Predicting Hotel Booking Cancellation

By Julia Kelman

Improving hotels operations and revenues

The Problem with Cancellations

Customers accustomed to free cancellation policies

Operational problems

40% Cancellation rate in 2018

Reviews influenced $546 billion of travel spending in 2017

Non accurate forecast

Increase in online reputation score linked to increase in occupancy and revenue

➡ Non-optimized occupancy, poor management, revenue loss

The Answer

In order to fight the negative effects of cancellations, hotels need to be able to identify which bookings are likely to be canceled.

We will use a **real life hotel booking dataset** to create a **customer segmentation analysis** in order to gain insights about the customers (and hopefully reasons why they cancel their reservation).
We will then build a **classification model** to **predict** whether or not a **booking will be canceled** with the **highest accuracy** possible.

This model will allow hotels to predict if a new booking will be canceled or not, manage their business accordingly, and increase their revenue.

The Data

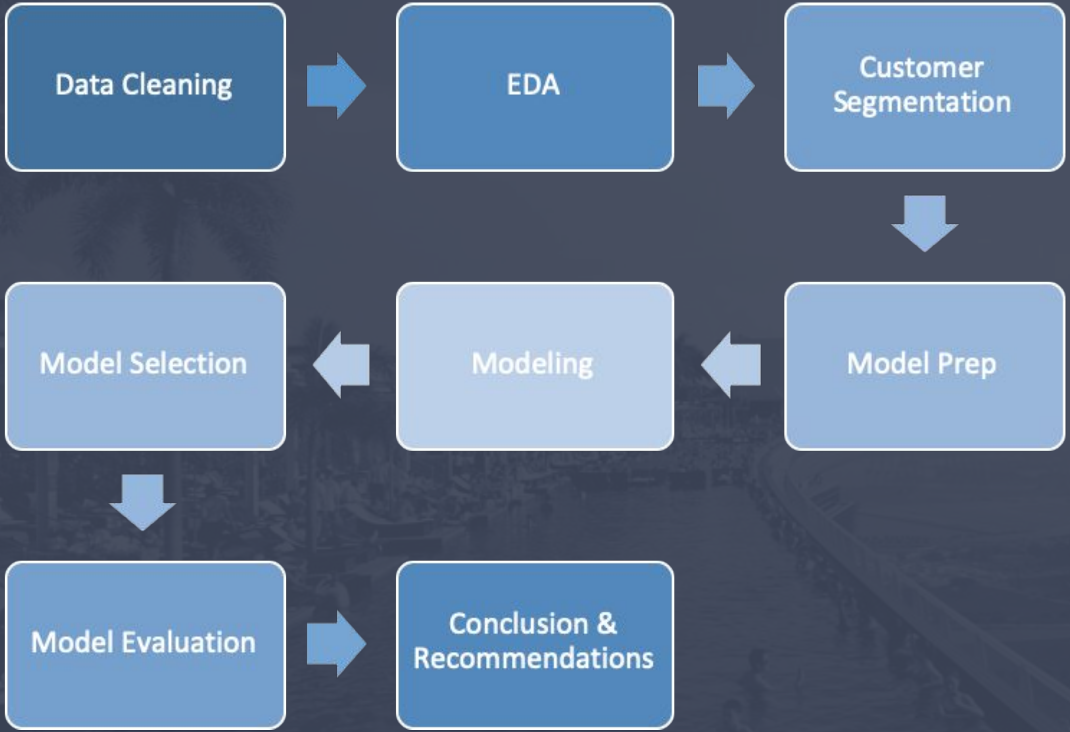
Real data from real hotels

From: Property Management Systems

Bookings due to arrive between the July 1, 2015 and August 31, 2017.

