# 



# School of InfoComm Technology

**Data Visualisation & Story Telling**

Specialist Diploma in Data Analytics

October 2020 Semester

**ASSIGNMENT 1**

**(Individual Assignment)**

**Submission Deadline:**

**20th December 2020 (Sunday), 11:59PM**

|  |  |  |
| --- | --- | --- |
| **Tutorial Group** | **:** | **P01 / P02 / P03** |
| **Student Name** | **:** | Tay Cong Run |
| **Student Number** | **:** |  |

**Penalty for late submission:**

10% of the marks will be deducted every calendar day after the deadline.

**NO** submission will be accepted after 24th Dec 2020 (Thursday), 11:59PM.

1. **Objectives**

In this project, I will be attempting to present data insights that might be helpful in improving the hotels’ operation. I chose to appeal to hotel managements instead of consumers as the target audience because the dataset is collected from one resort hotel and one city hotel. Due to the small sample size of hotels in this dataset, conclusions presented to consumers would have little extensibility to other hotels. Furthermore, the dataset contains data pertaining to the hotel booking system, source of bookings, as well as customers’ needs and preferences.

When properly analyzed, this dataset could reveal lapses in the hotel booking system, if any; better coordinate hotel bookings and available hotel resources; provide an informative breakdown of hotel customers; and reveal subtle customer behavioral patterns.

**Exploratory Questions**

1. How have check-in patterns and spending patterns varied over the years?
2. Which group of customers contributed to the lengthiest stays, and when was this period?
3. What is the breakdown of customers by country, market segment, and their means of booking the hotel?
4. How does no-shows or cancellations present themselves in the hotel booking system, if they are predictable from it at all? Was it because they were not assigned their desired rooms or held up too long on the waiting list?
5. **Data Preparation**

As can be seen from the Exploratory Questions above, this dataset can set out to answer many questions because its columns are comprehensive to our context. It is rather complete for the exploratory questions we have set out to answer, with one exception being that it is unclear whether ad-hoc compensations made to customers are duly backlogged into and reflected in the data. Also, the metadata provided alongside the dataset accurately describes the domain of categorical data, so the need for further cleaning is minimal.

With that being said, a few tweaks to the data can be made to make it easier to explore the data in Tableau:

1. To facilitate plotting visualizations that require the lowest granularity, I added to **hotel\_bookings.csv** a new column containing the row number, prior to loading it into Tableau.
2. Arrival Date was split into Arrival Date Year, Arrival Date Month, Arrival Date Week Number and Arrival Date Day of Month. These columns can be aggregated into a single Arrival Date column and typed appropriately as Date. The aggregated columns can be hidden to free up workspace in the side panel. It will enable us to fully utilize Tableau’s built-in date drilldowns and functions, all within a single column.
3. Meal packages has two aliases (SC, Undefined) that points to No Meal Package.
4. is\_repeated\_guest and is\_cancelled can be aliased with True (1) and False (0) to convert into a more sensible Boolean dimension.
5. Fields that were replaced with digit identifiers (e.g. Agent, Company) would be interpreted by Tableau as integer Measures. They should be manually converted into Dimensions.
6. **Exploratory Data Analysis and Visualisation**

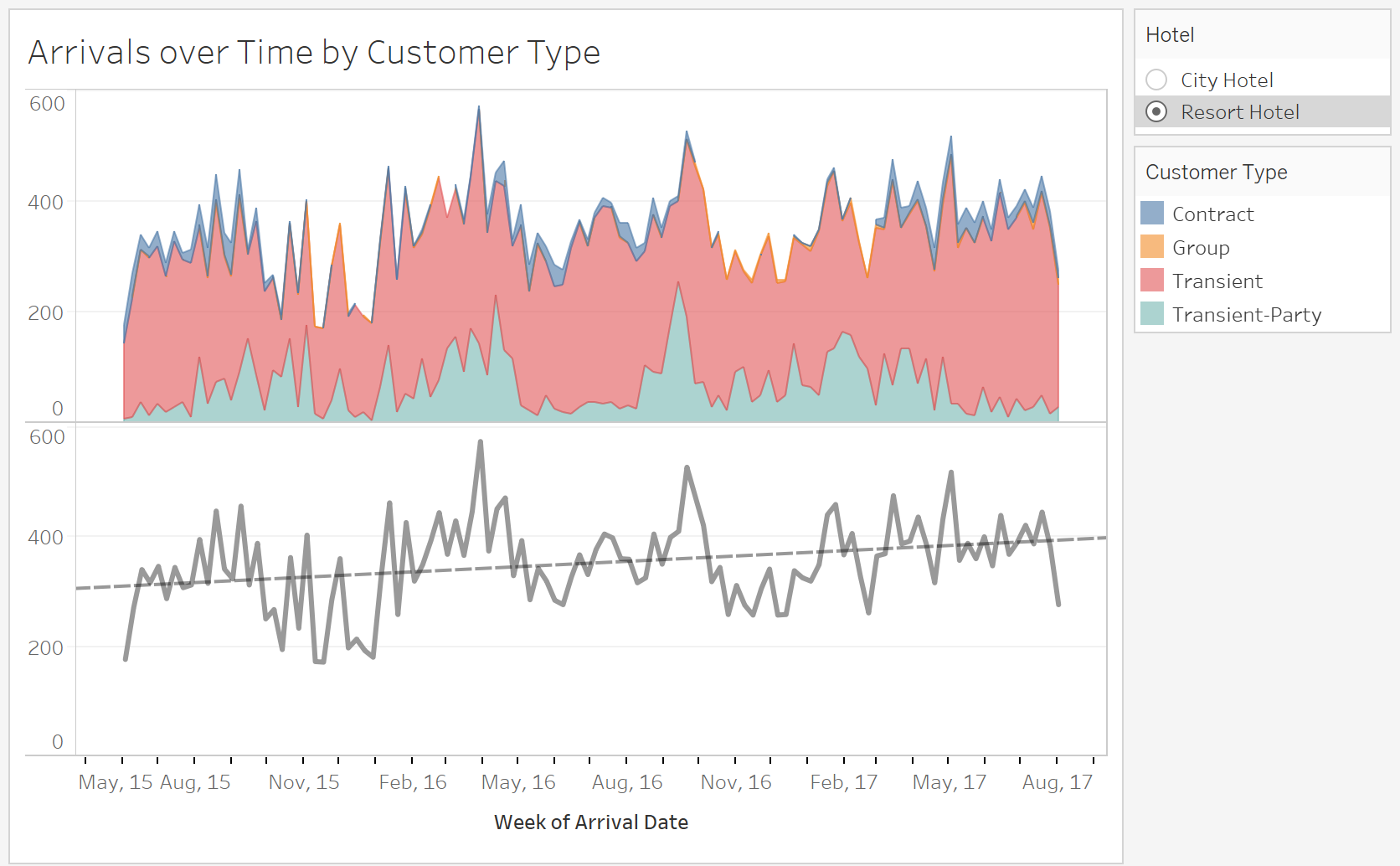
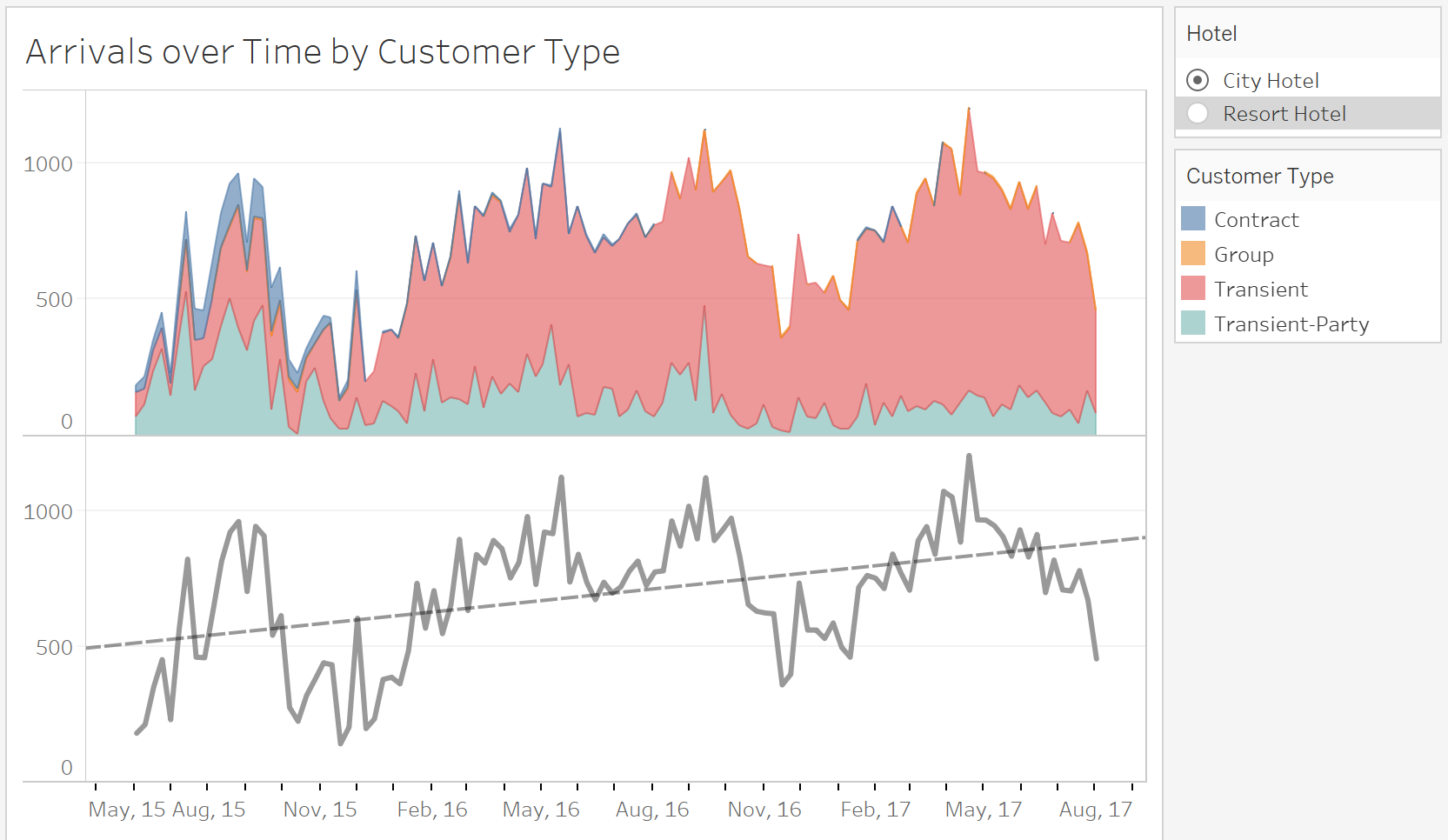


