Visualization Techniques to Gain Actionable Insights – Tableau and ggplot

Katranitsiotis Panagiotis

Abstract. It is widely accepted that Big Data Analytics have become a globally researchable field for scientists. The rapid evolution of technology and internet led both to the significant increase of the available amount of data and the breakthrough of the statistical tools. Visualization of data is one of the most common methods used for an efficient and accurate statistical analysis. This paper describes the usage of the two most remarkable visualization tools, R and Tableau. R is the most preferable open source language for handling Big Data and produces plethora of useful and complex plots using the library Tidy Verse and ggplot. Tableau is a commonly used software especially for users who are not familiar with programming languages like R. Its most significant advantage is its drag and drop interface which provides a more friendly environment for users comparing to R. For this visualization analysis, a hotel demanding dataset will be used in order to understand better the relations among the variables of hotel bookings and cancellations.

Keywords: Big Data; R; Tableau; Visualization; TidyVerse; ggplot2; Hotel Data;

Contents

1 Introduction

Nowadays Big Data is one of the most researchable field for scientists. Due to the rapid evolution of technology, people tend to enter IT technologies more and more in their daily lives. As a result, compared to previous years a tremendous amount of data coming from mobile devices, applications, emails, cars etc. is now available for researchers [1]. This led scientists to create the necessary tools for processing them in order to improve people's life and routine.

The way of processing all these data is an extremely difficult procedure. A few years ago, analyzing and even storing Big Data were complicated problems for scientists [2]. Owing to the development of computers and the breakthrough of CPUs speed and intelligence, now the storage of massive data became an easier procedure.

Big Data Analytics have an important impact on numerous fields. In recent years, big datasets have been processed by academics, industries, businesses, etc. in order to extract accurate statistical results and in many cases to use forecasting methods such as profit prediction models.

The basic characteristics of Big Data are Volume, Velocity and Variety. More specifically, compared to the most common belief, a problem with a huge amount of data is not necessarily categorized as a Big Data one. The reason why high velocity and variety of data are decisive for Big Data problems, is their crucial role for producing effective prediction models and accurate statistical results. For instance, the visualization of data is an extremely effective tool to understand the relation among variables of the dataset and even producing significant results.

The purpose of this paper is to use the two most common visualization tools, Tableau software and the open source R language. In Big Datasets, the visualization process plays a vital role in concluding accurate results because it contributes to a better understanding of the relations among variables. The selected dataset for this purpose is based on hotel demands and more specifically on booking cancellation data.

2 Dataset

The selected dataset is referred to hotel booking demands. It is retrieved from Kaggle which is a reliable and widely known site for its datasets for analysts [3]. This dataset provides several variables that have a decisive role for statistical analysis. For instance, the capacity of rooms, reservation status, the number of adults and children, country as well as the reservation dates are some of the information that contribute to an accurate analysis. In addition, these variables are highly important for visualization results, such as booking cancellation rate. As a result, the "Hotel booking demand dataset" is a preferable chosen one for using Tableau and R for the visualization procedure.

This dataset describes the demand of two hotels, Resort and City Hotel. Both hotels are established in Portugal and more specifically the city hotel is located in Lisbon and the resort one in Algarve. The dataset consists of 40,060 observations of the Resort Hotel and 79,330 of the City one. Each observation represents a reservation in one of

the two hotels. In addition, this data set is composed of 32 variables related to hotel bookings. All the data represent hotel reservations from arrival dates between 1st of July 2015 and 31st of August 2017 [4]. This means that from statistical perspective, all the 119,390 observations from 32 variables will be used to achieve actionable visualization insights.

3 Evaluation of the two methodologies

The purpose of this paper is to provide visualization results based on graphs by comparing R and Tableau for their usage in the visualization procedure. More specifically, the presented paper will provide the necessary plots both creating by R and Tableau. Comparing the same plot depending on the program that was created, we will accomplish an accurate evaluation of these two tools. R as an open source programming language, offers more flexibility to the user to handle Big Data problems and consequently to achieve the desired visualization results. On the contrary, Tableau because of its drag and drop interface provides friendlier environment for users. This means, that users who are not familiar with the programming procedure will be able to provide visualization results as accurate as the ones by using R.

For this procedure the selected dataset focuses on the demanding of two hotels. The paper will use both R and Tableau to create plots in order to answer the specific questions created by the variables of the selected dataset. More specifically, the main purpose of this paper is to understand which variables are related to cancellations and consequently which of the two hotels has the largest cancellation rate. In order to conduct this result, some questions have to be answered first: (a) which of the two hotels (City or Resort Hotel) have the most reservations and cancellations, (b) from which countries are the guests of its hotel, (c) is average daily rate based on customer type, (d) is average daily rate differs on repeated guests and finally (e) does the cancellation rate differs for each year and each hotel?.

4 Analysis of dataset

As mentioned, the selected dataset contains booking data from two hotels in Portugal, the City Hotel and the Resort Hotel. In order to provide useful insights based on visualization tools, the importance of this dataset has to be described.

The main reason that this dataset has a tremendous impact on the society, is the importance of tourism to each country's economy. Tourism is one of the most economic developed industries globally. In many countries, such as Portugal, their basic income comes from tourism, especially in seaside areas, mountains, areas with important historic interests, etc. More specifically, the tourism industry provides income to countries coming from tourists, activities, hotels, taxes, souvenirs, etc [5]. From hotels aspect, each year's cancellations have a serious economic impact not only from business perspective but the state's economy as well. Therefore, the importance of the selected

dataset is decisive for both analysis and visualization as it contributes to a better understanding of the cancellation rate by finding which variables are related to this problem.

The booking demand data coming from this specific dataset is referred to two hotels in Portugal for years 2015, 2016 and 2017. More specifically, the Resort Hotel is from Algarve and the City Hotel is from Lisbon.

Based on the figures 1 & 2 *(Appendix C)*, it is concluded that in both hotels the most amount of reservations, made from period 1st July 2015 to 31st August 2017, were not canceled. However, the cancellations of each hotel are not a negligible percent. More specifically, there is only approximately 10% difference between cancellations and reservations. Consequently, this percentage remains a sizable one as it is big enough for crucial financial effects.

For an accurate conclusion of which of the two hotels has the most cancellations, several factors must be taken into account. For instance, to achieve an objective result about cancellation rate, each of the compared hotels must have tourists coming from similar countries. As it is concluded from figure 3 *(Appendix D)*, both Resort and City Hotel have costumers coming from the same countries, and in fact from a large range of countries around the world. More specifically, compared to the Resort Hotel, the City Hotel has most of its guests coming from Europe.

In addition, a determinant factor that affects the cancellation rate of hotels is their Average Daily Rate (ADR). ADR is the ratio of rooms revenue and number of rooms booked, which makes it a significant variable for customers to proceed with the confirmation or cancellation for their reservation [6]. Consequently, the booked meal as well as the customer type of each reservation are vital for ADR. It is expectable that in cases such as Group bookings or reservations with full board meal, will have a larger average daily rate than simple reservations. Figures 4 & 5 *(Appendix E)* which provide an aggregated graph among ADR, Customer Type and booked Meal, result that eventually there is no significant difference in daily rate between these variables. Examining more closely figures 6 & 7 *(Appendix E)*, we can understand that the largest Average Daily Rate is coming from Transient Customer Types bookings with Half Meal booked in their reservation. However, this difference compared to the other types of customer and meal is negligible. As a result, the ADR from this dataset is an objective variable to provide reliable insights for the cancellation rate.

