MOHAMED ALGHALY

EXCEL DATA ANALYSIS PROJECT

Data Wrangling

let's start by getting to know our data.

Start with 2018

- We have 2 hotels (City - Resort).
- Is cancelled columns is a 0/1 columns with no missing values.
- Lead Time between 0 and 737.
- Arrival date year is always 2018.
- Arrival moth is available for only 6 months.
- Week number is between 27 & 53.
    - 7 * 52 = 364 out of 365/366 so the year can have 53 weeks (the last one will only have few days)
    - 739 rows are in week 53 out of 22k (3 %), that could happen.
- Day of month between 1 – 31 which is good.
- Adults between 0 & 53.
- Children between 0 and 10 with 4 missing values.
- Babies between 0 and 9.
- Meals have the right values with 170 values missing, replace them with None.
- Countries look good.
- Market segment has two missing values, same rows as children, channel, and company.
- Channel has 5 missing values.
- Is Repeated Guest is a 0/1 column.
- Previous Cancellation has values between 0 & 26.
- Previous Uncancelled Bookings between 0 & 14.
- Reserved Room from A to L and so is assigned.
- Booking Changes between 0 & 20.
- Deposit Type looks good.
- Agent ID has 3099 missing values, replace it with 0 indicating undefined.
- Company ID has 20691 missing values.
- Days in waiting list between 0 & 122.
- Customer Type is a 4-value categorical column.

- I am interested in knowing if the customer has got the type of room he has asked for, so I will add 0/1 column for that.

Aside from the company column we have about 10 missing values most out of 22k, so we will just drop any missing values and drop the company column.

Even if we decided to make the company column a 0/1 column indicating whether the ticket is for individuals or company, 1305 company rows Vs 20600 for individuals will make the analysis biased so just ignore this feature.

After Removing rows with missing values (except for the meal column we will assume missing values as no meals registered), WE HAVE LOST ONLY 5 ROWS.

Time to change data types to the appropriate values.

Same for 2019, 2020.

Then, Join the 3 sheets.

Done with Data Wrangling phase, now the data is clean and good to go.

The analysis will be divided by 2 factors, Year (2018, 2019, and 2020) and Booking State (Cancelled or Not), and I will connect 2 filters to be able to explore each possible option.

Data Analysis

We have 141937 Bookings, 52831 (37.22%) of which are cancelled.

As we have different cancellation rates for each hotel, we will add a filter on that to be able to analyze the trends for each separately.
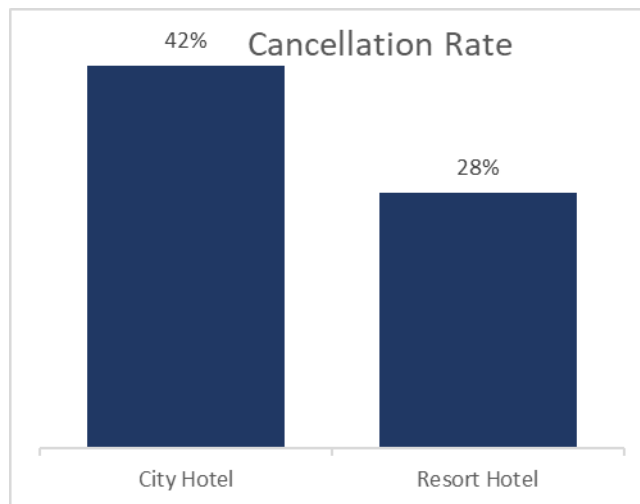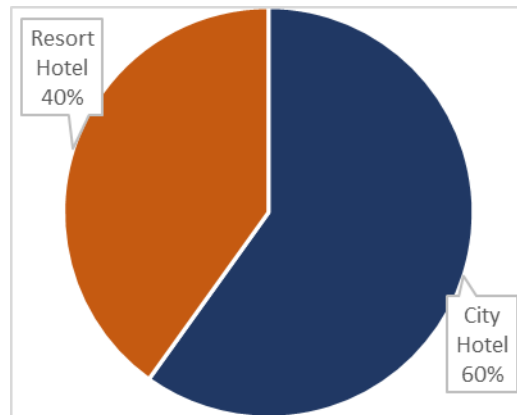

As we see the cancellation rate is higher for City Hotel, which can mean **we are not a family place**, but let's not rush into anything.

NOTE:

I will analyze each factor order by its appearance in the dictionary file not by importance, but there will be insights and recommendations sections at the end of the project docs.
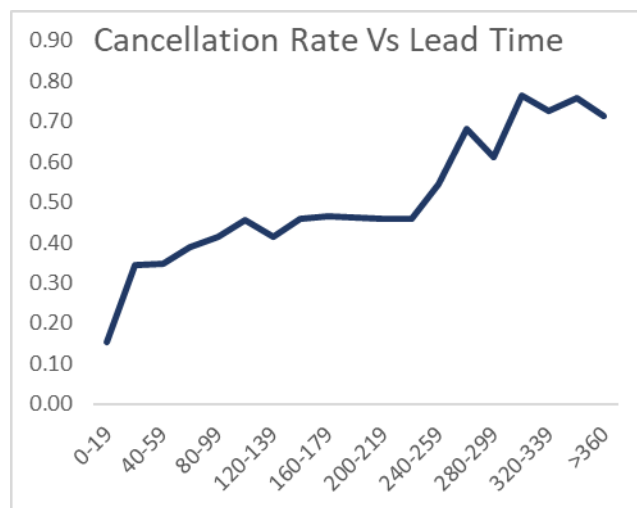
Hotel:

Here I want to see which hotel is the most popular.

It looks like the bookings' rates are **higher for City Hotel**.
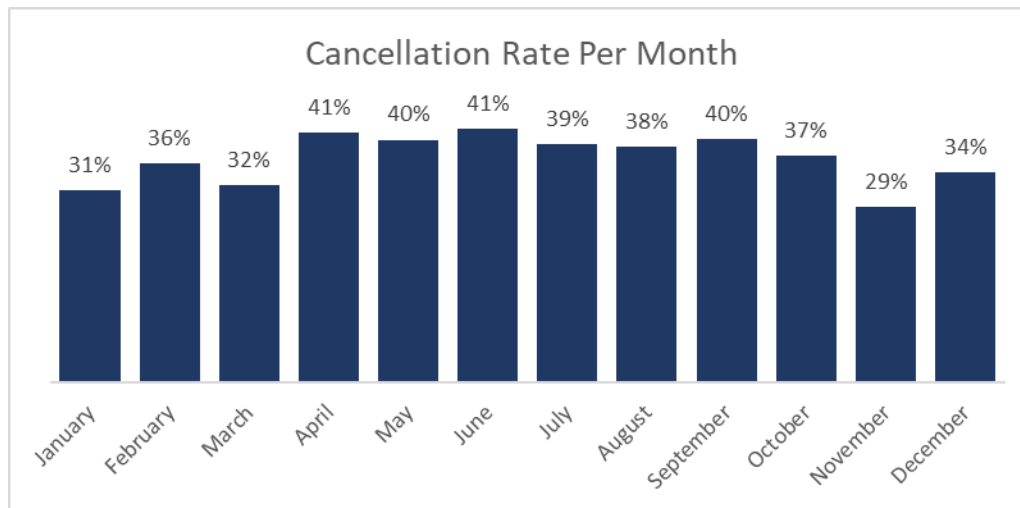
Lead Time:

Here I want to see if the time between booking the ticket and the arrival time affects the cancellation rate.

