

# DATA VISUALIZATION PROJECT

*Team members:*

Name
Rajab
Diaa
Yusuf
Ziad

In this Project we will talk about hotel booking demand dataset. We have divided our Report into 4 steps

1. Intro about our Dataset: In which we will talk a little bit about our dataset
2. Problem which Is needed to be solved
3. Questions formed to answer the Problem
4. Some solution trying to solve these problems

## **1-Brief About Data set**

This data article describes two data sets wit hotel demand data. One of the hotels(H1) is a resort hotel and the other is a city hotel (H2). Both data sets share the same structure, with 31 variables describing the 40,060 observations of H1 and 79,330 observations of H2. Each observation represents a hotel booking. Both datasets comprehend bookingsduetoarrivebetweenthe1stofJulyof2015 and the 31<sup>st</sup> of August 2017, including bookings that effectively arrived and bookings that were canceled. Since this is hotel real data, all data elements pertaining hotel or costumer identification were deleted. Due to the scarcity of real business data for scientific and educational purposes, these datasets can have an important role for research and education in revenue management, machine learning, or datamining, as well as in other fields. And we choose the aim of Descriptive analysis to talk about in our Project.

## **How Data was collected**

## 1. Data

In tourism and travel related industries, most of the research on Revenue Management demand forecasting and prediction problems employ data from the aviation industry, in the format known as the Passenger Name Record (PNR). This is a format developed by the aviation industry [2]. However, the remaining tourism and travel industries like hospitality, cruising, theme parks, etc., have different requirements and particularities that cannot be fully explored without industry's specific data. Hence, two hotel datasets with demand data are shared to help in overcoming this limitation.

The datasets now made available were collected aiming at the development of prediction models to classify a hotel booking's likelihood to be canceled. Nevertheless, due to the characteristics of the variables included in these datasets, their use goes beyond this cancellation prediction problem.

One of the most important properties in data for prediction models is not to promote leakage of future information [3]. In order to prevent this from happening, the timestamp of the target variable must occur after the input variables' timestamp. Thus, instead of directly extracting variables from the bookings database table, when available, the variables' values were extracted from the bookings change log, with a timestamp relative to the day prior to arrival date (for all the bookings created before their arrival date).

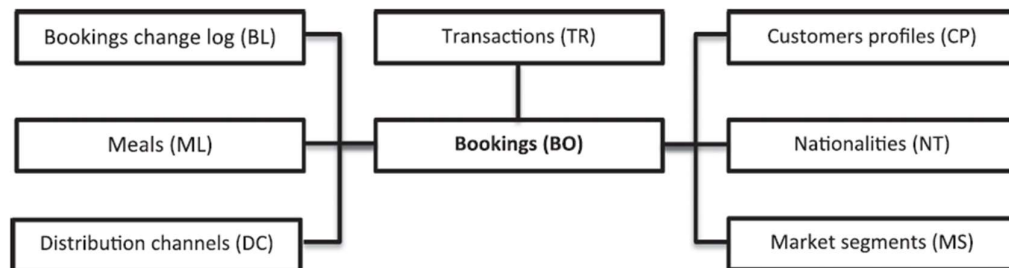


Fig. 1. Diagram of PMS database tables where variables were extracted from.

And now let's talk a little bit and **define some expressions about the business** of hotels:

### - Hotel Market Segmentation

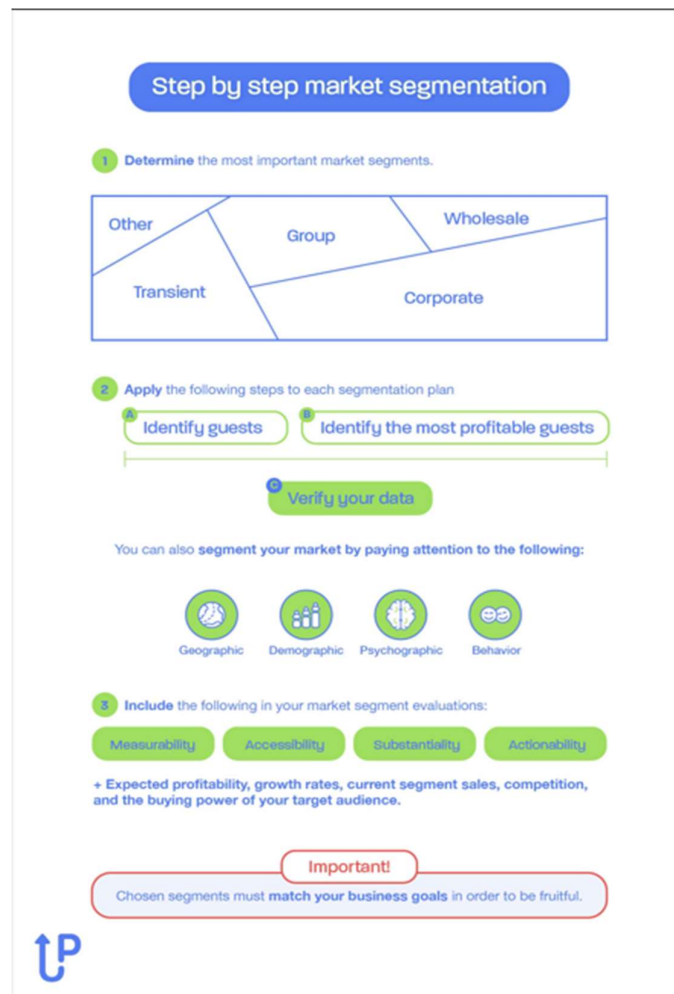
Hotel market segmentation is a strategy that allows hotel owners to better understand the purpose that drives the consumers and identify each trip's goal: either leisure or business. Many hotel owners wrongly believe that the price dictates the market segmentation rules, but it doesn't.

