See a Need, Fill a Need:

An inside look at the hotel industry from an experience standpoint.

Introduction and Background

The hospitality industry is one of the leaders in groundbreaking analytics collections. Businesses want to know how to make their guests happy, and in return, the guest wants to know they are receiving quality accommodation. Understanding the dataset of hotel bookings from Kaggle (Mojtaba, 2021) can give a glimpse into how analytics can create formidable changes in company operations and increase profits along with success. Before the COVID-19 pandemic, it was cited by the International Trade Administration (ITA, 2022) that the travel industry generated 1.9 trillion dollars in economic output and supported 9.5 million American jobs. However, as the world continues to recover from the lack of travel post-2020, we must understand that the travel industry will return, and when it does, the customers' needs.

The data presented tells the story of two different hotels called "City Hotel" and "Resort Hotel" as well as the visitors, booking information, and the overall needs of their stay. The types of questions our team is attempting to answer are as follows. Which hotel is more favorable for visitors? What season or time of the year is more profitable for a hotel? What is the customer type that frequents a given hotel? All in effort to understand the hotel visitor demographic and their needs.

The article "Competing on Analytics" by Thomas Davenport (Davenport, 2022) provides an inside look at where the study of business analytics originated from. He further discusses that Marriott is one of the biggest players in the field right now, crediting their predictive models to understand a guest's needs before they occur. As students in the analytics program, it becomes our sole task to understand where the industry is going and provide new insights to expand it. Like the data Marriott collects, this dataset can help our group practice the same concepts that are key to finding a successful career after graduation.

Objectives and Goals

Our group would like to learn how to better dissect datasets that exist in real-world settings. (HBS, 2017). This trial teaches us how to try new things, be innovative in solving problems, and understand the world around us, which is the overall goal of our project as a team. Learning effective data collection methods and forming them into coherent ideas that can be used for the benefit of furthering business practice is our overall goal. The motivation for the analysis of the hotel booking data specifically is because of three converging factors: efficiency in resource allocation, better forecasting, and customer segmentation—all very critical areas in many fields that depend on bookings and reservations. Base sources talk about data's role in the optimization of operational planning, pricing strategies, and understanding customer behavior.

The target of the project is to help the recovering travel industry in a post-pandemic society to understand how we measure the needs of the guests. The insight provided is beneficial to increase the future success of the hotel industry and guides us to consider consumer experiences in future projects. A knowledge of booking patterns and trends of customer behavior is vital for any business that depends on reservations to plan and balance the demand for resources. It can give insights into strategic decisions, optimize resources, and generally enhance efficiency in bookings and customer interactions. This project could allow the stakeholders to make data-driven decisions for optimization of resource allocation, reduction of cancellations, and enhancement of customer satisfaction by understanding the booking behaviors and further calibrating their strategies. The visualizations are supposed to highlight insights such as booking volume trends, lead times, and segmentation by customer demographics. Expected outcomes include a clearer identification of booking patterns, customer segmentation, and an understanding of the drivers of cancellations. Impact: This project could provide operations teams, strategists, and analysts with very valuable insights to inform resource planning and demand forecasting. The findings may

drive the creation of strategies to reduce cancellations or efficiently manage peak booking times. This could be translated into success via metrics showing better resource utilization, lower cancellations, or better customer satisfaction.

Datasets

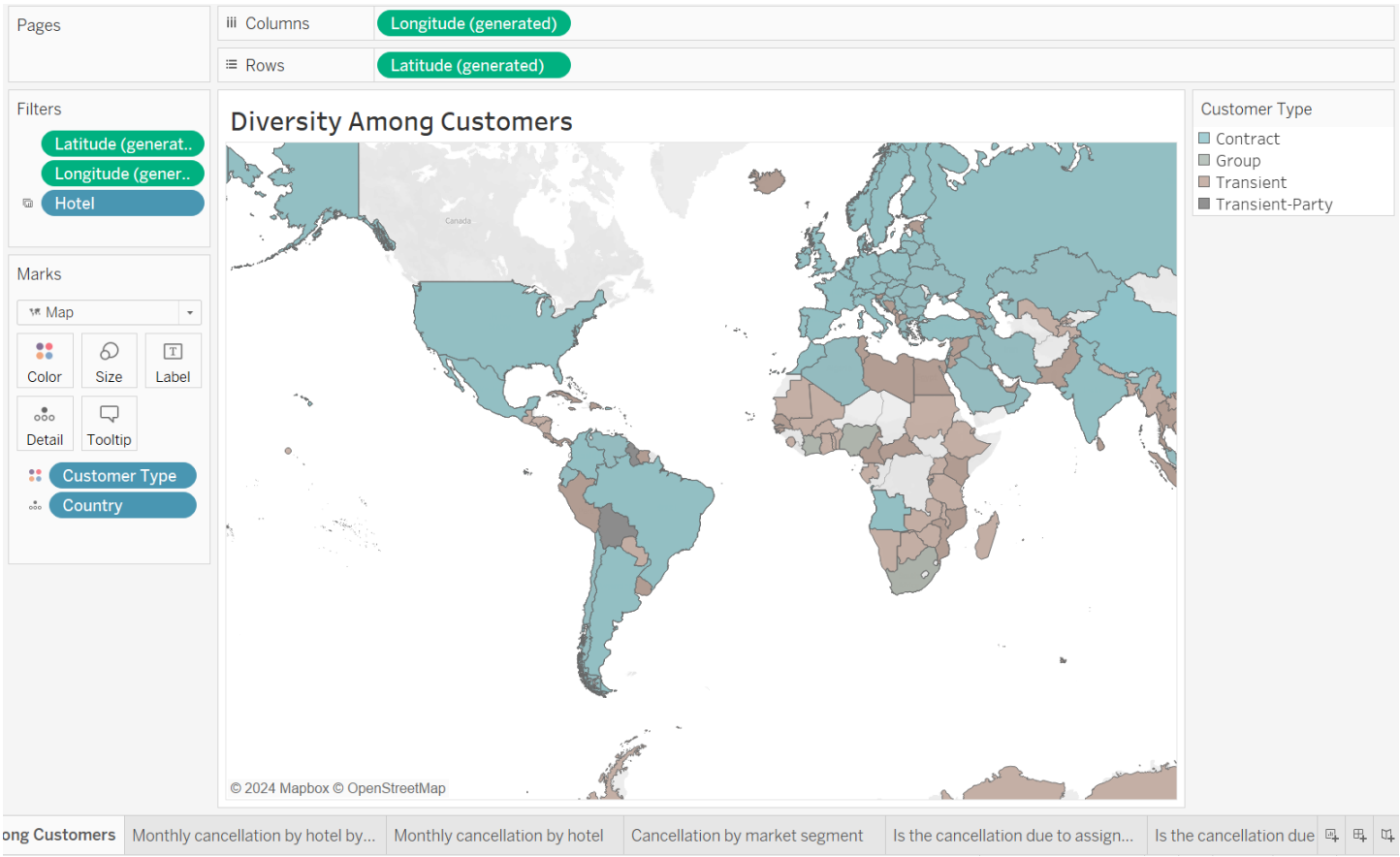
The data set chosen by our team contains 119390 observations for a City Hotel and a Resort Hotel. Each observation represents a hotel booking between the 1st of July 2015 and the 31st of August 2017, including bookings that effectively arrived and canceled. Since this is the hotel's real data, all data elements about the hotel or customer identification were deleted. The Four Columns, 'name', 'email', 'phone number', and 'credit card', have been artificially created and added to the dataset. The earlier document format was in CSV but was converted to Excel to be most effective for experimentation and discovery. The variables we are most interested in are based on understanding the customers most likely to visit the City Hotel and Resort Hotel. The data is originally from the article Hotel Booking Demand Datasets, written by Nuno Antonio, Ana Almeida, and Luis Nunes for Data in Brief, Volume 22, February 2019. However, there is very little information on the initial collection process provided to us by Kaggle.

The dataset chosen is broken down into 119390 rows and 36 columns from the years 2015 to 2017. The range of information includes the customer type, the customer's line of purchase, the need for parking, meal accommodation, children visiting, babies visiting, country, and various other areas of study to understand what their goals might be. The data included by country gives us an idea of the customer's origin and where the visitors are traveling from. Such data can tell a story in the visual format of mapping and create a recognizable form of a chart.

Visualization and Data Preprocessing Plan

Our team's desire to use this dataset is supported by the many fields of measurement that are contained in the document. As a group, we have tested the strength of these fields and created calculated fields to increase the effectiveness of showcasing the information. Our plan will be beneficial in supporting the final 10 charts needed to complete the final project. The following visualizations are a representation of what we would like to complete in the final project and a few of the questions we intend to ask.

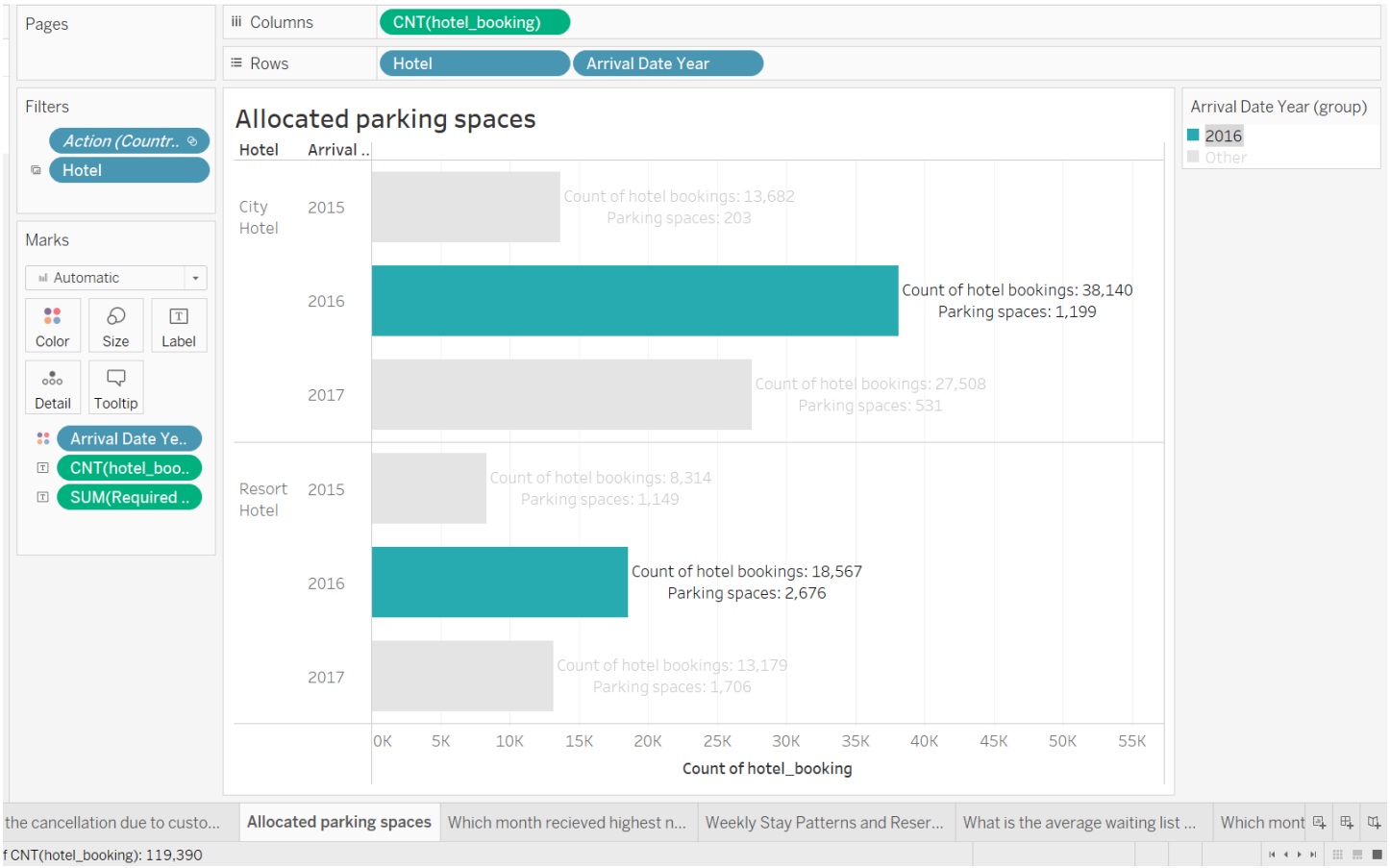
What are the diverse types of customer groups from all around the globe?