Moreover, ADR is associated with other variables as well such as room type. In the selected dataset, this factor indicates the type of room of each reservation referring to type from the letter A to P instead of the actual number for ethical reasons. In addition, there are several cases that hotels change the accommodation cost per night for repeated guests and thus ADR is also affected by this factor. Based on figures 8 & 9 *(Appendix F)* it is concluded that city hotel has a larger range of daily rate for non-repeated guests than repeated customers. In fact, the difference between them is considerable as the maximum rate for repeated guests is approximately 100 comparing to non-repeated ones which is over 200. On the contrary, in the Resort hotel there is not a notable difference neither for the range of ADR nor the maximum level of it. Last but not least, for this boxplot we can observe that the maximum levels of ADR are basically for the medium types of rooms such as E, F and G.

Finally, in order to achieve a reliable insight of Cancellation Rate, we should consider that this value may differ among years. In figures 10 & 11 *(Appendix G)* we can observe that the most cancellations were made in May 2017 with 2762 canceled reservations. More specifically, the largest number of cancellations in these three years is approximately 2000 to 3000. However, based on figures 12 & 13 *(Appendix G)* we notice that the cancellation rate is not similar to each hotel. The cancellation rate in City Hotel is not fixed and in fact its highest levels differ each year. On the other hand, the Resort Hotel has a standard amount of cancellations and significantly in lower levels comparing to the City Hotel.

5 Conclusion and Tools Comparison

The purpose of this paper was to provide an accurate and a notable insight of the cancellation rate of the City and the Resort Hotel and compare R and Tableau for its usage to this accomplishment. Considering the presented graphs, it is concluded that cancellation rate is remarkably higher on the City Hotel. Resort Hotel not only has the lowest cancellations but also this rate remains on similar levels thorough months and years. To achieve the above result both graphs from R and Tableau were considerable.

From technical perspective, R is one of the most common and reliable open source languages. Apart from its highly performance in data analysis, it can also provide significant graphs by using TidyVerse and ggplot2. More specifically, it can handle high complexity problems which is really important for big datasets. Nevertheless, it needs a high knowledge of programming to achieve a desirable graph as well as the installing a variety of packages for this accomplishment. Its environment is not friendly for users who are unfamiliar with programming languages. However, one of the most significant advantage in writing code for producing graphs, is the flexibility for the user to produce the graph exactly as he desires it.

On the other hand, Tableau provides a very friendly environment for users due to its drag and drop interface. Especially people who are not familiar with programming, can easily produce remarkable graphs in which can easily change the colors, the labels and even enter new variables on the same graph just by pressing a button. However, an important disadvantage of Tableau is that it provides limited visualizations which occasionally makes it impossible for user to produce a desirable graph. It is a very useful tool mainly for people less skillful with statistics and programming languages.

As a conclusion, both R and Tableau are really useful tools for visualization. Tableau is more useful for non-programmers because of its interface which makes it a quick tool for creating notable graphs. Even for programmers this feature is sometimes extremely helpful for easy and quick graphs. For instance, figure 3 *(Appendix D)* which was created very easily by Tableau and was necessary for our story, was almost impossible to be created by R. However, because of the complexity of Big Data problems, R is more preferable as it provides more flexibility to users and particularly to data scientists because it can also be combined with statistical analysis.

References

1. Sagiroglu, S., Sinanc, D. “Big data: A review”, *2013 International Conference on Collaboration Technologies and Systems (CTS)*, San Diego, CA, pp. 42-47, (2013).
2. Russom, Ph. “Big Data Analytics”, TDWI Best practices report. Fourth Quarter, pp.4-6 (2011).
3. Kaggle, (2019) *Hotel Booking Demand*. [Online] Available from: https://www.kaggle.com/jessemostipak/hotel-booking-demand. [Accessed: 10th November 2020]
4. Nuno, A., Almeida, A., Nunes, L. “Hotel booking demand datasets”, *Data in Brief*. Elsevier, 22, pp. 41-49, (2019).
5. Sanchez-Medina, A., C-Sanchez, E. “Using machine learning and big data for efficient forecasting of hotel booking cancellations”, *International Journal of Hospitality Management*. Elsevier. (2020)
6. Chattopadhyay, M., Mitra, S. “Determinants of revenue per available room: Influential roles of average daily rate, demand, seasonality and yearly trend. *International Journal of Hospitality Management*. Elsevier. (2018)

Appendix A: Description of Dataset

The dataset consists of 119,390 records with 32 variables about the booking demanding in two hotels.

Table 1. Description of Variables

	Variable Name	Type	Description
1.	Hotel	Categorical	Type of the Hotel: Resort or City hotel
2.	Is Canceled	Categorical	Represents whether a booking was canceled (1) or not (0)
3.	Lead Time	Integer	Passed days between PMS and arrival
4.	Arrival Date Year	Integer	Year of the arrival
5.	Arrival Date Month	Categorical	Month of the arrival
6.	Arrival Date Week Number	Integer	Week number of the arrival
7.	Arrival Date Day of Month	Integer	Date day of the arrival month
8.	Stays in Weekend Nights	Integer	Number of stayed nights in Saturdays or Sundays
9.	Stays in Week Nights	Integer	Number of stayed nights from Monday to Friday
10.	Adults	Integer	The amount of adults
11.	Children	Integer	Number of Children
12.	Babies	Integer	Number of babies
13.	Meal	Categorical	If meal was included: BB: Bed and Breakfast HB: Half Board FB: Full board
14.	Country	Categorical	Originating Country
15.	Market Segment	Categorical	The segment of Market: TA: Travel Agents TO: Tour Operators
16.	Distribution Channel	Categorical	Reservation distribution channel: TA: Travel Agents TO: Tour Operators
17.	Is Repeated Guest	Categorical	Represents whether a reservation made from a repeated guest (1) or not (0)
18.	Previous Cancellations	Integer	Number of previous cancellations from the same guest
19.	Previous Bookings Not Canceled	Integer	Number of previous reservations not canceled

20.	Reserved Room Type	Categorical	Type of room instead of the actual number for ethical reasons
21.	Assigned Room Type	Categorical	The type of assigned room
22.	Booking Changes	Integer	Number of changes in the reservation
23.	Deposit Type	Categorical	Represents if the guest deposited a prepayment
24.	Agent	Categorical	The ID of the travel agency
25.	Company	Categorical	The ID of the company that proceeded with the reservation
26.	Days In Waiting List	Integer	Number of days that the reservation was on the waiting list until it was confirmed
27.	Costumer Type	Categorical	Type of costumer: Contract: the booking resulting by contract Group: the booking is for a group Transient: the booking was not made by contract or group Transient-party: the booking is transient but is associated with other transient reservation
28.	ADR	Numeric	$\text{Average Daily Rate} = \frac{\text{Rooms Revenue}}{\text{Number of booked rooms}}$
29.	Required Car Parking Spaces	Integer	Number of cars needed for parking spots
30.	Total of Special Requests	Integer	Number of special request made the reservation
31.	Reservation Status	Categorical	Canceled: Reservation was canceled Check-out: Guest has departed No-show: Costumer didn't make the check-in
32.	Reservation Status Date	Date	The date that costumer canceled the reservation or made the check out

