BlackTECH

Hotel Executive Summary
Loan Prediction

Presentation

Presenter by:

Patrick Ugheoke Sebastine

Overview

- BlackTECH hotel executive summary database was provided by the revenue management team. The data were analysed to provide insights for the hotel's growth. The insights include revenues and shortfalls, with recommendations on how to improve for the directors and shareholders.

- In addition, the collections team of BlackTECH financial institution reported their challenges with respect to loans that were granted to management. The cohort 4 analytics team were provided data for the last 6 months and tasked with developing a prediction model that when a customer's information is inputted it can predict whether the loan will be rolled (1) or not rolled (0).

Hotel Executive Summary Observations

- 3 date entries on the booking table had data integrity/anomaly issues. It was found that the dates were wrongly captured and corrected with the help of the request table.

(Booking_id	Room_#	Check_in_date	check_out_date	Request_id)
(1919	L11	1/7/1916	1/17/1916	938)
(6456	S91	2/29/1916	3/9/1916	3101)
(9516	X75	3/1/1916	3/3/1916	4604)

- Inconsistent column names were found during data import in SQL but were corrected to maintain consistency across tables.

- The Food Order table had a stable join with the Menu table, but joining with the Booking, Request, and Room tables resulted in too many null values, repetitions of food orders, and order dates.

- We split the final table into two, the Restaurant and Reservation tables, as a solution.

- 281 missing records were found during the join of the Request and Booking tables. Analysis suggests that these were "not confirmed" or "cancelled" bookings, a practice that is normal in hotel management systems.

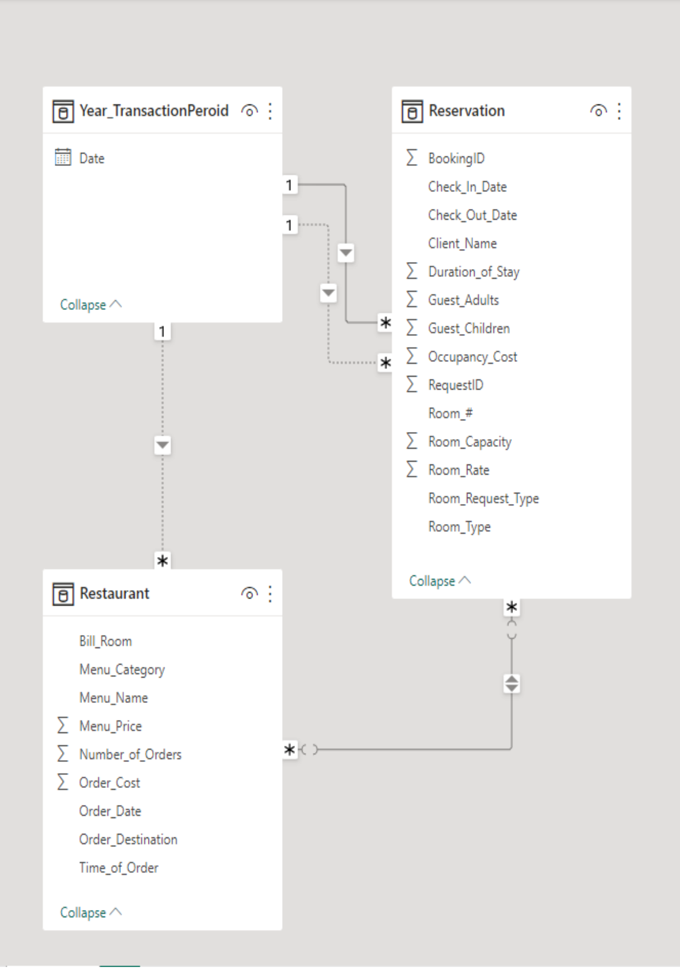
- To visualize accurate time series and trends, we created a new table using DAX on Power BI, linking check-in date, check-out date and order date. (The transaction period is from 1st January 2016 to 13th April 2016)

