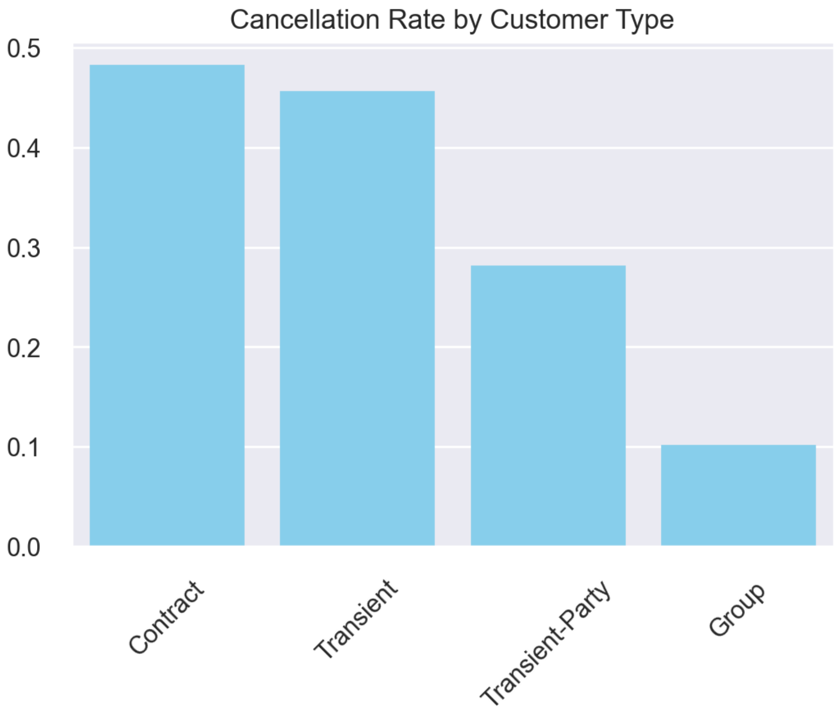
22%

Contract

3%

Group

0.4%



Cancellation Rate by Customer Type

Guests: Market Segment

% Distribution of Market Segments

Online TA

48%

Offline TA/TO

21%

Groups

17%

Direct

7%

