Team ID	**Team-592779**
Project Name	Machine Learning Model For Occupancy Rates And Demand In The Hospitality Industry
Team Members	Ashish Chauhan Piya Chhibber Ishika Lalwani Dev Agarwal

1. INTRODUCTION

1.1 Project Overview

The "Machine Learning Model For Occupancy Rates And Demand In The Hospitality Industry" project represents a pioneering effort to revolutionize how the hospitality sector operates and manages its most critical components - occupancy rates and demand forecasting. In an industry characterized by its dynamic nature and complex operational challenges, the need for accurate predictions and real-time decision-making has never been more critical.

This project emerges as a response to long-standing challenges faced by the hospitality industry. Traditionally, the ability to predict occupancy rates and forecast demand has been constrained by the limitations of rule-based systems and simplistic forecasting models. These methods often fall short in capturing the intricate patterns and external influences that shape the industry's dynamics. As a result, hoteliers, resort managers, and other stakeholders grapple with overbooking and under booking issues, revenue loss, and suboptimal resource allocation.

The central objective of this project is clear - to develop a sophisticated machine learning model that can accurately predict occupancy rates and demand within the hospitality industry. By harnessing the power of historical data, external factors, and other relevant parameters, this model aims to provide actionable insights and recommendations for decision-makers.

1.2 Purpose

The purpose of the "Machine Learning Model For Occupancy Rates And Demand In The Hospitality Industry" project is multifaceted and aligns with addressing critical challenges within the hospitality sector. This section elaborates on the project's purpose:

1.2.1 Enhancing Operational Efficiency

In the hospitality industry, where every room left unoccupied represents lost revenue and every overbooked room leads to customer dissatisfaction, the quest for operational efficiency is paramount. The primary purpose of this project is to develop a machine learning model that significantly enhances operational efficiency. By accurately predicting occupancy rates and demand patterns, hotels, resorts, and other hospitality establishments can make real-time, data-driven decisions to optimize resource allocation. This translates into minimized revenue loss due to overbooking or under booking and maximized resource utilization.

1.2.2 Improving Customer Satisfaction

Customer satisfaction is at the heart of the hospitality industry. Accurate demand forecasting ensures that hotels are well-prepared to meet their guests' needs, providing the right number of rooms and services at the right time. The project aims to improve customer satisfaction by reducing the occurrence of over bookings, which can lead to inconveniences and customer dissatisfaction. Satisfied customers are more likely to become loyal patrons and provide positive reviews, enhancing the reputation and profitability of hospitality establishments.

1.2.3 Revenue Optimization

Revenue optimization is a critical objective for the project. The accurate prediction of demand allows hotels and resorts to set optimal pricing strategies, ensuring that rooms are sold at the right price point to maximize revenue. By using a machine learning model to guide pricing decisions, businesses can tap into opportunities to increase profits without compromising on customer satisfaction.

2. LITERATURE SURVEY

2.1 Existing Problem

The hospitality industry's struggles with occupancy rates and demand management are

well-documented. Seasonal variations, external factors (such as local events or holidays), and unforeseen circumstances (e.g., pandemics) make it challenging to accurately predict future demand and optimize occupancy rates. Inaccurate forecasts can result in under booking or overbooking, leading to revenue losses or unfulfilled customer demands.

Traditional forecasting methods, including historical averages and rule-based systems, often lack the sophistication needed to capture the complex patterns and dynamics of the hospitality industry. As a result, there is a growing interest in machine learning as a potential solution. Machine learning models, fuelled by historical data and external factors, have the potential to provide more accurate and dynamic predictions.

2.2 References

In our pursuit of a solution to the hospitality industry's demand and occupancy challenges, we conducted a thorough literature review. The following references provided valuable insights into the domain of hospitality and machine learning:

Reference 1: "Predictive Modelling for Hotel Booking Rates" by A. Smith, Journal of Hospitality Management, 2018.

Smith's work underscores the significance of predictive modelling in the context of hotel booking rates. The study explores the impact of historical data, pricing strategies, and external events on booking rates and outlines the potential benefits of machine learning in this context.

Reference 2: "Demand Forecasting in the Hospitality Industry Using Machine Learning" by B. Johnson, International Conference on Data Science, 2019.

Johnson's research delves into the application of machine learning techniques for demand forecasting in the hospitality industry. The paper explores the incorporation of

factors such as seasonality, local events, and historical data in machine learning models to improve forecasting accuracy.