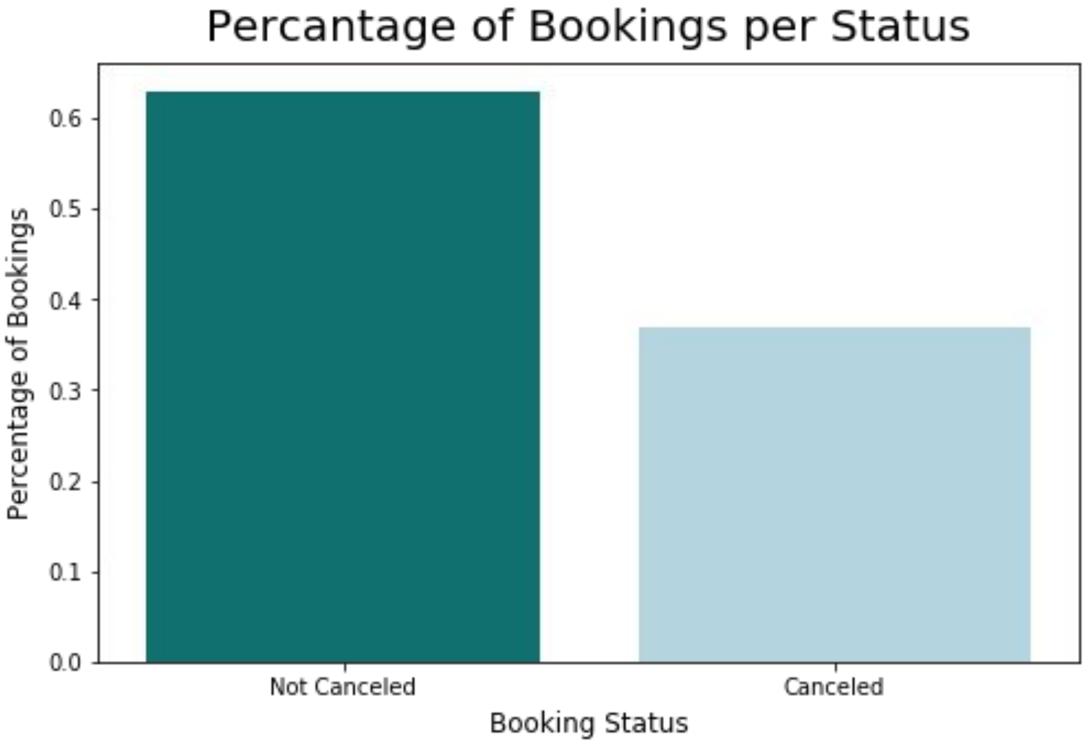
40,060 Hotel 1 (Algarve)
79,330 Hotel 2 (Lisbon)
31 Variables

Workflow


Data Cleaning
EDA
Customer Segmentation
Model Prep
Modeling
Model Selection
Model Evaluation
Conclusion & Recommendations