Corporate

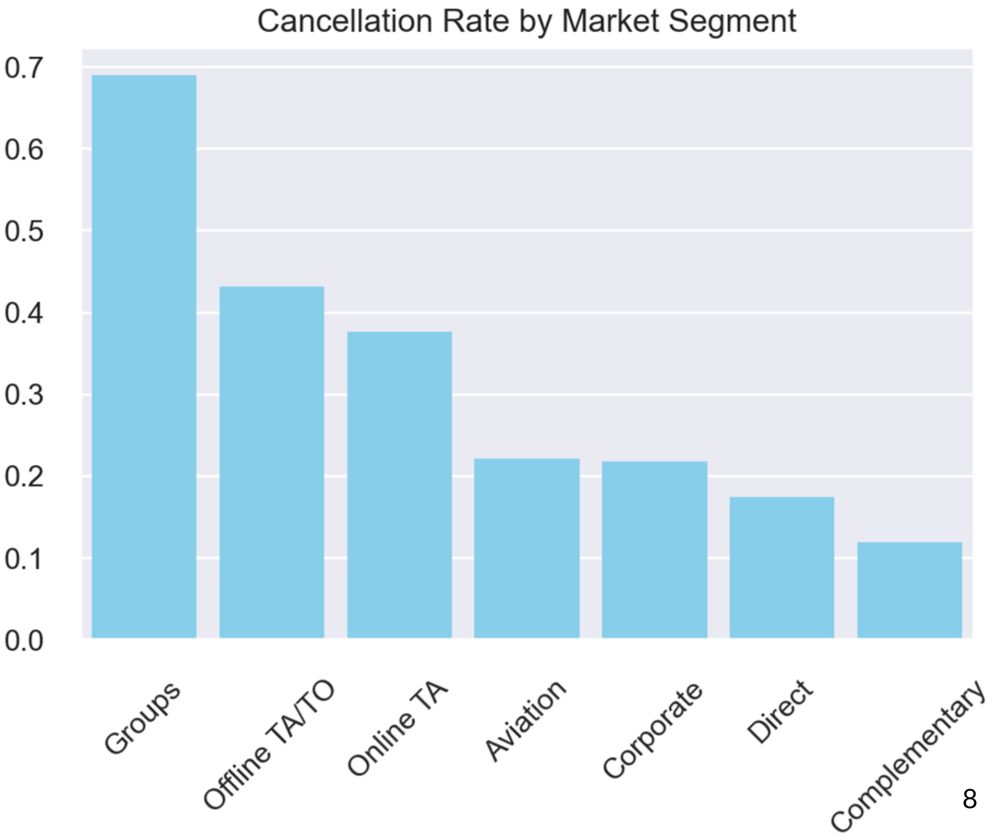
3%

Complementary

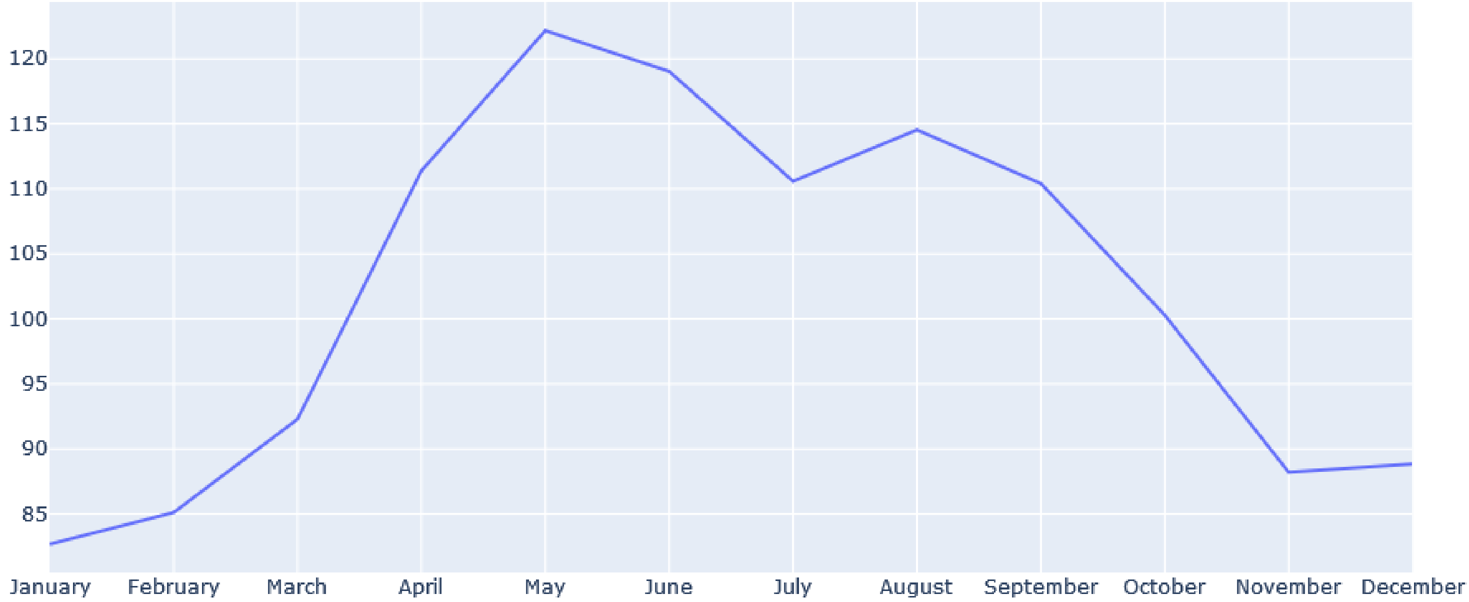
0.7%

Aviation

0.3%



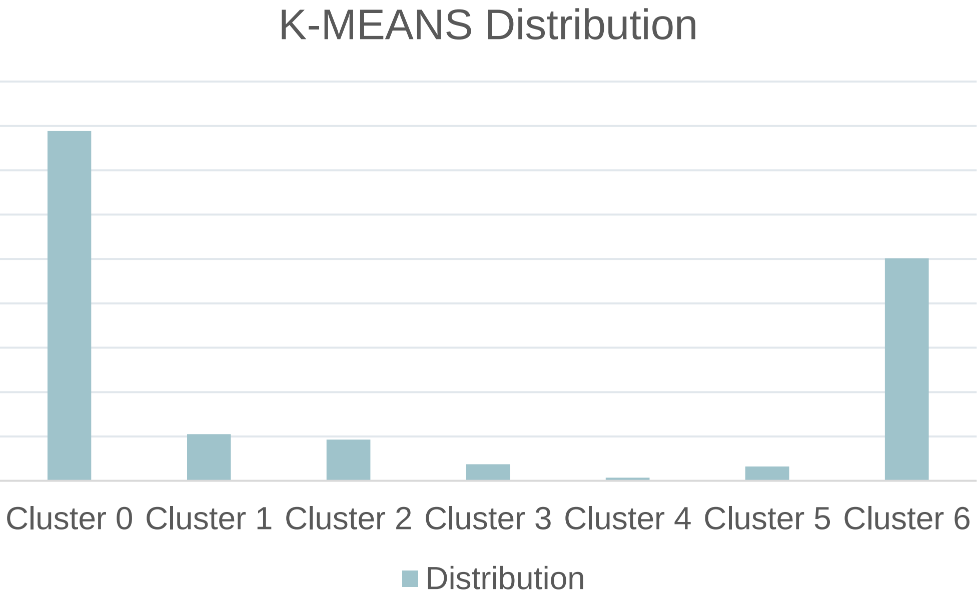
Average Daily Rate (ADR) per month



Clustering Techniques