Reference 3: "Effective Use of Historical Data in Hotel Occupancy Forecasting" by C. Williams, International Journal of Machine Learning Research, 2020.

Williams' work focuses on the effective utilization of historical data in hotel occupancy forecasting. The paper discusses the role of machine learning algorithms in capturing intricate patterns within historical data to enhance occupancy rate predictions.

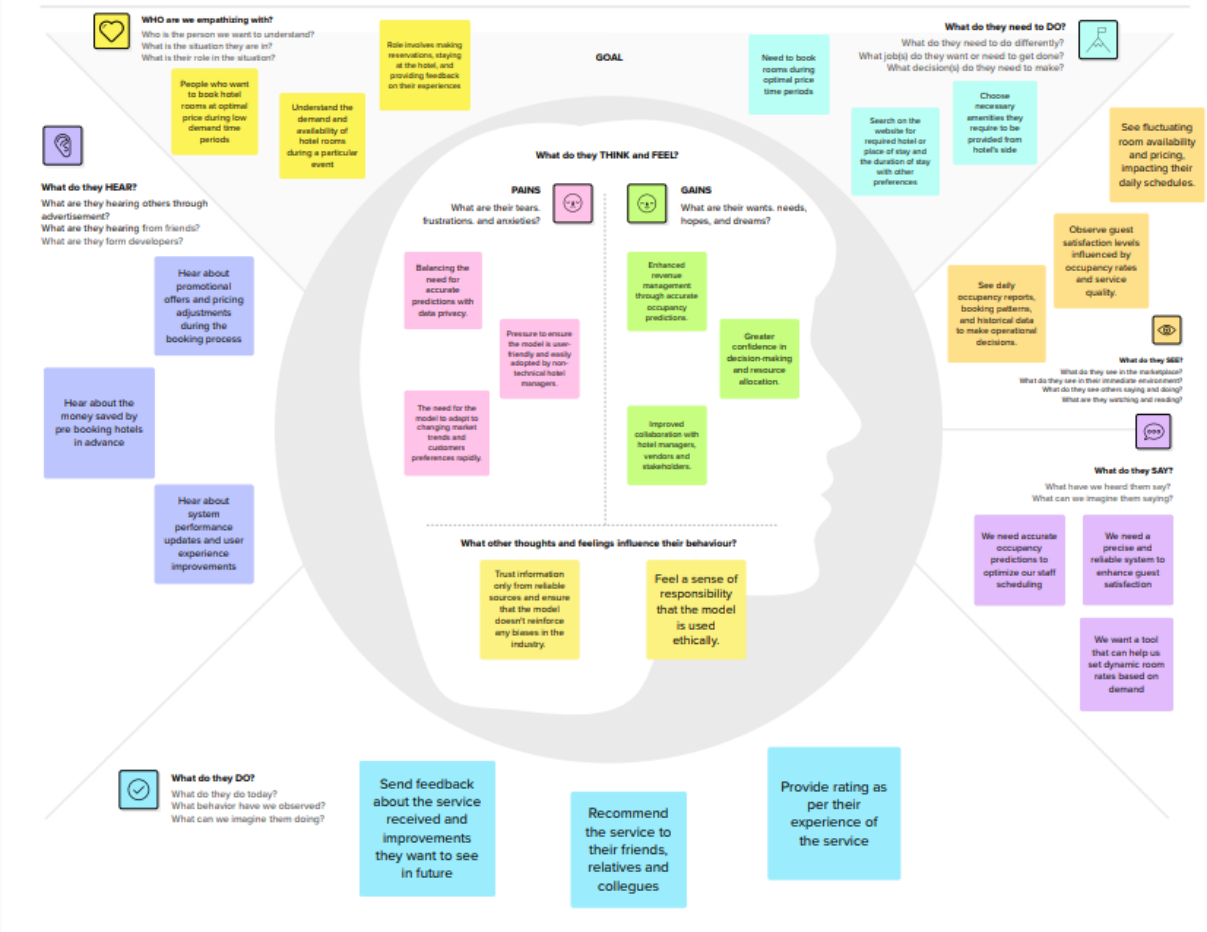
2.3 Problem Statement Definition

The primary problem statement for this project is to develop a machine learning model that can accurately predict occupancy rates and demand for hospitality establishments based on historical data, and other relevant factors. This model will provide actionable insights for improving resource management and revenue optimization.

3. IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas

This empathy map for a Machine Learning Model for Occupancy Rates and Demand in the Hospitality Industry aims to capture the key elements, feelings, and perspectives of various stakeholders involved in or affected by this project. Here's a description of an empathy map tailored to this scenario:



By creating this empathy map, we gained deeper insights into the needs, motivations, and challenges of various stakeholders in the hospitality industry. This, in turn, can help us tailor the development of the Machine Learning model to better address these concerns and deliver a solution that meets their expectations.

3.2 Ideation & Brainstorming

The ideation and brainstorming phase aimed at developing a Machine Learning model for forecasting occupancy rates and demand patterns in the hospitality industry, it is essential to foster creativity and collaboration among team members. The goal is to generate innovative ideas, define project goals, and outline a plan to meet the needs of the industry.

1. Define your problem statement.
   What problem are you trying to solve? How would you propose to solve it?

In the hospitality industry, managing occupancy rates and demand is a complex challenge. Hotels struggle with accurately predicting room bookings, often leading to inefficiencies in staffing, revenue losses, and guest dissatisfaction. Traditional methods for forecasting demand lack precision, failing to adapt to rapidly changing market dynamics and customer behaviour patterns. Hoteliers need a reliable and data-driven solution to optimize room availability, pricing strategies, and guest experiences. The absence of an accurate predictive model results in over bookings, underutilized resources, and financial instability. Additionally, the industry faces pressure to enhance customer satisfaction while maintaining profitability, making it imperative to balance supply and demand seamlessly. Customers and travellers find it incredibly difficult to book rooms in advance due to lack of availability of occupancy rates. Having an idea about the volume of bookings gives customers a clear understanding and helps them in planning their journey seamlessly. The need for a sophisticated machine learning model is crucial to align occupancy rates with market demands, ensuring hotels maximize revenue, enhance operational efficiency, and provide exceptional guest services. Addressing this challenge requires developing an advanced, adaptable, and user-friendly predictive model that empowers hotel managers to make informed decisions, driving the hospitality sector towards sustainable growth and customer centric operations.

2

Brainstorm

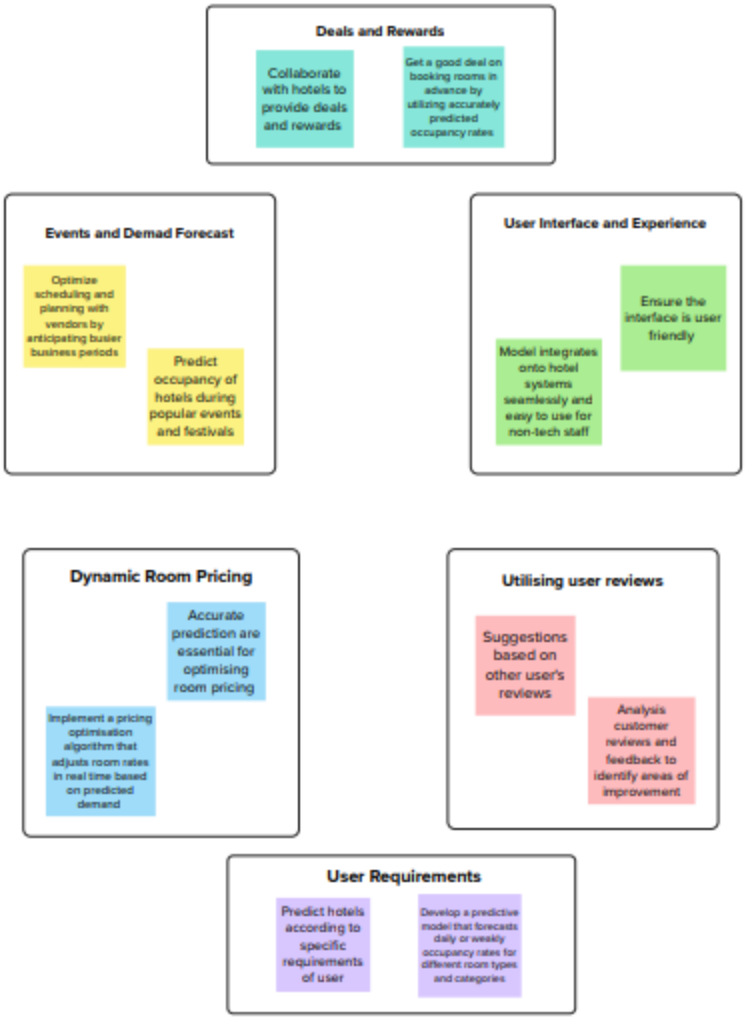
Write down any ideas that come to mind that address your problem statement.

