

ISSS602 Data Analytics Lab G3

AY2020-21T1

Assignment 1: Show Me the Numbers

The Cost of Booking Cancellations



Gabriella Pauline DJOJOSAPUTRO

Table of Contents

Overview	3
Data Preparation.....	3
Data Source	3
Data Quality Issues.....	3
1. Inconsistent way of storing date variables.....	3
2. Inconsistent way of documenting missing values	6
3. Non-missing values labelled as “NULL”	6
4. Erroneous data: more than four adults in one booking.....	7
5. Incorrect modelling type	8
Data Analysis.....	9
Insights	9
1. Cancellation proportions in H1 and H2 are different from each other.....	9
2. H1 have lower proportion of bookings from TA/TO compared to H2	11
3. Bookings from TA/TO have higher cancellation proportions.....	12
4. Non-refundable deposit has nearly 100% cancellation.	13
5. Cancelled reservations has higher mean of previous cancellations than those that are not cancelled.....	14
6. The mean lead time in cancelled reservations is higher than the ones that are not cancelled	15
7. Reservations that are cancelled has higher average of days in waiting list compared to those that are not cancelled and the average for H2 is higher than H1.	18
8. Transient customers have the highest proportion of cancellation H1, but contract customers are the highest in H2.	19
9. Repeated guests have lower proportion of cancellations	20
10. Bookings that are not made by or paid by companies have higher rate of cancellation	22
Interpretation of Analysis Results.....	23
Annex A – Data Preparation Change Log.....	24
References	29

Overview

Hotel booking cancellation is one of the challenges that affect tourism and hospitality industry. Once the reservation is cancelled, it would not be easy to find a replacement booking, especially if the cancellation is performed last-minute. This technical report aims to discover insights regarding the patterns in the cancellation so that hotel managements can use the new understanding to drive efforts in reducing the possibility of cancellation.

Data Preparation

Data Source

The data sets provided by Antonio, Almeida, and Nunes (2019) are used in this analysis. It contains data from two hotels in Portugal—one in the city and the other at a resort area.

Data Quality Issues

1. Inconsistent way of storing date variables

The data sets contain columns to record the arrival date and the reservation status date, but these two are stored differently. Arrival date is stored in four columns—each for the year, month, week, and day of month. Meanwhile, the reservation status date is stored in a single column as character. Without solving the inconsistency, it will be hard to compare the dates. The steps taken to resolve the issue are:

a) Change ReservationStatusDate data type

- Right-click on ReservationStatusDate column > **Column Info**
- Data Type > Numeric
- Modeling Type > Continuous
- Format > Date > yyyy-mm-dd
- Click **OK**

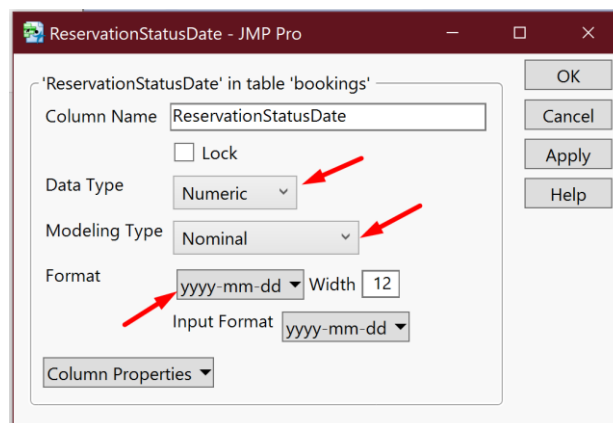


Figure 1. Converting ReservationStatusDate into proper format

b) Recode ArrivalDateMonth

- Right-click on ArrivalDateMonth column > **Recode**
- Enter corresponding number in **New Values** for each month specified in **Old Values**
- Enter the new column name for recoded values in **Name**
- Click **Recode**

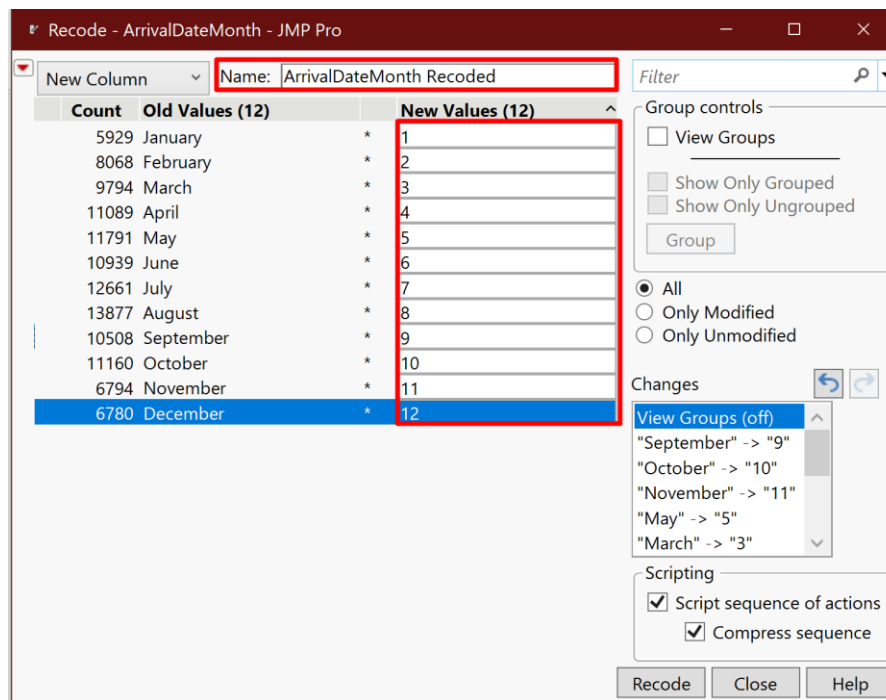


Figure 2. Recoding ArrivalDateMonth

c) Create a new formula column

- Right-click on any column > **Insert columns**
- Right-click on the new column > **Column info**
- Enter "ArrivalDate" as the column name
- **Column Properties** > **Edit Formula** > Insert the formula shown in Figure 3
- Change to date format as in step 1A

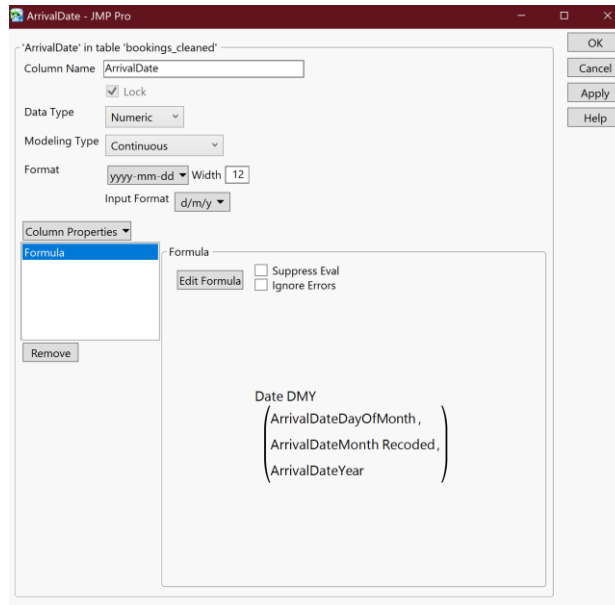


Figure 3. Creating new formula column to combine arrival date

- d) Hide and exclude original columns for arrival date
- In columns panel, right-click on ArrivalDateYear
 - Press shift and click on ArrivalDateDayOfMonth
 - Right-click > **Hide/Unhide**
 - Right-click > **Exclude/Unexclude**

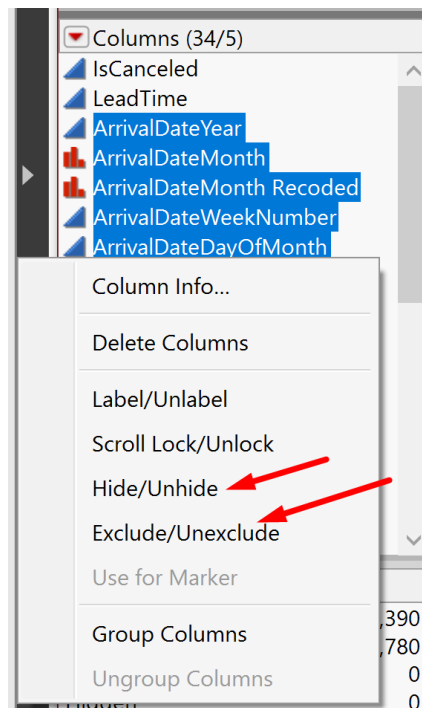


Figure 4. Hide and exclude columns

2. Inconsistent way of documenting missing values

There are missing values in various different columns of the data, but they are not stored consistently. They are recorded as “NA” in Children, “Undefined” in Meal, MarketSegment, and DistributionChannel, and “NULL” in Country (Figure 5 and 6).

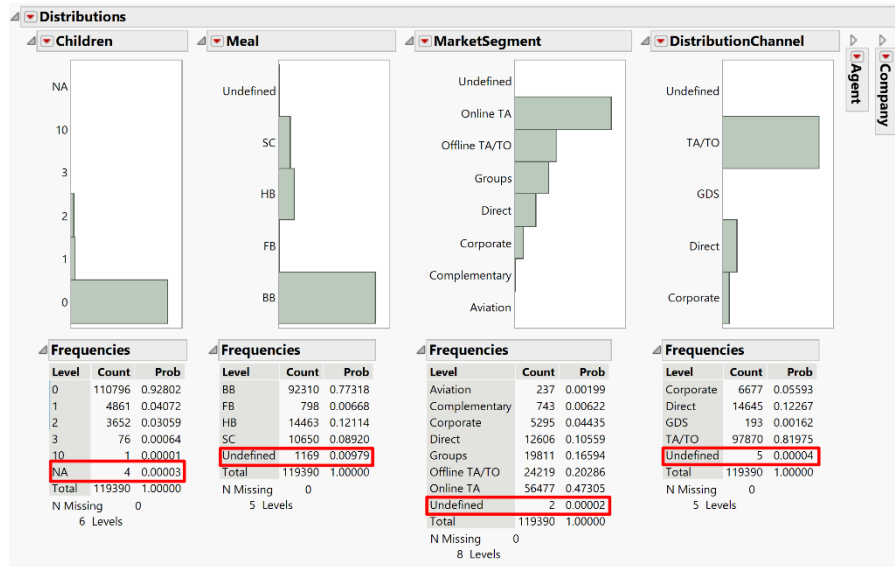


Figure 5. Missing values in Children, Meal, MarketSegment, and DistributionChannel

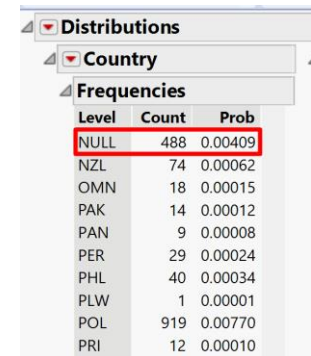


Figure 6. Missing values in Country

To prevent the missing values to be treated as another category during the analysis, all these columns are recoded to an empty value.

