CS5346
Project Interim report

OPTIMIZING HOTEL BOOKING DEMAND:
A DATA-DRIVEN APPROACH TO ENHANCE OCCUPANCY AND CUSTOMER SATISFACTION

GROUP 11

DENG XINRUI(A0274719R) ZHANG YI(A0276634W) ZHAO SHANHE(A0287084U)

Background

Introduction

🤔 What is the background and purpose of our project?

The hospitality industry is a cornerstone of the global economy, offering services to travelers and tourists worldwide. Analyzing hotel booking demand is crucial for understanding market dynamics, customer preferences, and operational challenges within this sector. The dataset here we chose, "*Hotel Booking Demand*," provides a comprehensive overview of booking information for a city hotel and a resort hotel, including details like booking time, length of stay, the number of guests, cancellations, and more.

This analysis aims to unlock insights into consumer behavior, identify trends and patterns, and optimize hotel operations for better financial performance and customer satisfaction.

Introduction

🤔 Why are we interested in booking data for hotels?

All three of us love traveling, and booking hotels is something we do regularly. It's a part of our lifes that we can connect with on a personal level, and it's a great way to blend our interests with academic study, making us more enthusiastic about data analysis. Out of all the datasets out there, this one is one of the few that gives us plenty of details to work with. We believe that combining these details will give us some really interesting ideas.

By looking at the data, we can really understand how people behave when booking hotels, what they like, and how the market changes. This gives us valuable insights that we can share with hotel owners.

Problem Statement

🤔 What is the core objective of this research?

The primary issue this project aims to tackle is the optimization of hotel booking demand to maximize occupancy rates, minimize cancellations, and enhance overall customer satisfaction. Despite the vast amount of data available, hotels often struggle to predict demand accurately, leading to lost revenue opportunities and operational inefficiencies.

🤔 What do we expect we can bring to the business operator?

For hoteliers, understanding and predicting booking patterns is vital for strategic planning and operational success. Accurate demand forecasts enable better staffing, inventory management, and pricing strategies.

Questions about the dataset

🤔 What is the dataset source? Is it reliable?

The dataset is originally from the article ***Hotel Booking Demand Datasets***. And this dataset was extracted from the SQL database of the hotel Property Management System (PMS).

The author only did very simple processing of the dataset. The dataset was also open-sourced with the article and was accessible to subsequent researchers. Therefore it is recognized as trustworthy dataset.

🤔 What does this dataset contain?

The dataset contains bookings arriving between July 1,2015, and August 31,2017, including valid arrivals and cancellations. Since this is the real data of the hotel, all data elements related to the identity of the hotel or the customer are removed.

Questions about datasets

🤔 What are the characteristics of this dataset?

The main types of this dataset are as follows, it contains 32columns and more than 119k line:

Data	Data Type	Description
hotel	Nominal	H1 = Resort Hotel / H2 = City Hotel
is_canceled	Nominal	Value indicating if the booking was canceled (1) or not (0)
lead_time	Quantitative	Number of days that elapsed between the entering date of the booking into the PMS and the arrival date

Questions about datasets

🤔 What are the characteristics of this dataset?

The main types of this dataset are as follows:

Data	Data Type	Description
arrival_date_year/month/week_number/day_of_month	Quantitative	Year/Month/Week_number/Day_of_month of arrival date
stays_in_weekend/week_nights	Quantitative	Number of weekend / week nights the guest stayed or booked to stay at the hotel
adults / children /babies	Quantitative	Number of adults / children / babies

Questions about datasets

🤔 What are the characteristics of this dataset?

The main types of this dataset are as follows:

Data	Data Type	Description
meal	Nominal	Type of meal booked. Categories are presented in standard hospitality meal packages: Undefined/SC – no meal
country	Nominal	Country of origin
market_segment / distribution_channel	Nominal	market_segment Market segment designation / Booking distribution channel. “TA” means “Travel Agents” and “TO” means “Tour Operators”

Questions about datasets

🤔 What are the characteristics of this dataset?

The main types of this dataset are as follows:

Data	Data Type	Description
is_repeated_guest	Nominal	Value indicating if the booking name was from a repeated guest (1) or not (0)
previous_cancellations/booking	Quantitative	Number of previous bookings that were cancelled / not cancelled by the customer prior to the current booking
reserved/assigned_room_type	Nominal	Code of room type reserved / assigned. Code is presented instead of designation for anonymity reasons.

Questions about datasets

🤔 What are the characteristics of this dataset?

The main types of this dataset are as follows:

Data	Data Type	Description
booking_changes	Quantitative	Number of changes/amendments made to the booking from the moment the booking was entered on the PMS
deposit_type	Nominal	Indication on if the customer made a deposit to guarantee the booking.
agent	Nominal	ID of the travel agency that made the booking
company	Nominal	ID of the company/entity that made the booking or responsible for paying the booking

Questions about datasets

🤔 What are the characteristics of this dataset?

The main types of this dataset are as follows:

Data	Data Type	Description
days_in_waiting_list	Quantitative	Number of days the booking was in the waiting list before it was confirmed to the customer
customer_type	Nominal	Type of booking
adr	Continuous	Average Daily Rate as defined by dividing the sum of all lodging transactions by the total number of staying nights
required_car_parking_spaces	Quantitative	Number of car parking spaces required by the customer

Questions about datasets

🤔 What are the characteristics of this dataset?

The main types of this dataset are as follows:

Data	Data Type	Description
total_of_special_requests	Quantitative	Number of special requests made by the customer (e.g. twin bed or high floor)
reservation_status	Nominal	Reservation last status (Three types)
reservation_status_date	Temporal	Date at which the last status was set. This variable can be used in conjunction with the ReservationStatus to

Questions about datasets

🤔 What are the characteristics of this dataset?