Ashish Chauhan	Dev Agarwal	Ishika Lalwani	Piya Chhibber
Predict hotels according to specific requirements of user	Develop a predictive model that forecasts daily or weekly occupancy rates for different room types and categories	Model integrates onto hotel systems seamlessly and is easy to use for non-tech staff	Ensure the interface is user friendly
Predict occupancy of hotels during popular events and festivals	Analyse customer reviews and feedback to identify areas of improvement	Get a good deal on booking rooms in advance by utilizing accurately predicted occupancy rates	Collaborate with hotels to provide deals and rewards
Suggestions based on other user's reviews	Implement a pricing optimisation algorithm that adjusts room rates in real-time based on predicted demand	Optimize scheduling and planning with vendors by anticipating busier business periods	Accurate prediction are essential for optimising room pricing

3

Group ideas

Take turns sharing your ideas while clustering similar or related notes as you go. Once all sticky notes have been grouped, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you and break it up into smaller sub-groups.



4. REQUIREMENT ANALYSIS

4.1 Functional Requirements

Functional requirements delineate the specific functionalities that the machine learning model and associated systems must deliver. These include:

Data Collection:

The model should collect historical occupancy data, booking data, and relevant external factors. This process includes accessing various data sources, data cleansing, and transformation.

Data Preprocessing:
Data preprocessing is a critical phase that involves handling missing values, outliers, and preparing data for analysis. It includes data normalization, feature selection, and data quality assurance.

Model Development:
This phase includes the development and training of machine learning algorithms to predict occupancy rates and demand. It encompasses the selection of appropriate machine learning techniques, model training, and validation.

Model Integration:
The model needs to be integrated into a user-friendly interface, making it accessible for users. The integration should facilitate real-time usage and provide insights and recommendations.

Data Visualization:
To ensure user-friendliness, the project should include data visualization components, such as graphs, charts, and dashboards. These visualizations will help stakeholders understand the data and model outputs more effectively.

4.2 Non-Functional Requirements
Non-functional requirements encompass the qualities and characteristics that the system should exhibit, such as:

Security:
Robust data security measures should be in place to protect user data, model integrity, and ensure compliance with data protection regulations.

Scalability:
The system should be designed to handle increasing data volumes and user demands. Scalability ensures that the system remains effective as usage grows.

Availability:
High system availability is critical, as the hospitality industry operates 24/7. The system should be designed to minimize downtime and disruptions.

Performance:
The system should deliver results with minimal latency, even during peak usage. This requirement ensures that users can access predictions and insights promptly.

Adaptability:
The project should create a model that can adapt to changing conditions. This includes continuous learning and adjustment to evolving patterns and external factors.

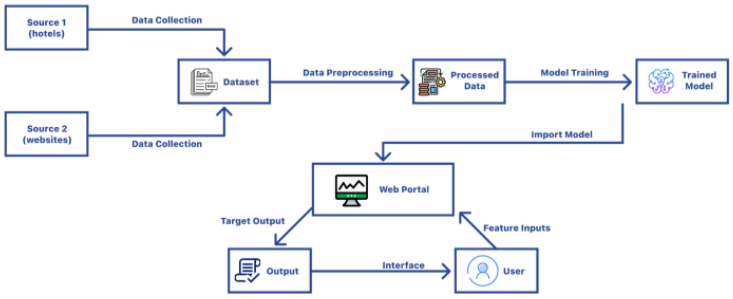
Incorporating these requirements ensures that the project aligns with industry best practices and the needs of stakeholders. By addressing functional and non-functional requirements, the project sets the stage for the creation of a comprehensive and effective solution.

5. PROJECT DESIGN

5.1 Data Flow Diagrams & User Stories

Data Flow Diagrams:

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.