- Recode the values into missing value (see steps in 1B)
- Hide and exclude the original columns (see steps in 1D)

3. Non-missing values labelled as “NULL”

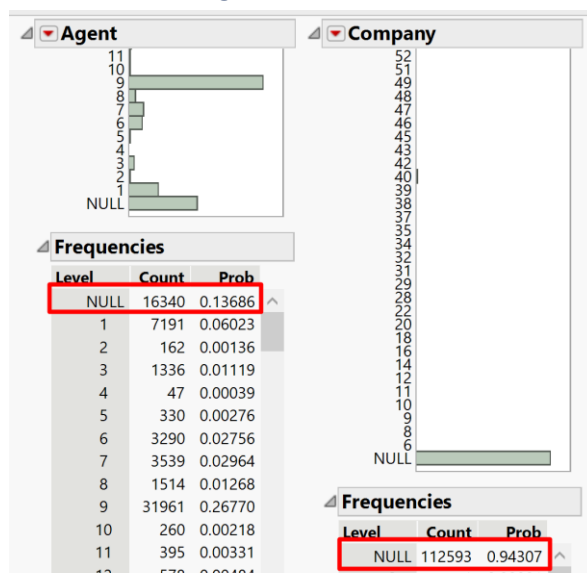


Figure 7. "NULL" in Agent and Company

Agent and Company columns have considerable amount of data with the value “NULL” (Figure 7), but it was explicitly mentioned by Antonio et al. (2019) that this means the booking did not come from an agent or a company. Using the category “NULL” suggests that these values are missing, instead of being not applicable. Therefore, to prevent misunderstanding, these values are recoded as “No Agent” and “No Company”.

- Recode the values (see steps in 1B)
- Hide and exclude the original columns (see steps in 1D)

4. Erroneous data: more than four adults in one booking

			Cancellation	
Hotel			Cancelled	Not Cancelled
H1	Adults	Min	0	0
		Max	55	4
H2	Adults	Min	0	0
		Max	4	4

Figure 8. Tabulation of Adults by Cancellation and Hotel

The number of adults seems to be suspicious with some values being unusually high while most of them are 4 and below. Tabulating the data (Figure 8) further reveals that the number of adults being higher than 4 only occurs on cancelled bookings in H1.

To examine these records in a separate table:

- Press right-click on Adults column > **Sort > Descending**
- Select the records that have more than 4 adults, select **Tables > Subset**

	Cancellation	LeadTime	ArrivalDate	TotalStay	BookingDate	Adults	Country	MarketSegment	IsRepeatedGuest	Agent	ReservationStatus	ReservationStatusDate	LeadTimeCancelled	Hotel
1	Cancelled	338	2015-10-04	2	31/10/2014	55	PRT	Direct	New Guest		Canceled	2015-01-02	275	H1
2	Cancelled	336	2015-09-07	3	06/10/2014	50	PRT	Direct	New Guest		Canceled	2015-01-18	232	H1
3	Cancelled	304	2015-09-03	3	03/11/2014	40	PRT	Direct	New Guest		Canceled	2015-01-02	244	H1
4	Cancelled	352	2015-09-24	4	07/10/2014	27	PRT	Direct	New Guest		Canceled	2015-01-02	265	H1
5	Cancelled	349	2015-09-21	4	07/10/2014	27	PRT	Direct	New Guest		Canceled	2015-01-02	262	H1
6	Cancelled	361	2015-10-03	7	07/10/2014	26	PRT	Offline TA/TO	New Guest	96	Canceled	2015-01-02	274	H1
7	Cancelled	354	2015-09-26	7	07/10/2014	26	PRT	Offline TA/TO	New Guest	96	Canceled	2015-01-02	267	H1
8	Cancelled	347	2015-09-19	7	07/10/2014	26	PRT	Offline TA/TO	New Guest	96	Canceled	2015-01-02	260	H1
9	Cancelled	340	2015-09-12	7	07/10/2014	26	PRT	Offline TA/TO	New Guest	96	Canceled	2015-01-02	253	H1
10	Cancelled	333	2015-09-05	7	07/10/2014	26	PRT	Offline TA/TO	New Guest	96	Canceled	2015-01-02	246	H1
11	Cancelled	328	2015-10-06	7	12/11/2014	20	PRT	Direct	New Guest		Canceled	2015-09-03	33	H1
12	Cancelled	334	2015-10-12	7	12/11/2014	20	PRT	Direct	New Guest		Canceled	2015-01-02	283	H1
13	Cancelled	334	2015-10-12	7	12/11/2014	10	PRT	Direct	New Guest		Canceled	2015-01-02	283	H1
14	Cancelled	328	2015-10-06	7	12/11/2014	6	PRT	Direct	New Guest		Canceled	2015-01-02	277	H1
15	Cancelled	334	2015-10-12	7	12/11/2014	5	PRT	Direct	New Guest		Canceled	2015-01-02	283	H1
16	Cancelled	328	2015-10-06	7	12/11/2014	5	PRT	Direct	New Guest		Canceled	2015-01-02	277	H1

Figure 9. Subset of records with more than 4 adults

An interesting observation from these records are that most of these records have similar booking and cancellation date. Therefore, there is a possibility that these records are cancelled because the hotel or travel agent (TA) realized there was an error with these bookings. For this reason, the 16 records with more than 4 adults are hidden and excluded from the analysis.

- In the main data table, select the 16 records with more than 4 adults
- Right-click on the column > **Hide and Exclude**