In short, there are four types:

1. Nominal: Hotel type (city hotel and resort hotel), meal, etc
2. Quantitative: Number of adults, number of children, demand for parking spaces, etc
3. Temporal Data: Year, month, week, day, etc
4. Continuous Data: Average Daily Rate

Beneficiaries

🤔 Who are the beneficiaries of our project?

By leveraging Tableau for data visualization and analysis, we can uncover valuable findings that can inform strategic decisions for Hotel Managers and Owners

🤔 What can they obtain from our project?

By gaining insights into customer behavior patterns, this enables the formulation of staffing, pricing strategies, and resource allocation plans, thereby improving operational efficiency.

Insights gleaned from the dataset can also inform enhanced revenue management strategies, improved customer experiences, and more efficient resource allocation.

Moreover, this project has the potential to bolster the sustainability and resilience of the hotel industry in navigating economic fluctuations and evolving consumer behaviors.

Plan

🤔 How can we complete our plan?

Utilizing the powerful visualization tool Tableau, we can present our analysis results in the form of interactive charts and visual maps, making the data more intuitive and comprehensible. For instance, we can create dynamic charts using Tableau to illustrate trends in booking durations and cancellation rates, or utilize its mapping features to display booking patterns across different countries and regions. This aids in optimizing hotel operations, devising targeted marketing strategies, formulating supportive policies, fostering local economic activities, and enhancing customer experiences and satisfaction.

What we have done

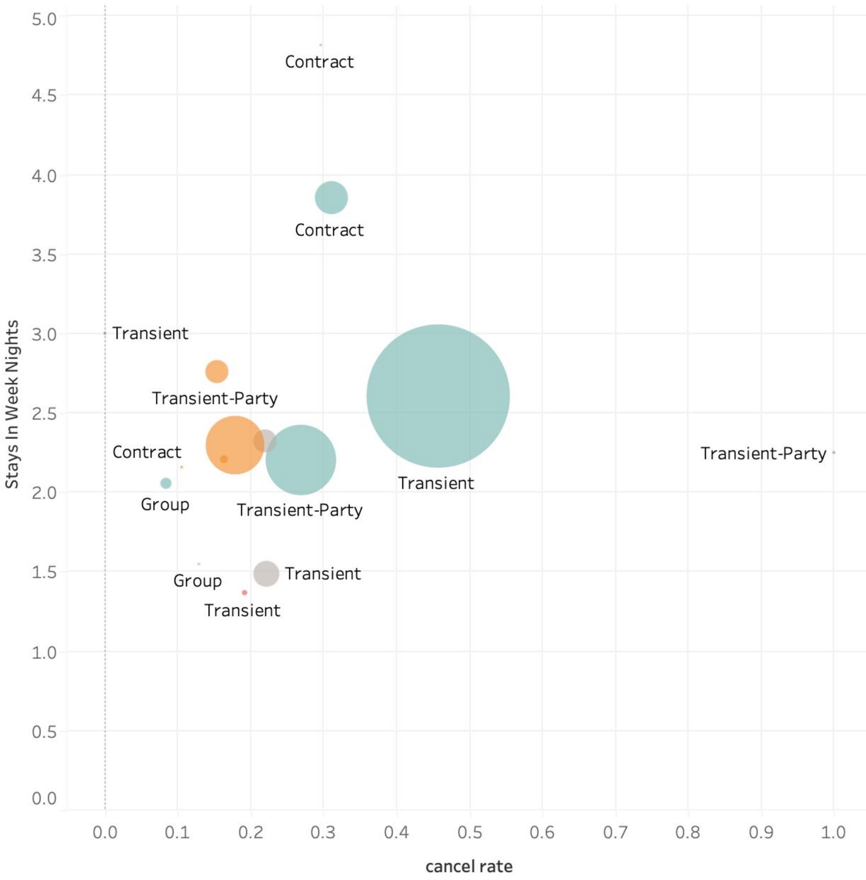
Visual Exploration

Bubble chart based on cancel rate

Focus: Cancel Rate

We want to study the relationship between cancellation rates and other attributes, for the following reasons:

- ❖ Improving resource usage: By knowing why bookings are canceled, hotels can manage rooms and staff more efficiently.
- ❖ Enhancing guest satisfaction: Understanding the reasons for cancellations allows hotels to improve services and products, leading to happier guests.
- ❖ Refining marketing strategies: Analyzing cancellation patterns enables hotels to tailor promotions and pricing strategies, attracting more guests and increasing revenue.