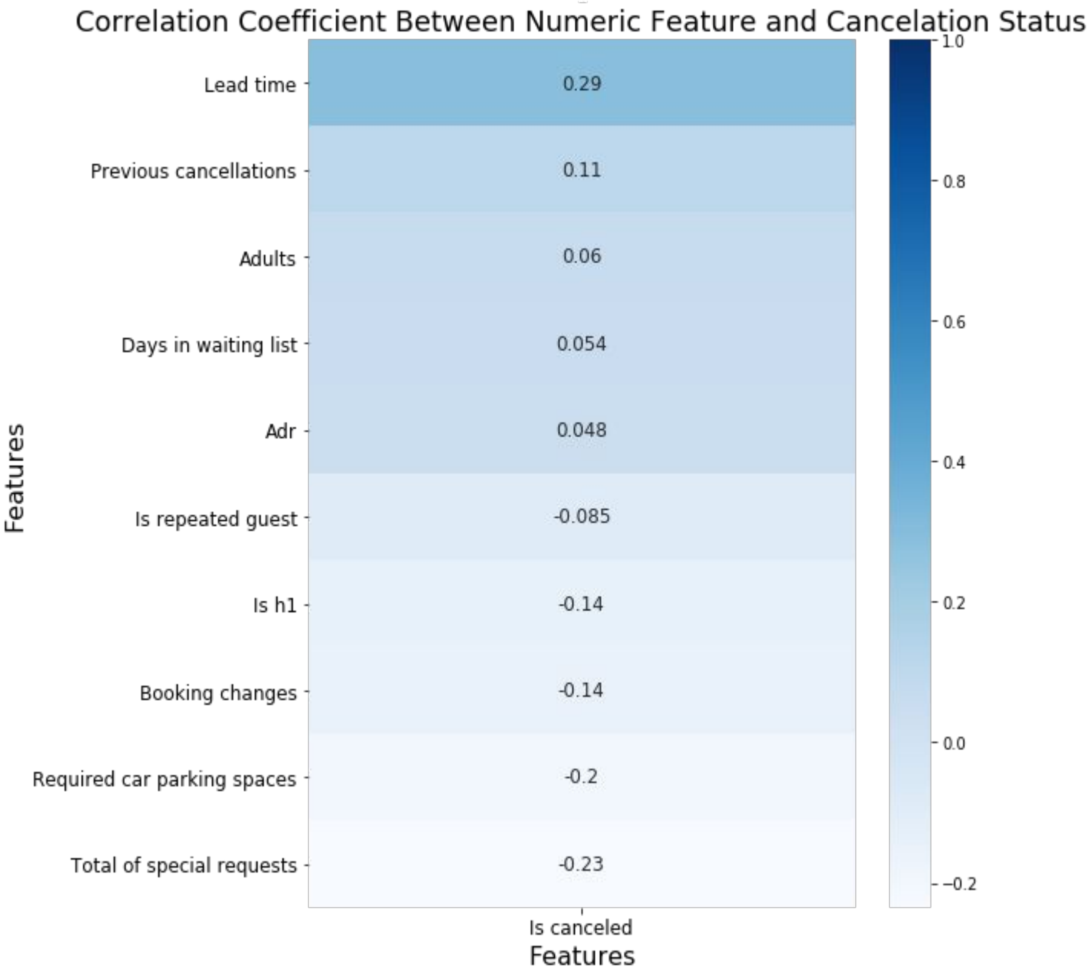
Percentage of Canceled Bookings

- 37% of bookings canceled



Features correlated with cancellation

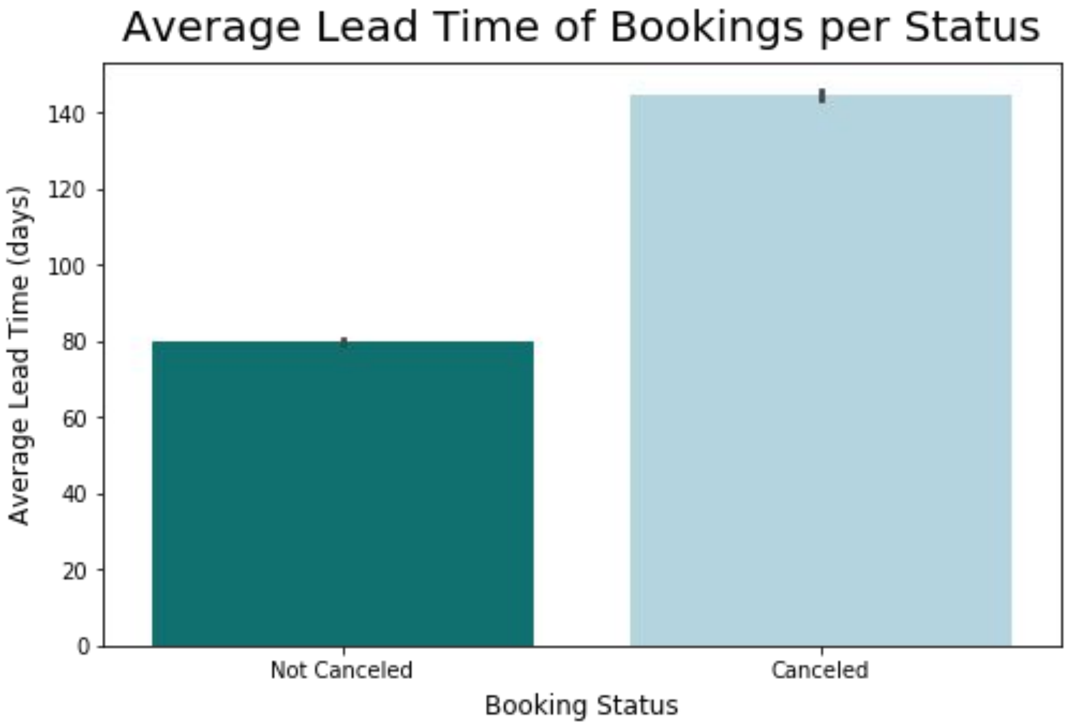
- Lead Time
- Special Requests
- Parking Spaces
- Booking Changes
- Previous Cancellations