It is clear that bookings made way before the arrival time are **most likely to get cancelled.**
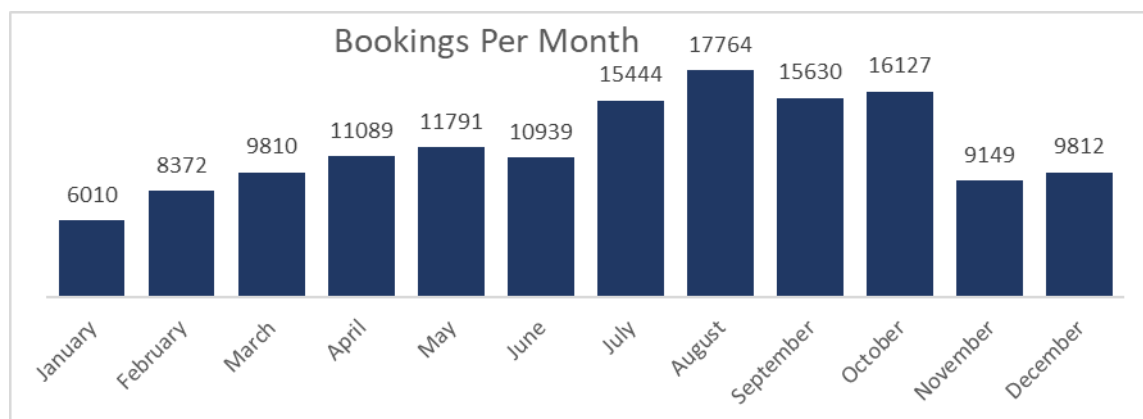
Month:

We want to see the booking/cancellation trends over each month.

There is no clear pattern here, but it is obvious that for each year or over the past 3 years for each Hotel or over both hotels the cancellation rates are **lower in November**, which needs deeper investigation.

Let's have a look at booking trends.

Booking activations are **most active in July – August – October**, which is weird, as I have expected to see higher rates in December due to holidays which means we need deeper investigations.

Week Number:

Will add nothing to the combination of month/day of month.

Day of month:

We are concerned with the change in bookings/cancellations trends over each month.
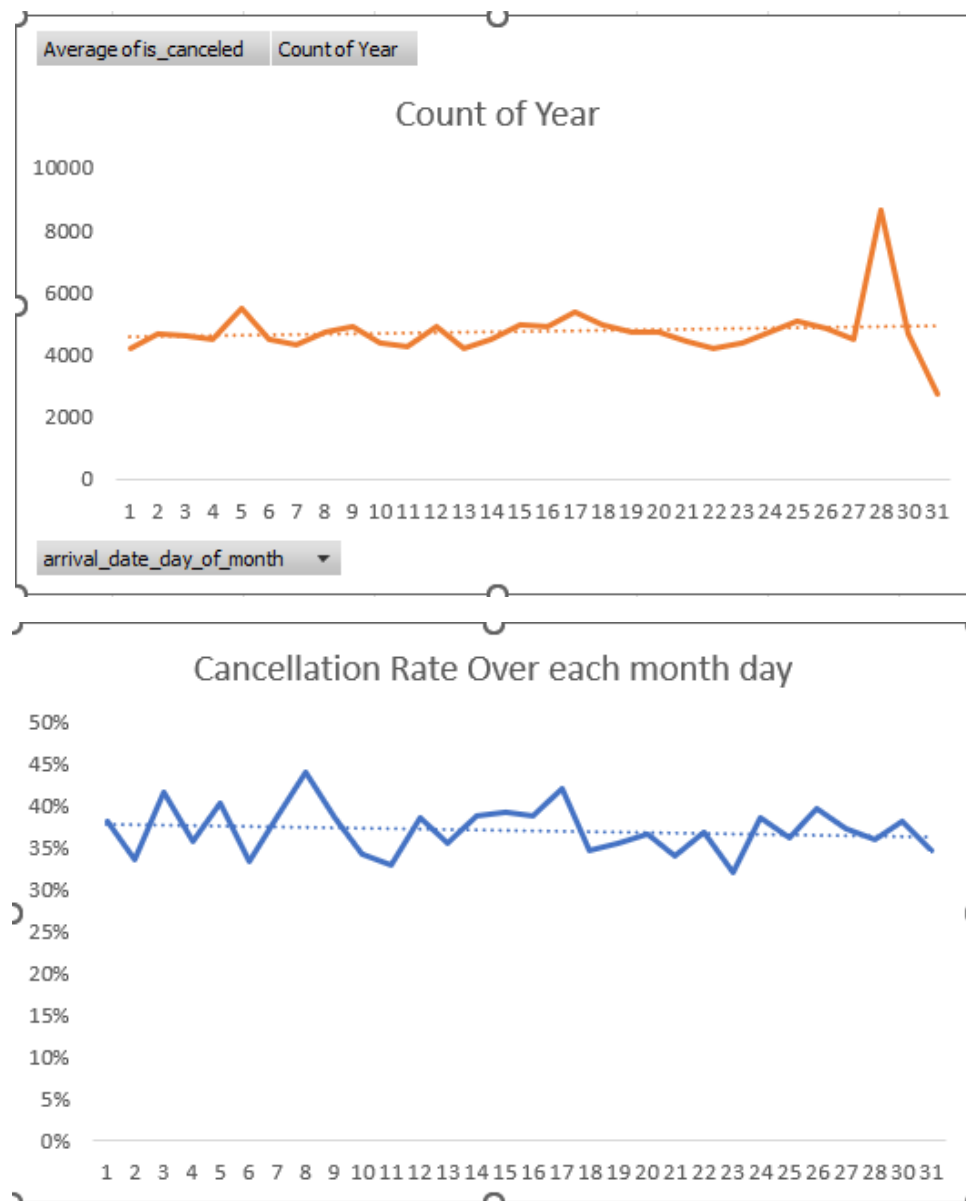
Booking/Cancellation activations are **not affected by the day number**, which make sense.
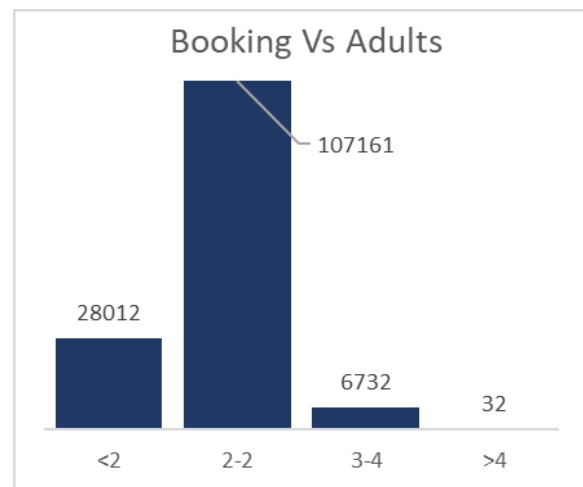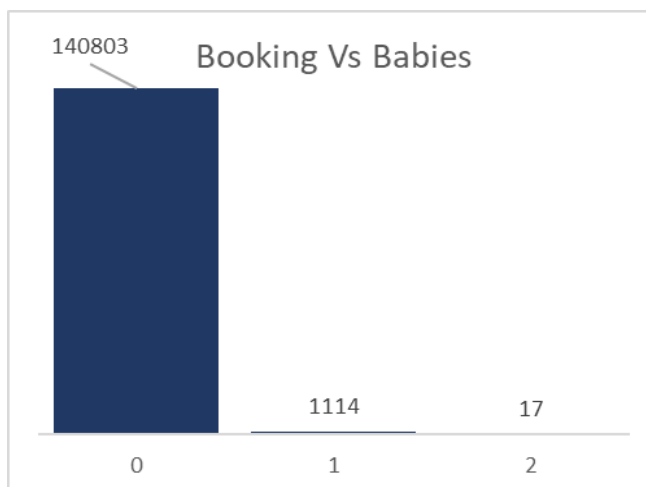
As we saw the booking's rate is higher only at the end of the month.

