



Hotel Booking Demand Analysis & Cancellation prediction

Data Science Use case

START

© 2020 Petroliaam Nasional Berhad (PETRONAS)

All rights reserved. No part of this document may be reproduced in any form possible, stored in a retrieval system, transmitted and/or disseminated in any form or by any means (digital, mechanical, hard copy, recording or otherwise) without the permission of the copyright owner.

Hotel Booking Demand Analysis & Cancellation prediction

Presented by
Dr Booma



Content

01. Tools & Technologies	03
02. Use Case – Hotel Booking	05
a. Hotel Booking - Overview	
b. Hotel Booking Pain Points	
c. Data	
d. Sample Examples	

Tools & Technologies

Tools & Technologies

1. Programming Languages –

- a. Preferably use **Python** or R
- b. If you are not comfortable with above, then can use any other programming languages

2. IDE –

- a. Preferably use **Jupyter**, Rstudio, Pycharm, Spyder, VS Code etc
- b. If you are not comfortable with above, then can use Azure ML or any other platforms

3. Visualization/ EDA –

- a. Preferably use **Python** or R
- b. If you are not comfortable with above, then can use Power BI, Tableau, Excel or any other tools

4. Documentation

- a. Preferably use Markdown languages
- b. If you are not comfortable with above, then can use any other tools/languages

5. Final Report –

- a. Share your development code base (if in a separate script)
- b. Provide the final report and approach analysis in Jupyter notebook, Google Colab or R markdown
- c. If you are not comfortable with above, use your preferred reporting tool (Word, PDF etc)

Use Case – Hotel Booking

Hotel Booking - Overview

Have you ever wondered:

- When is the best time of the year to book a hotel room?
- What is the optimal length of stay in order to get the best daily rate?
- What if you wanted to predict whether or not a hotel is likely to receive a disproportionately high number of special requests?
- Whether a booking is going to be cancelled?

This hotel booking dataset can help you to explore those questions!



Hotel booking pain points from the perspectives of Hotel, Guest and Data Science

- Hotels currently don't know about the guests' preferences, booking / cancelation patterns, deeper insights etc.
- It would be nicer for the hotels to have a model to predict if a guest will come.



Hotel

- Descriptive analytics

Q1. Provide various descriptive analytics and insights

- Data Science Use Cases

Q2. Suggest the use cases suitable here from the perspective of the hotel owner?



Guest

- Data Science Use Cases

Q3. Suggest the use cases suitable here from the perspective of the guests?



Data Science

Q4. Develop an unsupervised model to provide hidden insights and kinds of persona and explain the model as well as it's output.

Q5. Develop a classification model to predict bookings cancellation and explain the model as well as it's output.

Q6. Develop a regression model to predict the number of booking nights for any new booking.

Data

Hotel Dataset



hotel_bookings.csv.zip

Description

This data set contains booking information for a city hotel and a resort hotel and includes information such as when the booking was made, length of stay, the number of adults, children, and/or babies, and the number of available parking spaces, among other things.

Data

Dataset Features

Dataset contains following features:

- 1.hotel
- 2.is_canceled
- 3.lead_time
- 4.arrival_date_year
- 5.arrival_date_month
- 6.arrival_date_week_number
- 7.arrival_date_day_of_month
- 8.stays_in_weekend_nights
- 9.stays_in_week_nights
- 10.adults
- 11.children
- 12.babies
- 13.meal
- 14.country
- 15.market_segment
- 16.distribution_channel
- 17.is_repeated_guest
- 18.previous_cancellations
- 19.previous_bookings_not_canceled
- 20.reserved_room_type
- 21.assigned_room_type
- 22.booking_changes
- 23.deposit_type
- 24.agent
- 25.company
- 26.days_in_waiting_list
- 27.customer_type
- 28.adr
- 29.required_car_parking_spaces
- 30.total_of_special_requests
- 31.reservation_status
- 32.reservation_status_date

The background of the slide is a whiteboard filled with various hand-drawn business diagrams. These include flowcharts, organizational charts, line graphs, bar charts, pie charts, and circular diagrams. Some diagrams are labeled with words like 'PLAN', 'MANAGEMENT', 'IDEA', 'PROGRESS', 'NEXT', 'START', 'OPTIONS', 'UP', 'down', and 'MAX'. There are also small human figures and arrows indicating flow and direction. The overall theme is business strategy and analysis.

Descriptive analytics

Q1. Provide various descriptive analytics and insights

Determining Hotel Reservation Cancellation Inferences Using Descriptive Analysis

Goal

The primary objective of this study is to discover the inferences on booking cancellation in Hotel side using descriptive analysis.