Figure 1: Arrivals over Time by Customer Type (City Hotel) Figure 2: Arrivals over Time by Customer Type (Resort Hotel)

Figure 1 & 2 illustrates the number of hotel stays (customer arrived and checked-out subsequently) at the City and Resort Hotel respectively. It further breaks down the stays by Customer Type, revealing that the bulk of hotel stays consists of Transient followed by Transient-Party customers, for both hotels. By fitting a trend line, it can be seen that both hotels are experiencing growth in hotel stays, with the City Hotel growing at a more remarkable rate. Additionally, either visually or by reading the R-squared values, it can be inferred that the City Hotel experiences more fluctuations in hotel stays whereas the Resort Hotel has a more consistent rate.

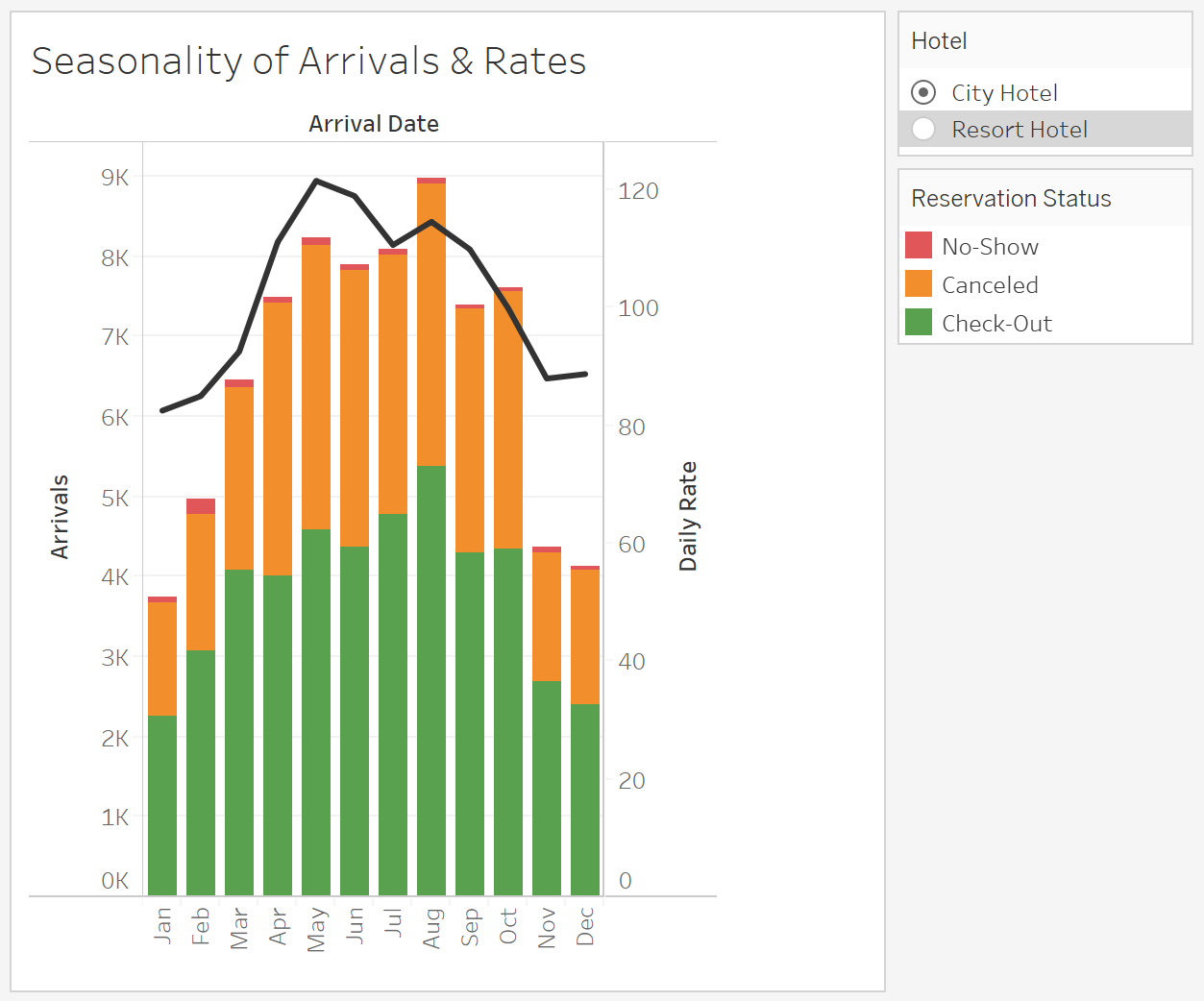
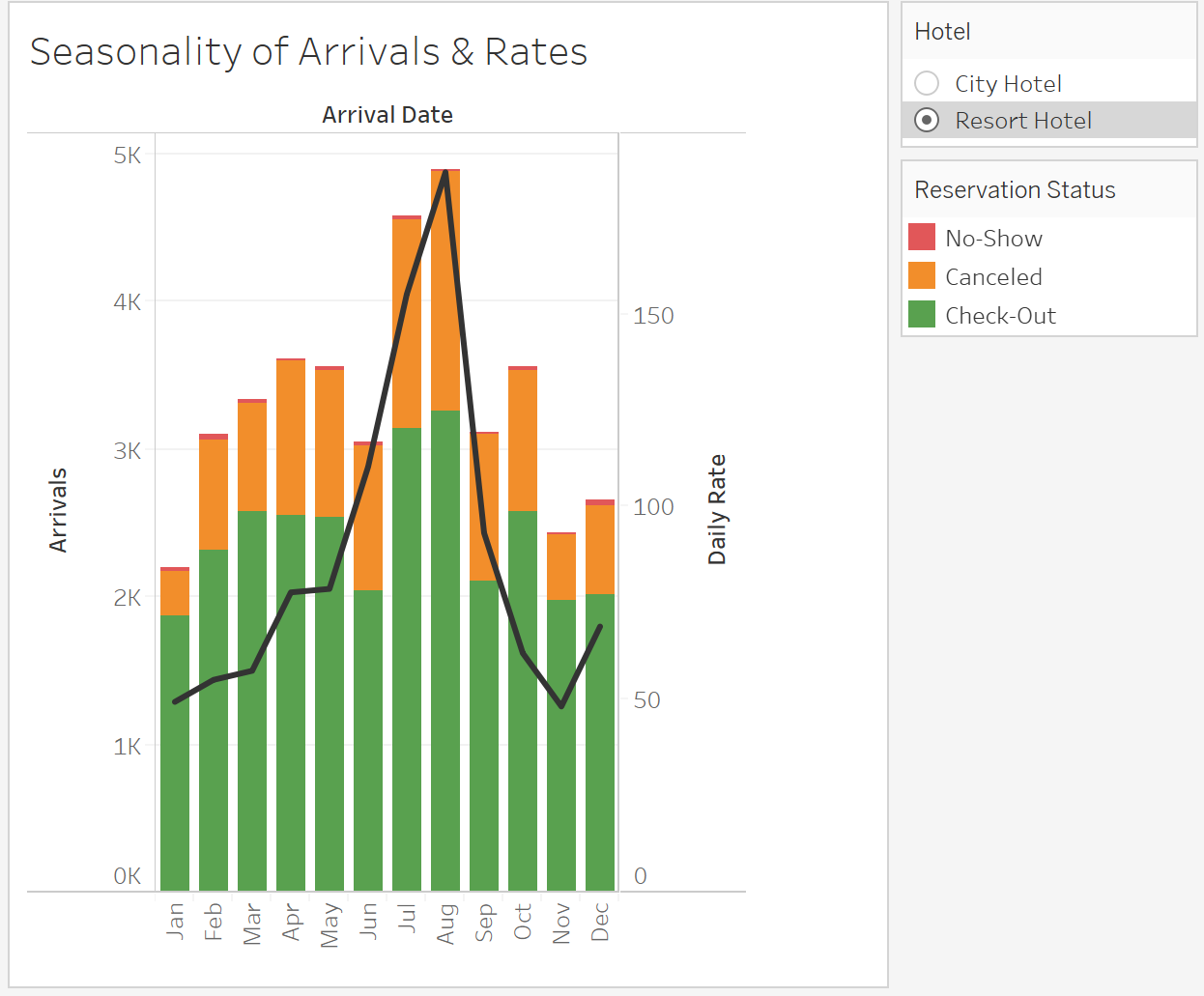
 