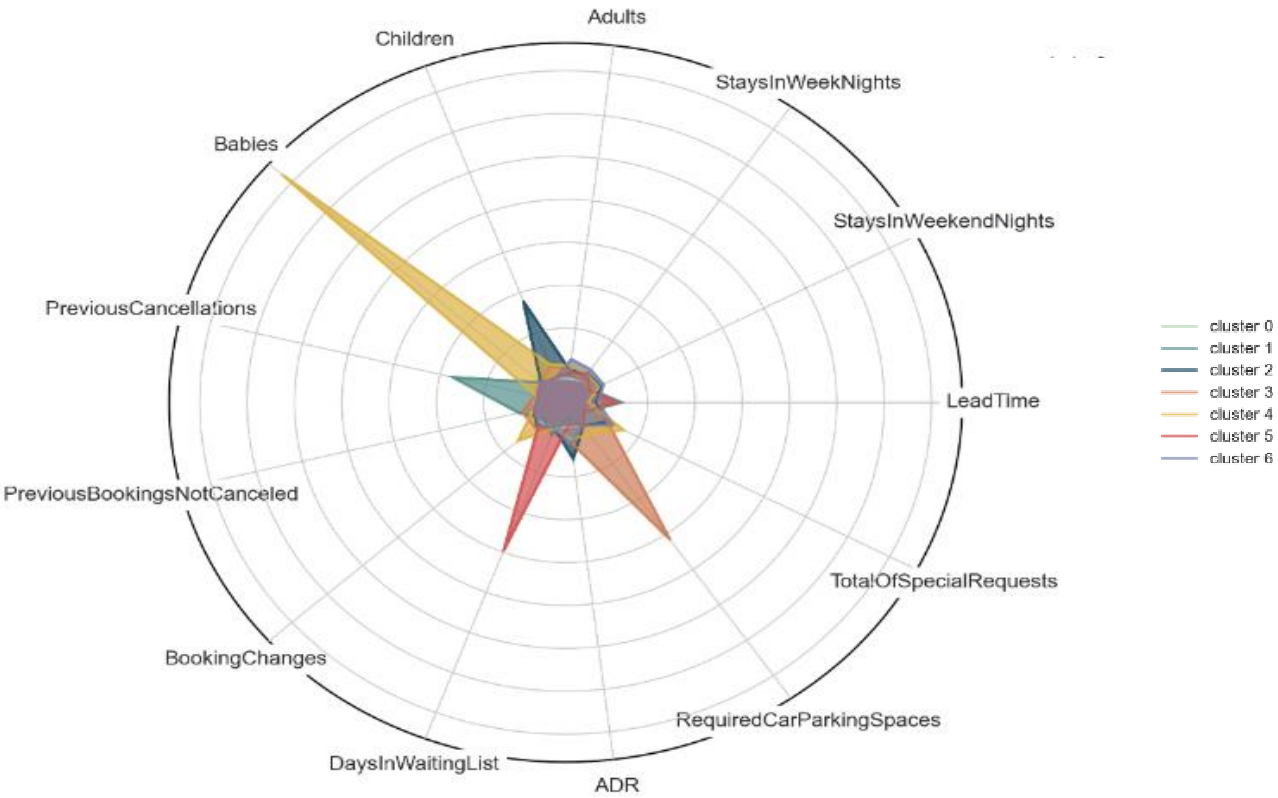
	● K-MEANS with PCA	● K-MEANS	● K-Medoids
Calinski Harabasz Index	8814	13457	5846

A higher CH index value indicates better clustering quality, meaning more distinct and well-separated clusters.