**Market segmentation is also used to determine clear differences between group and individual businesses.** If it's done the right way, market segmentation can be extremely useful in recognizing new trends in the market and hotel industry, staying up to date with the latest events in your business, getting ahead of the

competition curve, and providing your consumers with exceptional user/customer experience.

If you're running a luxury hotel, market segmentation can help you get ahead of the tourism game by knowing everything about your target market. More importantly, **it can help you identify the crucial hotel business concerns**, such as:

- Length of stay
- Day of weeks stays
- Total revenue per room
- Total revenue per client
- Booking lead time
- Cancellation %
- No show ratio



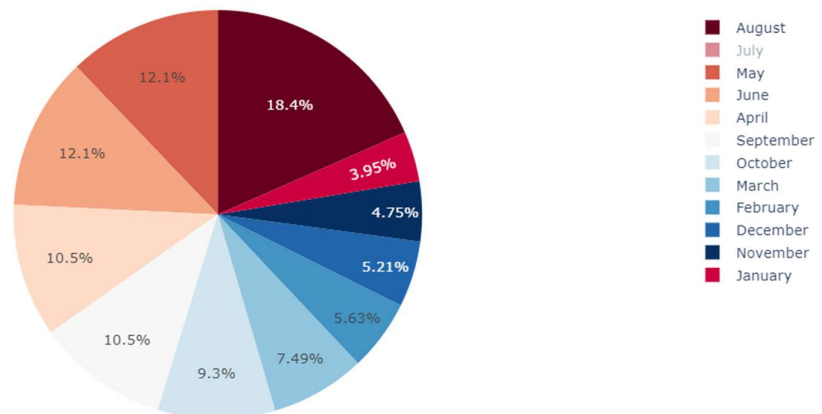
## Hotel Transactions

means a hotel booking/reservation per traveler, made, regardless of whether, the traveler uses the hotel reservation or not.

To clear some stuffs we will ask question then we answer by graph and comments.  
Now let's begin our **visualization journey**

**1-what is the most commn month that has highst average daily rate?**

Average daily rate by month



Idiom: Pie Chart

Mark: interlocking area

Channel:2D area

Data: 1 categorical Key (arrival date of month), 1 quantitative value = Average daily rate

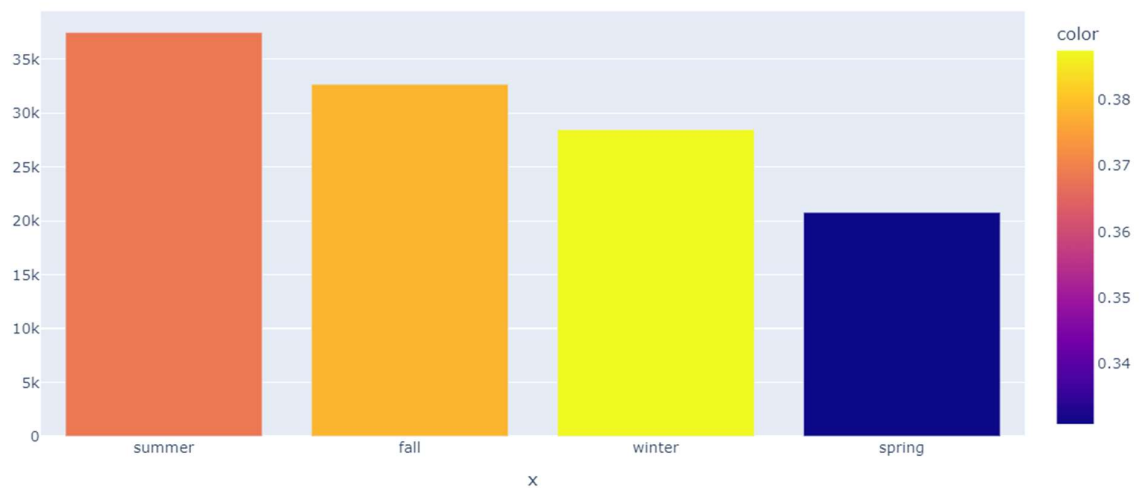
As we can see August as the highest percentages of average daily rate that means people mostly to travel to our hotels in August. also we have noticed that the smallest percentages in January.

how is this graph can help us?

we can make offers January to encourage travelers also we can study why people don't prefer to come in January

**2. Which season has the most reservations and what is the most season has huge cancellation?**

Seasons and percentages of cancelation



Idiom: Bar Chart

Mark lines

Channels: length to express quant value, spatial regions: one per mark and color for average cancelation