Hotel Executive Summary Insights

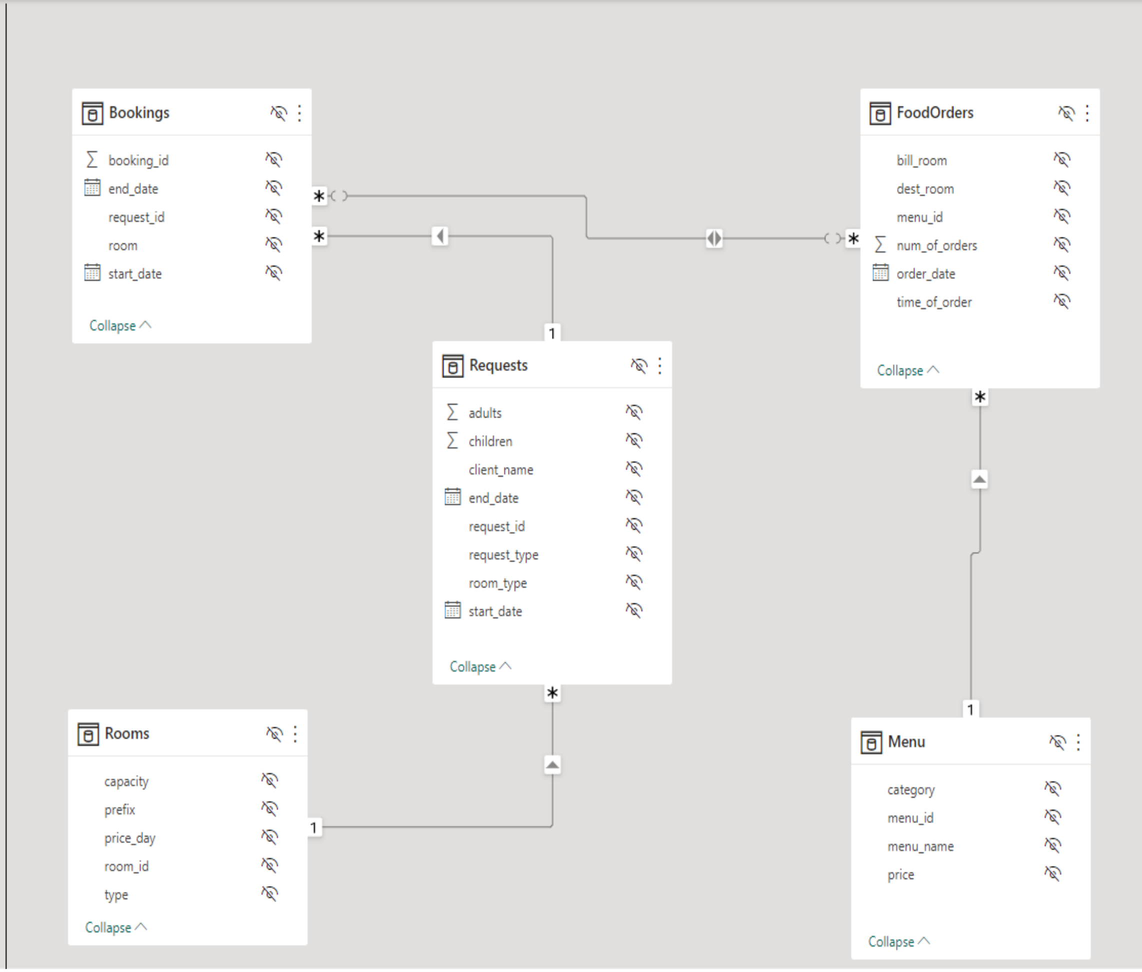
- In order to improve the efficiency of data storage and entry, we propose implementing a revamped system for inputting information into the hotel's database that includes a data validation component to notify users of any inaccuracies and incorporating SQL to normalize the data.
- Based on our analysis, normal rooms generated the highest revenue at $1,215,280, accounting for 26.08% of total booking revenue. However, Tuesdays were identified as the lowest booking days for normal rooms. We recommend offering discounted rates mid-week to further increase the revenue it generates.
- Based on our analysis, we found that the menu item 'steak 'n stuff' had the highest sales at $9,990, and soft drinks had the lowest sales at $1,365. To increase sales, we recommend offering combo menu specials that combine items from our top 5 and bottom 5 menu categories.
- Our data shows that the busiest date was January 30th, 2016, and a similar trend occurs at the end of each month. To meet the increased demand at the end of the month, it would be advisable to offer temp contracts to additional staff.
- Given the high demand for both holiday bookings and normal room bookings, it would be beneficial to open an additional branch specifically catering to holiday makers.
- To reward our frequent and top paying customers, it would be beneficial to introduce a loyalty program for clients who spend above a certain amount.