Outcome - Clustering Insights

- K-MEANS



Outcome - Predictions

XGB Classifier

Train Scores

Precision

89%

Recall

89%

F1-score

89%

Test Scores

Precision

87%

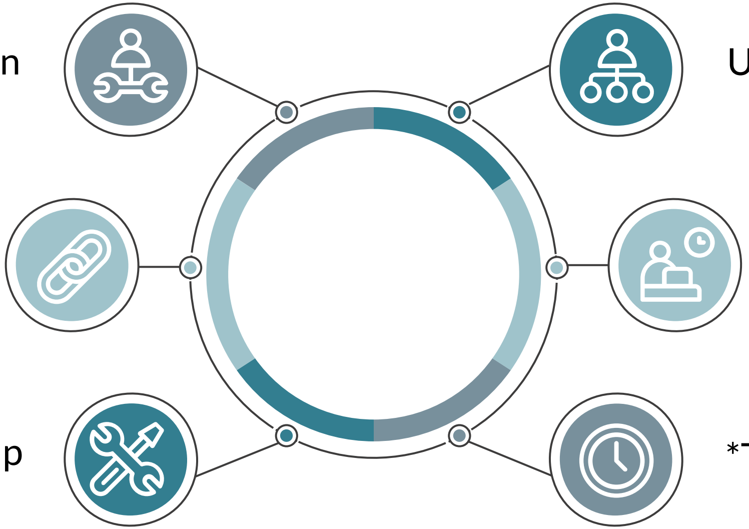
Recall

87%

F1-score

87%

Deployment



System Integration

API Development

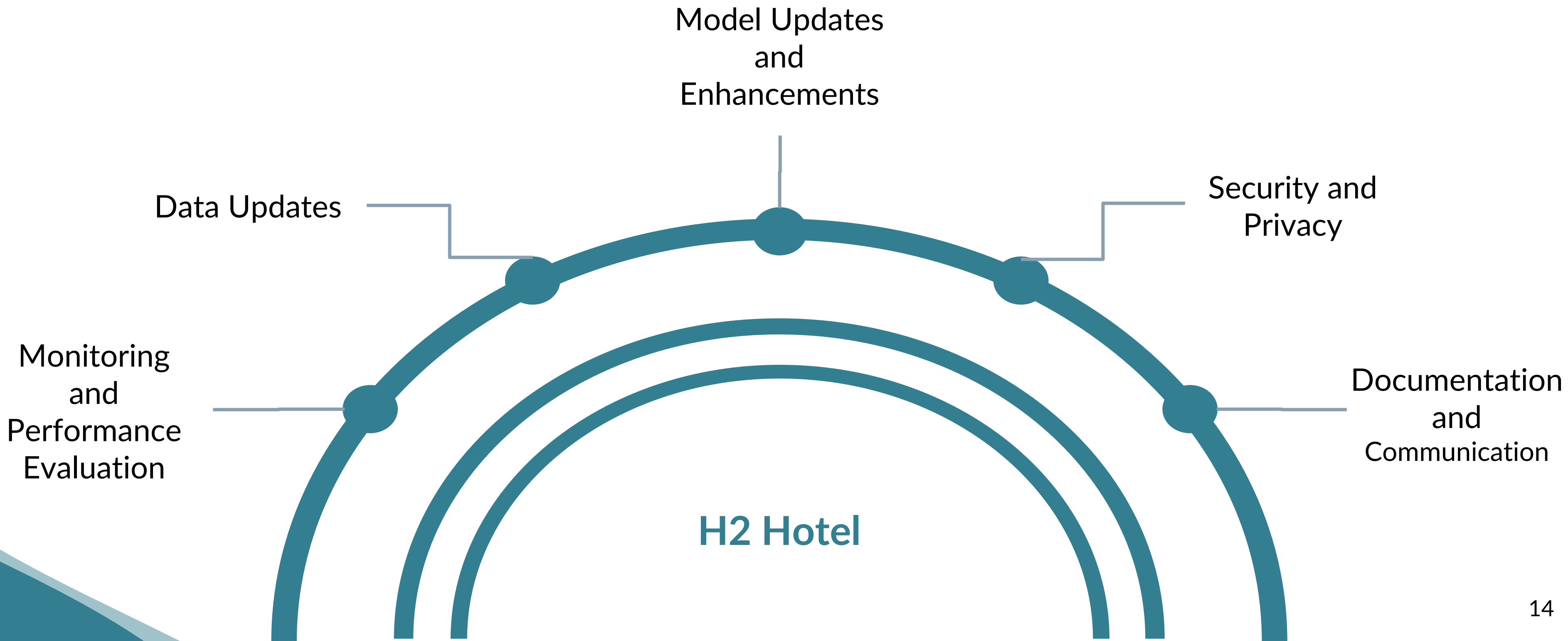
Infrastructure Setup

User Interface

Testing and Quality Assurance

*Typically takes around 2-4 weeks

Maintenance



Model Updates and Enhancements

Data Updates

Security and Privacy

Monitoring and Performance Evaluation

Documentation and Communication

H2 Hotel

THANKS!

DOES ANYONE HAVE ANY QUESTIONS?

Group D

Jaime Duarte:	20220675
Shanjida Roman:	20221395
Yousef Ebrahimi:	20221382
Diogo Martins:	20221361
David Martins:	20221006

NOVA IMS Information Management School

Credits: The original presentation template was created by Slidesgo, and Includes icons by Flation, and infographics & images by Freepik