Derived attribute: season by combine months in arrival date of month attribute

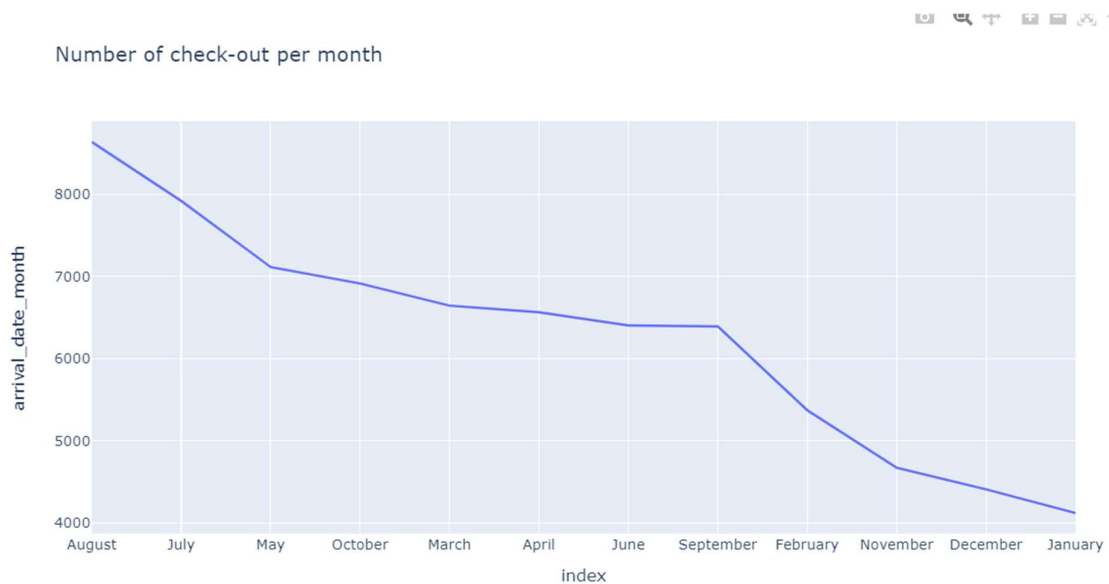
Data: 1 categorical attribute (Seasons), 1 quantitative value = frequency of each season

As you can see summer has the most frequency of reservation is that because our hotels in Portuguese and Portuguese is a cost country.

also, we have noticed that the most season average cancellation is winter meanwhile spring is the smallest average cancellation.

is that helping us by discovering reasons why people cancel in winter and fixing this can help us.

**3-Which months has most frequent check-out reservation?**



Idiom: line Chart

Mark: points and line connection marks between them

Channels: aligned lengths to express quant value and separated and ordered by key attribute into

as you can see here August has the most frequent check out.

That means the people that make reservation in August come and they are not cancelled the most. Even though May in spring but because the third most frequently checked out that because there is a lot of festivals in Portuguese in May.

The most important point we want to care about is to decrease expenses

Increase the profit from the two hotels so we have made a lot of visualization to see the relation between factors based on that stake holder will take his action.

Here we are in the discovery step of the data we want to know each

- First, we want to have a look about some attributes as Market Segments, Meal, percentage isCanceled and Customer Type to know initial values and have a good background about the data.

**In this section we have used idiom : Pie chart**

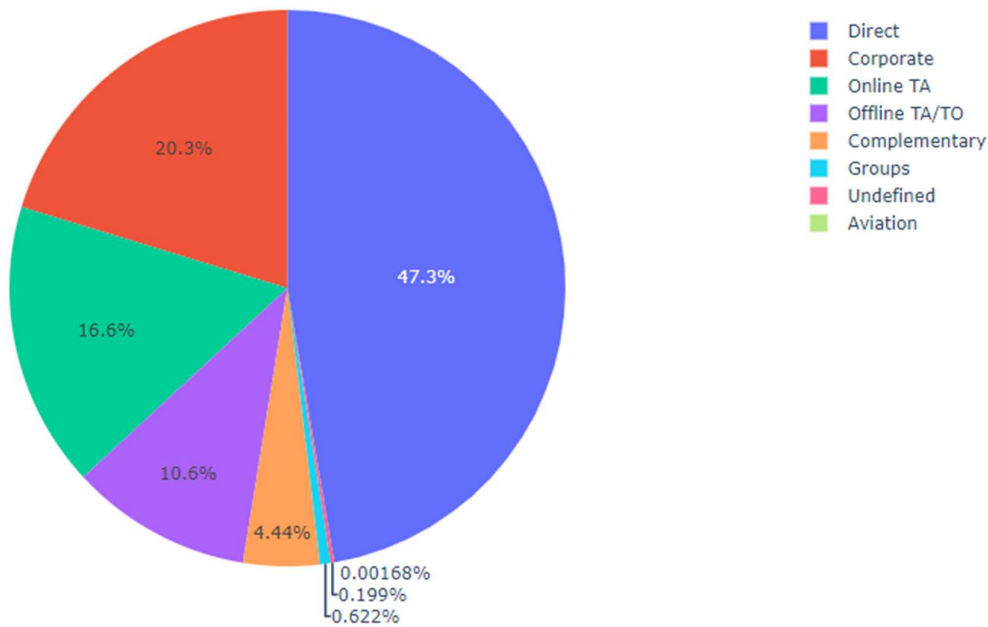
It is an interlocking area marks with angle channel