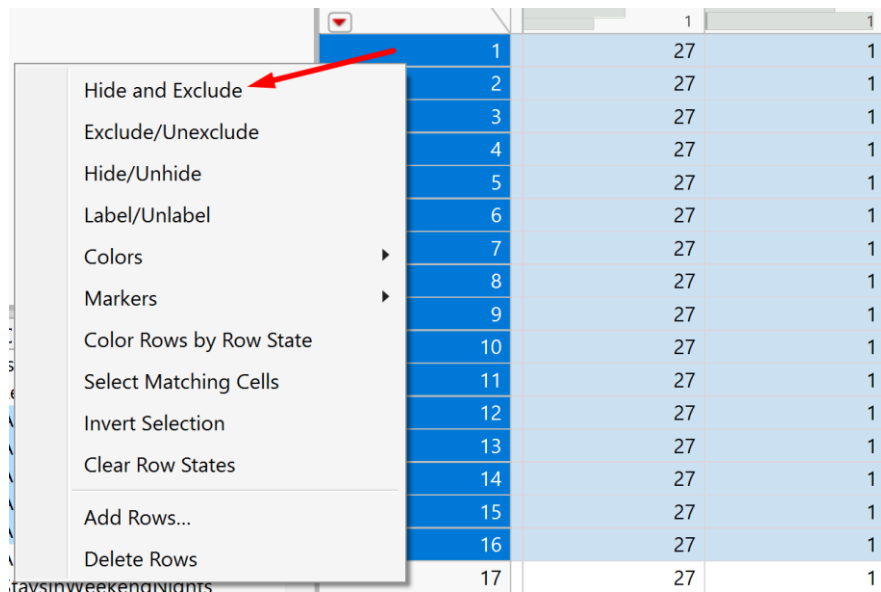


Figure 10. Hide and Exclude rows

5. Incorrect modelling type

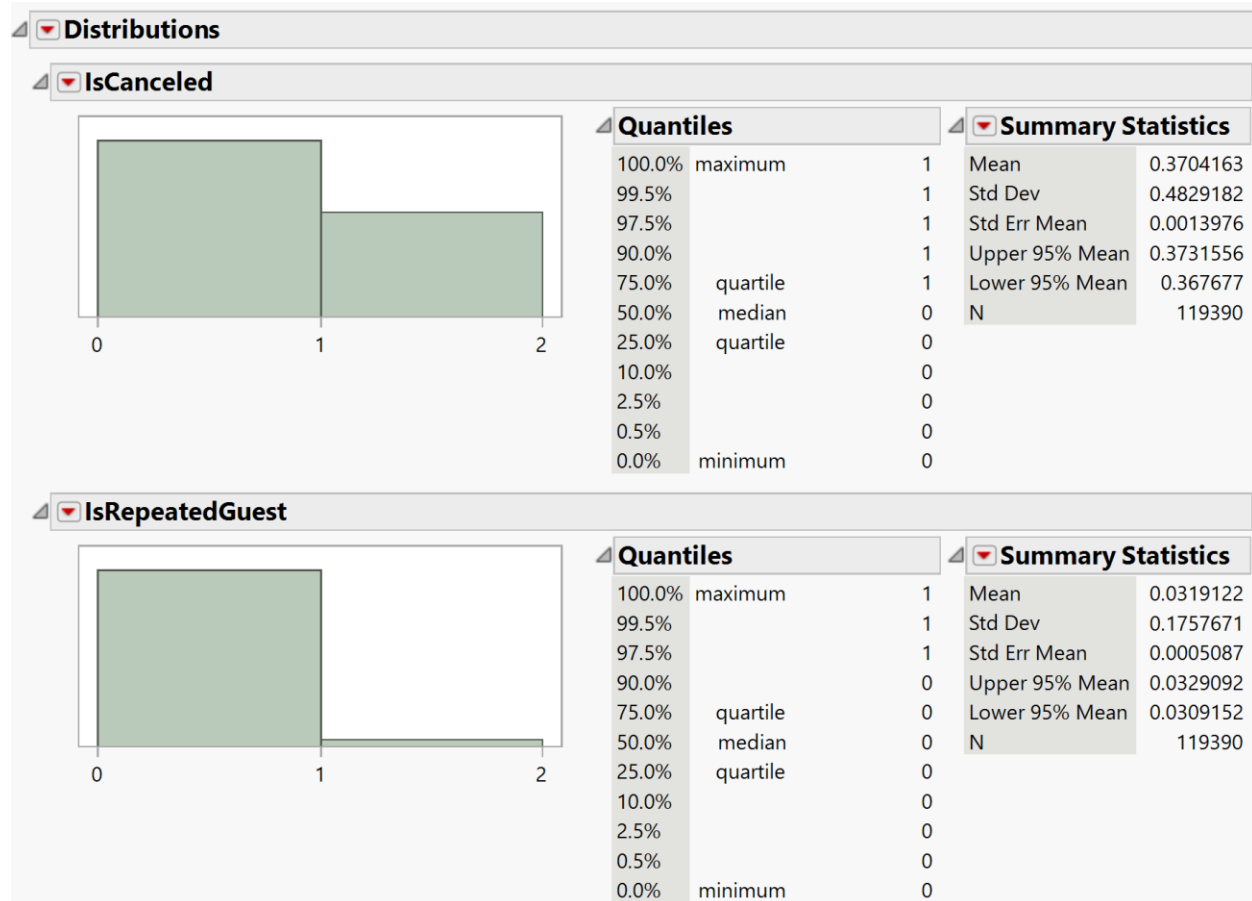


Figure 11. Distribution of IsCanceled and IsRepeatedGuest

IsCanceled and IsRepeatedGuest columns are using 1 and 0 to represent true or false. Therefore, they are actually nominal, instead of continuous as it is shown in the default. Not specifying the proper modeling type can make the analysis produce meaningless results, such as a standard deviation (Figure 11). To foster easier understanding during analysis, the values are changed to nominal with a more self-explanatory value. The values are recoded as “Cancelled” and “Not Cancelled” for IsCanceled and “Repeated Guest” and “New Guest” for IsRepeatedGuest.

- Recode the values (see steps in 1B)
- Hide and exclude original columns (see steps in 1D)
- Click **OK** when there is a warning that character is being entered to a numeric variable. JMP will automatically select the appropriate data and modelling type

Data Analysis

Insights

1. Cancellation proportions in H1 and H2 are different from each other

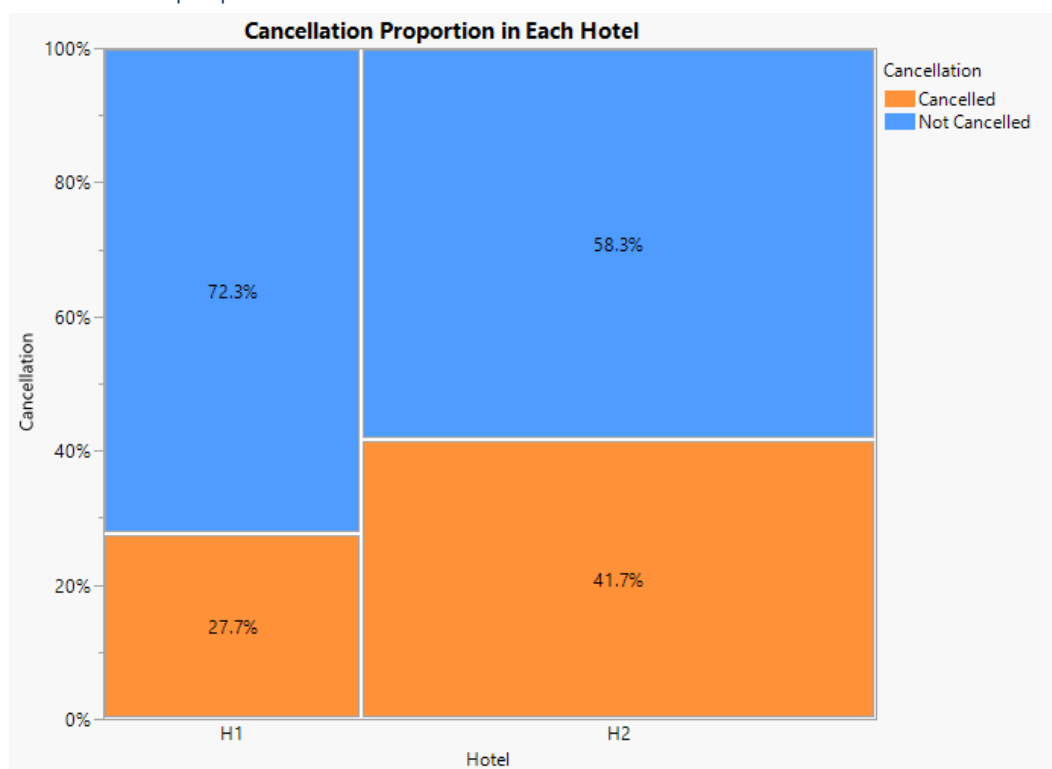


Figure 12. Cancellation proportions in H1 and H2

As the data sets came from two different hotels, it is necessary to investigate whether the booking cancellation proportions are similar enough, so we can determine whether the analysis can be performed together or must be separated for each hotel. The proportion shown in Figure

12 suggests that the cancellation proportion in H2, around 40%, is higher than H1 which value is slightly less than 25%.

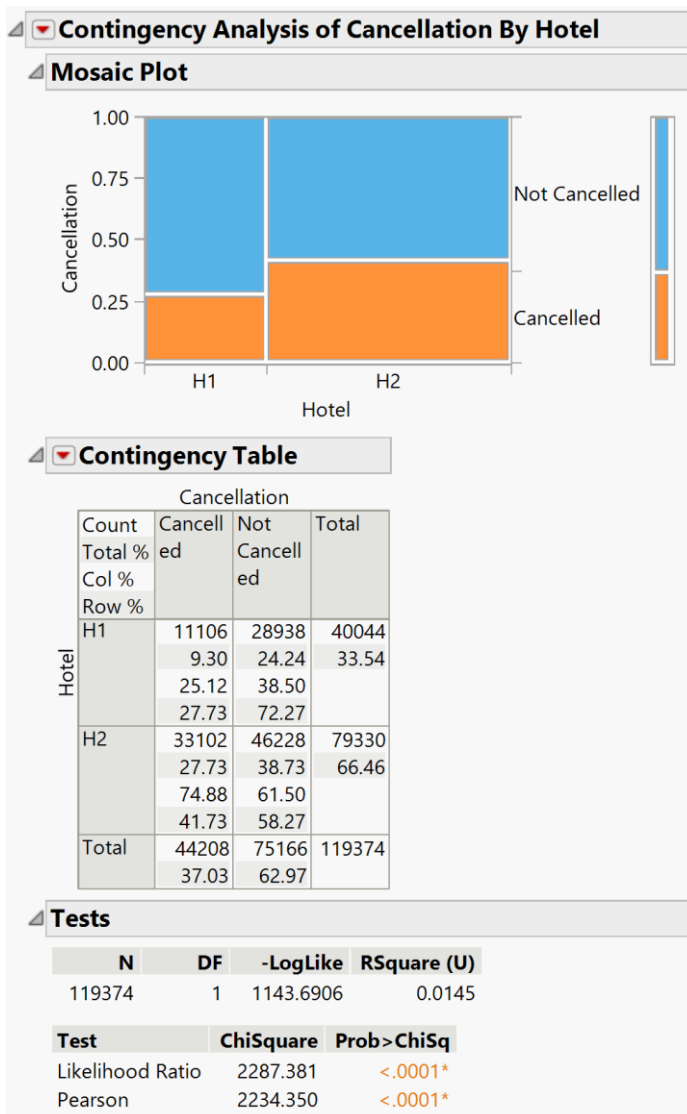
Hypothesis Testing

To confirm this observation, hypothesis testing was conducted with the following hypotheses and confidence level:

H_0 : There is no difference between the proportion of hotel booking cancellations in H1 and H2.

H_1 : There is a difference between the proportion of hotel booking cancellations in H1 and H2.

Confidence level: 95%



By observing the contingency table (Figure 13), we know that there are more than 5 records for each of the categories. Therefore, the assumptions for using Chi-Square test is fulfilled.

Figure 13 also shows the p-value of the Chi-Square test, which is less than the critical value of 0.05. Hence, we reject the null hypothesis that “There is no difference between the proportion of hotel booking cancellations in H1 and H2”. As such, the two data sets should be analyzed separately to discover what makes them different from one another, and identify consistent patterns in both hotels that affect cancellations of hotel bookings.