Appendix B: Insert dataset in R

Code for downloading the necessary libraries (TidyVerse and ggplot2) as well as for inserting the dataset in R. In addition, by looking the dataset's format we can observe that some of the qualitative variables were wrongly categorized as measured ones so they had to be changed in order to achieve accurate visualization insights.

```
#### installation of the library for ggplot ####
install.packages("tidyverse")
library(tidyverse)
library(ggplot2)

#### enter dataset in R ####
library(readr)
hb <- read_csv("hotel_bookings.csv")
hb<-as.data.frame(hb)
View(hb)
attach(hb)

#observe whether the variables are categorized correctly to measured and
#factored or not
str(hb)
hb$is_canceled<-as.factor(is_canceled)
hb$is_repeated_guest<-as.factor(is_repeated_guest)
```

Appendix C: Plot – Reservation status of each hotel

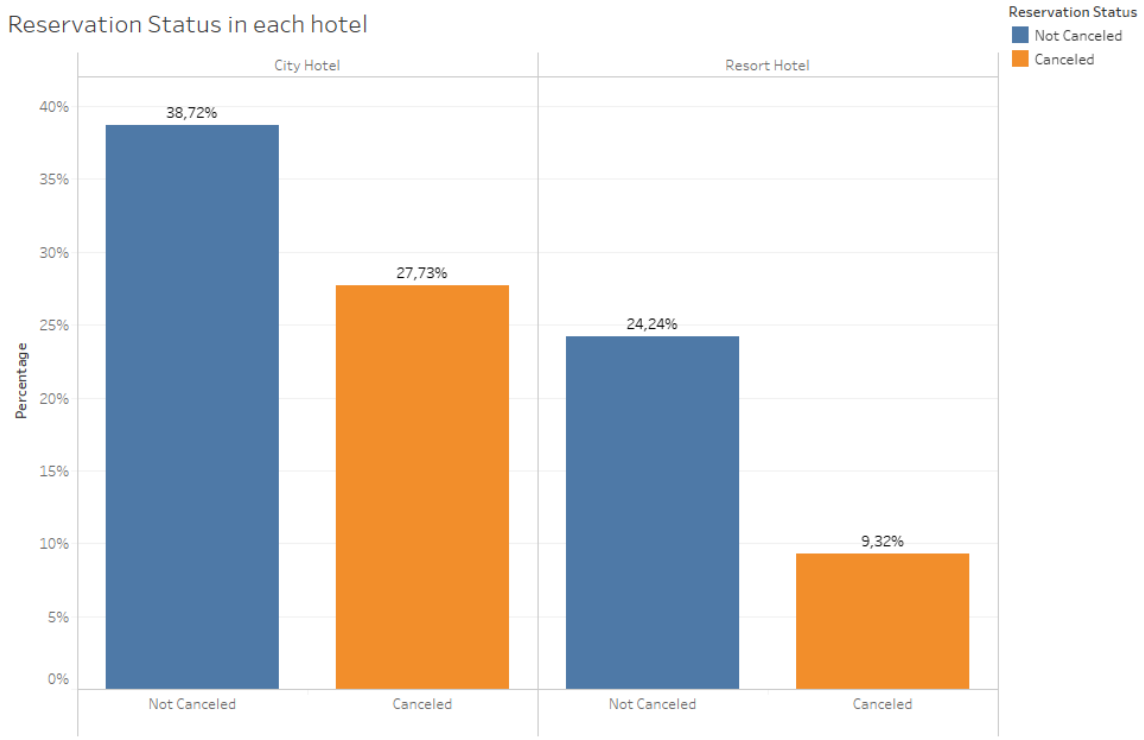

Figure 1. Booking Status for each hotel (using Tableau)

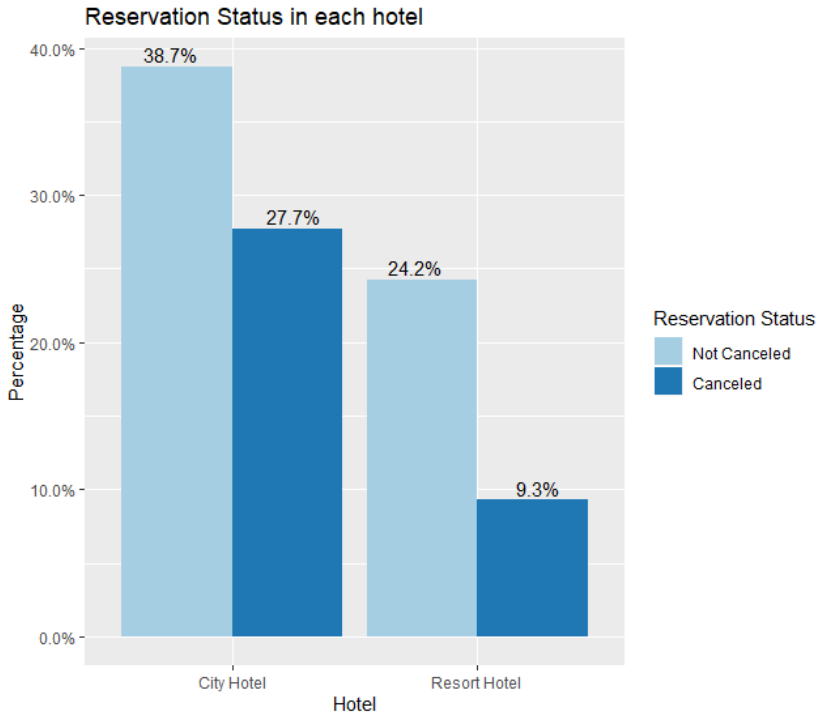

Figure 2. Reservation status for each hotel (using R)

Code 1. Code in R for figure 2

```
##### plot - Reservation Status ####

#preparation of plot data
temp<-factor(is_canceled,labels = c("Not Canceled","Canceled"))

#creation of barplot
ggplot(data=hb,
aes(
x=hotel,
y=prop.table(stat(count)),
fill=factor(temp),
label=scales::percent(prop.table(stat(count)))
))+
geom_bar(position = position_dodge())+
geom_text(stat="count",position = position_dodge(1),vjust=-
0.3,hjust=0.5)+
scale_fill_brewer(palette="Paired")+
scale_y_continuous(labels = scales::percent)+
labs(title = "Reservation Status in each hotel", x="Hotel", y="Percent-
age", fill="Reservation Status",
  levels=c("Not","Yes"))
```