Pie chart is a kind of chart that takes a key attribute and quantative value

**Market segmentation:**

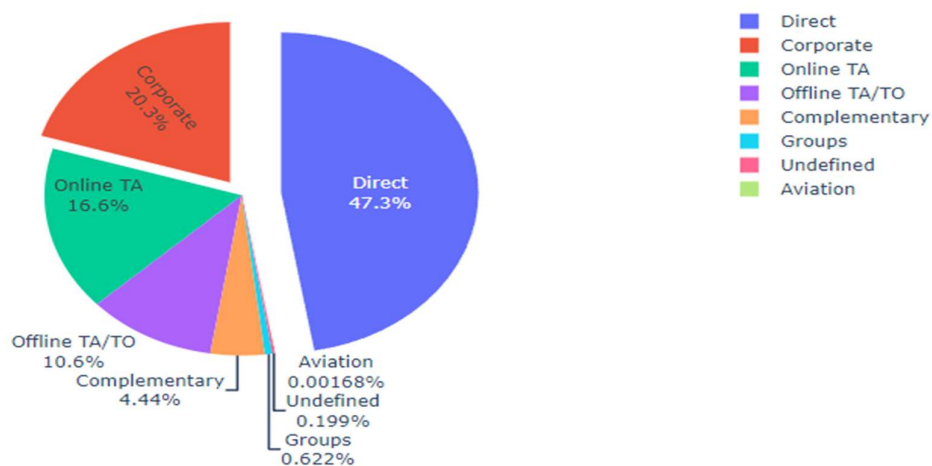
- Using Plotly Express(ex) library we draw this pie chart





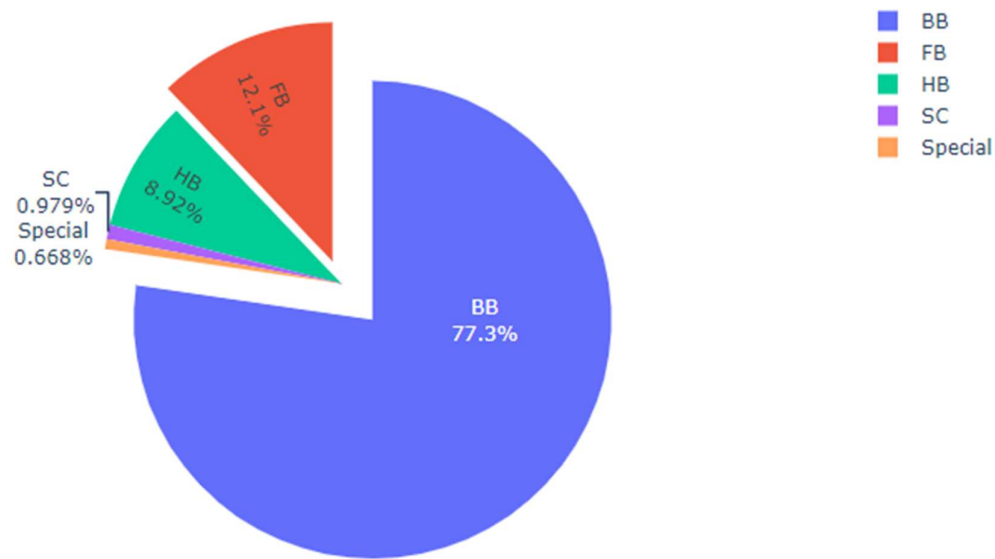
From the previous graph we notice that most of our customers books directly which means the hotel may have a good reputation in the second place we have Corporate section so if we are going to have more deals and offers for Companies section in the future will be better in the third place we have online and then off line Travel Agents and Operators which depends on feedbacks of customers , offline & online ads and offers which attract people then at last we have groups which depends on seasonal events , holidays , parties ...etc. At last, we have aviation which are waiting for their flights and Undefined which haven't filled segmentation part in the form.

- And here is another pie chart but this time using Graph Objects(go) library we draw this graph



**Meals:**

- Here is another pie chart using Plotly Graph Objects(go) library we draw this graph separating it's slices by attributes freq



---

Special/SC – no meal package.

BB – Bed & Breakfast.

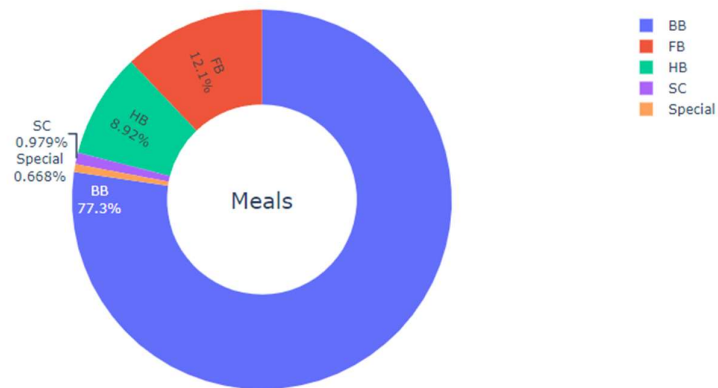
HB – Half board (breakfast and one other meal – usually dinner);

FB – Full board (breakfast, lunch and dinner)

From this pie chart we notice that more than 75% of food section is from BB type which means most of our customer among customer type and Market Segmentation spend most of their time outside the hotel as we have mention in market segmentation, they may be Tourists, Business Agents, Event Coordinator .... etc. So, they take they spend Morning time on planning and having their breakfast and get back to the hotel in the Evening to rest having their Bed meal and then sleep So we suggest directing our attention to this type of meal and try to offer a variety of this meal in its menu . Second and Third places is for FB and HB for customers who stay most of their time in that hotel may be they are waiting for aviation or waiting for someone so we want to make a lot of offers for these two types to attract more customer

- Here is another pie chart using Plotly Graph Objects(go) library we draw this graph as Donut Graph

