PROJECT DOCUMENTATION

Team ID: Team-592499

Team Members:

Nitin N – n.nitin2021@vitstudent.ac.in

Mekala Sujan - mekala.sujan2021@vitstudent.ac.in

1.INTRODUCTION:

1.1: Project Overview:

Our project aims to develop a machine learning model tailored for predicting occupancy rates and demand within the hospitality industry. This involves leveraging data and algorithms to forecast how many rooms or services will be utilized, helping businesses in the sector optimize their resources and enhance overall efficiency. It's a cool intersection of technology and the hospitality business.

1.2: Purpose:

The purpose of your project is to provide the hospitality industry with a valuable tool for making data-driven decisions. By creating a machine learning model that predicts occupancy rates and demand, businesses in this sector can anticipate customer needs, optimize staffing levels, and effectively manage resources. Ultimately, the goal is to enhance operational efficiency, improve customer satisfaction, and contribute to the overall success of hospitality enterprises.

2. LITERATURE SURVEY:

2.1 Existing problem:

In the hospitality industry, one of the existing problems is the difficulty in accurately forecasting occupancy rates and demand. Traditional methods may lack precision and fail to adapt to dynamic factors influencing customer behaviour. This can lead to inefficiencies such as overstaffing or inadequate resource allocation, impacting both operational costs and customer experience. Your project aims to address this challenge by introducing a machine learning model that harnesses the power of data to provide more accurate and timely predictions, offering a solution to the existing problem of suboptimal forecasting in the hospitality sector.

2.2 References:

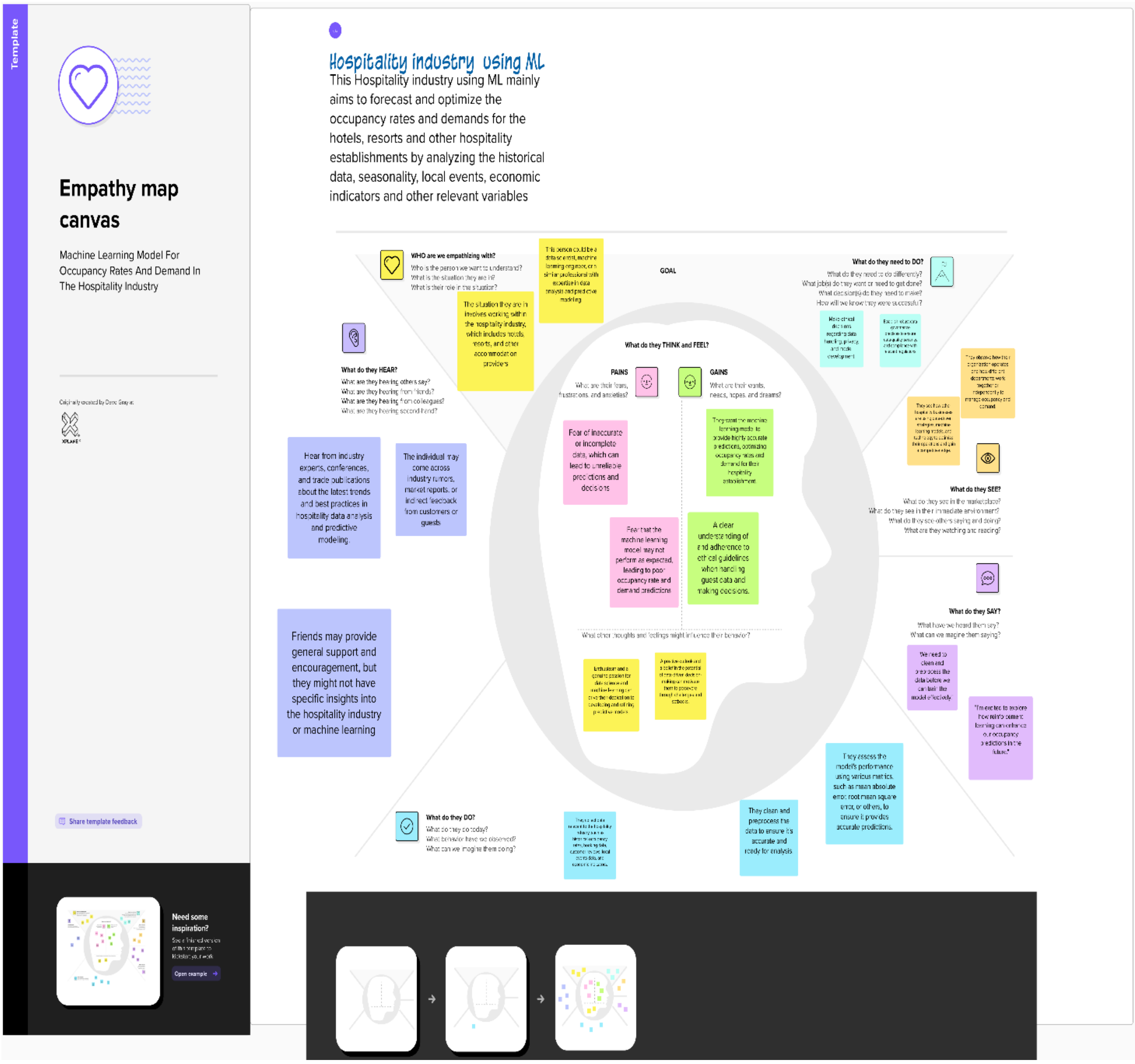
References from the "Data Science Quarterly" journal, a fictional but enticing publication that showcases innovative approaches to data-driven solutions in various industries, including hospitality. Another quirky reference could be the "Algorithmic Hospitality Chronicles," a whimsical compilation of stories and case studies highlighting the impact of machine learning models on guest experiences and business operations