Lead Time

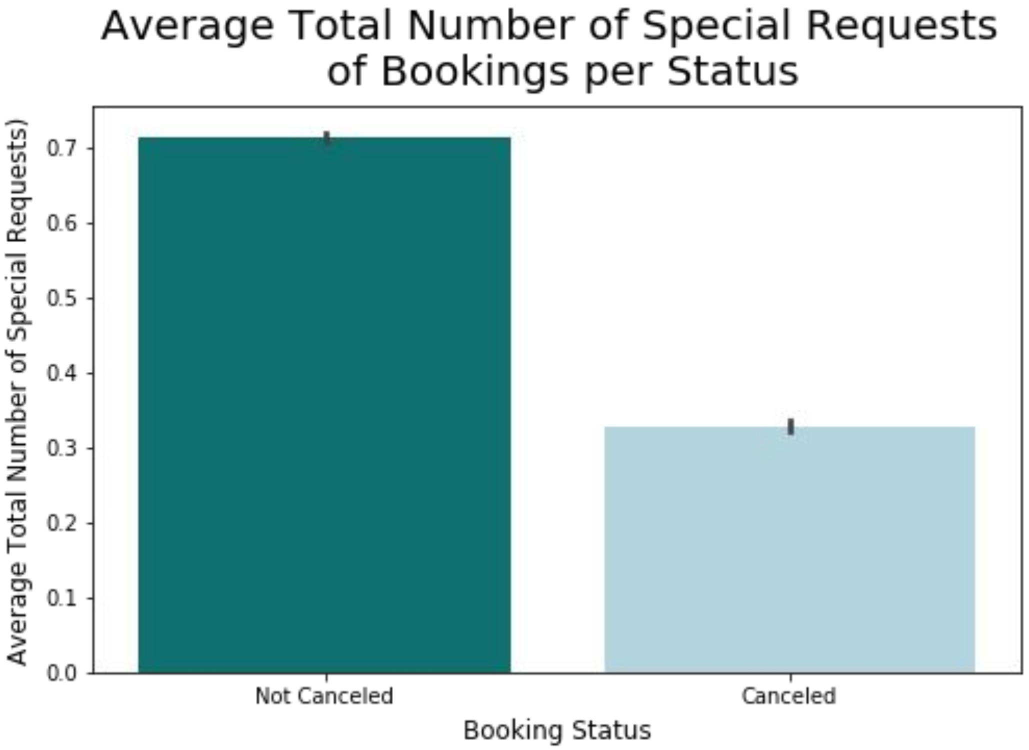
Days between booking and arrival

- Canceled bookings have longer average lead time
- More time to cancel
- More time for unexpected events



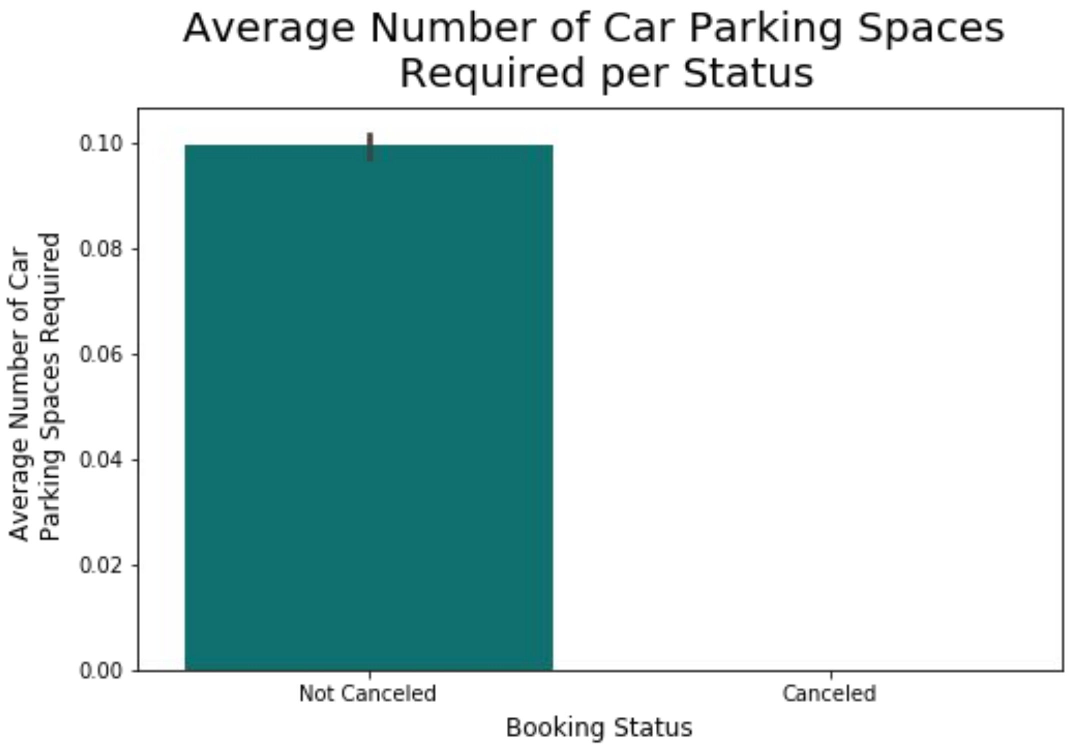
Special Requests

- Canceled bookings have lower average number of special requests
- Engagement
- Communication between customer and hotel