[Python Script link](#)

Conclusion

Based on descriptive analysis, it is clear that a third of all bookings have been canceled by customers. Over 50% of cancellations are made by group segment, and offline and internet TA/TOs have cancellation rates above 33%. However, lower cancellation rates are made from live segments. According to this analysis, it is easier to understand that the majority of individuals do not seem to like a stay in a hotel longer than a week. However, hotel stays of up to 12-13 days seem to be common. Staying longer than 15 days will inevitably generate outlier values for each segment, although this varies by segment. This analysis also shows that the number of guests staying at hotels in the city is increasing every year at a rate of . reservation rate over 60% of the population. Portugal, UK, France, Spain, Germany, Ireland, Italy and Belgium are ranked in the top 15 countries based on guest bookings and length of stay at both hotels. In 2016, the number of reservations doubled compared to 2015. However, the number of reservations in 2017 decreased by about 15%. As observed, July and August are the busiest months although prices are quite high compared to other months.



Data Science Use Cases

Q2. Suggest the use cases suitable here from the perspective of the hotel owner

Suggest the use cases suitable here from the perspective of the hotel owner

➤ Hotel Website

- ✓ Make Hotel name, website consistent and secure across all OTAs
- ✓ Make Hotel website with updated pictures
- ✓ Make Hotel website User-friendly to view details and do booking
- ✓ Identify the -new guests and repeated guests visited the website, track their interest to make them hooked on the website(based on CDP)

➤ Brand Monitoring

- ✓ Gather review to offer possible revenue streams apart from room selling and also to identifying the guests satisfaction scale to improve their facilities(customer Behavior)

➤ Booking Promotions/Revenue Management

- ✓ Identify most guests come from which country
- ✓ Identify the busiest month to manage inventory, meals, capacity of the hotel,etc
- ✓ Identify the length of stay