Figure 3: Seasonality of Arrivals & Daily Rates Figure 4: Seasonality of Arrivals & Daily Rates

Figure 3 & 4 explores the seasonality of hotel bookings and daily rates, which are likely for the hotel industry. Based on this visualization, hotel bookings and stays peak around August and slow down towards the start and end of the year, for both hotels. Daily rates follow this pattern closely as well. However, the Resort Hotel rates peaks higher in the peak season and falls lower in the slow season, compared to the City Hotel. This visualization also highlights that no-shows and cancellations are a much bigger problem for the City Hotel throughout the year, while it only plagues the Resort Hotel closer to the peak season.

With that being said, one ought to be cautious before deriving serious conclusions about seasonality from this visualization, because the dataset is incomplete for 2015 and 2017.

From here on out, we will place our attention on the City Hotel, in an effort to look into the cancellations and no-shows plaguing it and to minimize inter-hotel comparisons (reason stated above).

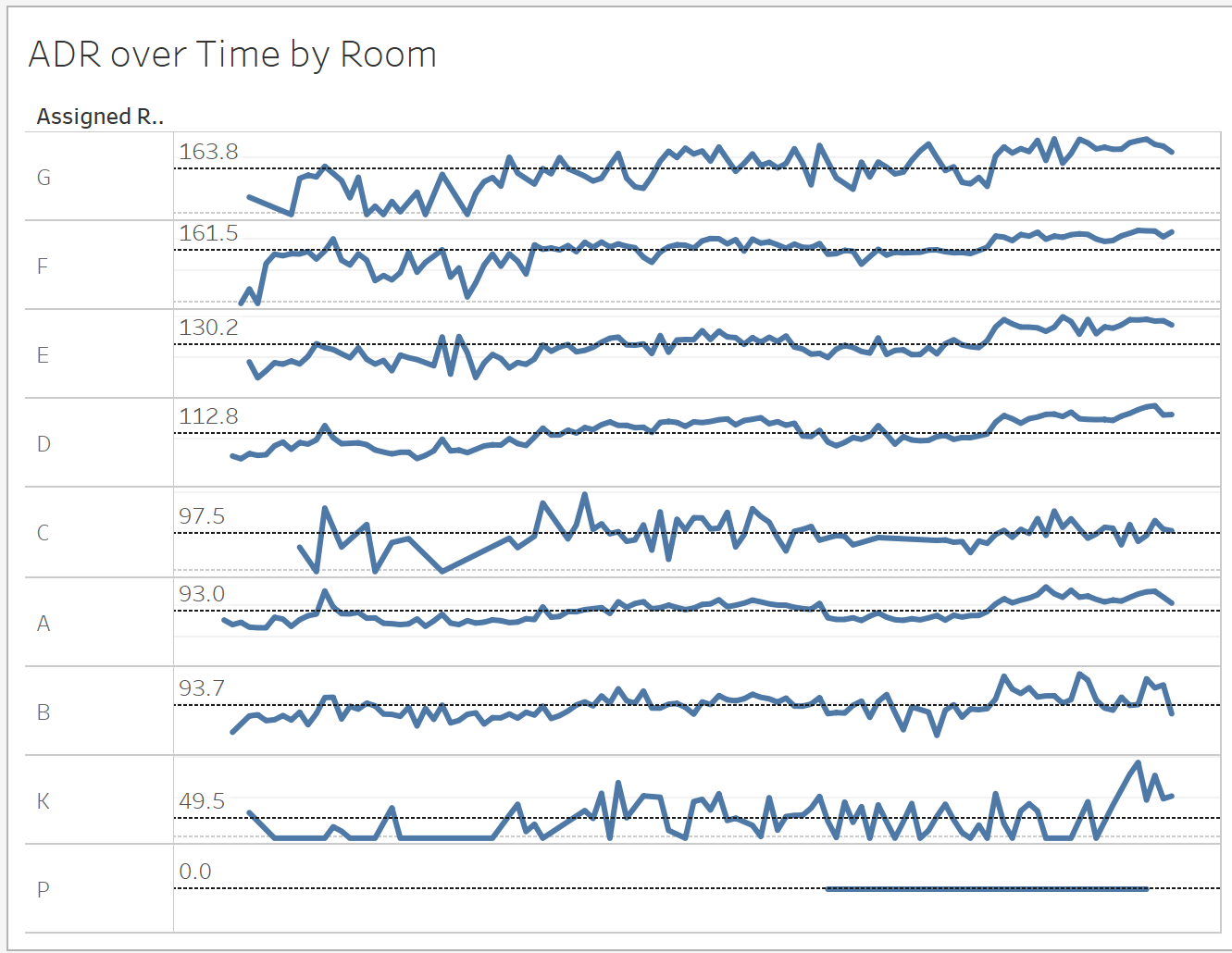


Figure 5: Average Daily Rates over Time by Room Type

Figure 5 provides information about how the hotel rates by room type have varied across time. At one glance, it is easy to tell which are the most costly or least costly rooms, as well as which room rates does not follow the seasonality pattern like most rooms would.

What might be of interest here are Room P which seems like a complimentary room and Room K which has had fairly few bookings at vastly different rates.