Appendix D: Plot – Stays in Weeknights per Country for each hotel

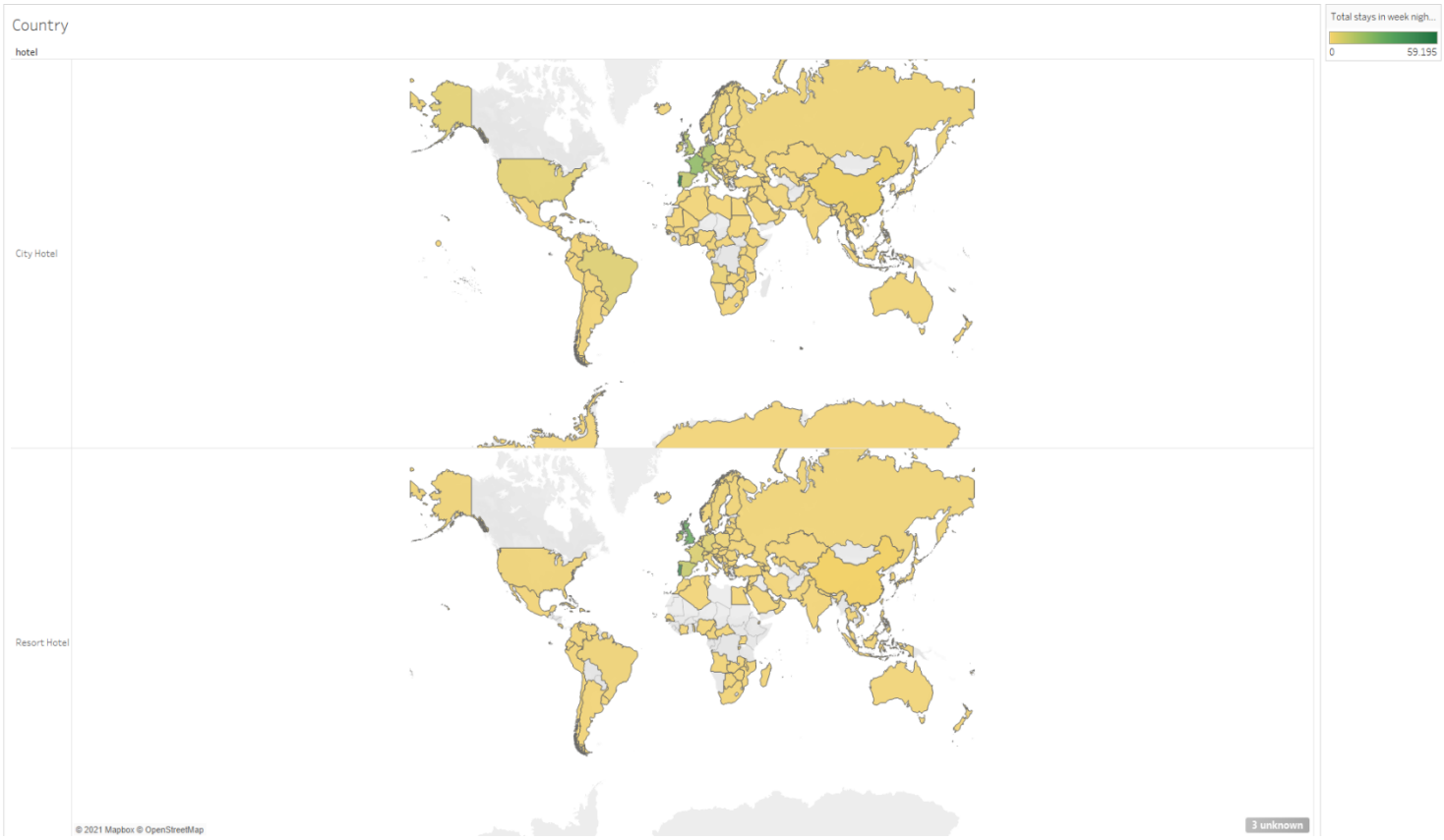

Figure 3. Stays in weeknights for each hotel per country (using Tableau)

Appendix E: Plot - Average Daily Rate for each Costumer Type

- Aggregated Graph

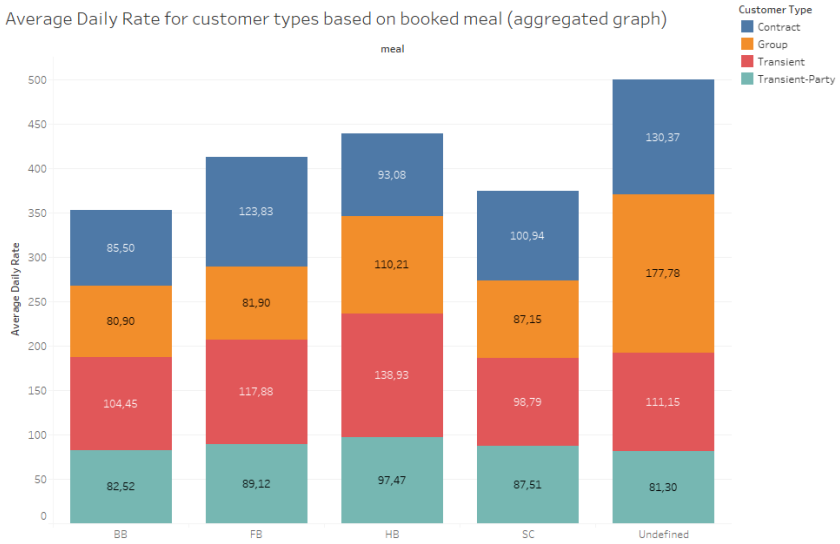

Figure 4. Average Daily Rate for all costumers' types based on booked meal (using Tableau)

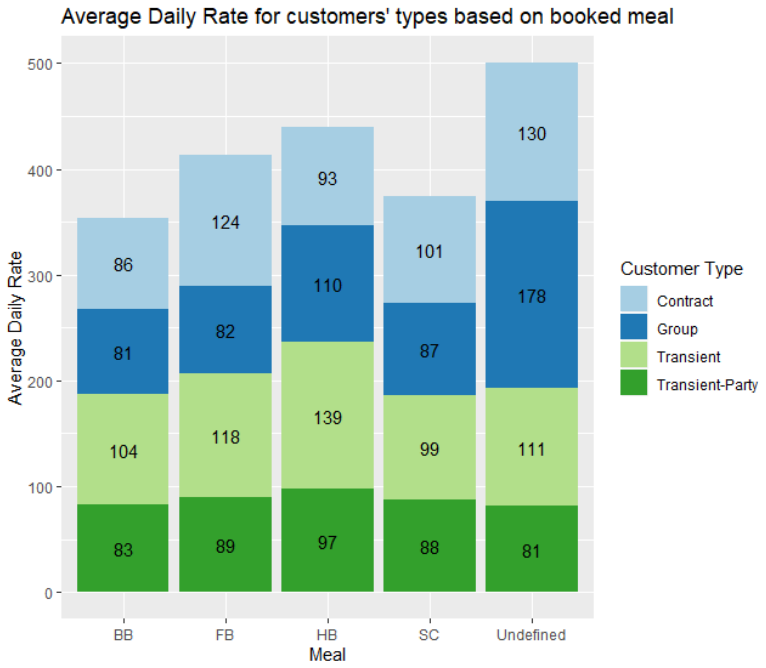

Figure 5. Average Daily Rate for all costumers' types based on booked meal (using R)

Code 2. Code in R for figure 5

```
#####plot - Avarage Daily Rate for each costumer type based on booked
#meal ####

#creating a temporary variable in order present the average value of adr

temp1 = aggregate(list(adr = hb$adr), list(meal = factor(hb$meal),
customer_type=hb$customer_type), mean)

#creation of graph with all costumer types simultanusely

ggplot(temp1,
aes(x=meal,
y= adr, fill=customer_type)) +
geom_bar(stat="identity") +
geom_text(aes(label=round(adr)),position = position_stack(vjust = 0.5))+
scale_fill_brewer(palette="Paired")+
labs(title="Average Daily Rate for customers' types based on booked
meal",
y = "Average Daily Rate",
x="Meal",
fill="Customer Type")
```