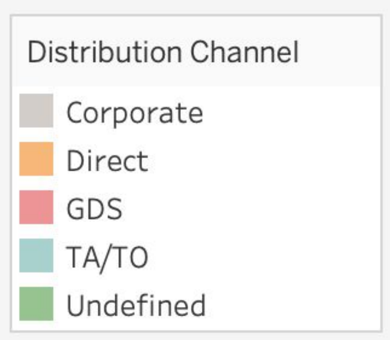


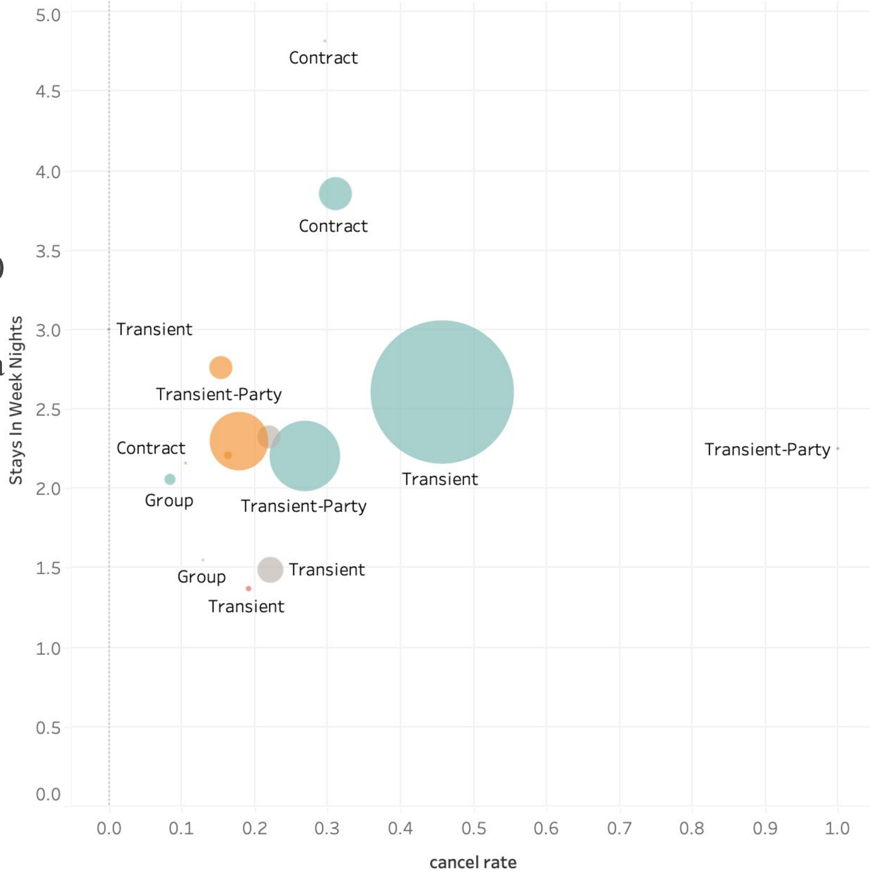
Visual Exploration

Bubble chart based on cancel rate

Encoding:

- ❖ Size: Represents numbers of people
- ❖ Color: Represents Distribution Channel
- ❖ X-axis: Represents Cancel Rate (ranging from 0 1)
- ❖ Y-axis: Represents Stays in Week Nights (Avera ranging from 1 to 5)
- ❖ Label: Represents Customer Type

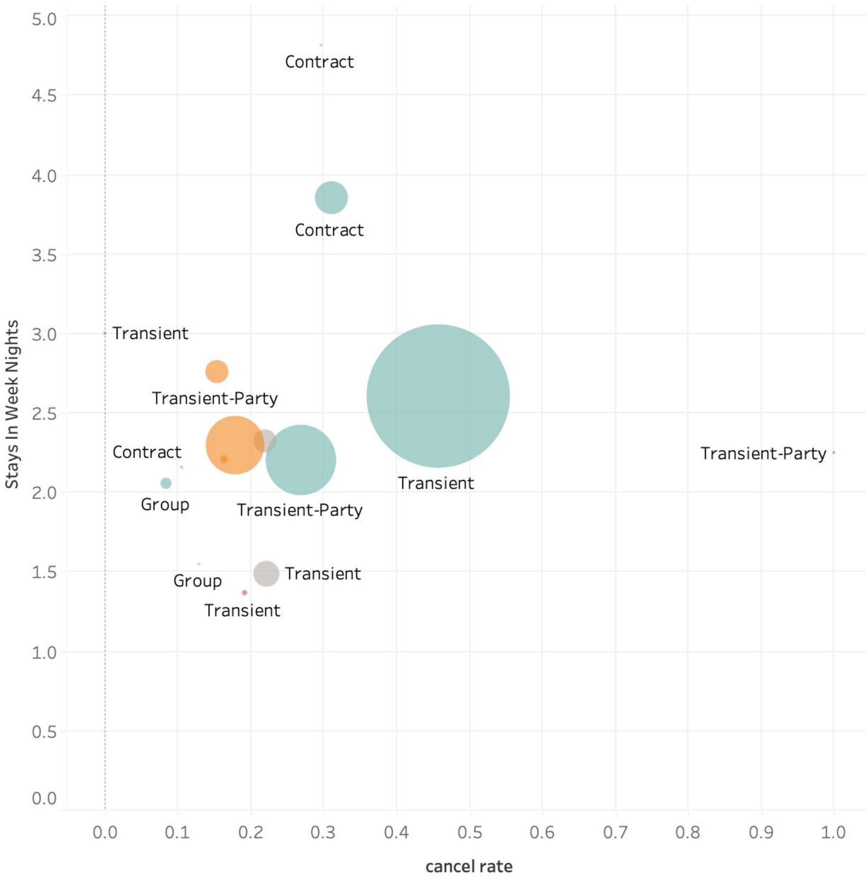




Visual Exploration
Bubble chart based on cancel rate

Finding :