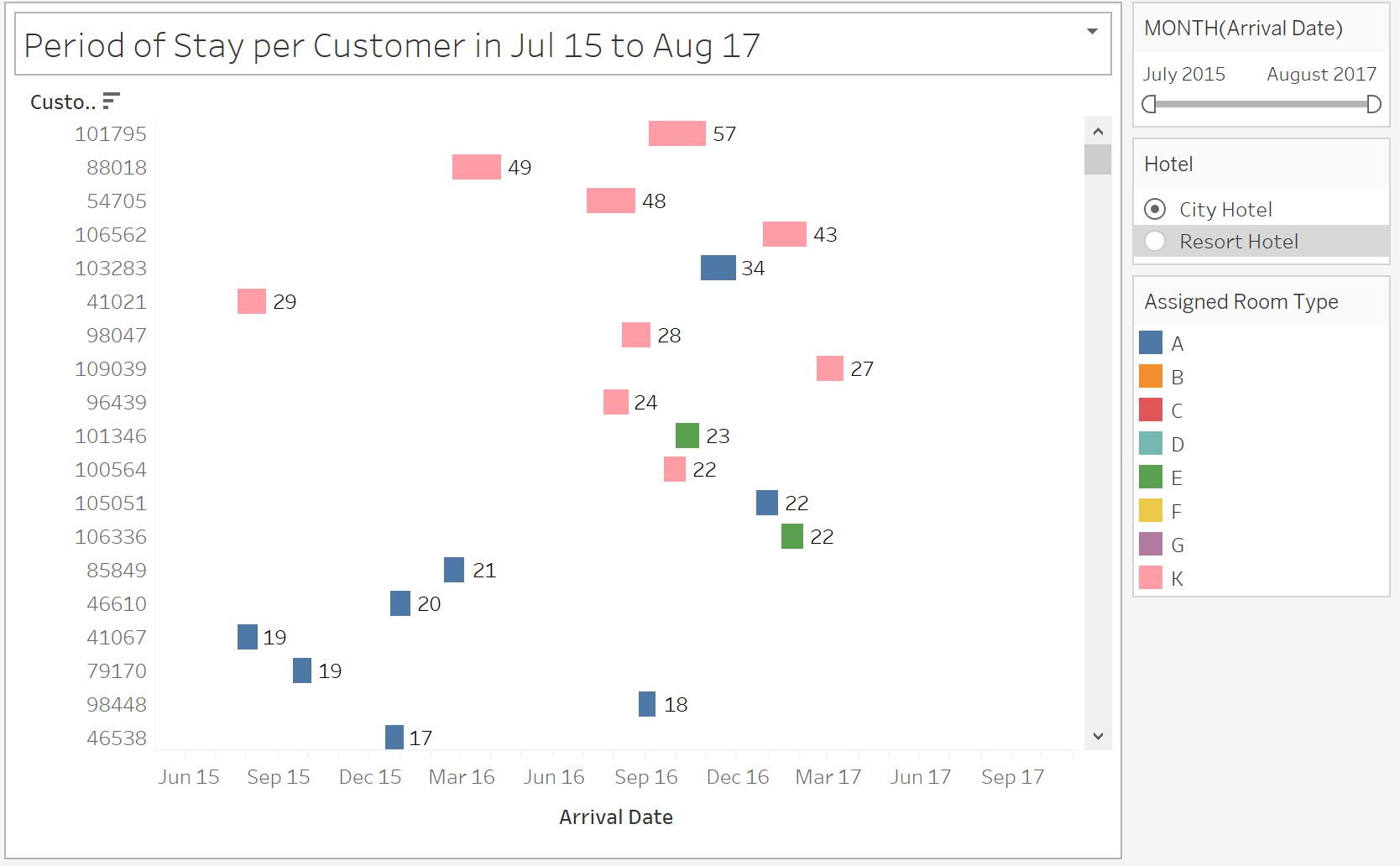
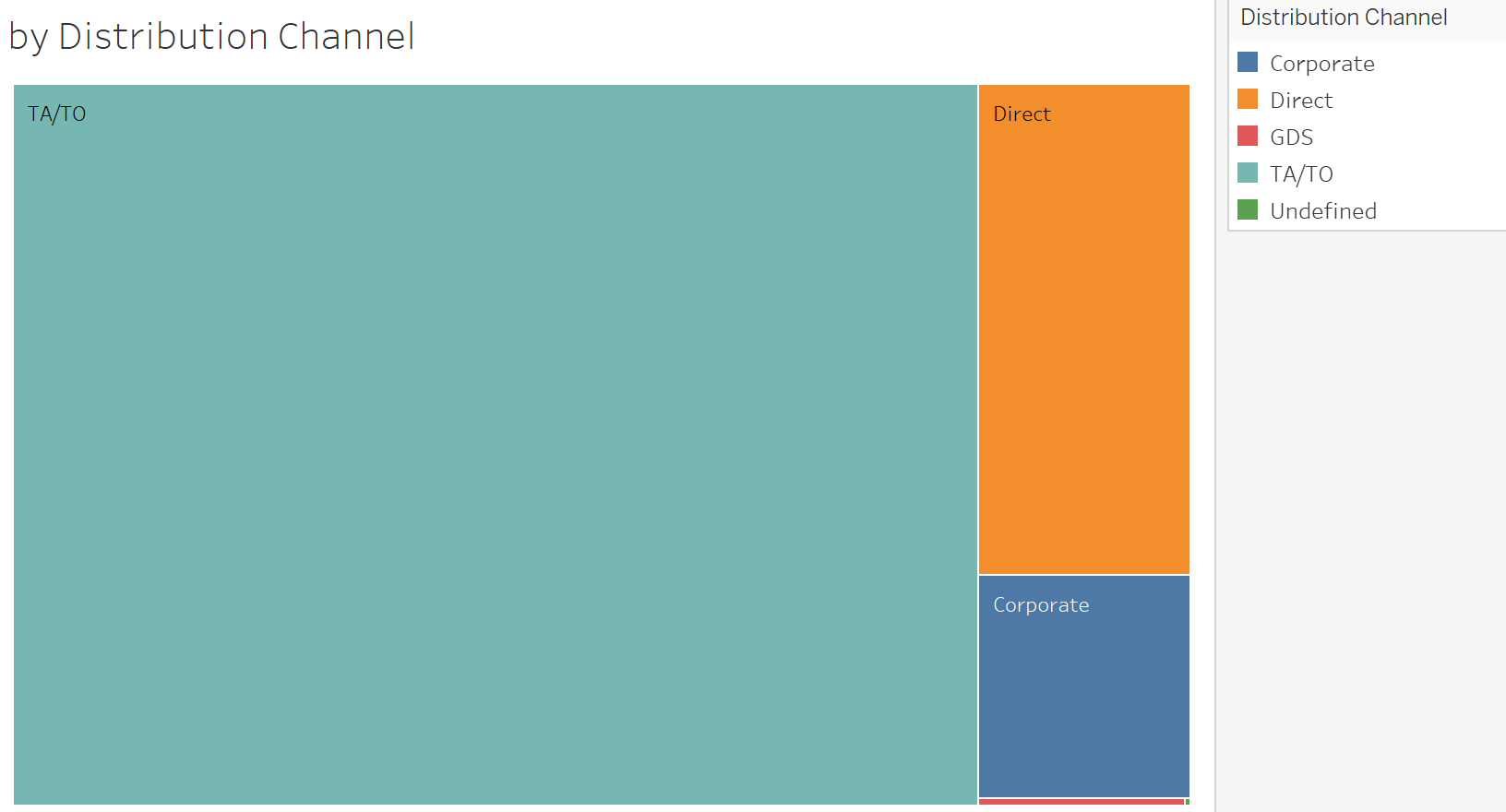
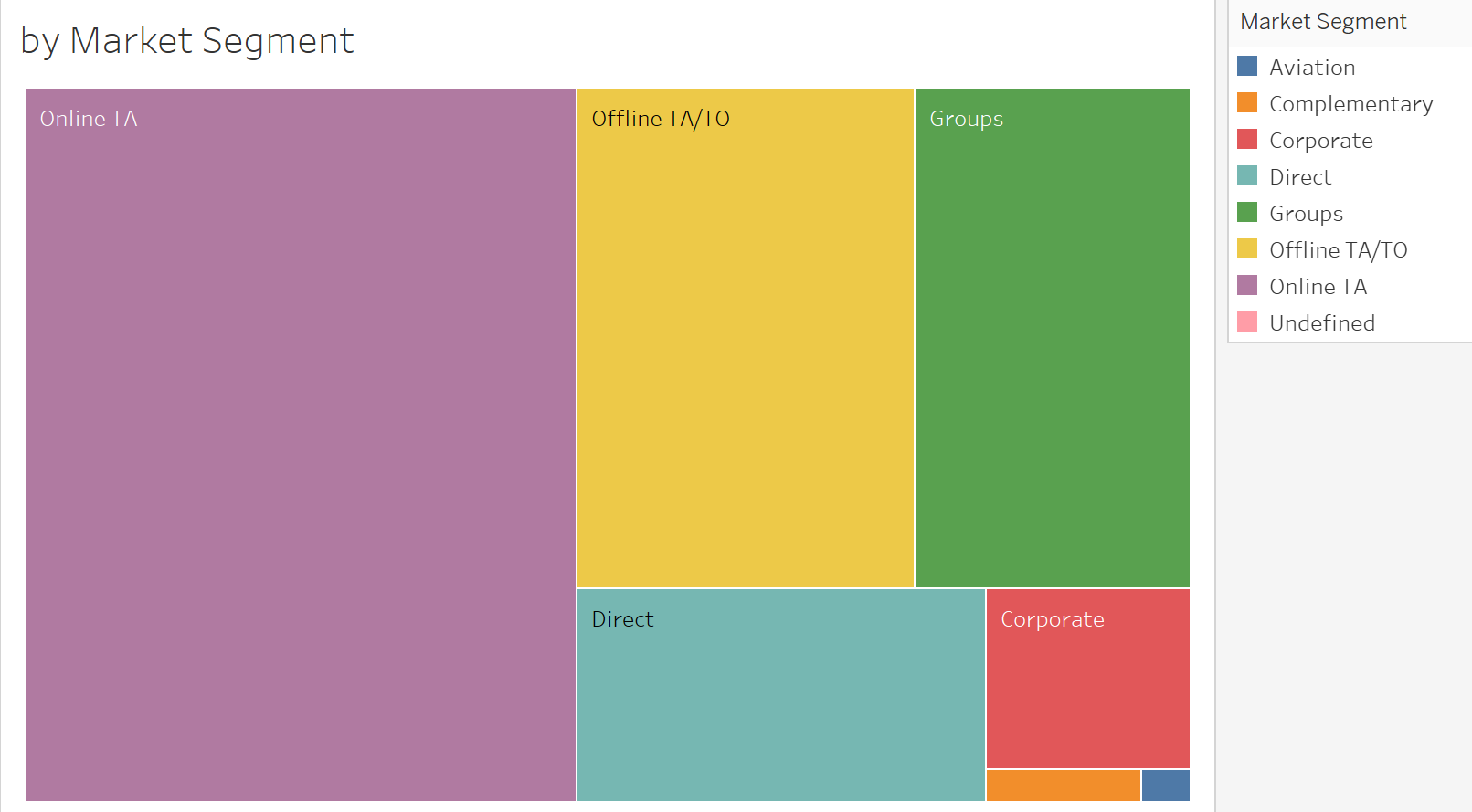