Figure 13. Contingency Table and Chi-Square test results for cancellation rate by hotel

2. H1 have lower proportion of bookings from TA/TO compared to H2

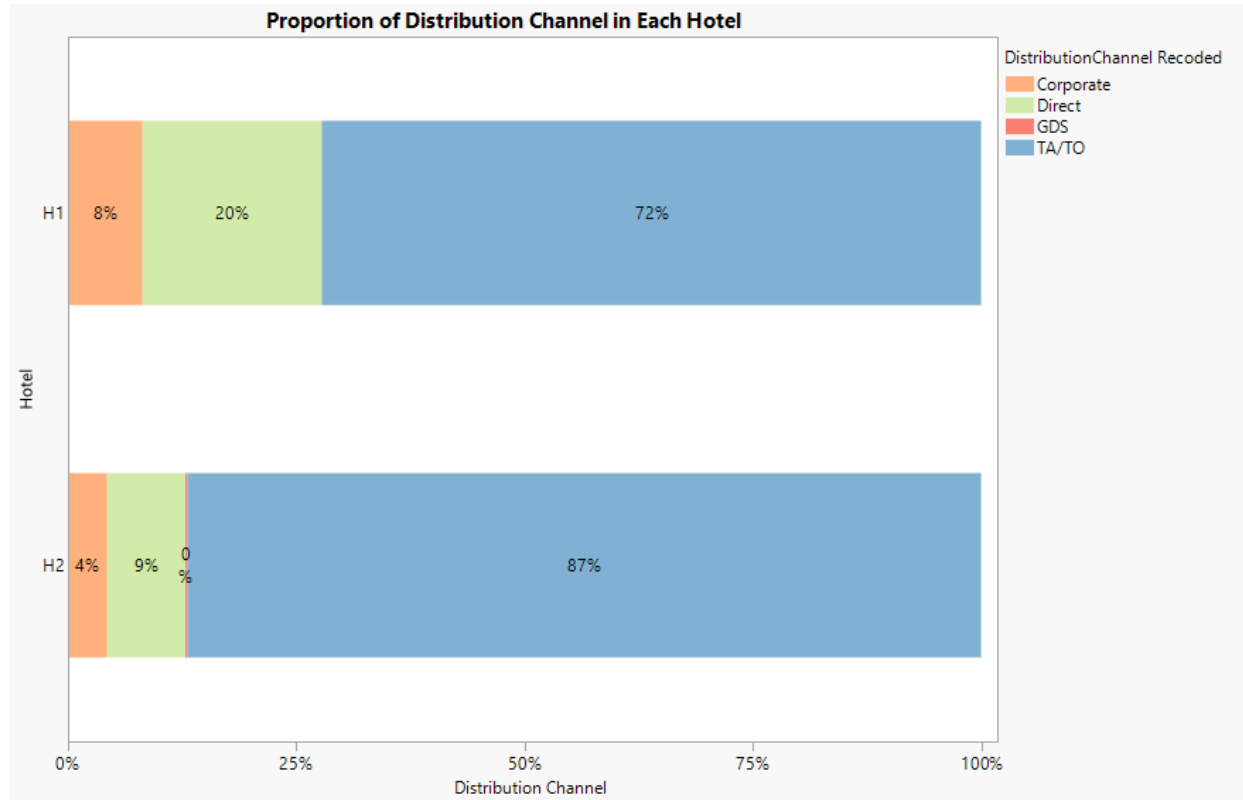


Figure 14. Proportion of Distribution Channel in Each Hotel

As seen in Figure 14, H1 only has 72% of bookings from Travel Agents and Tour Operators (TA/TO), while H2 has 87%. The proportion of bookings from corporate and direct channel are also twice higher in H1 than H2—accounting for 8% and 20% in H1, but only 4% and 9% in H2.

DistributionChannel Recoded	Hotel			
	H1		H2	
	N	Column %	N	Column %
Corporate	3269	8.16%	3408	4.30%
Direct	7854	19.61%	6780	8.55%
GDS	0	0.00%	193	0.24%
TA/TO	28920	72.22%	68945	86.91%

Figure 15. Tabulation of Distribution Channel by Hotel

there are only 3 sections shown for H1 in Figure 14, but H2 have a thin line marked as 0% due to rounding.

Figure 15 further illustrate the proportion of bookings from each distribution channel in each hotel. H1 does not have any booking from Global Distribution System (GDS), but H2 have 0.24% bookings from GDS. So,

3. Bookings from TA/TO have higher cancellation proportions

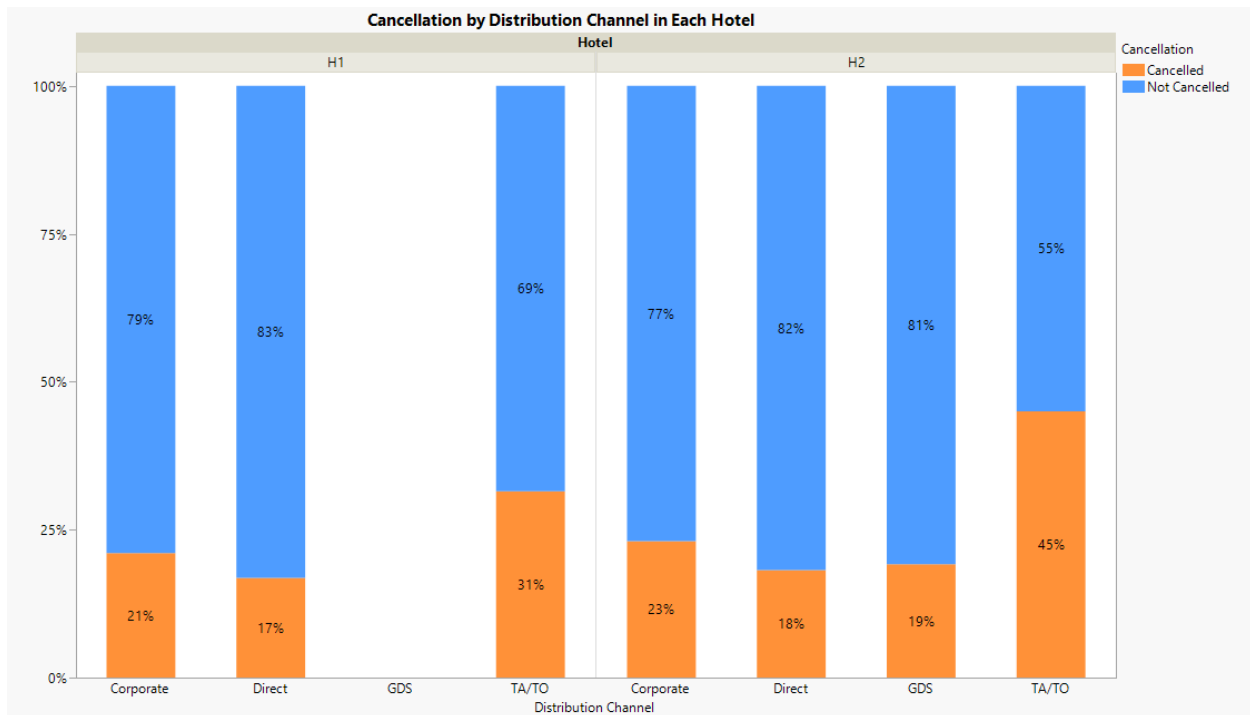


Figure 16. Cancellation by Distribution Channel in Each Hotel

As what Funnell (2019) and Orange Hotel Marketing (2016) argue, in both hotels, bookings from TA/TO have higher cancellation proportions than the other channels. The difference is more pronounced in H2, in which the cancellation proportion in TA/TO bookings (45%) is two-folds greater than Corporate, Direct, and GDS channels with 23%, 18%, and 19% cancellation each. H1 has 31% cancellation for TA/TO, although the cancellation for Corporate and Direct are comparable to H2 on 21% and 17%. This can be attributed to the lower overall cancellation rate in H1 than H2 as mentioned in the first insight.

4. Non-refundable deposit has nearly 100% cancellation.

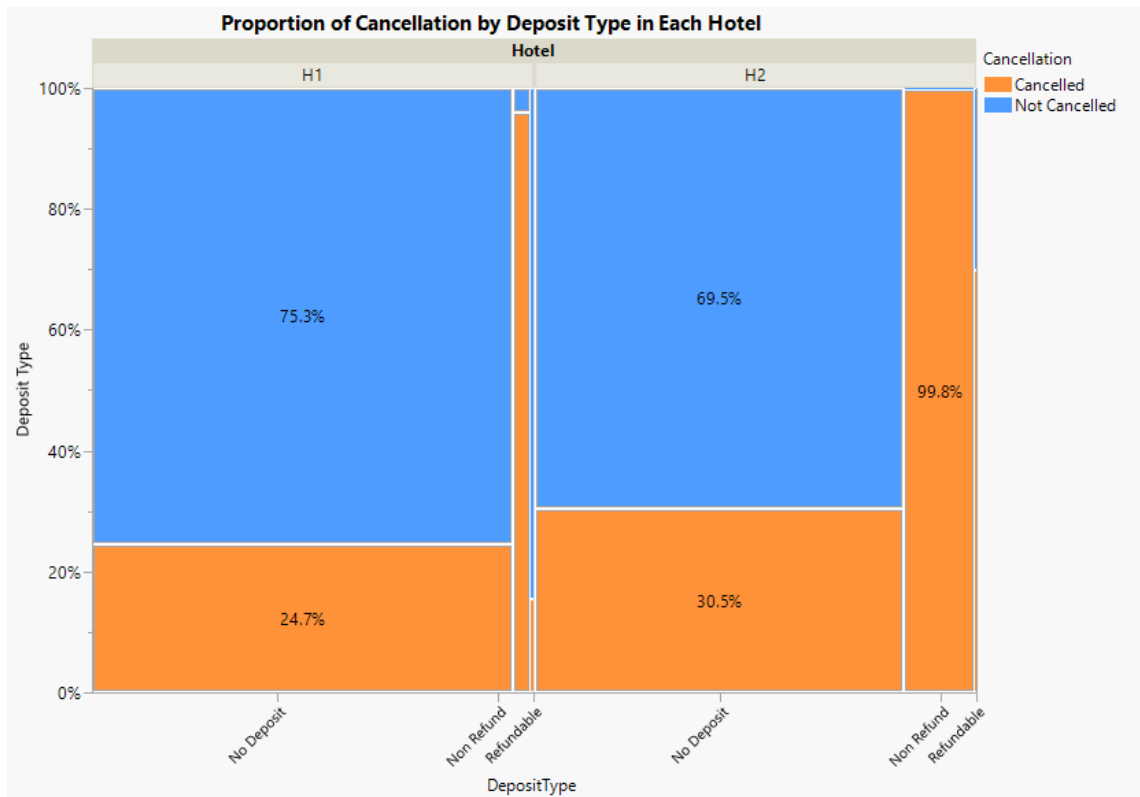


Figure 17. Proportion of Cancellation by Deposit Type in Each Hotel

Figure 17 shows there are higher number of bookings made with no deposits for both hotels, consistent with Funnell's (2019) observation that bookings with 'free' cancellations is getting more prevalent. However, it is interesting to note that the proportion of cancellation is nearly 100% for non-refundable bookings.