Stay In Weekend/Week Nights:

Seems to me there is nothing interesting here.

Adults/Children/Babies:

Let's see how we handle each age: we will start with the cancellation rate.
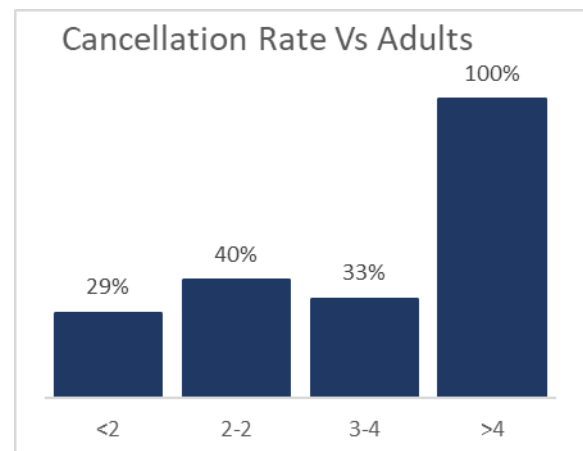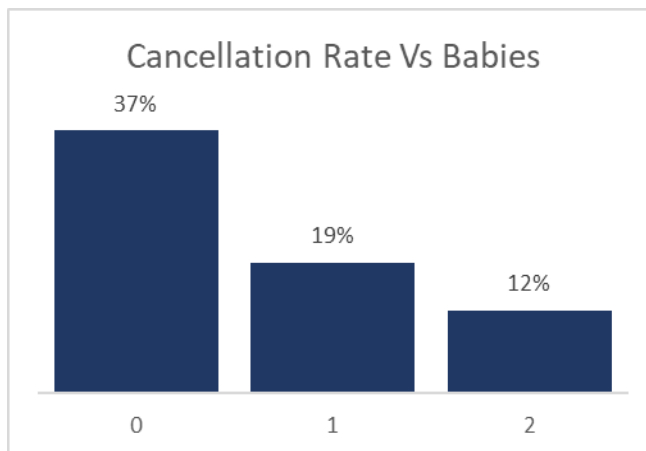
Booking rates obviously **decrease** as the number of people increases (babies, children, or adults), which means we have **problems dealing with large parties.**

From those charts we could clearly get the feeling that we **ARE NOT** a family place.

Booking rates is low for < 2 because no one will book a hotel room to enjoy quality time with himself, so the pattern starts from 2.

Cancellation rates are 100% for more than 4 adults, which is evidence that we **can't handle groups**.

The cancellation rates pattern for kids is not trustworthy as the number of bookings with babies/children are too low.

Meal:

	BB	FB	HB	None	SC
■ Can	41041	779	6671	338	4002
■ All	109870	1228	18400	1372	11067
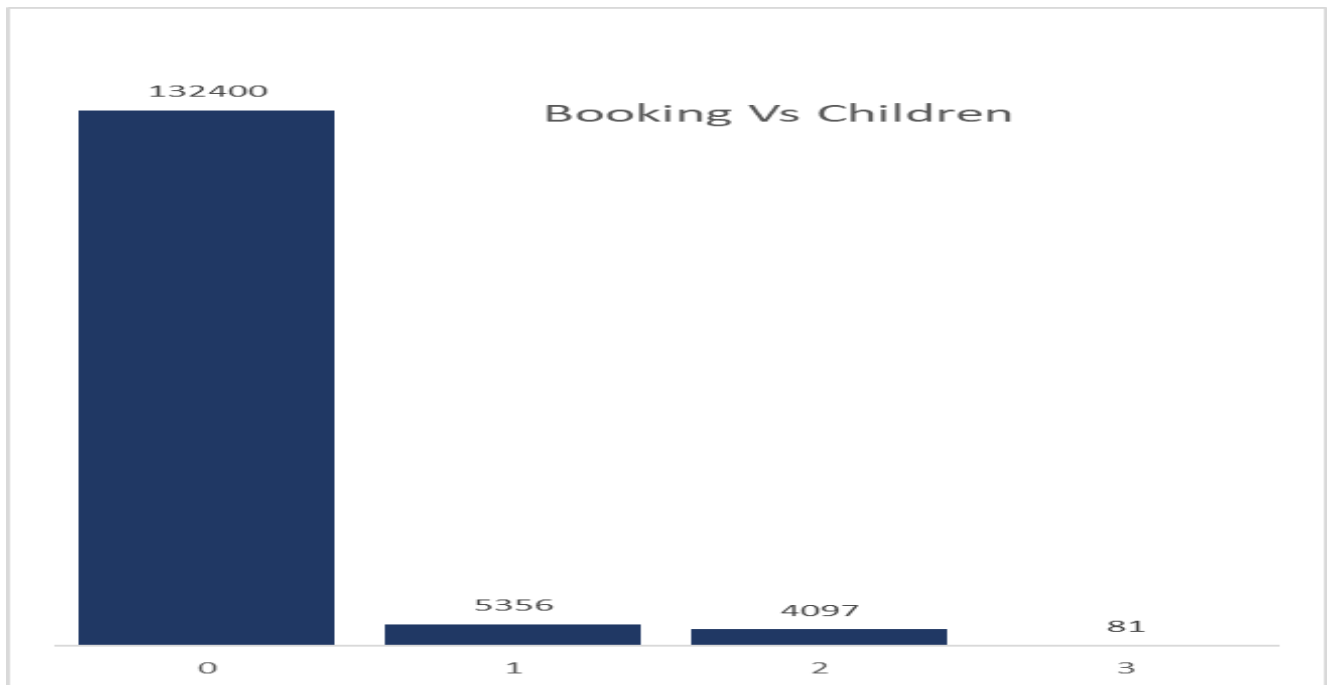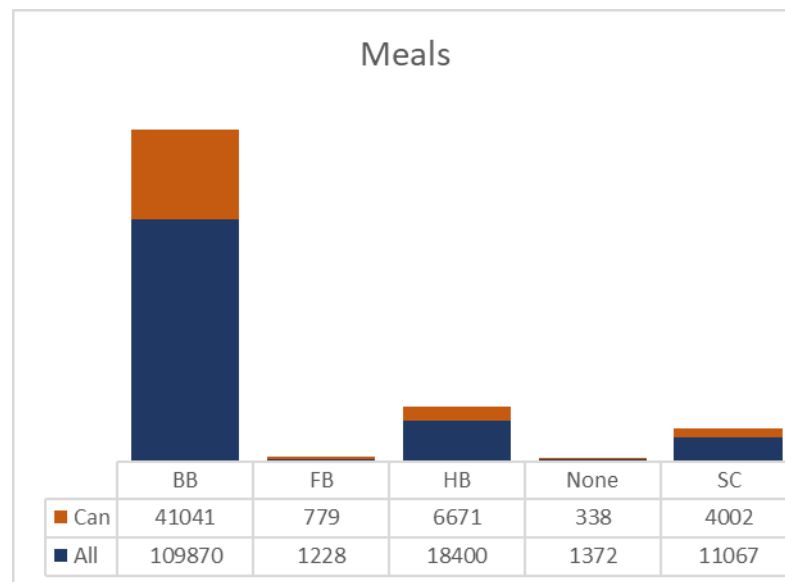
It looks like only the lowest cost meal is acceptable.