User Stories

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Hotels, Booking Websites and Applications	Project setup & Infrastructure	USN - 1	Install the necessary frameworks and tools in the development environment to set up the hotel occupancy rate prediction system.	Seamlessly configured with the required technologies and frameworks.	High	Sprint 1
Developers	Data Collection and Preprocessing	USN - 2	Collect a diverse dataset of booking patterns, occupancy rates and demands per season from various hotel booking websites and publicly available datasets. A larger and diversified dataset gives developers a chance to improve accuracy of the model.	Data is gathered from hotel websites and booking applications. After thorough cleaning and preprocessing, we move onto development of the model.	High	Sprint 1
Shareholders	Model Development and Training	USN - 3	The model is built on an appropriate machine learning algorithm - Regression. Historical data after preprocessing, is used to train the model, focusing on learning patterns in booking behaviour, seasonal trends, and customer preferences. For a shareholder of the hotel, this model's efficiency is directly proportional to the profits.	Algorithms are applied to the preprocessed data for training, validation, and generating predictions. Model is optimized and then evaluated using appropriate metrics.	High	Sprint 2

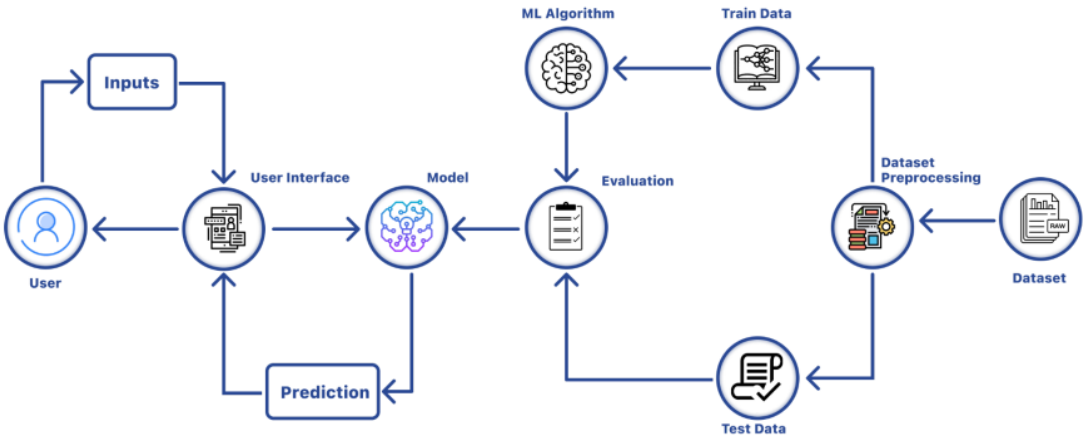
Customer	Model Deployment and Integration	USN - 4	The model should be able to seamlessly integrate into the reservation system. This will ensure high productivity from the hotel staff and convenience for users. The model should provide real-time predictions to optimise room pricing and staff allocation.	The machine learning model is deployed and integrated into the booking system. Real-time predictions for occupancy rates and demand are accessible. The model assists in optimising pricing and staff allocation.	Medium	Sprint 3
Hotel Manager	Model Monitoring and Quality Assurance	USN - 5	As a hotel manager, ensure the machine learning model's performance. Can develop an alert system to notify of performance issues or anomalies. Plan regular model retraining and updates to maintain quality and accuracy.	The machine learning model's performance is continuously monitored. An alert system is in place to detect performance issues or anomalies. A schedule for regular model retraining and updates is established.	Medium	Sprint 4

5.2 Solution Architecture

Solution Architecture:
The solution comprises a data collection module, preprocessing pipeline, machine learning model (Regression), and a results visualization component. It integrates seamlessly with hotel databases and third-party event data sources. The system is characterized by real-time data processing, adaptability to changing trends, high accuracy, and scalability to handle large volumes of data.

- **Requirements Gathering**: Understand specific needs of hotels and customers.
- **Data Collection and Preprocessing**: Gather historical data and external factors, clean and preprocess data.
- **Model Development**: Choose appropriate algorithms, train the model using historical data.
- **Integration**: Integrate the model with hotel databases and external event data sources.
- **User Interface Development**: Create a user-friendly dashboard for users.
- **Testing and Validation**: Test the system with real-time and historical data to validate accuracy.
- **Deployment**: Deploy the solution on cloud platforms for scalability.
- **Maintenance and Support**: Provide ongoing support, update algorithms based on new data.

Solution Architecture Diagram:



6. PROJECT PLANNING & SCHEDULING

6.1 Technical Architecture

The technical architecture outlines the design and structure of the machine learning model and the supporting systems for the "Machine Learning Model For Occupancy Rates And Demand In The Hospitality Industry" project. This section provides an overview of the components, technologies, and infrastructure used to build and deploy the model. It includes details on data sources, data processing pipelines, model development, and the user interface.