Insights:

To understand the diverse customer types around the globe we must consider taking geographical data available which is latitude and longitudinal coordinates, the information on the names of the countries that customers come from and types of customers.

How many parking spaces were allocated for each hotel?



Insights:

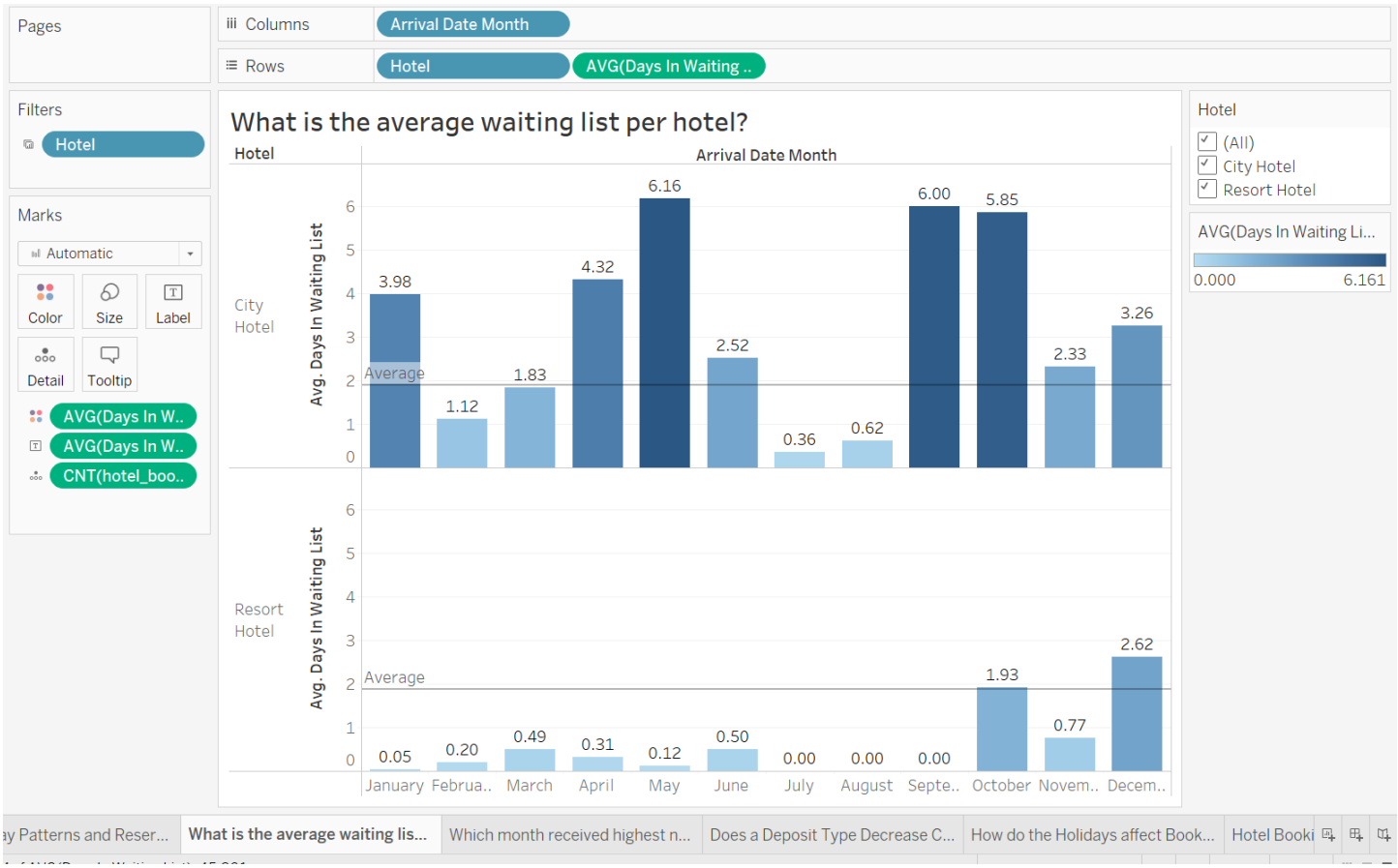
Improvements provided:

- Added the year of arrival for each hotel.
- Used labels.
- Provided grouping which will highlight the highest number of bookings with their allocated parking spaces with respect to year.

The above visualization draws an inference based on the number of parking spaces allocated for each hotel with respect to year. It can be observed that the year 2016 had an excellent reservation booking for each hotel. Of the above two hotels the city hotel had the highest booking, but their parking spaces were very less.

This could be because the guests did not require parking spaces, or they could be international guests. There can be a possibility that the guests would come from far locations where they did not require parking spaces.

What is the average waiting list per hotel?



Insights:

This plot draws an insight into average waiting days for each hotel. This information gives a deeper understanding of the seasonal period for each hotel.

The average waiting days for city hotels are very high in May, September and October. This information shows that these months there might be a very good number of bookings with depicts that the best seasonal month/period for the city hotel is May, September and October. Hence, it gives a clear indication to book the reservation prior.

The Resort hotel has on average 3 days of waiting in the month of December. This suggests that the people should be prompt in making the reservations on time to visit the hotel in the month of December.

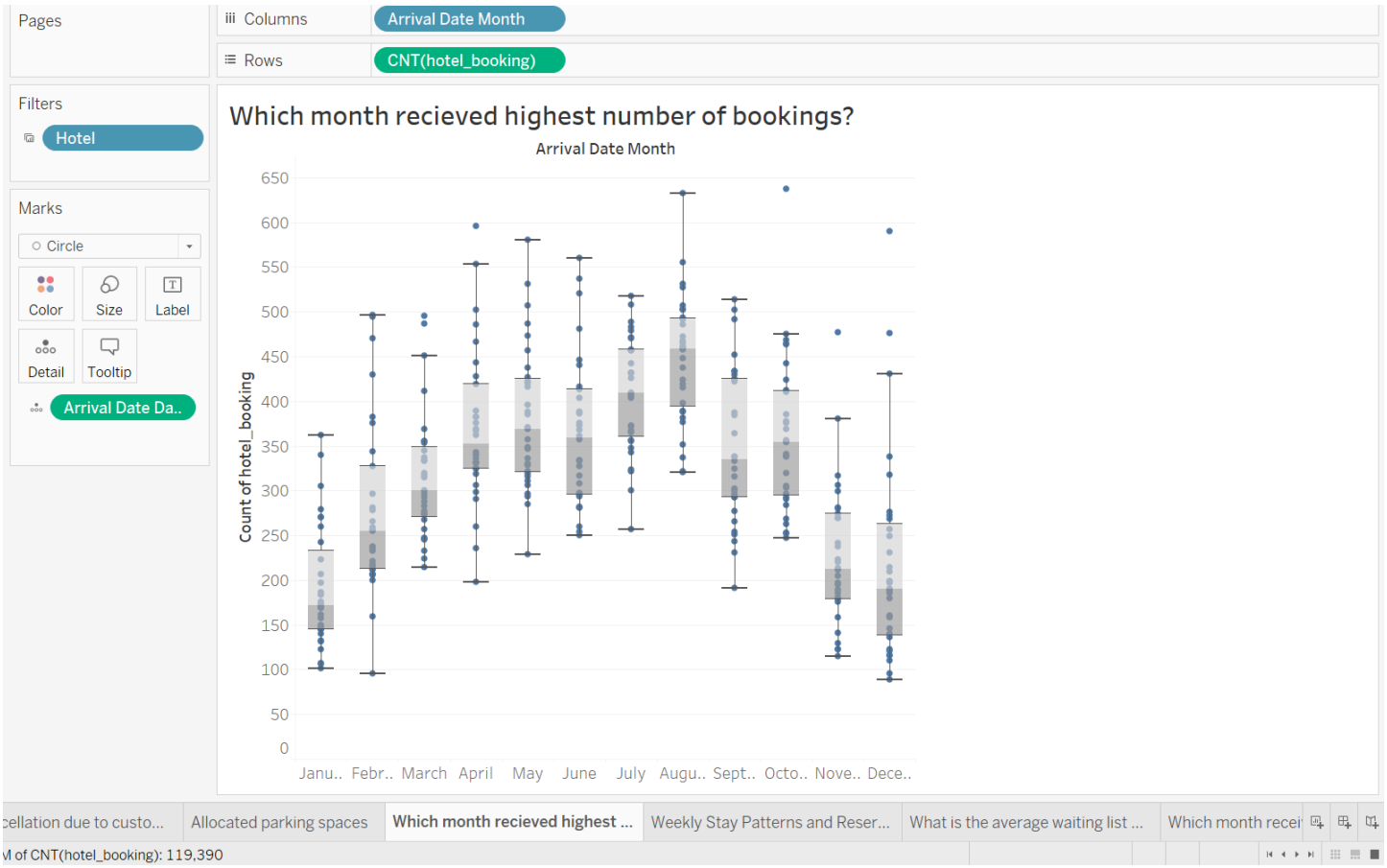
Which month received the highest number of special requests?

Insights:

The Line chart is used to get information on the number of special requests for the City Hotel.

The above plot gives an insight into which month the highest number of special requests for the City Hotel received. From the plot it is very clear that the month of August received the highest number of special requests. These special requests for bookings might be made based on seasonality. Around 6K number of special requests were made in the month of August. This is annotated using a green color. The minimum special requests are around 2K in the month of January and annotated in orange color.

