***Hotel accommodation bookings and the effect of cancellations***

*Institute of Data – Capstone Project*

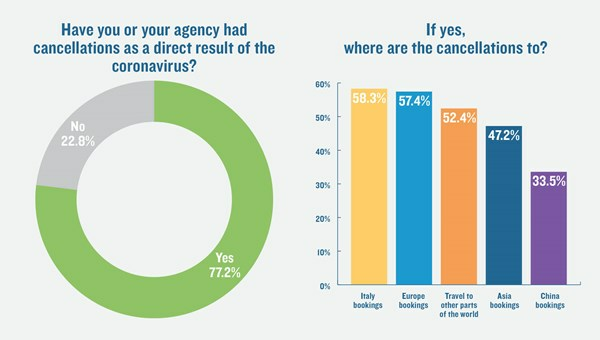
By Anthony Tan

Problem Statement



(Source: <https://www.ahla.com/covid-19s-impact-hotel-industry>)

The Covid-19 pandemic has greatly impacted many businesses and travel/tourism is amongst the most affected sectors. This current year is projected to be the lowest on record for hotel occupancy with greater than 50% revenue decline (approx. $112 billion lost).



(Source: <https://www.travelweekly.com/Travel-News/Travel-Agent-Issues/Travel-Weekly-readers-poll-Cancellations-piling-up-coronavirus>)

Cancellations are greatly affecting the tourism/travel industry with Europe being one of the most affected regions. Any potential cancellations of existing hotel bookings can have significant financial impact to the business. Although this does not include any data for 2020 as yet, the findings can still be applied for the future where it will be even more imperative to retain bookings.

Stakeholders

* Hotel owners
* Investors
* Franchisors
* Hotel managers

Business Question

* Business Question: How can we minimize hotel booking cancellations and retain business on the books?
* Data Question: What specific features has most impact on the target?
* Rationale: Knowing the likelihood of a guest cancellation can help with planning of pricing strategies, forecasting, and many other decisions that affect a hotel's bottom line in the interest of the key stakeholders.

Data

Data is extracted from reliable academic source of Instituto Universitário de Lisboa (ISCTE-IUL), Lisbon, Portugal. Comprises of 79,330 records and 32 features.

Source: <https://www.sciencedirect.com/science/article/pii/S2352340918315191>

EDA

Performed data analysis on the following:

* Distribution channel
* Monthly booking volume
* Average daily room rate
* Percentage of cancellations per month
* Correlation of lead time to cancellation rate

Machine Learning Models

* 19 out of 32 features were selected and inputted into the machine learning algorithm
* Tested five different classification models (Logistic Regression, Decision Tree, Random Forest, Boosting, AdaBoost and XG Boost)
* Performance evaluation metrics: Accuracy and Precision
* Random Forest had best performance with 85% precision in predicting true positive outcome

Findings

Calculated Feature Importance from Random Forest model to find most influential features:

* Lead Time
* Deposit Type
* Average Daily Rate
* Arrival Date
* Special Request

Recommendations

* Prepaid discount for booking a certain number number of days in advance. This incorporates lead time and deposit into solution.
* Stricter cancellation policies instead of "free" during high demand seasons to reduce the risk of cancellations on high rates.
* If booking meets specific criteria that fits the profile of a high probability cancellation, the hotel can make contact with guest provide personalized customer service in the hopes they will retain their booking.
* Pipeline can be re-produced and re-applied to any hotel, but may have to be tweaked depending on market mix of property.

References

Dataset and Jupyter Notebook with code can be found at the below repository:

<https://github.com/AnferneeTan/Data-Science>