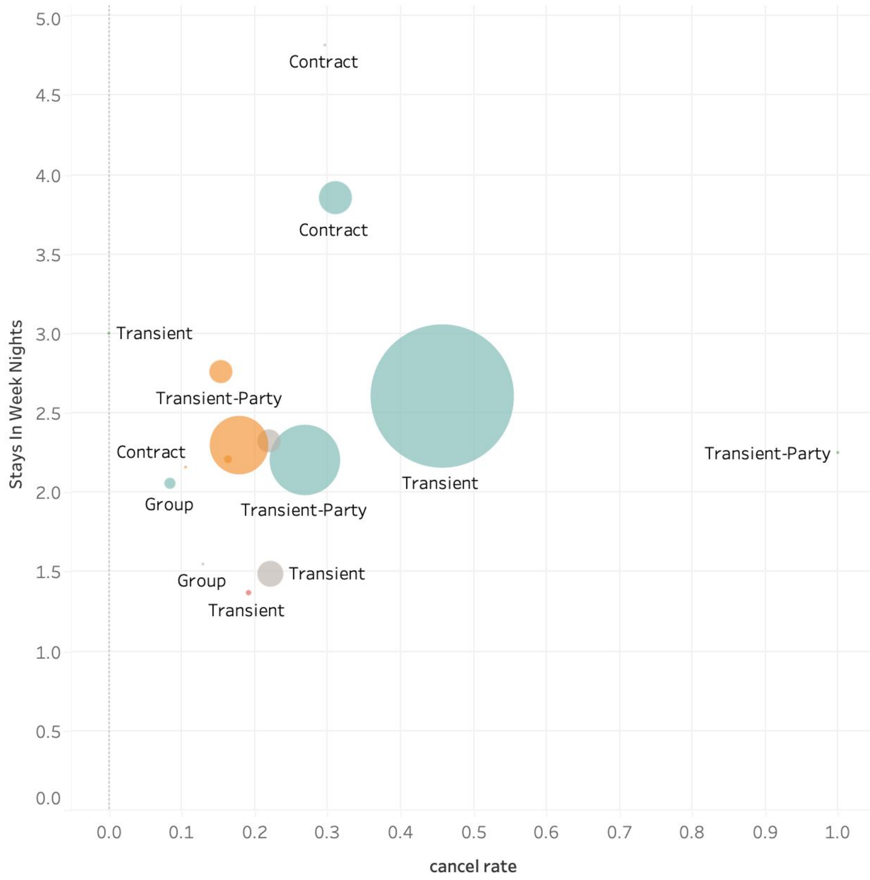
- **Y-axis Analysis:** The majority of individuals tend to book hotels for 2-3 days, with bookings rarely exceeding 5 days.
- **Color Analysis:** Blue data points (representing the TA/TO distribution channel) are clustered towards the right side, indicating a generally higher cancellation rate for TA/TO bookings. Gray data points (corporate bookings) follow, with cancellation rates falling between those of TA/TO and direct bookings. Orange data points (direct bookings) exhibit the lowest cancellation rates among the three distribution channels.
- **Label Analysis:** Transient bookings display significant variability, with cancellation rates ranging from 0 to 1. Contract bookings are more concentrated around a cancellation rate of 0.3, while group bookings have the lowest cancellation rates, with most falling around a cancellation rate of 0.1.



Visual Exploration

Conclusion

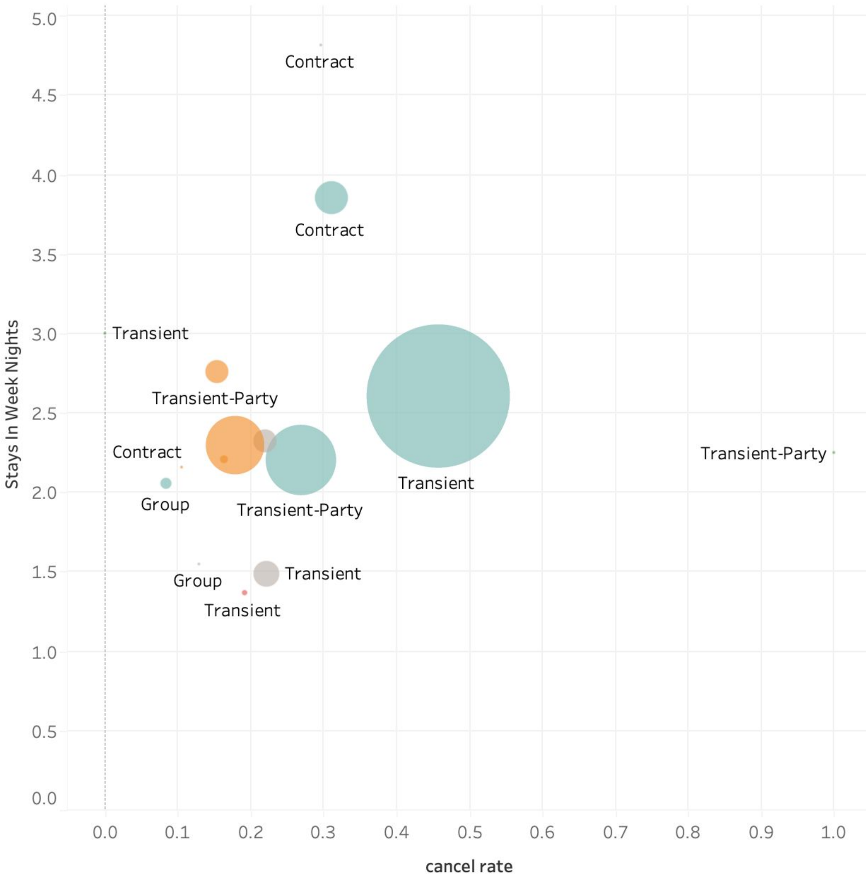
- Based on the comprehensive analysis, it can be inferred that customers booked through the TA/TO distribution channel tend to have higher cancellation rates, while those booking directly have lower cancellation rates. Additionally, cancellations for short-term guests fluctuate more compared to contract and group bookings, which exhibit relatively stable cancellation rates. These conclusions are instrumental in devising cancellation policies and marketing strategies tailored to different customer segments, thereby minimizing losses and enhancing profitability.



Visual Exploration
Bubble chart based on cancel rate

Suggestion

- ❖ TA/TO Distribution Channel: Take steps to reduce cancellation rates, such as offering incentives for non-refundable bookings or improving communication methods.
- ❖ Direct Bookings: Continue to promote direct bookings, emphasizing the lower cancellation rates and implementing strategies to enhance customer loyalty.
- ❖ Short-Term Guests: Establish flexible cancellation policies tailored to short-term guests and provide last-minute discounts to encourage bookings closer to the check-in date.
- ❖ Contract Customers: Ensure clear contract terms and provide personalized services to maintain stable cancellation rates for contract customers.
- ❖ Group Bookings: Simplify the group booking process, offer group discounts or exclusive benefits to attract bookings, and ensure attendees have a memorable experience.