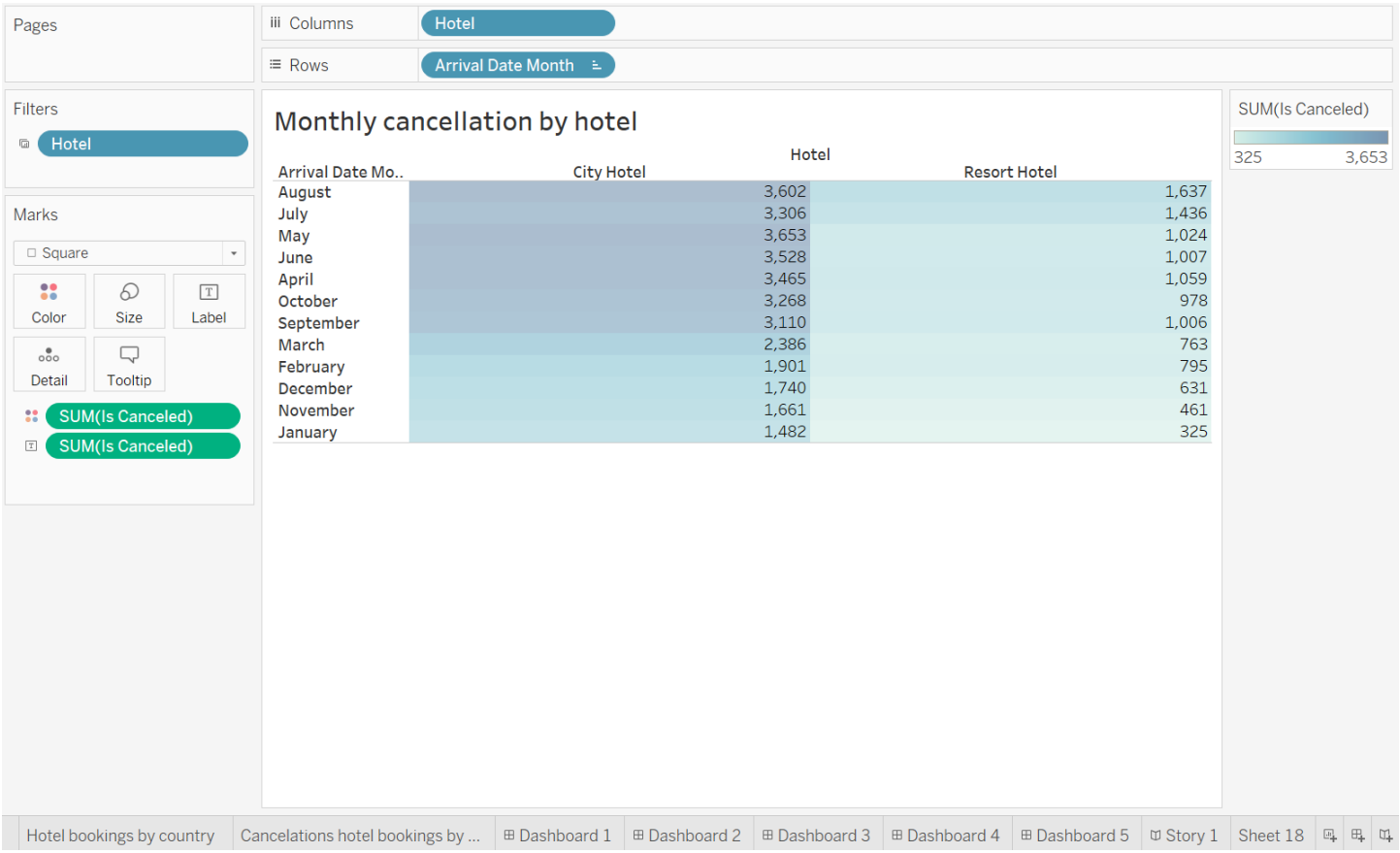
Which month received the highest number of bookings?



Insights:

The highest number of bookings were received in the month of October. A box plot helps to understand the total highest hotel bookings for the corresponding month. From the plot it is very clear that July and September months had low bookings whereas months ranging from October to December (winter months) were the busiest ones. Outliers depict the occasionality. The occasional spikes in demand were observed during the peak months. This plot clearly helps customers to book their tickets or make their reservations accordingly.

What is the monthly trend for cancellations for the City Hotel and Resort Hotel?



Insights:

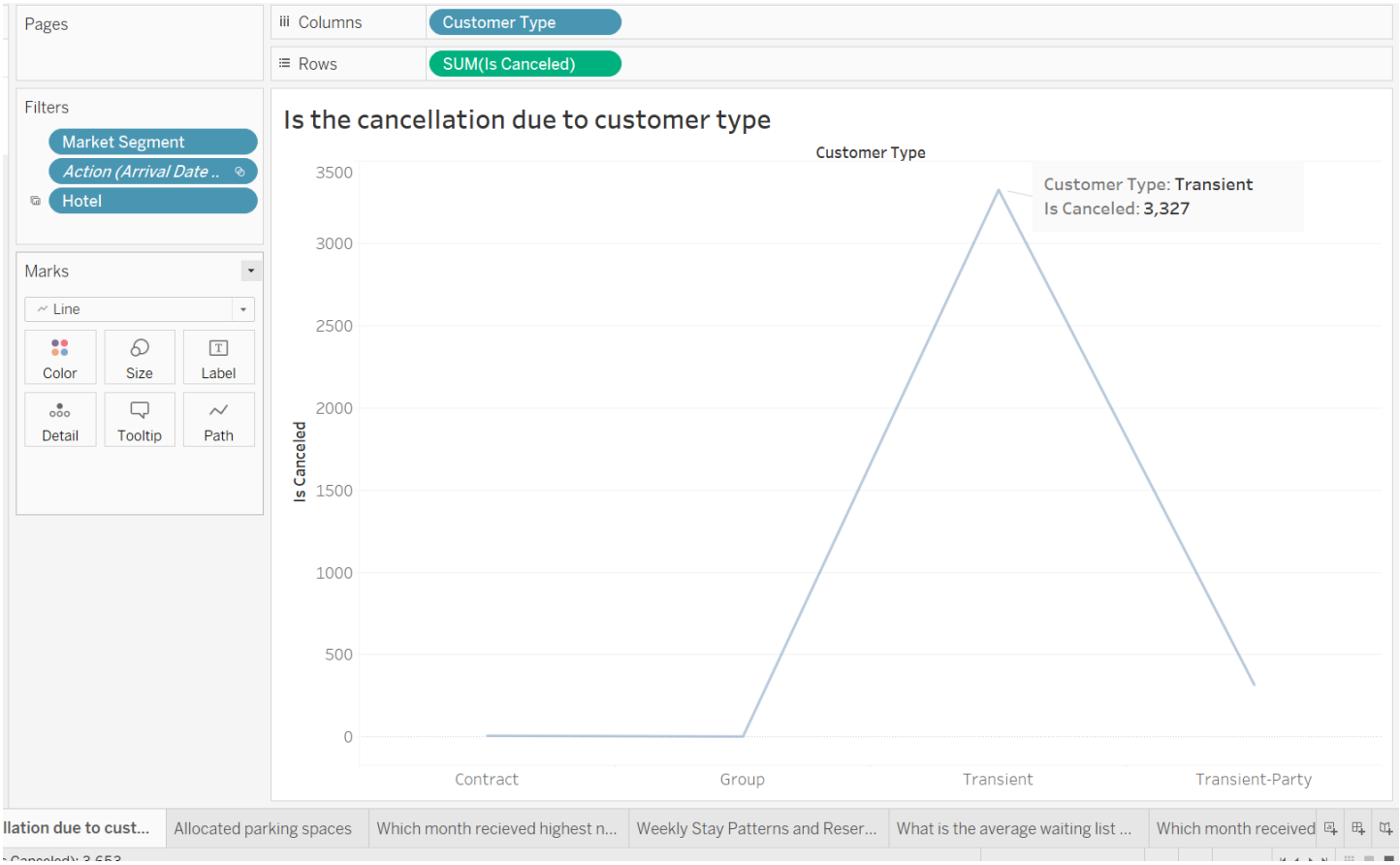
Improvements provided:

Including the customer type, to shed light on more aspects. It would allow users to focus their attention and make determinations based on which methods of booking or type of customers are likely to cancel bookings. Given a title to it. Sorted it by descending order.

The above graph helps to draw the information based on the cancelations for each hotel. This table helps to get the statistics. The numbers suggest that the highest cancelations are made in the month of May for City Hotel and August for Resort Hotel.

Few factors are taken into consideration to understand the reason of their cancelation in the bookings. These are shown below.

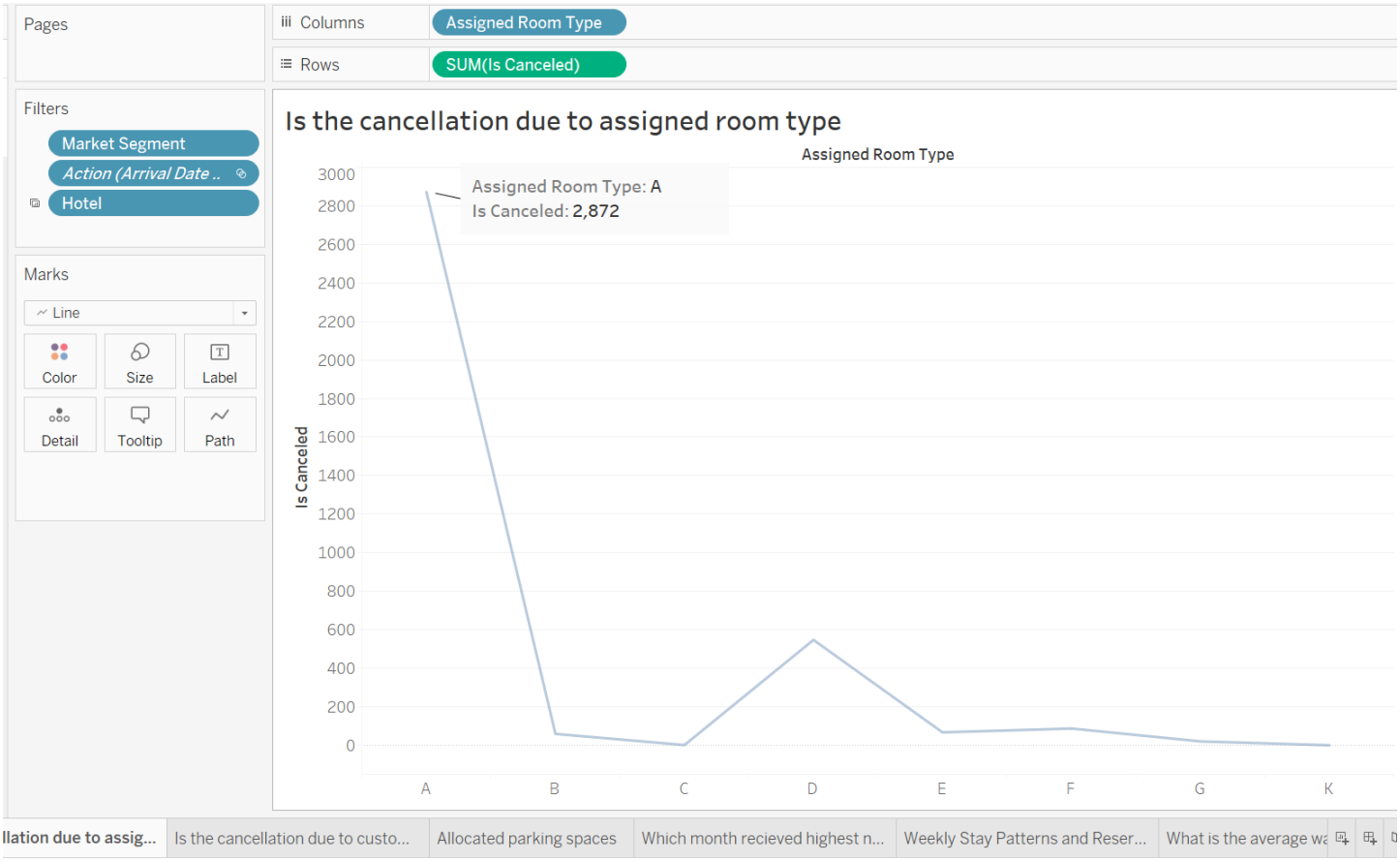
Is the cancelation due to customer type?



Insights:

The above line graph helps to draw an insight into the customer type for each hotel. This will help to draw an insight into whether the cancelations were based on the customer type or which customer type faced the maximum number of cancelations. From the graph it is very evident that the customer type of Transient faced the highest number of cancelations for both the hotels. This is shown with the help of a marking annotation to point out the highest number of cancellations for the customer type.