This insight is inconsistent with Funnell's (2019) argument that the increasing rate of cancellation in reservations are due to the "free cancellation" schemes offered by online travel agents (OTAs). For bookings with no deposit, the proportion of cancellation is 24.71% and 30.47% in H1 and H2 respectively. Yet, for non-refundable deposit, the cancellation proportion shoots up to 95.99% and 99.81%.

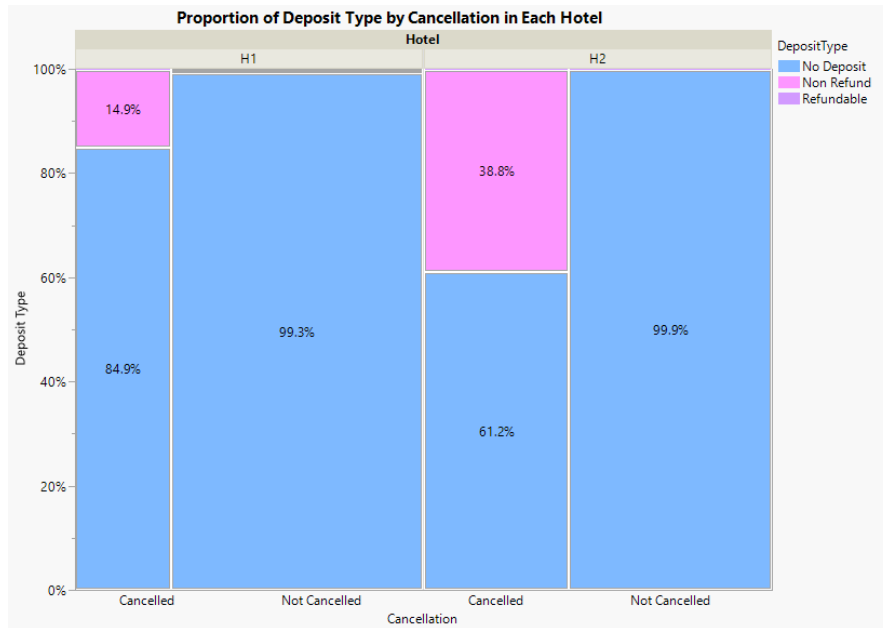


Figure 18. Proportion of Deposit Type by Cancellation in Each Hotel

Swapping the cancellation and deposit type gives another perspective (Figure 18). In both hotels, non-refundable deposit almost entirely exists in the cancelled bookings. It accounts 14.9% of cancelled bookings in H1 and 38.8% in H2.

- Cancelled reservations has higher mean of previous cancellations than those that are not cancelled.

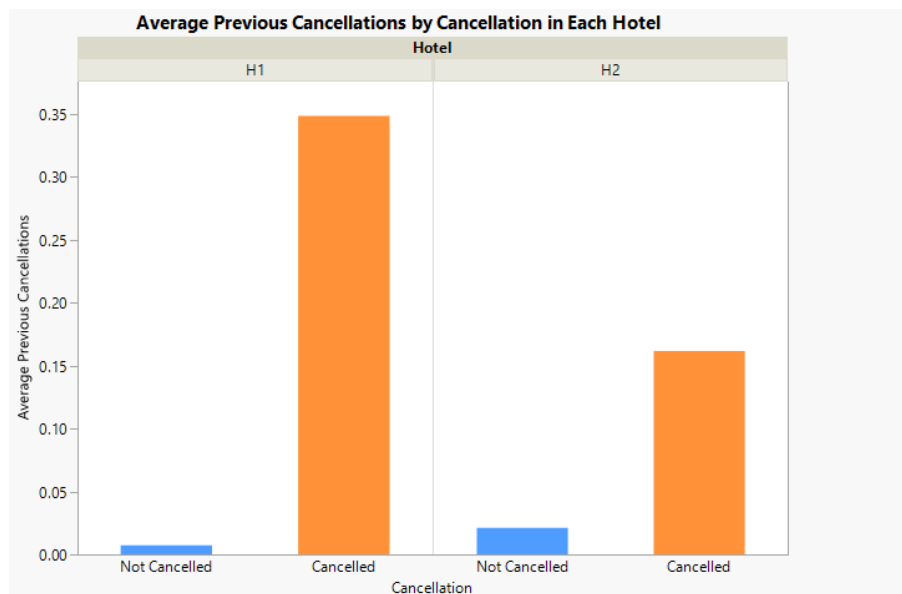


Figure 19. Average Previous Cancellations by Cancellation in Each Hotel

The average number of previous cancellations is higher for cancelled reservations in both hotels (Figure 19). Average previous cancellations for non-cancelled and cancelled bookings in H1 are 0.007 and 0.348 respectively, while for H2 they are 0.021 and 0.162.

6. The mean lead time in cancelled reservations is higher than the ones that are not cancelled

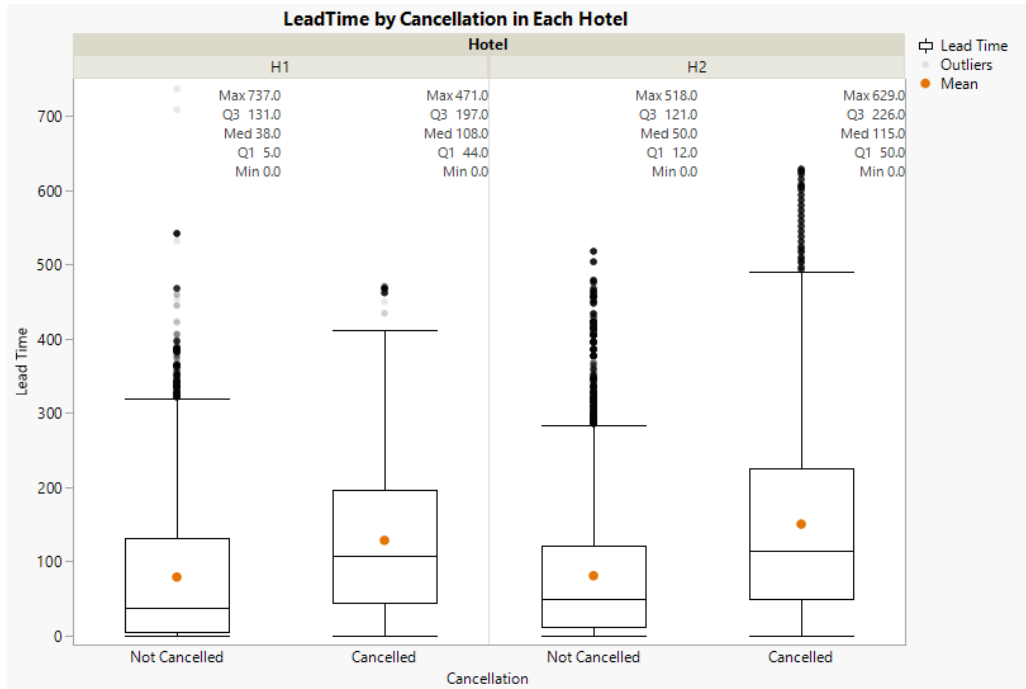


Figure 20. LeadTime by Cancellation in Each Hotel

The mean and median of lead time are higher in cancelled reservations than in those that are not cancelled (Figure 20). Cancelled and non-cancelled reservations have the mean lead time of 128.68 days and 78.84 days in H1, and 150.28 days and 80.70 days in H2.

Hypothesis Testing

Hypothesis testing is performed to verify this observation.

H_0 : There is no difference between the means of lead time for cancelled reservations and those that are not cancelled.

H_1 : There is a difference between the means of lead time for cancelled reservations and those that are not cancelled.

Confidence level: 95%

To select the proper statistical methods, assumptions are verified. Normal distribution is tested using Anderson-Darling with the null hypothesis that the distribution revealed resembles normal distribution. Equal variance is also tested with the null hypothesis that the distributions have equal variance.

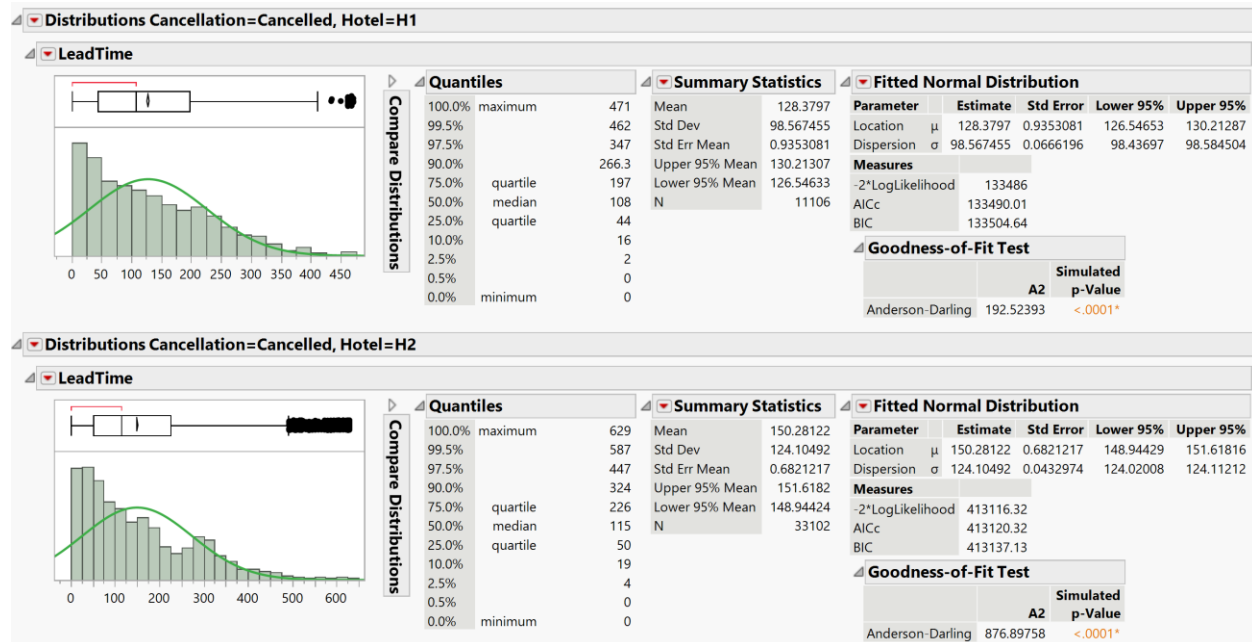


Figure 21. Distribution of Lead Time (1)