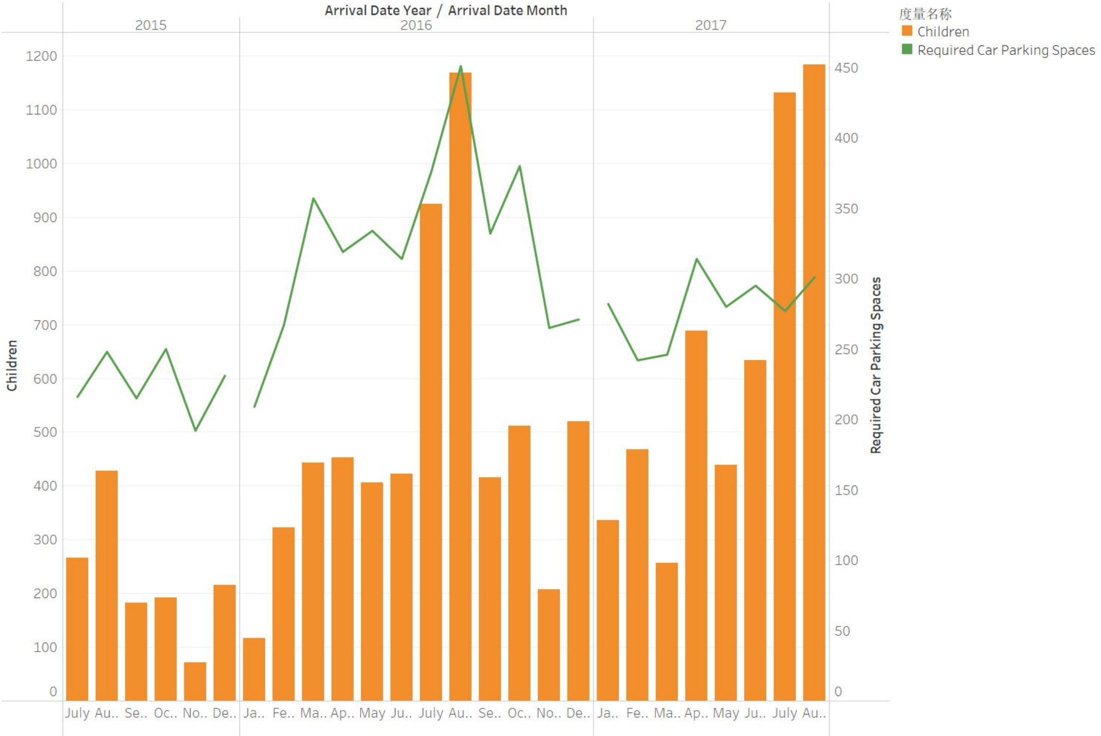
Visual Exploration

Graph between children and the demand for parking spaces

Children-Car Parking

Arrival Date Year	Arrival Date Month	Children	Required Car Parking Spaces
2015	July	266	216
	August	428	248
	September	183	215
	October	192	250
	November	72	192
	December	216	231
2016	January	116	209
	February	322	267
	March	443	357
	April	453	319
	May	406	334
	June	423	314
	July	924	376
	August	1,168	451
	September	415	332
	October	511	380
	November	207	265
	December	520	271
2017	January	336	282
	February	468	242
	March	257	246
	April	688	314
	May	439	280
	June	634	295
	July	1,132	277
	August	1,184	301

Children-Car Parking, r=0.644



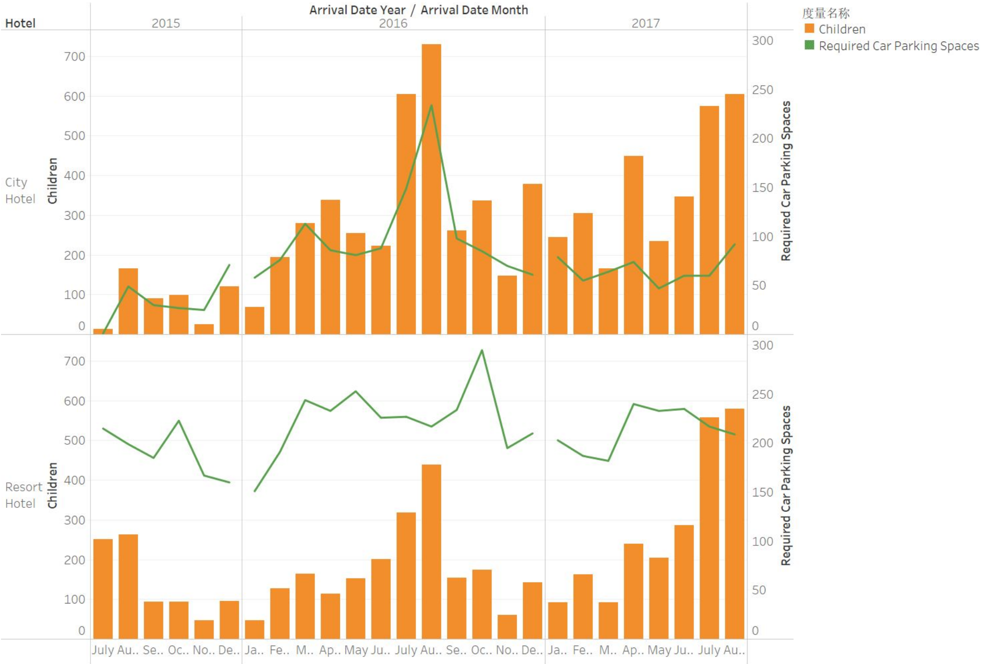
Visual Exploration

Graph between children and the demand for car parking spaces (in different hotels)