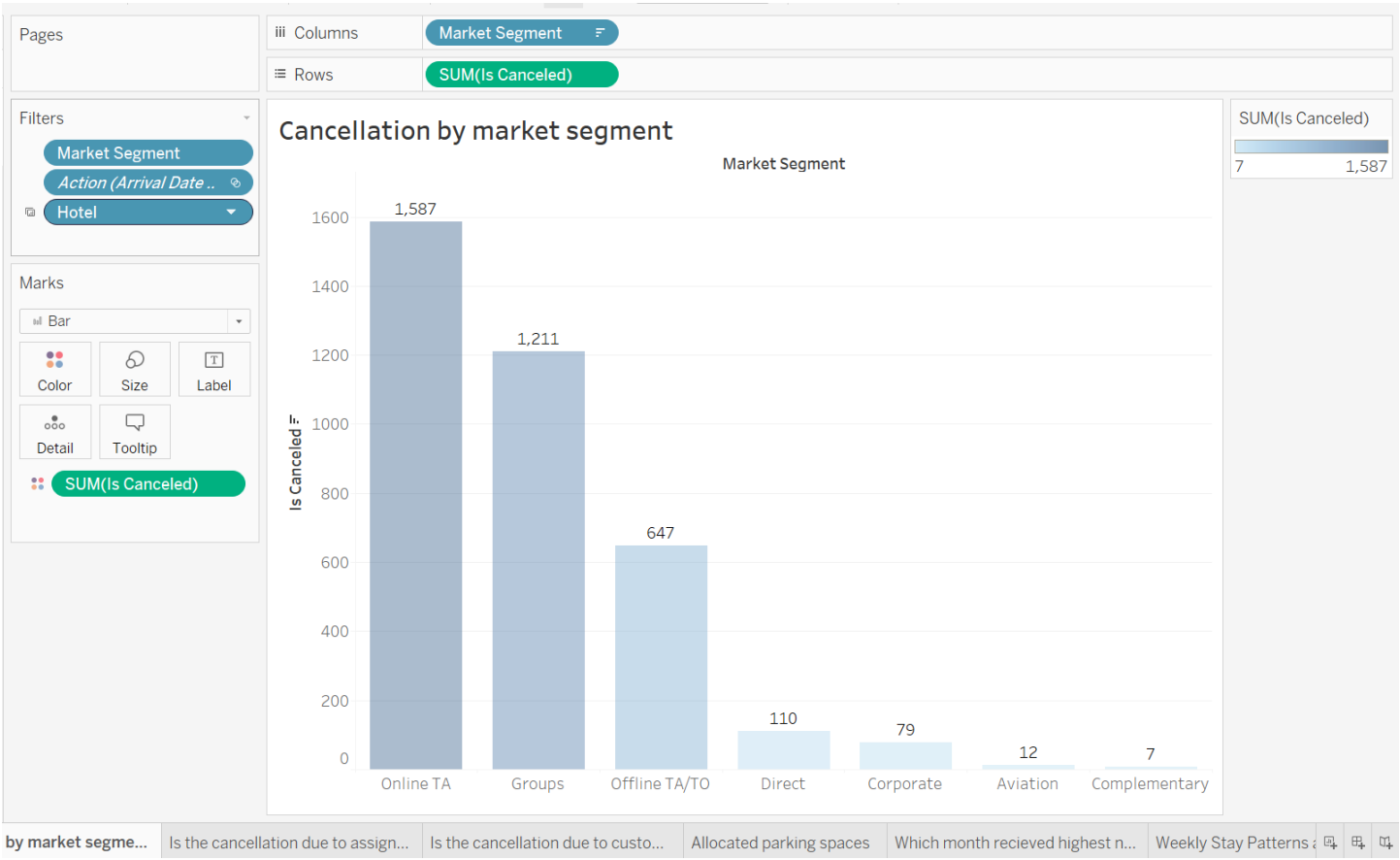
Were the cancelations based on assigned room type?



Insights:

The above line graph points out the cancelations based on the assigned room type. This graph helps us to understand and get a deeper understanding of the cancelation's reasons. From the graph, guests with an assigned room type A faced many cancelations. This could be the reason to understand that yes, the cancelations can be because of assigned room type for both the hotels. Room type A is shown using a mark annotation.

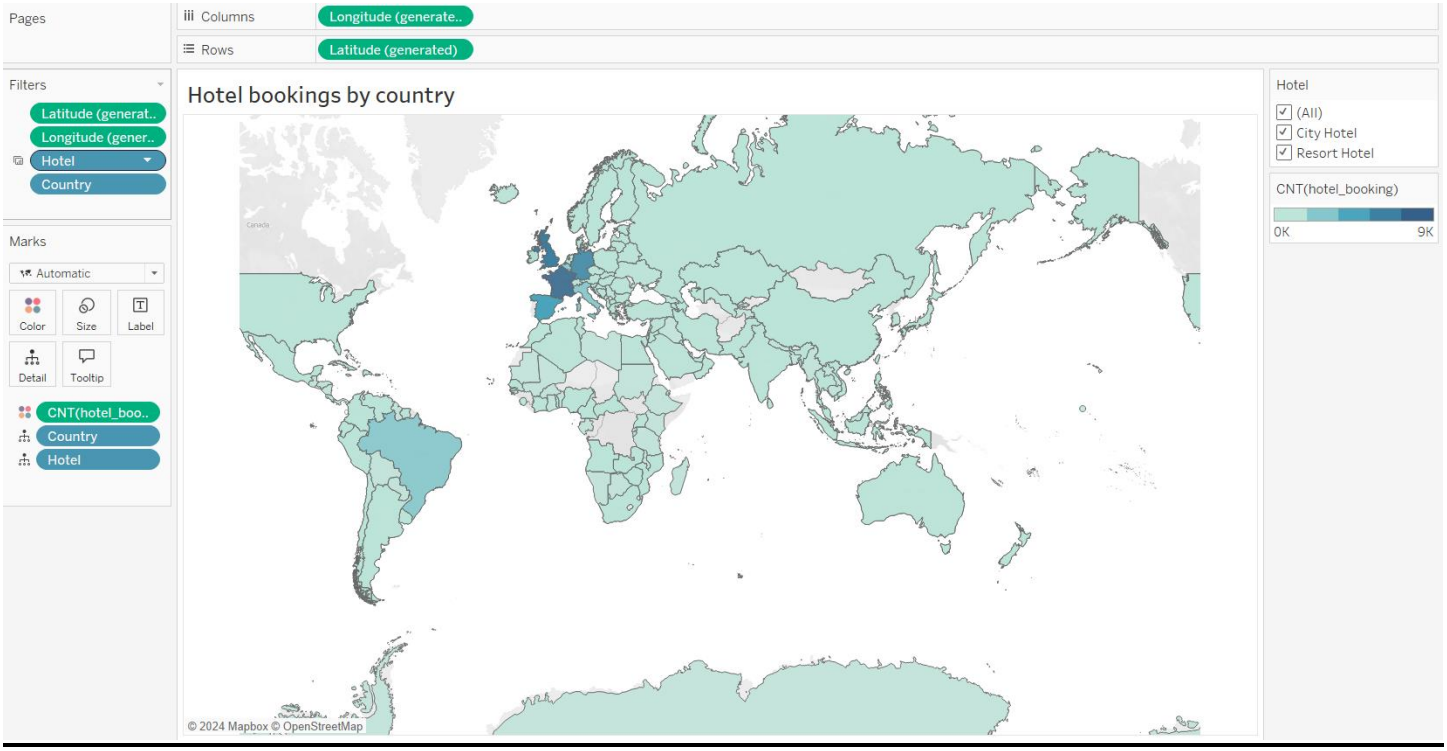
Which market segment received the highest number of cancelations?



Insights:

The above bar graph helps to draw up the information regarding the cancelations for each market segment. This helps us to understand which market segment faced the highest cancelations for each hotel. The Online TA market segment faced the highest cancellations as compared to others for each hotel.

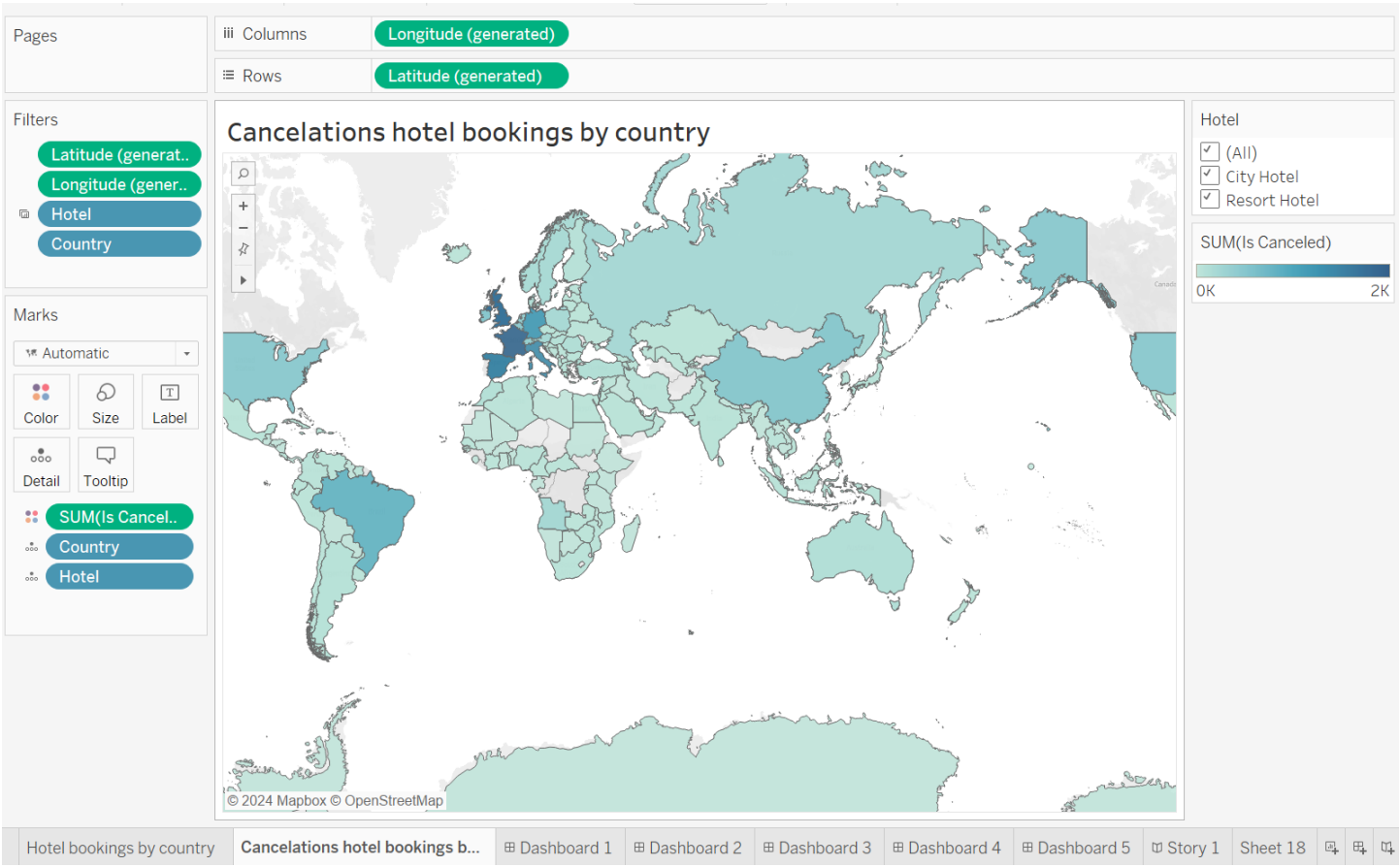
Which country has the highest hotel bookings geographically?



Insights:

The above map helps to draw the total count of hotel bookings that happened geographically. This map suggests that the Europe region has the highest number of hotel bookings. For City hotel, around 9K bookings were recorded in France. Whereas, Great Britian has the highest number of hotel bookings of around 7K.