Figure 6: Customers Period of Stay

Figure 5 is a Gantt chart of each customer’s period of stay (from check-in till check-out). When sorted descending by the Length of Stay (the difference between Arrival Date and Reservation Status Date in days), it is immediately clear which customers contribute to the lengthiest hotel stays and in which period. Furthermore, the color coding by Assigned Room Type reveals that Room Type A & K are among the most popular rooms for extended stay. If we recall our findings from Figure 4, Room Type A is a medium rate room while Room Type K is a budget room. The cheap rate and extended stays are sound explanations for the few arrivals for Room Type K and its hike in rates towards 2017.

When we sort the Gantt chart ascending by the Minimum Arrival Date, what we get is an operational visualization that tells the user customers’ check-ins, check-outs, period of stay and in what room type, in chronological sequence. Coupled with a time bar to filter down the vast number of entries, it is ready for operational reference for both historical and real-time use.

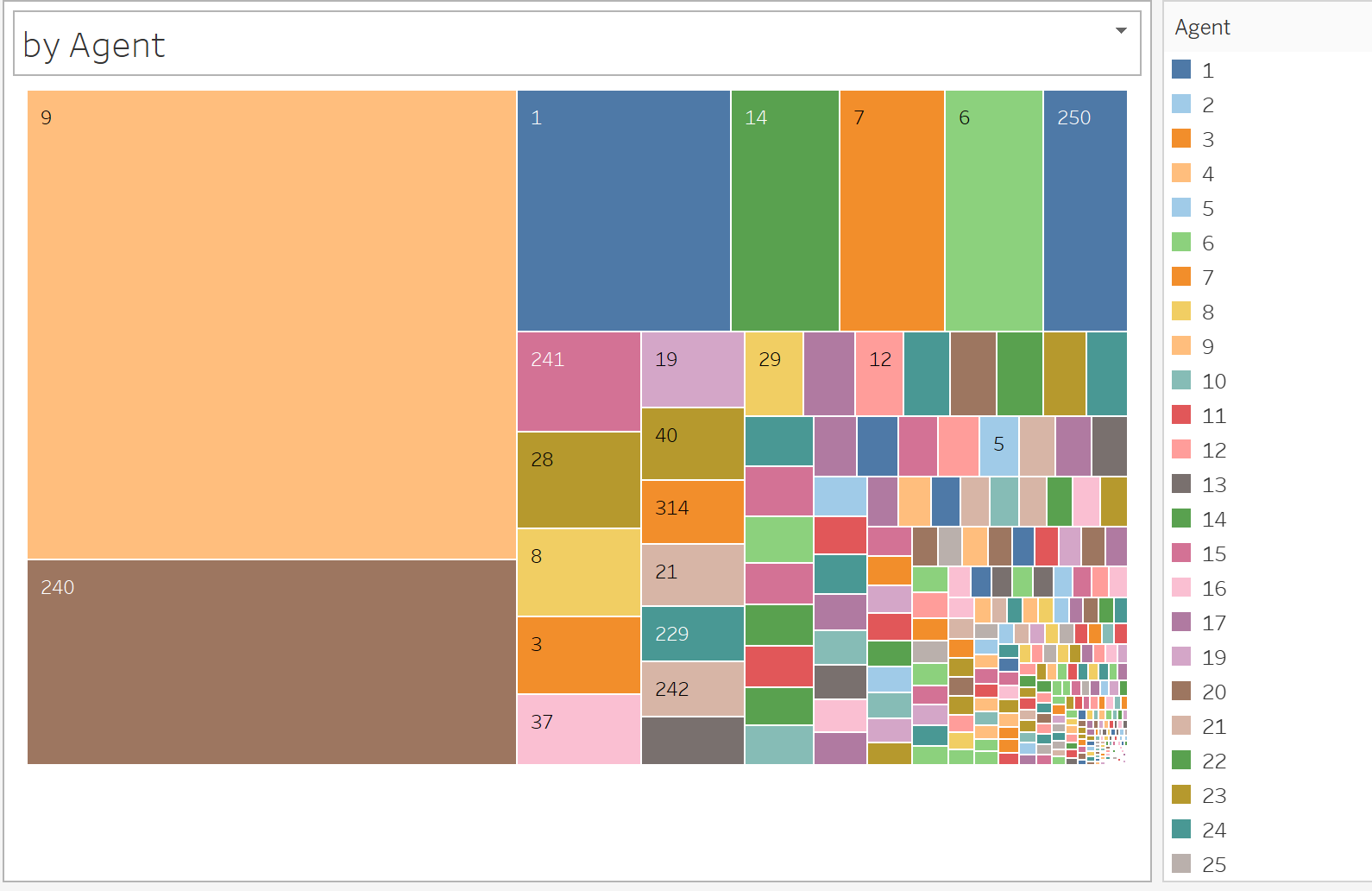
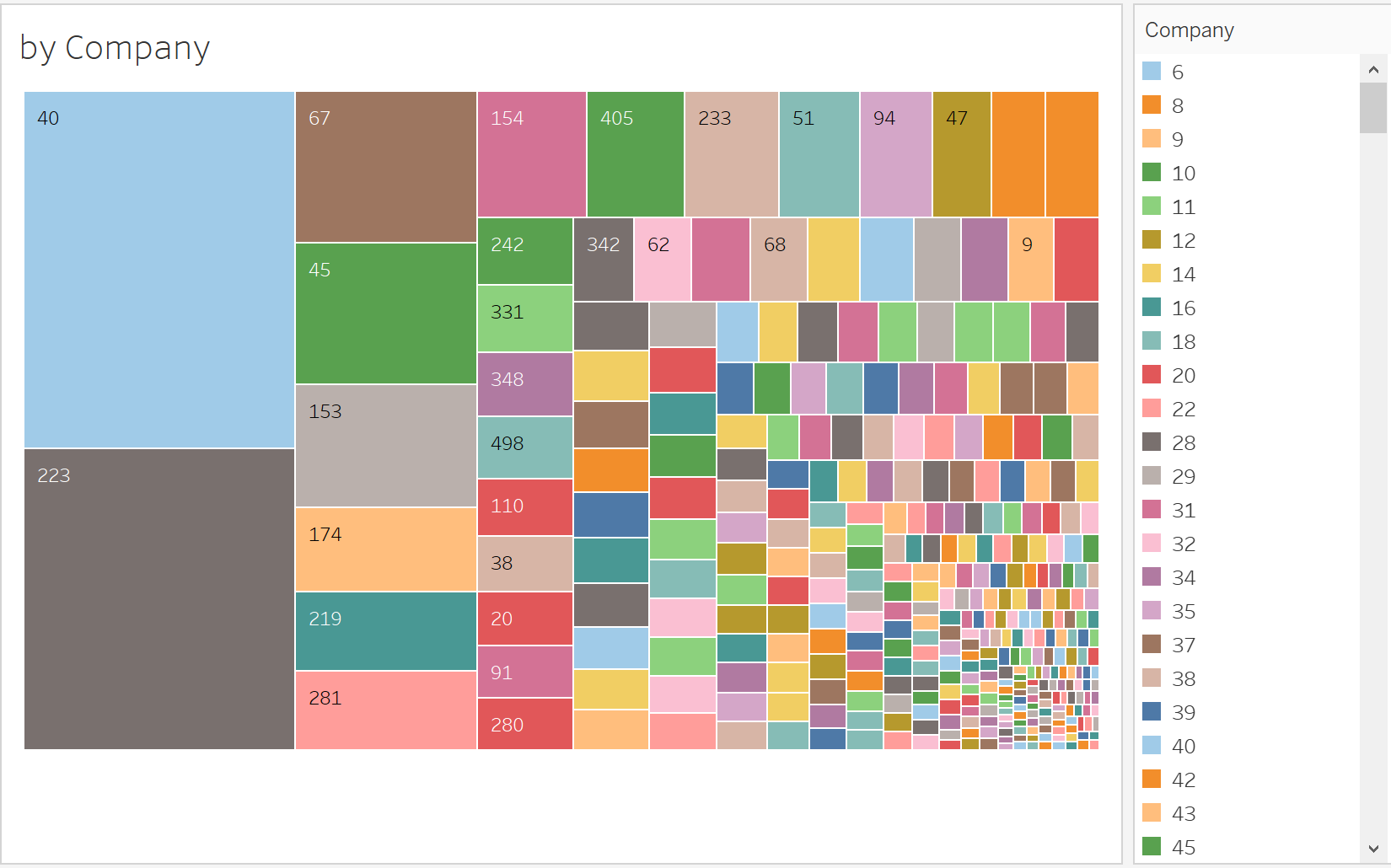
 

