MGT 6203 Group Project Proposal

TEAM INFORMATION (1 point)

Team #: 60

Team Members:

1. Andrea Kirksey; GT Id: akirksey6

I am currently working in Software Development. At work, I am working with AI and machine learning about cars and deploying different applications to them. I graduated with a B.S. in Information Technology.

2. Dakota Coomes; GT Id: dcoomes3

I am currently an EH&S Data Analyst for a global manufacturing corporation and am tasked with data analytics projects for external reporting requirements and dashboarding for internal stakeholders. I have an M.S. and B.S. in Biology.

3. Gurman Singh; GT Id: gsingh345

I am a Systems Architecture Engineer in the wireless communications industry working on NR 5G network, protocols, design, solutions, deployment, and performance. I graduated with B.S. in Electrical Engineering. I worked on analytics projects where I analyzed and predicted coverage and performance (KPIs).

4. Jiangqin Ma; GT Id: jma416

I have an engineering master's degree and a one year's certificate for accounting. I had even been a document engineer and insurance agent.

5. Duarte-Nuno De Sousa; GT Id: dsousa6

I have a B.E. in Engineering Mgmt. and 10+ years in IT project management in the Financial Services industry. My data analysis has been used in Financial Forecasting, Merger Synergies, and Product Development.

OBJECTIVE/PROBLEM (5 points)

Project Title: Modelling and Forecasting Hotel Booking Cancellations

Background Information on chosen project topic:

Increased hotel reservation options have changed booking behavior among potential customers. Due to the ease of bookings and cancellations associated with online reservation channels, hotels take on more booking risk, often with no financial recourse if bookings are cancelled. This risk can be mitigated substantially if it can be modelled and forecasted.

To address this consideration, we are analyzing two related datasets that include customer profiles, hotel profiles, and time-series data along with cancellation status for multiple bookings.

Problem Statement:

Determine what factors impact hotel booking cancellations and forecast future booking cancellations.

State your Primary Research Question (RQ):

Can hotel booking cancellations be modelled and forecasted using data on customer profiles, booking profiles, and time of year?

Supporting Research Questions:

1. Are hotel booking cancellations seasonal?
2. Are hotel booking cancellations affected by hotel profile?
3. Does family size/make-up impact booking cancellation propensity?
4. Do hotel booking cancellations show an increasing/decreasing trend?

Business Justification:
Hotel booking cancellations can leave hotels with unoccupied rooms that end up costing overhead for no return. If booking cancellations can be modelled, hotels can predict when cancellations are likely to occur and plan accordingly to maximize revenue/room utilization and operational staff to attend to these rooms/customers.

DATASET/PLAN FOR DATA (4 points)

Data Sources and descriptions:

https://www.kaggle.com/datasets/jessemostipak/hotel-booking-demand?resource=download&select=hotel_bookings.csv

https://www.dropbox.com/s/h6mfcq5ax2es2i8/hotel_bookings.csv.zip?dl=0

hotel	is_cancele	lead_time	arrival_d	arrival_dat	arrival_	arrival_	stays_i	stays_i	adults	childre	babies	meal	country
Resort Hotel	0	342	2015	July	27	1	0	0	2	0	0	BB	PRT
Resort Hotel	0	737	2015	July	27	1	0	0	2	0	0	BB	PRT
Resort Hotel	0	7	2015	July	27	1	0	1	1	0	0	BB	GBR
Resort Hotel	0	13	2015	July	27	1	0	1	1	0	0	BB	GBR
Resort Hotel	0	14	2015	July	27	1	0	2	2	0	0	BB	GBR

This dataset contains booking information from two separate datasets, one for a city hotel and one for a resort hotel. Some attributes include information on when the booking was made, length of stay, the number of adults, children, and/or babies.

https://www.kaggle.com/datasets/ahsan81/hotel-reservations-classification-dataset

https://www.dropbox.com/s/b4r6aahnn0ne77f/Hotel%20Reservations.csv?dl=0

Booking_ID	no_of_adı	no_of_chi	no_of_we	no_of_we	type_of_mea	required_	room_type_	lead_time	arrival_ye	arrival_mo	arrival_da	market_s
INN00001	2	0	1	2	Meal Plan 1	0	Room_Type	224	2017	10	2	Offline
INN00002	2	0	2	3	Not Selected	0	Room_Type	5	2018	11	6	Online
INN00003	1	0	2	1	Meal Plan 1	0	Room_Type	1	2018	2	28	Online
INN00004	2	0	0	2	Meal Plan 1	0	Room_Type	211	2018	5	20	Online
INN00005	2	0	1	1	Not Selected	0	Room_Type	48	2018	4	11	Online

This dataset contains various attributes of customers' reservation details. Some of the attributes in this dataset include room type, average price per room, booking status, previous booking cancellations, etc.

Key Variables:

Descriptive and predictive analytics - Logistic regression and classification

- Predictors: number of adults, number of children, lead time for reservation, arrival month, arrival date, booking price, location, deposit type
    - Important predictors: Arrival month, arrival date, lead time, booking price, location, deposit type
- Response: Booking status (cancelled or not)

Predictive analytics – time-series analysis/Holt-Winters forecasting

- Arrival date for cancelled hotel booking reservations
    - Total cancelled hotel booking reservations by week or month and hotel profile

APPROACH/METHODOLOGY (8 points)

Logistic regression to determine how parameters affect likelihood of hotel booking cancellation.

- Create a confusion matrix to summarize the classifications.
- Plot ROC Curve and AUC to analyze sensitivity and specificity with different thresholds.

KNN and SVM model for classification

- Tune C and K hyper-parameters.
- Split the dataset into training, validation, test sets. Use predict accuracy to compare the models.

Holt-Winters forecasting and exponential smoothing to determine periodicity and trend in booking cancellations throughout the year

Anticipated Conclusions/Hypothesis

More cancellations happen during the winter (December, January, February) due to inclement weather affecting travel itineraries.

City hotels have more options. Therefore, the booking cancellations would be higher due to customers having more options in the city. Whereas, when traveling to a specific destination, they have fewer hotel options.

The hotel will have an increasing booking cancellations trend due to the gradual increase in booking reservations overtime. Holt-Winters forecasting approach.

Large groups or families may have less cancellation than individuals due to logistical complexity associated with large-group planning.

What business decisions will be impacted by the results of your analysis? What could be some benefits?

After predicting the hotel cancellations, the hotel can minimize overhead costs, maximize revenue and room utilization. If the hotel has a decreasing booking cancellation trend, it may need to expand its business.

PROJECT TIMELINE/PLANNING (2 points)

Project Timeline/Mention key dates you hope to achieve certain milestones by:

- Data cleaning, exploratory analysis, visualization – March 19th
- Research articles on hotel reservations and cancellations – March 19th
- Finish up PowerPoint slides – March 22nd
- Do proposal video – March 24th
- Have analysis, and models done. Data is trained – March 24th
- Progress Report finished – March 29th
- Come up with final hypotheses, findings, and evaluate results – April 7th
- Final Report finished – April 10th
- Final Presentation Video, along with slides finished – April 17th