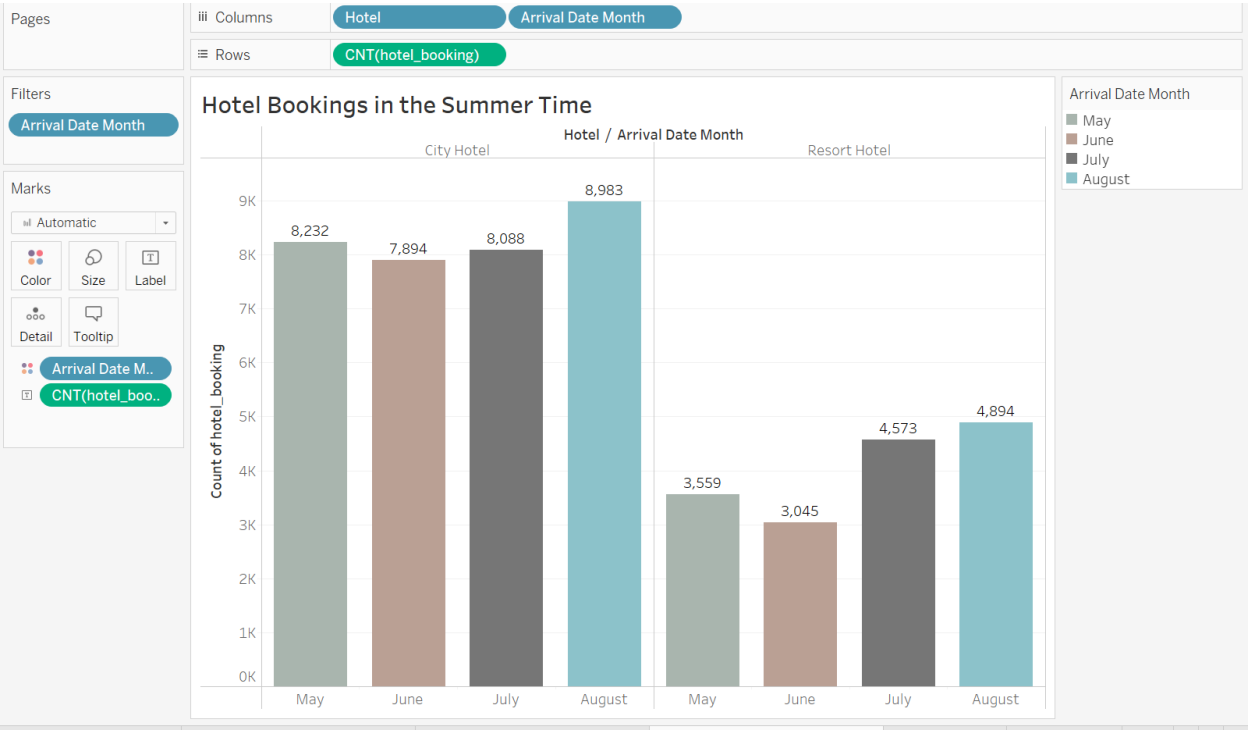
Cancelations in hotel bookings geographically.



Insights:

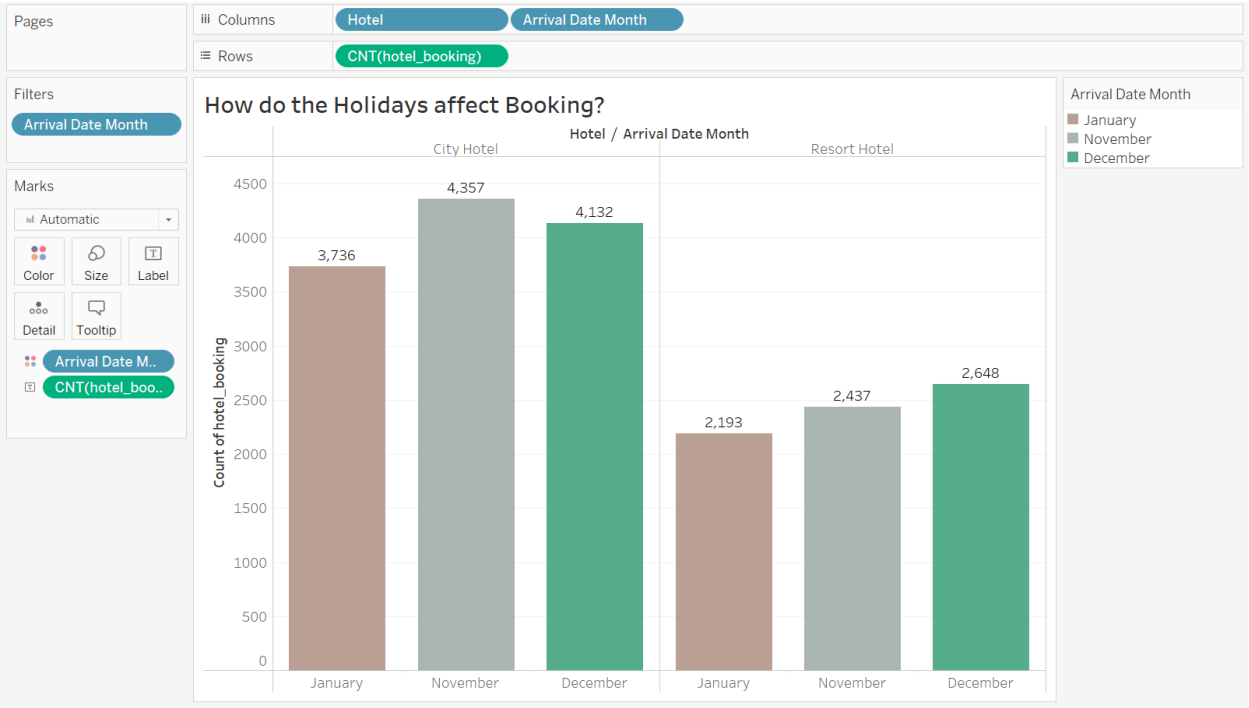
The above map gives a cancelation of hotel bookings for each hotel by country. The country is the only geographic factor in the dataset hence, exploring with respect to countries helps to understand that most of the Europe region faced the highest number of cancelations in bookings. France faced the highest number of cancelations globally. For City hotel, around 2K cancelations happened in France. For Resort hotel, around 1K cancellations happened in Great Britian and Spain.

Does the summertime have any impact on the number of bookings a hotel receives?



Insights: The purpose of this chart is to compare the number of bookings between the two hotels during the summertime. Based on the chart city hotels see a significant peak in booking as opposed to resort hotels. However, both hotels are at their busiest times in August.

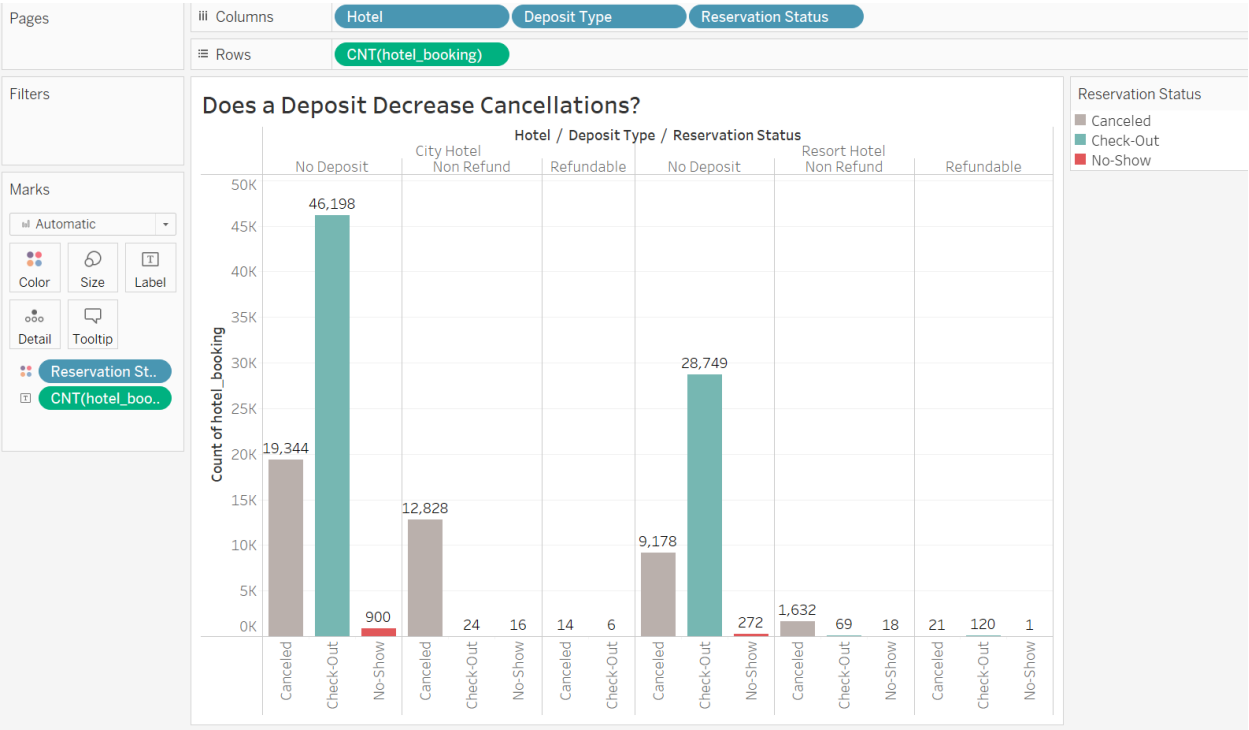
How can the holidays affect the type of hotel being booked as well as the number of bookings?



Insights:

This bar chart shows how holidays affect booking. For city hotels they're the busiest in November as opposed to resorts hotel who see a peak in reservation during December. Both hotels have the least number of bookings during January when the holidays are coming to an end.

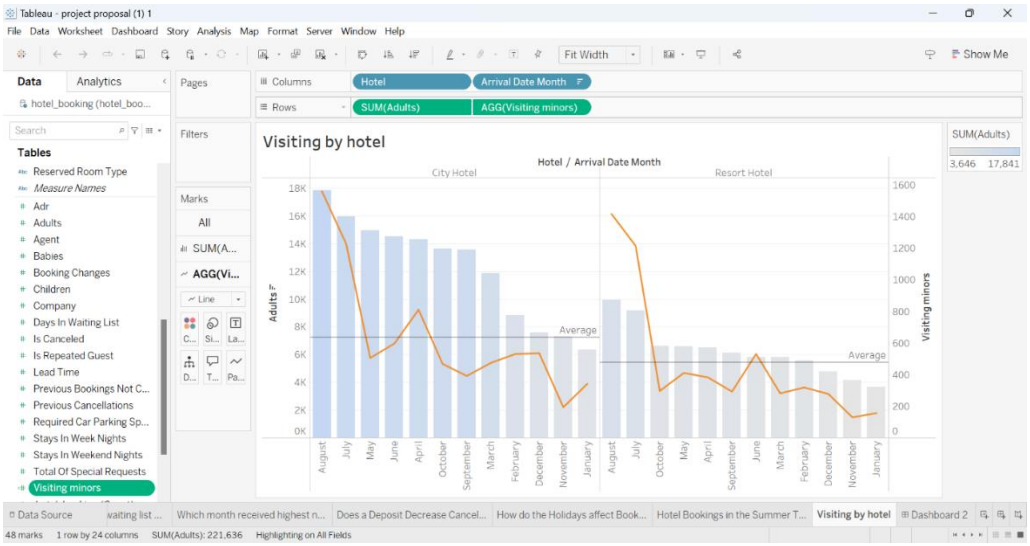
How can a requiring a deposit affect the odds of customers cancelling their reservations as opposed to checking in?