Figure 7: Breakdown of Bookings

The tree map visualizations in Figure 6 seeks to breakdown the source of hotel bookings. The main channels the hotel receives booking are through Travel Agents and Tour Operators or via Direct booking. It is also interesting to see that Offline Travel Agents and Tour Operators remain very relevant despite the advent of technology and the Internet.

The interactive prowess of the tree map visualizations on a dashboard will value-add in the form of insights hotel management can extract. It can act as both a breakdown and a filter for the dashboard in question.

When it comes to the breakdown by Company and Agent, the tree maps are cluttered. Despite that, it remains relevant as the various filter actions on a dashboard can improve the cluttering and provide a clearer picture of the Companies and Agents in question.

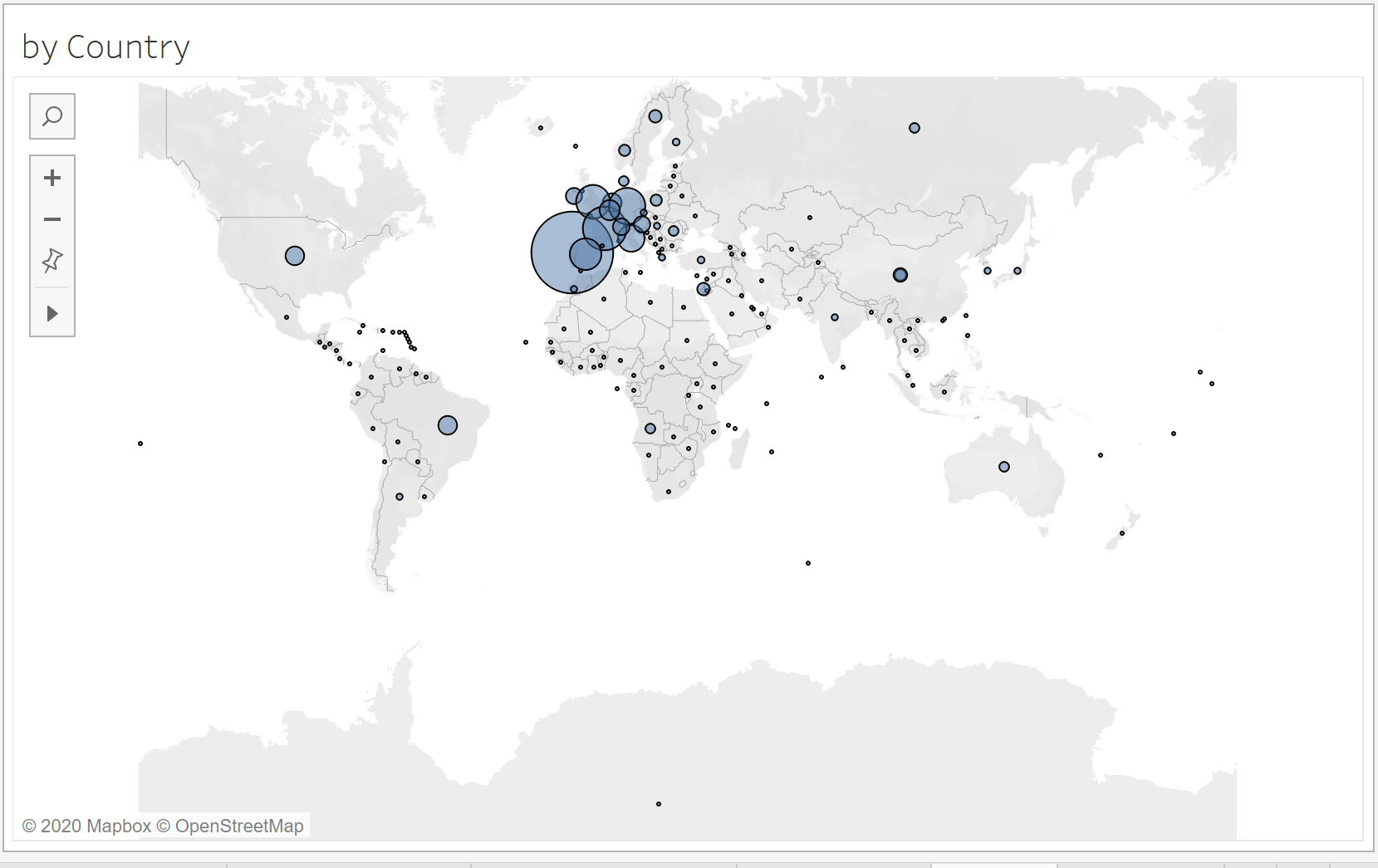


Figure 8: Origins of Bookings

Figure 7 provides a map view of the country of origin of hotel bookings. It can give the user a sense of what the global coverage of the hotel currently is. In the future, this will be helpful in planning advertisements or engaging vendors, in a more targeted fashion.