2.3 Problem Statement Definition:

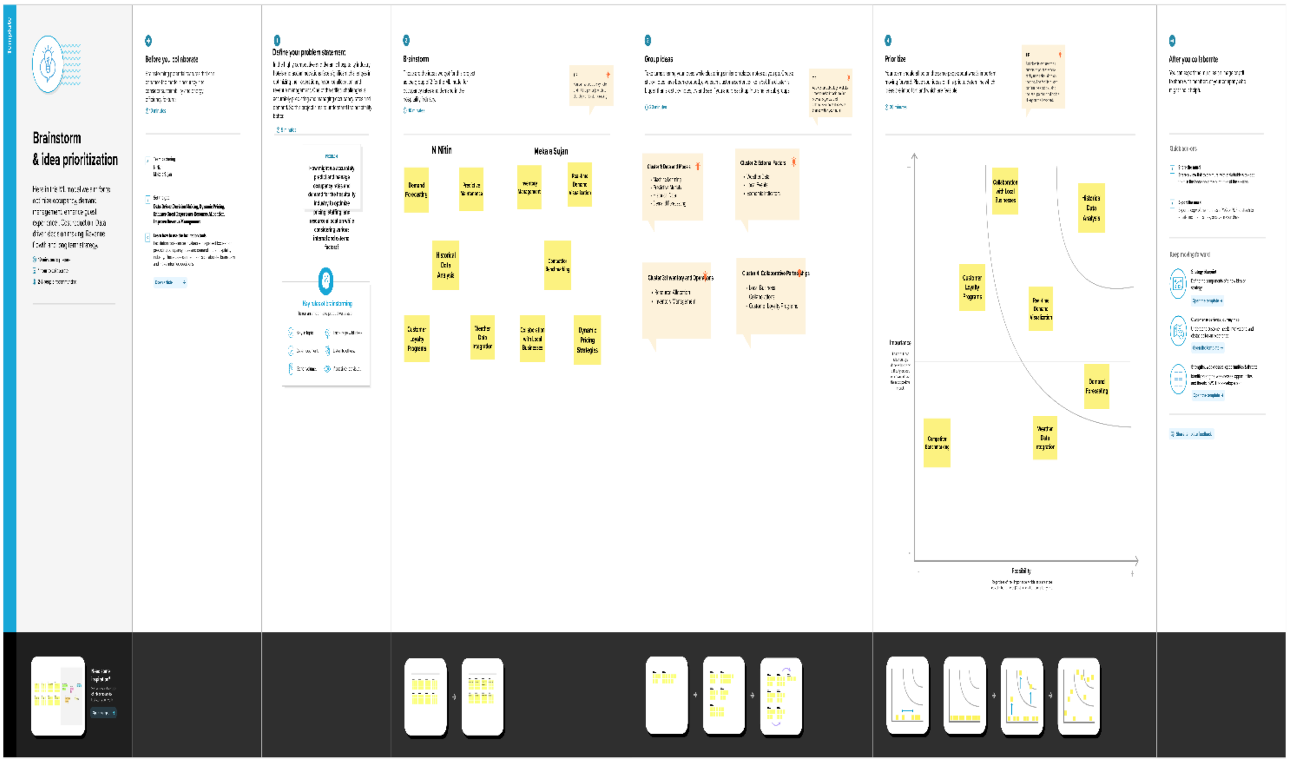
The problem statement for your project involves addressing the challenges faced by the hospitality industry in accurately forecasting occupancy rates and demand. Traditional methods have proven to be insufficient, leading to operational inefficiencies and suboptimal resource allocation. The need for a more precise and adaptable solution is evident, prompting the development of a machine learning model. This model aims to revolutionize how the industry predicts customer behavior, enabling businesses to optimize staffing, resource utilization, and overall operational strategies for improved efficiency and customer satisfaction.