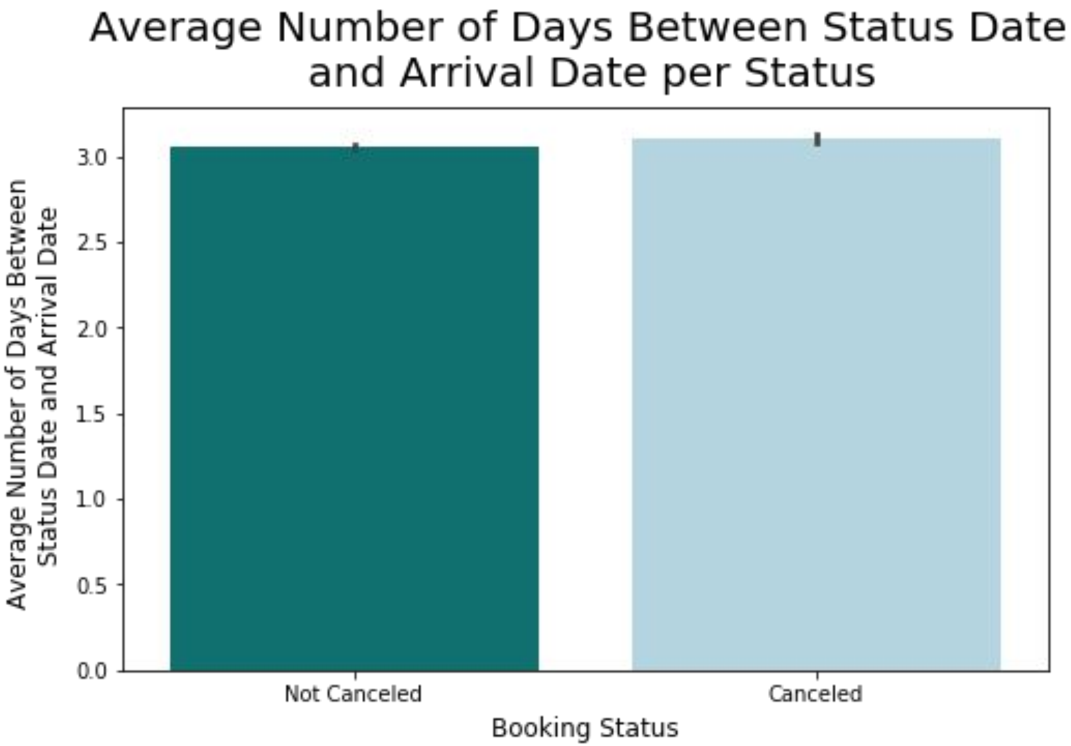
Parking Spaces

- Canceled bookings have lower average number of required parking spaces
- Engagement
- Shows commitment to destination
- Limit customer hotel option



Difference Between Status Date and Arrival Date

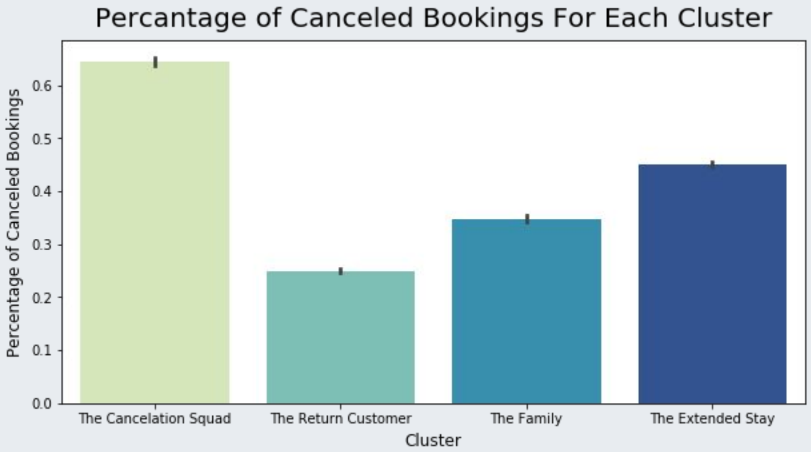
- Stays 3 nights in hotel on average
- Cancel 3 days prior to arrival on average

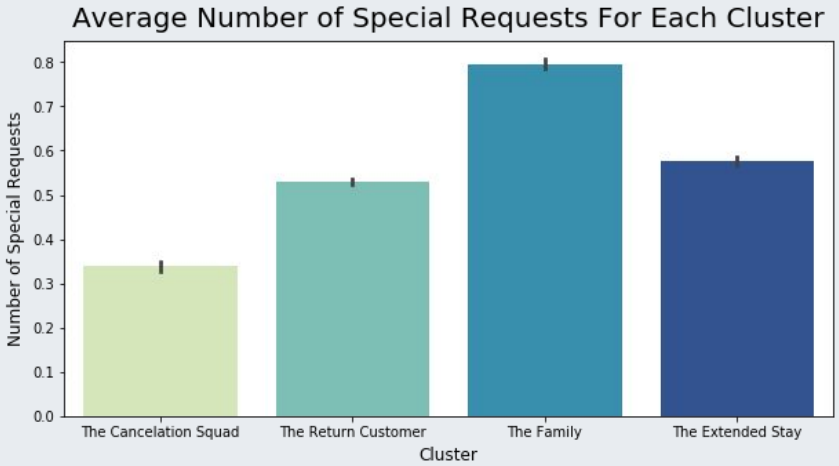


Deposit Type

- Customers who pay a non-refundable deposit have a much higher percentage of canceled reservations
- Transient groups who use a travel agent
- Hotel deposit policies

Customer Segmentation





The Cancellation Squad
- High cancellation
- Long lead time
- Higher previous cancellations
- September arrival date
- Groups