Entity Relationship Diagram (ERD)

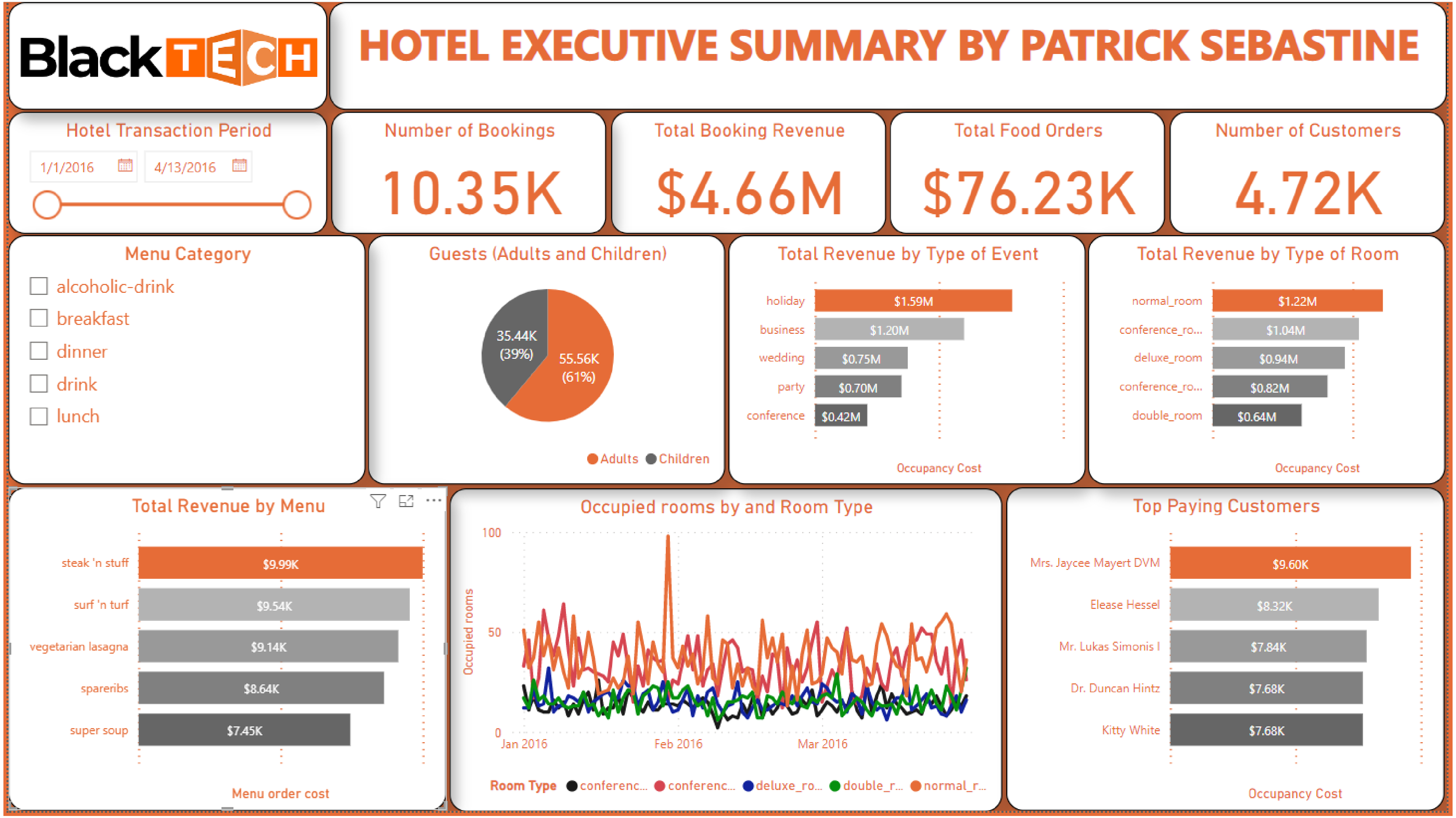
Joined tables ERD



Original tables ERD



Dashboard


BlackTECH
HOTEL EXECUTIVE SUMMARY BY PATRICK SEBASTINE
Hotel Transaction Period
1/1/2016
4/13/2016
Number of Bookings
10.35K
Total Booking Revenue
$4.66M
Total Food Orders
$76.23K
Number of Customers
4.72K
Menu Category
alcoholic-drink
breakfast
dinner
drink
lunch
Guests (Adults and Children)
35.44K (39%)
55.56K (61%)
Adults
Children
Total Revenue by Type of Event
holiday
$1.59M
business
$1.20M
wedding
$0.75M
party
$0.70M
conference
$0.42M
Occupancy Cost
Total Revenue by Type of Room
normal_room
$1.22M
conference_ro...
$1.04M
deluxe_room
$0.94M
conference_ro...
$0.82M
double_room
$0.64M
Occupancy Cost
Total Revenue by Menu
steak 'n stuff
$9.99K
surf 'n turf
$9.54K
vegetarian lasagna
$9.14K
spareribs
$8.64K
super soup
$7.45K
Menu order cost
Occupied rooms by and Room Type
100
50
0
Occupied rooms
Jan 2016
Feb 2016
Mar 2016
Room Type
conferenc...
conferenc...
deluxe_ro...
double_r...
normal_r...
Top Paying Customers
Mrs. Jaycee Mayert DVM
$9.60K
Elease Hessel
$8.32K
Mr. Lukas Simonis I
$7.84K
Dr. Duncan Hintz
$7.68K
Kitty White
$7.68K
Occupancy Cost

Loan Prediction Observations

- Dataframe imported had 5783 records (rows) and 11 fields (columns).
- There were 377 records, translating to 6.5% of null values. These null values were subsequently dropped.
- There were 533 duplicate values. The first values were retained, and the second values were subsequently dropped from the Dataframe.
- The 'loan_id' and 'delq_history' fields were not needed for our analysis and were therefore dropped.
- The following fields were derived to assist our prediction model in making better predictions:

  Debt-to-income ratio: This ratio is a measure of a borrower's financial stability, calculated by dividing the borrower's monthly debt payments by their monthly income.

  Interest rate: The interest rate on the loan may affect the risk of default. Interest rate can be calculated using the formula ((closing_principal_balance/original_loan_amount)^(1/original_loan_term)-1).