- Separate graph for each costumer Type

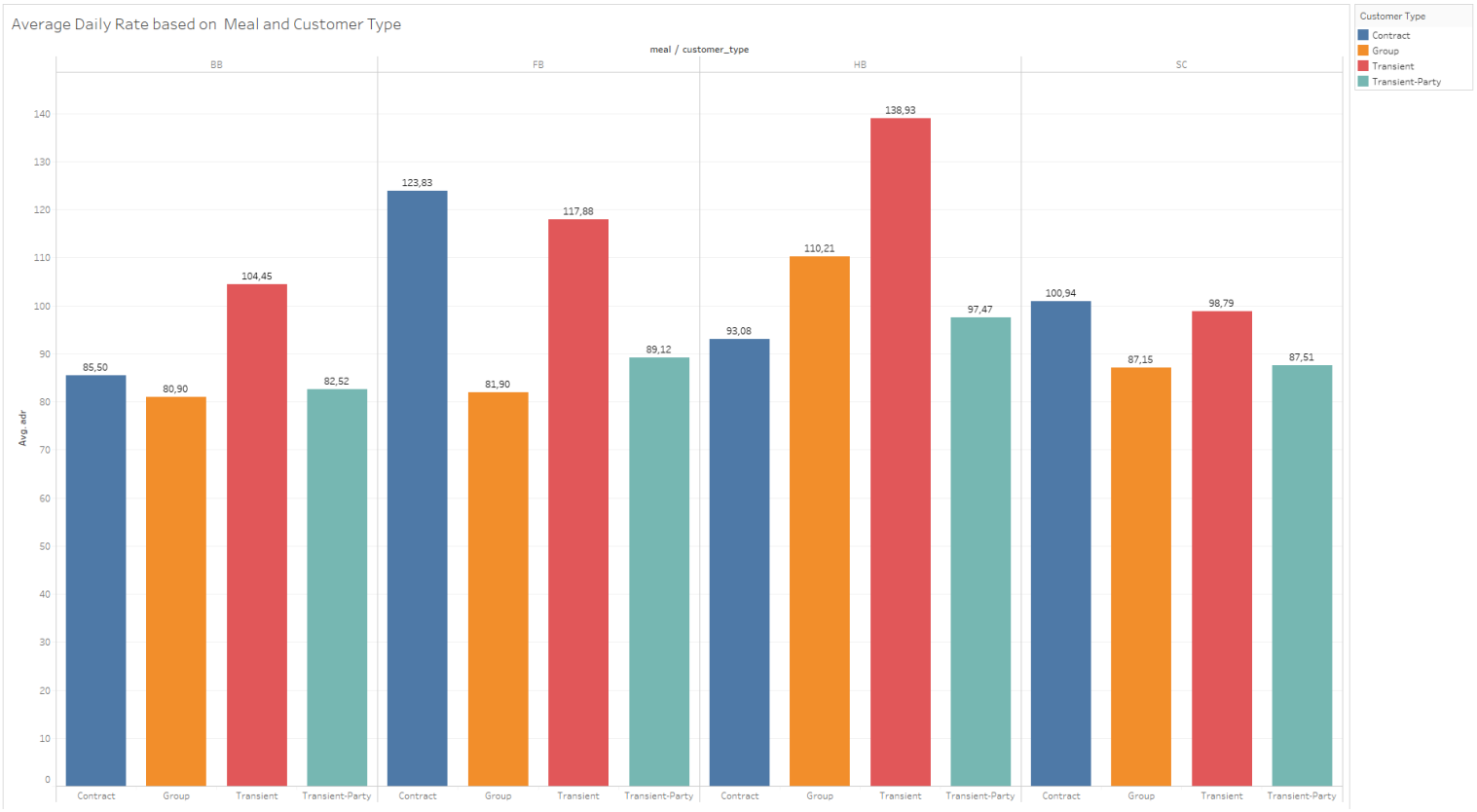

Figure 6. Average Daily Rate for each Costumer Type based on booked Meal (using Tableau)

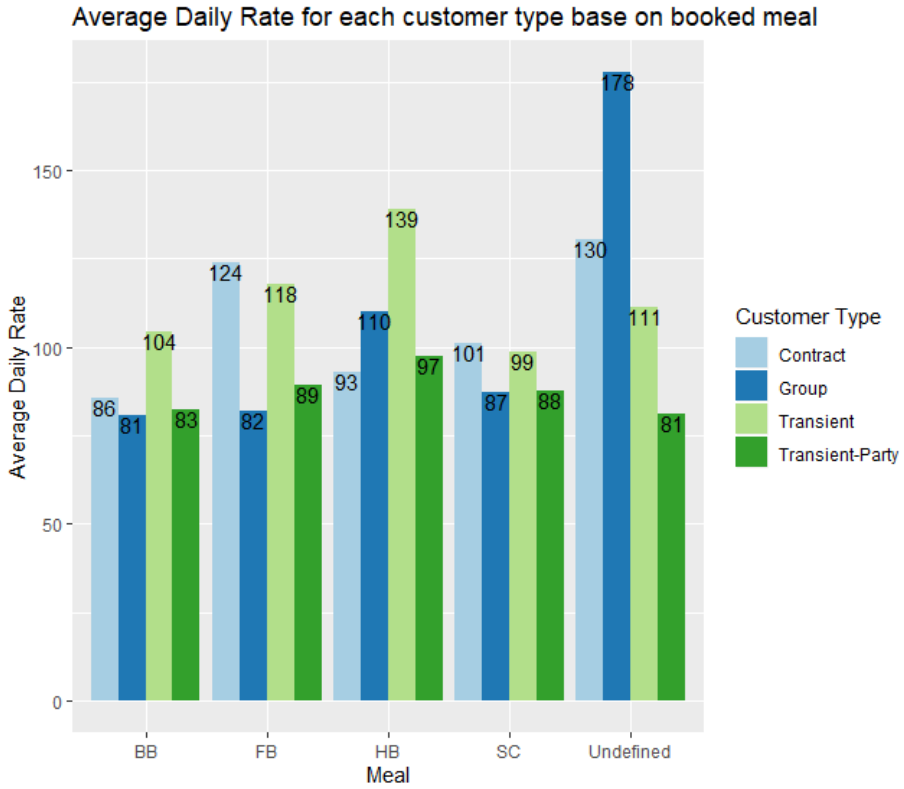

Figure 7. Average Daily Rate for each Costumer Type based on booked Meal (using R)

Code 3. Code in R for figure 7

```
#creation of graph for each costumer type (separately)

ggplot(temp1,
aes(x=meal,
y= adr, fill=customer_type)) +
geom_bar(stat="identity",
position = position_dodge()) +
geom_text(aes(label=round(adr)),position = posi-
tion_dodge(width=0.9),hjust=0.5,vjust=1)+
scale_fill_brewer(palette="Paired")+
labs(title="Average Daily Rate for each customer type base on booked
meal",
y = "Average Daily Rate",
x="Meal",
fill="Customer Type")
```