Data Science Use Cases

Q3. Suggest the use cases suitable here from the perspective of the guests

Suggest the use cases suitable here from the perspective of the guests

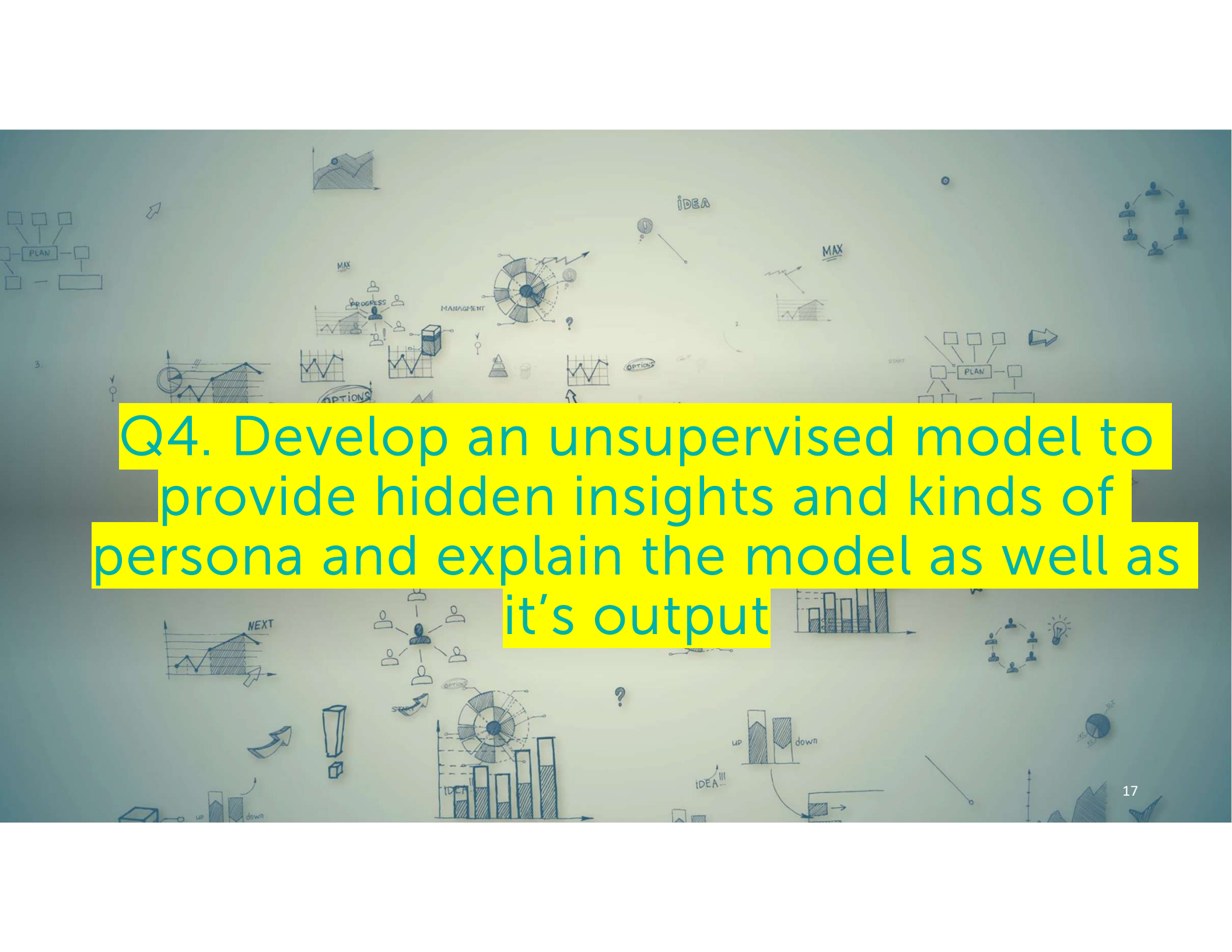
➤ **Booking Engine**

- Attractive website / Mobile app with all necessary information
- Updated information and reliable Ratings

➤ **Identify best time to book the hotel based on price, location, amenities ,weather etc**

➤ **Price Management**

➤ **Notification system**



Q4. Develop an unsupervised model to provide hidden insights and kinds of persona and explain the model as well as it's output

Develop an Unsupervised Learning method using Customer Segmentation to Identify Hidden Insights in Hotel Booking Cancellation

Goal

The aim of this study is to unwrap the insights on booking cancellation using unsupervised learning

[Python Script link](#)

Conclusion

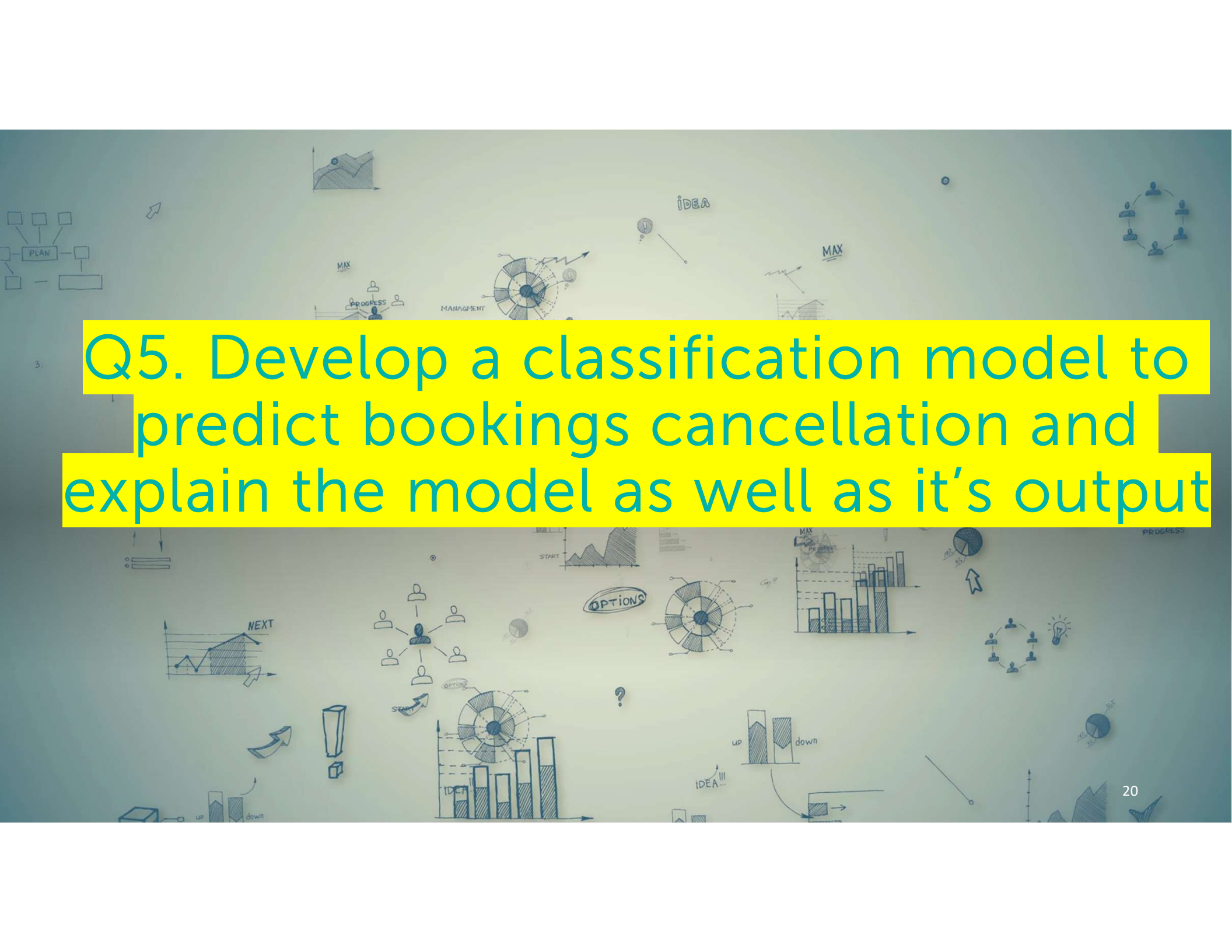
This study conducted to unwrap the information using unsupervised learning - clustering in the dataset regarding the hotel booking cancellation. Both k-means and k-prototype clustering was used in this study. In order to handle both numerical and categorical data, I have decided to go with K- prototype than K-means. K-prototpe measures the distance between numerical features using the Euclidean distance , but also measures the distance between categorical features using the number of corresponding classes. The clusters generated by K-Prototypes outperformed than K-means in both visually and empirically. Although both algorithms divide customers into distinct groups, the groups created by k prototype are more informative and therefore potentially more useful to marketers.