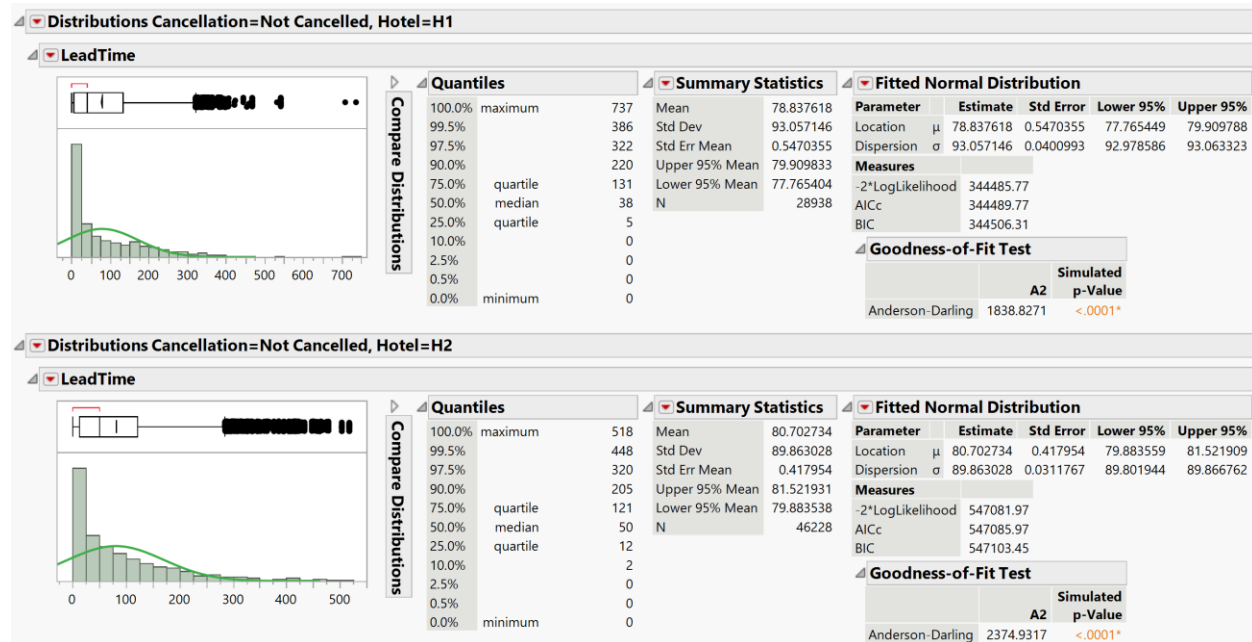


Figure 22. Distribution of Lead Time (2)

For all four distributions, the p-value are less than the critical value of 0.05 (Figure 21 and 22), so there is 95% confidence that there is not enough evidence to support that the distribution patterns shown resemble normal distribution.

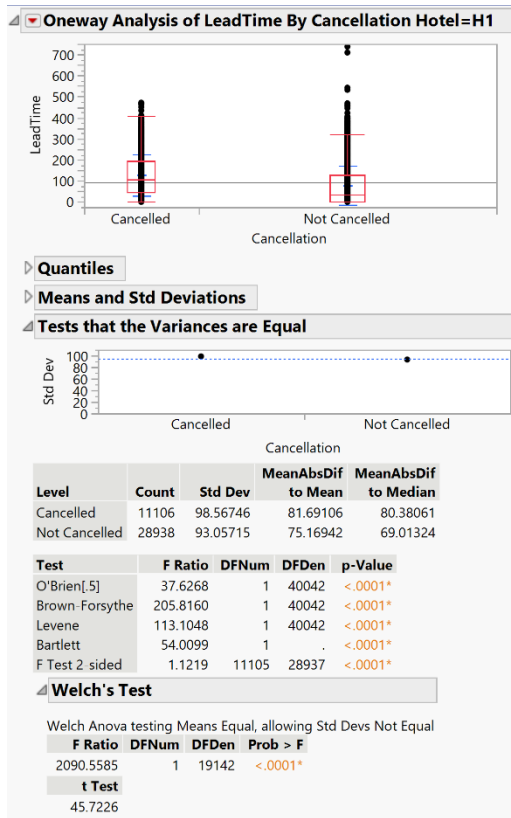


Figure 23. Equal variance and Welch Test results for lead time by cancellation in H1

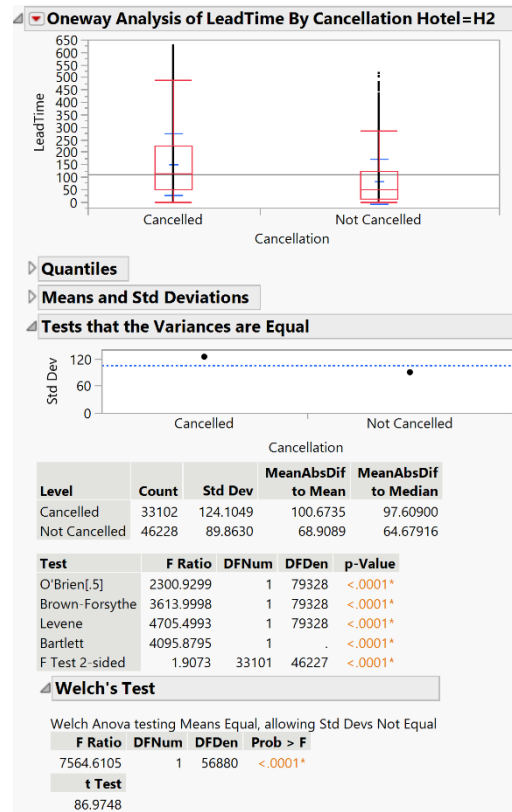


Figure 24. Equal variance and Welch Test results for lead time by cancellation in H2

Tests for equal variance between cancelled reservations and non-cancelled reservations in each hotel also obtain p-values of less than 0.05 (Figure 23 and 24), so we reject the null hypothesis that there is equal variance between the groups. Therefore, Welch test was used to test whether there is a difference in the means of lead time between the groups.

The p-value of the Welch tests are also less than 0.05 (Figure 23 and 24), which means in both hotels, there is not enough evidence to show that "There is no difference between the means of lead time for cancelled reservations and those that are not cancelled."

7. Reservations that are cancelled has higher average of days in waiting list compared to those that are not cancelled and the average for H2 is higher than H1.

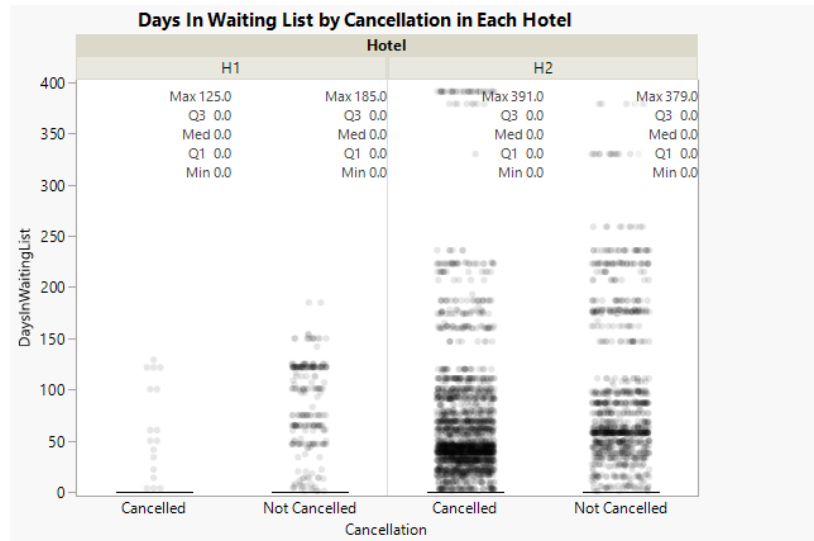


Figure 25. DaysInWaitingList Box Plot

Figure 25 shows that more than 75% of the reservations has 0 days in waiting list, but there seem to be more outliers for H2 and they seem to have a higher number.

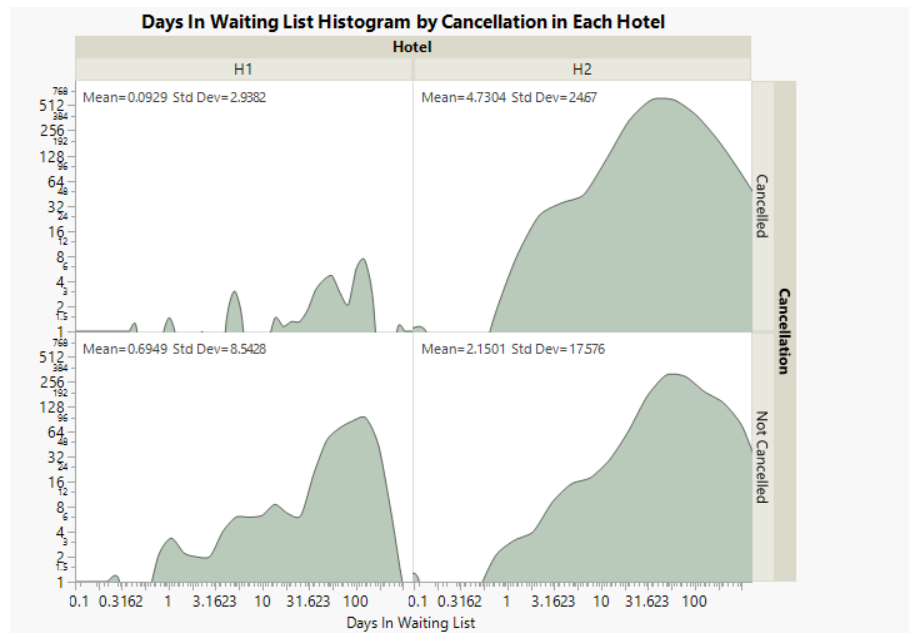


Figure 26. DaysInWaitingList Histogram

Therefore, kernel density histogram for days in waiting list by cancellation in each hotel was constructed with log 10 transformation for y-axis and log 2 for x-axis (Figure 26). For both cancelled and non-cancelled bookings, H2 has higher means for the days in waiting list than H1. However, H1 have a higher average of days in waiting list for cancelled bookings (0.09) than the non-cancelled ones (0.65).