Insights:

The bar chart shows how deposits can affect the number of cancellations. Hotels that have no deposit have a higher cancellation as opposed to those that do. But another thing to consider is, despite the cancellation being higher for hotels without deposits, so is their checkout.

Which months see the highest number of adult and minor visitors in city and resort hotels, and how does the visits pattern vary between the two guest groups over the year?



Insights:

Monthly Peaks in Adult Visitation:

City Hotel: During the month of August, the adult visitor peaks and reaches as many as 17,841 visitors. Without a doubt, summer holidays are one of the causes for this peak, increasing the solo or family travelers.

Resort Hotel: The peak month in Resort Hotels is August, although at a far lower rate of adult visitation than that of City Hotels, with 1,416 visiting minors. This further suggests that fewer visiting minors, overall, visit Resort Hotels.

Stable Seasonality Patterns:

From the month of August onwards, the graph for visitors in both City and Resort hotels for adults shows a steep decline, slowly decreasing from September onwards. The visitation is lowest during Christmas month and during the month of January. This might be the season when demand is at its lowest because of winter or holiday reasons associated with these months.

Minor Visitation Trends:

The maximum visitation of minors to City Hotels is during the month of August, with 1560 minors. Thus, it can be observed that families travel more during summer months. While Resort Hotels record a minor peak in August with about 1,416 minors, this is still way lower compared to City Hotels, which evidence that family tourists prefer urban destinations or that family-friendly amenities are better in the case of City Hotels.

Notable Differences Between City and Resort Hotels:

City Hotels consistently attract higher numbers of adults and minors than Resort Hotels, which signifies that they are more popular year-round.

Resort Hotels are more seasonal in their appeal and attract fewer people, and this may be due to location, amenities, or the types of activities available.

Implications for Marketing and Promotions:

City Hotels can leverage their popularity and offer off-season discounts or incentives from September to January to sustain visitor numbers.

Resort Hotels can offer focused family promotions or seasonal events during the summer period to enhance minor and family visitation.

Both types of hotels can take advantage of the quieter winter months by developing holiday-themed packages or discount packages to attract more visitors.

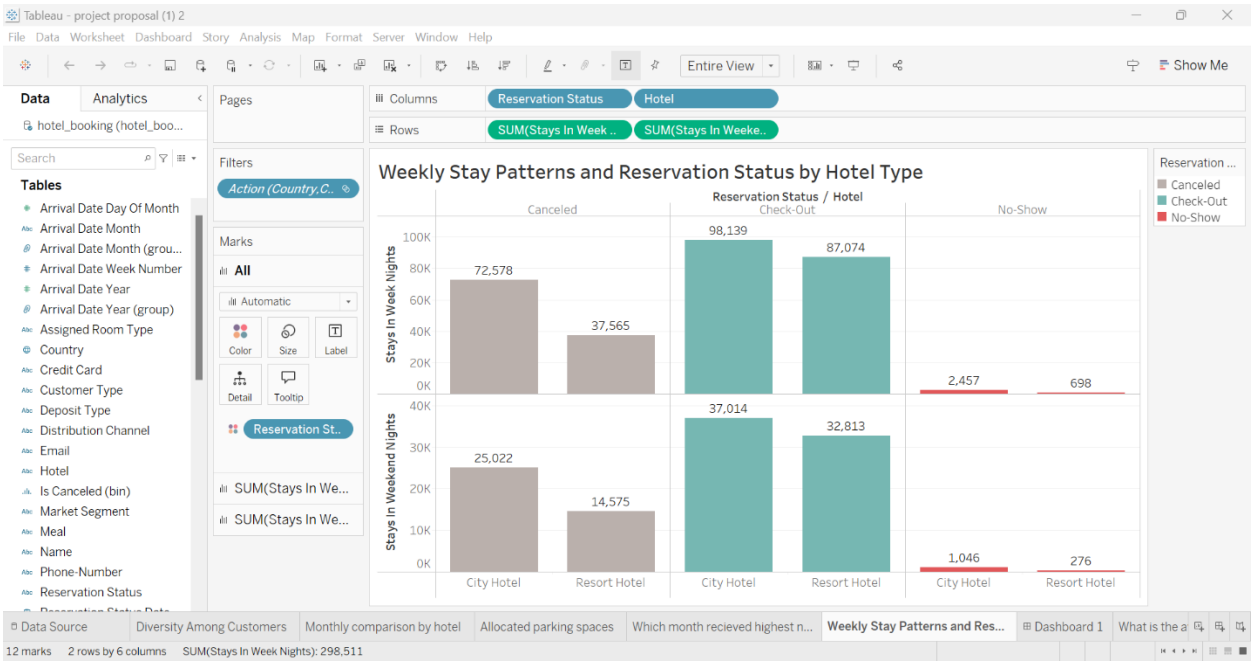
Visitation of Minors Proportion:

The proportion of minors, overall, during the year is much lower than that of adults, which indicates that most bookings are either singles or adults. However, this summer peak in minor visitation would also signal that both types of hotels could provide more family packages or activities in summer.

These insights, furthermore, help outline seasonal trends and visitation demographics that may inform possible marketing strategies and operational changes to maximize occupancy throughout the year. Every month more adults and minors (babies + children) visit City hotel more than the Resort hotel, it can be because that in City hotel there might be more activities for both adults and minors or many famous places for activities might be near it like skiing, waterpark. So, to have more business, make your hotel near places like theirs.

What is the distribution of stays in weeknights and weekend nights for different reservation statuses across both the hotels?

To visualize it we need, Reservation status, hotel in **Columns** & Stays In weeknights and weekend nights in rows also Reservation Status in color.



Insights:

Check-out Status Dominates:

Most of the stays, in both City Hotels and Resort Hotels, are in “Check-Out" status; that is, most reservations ended up translating into actual stays.

During weeknights, the number of checked-out stays is significantly higher for City Hotels 98,139 compared to Resort Hotels 87,074. The same trend continues for weekends, albeit with a close margin: City Hotel – 37,014; Resort Hotel – 32,813.

High Cancellation Rates:

Cancellations are the second-highest statuses for both types of hotels.

Weekday cancellations are dramatically more significant in City Hotels, 72,578, than Resort Hotels, at 37,565. This may reflect the high demand and transition rate during the weekdays of the city hotels.

At weekends, cancellations are lower but still follow a similar pattern: 25,022 in the case of the City Hotel, 14,575 with the Resort Hotel.

Limited number of No-Show cases

The "No-Show" cases are one of the least frequent cases in the same two types of hotels and at various times of the week.

No shows are relatively higher in City Hotels' case, which stand at 2,457, compared to Resort Hotels, which stand at 698.

No-shows during the weekend for both types are quite similar and low: 1,046 in the case of a City Hotel and 276 for a Resort Hotel. It would thus appear that guests tend to respect weekend bookings with greater consistency.

Weekday Stay preferencing in City Hotels:

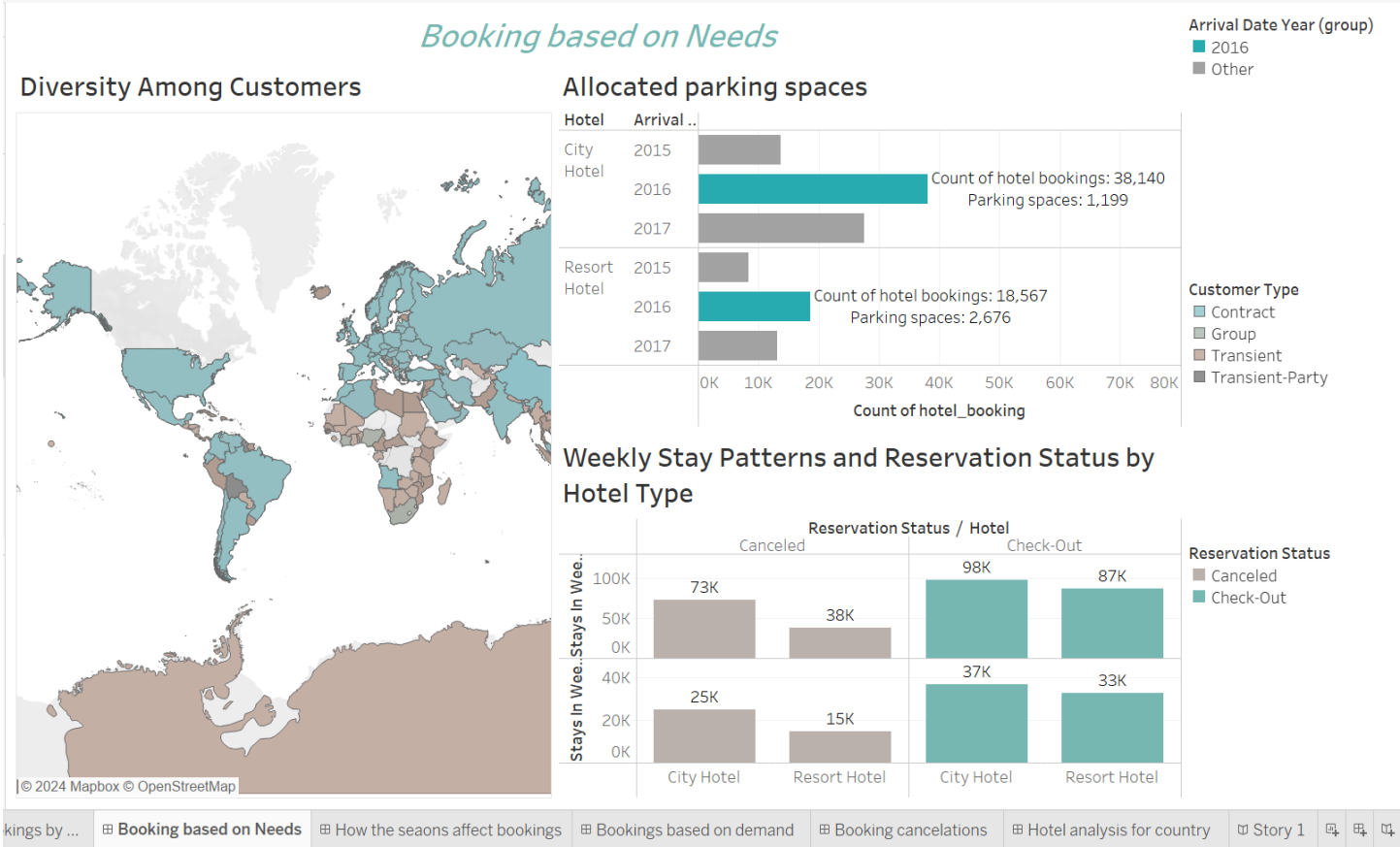
The check-outs and cancellations recorded on weekdays being really higher as compared to weekends, city hotels seem to be preferred for a stay during the week.
This could be because of business travelers or customers who only come during weekends.

Demand Patterns for Resort Hotels Are Consistent:
Occupancy during the week varies significantly less in Resort Hotels as compared to City Hotels, which suggests that demand for Resort Hotels is more evenly distributed across the week.
This may indicate that resorts attract leisure customers whose length of stay is not necessarily restricted to weekends.

Total Stay Nights across All Hotels:
Adding up all statuses and hotel types, the whole dataset amounts to approximately 298,511 nights in total, which gives an idea of the scale of this data.
From this analysis, it appears that the City Hotels seem to be more affected by business travelers who are not showing up on weekdays, while Resort Hotels are still somewhat stable. In both cases of hotels, there might be strategies to control and minimize cancellations and no-shows, probably by adjusting policies or confirmation of bookings.
Half the amount of No-show has been noticed for city hotels in weekend nights which is 1,046 for City hotel and 2,457 in case of weeknights which shows that as people are working in weekdays so due to any urgent work, so people didn’t show up at the hotel for weeknights.
Almost a little more than double are the check-outs for resort hotel from the cancelled bookings in Weekend and weeknights whereas almost 15k to 20k the check-outs are more than cancelled for city hotel in weeknights and weekend nights.