Cancellation rates with **FB meals are 63%!**

Country:

Seems to me there is nothing interesting here.

Market Segment:
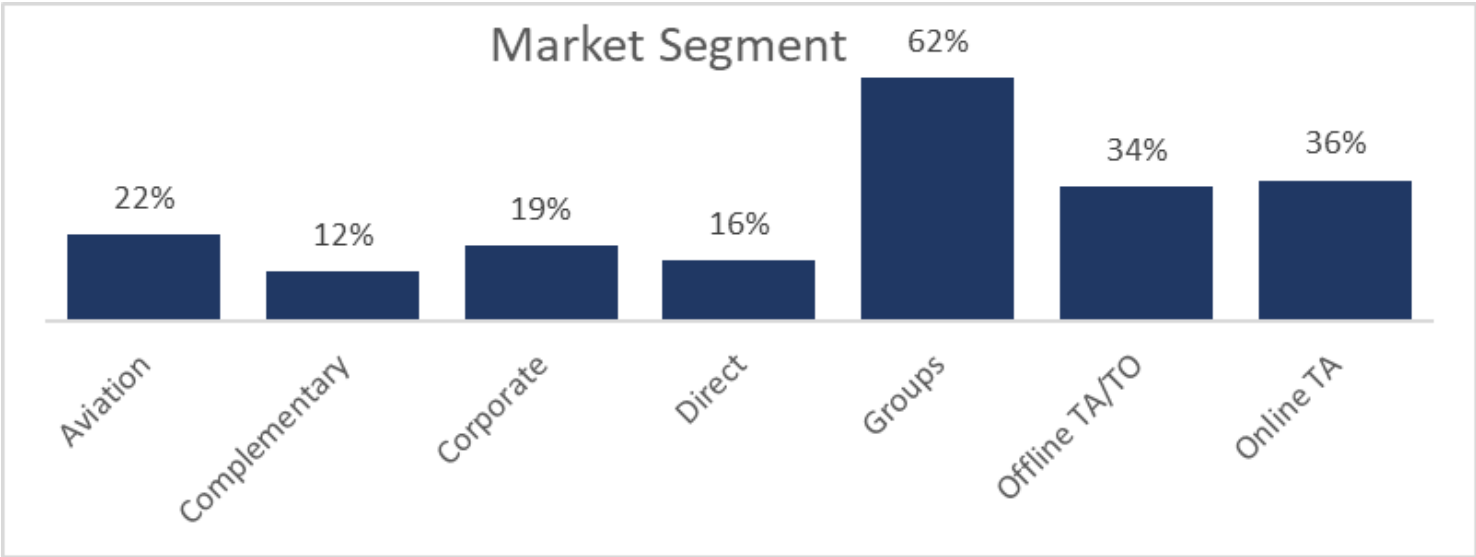
As we have stated before, when it comes to **groups, we have a problem**.

Channel:

	Corporate	Direct	GDS	TA/TO
■ Can	1874	3142	37	47778
■ All	8167	17534	194	116042
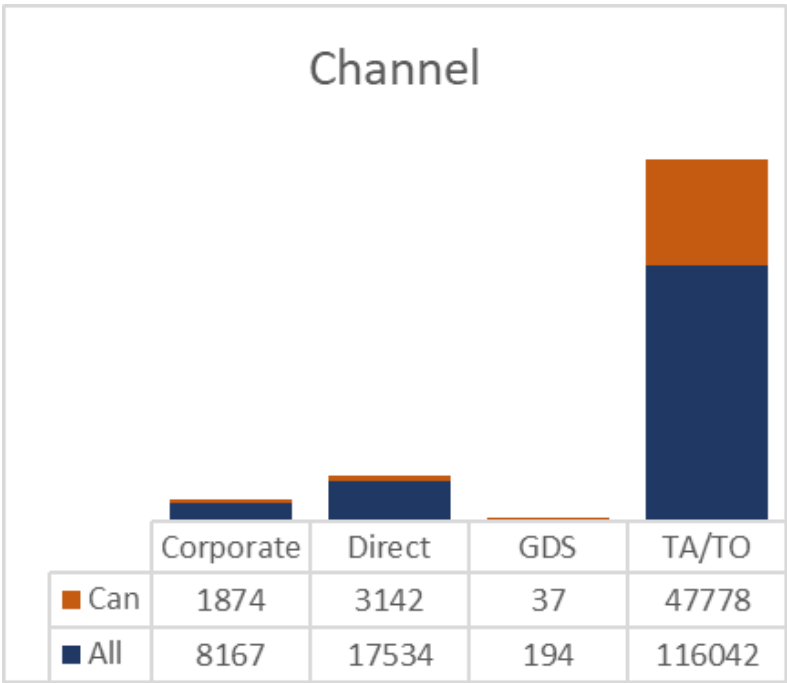
Only TA/TO Channel is working as expected, but with **higher cancellation rates.**

Days in waiting list:

Most of the values are 0 so ignore this feature.