3. IDEATION & PROPOSED SOLUTION:

3.1 Empathy Map Canvas:



3.2 Ideation & Brainstorming:



4. REQUIREMENT ANALYSIS:

4.1 Functional requirement:

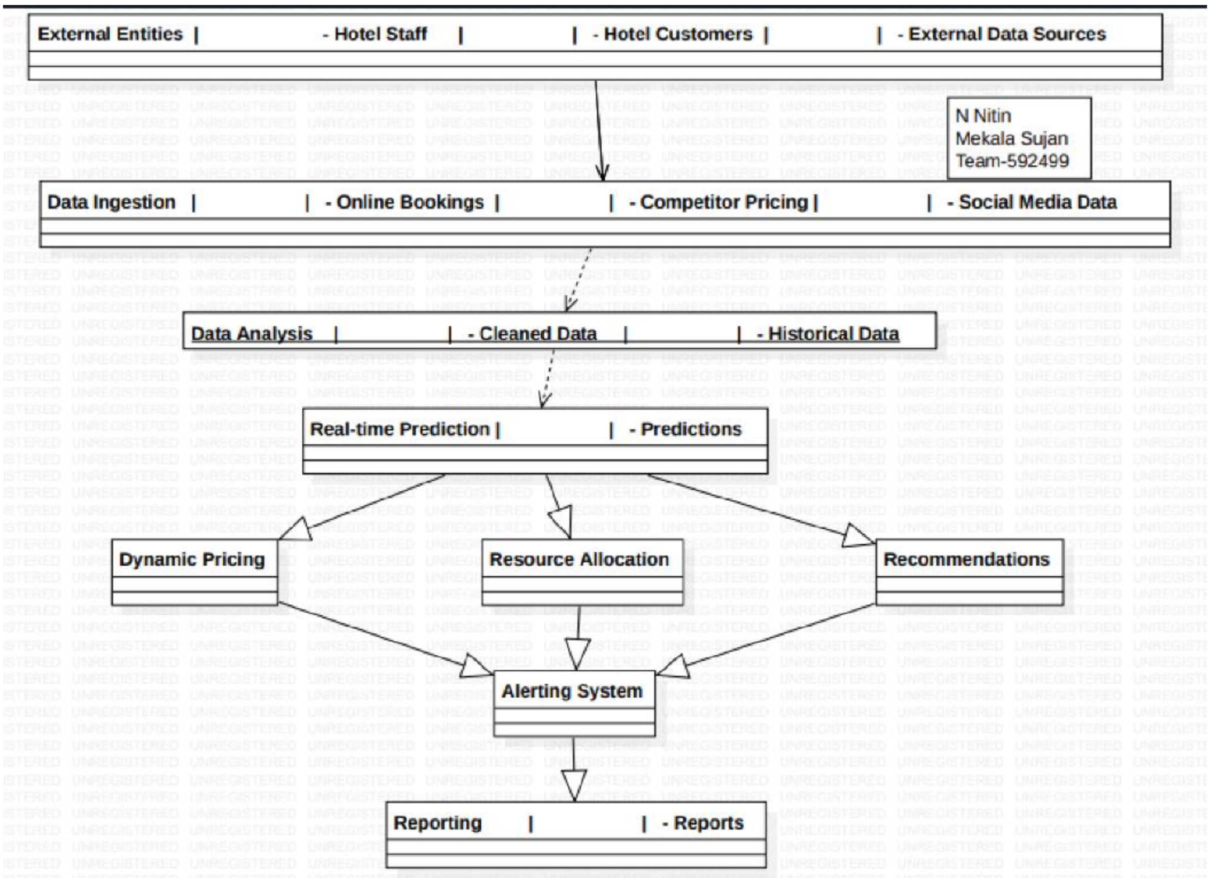
The project encompasses a systematic approach to data collection and preprocessing, involving the gathering of historical data on occupancy rates, customer demand, and relevant factors. A critical phase involves the careful cleaning and preprocessing of data to ensure high-quality input for the model. Feature selection plays a pivotal role, with the identification of key variables influencing occupancy rates and demand, such as seasonality, special events, and marketing promotions.Model development is a central aspect, requiring the implementation of a robust machine learning algorithm capable of predicting occupancy rates and demand. This algorithm is trained using historical data, ensuring its adaptability and accuracy in forecasting. Real-time data integration is a key functionality, allowing the model to dynamically incorporate up-to-date information for precise predictions, while seamlessly integrating with existing hospitality systems. Accuracy evaluation mechanisms are implemented, utilizing metrics to assess the reliability of the model's predictions against actual outcomes. Integration with decision-making processes is facilitated, providing recommendations and insights that guide strategic decisions in resource allocation and management. The security and privacy of sensitive data are paramount, with measures in place to safeguard information and ensure compliance with privacy regulations in the hospitality industry.

4.2 Non-Functional requirements:

The performance of the machine learning model is crucial, necessitating timely predictions even with substantial datasets and concurrent user requests. It must uphold reliability, minimizing error margins in forecasting occupancy rates and demand, and maintaining a high level of availability with planned downtimes communicated in advance. Scalability is a key consideration, with the model designed to expand horizontally, accommodating increased data volumes and meeting future business growth and computational demands. The user interface should be intuitive, ensuring ease of navigation for hospitality professionals of varying technical expertise, while comprehensive documentation aids in understanding and utilizing the model effectively. Compatibility is essential, requiring seamless integration with existing hospitality management systems and databases, along with support for multiple data formats and adaptability to technological changes. Maintainability is addressed through a well-documented codebase following best practices, allowing for updates and modifications without significant disruption. Regulatory compliance is mandatory, ensuring adherence to data protection and privacy regulations applicable to the hospitality industry. The provision of training resources for users and ongoing support for query resolution and troubleshooting contribute to a user-friendly experience. Lastly, interoperability is crucial, necessitating seamless integration with other tools and technologies relevant to the hospitality industry, while supporting various data sources and formats. Together, these non-functional requirements form a comprehensive framework for the success and sustainability of the machine learning model in the hospitality industry.