Dashboard insights:

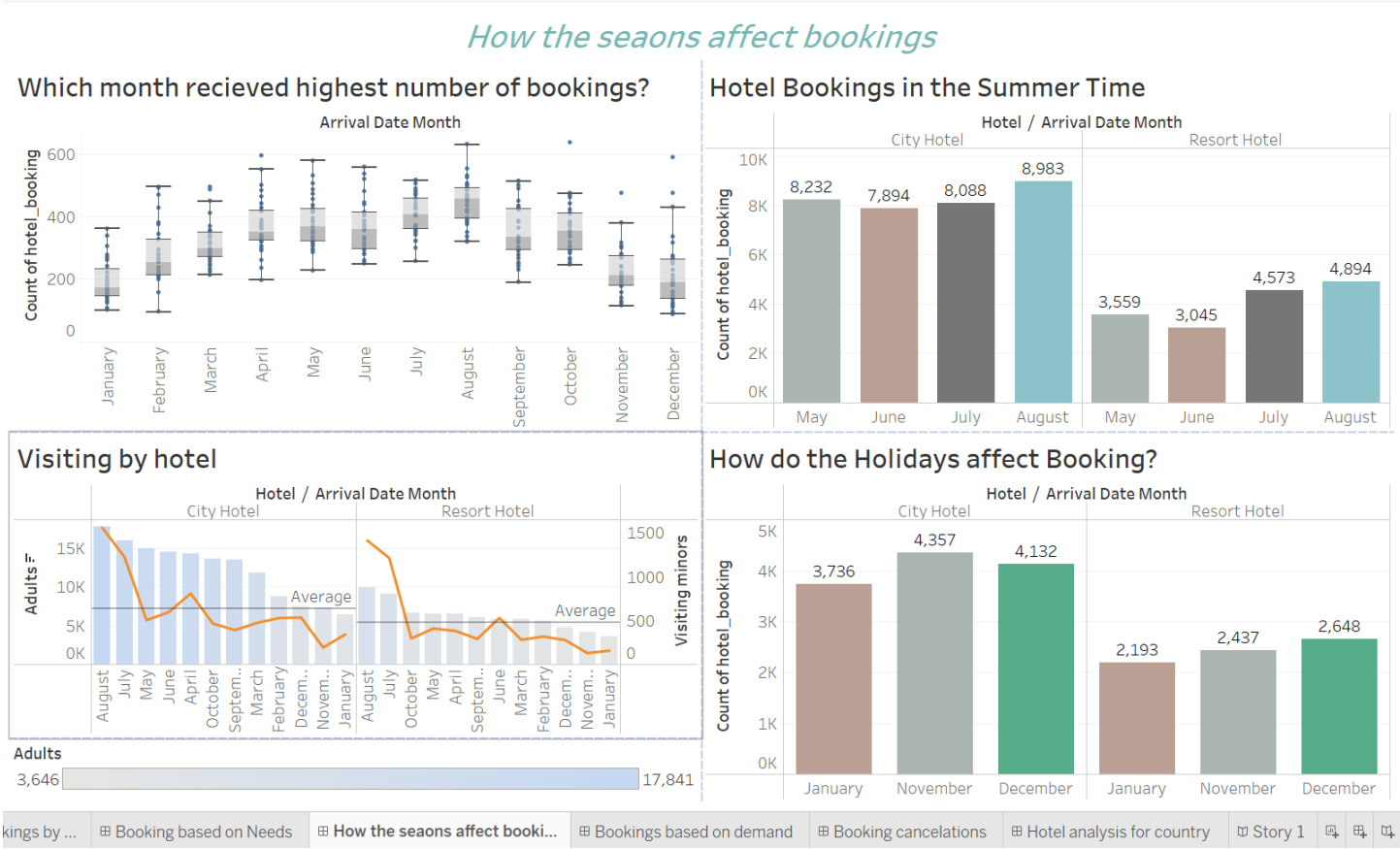
Dashboard 1: Bookings based on needs



Insights:

This dashboard was created to represent the type of customers that are booking each hotel and what factors they need to consider before booking. For example, the number of parking spaces available is crucial depending on the purpose of the booking as well as what day of the week each party needs to reserve.

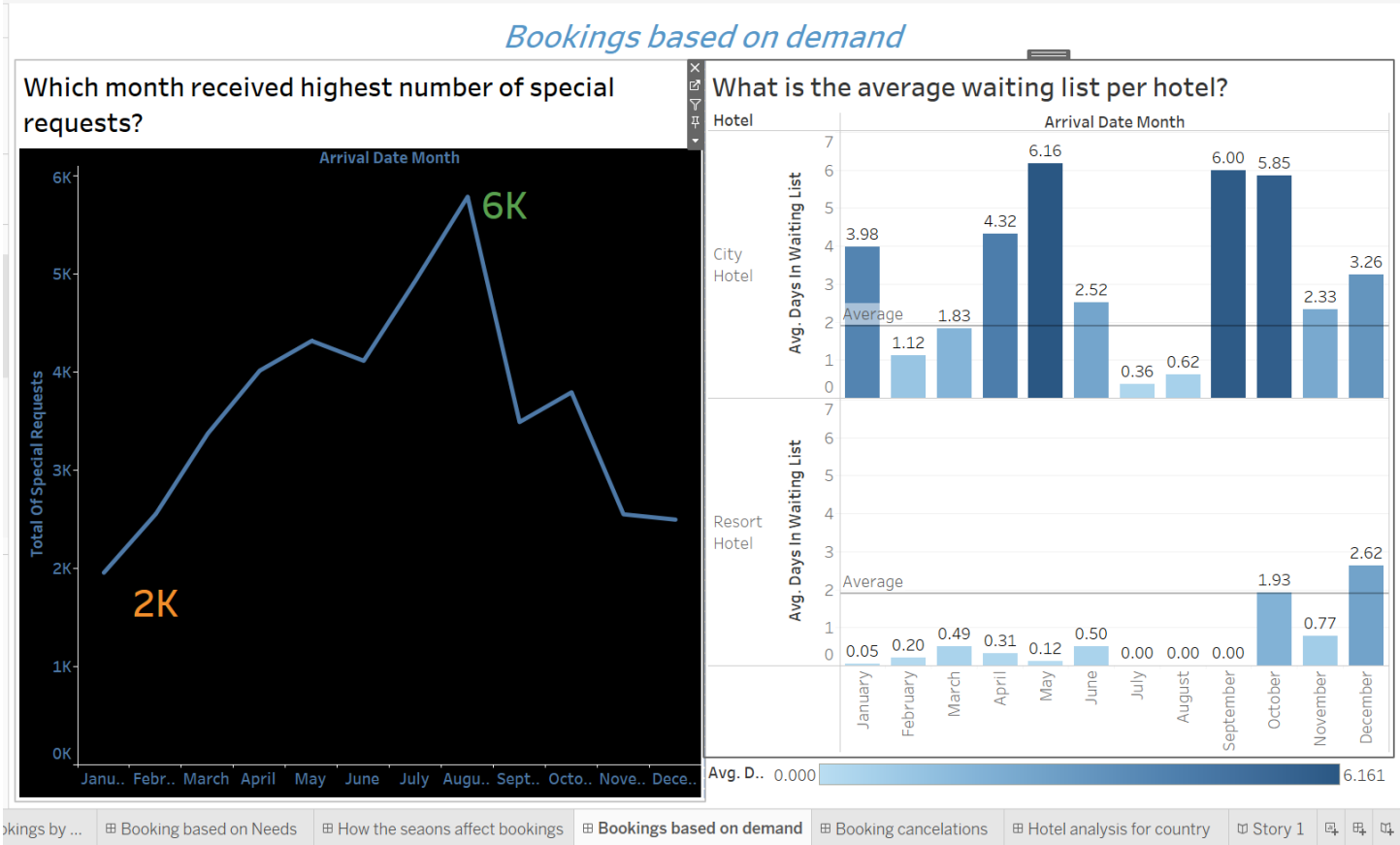
Dashboard 2: How the seasons affect bookings?



Insight:

This dashboard is about how the seasons can affect booking. The summertime has a significantly higher number of bookings as opposed to the holidays season. However, during the holidays December is the peak season and for the summertime August is the peak booking season. Based on the other two visuals August is the busiest month, and summertime is the busiest season overall for both City and Resort Hotels.

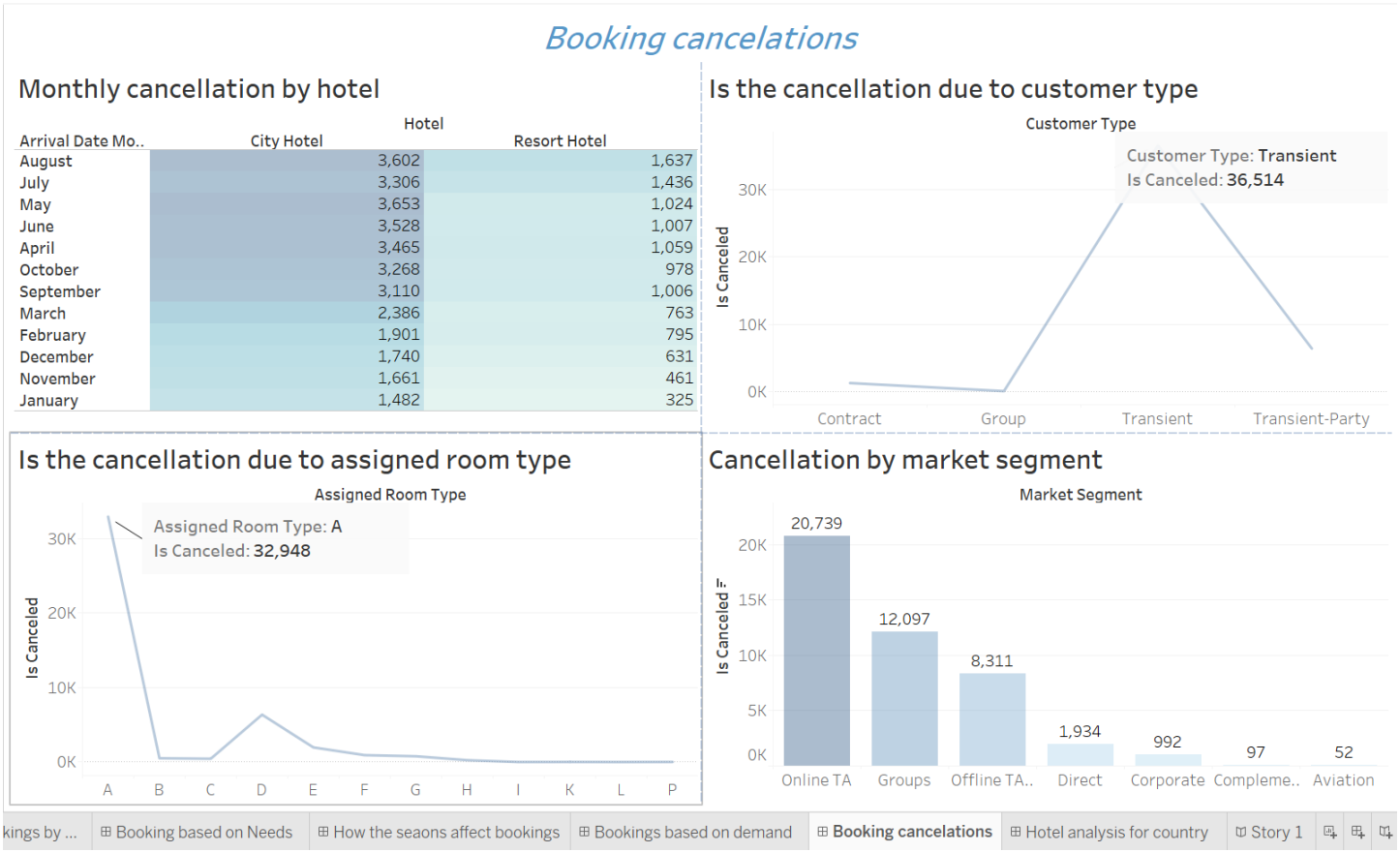
Dashboard 3: Bookings based on demands.