Meals reivew

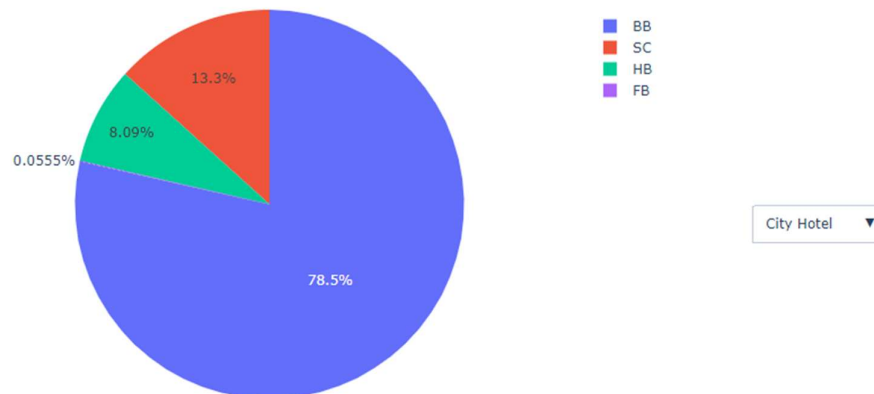


### ***We are still in Meal's section***

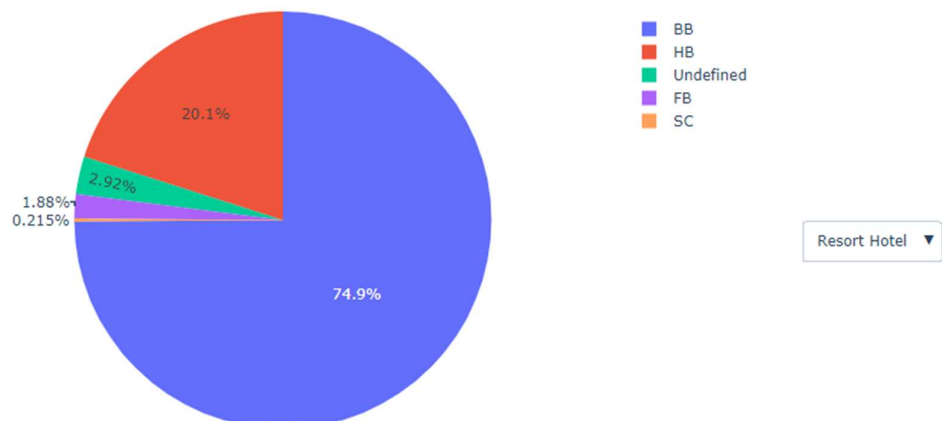
***We want to make an interactive graph representing percentages each type of meals in each hotel as we have mentioned before this data is merged from two hotels resort hotel and City Hotel***

- Here is another pie chart using Plotly Graph Objects(go) library, but this time add some interactivity to the graph separating it based on which hotel data come from

City Hotel

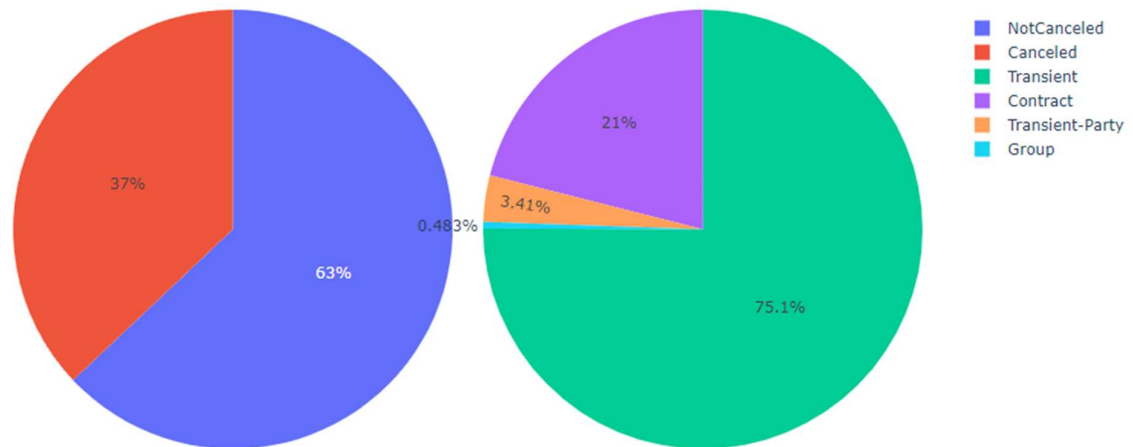


Resort Hotel



## - Using Plotly Graph Objects(go)

Pie Charts of Cancellation and Customer types



Contract – when the booking has an allotment or other type of contract associated to it.

Group – when the booking is associated to a group.

Transient – when the booking is not part of a group or contract and is not associated to other transient booking.

Transient-party – when the booking is transient, but is associated to at least other transient booking

From the pie chart above on the right side we see that most of our customers has booked as Transient nearly 75% which means they are not permanent customer if we manage to move them to the second place Contract we will make a great improvement and guarantee they are coming again.

From the pie chart above on the left side we notice the rate of cancellation and we can assure that most of our customers didn't cancel their booking but we must find a solution for persons who cancel it by making them leave guarantee, updating deposit time system....etc.