8. Transient customers have the highest proportion of cancellation H1, but contract customers are the highest in H2.

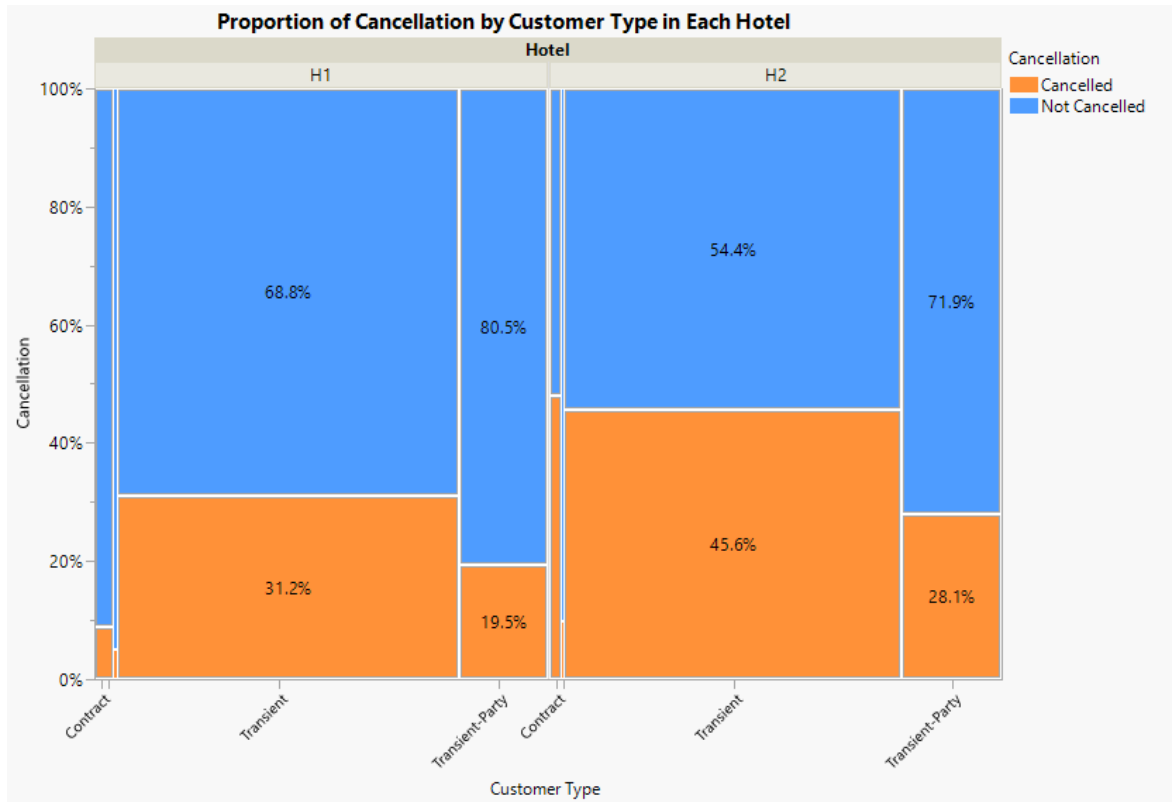


Figure 27. Proportion of Cancellation by Customer Type in Each Hotel

Different customer types seem to have different proportion of cancellation, but it does not seem to be consistent between H1 and H2. Transient customers have the highest cancellation rate in H1, but the highest for H2 is contract customers.

9. Repeated guests have lower proportion of cancellations

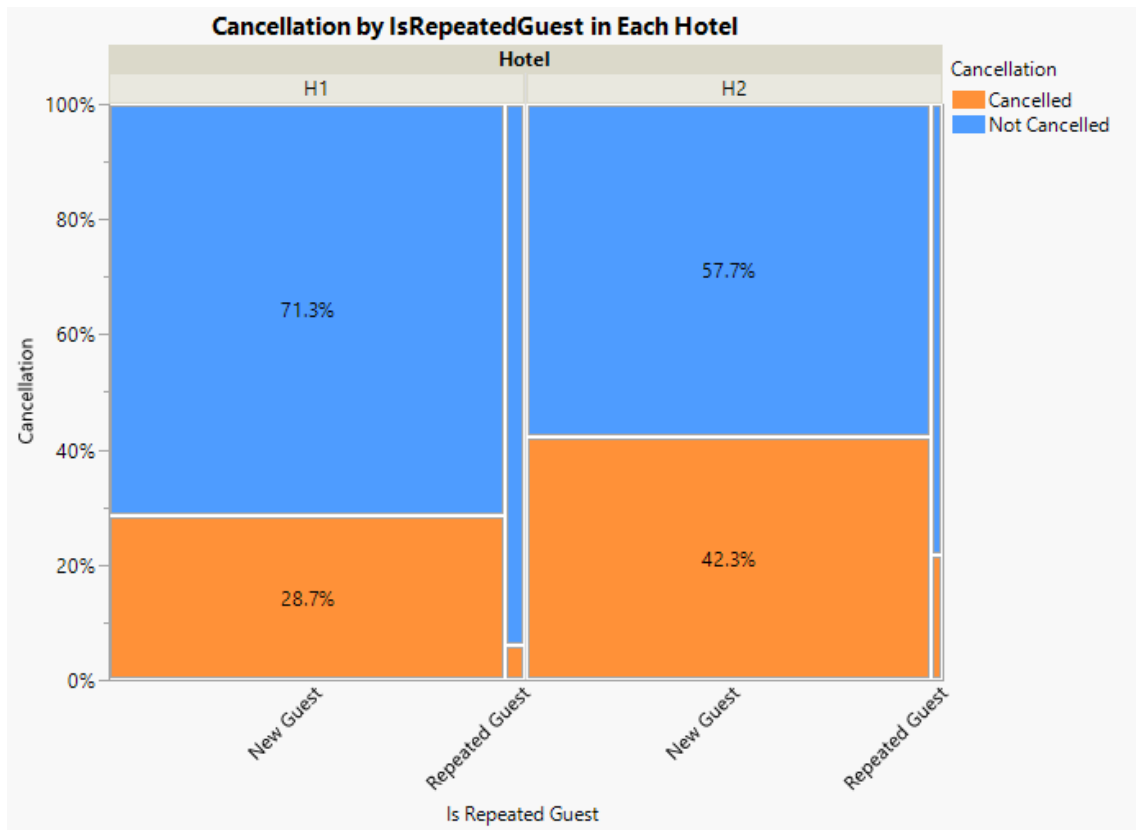


Figure 28. Cancellation by IsRepeatedGuest in Each Hotel

Figure 28 highlights that new guests have higher proportion of cancellation in both H1 (28.73%) and H2 (42.25%) than repeated guests (6.24% and 21.70%), but they also make up the most of the bookings.

Hypothesis Testing

The hypothesis and confidence level for this test are:

H_0 : There is no difference in the proportion of cancellation between new guest and repeated guest.

H_1 : There is a difference in the proportion of cancellation between new guest and repeated guest.

Confidence level: 95%

Assumption that there is no categories with less than 5 observations is verified (Figure 29 and 30), so Chi-Square test is used.

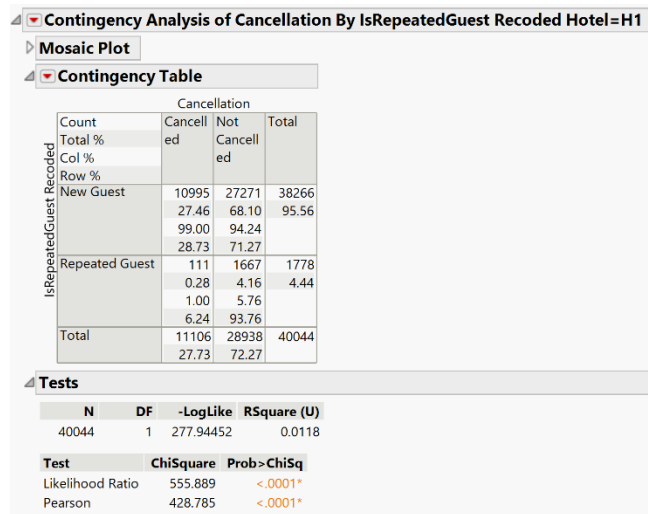


Figure 30. Contingency table and Chi-Square test result for H1

For both H1 and H2, the p-value of Chi-Square test is lower than the critical value of 0.05, so we reject the null hypothesis that “There is no difference in the proportion of cancellation between new guest and repeated guest.”