- After the data cleaning process, our Dataframe was left with 4873 records and 12 fields.

Loan Prediction Insights (1)

Machine learning prediction using Python.

For our independent features (X), the following fields were used;

#	Fields	Non-Null Count	Datatype
1	monthly_income	4873 non-null	float64
2	origination_score_band	4873 non-null	int64
3	TOB_months	4873 non-null	float64
4	closing_principal_balance	4873 non-null	float64
5	original_loan_amount	4873 non-null	float64
6	product	4873 non-null	float64
7	original_loan_term	4873 non-null	int64
8	remaining_loan_term	4873 non-null	int64
9	interest_rate	4873 non-null	float64

'Product' field was converted from an object datatype to float64 datatype using ordinal encoder, as label encoder did not predict the best outcome.

And the 'target' field was used for our dependent features (y)

#	Fields	Non-Null Count	Datatype
1	target	4873 non-null	int64

Loan Prediction Insights (2)

Logistics Regression Classification Model		
1	Test data size and Random state	test_size=15% and random_state=190
2	Model parameters	C=1.0, penalty='12', max_iter=200
3	Accuracy score	72.78%
4	Model score	69.85%
5	Precision score	80.00%
6	Confusion Matrix	array([[528, 1], [198, 4]], dtype=int64)

Random Forest Classification Model		
1	Test data size and Random state	test_size=0.25, random_state=42
2	Model parameters	n_estimators=200, random_state=100
3	Accuracy score	100.0%
4	Model score	100.0%
5	Precision score	100.0%
6	Confusion Matrix	array([[841, 0], [0, 378]], dtype=int64

XGBoost (eXtreme Gradient Boosting) Model		
1	Test data size and Random state	test_size=0.2, random_state=42
2	Model parameters	objective="binary:logistic", random_state=42
3	Accuracy score	100.0%
4	Model score	100.0%
5	Precision score	100.0%
6	Confusion Matrix	array([[668, 0], [0, 307]], dtype=int64)

Presented by:
Patrick Ugheoke Sebastine

LinkedIn
https://www.linkedin.com/in/patrickugheokesebastine/

Twitter
https://twitter.com/iamreddsebastin

GitHub
https://github.com/PatrickSebastine

Kaggle
https://www.kaggle.com/patricksebastine

Thank you

BlackTECH