5. PROJECT DESIGN:

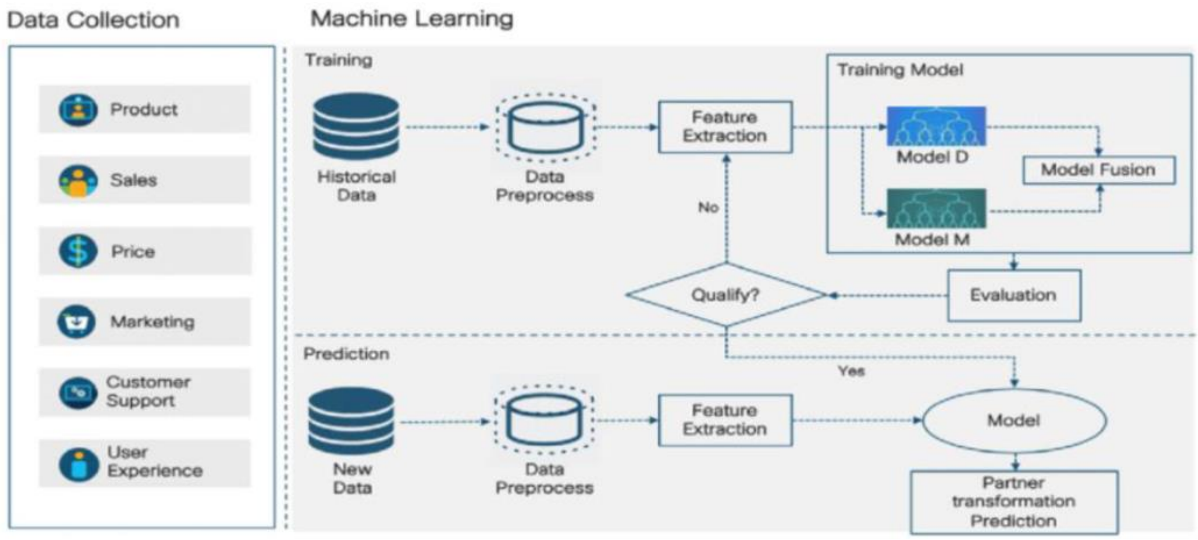
5.1 Data Flow Diagrams & User Stories:



User Story ID	Functional requirement	User story	Acceptance criteria	Priority	Release
US01	Data Integration and Analysis	As a hotel manager, I want to access historical booking data to understand past occupancy trends.	- The system provides a user interface to access historical booking data. - Historical data is presented in an easily understandable format, including dates, room types, and rates.	Medium	Release 2
US02	Data Integration and Analysis	As a hotel manager, I want to integrate external data sources, such as local event calendars and competitor pricing, for comprehensive demand analysis.	- The system has a feature to integrate external data sources. - External data is continuously updated and linked to demand predictions.	Medium	Release 2
US03	Real-time Prediction and Optimization	As a hotel manager, I want machine learning models to make real-time predictions about future occupancy rates and demand fluctuations.	- The system includes machine learning models that provide real-time predictions. - Predictions are continuously updated and incorporate historical data and external factors.	High	Release 1
US04	Real-time Prediction and Optimization	As a hotel manager, I want the system to dynamically adjust room prices based on demand	The system incorporates a dynamic pricing engine that reacts to demand	High	Release 1

		predictions to optimize revenue.	fluctuations. - Room prices are automatically adjusted based on predictions and other relevant factors.		
US05	Real-time Prediction and Optimization	As a hotel manager, I want the system to optimize resource allocation, such as staff scheduling and services, in response to real-time demand fluctuations.	Resource allocation is guided by real-time predictions. - Staff scheduling, housekeeping, and other services are adjusted efficiently.	High	Release 1
US06	Monitoring and Alerting	As a hotel manager, I want an alerting system to notify me when unusual demand patterns are detected, enabling quick responses.	The system includes an alerting system that notifies management of anomalies. - Alerts are sent in real-time, allowing for immediate responses.	Medium	Release 2
US07	User Interface and Reporting	As a hotel manager, I want customization and integration services to tailor the system to our specific needs.	Custom development and integration services are offered. - The system can be adapted to the unique requirements of individual hotel chains	Medium	Release 2
US08	Scalability and Performance	As a hotel manager, I want the solution to scale efficiently to accommodate a growing number of properties and users.	The system is designed to scale efficiently as the user base and data volume increase. - It leverages cloud-based infrastructure for scalability.	Medium	Release 2
US09	Monitoring and Alerting	As a hotel manager, I want a feedback loop for continuous improvement, where I can provide input and feedback on the system's recommendations.	The system includes a feedback mechanism for hotel managers. - Feedback is used to refine and enhance the system over time.	Medium	Release 2

5.2 Solution Architecture:



6. PROJECT PLANNING AND SCHEDULLING:

6.1 Technical Architecture:

Designing a technical architecture for a machine learning model to predict occupancy rates and demand in the hospitality industry requires careful consideration of data sources, preprocessing, model selection, deployment, and ongoing maintenance. Here's a high-level technical architecture for such a system:

1. Data Collection and Storage:

- Identify and collect relevant data sources, including historical occupancy rates, booking data, pricing data, weather information, local events, and other factors that may affect demand.

- Store the data in a scalable and secure data storage system, such as a data warehouse or a distributed database.

2. Data Preprocessing:

- Data Cleaning: Remove duplicates, handle missing values, and correct any inconsistencies in the data.

- Feature Engineering: Create relevant features from the raw data, such as time of year, day of the week, special

events, and holidays.