- Know we want to know which customer cancels the most

**In this section we have used idiom : Bar Chart and Stacked Bar Chart**

Bar: - It takes 1 categorical attribute and 1 qualitative attribuet

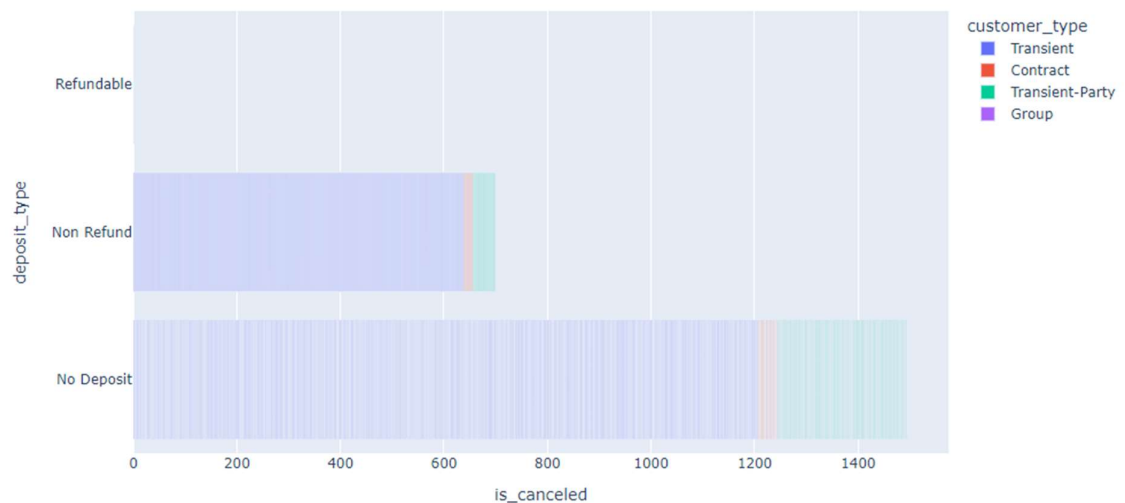
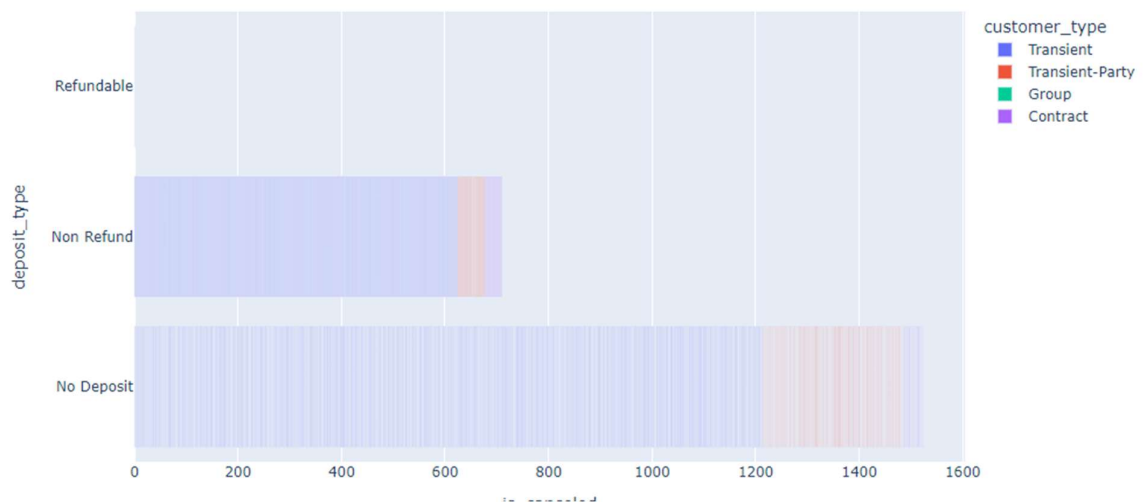
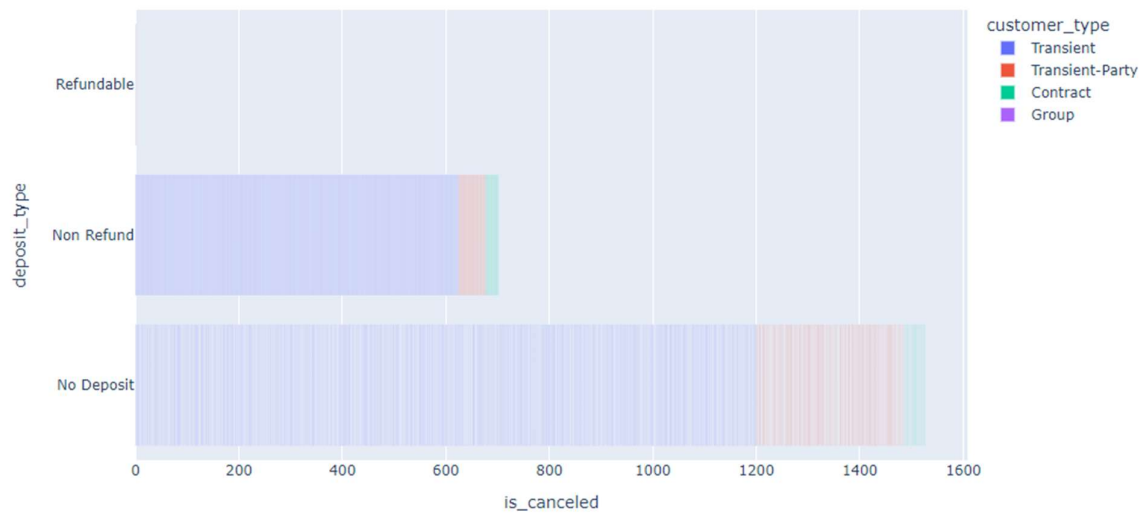
- Its mark is lines
- Task to compare, lookup values

Stacked Bar: - It takes 2 categorical attributes and 1 qualitative attribuet

- Its mark vertical stack of line marks, glyph: composite object, internal

## Cancellation, Deposit type, customer type and Relation between them:

- Using Plotly Express(ex) library we draw this stacked bar chart



In the previous images we recognize three stacked bar charts as the huge amount of

data set, which is 119390 rows, we have taken a sample 6000 rows as the computational power is not very high so results may be slightly change.