Findings:

This approach helped to identify profitable “guests based on Customer Segmentation” . As per the study conducted,, cancellation, total_guest, repeat guest, and the LOS were the other 4 client clusters that we were able to pinpoint. Hotels can use those clusters to better anticipate their visitors, interact with them in a more focused way, and estimate the risk of cancellation. in order to attract guests and increase the business, Hotel owner could offer special discount / promotions for online segment and groups segment guests.

Future Enhancement:

As a future enhancement, the clustering result could be used as a target for the best fit classification models, in order to predict the booking cancellation

The background of the slide is a wall covered in various hand-drawn business sketches and diagrams. These include flowcharts, pie charts, bar graphs, and circular diagrams. Some of the visible text in the sketches includes 'PLAN', 'PROGRESS', 'MANAGEMENT', 'IDEA', 'MAX', 'START', 'NEXT', 'OPTIONS', 'UP', 'down', and 'IDEA III'. There are also small icons of people and arrows indicating flow or direction.

Q5. Develop a classification model to predict bookings cancellation and explain the model as well as it's output

Forecasting Hotel Reservation Cancellation to predict Length of Night Stay


Goal

The primary objective of this study is to identify appropriate machine learning regression model to forecast length of night stays.

[Python Script link](#)

Conclusion

Now, I can predict the length of a night stay in hotel for a new data using data with the highly important features. Such a model has the potential to profoundly improve hotel management. The most surprising aspect of this work was how the lead_time played a more important role than age when predicting the length-of-stay. By far, the most challenging aspect of this project was the feature engineering. Given that the lead_time have such a strong feature importance, it would be worth evaluating wwhich would yield a better prediction model. My theory is that the prediction model would become more accurate (lower MAE) with this optimization.



Q6. Develop a regression model to predict the number of booking nights for any new booking.

A Classification Approach to Predict Hotel Booking Cancellation Using Classification Techniques

Goal

The goal of the study is to forecast hotel cancellations using machine learning and examine the most significant influencing factors.

[Python Script Link](#)

Conclusion

This study addresses the problem of class imbalance in predicting hotel cancellations. I have used the SMOTE oversampling technique to solve this problem. This study shows that after solving this problem with SMOTE, the performance of machine learning classifiers is significantly improved. All classifier models show significant improvements over the unbalanced and balanced data. Among them, Adaptive boosting along with SMOTE, which produces a score as high as 99.77% in all performance metrics included in this study. The proposed method can solve the problem of unbalanced data sets, and predictive models can allow hotel managers to calculate their losses based on various scenarios.

The prediction model can be implemented to the hotels booking system to help with the process, as the most significant factors in the Adaptive boosting model is Reservation status a categorical variable represents booking status.

A cancellation model could allow hotel owners to implement less permissive tactics without amplifying their vulnerabilities. This could lead to more offers, as a more flexible booking strategy creates more customers. In addition, these classifiers allow hotel management to predict and prepare for possible reservation cancellations. Additionally, the hospitality industry can benefit from this kind of approach to increase revenue by improving classifier performance through more accurate demand forecasting

The study conducted has some constraints. Firstly, the majority of the dataset features describe customers. The current system is acceptable, despite the fact that guests actions frequently reveal details about the hotels they have reserved. However, the forecast can be more accurate by including more hotel-specific details in the algorithm. Additionally, the dataset only includes data on reservations made between July 1, 2015, and August 31, 2017, which is not the most recent data.

Thank you for your passion!