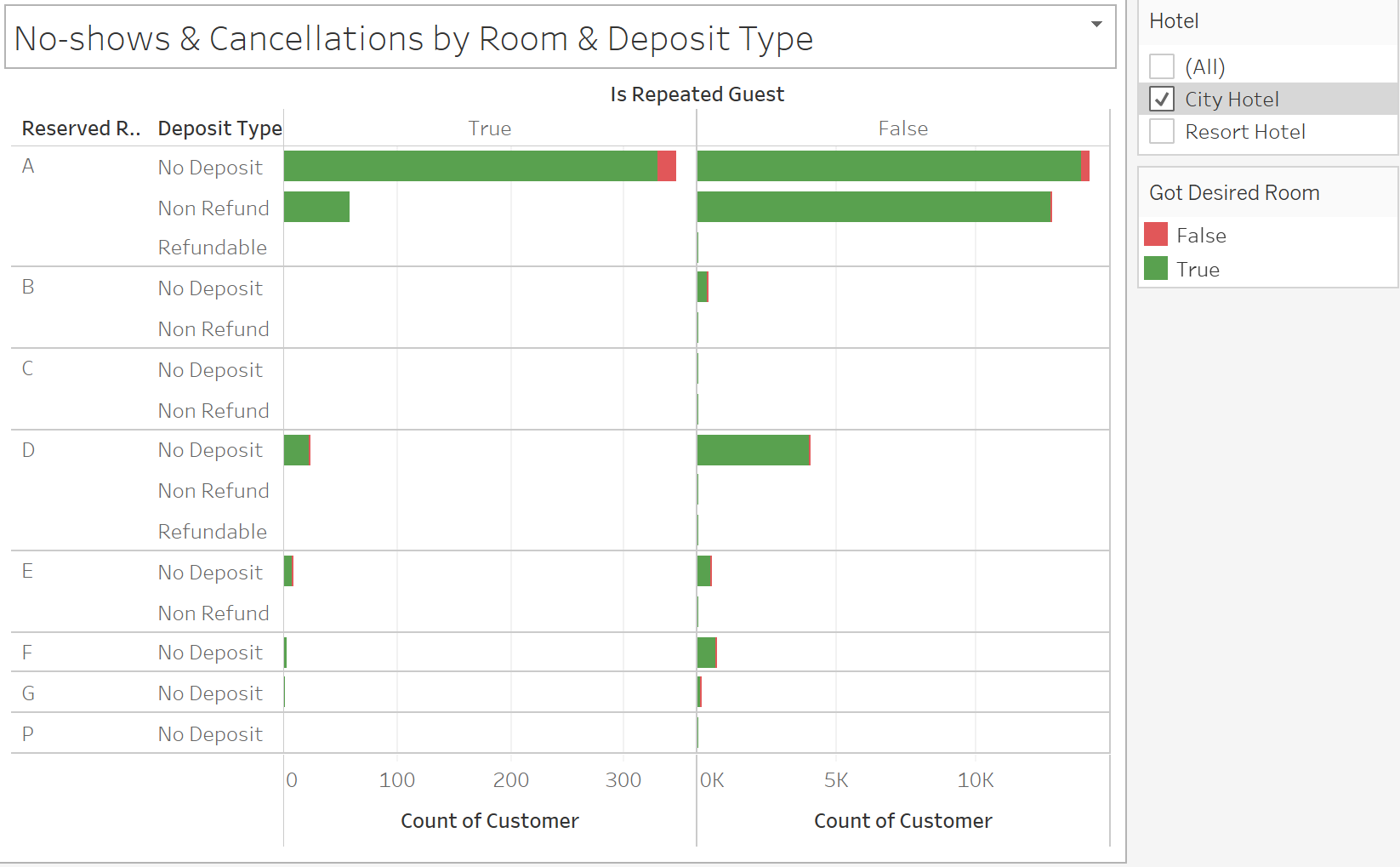


Figure 9: No-shows & Cancellations by Room Type and Deposit Type

Figure 9 is a simple bar chart used to answer the number of no-shows and cancellations experienced for each Room Type and Deposit Type. It is further segregated into Repeated Guests and Non-repeated Guests, to determine if there is a significant difference in no-shows & cancellations between the two. In addition, the color-coding tells us whether the customer that cancelled or no-showed was assigned their Desired Room Type.

Room Type A is most prone to no-shows & cancellations by all guests, and Room Type D is prone to no-shows & cancellations by Non-repeated Guests. A larger proportion of no-shows & cancellations come from Non-repeated Guests, as can be seen from the difference in scale of the axes.

Generally, implementing a Refundable or Non Refundable Deposit is sufficient to bring down discourage no-shows & cancellations. However, this is not the case for Room Type A. It is likely that Room Type A, which is a medium-priced room, is offered to customers to book more flexibly by accepting little to no deposits.

From the color coding in Figure 9, it is safe to say that cancellations and no-shows are rarely caused by the hotel’s inability to assign customers their Desired Room Type.

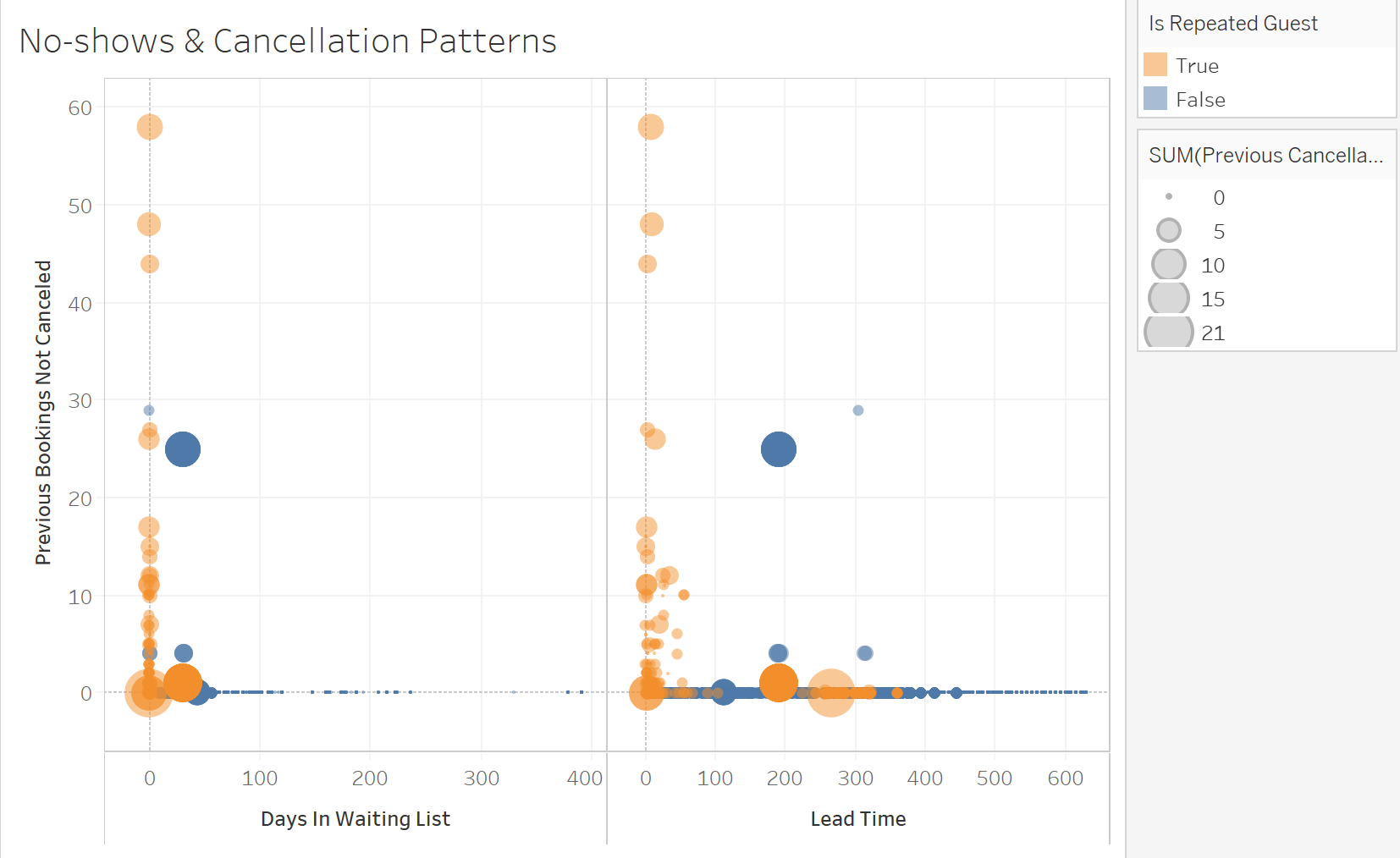


Figure 10: Cancellation & No-show Patterns

Figure 10 presents two scatterplots of customers that have cancelled or no-show, by plotting Previous Bookings Not Canceled against Days in Waiting List and Lead Time respectively. To further take Previous Cancellations and whether the customer Is Repeated Guest into account, we size the data points by the former and color-code them by the latter.

Previous Bookings Not Cancelled can be interpreted in two different ways. The first being a good track record of Repeated Guests for following through with prior bookings (namely Orange data points that are higher up along the y-axis). The second being a red flag that Non-repeated Guests have bookings in-force and is cancelling this booking (namely Blue data points that are higher up along the y-axis). Interestingly, there is a dark Blue clutter representing many Non-repeated Guests that have both a large number of Previous Cancellations and Previous Bookings Not Cancelled. Hotel management might want to profile these customers and keep an eye out, as they might be bookings made by bots to jam up the hotel booking system.

Looking at the left scatterplot, the hotel has done well to not keep Repeated Guests in the Waiting List for long. However, there are many first-time guests with few Previous Cancellations that were kept on the waiting list for extended periods of time (small, Blue data points along the x-axis). Perhaps more can be done to serve these potential first-time customers of the hotel.

Looking at the right scatterplot, customers who tend to make bookings further in advance have fewer Previous Cancellations (smaller data points towards the right). There are, however, several outliers. More notably, there are Repeated Guests who booked in advance but still has many Previous Cancellations. Perhaps the reception can check-in more regularly with these Repeated Guests to affirm their advanced bookings.