Children-Car Parking (Hotel)

Arrival Date Year	Arrival Date Month	Children: City Hotel	Children: Resort Hotel	Required Car Parking Spaces: City Hotel	Required Car Parking Spaces: Resort Hotel
2015	July	14.0	252.0	1.0	215.0
	August	165.0	263.0	49.0	199.0
	September	90.0	93.0	30.0	185.0
	October	98.0	94.0	27.0	223.0
	November	25.0	47.0	25.0	167.0
	December	120.0	96.0	71.0	160.0
2016	January	69.0	47.0	58.0	151.0
	February	194.0	128.0	76.0	191.0
	March	279.0	164.0	113.0	244.0
	April	339.0	114.0	86.0	233.0
	May	254.0	152.0	81.0	253.0
	June	222.0	201.0	88.0	226.0
	July	605.0	319.0	149.0	227.0
	August	730.0	438.0	234.0	217.0
	September	261.0	154.0	98.0	234.0
	October	336.0	175.0	85.0	295.0
	November	147.0	60.0	70.0	195.0
	December	378.0	142.0	61.0	210.0
2017	January	244.0	92.0	79.0	203.0
	February	305.0	163.0	55.0	187.0
	March	165.0	92.0	64.0	182.0
	April	448.0	240.0	74.0	240.0
	May	235.0	204.0	47.0	233.0
	June	347.0	287.0	60.0	235.0
	July	574.0	558.0	60.0	217.0
	August	604.0	580.0	92.0	209.0

Children-Car Parking (Hotel), r(city)=0.739, r(resort)=0.275



PPMCC

PPMCC is used to measure the degree of linear correlation between variables X and Y in two sets of data. It is the ratio of the covariance of two variables times their standard deviation; therefore, it is essentially a normalized measure of covariance, so the result always has a value between -1 and 1.

Value of X
Value of Y
Summation: "Take The Sum Of"
Mean of X Variable
Mean of Y Variable
Sample Correlation Coefficient

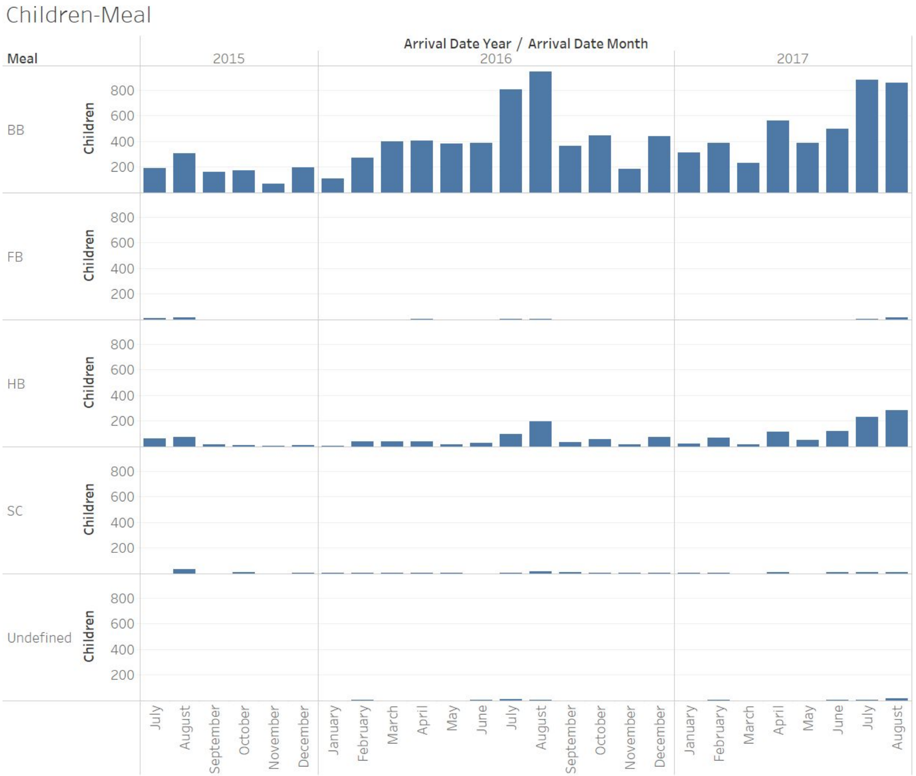
$$r = \frac{\sum\left[\left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)\right]}{\sqrt{\underbrace{\Sigma\left(x_i - \overline{x}\right)^2}_{\text{Sum of the squared deviations for X}} * \underbrace{\Sigma\left(y_i - \overline{y}\right)^2}_{\text{Sum of the squared deviations for Y}}}}$$

Square Root

|PPMCC| (|r|)	Correlation
0.1~0.3	Weak Correlation
0.3~0.5	Moderate Correlation
0.5~1	Strong Correlation