Previous Cancellations:
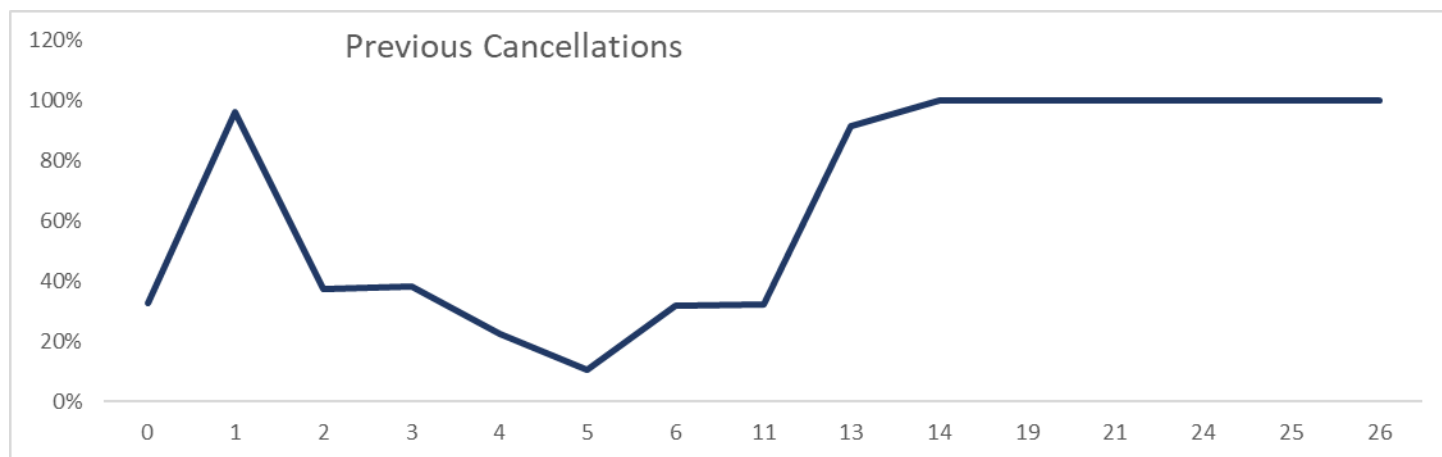
People who have cancelled on us more than 14 times **will cancel on us again**, again, and again.

People with previous cancellations of **more than 10 are a danger.**

Previous Bookings not Canceled:
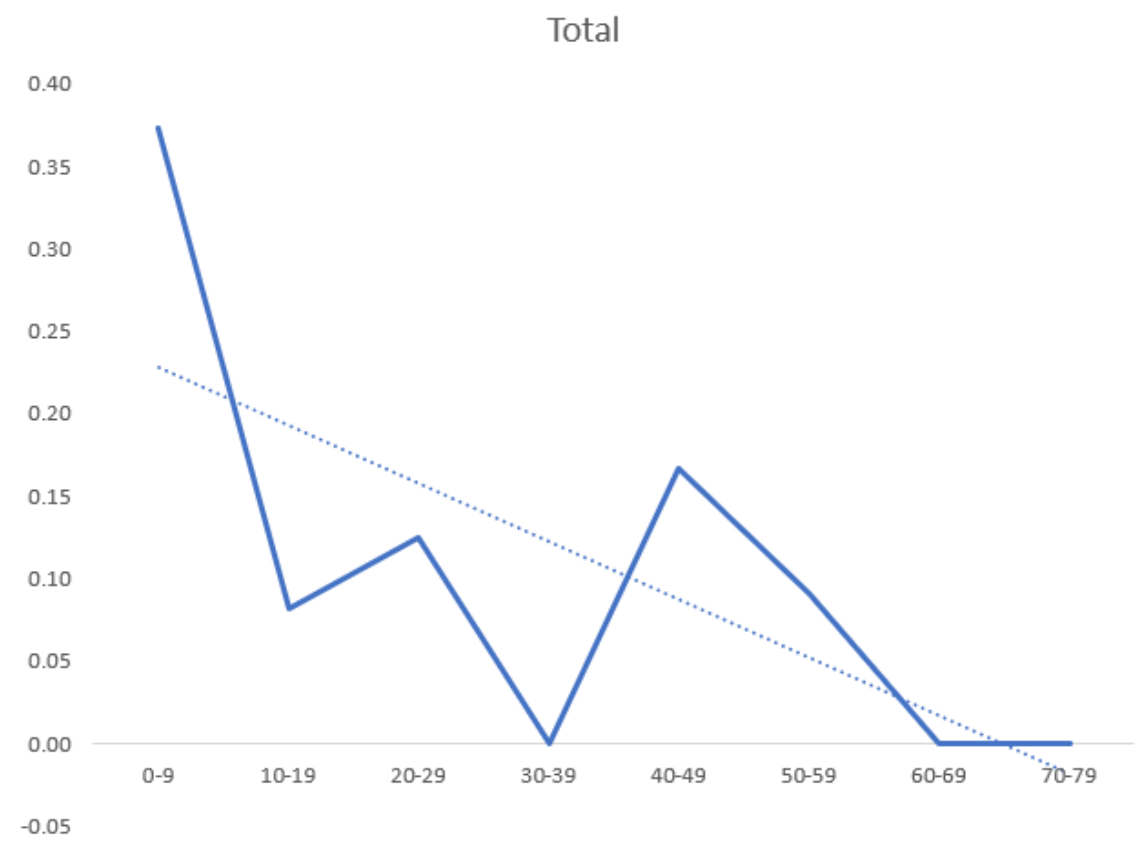
People with previous cancellations of higher previously uncancelled bookings **are most likely to not cancel on us**.

Same:

We want to see if the cancellation rate is higher when the customer gets a room different from what he has asked for.
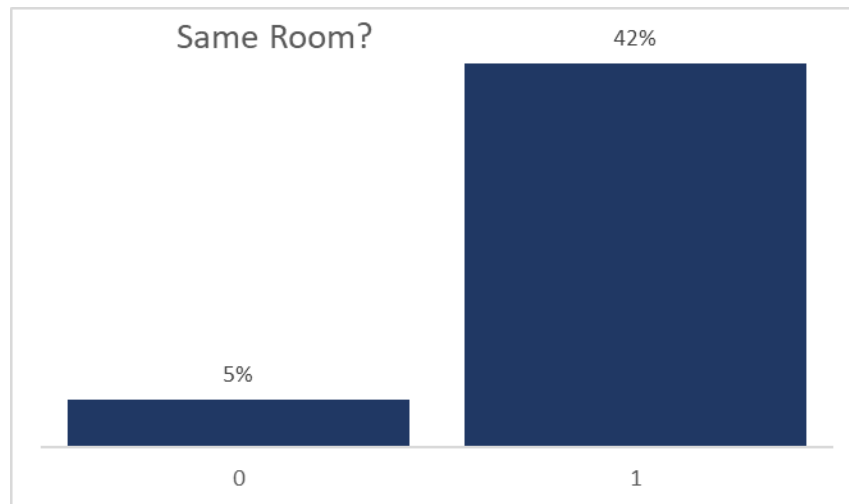
As expected, when the customer does not get the room, he has asked for he will most likely **cancel the reservation.**