1. **Exploratory Data Analysis and Visualisation**

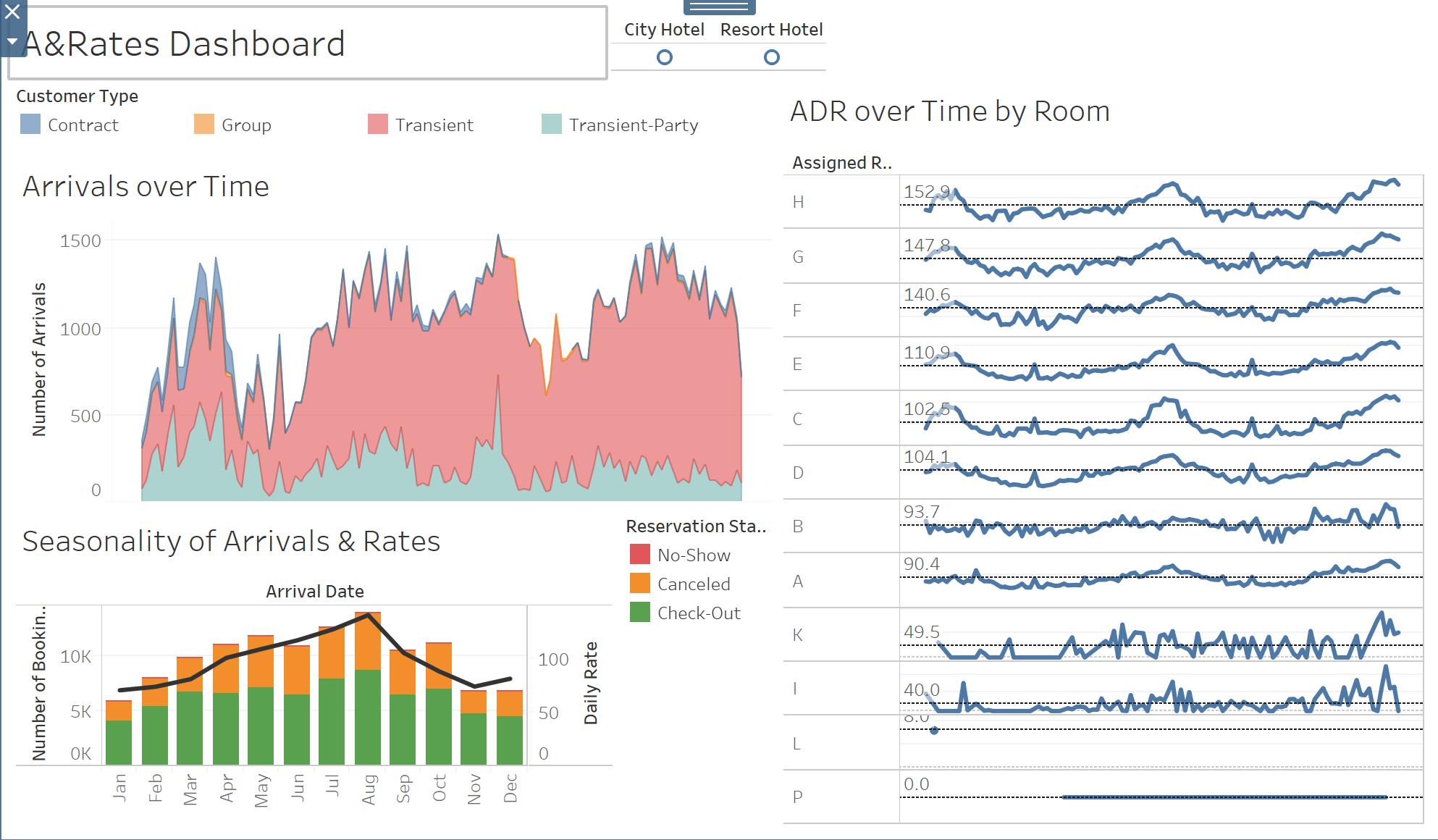


Figure 11: Arrivals & Rates Dashboard

Figure 11 serves as an informative summary dashboard for hotel management to track the patterns of Number of Arrivals and Daily Rates over time. It combines Figure 1/2, 3/4 and 5, coupled with a button for user to choose between City and Resort Hotel.

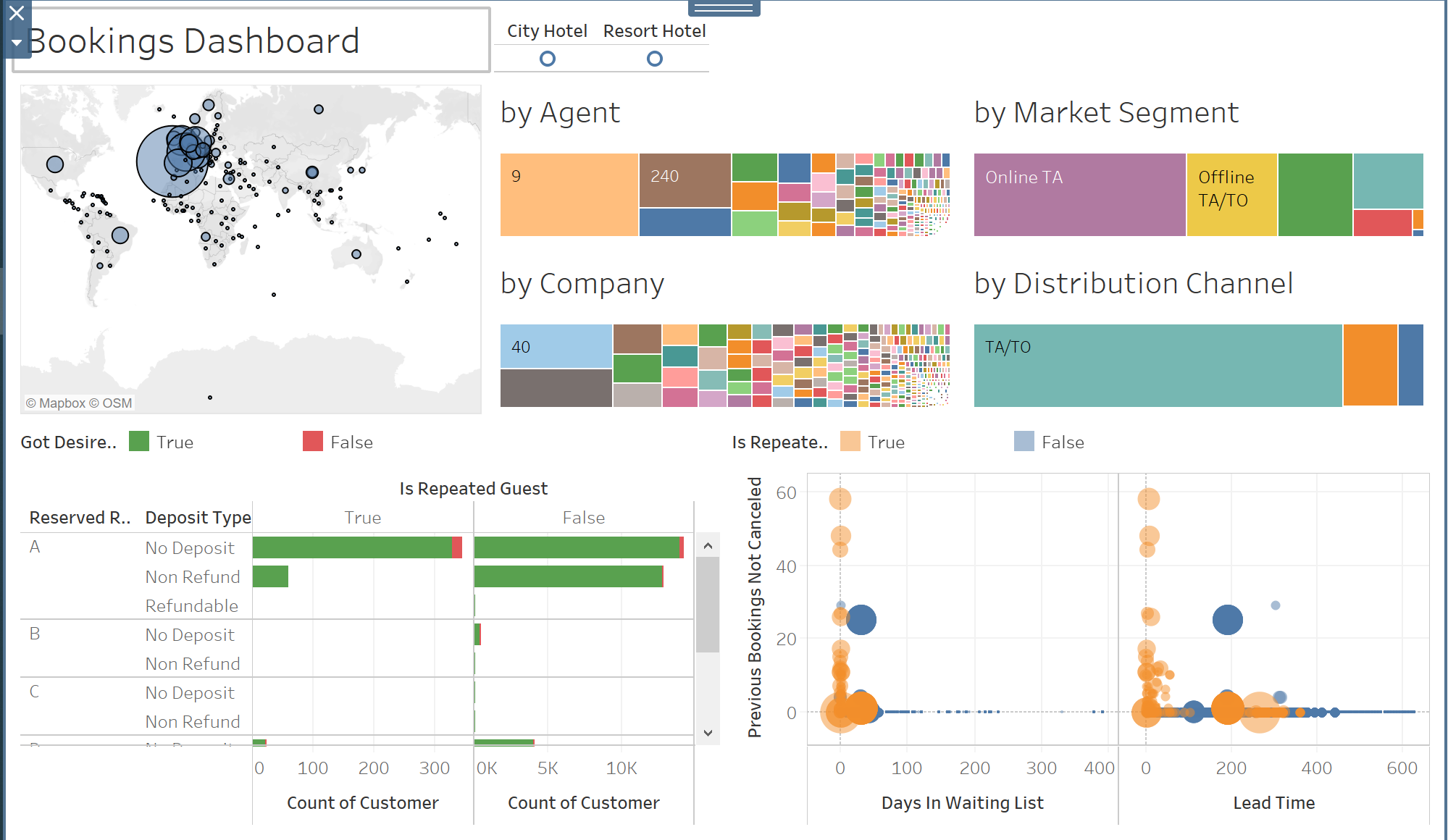


Figure 12: Bookings Dashboard

Figure 12 is an interactive dashboard for users to explore the origins of hotel bookings. It includes Figures 7 through 10. Every visualization here are used as filters, to allow users to filter the vast number of data points based on various aspects while maintaining the informative breakdown in each visualization. Similarly, there is a button for user to choose between City and Resort Hotel.