Visual Exploration

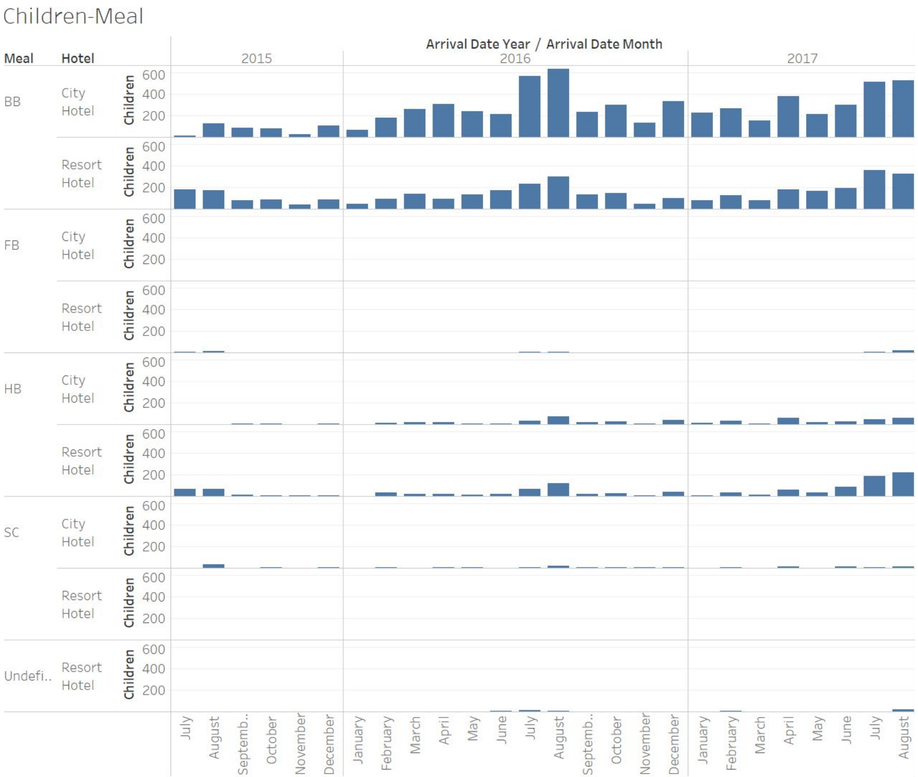
Graph between children and meal



- Most children staying in hotels like breakfast-only package service.
- Most of the other children choose the package service that provides breakfast, lunch and dinner.

Visual Exploration

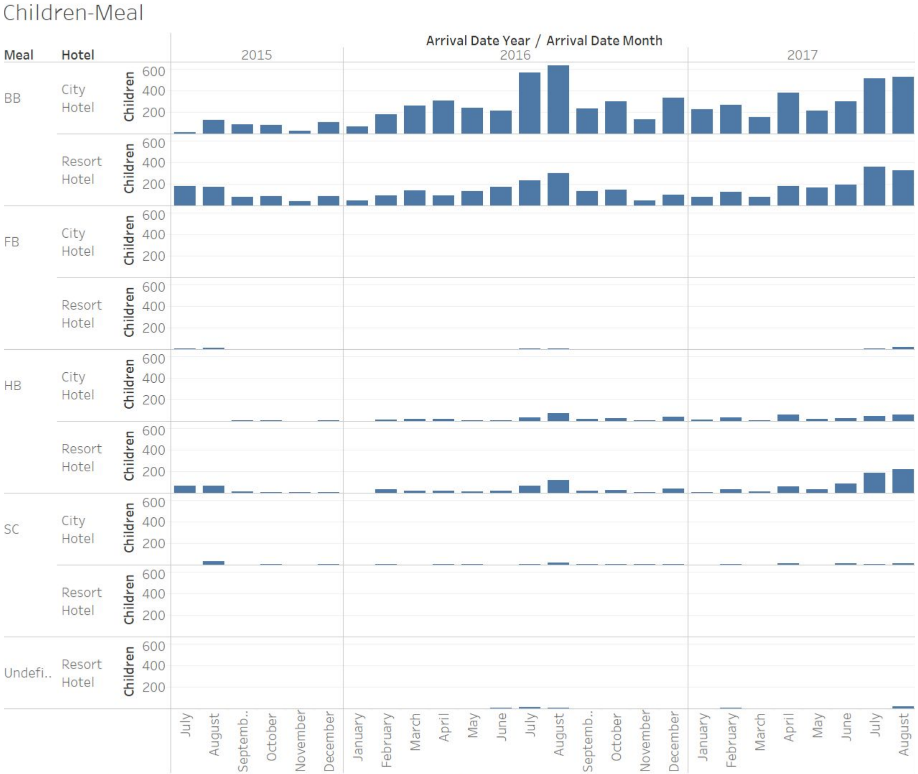
Graph between children and meal (in different hotels)



- Relative to resort hotels, the children-meal relationship graph shows stronger cyclical volatility for city hotels.
- Whether they are city hotels or resort hotels, the peak number of meals all occurs in July and August.

Visual Exploration

Graph between children and meal (in different hotels)



- In addition to the peak data, there is also a small peak in the city hotel meal data in April.

Visual Exploration

Map based on Customer Type - Country - Quarter

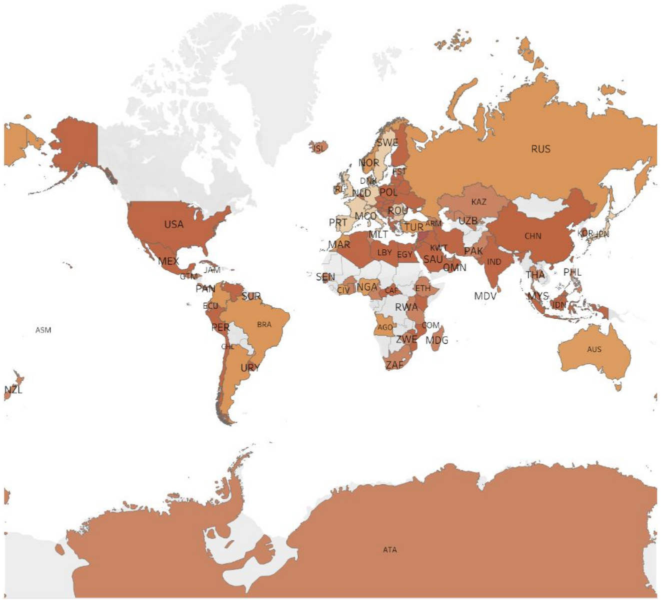
Research background: When exploring the time dimension, it is found that the data only exist in 3 years, thus the trend line cannot be drawn, and the change trend in the unit of month or day is not significant.