Changes:
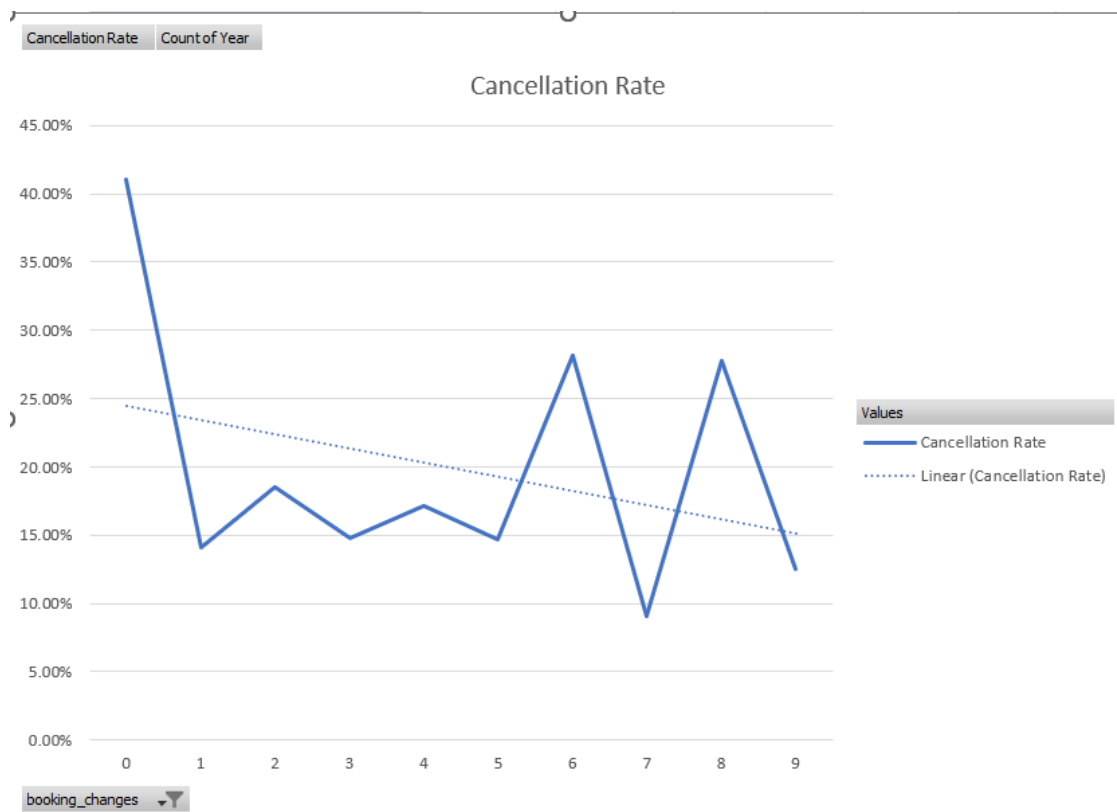
There is no pattern here.

Deposit Type:
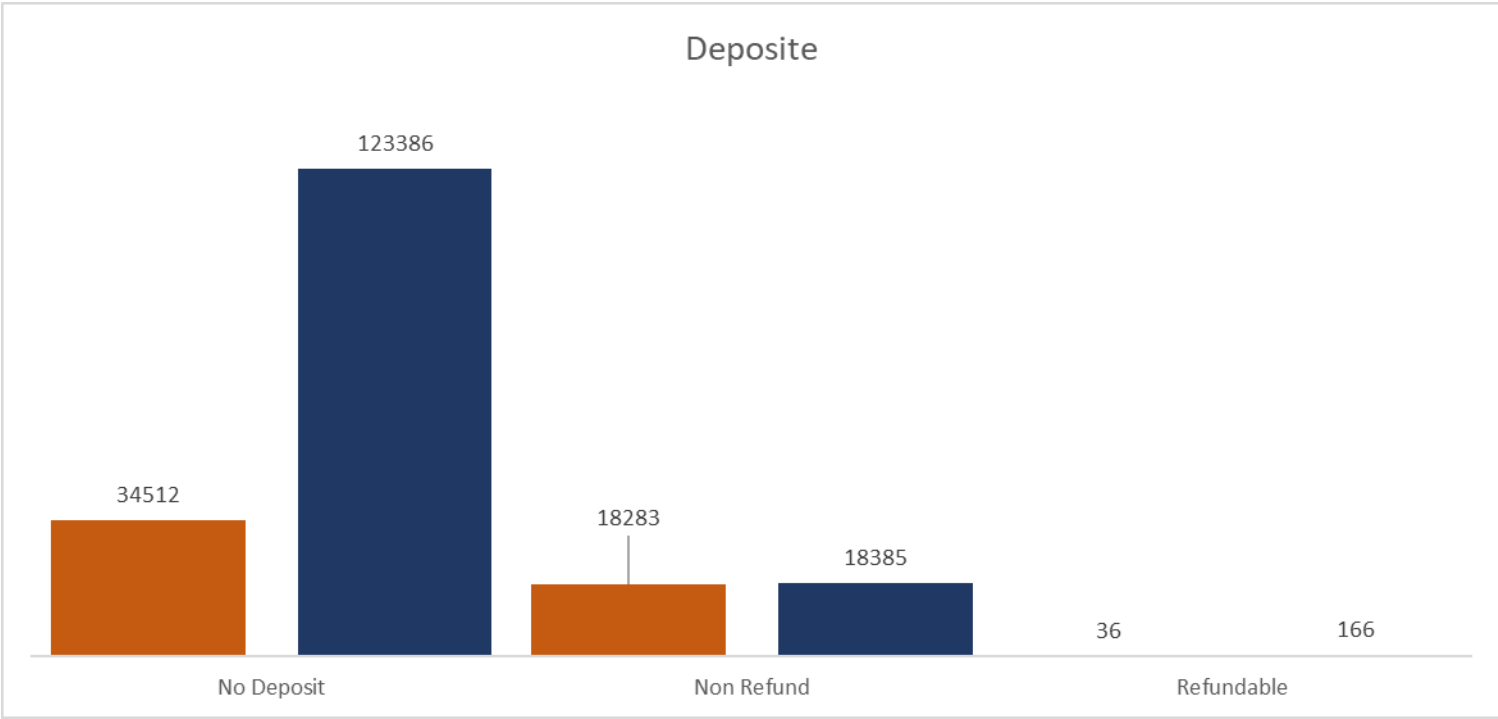
When it comes to No Deposit/Refundable **we are good.**

But for Non-Refundable deposits, there **is an issue as cancellation rate is almost 100 %.**