Insights: The peak number of special requests is 6,000 in August. Special requests increase gradually from January and reach their peak in August, then decline towards the end of the year.

For the city hotel May month is experienced the highest waiting list. This suggests that most of the people visit the hotel in the summer. This helps us to gain managerial insights, stating that May month usually experiences a high demand as it is during the summer.

Dashboard 4: Booking cancelations.



Insights:

Seasonality: The high cancellations in peak travel months, such as August and July, signal increased demand and a degree of customer indecision.

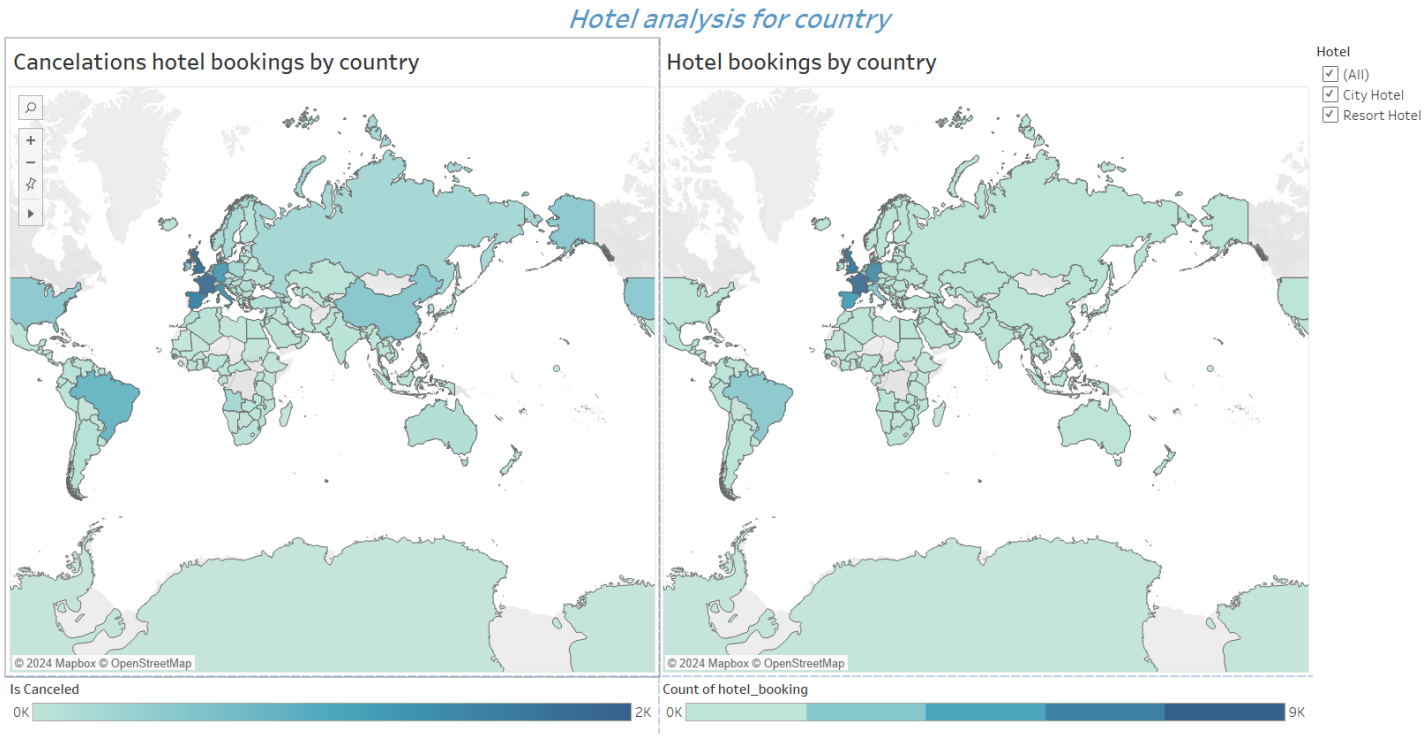
Customer Profile: Transient travelers are the major contributors to cancellations, thus requiring a retention strategy.

Room Type: Room Type A faces disproportionately high cancellations, signaling potential dissatisfaction or overbooking.

Channel Dynamics: From a market perspective, the cancellation landscape is dominated by Online TAs, who contribute more than 20,000 cancellations. This goes to say that while OTAs drive a lot of bookings, they also have a higher rate of volatility. On the other hand, direct bookings and Aviations segments are far fewer in cancellations, which underlines the stability of these customer bases.

Used an action filter on monthly cancelation by hotel to highlight cancelations for each hotel.

Dashboard 5: Hotel analysis geographically.

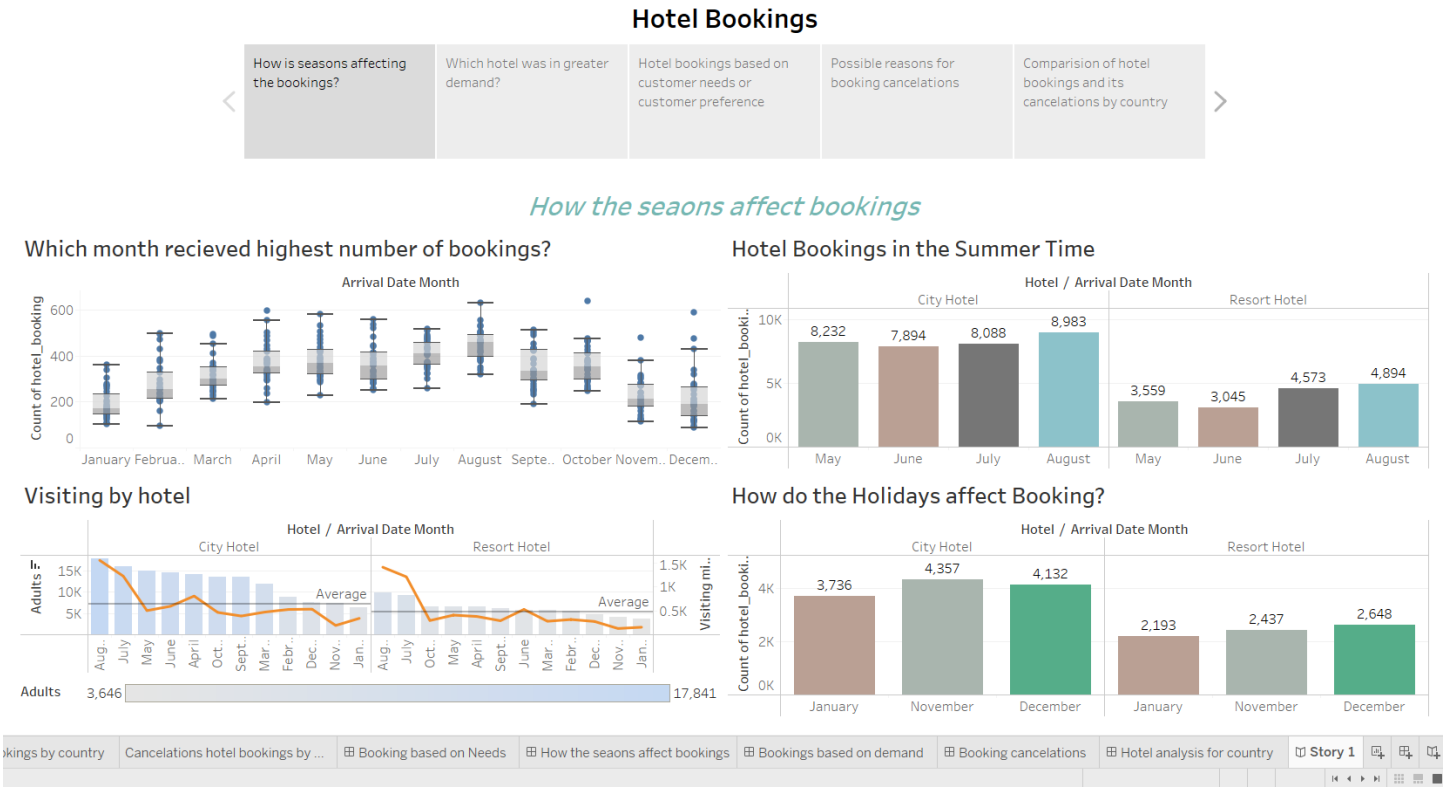


Insights:

This dashboard visualizes the country breakdown of hotel bookings and cancellations. Europe has the highest in both bookings and cancellations, while South America shows high bookings with fewer cancellations, indicating more stable travel behavior. The filters "City Hotel" and "Resort Hotel" indicate that trends may differ by hotel type, with city hotels likely to face higher cancellations due to business travel uncertainties. Highly cancelled regions could consider policies such as non-refundable bookings or customer incentives to increase retention. On the other hand, countries with consistent bookings and low cancellations represent reliable markets, ideal for loyalty programs. Insights enable targeted strategies to reduce cancellations and enhance profitability.

Limitations: The dataset did not have a city or a zip code feature to dive deep into the region as to understand why Europe region were having the highest hotel bookings and maximum number of cancelations.

Story:



A story gives us overall insights from the data and the visualizations created, it is a way of putting the things and conveying the population with the relevant information that will help them to plan their reservations for a specific type of hotel.

This story has a total of 5 views, each view consists of one dashboard, which displays information regarding seasonality, cancellations, geographic involvement of tourists and based on people's needs.

Conclusion

Visualizations brought forth a whole lot of insight into booking trends, the top periods of the year for booking, lead times of bookings, and customer segmentation based on demographics or booking behavior. All these patterns helped in understanding the volume of bookings, cancellation rates, and customer preferences. The project largely achieved its results by answering the initial research questions. Visualization serves core business issues, for instance, resource allocation during peak times and cancellation trends. For example, specific segments with a higher cancellation rate were identified; hence, strategies to reduce cancellation in those segments can be made. One unexpected finding was the strong influence of booking lead time on cancellation rates, suggesting that last-minute bookings were more prone to cancellations. A challenge encountered was handling missing data for certain customer details, which was resolved by imputing averages or excluding irrelevant records to maintain data integrity. These findings will be relevant to strategic decision-making by identifying actionable areas, such as adjusting the price levels or availability to balance demand during peak times or targeting specific customer segments with tailor-made offers. Therefore, it is recommended to include external factors like seasonality or promotions in analysis for further refinement in understanding and improvement in predictability of booking behavior.

References:

Mojtaba. (2021, June 29). *Hotel booking*. Kaggle. https://www.kaggle.com/datasets/mojtaba142/hotel-booking?select=hotel_booking.csv

Travel & Tourism Industry. International Trade Administration | Trade.gov. (n.d.). https://www.trade.gov/travel-tourism-industry#:~:text=The%20U.S.%20travel%20and%20tourism%20industry%20generated%20%241.9%20trillion%20in,for%202.9%25%20of%20U.S.%20GDP.

Davenport, T. (2024, February 16). *Competing on analytics*. Harvard Business Review. https://hbr.org/2006/01/competing-on-analytics

How to analyze a dataset: 6 steps: HBS Online. Business Insights Blog. (2017, April 5). https://online.hbs.edu/blog/post/how-to-analyze-datasets