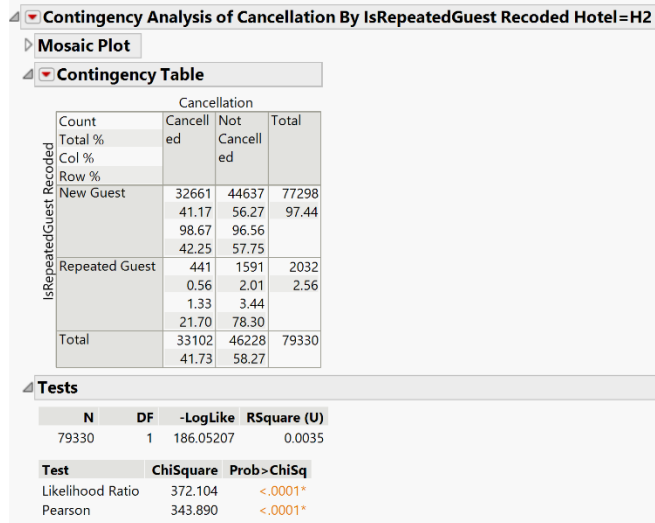


Figure 29. Contingency table and Chi-Square test result for H2

10. Bookings that are not made by or paid by companies have higher rate of cancellation

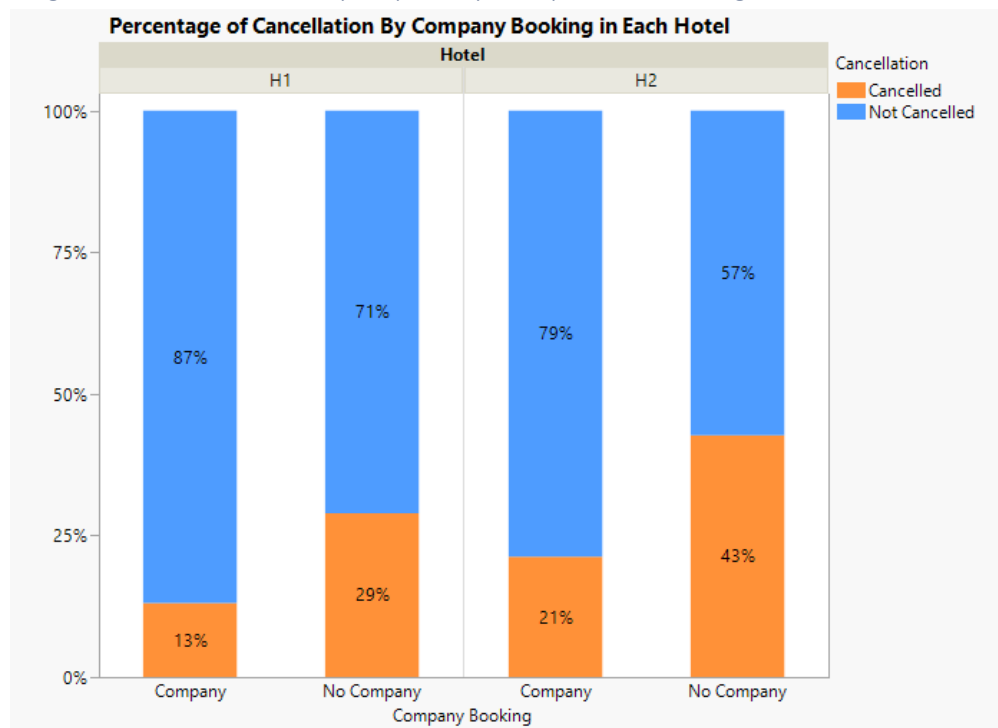


Figure 31. Percentage of Cancellation by Company Booking in Each Hotel

For both H1 and H2, bookings that do not have a company ID associated with it have higher cancellation rate (Figure 31). The company ID indicates the company that made or paid the booking. For H1, 13% of company bookings are cancelled and 29% of non-company bookings are cancelled. The number is higher for H2 at 21% and 43%.

Interpretation of Analysis Results

The analysis shows that the cancellation rate in H1 and H2 are different from one another*, in which H2 has higher proportion of cancelled bookings (41.7%) than H1 (27.7%). Another difference observed between H1 and H2 is their distribution channel. H2 have larger stream of bookings from TA/TO (87%) than H1 (72%).

Some characteristics of guest booking consistently have higher cancellation proportion in both H1 and H2, despite being a different type of hotel in a different location in Portugal. These characteristics are:

1. Being made from TA/TO distribution channel rather than corporate, direct, or GDS channels.
2. Using non-refundable deposit.#
3. Having higher number of previous booking cancellations.
4. Making the bookings further away from the arrival date, i.e. having a higher lead time.*
5. Being a new guest.*
6. Not being made by or paid by a company.

Some findings are inconsistent between H1 and H2:

1. Cancelled bookings have higher average of days in waiting list than non-cancelled bookings in H2, but it is lower in H1.
2. Transient customers have the highest proportion of cancelled bookings for H1, while it is contract customers for H2.

Insights marked with an asterisk (*) have been confirmed using statistical hypothesis testing, while insight marked with a pound sign (#) needs further investigation. Ninety-nine percent of non-refundable bookings are cancelled in both hotels, which is counterintuitive. There is possibility that logic behind categorization of deposit type has errors, which is currently done by determining whether there are any payments made before the arrival date. For example, if only the cancellation fee will be directly paid to the hotel while payments for booking via TA/TO is only paid in lump sum periodically, TA/TO bookings will be only categorized as non-refundable when they are cancelled.

Annex A – Data Preparation Change Log

Note: for every recoded items, the results are stored in a new column with “Recoded” appended to the name of the original column. When there is additional action to hide and exclude the column, it means the actions are performed to the original column and only the new column is used.

Item	Field Name	Comments	Action
1	ReservationStatusDate	Too many subclasses because it has character data type.	Converted the data type into numeric, with continuous modeling type and date format (yyyy-mm-dd).
2	Company	Large proportion of data is missing (94.3%).	No action taken. As we are not using the data to create a model, we only need to be careful when including the column into our analysis. It might still be able to provide insight.
3	StaysInWeekendNights	Has unusually large value (4 in 99.5% quantile but the maximum is 19).	The same records that are outliers for StaysInWeekNights. No action taken. It means they are just making a long-term booking, probably an apartment suite booking.
4	StaysInWeekNights	Has unusually large value (10 in 99.5% quantile but the maximum is 50).	The same records that are outliers for StaysInWeekendNights. No action taken. It means they are just making a long-term booking, probably an apartment suite booking.
5	Adults	Has unusually large value (3 in 99.5%	-

		quantile but the maximum is 55).	
6	Babies	Has unusually large value (1 in 99.5% quantile but the maximum is 10).	-
7	Children	Not in the correct data type because null is recorded as "NA".	Recoded "NA" as missing values. Hide and exclude from analysis.
8	LeadTime	Has unusually large value (478 in 99.5% quantile but the maximum is 737).	-
9	Meal	"SC" and "Undefined" are both described as "no meal package" in the data dictionary.	Recoded "Undefined" as missing values. Hide and exclude from analysis.
10	MarketSegment	Missing values are recorded as "Undefined".	Recoded "Undefined" as missing values. Hide and exclude from analysis.
11	DistributionChannel	Missing values are recorded as "Undefined".	Recoded "Undefined" as missing values. Hide and exclude from analysis.
12	Country	Missing values are recorded as "NULL".	Recoded "NULL" as missing values. Hide and exclude from analysis.
13	PreviousCancellations	Has unusually large value (1 in 99.5% quantile but the maximum is 26).	-
14	PreviousBookingsNotCanceled	Has unusually large value (7 in 99.5%	-

		quantile but the maximum is 72).	
15	ReservedRoomType	Has two less categories than AssignedRoomType	Not enough information to determine the cause of the discrepancy. No action taken.
16	BookingChanges	Has unusually large value (4 in 99.5% quantile but the maximum is 21).	-
17	Agent	Non-missing values are recorded as "NULL".	Recoded "NULL" as "No Agent".
			Hide and exclude from analysis.
18	Company	Non-missing values are recorded as "NULL".	Recoded "NULL" as "No Company".
			Hide and exclude from analysis.
19	ADR	Has unusually large value (275 in 99.5% quantile but the maximum is 5400).	-
20	RequiredCarParkingSpaces	Has unusually large value (1 in 99.5% quantile but the maximum is 8).	-
21	File Name	Can be renamed to "Hotel" and have the ".csv" removed from the value to promote easier understanding.	Recoded to remove ".csv" from the values into a new column called "Hotel".
			Hide and exclude from analysis.
22	ArrivalDateYear, ArrivalDateMonth, ArrivalDayOfMonth	Do not have a combined column with continuous data type.	Addressed in item 23 and 24.
23	ArrivalDateMonth	Character column cannot be joined in a formula to concatenate date.	Recoded ArrivalDateMonth into number (e.g. January → 1) .

			Converted data type into numeric and modelling type into ordinal.
24	ArrivalDate	New formula column.	Calculated the combined value of ArrivalDateYear, ArrivalDateMonth, and ArrivalDayOfMonth.
			Converted modelling type into date (d/m/y format).
25	TotalStay	New formula column.	Calculated total number of nights the reservation was for.
26	BookingDate	New formula column.	Calculated the date when the booking was made by subtracting LeadTime from ArrivalDate.
27	LeadTimeCancelled	New formula column.	Calculated the difference between ArrivalDate and ReservationStatusDate if the reservation status is not "Check-Out".
28	IsCancelled	Not in the correct modelling type because it is coded as 0 and 1.	Recoded into "Cancelled" for 1 and "Not Cancelled" for 0 to promote easier understanding. Recoded values are stored in a new column called "Cancellation". Automatically converted into character data type and nominal modeling type.
			Hide and exclude from analysis
29	IsRepeatedGuest	Not in the correct modelling type because it is coded as 0 and 1.	Recoded into "Repeated Guest" for 1 and "New Guest" for 0 to promote easier understanding.

			Automatically converted into character data type and nominal modeling type.
30	CompanyBooking	New formula column.	Record "Company" if there is a company ID in the Company column, or else record "No Company".
31	AgentBooking	New formula column.	Record "Agent" if there is an agent ID in the Agent column, or else record "No Agent".

References

- Antonio, N., de Almeida, A., & Nunes, L. (2019). Hotel booking demand datasets. *Data in Brief*, 22(2019), 41-49. doi:10.1016/j.dib.2018.11.126
- Funnell, R. (2019, May 8). *The real cost of 'free' cancellations*. Retrieved from Triptease: https://triptease.com/blog/the-real-cost-of-free-cancellations/?utm_source=MediaPartner&utm_medium=HotelSpeak
- Orange Hotel Marketing. (2016, November 15). *Increasing cancellation rates are hurting hotels*. Retrieved from Orange Hotel Marketing: <https://www.orangehotelmarketing.com/increasing-cancellation-rates-hurting-hotels/>