So, we want to know who cancels the most, his deposit and his type as customer:

After understanding the meaning of bar charts we see that most of booking cancellation came from Transient customer

In the 3 charts they are the customers who cancels the most and we see that they haven't leave any deposit then cancels which cause us a lot of expenses and may some other customer book in this time so as we have mention previously we must find a solution for this issue especially for Transient customers

We should also care about Deposit type as most of cancellation come from no leaving any guarantee which make it easier for any customer to cancel at any time and at the end more expenses

*what is the Percentage of booking in the hotel from Europe and Outside Eu so the hotel will know the market segmentation of customers?*

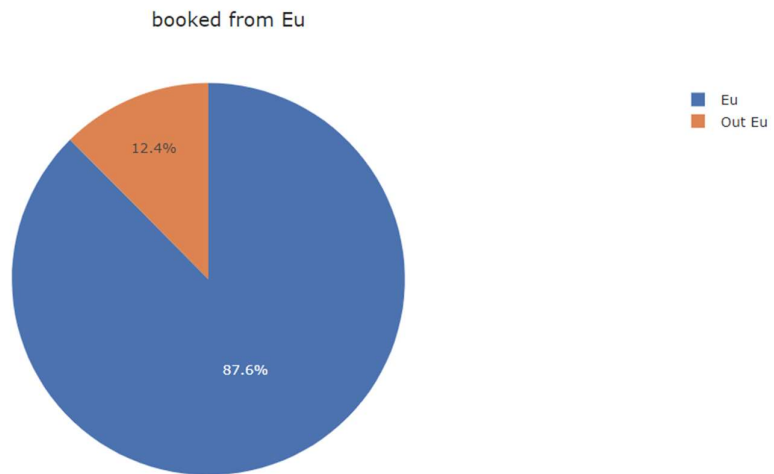
Pie Chart

Mark: interlocking area

Channel: 2D area

Derived Attribute: is EuCountry

Data: 1 categorical Key (Eu, outside Eu), 1 quantitative value = Number of Customers



The hotel has high Number of Customers in Europe

*What is the average cancellation between Countries in EU and outside Eu ? and what is the highest average of cancellation between countries, and if the season makes a difference?*

The idea to make hierarchal plot refers to:

parent (Compare average cancellation between Eu and Outside),

child (Compare average cancellation in countries for each region)

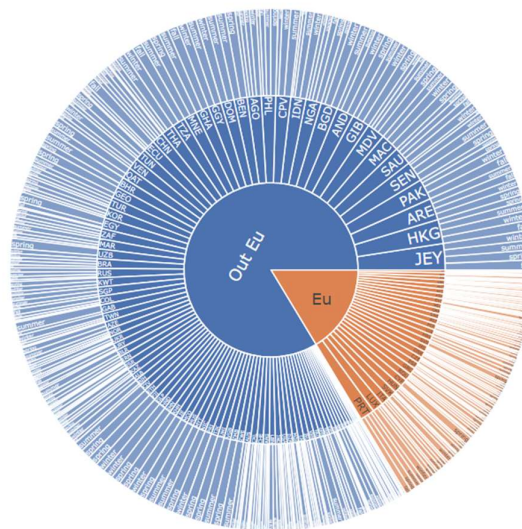
leaf node: (Compare average season cancellation)

sunburst chart:

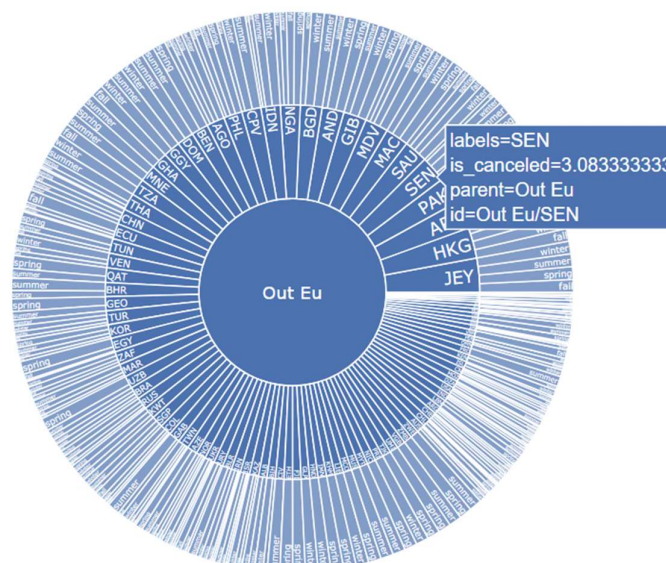
Mark: interlocking area

Channel: 2D area

Average of Hotel cancellation

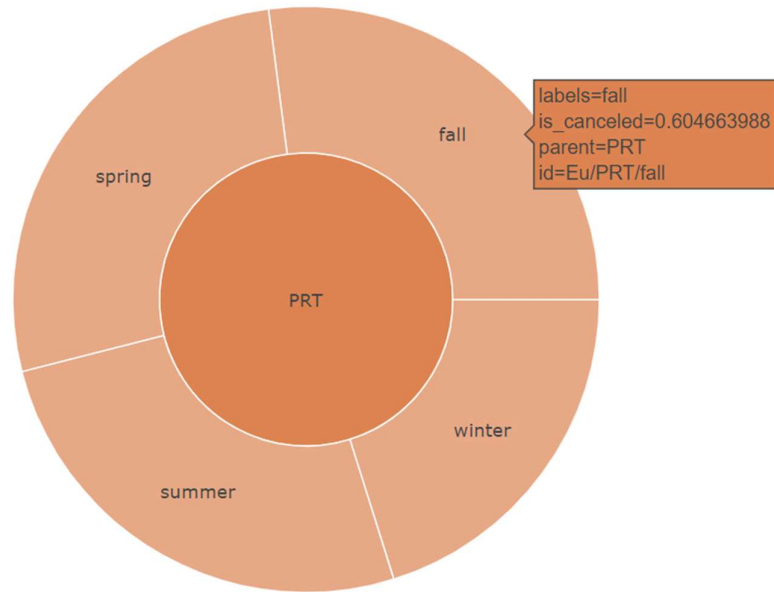


Average of Hotel cancellation





### Average of Hotel cancellation



Countries outside EU have an average cancellation reach to 127 , and Eu countries have an average of 26.

Portugal has most cancellation rate in Europe mostly in Spring and fall

this hotel has a high percentage of cancellations so it may have problems with pricing, location, or Competitors offer better offers

## Icicle Plot:

This icicle plot shows the difference in number of bookings between each season and month and the percentage of children coming with each booking



Task abstraction:

- Analyze: Showing which season has the greatest number of bookings
- Search: lookup the percentage of children coming each month

Markers and Channels:

- Color: Represents the percentage of children in the specified month and season
- Size: Represents the number of customers that booked in the specified month and season

## Tree Map:

This tree map counts adults coming from each country and the percentage of those adults that canceled their booking



Task abstraction:

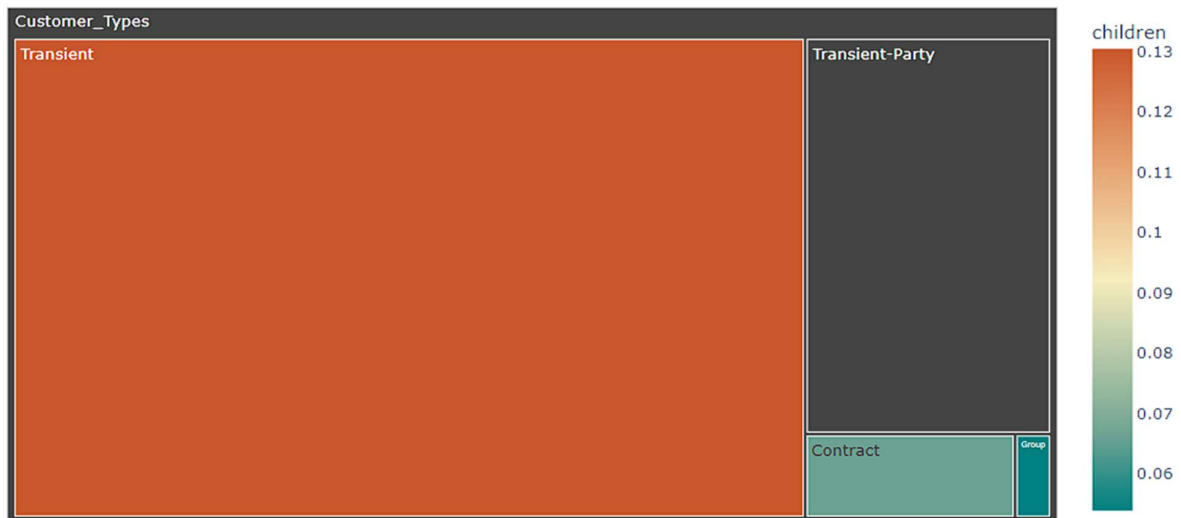
- Analyze: Showing which countries do customers come from most to adjust the marketing strategy for them
- Search: Looking up how many customers cancel their booking from each country

Markers and Channels:

- Color: Represents the percentage of cancellations
- Size: Represents the number of customers from the specified country

## Tree Map:

This tree map shows the number of customers in each customer type and the percentage of children each customer type brings



Task abstraction:

- Analyze: Showing which customer has the greatest number of bookings
- Search: lookup the percentage of children each customer type brings to adjust meal types accordingly

Marks and Channels:

- Color: Represents the percentage of children brought along
- Size: Represents the number of customers from the specified customer type

## **Tree Map:**

This tree map shows the number of customers that come from each company  
(Companies are represented by their IDs)

Companies



Task abstraction:

- Analyze: Showing which company has the greatest number of bookings
- Search: lookup if the customer came with a company or individually

Marks and Channels:

- Size: Represents the number of customers from each company

## Word cloud

Word clouds (also known as tag clouds or word art) are used to visualize and summarize all sorts of data, from voice of the customer feedback to academic papers, and everything in between.



We used word cloud to represent countries that travelers come from