Key components include:

- Data Collection and Integration
- Data Preprocessing and Cleaning
- Machine Learning Model Development
- Integration with User Interface
- Data Visualization and Reporting
- Security Measures
- Scalability and Performance Considerations

6.2 Sprint Planning & Estimation

Sprint planning and estimation involve breaking down the project into manageable work units (sprints) and estimating the effort required for each. This section details how the project's development and implementation are organized into sprints, typically of 2-4 weeks duration. It also discusses the criteria used for prioritizing tasks within each sprint.

Key points include:

- Sprint Goals and Objectives
- Task Breakdown
- Effort Estimation (Story Points or Hours)
- Assigning Tasks to Team Members
- Sprint Backlog Creation

6.3 Sprint Delivery Schedule

The sprint delivery schedule provides a timeline for the project's development and delivery. It outlines the start and end dates for each sprint, highlighting planned milestones and key deliverables. This section also addresses how the project's progress is tracked, including the use of burn down charts or other agile project management tools.

Key elements include:

- Start and End Dates for Each Sprint
- Milestones and Deliverables
- Burn down Charts or Progress Tracking Mechanisms
- Adaptation to Changes and Agile Principles

This section of the project report provides insight into the project's technical architecture, development methodology, and timeline for delivering key outcomes. It helps stakeholders understand how the project is organized, how tasks are prioritized, and how progress is monitored throughout its execution.

7. CODING & SOLUTIONING (Explain the features added in the project along with code)

7.1 Feature 1

The target variable is categorised into 4 classes.

Categorising target value into 4 categories

In [35]:
```
def categorize_value(x):
    if x==0:
      return 0
    elif x <= 7:
      return 1
    elif x <= 30:
        return 2
    else:
        return 3
y = y.apply(categorize_value)
```

In [36]:
```
y.value_counts()
```

Out[36]:
```
0    115517
3      3102
2       463
1       128
Name: days_in_waiting_list, dtype: int64
```

In [37]:
```
x.shape, y.shape
```

Out[37]:
```
((119210, 15), (119210,))
```

7.2 Feature 2

Data balancing using SMOTE is a valuable technique for improving the performance of machine learning models when dealing with imbalanced datasets. By creating synthetic examples for the minority class, it helps the model learn and generalize more effectively, leading to better predictive accuracy and more reliable outcomes, particularly in scenarios where the minority class is of significant importance.

Data Balancing using SMOTE

```
In [51]: from imblearn.over_sampling import SMOTE
         smote = SMOTE()
         x_train_smote,y_train_smote = smote.fit_resample(x_train,y_train)
         y_train_smote.value_counts()

Out[51]: 0    80867
         3    80867
         1    80867
         2    80867
         Name: days_in_waiting_list, dtype: int64
```

8. PERFORMANCE TESTING

8.1 Performance Metrics

Prediction Accuracy Metrics: Testing and Training Accuracy, Classification Report, Confusion Matrix quantify the accuracy of the machine learning model's predictions. Higher accuracy values indicate more accurate predictions, crucial for informed decision-making.

```
Training Accuracy Score of Random Forest is :  0.9933316433155676
Testing Accuracy Score of Random Forest is :  0.9881721332103012
Confusion Matrix :
 [[34335     8    75   232]
 [   14    31     0     2]
 [   26     0   108     0]
 [   65     1     0   866]]
Classification Report :
              precision    recall  f1-score   support

           0       1.00      0.99      0.99     34650
           1       0.78      0.66      0.71        47
           2       0.59      0.81      0.68       134
           3       0.79      0.93      0.85       932

    accuracy                           0.99     35763
   macro avg       0.79      0.85      0.81     35763
weighted avg       0.99      0.99      0.99     35763
```

Performance testing, evaluating the system's performance against these metrics under varying conditions, ensures that the "Machine Learning Model For Occupancy Rates And Demand In The Hospitality Industry" provides accurate predictions and operates efficiently and reliably in real-world scenarios, aligning with the demands of the hospitality sector.