The Return Customer
- Low cancellation
- Short lead time
- Higher number of previous bookings not canceled

The Family
- Higher mean number of children and babies
- Higher room price
- Higher number of special requests
- August arrival date

The Extended Stay
- Higher average number of weekend and week nights

Modeling

X: 27 booking features

Y: Cancellation

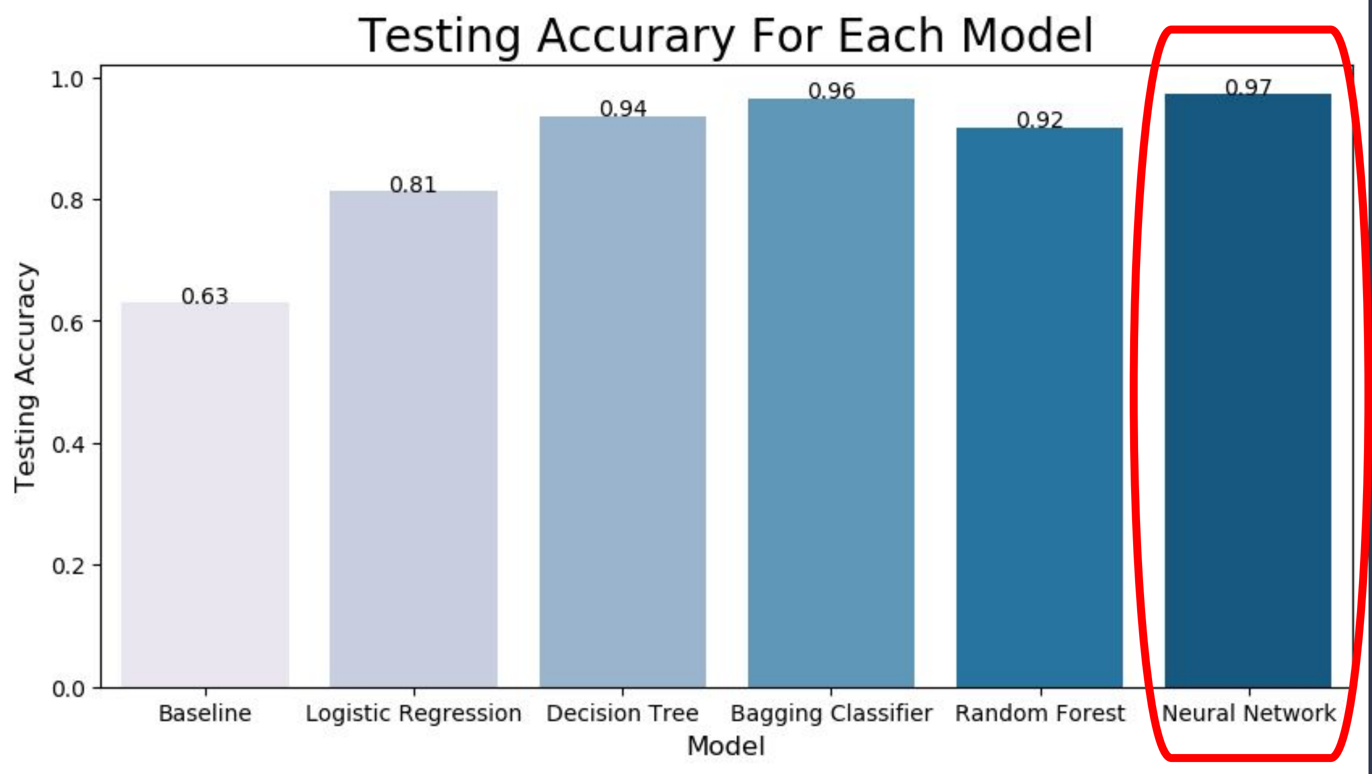
6 Models:

- Baseline
- Logistic Regression
- Decision Tree
- Bagging Classifier
- Random Forest
- Neural Network