Appendix F: Plot – Average Daily Rate for Repeated Guests

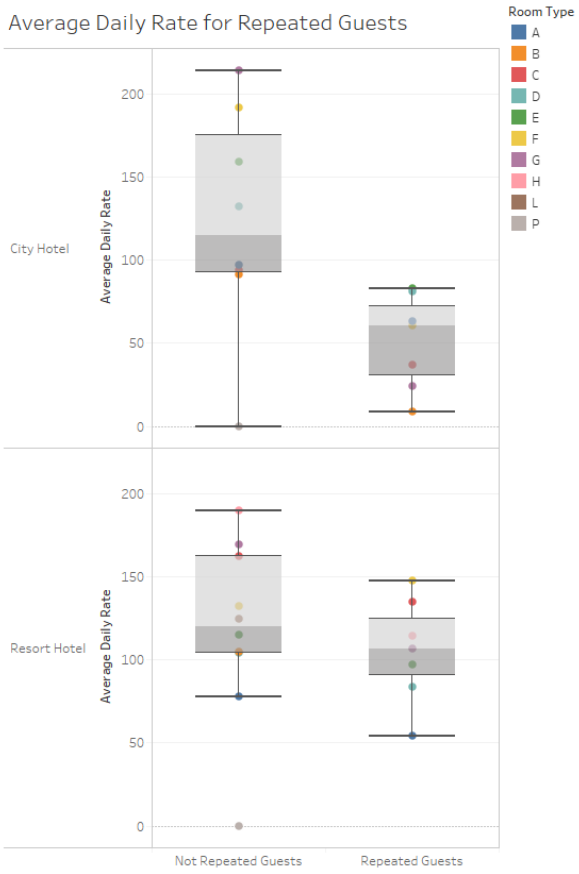

Figure 8. Average Daily Rate for repeated and not repeated guests (using Tableau)

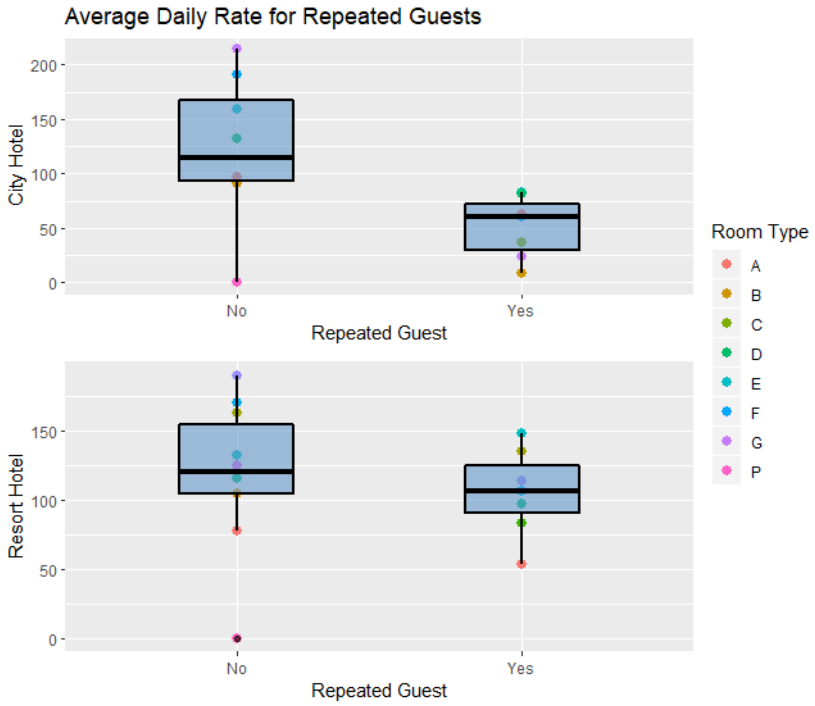

Figure 9. Average Daily Rate for repeated and not repeated guests (using R)

Code 4. Code in R for figure 9

```
##### plot - Average Daily Rate for Repeated Guests #####
library(dplyr)
library(ggpubr)

#preparation of plot data
boxplot.data<-group_by(hb,reserved_room_type,is_repeated_guest,hotel)
%>% summarise(avg.adr=mean(adr))
head(boxplot.data)

#creation of plot for City Hotel
boxplot.data.ch<-filter(boxplot.data,hotel=="City Hotel")
bp1<-ggplot(boxplot.data.ch,
aes(
factor(is_repeated_guest,labels=c("No","Yes")),
avg.adr,
color=reserved_room_type))+
geom_jitter(size=2.5,position=position_dodge(0))+
guides(col=guide_legend("Room Type"))+
geom_box-
plot(fill="#6699CC",alpha=0.6,colour="black",size=0.8,width=0.4)+
labs(x="Repeated Guest",y="City Hotel",title = "Average Daily Rate for
Repeated Guests")

#creation of plot for Resort Hotel
boxplot.data.rh<-filter(boxplot.data,hotel=="Resort Hotel")
```

```
bp2<-ggplot(boxplot.data.rh,
aes(
factor(is_repeated_guest,labels=c("No","Yes")),
avg.adr,
color=reserved_room_type))+
geom_jitter(size=2.5,position=position_dodge(0))+
guides(col=guide_legend("Room Type"))+
geom_box-
plot(fill="#6699CC",alpha=0.6,colour="black",size=0.8,width=0.4)+
labs(x="Repeated Guest",y="Resort Hotel")

#creation of aggregated graph
  ggarrange(bp1,bp2,nrow  =  2,ncol=1,common.legend  =  TRUE,legend  =
"right")
```

Appendix G: Plot – Cancellation Rate for each year

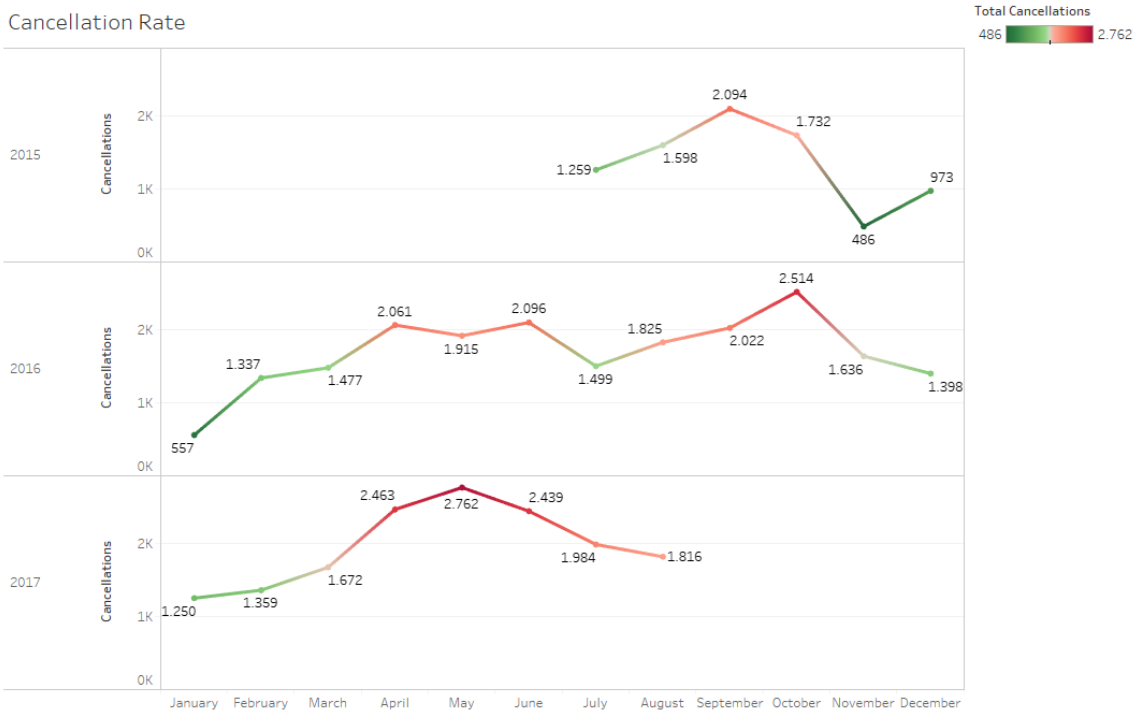

Figure 10. Cancellation Rate for each year (using Tableau)