Customer Type:
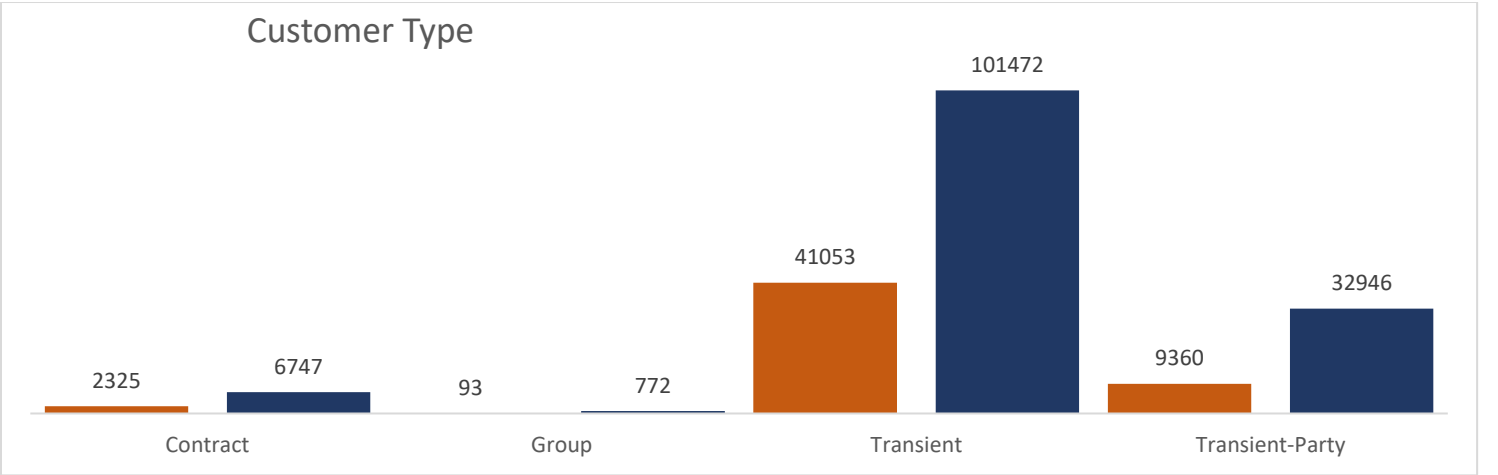
There is nothing interesting going on here.

Is Repeated Guest:

We are concerned with the cancellation rate for regular/unregular customers.
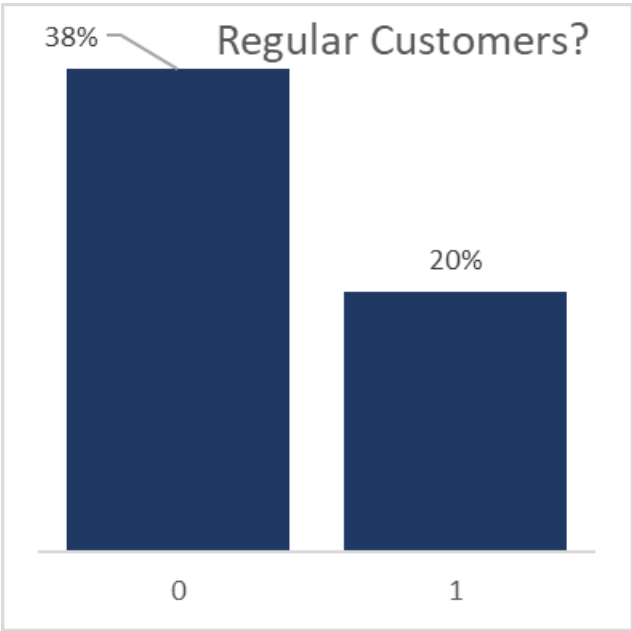
Seems like we know how to deal with regular customers, as the cancellation rate are way higher for unregular customers.

Agent:

We want to know who is the best performing salesman.

	0	1	6	7	9	14	28	240	241	250
■ All	19549	11395	4506	3878	34290	4019	1977	16686	2052	3357
■ Can	4800	8229	1650	516	13895	712	122	6436	281	602
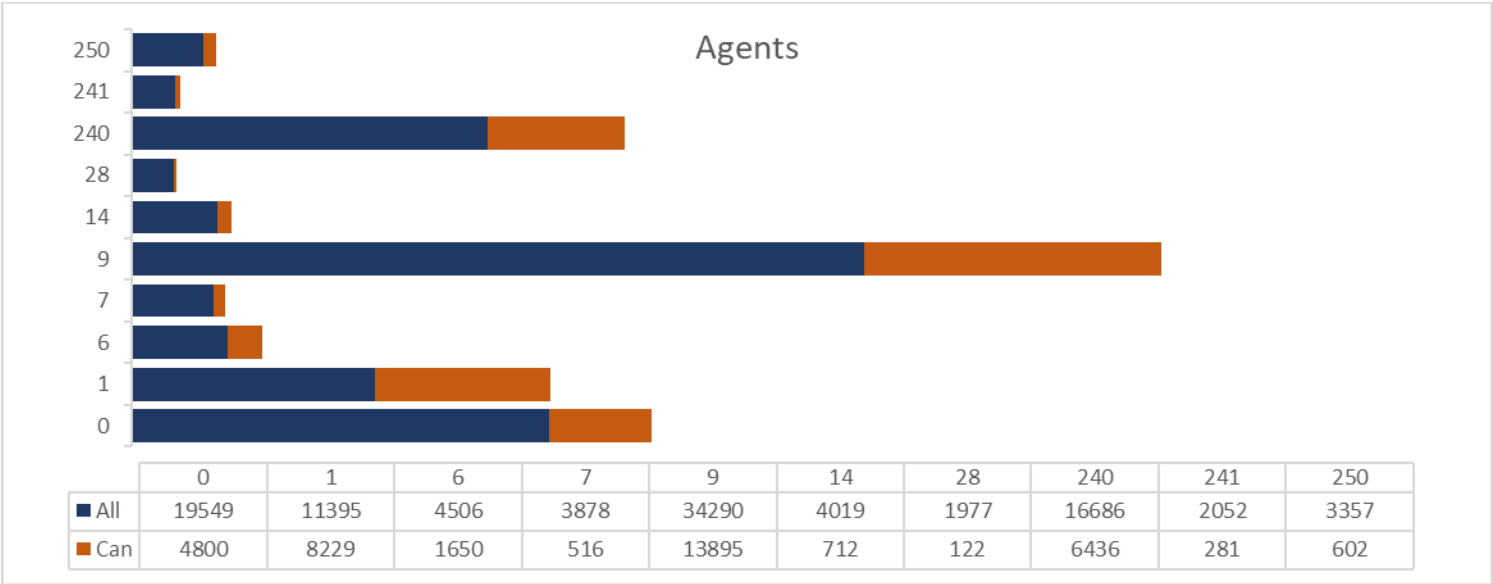
I would go with agent 250 and agent 9.

Dashboards

Bookings Analysis

Insights on reservations demand and key factors that impact the bookings.