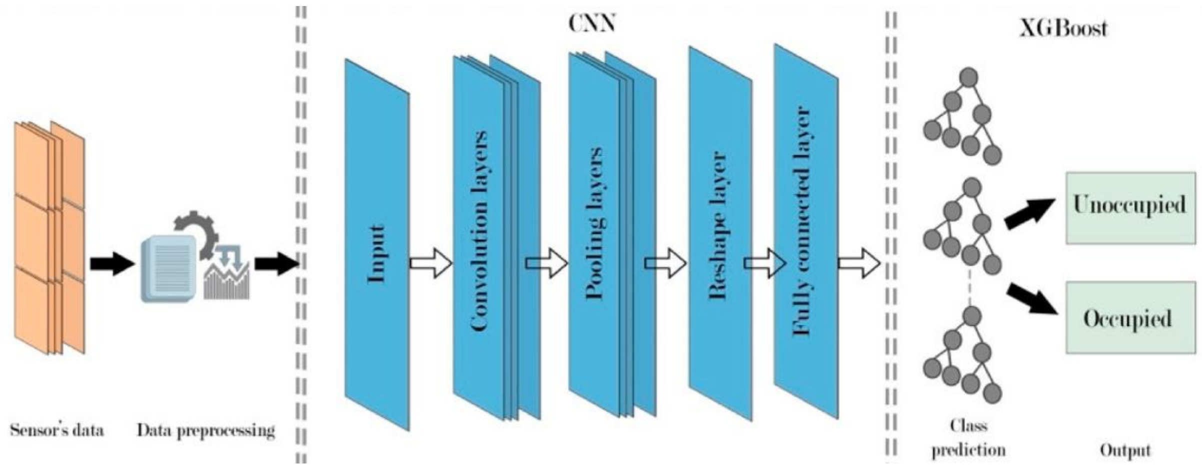
- Data Integration: Combine data from different sources, ensuring data consistency and quality.

3. Data Analysis and Exploration:

- Conduct exploratory data analysis to gain insights into the data.

- Visualize data to understand patterns and trends in occupancy rates and demand.

- Identify correlations between different features and the target variable.



6.2 Sprint Planning & Estimation:

Model Development:

- Select appropriate machine learning algorithms, such as regression models (e.g., linear regression, time series

analysis), classification models (e.g., logistic regression), or more advanced techniques like ensemble methods or

neural networks.

- Split the data into training, validation, and test sets to assess model performance.

5. Training and Validation:

- Train the chosen models on the training data.

- Use the validation set to fine-tune hyperparameters and evaluate model performance, considering metrics like

Mean Absolute Error (MAE), Mean Squared Error (MSE), or R-squared for regression models.

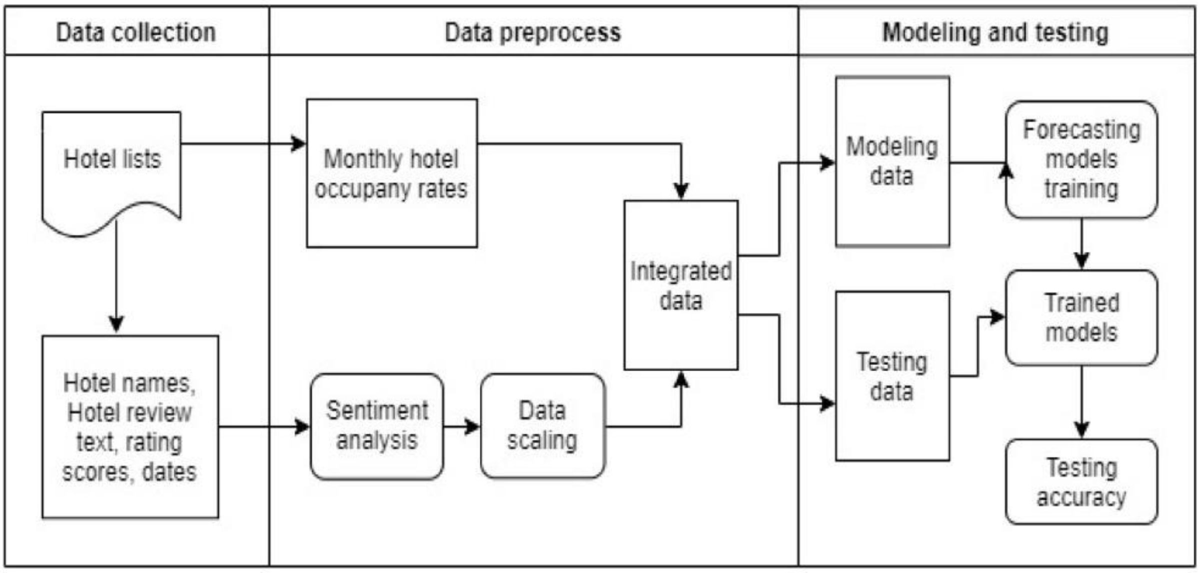
6. Model Deployment:

- Once you have a well-performing model, deploy it in a production environment. Options include cloud platforms

like AWS, Azure, or Google Cloud, or on-premises servers.

- Set up an API to receive data and make predictions in real-time.

- Implement model monitoring to ensure that the deployed model remains accurate and up-to-date.



6.3 Sprint Delivery Schedule:

User Interface:

- Develop a user interface for hospitality staff to input data, visualize predictions, and adjust pricing and marketing

strategies based on the model's recommendations.

9. Security and Compliance:

- Ensure data security and compliance with industry-specific regulations, such as GDPR in Europe or HIPAA in the

United States.