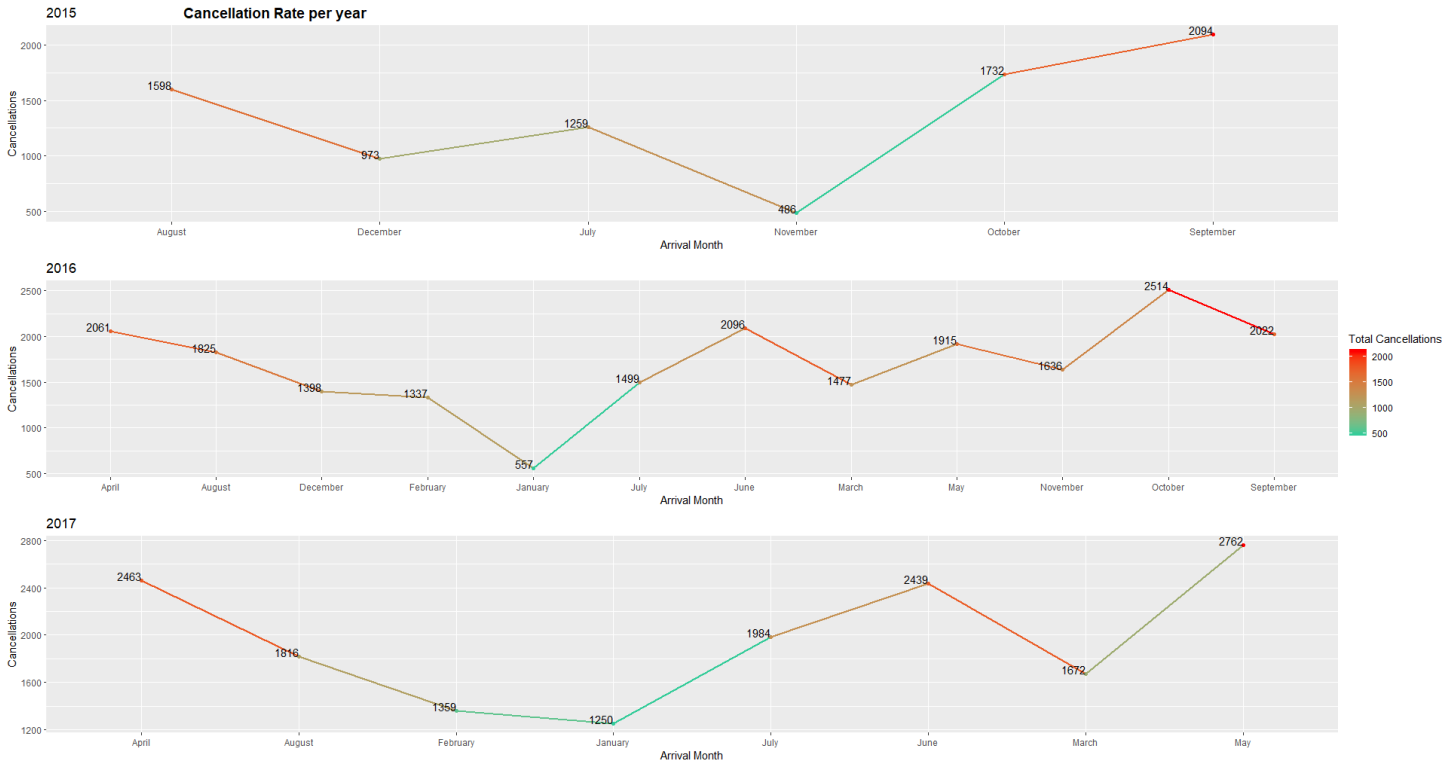

Figure 11. Cancellation Rate for each hotel (using R)

Code 5. Code in R for figure 11

```
######## plot - Cancellation Rate per year ######
library(dplyr)
library(ggpubr)

#preparation of plot data
temp2<-hb %>%
group_by(is_canceled,arrival_date_year,arrival_date_month) %>%
summarize(n=n())
temp2<-filter(temp2,is_canceled==1)
head(temp2)

###2015#
line.plot.data.2015<-filter(temp2, arrival_date_year==2015)
head(line.plot.data.2015)
p2015<-ggplot(line.plot.data.2015,aes(
x=arrival_date_month,
y=n,
col=n,
group=1
```

```
))+
scale_colour_gradient(low="#33CC99",high="red",name="Total Cancella-
tions")+
geom_line(size=1)+
geom_point()+
geom_text(aes(label=n,vjust=0,hjust=1),colour="black")+
labs(x="Arrival Month", y="Cancellations", title = "2015")

###2016
line.plot.data.2016<-filter(temp2, arrival_date_year==2016)
head(line.plot.data.2016)
p2016<-ggplot(line.plot.data.2016,aes(
x=arrival_date_month,
y=n,
col=n,
group=1
))+
scale_colour_gradient(low="#33CC99",high="red",name="Total Cancella-
tions")+
geom_line(size=1)+
geom_point()+
geom_text(aes(label=n,vjust=0,hjust=1),colour="black")+
labs(x="Arrival Month", y="Cancellations", title = "2016")

###2017
line.plot.data.2017<-filter(temp2, arrival_date_year==2017)
p2017<-ggplot(line.plot.data.2017,aes(
x=arrival_date_month,
y=n,
col=n,
group=1
))+
scale_colour_gradient(low="#33CC99",high="red",name="Total Cancella-
tions")+
geom_line(size=1)+
geom_point()+
geom_text(aes(label=n,vjust=0,hjust=1),colour="black")+
labs(x="Arrival Month", y="Cancellations", title = "2017")

##creation of line plot
ggarrange(p2015,p2016,p2017,ncol=1,nrow=3,labels="Cancellation Rate per
year",hjust=-1,common.legend = TRUE,legend =
"right")
```

Figure 12. Cancellation Rate per year for each hotel (using Tableau)