Booking Analysis

Bookings	Cancellation Rate	Cancellations
141937	37.22%	52831

Pie Chart
Resort Hotel 40%
City Hotel 60%

Bookings Per Month

Month	Bookings
January	6010
February	8372
March	9810
April	11089
May	11791
June	10939
July	15444
August	17764
September	15630
October	16127
November	9149
December	9812

Agents

	0	1	6	7	9	14	28	240	241	250
All	19549	11395	4506	3878	34290	4019	1977	16686	2052	3357
Can	4800	8229	1650	516	13895	712	122	6436	281	602

Meals

	BB	FB	HB	None	SC
Can	41041	779	6671	338	4002
All	109870	1228	18400	1372	11067

Channel

	Corporate	Direct	GDS	TA/TO
Can	1874	3142	37	47778
All	8167	17534	194	116042

Deposite

No Deposit	Non Refund	Refundable
34512 / 123386	18283 / 18385	36 / 166

Customer Type

Contract	Group	Transient	Transient-Party
2325 / 6747	93 / 772	41053 / 101472	9360 / 32946

Cancellation Analysis

Insights on what are the factors that affect bookings' cancelation .

Cancellations	Cancellation Analysis	Cancellation Rate
52831		37.22%

Cancellation Rate

- City Hotel: 42%
- Resort Hotel: 28%

Cancellation Rate Per Month

Month	Rate
January	31%
February	36%
March	32%
April	41%
May	40%
June	41%
July	39%
August	38%
September	40%
October	37%
November	29%
December	34%

Cancellation Rate Vs Lead Time

(line chart, axis 0.00 to 0.90; x-axis: 0-19, 40-59, 80-99, 120-139, 160-179, 200-219, 240-259, 280-299, 320-339, >360)

Same Room?

- 0: 5%
- 1: 42%

Regular Customers?

- 0: 38%
- 1: 20%

Previous Cancellations

(line chart, axis 0% to 120%; x-axis: 0 1 2 3 4 5 6 11 13 14 19 21 24 25 26)

Market Segment

Segment	Rate
Aviation	22%
Complementary	12%
Corporate	19%
Direct	16%
Groups	62%
Offline TA/TO	34%
Online TA	36%

Age Analysis

Analyze customer satisfaction over each age.

Age Booking/Cancellation Analysis
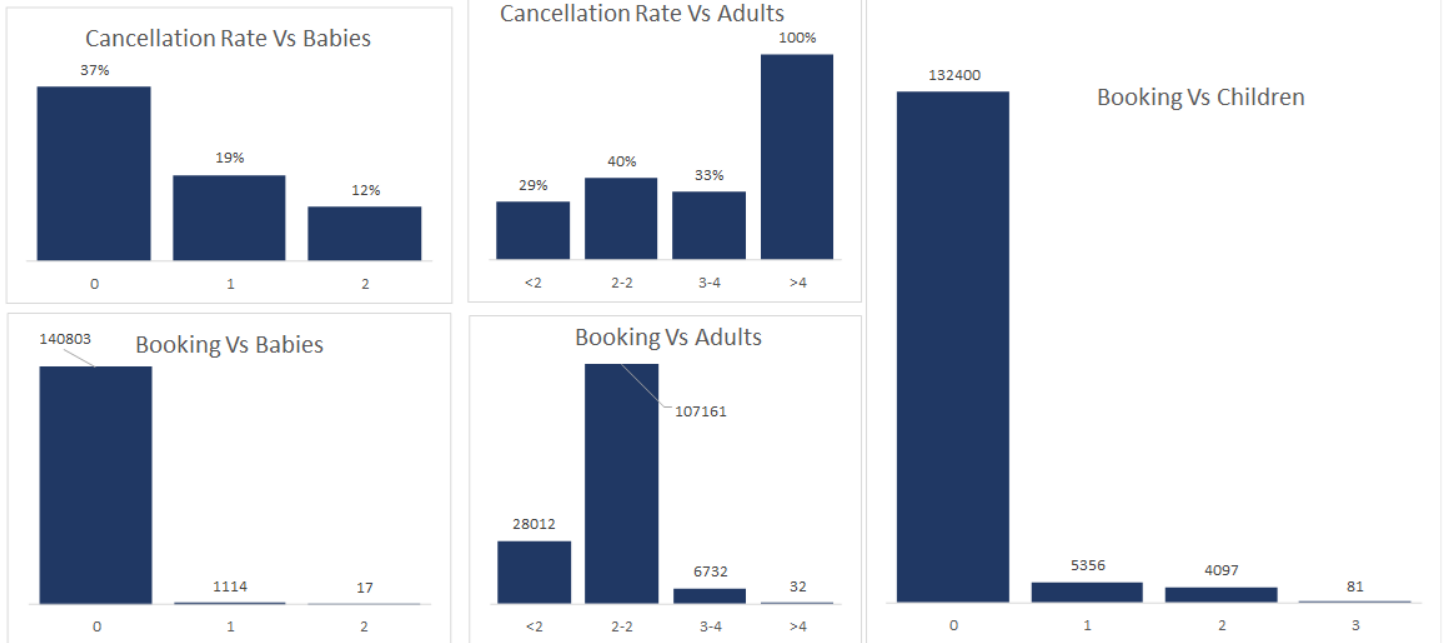
INSIGHTS & RECOMMENDATIONS

1- Cancellation rates and booking rates are higher for City Hotel.

Recommendations:

- As cancellation rates are higher for City Hotel, we need to make some extra effort to improve its performance.

2- The cancellation rate increases as the gap between booking and arrival increases.

Recommendations:

- We should set a limit on the lead time, like you can't book a room before a month of arrival.
- We can set the limit to 2 weeks max.

3- Cancellation rates are lower in November.

Recommendations:

- We need to communicate with the management team to understand why cancellation rates are lower in November.

4- Bookings are higher in July – August – October.

Recommendations:

- We should do some marketing and offers in other months.

5- Bookings are high at the end of the month.

Recommendations:

- We can offer more products or services to benefit from rush days.

6- We have serious issues with groups, families, and large parties.

Recommendations:

- We should come up with a program to offer that suits families and groups.

7- Most of the meals we are offering are not popular.

Recommendations:

- We must upgrade our kitchen and improve our food.

8- Most of the booking channels are not active.

Recommendations:

- We need to hire a good marketing team.

9- People with previous cancellations of more than 10 are a danger.

Recommendations:

- We can refuse any booking when the customer has cancelled more than 10 bookings before.
- Or take a non-refundable deposit from him.

10- When the customer doesn't get the room, he has asked for, he is upset.

Recommendations:

- We should try our best to give him the room he wants.
- When a customer asks for any type of room give him the least popular one.
- We need to communicate with the management team to come up with other solutions for this issue.

11- Non-Refundable deposits are not acceptable.

Recommendations:

- We could limit that option as we can.

12- Cancellation rates are higher for non-regular customers.

Recommendations:

- We can make offers and deals for new customers.

13- The best performing salesmen are 250 and 9.

Recommendations:

- We can make offers for salesmen to improve their performance.