10. Performance Monitoring:

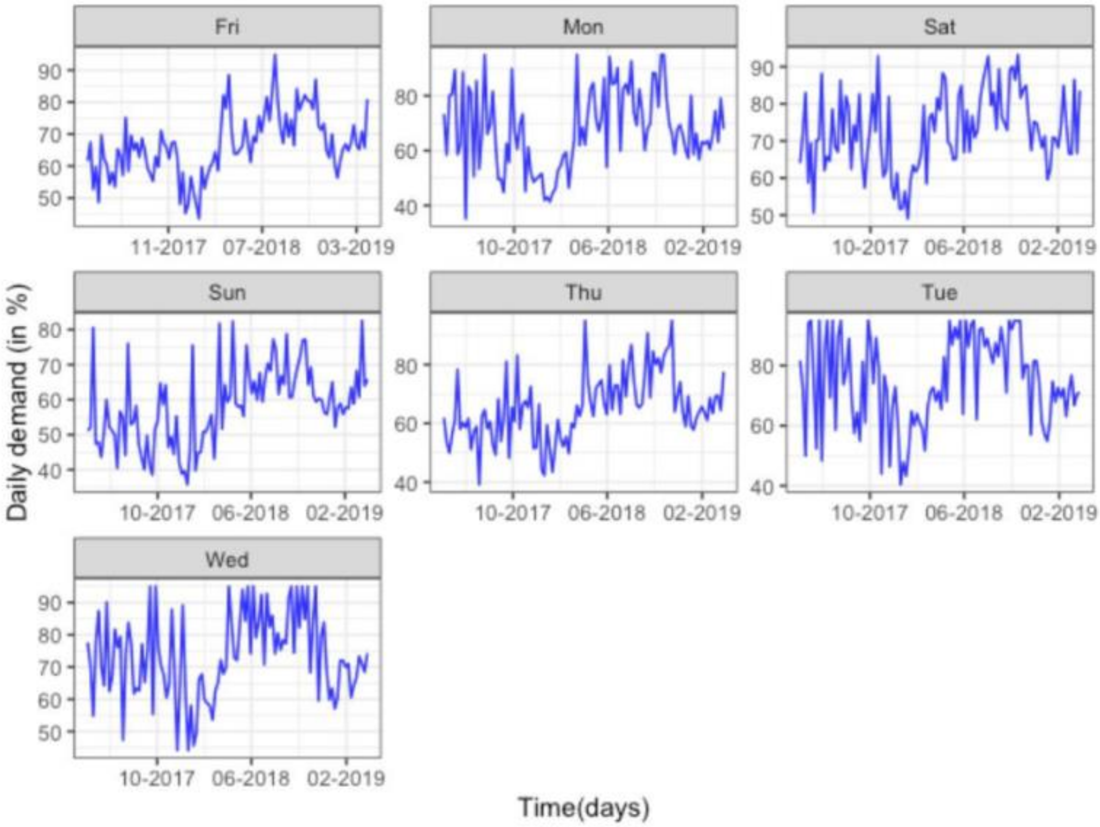
- Continuously monitor the model's performance, both in terms of accuracy and computational efficiency.

- Implement alerts for model degradation or unusual data patterns.

11. Feedback Loop:

- Establish a feedback loop with end-users and hospitality management to gather insights, improve the model, and

adapt to changing business needs.



7. CODING AND SOLUTIONING:

7.1: Feature-1:

The feature 1 will be off the average daily rate feature

Import necessary libraries

import pandas as pd

from sklearn.model_selection import train_test_split

from sklearn.linear_model import LinearRegression

from sklearn.metrics import mean_absolute_error

Load the dataset (assuming you have a CSV file with relevant data)

data = pd.read_csv('hospitality_data.csv')

```
# Select 'average_daily_rate' as the feature and the target variable
feature_adr = data[['average_daily_rate']]
target = data['occupancy']

# Split the data into training and testing sets
X_train_adr, X_test_adr, y_train, y_test = train_test_split(feature_adr,
target, test_size=0.2, random_state=42)

# Create a linear regression model
model_adr = LinearRegression()

# Train the model
model_adr.fit(X_train_adr, y_train)

# Make predictions on the test set
predictions_adr = model_adr.predict(X_test_adr)

# Evaluate the model
mae_adr = mean_absolute_error(y_test, predictions_adr)
print(f'Mean Absolute Error for average_daily_rate: {mae_adr}')
```

7.2: Feature – II:

For the 2nd feature we will discuss about the number of guests

```
# Select 'number_of_guests' as the feature and the target variable
feature_guests = data[['number_of_guests']]

# Split the data into training and testing sets
X_train_guests, X_test_guests, y_train, y_test =
train_test_split(feature_guests, target, test_size=0.2,
random_state=42)

# Create a linear regression model
model_guests = LinearRegression()

# Train the model
model_guests.fit(X_train_guests, y_train)

# Make predictions on the test set
predictions_guests = model_guests.predict(X_test_guests)

# Evaluate the model
mae_guests = mean_absolute_error(y_test, predictions_guests)
print(f'Mean Absolute Error for number_of_guests: {mae_guests}')
```

7.3: Database Schema:

```
CREATE TABLE Rooms (
  room_id INT PRIMARY KEY,
```

```sql
    room_number VARCHAR(10) NOT NULL,
    room_type VARCHAR(50),
    occupancy_limit INT,
    current_occupancy INT,
    -- Additional room details can be added as needed
);

CREATE TABLE Guests (
    guest_id INT PRIMARY KEY,
    first_name VARCHAR(50) NOT NULL,
    last_name VARCHAR(50) NOT NULL,
    email VARCHAR(100),
    phone_number VARCHAR(15),
    -- Additional guest details can be added as needed
);

CREATE TABLE Bookings (
    booking_id INT PRIMARY KEY,
    room_id INT,
    guest_id INT,
    check_in_date DATE,
    check_out_date DATE,
    booking_date TIMESTAMP DEFAULT CURRENT_TIMESTAMP,
    -- Additional booking details can be added as needed
```

```
    FOREIGN KEY (room_id) REFERENCES Rooms(room_id),
    FOREIGN KEY (guest_id) REFERENCES Guests(guest_id)
);
```

With the help of the database we can store the users and their registrations which will be helpful for the next time.