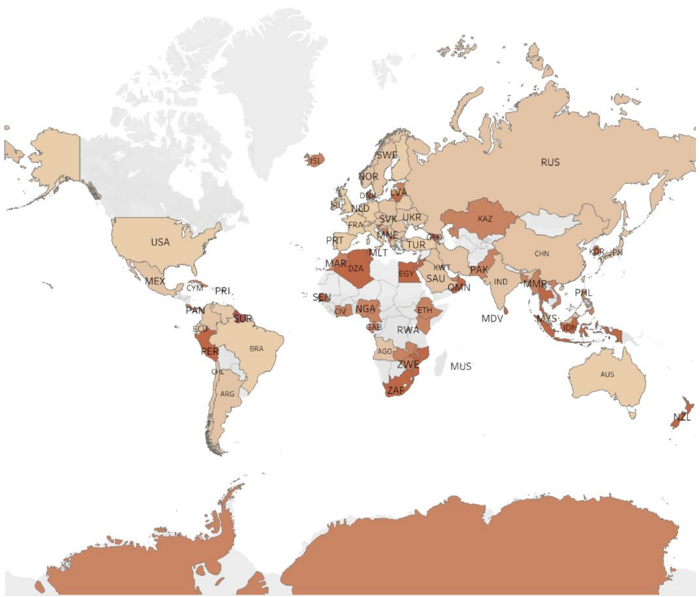
Advanced exploration: Try to correlate countries and customer types, add time latitude, and draw four colored maps of countries and customer types changing with the quarter (q1 ~ q4) as the basis of map segmentation.

The study found: In China, for example, in the last quarter, the main customer type was Transient, that is, individual tourists without contracts and not with groups, while in the first quarter the main customer type was Contract, that is, contract tourists. Australia, Russia and other countries also have their own relationship with seasonal changes in tourist types.

Visual Exploration

Map based on Customer Type - Country - Quarter





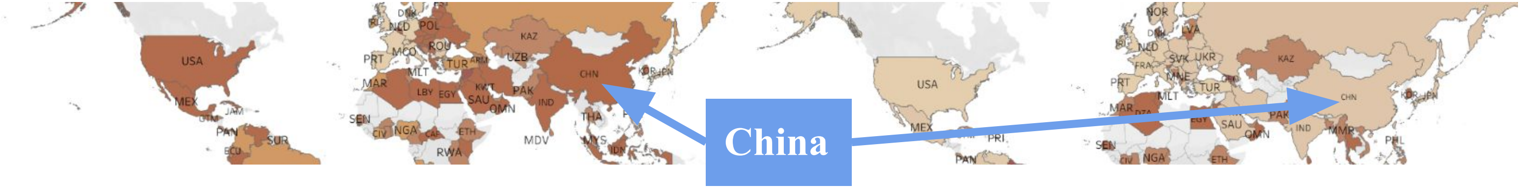
Here are the two most representative quarters in the data feedback: Q1 and Q4

Visual Exploration

Map based on Customer Type - Country - Quarter

Take China as an example, the first quarter includes the Chinese Spring Festival, which is a small peak for Chinese tourists in the past year, and the whole family takes a high proportion of trips, so contract customers are the main body.

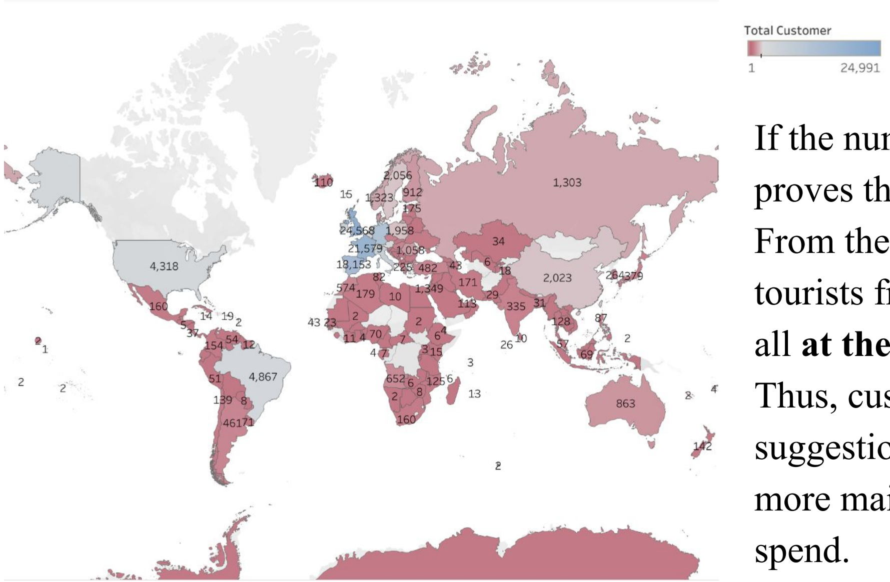
The last quarter covers China's National Day, when Chinese tourists are more inclined to scatter around the world in a free way.



Here are the two most representative quarters in the data feedback: Q1 and Q4

Visual Exploration

Map based on Count of Total Customer - Country



If the number of passengers is relatively large, it proves the significance of the study.

From the graph, we can see that the number of tourists from the previously mentioned countries are all **at the level of thousands**.

Thus, customers can provide customized suggestions for hotel service management to attract more main user groups (such as Chinese tourists) to spend.

Visual Exploration
Map based on Customer Type - Country - Quarter

Future progress:

✅ Quantitative research

For the relevant conclusions, such as Chinese passengers, other evidence is still needed to support, such as the number of passengers, age, to lay the foundation for the preliminary conclusions.

✅ Qualitative research

Continue to assist other objective conditions (draw other charts), continue to explore the causes of quarterly changes in tourist types in various countries, build tourist portraits, export effective services, reduce hotel costs, and stimulate customer consumption.