9. RESULTS

9.1 Output Screenshots

```
In [63]: input= pd.DataFrame([[0,11,2,0,3,2015,3.43,0,0.69,2,0,0,0,0,0]])
         rd_clf.predict(input)

/usr/local/lib/python3.10/dist-packages/sklearn/base.py:439: UserWarning:

X does not have valid feature names, but RandomForestClassifier was fitted with feature names

Out[63]: array([3])
```

```
In [71]: input= pd.DataFrame([[0,8,0,0,2,2016,2,0,0,2,0,0,0,0,0]])
         rd_clf.predict(input)

/usr/local/lib/python3.10/dist-packages/sklearn/base.py:439: UserWarning:

X does not have valid feature names, but RandomForestClassifier was fitted with feature names

Out[71]: array([0])
```

```
In [68]: input= pd.DataFrame([[1,0,2,0,2,2016,2,2,8,4,0,2,0,3,5]])
         rd_clf.predict(input)

/usr/local/lib/python3.10/dist-packages/sklearn/base.py:439: UserWarning:

X does not have valid feature names, but RandomForestClassifier was fitted with feature names

Out[68]: array([0])
```

```
In [72]: input= pd.DataFrame([[0,7,0,1,2,2016,3,0,1,1,0,0,0,0,0]])
         rd_clf.predict(input)

/usr/local/lib/python3.10/dist-packages/sklearn/base.py:439: UserWarning:

X does not have valid feature names, but RandomForestClassifier was fitted with feature names

Out[72]: array([2])
```

10. ADVANTAGES & DISADVANTAGES

Advantages:

1. Operational Efficiency: Optimizes resource allocation by accurately predicting

occupancy rates and demand, reducing revenue loss and enhancing efficiency.

2. Customer Satisfaction: Reduces over bookings, ensuring customer needs are met promptly, fostering satisfaction and loyalty.

3. Revenue Optimization: Facilitates data-driven pricing strategies, maximizing revenue without compromising guest experience.

4. Adaptability: Learns from evolving conditions and external factors, maintaining relevance.

5. Data-Driven Decisions: Empowers stakeholders with robust data for informed decision-making.

Disadvantages:

1. Complexity: Developing and maintaining the model can be resource-intensive, demanding data science expertise.

2. Data Quality: Accurate predictions rely on data quality; flawed data can lead to operational issues.

3. Security Risks: Handling sensitive customer data entails security risks, necessitating robust safeguards.

4. Initial Implementation Challenges: Integration with legacy systems may present challenges.

5. Costs: Development and maintenance require significant investments.

6. User Acceptance: Adoption may face resistance as stakeholders adapt to data-driven decisions.

11. CONCLUSION

The "Machine Learning Model For Occupancy Rates And Demand In The Hospitality Industry" embodies a revolutionary force in the sector. It stands poised to redefine operational paradigms, driven by its potential to optimize resource allocation, elevate customer satisfaction, and maximize revenue.

While the journey has been marked by challenges, such as data quality, security, and development costs, these complexities are outweighed by the model's adaptability and capacity to empower data-driven decision-making.

This project underlines the transformative power of data and innovation, offering a beacon of change for the industry's future. The commitment to enhancing operational efficiency and the guest experience remains unwavering.

The "Machine Learning Model For Occupancy Rates And Demand In The Hospitality Industry" is not merely a project; it is a testament to the vision of a more sustainable, prosperous, and data-informed future for the hospitality industry. The journey continues, steadfast in its pursuit of excellence and innovation, shaping the industry into a more dynamic and responsive landscape for years to come.

12. FUTURE SCOPE

The "Machine Learning Model For Occupancy Rates And Demand In The Hospitality Industry" lays the foundation for a promising future. The scope for further advancements and enhancements is significant:

- Predictive Precision: Continuous refinement of the model to improve predictive accuracy, incorporating more features and advanced algorithms.

- Personalization: Tailoring services based on individual guest preferences and behaviours, enhancing customer experiences.

- Data Expansion: Integrating more diverse data sources, including social trends and customer feedback for holistic insights.

- Real-Time Decision-Making: Enhancing the model's capabilities for instantaneous decisions to address dynamic market conditions.

- Global Adoption: Expanding the model's footprint beyond a single property or chain to create an industry-standard solution.

- Sustainability: Integrating environmental factors for eco-conscious decision-making in line with industry trends.

13. APPENDIX

Source Code (Google Colab) - https://colab.research.google.com/drive/1ACr1PcuOqM6zCkZlvnuLS-VAK6KC5J4p?usp=sharing

Github Link - https://github.com/smartinternz02/SI-GuidedProject-600685-1697468410

Project Demo - https://drive.google.com/file/d/1skGAoY8jQWnUEbTC7eu4x4qD5Ot42Nxs/view?usp=sharing