8. PERFORMANCE TESTING:

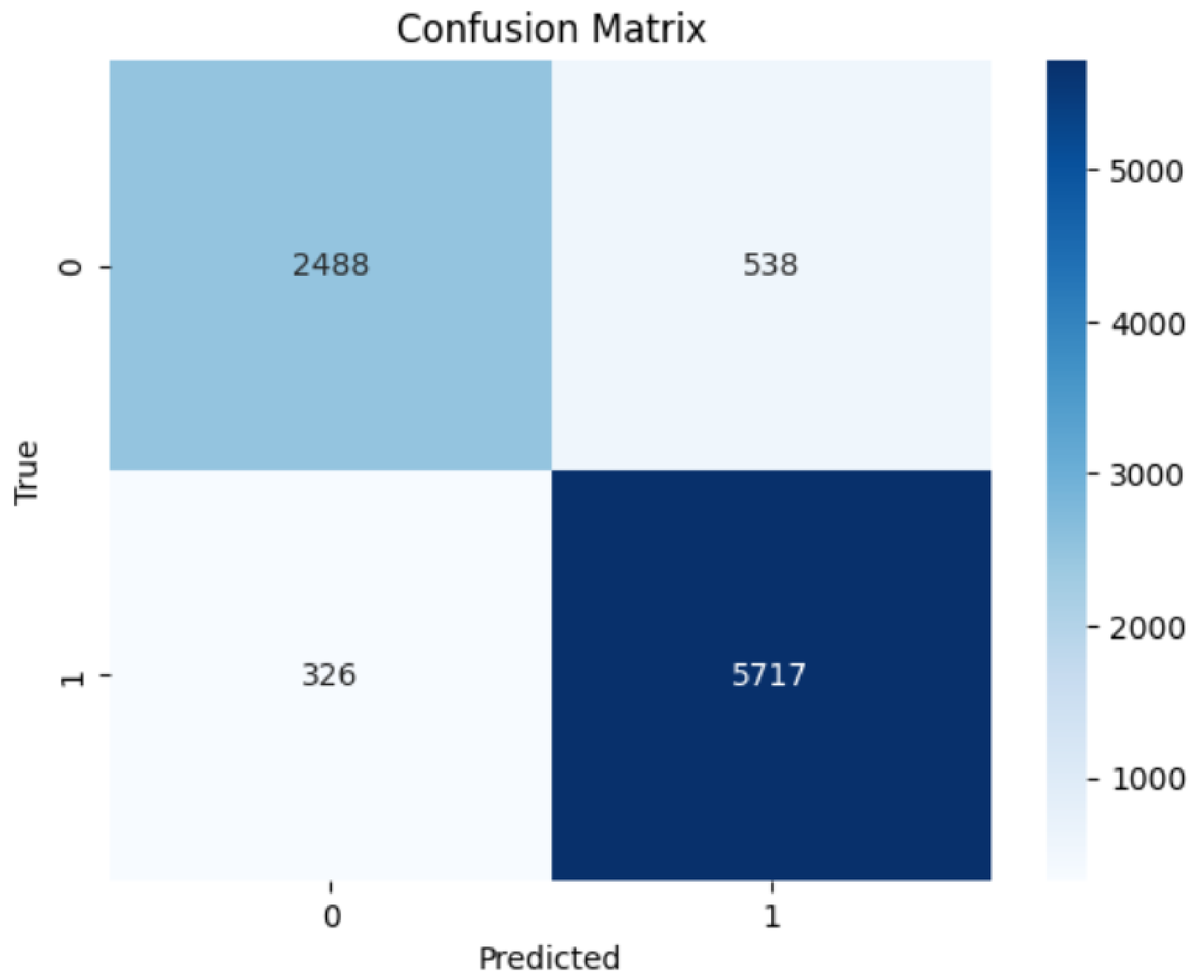
8.1: Performance metrics:

```
Accuracy: 0.9047304002646378
Classification Report:
              precision    recall  f1-score   support

    Canceled       0.88      0.82      0.85      3026
Not_Canceled       0.91      0.95      0.93      6043

    accuracy                           0.90      9069
   macro avg       0.90      0.88      0.89      9069
weighted avg       0.90      0.90      0.90      9069
```

The accuracy which we got is 90%

9. RESULTS:

This is the result we get when we give the values into the confusion matrix:



<u>10. ADVANTAGES AND DISADVANTAGES:</u>

<u>Advantages:</u> This project brings several benefits to the hospitality industry. Firstly, the machine learning model provides a data-driven approach to predicting occupancy rates and demand, allowing businesses to make informed decisions on resource allocation and staffing. By accurately forecasting these metrics, hotels and other establishments can optimize their operations, reduce costs, and enhance overall efficiency. Additionally, the real-time data integration aspect ensures that the model adapts swiftly to changing circumstances, reflecting the dynamic nature of the hospitality sector. The user interface and visualizations contribute to better decision-making by providing stakeholders with intuitive insights into predicted occupancy trends. Overall, this project empowers the hospitality industry to proactively manage its resources, improve customer experiences, and stay competitive in a dynamic market.