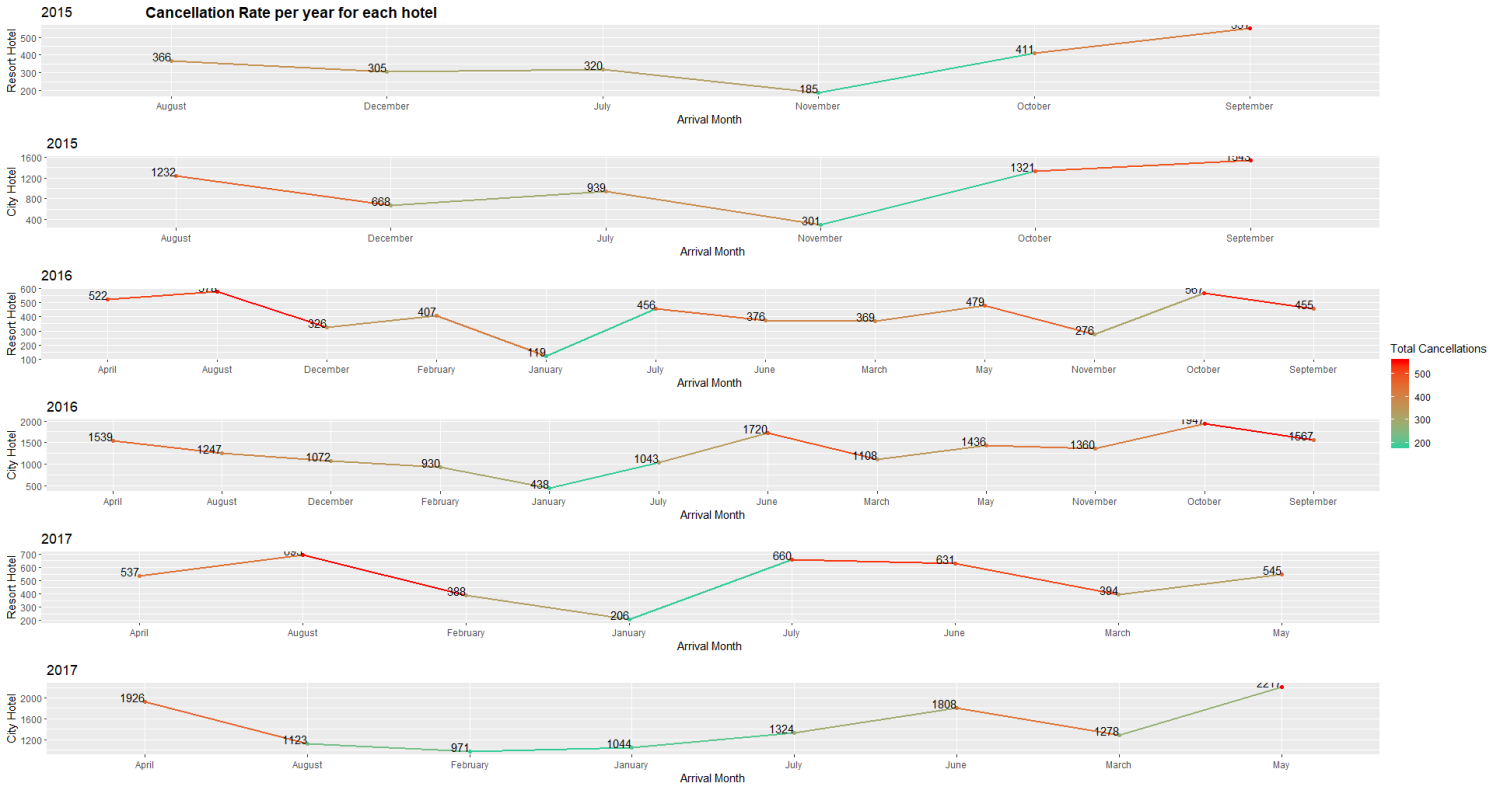

Figure 13. Cancellation Rate per year for each hotel (using R)

Code 6. Code in R for figure 13

```
#######For each hotel

temp3<-hb %>%
group_by(is_canceled,arrival_date_year,arrival_date_month,hotel) %>%
summarize(n=n())
temp3<-filter(temp3,is_canceled==1)
head(temp3)

###2015 for resort hotel#
line.plot.data.2015.rh<-filter(temp3,hotel=="Resort Hotel" & arri-
val_date_year==2015)
head(line.plot.data.2015.rh)
p2015.rh<-ggplot(line.plot.data.2015.rh,aes(
x=arrival_date_month,
y=n,
col=n,
group=1
))+
scale_colour_gradient(low="#33CC99",high="red",name="Total Cancella-
tions")+
geom_line(size=1)+
```

```
geom_point()+
geom_text(aes(label=n,vjust=0,hjust=1),colour="black")+
labs(x="Arrival Month", y="Resort Hotel", title = "2015")

###2016 for resort hotel#
line.plot.data.2016.rh<-filter(temp3,hotel=="Resort Hotel" & arri-
val_date_year==2016)
head(line.plot.data.2016.rh)
p2016.rh<-ggplot(line.plot.data.2016.rh,aes(
x=arrival_date_month,
y=n,
col=n,
group=1
))+
scale_colour_gradient(low="#33CC99",high="red",name="Total Canella-
tions")+
geom_line(size=1)+
geom_point()+
geom_text(aes(label=n,vjust=0,hjust=1),colour="black")+
labs(x="Arrival Month", y="Resort Hotel", title = "2016")

###2017 for resort hotel#
line.plot.data.2017.rh<-filter(temp3,hotel=="Resort Hotel" & arri-
val_date_year==2017)
head(line.plot.data.2017.rh)
p2017.rh<-ggplot(line.plot.data.2017.rh,aes(
x=arrival_date_month,
y=n,
col=n,
group=1
))+
scale_colour_gradient(low="#33CC99",high="red",name="Total Canella-
tions")+
geom_line(size=1)+
geom_point()+
geom_text(aes(label=n,vjust=0,hjust=1),colour="black")+
labs(x="Arrival Month", y="Resort Hotel", title = "2017")

###2015 for City hotel#
line.plot.data.2015.ch<-filter(temp3,hotel=="City Hotel" & arri-
val_date_year==2015)
head(line.plot.data.2015.ch)
p2015.ch<-ggplot(line.plot.data.2015.ch,aes(
x=arrival_date_month,
y=n,
col=n,
group=1
))+
scale_colour_gradient(low="#33CC99",high="red",name="Total Canella-
tions")+
geom_line(size=1)+
geom_point()+
geom_text(aes(label=n,vjust=0,hjust=1),colour="black")+
labs(x="Arrival Month", y="City Hotel", title = "2015")

###2016 for City hotel#
line.plot.data.2016.ch<-filter(temp3,hotel=="City Hotel" & arri-
val_date_year==2016)
head(line.plot.data.2016.ch)
p2016.ch<-ggplot(line.plot.data.2016.ch,aes(
```

```
x=arrival_date_month,
y=n,
col=n,
group=1
))+
scale_colour_gradient(low="#33CC99",high="red",name="Total Canella-
tions")+
geom_line(size=1)+
geom_point()+
geom_text(aes(label=n,vjust=0,hjust=1),colour="black")+
labs(x="Arrival Month", y="City Hotel", title = "2016")

###2017 for City hotel#
line.plot.data.2017.ch<-filter(temp3,hotel=="City Hotel" & arri-
val_date_year==2017)
head(line.plot.data.2017.ch)
p2017.ch<-ggplot(line.plot.data.2017.ch,aes(
x=arrival_date_month,
y=n,
col=n,
group=1
))+
scale_colour_gradient(low="#33CC99",high="red",name="Total Canella-
tions")+
geom_line(size=1)+
geom_point()+
geom_text(aes(label=n,vjust=0,hjust=1),colour="black")+
labs(x="Arrival Month", y="City Hotel", title = "2017")

#creation of the line plot for each hotel (per year)
ggarange(p2015.rh,p2015.ch,p2016.rh,p2016.ch,p2017.rh,p2017.ch,ncol=1,
nrow=6,labels ="Cancellation Rate per year for
           each hotel",common.legend = TRUE,legend = "right")
```