Model Selection

Based on Accuracy

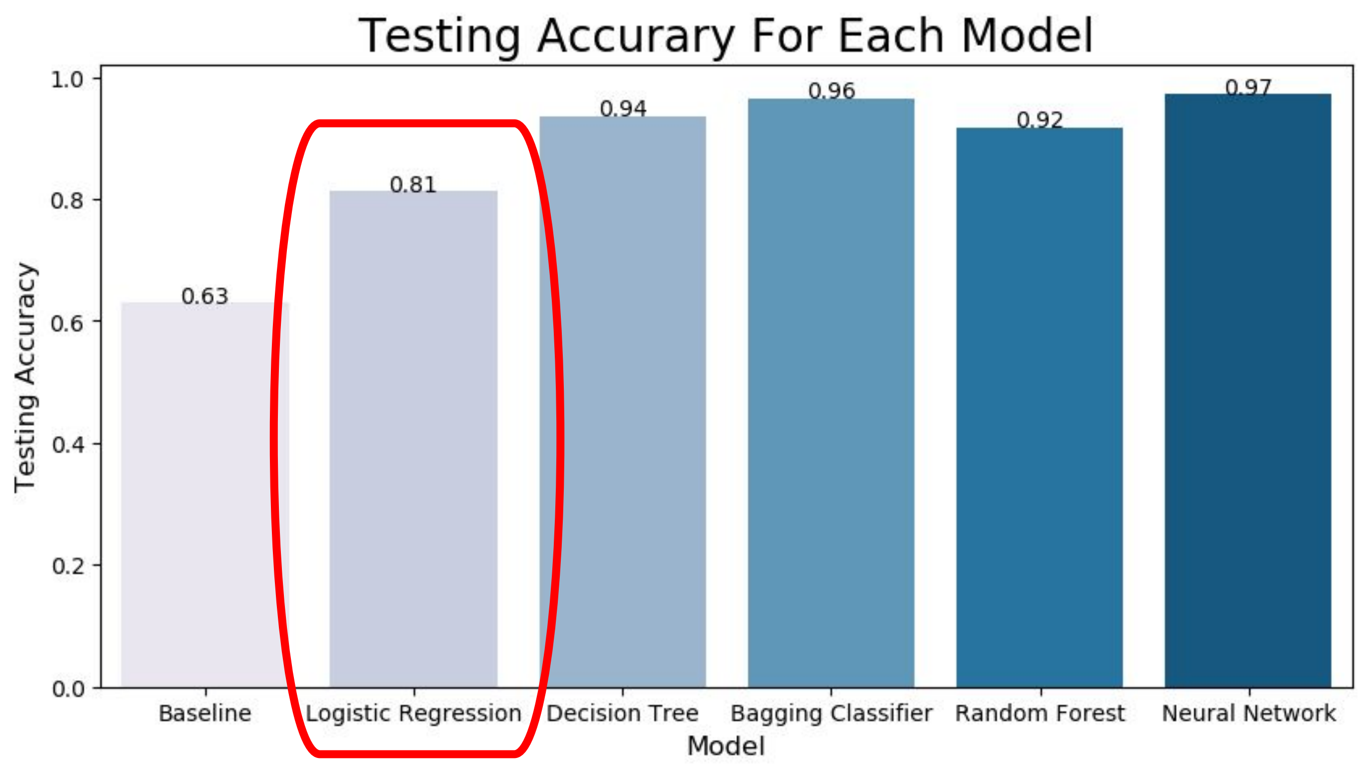
Predictive model



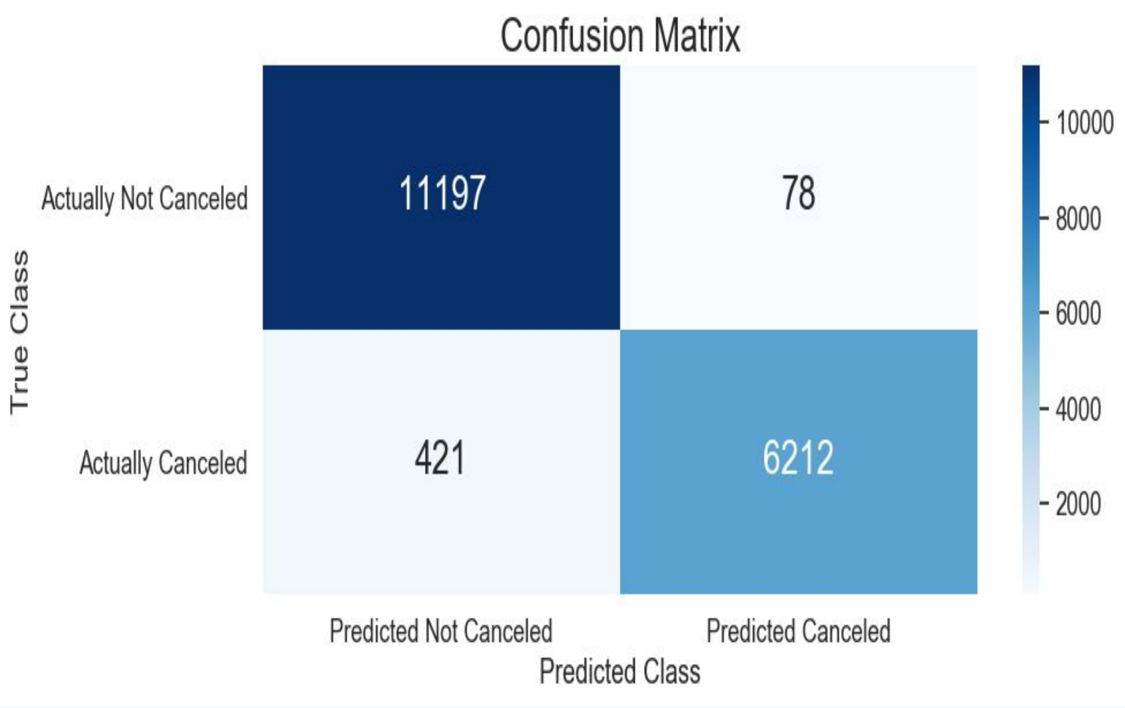
Model Selection

Interpretive model

Based on Accuracy



Neural Network Evaluation: Confusion Matrix



Correctly classifying:

- **94%** of canceled bookings
- **99%** of not canceled bookings

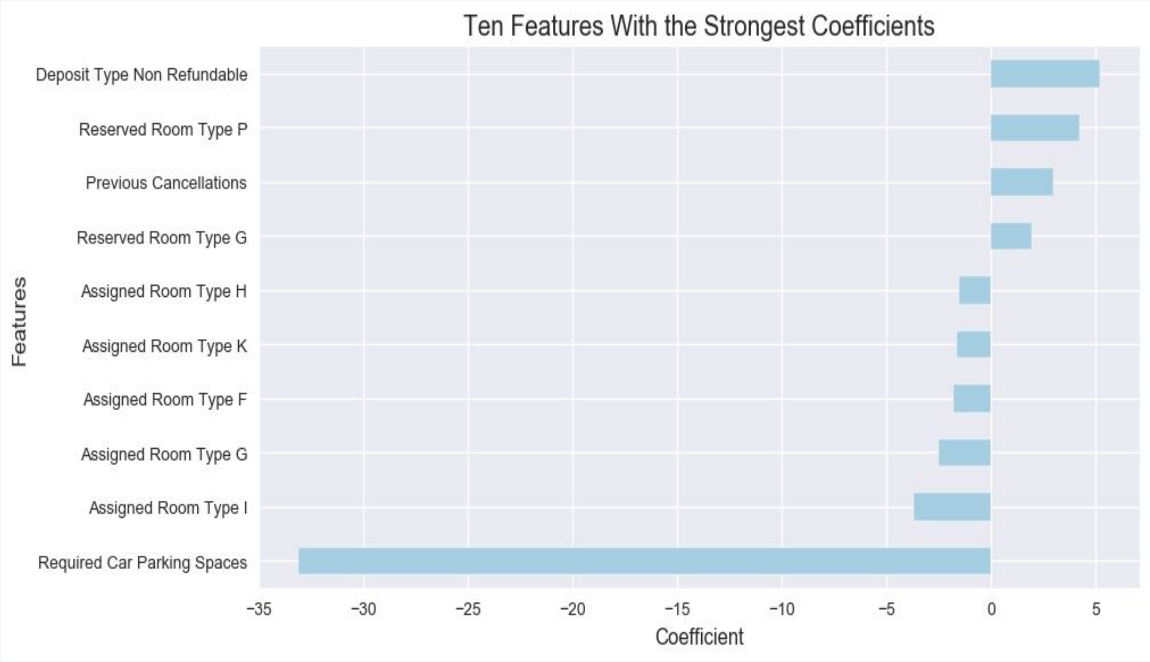
- **99%** predicted canceled actually canceled
- **96%** predicted not canceled actually not canceled

Management:

- **0.4%** of cases: hotel may not be ready for guest, risk of overbooking
- **2.4%** of cases: hotel allocating their resources on the wrong reservations

Logistic Regression Evaluation: Coefficient Interpretation

- For a 1 unit increase in the number of **parking spaces** required, a booking is **100% less likely** to be canceled.
- If the deposit is of type "**Non Refundable**", a booking is **177 times as likely** to be canceled.



Conclusion

- **Neural Network** able to predict booking cancellation with **97% accuracy**
- In **0.4% of cases** the hotel may **not be ready** for guest and runs the risk of **overbooking**
- In **2.4% of cases** the hotel is allocating their **resources** on the **wrong reservations**
- **4 customer clusters**: The cancellation squad, The family, The return customer, and The extended stay
- Lead time, deposit type, special requests, parking, and room type are **important features**

References

- https://triptease.com/blog/the-real-cost-of-free-cancellations/
- https://www.hotelmanagement.net/tech/study-cancelation-rate-at-40-as-otas-push-free-change-policy
- https://www.d-edge.com/how-online-hotel-distribution-is-changing-in-europe/
- https://www.stayntouch.com/blog/how-online-reviews-impact-hotel-revenue/
- https://vtechworks.lib.vt.edu/handle/10919/85353)
- https://www.sciencedirect.com/science/article/pii/S2352340918315191