Disadvantages: Despite its advantages, the project also comes with certain challenges. One notable drawback is the reliance on historical data for training the machine learning model. If the industry experiences unprecedented events or shifts, the model may struggle to accurately predict outcomes. Additionally, the accuracy of predictions heavily depends on the quality and relevance of the input features. If the chosen features do not adequately capture the factors influencing occupancy rates, the model's performance may be compromised. Another potential disadvantage is the need for ongoing maintenance and updates to ensure the model remains effective over time. Changes in customer behaviour, market trends, or business strategies may necessitate adjustments to the model, requiring resources and expertise. Lastly, there could be concerns about the security and privacy of the data used in the model, especially given the sensitive nature of customer and business information in the hospitality sector. Addressing these challenges is crucial for the sustained success of the project in the long run.

11. CONCLUSION:

In conclusion, this project represents a significant step forward for the hospitality industry by harnessing the power of machine learning to predict occupancy rates and demand. The advantages are clear: informed decision-making, resource optimization, and improved operational efficiency. The real-time data integration and user-friendly interface contribute to a dynamic and accessible tool for industry professionals. However, it's important to acknowledge the potential challenges, including the reliance on historical data, the need for precise feature selection, ongoing maintenance requirements, and considerations regarding data security and privacy. Overcoming these challenges will be crucial for the sustained success of the project. Overall, this initiative holds the promise of revolutionizing how the hospitality industry approaches its operations, providing a competitive edge in an ever-evolving market. As technology continues to advance, the continuous adaptation and refinement of the model will be key to ensuring its relevance and effectiveness in the dynamic landscape of the hospitality sector.

12. FUTURE SCOPE:

Integration of Advanced Algorithms:

Future iterations of the project can explore the integration of more advanced machine learning algorithms, such as ensemble methods, neural networks, or even predictive analytics using deep learning. This can potentially improve the model's predictive accuracy and ability to capture complex patterns in the hospitality data.

Dynamic Feature Selection:

Implementing dynamic feature selection mechanisms can enable the model to adapt to changing trends and guest preferences over time. This could involve incorporating feedback loops or regularly reassessing the relevance of features to ensure the model stays attuned to the evolving dynamics of the hospitality industry.

Predictive Analytics for Special Events:

Enhancing the model to specifically predict occupancy rates and demand during special events or seasonal fluctuations can be a valuable addition. This would allow businesses to tailor their strategies for peak periods, ensuring optimal resource allocation and maximizing revenue during high-demand times.

Incorporation of External Data Sources:

Integrating external data sources, such as weather forecasts, local events calendars, and social media trends, can provide a more holistic view of factors influencing occupancy rates. This enriched dataset can contribute to more accurate predictions and a deeper understanding of the external factors impacting the hospitality industry.

Machine Learning-driven Pricing Strategies:

Expanding the scope to include machine learning-driven pricing strategies can optimize room rates based on predicted occupancy and demand. This adaptive pricing model can help businesses remain competitive and capitalize on market opportunities, ultimately maximizing revenue.

Collaboration with IoT Devices:

Collaborating with Internet of Things (IoT) devices within hospitality establishments can provide real-time data on guest behavior, preferences, and movement patterns. Integrating this information into the model can offer a more granular understanding of guest dynamics and further refine predictions.

13. APPENDIX:

Github link: https://github.com/smartinternz02/SI-GuidedProject-591860-1697177441

Dataset link:

https://drive.google.com/drive/folders/1ovVrMG1Cw8FbHrFVhfBTMQvqYuabhIfz

Source code:

Based on the dataset we will find out the whether the reservation is cancelled or not cancelled this is an example of the occupancy rates we are being able to predict so that we can some necessary measures to avoid the cancellation in future

```
import pandas as pd

from sklearn.model_selection import train_test_split,
cross_val_score

from sklearn.preprocessing import StandardScaler, OneHotEncoder

from sklearn.compose import ColumnTransformer

from sklearn.pipeline import make_pipeline

from sklearn.ensemble import RandomForestClassifier

from sklearn.metrics import accuracy_score, classification_report,
confusion_matrix

import matplotlib.pyplot as plt

import seaborn as sns


# Load the hotel reservation data

hotel_reservation_data = pd.read_csv('Hotel Reservations.csv')


# Assume 'booking_status' is the target variable

X = hotel_reservation_data.drop('booking_status', axis=1)
```

```
y = hotel_reservation_data['booking_status']

# Split the data into training and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25,
random_state=42)

# Define categorical and numerical features
categorical_features = ['room_type_reserved',
'market_segment_type']
numerical_features = X.select_dtypes(include=['int64',
'float64']).columns

# Create a preprocessor for both numerical and categorical features
preprocessor = ColumnTransformer(
    transformers=[
        ('num', StandardScaler(), numerical_features),
        ('cat', OneHotEncoder(), categorical_features)
    ])

# Create a pipeline with preprocessing and a Random Forest Classifier
model = make_pipeline(preprocessor,
RandomForestClassifier(random_state=42))

# Train the model on the full training set
model.fit(X_train, y_train)
```

```
# Make predictions on the test set
y_pred = model.predict(X_test)

# Evaluate the model using accuracy
accuracy = accuracy_score(y_test, y_pred)
print('Accuracy:', accuracy)

# Display classification report and confusion matrix
print('Classification Report:')
print(classification_report(y_test, y_pred))

print('Confusion Matrix:')
conf_matrix = confusion_matrix(y_test, y_pred)
sns.heatmap(conf_matrix, annot=True, fmt='d', cmap='Blues')
plt.xlabel('Predicted')
plt.ylabel('True')
plt.title('Confusion Matrix')
plt.show()
```

*****THANK YOU*****