# CAPSTONE PROJECT-1
EDA ON HOTEL BOOKING ANALYSIS

BY SAMBIT PANI

PROBLEM STATEMENT

❖ For this project we have provided with a dataset having booking information of more than 1 lakh resort and city hotel bookings. Dataset comprises of booking details, booked members, Days to stay, Best foods etc. As we know hotel booking affects a lot of factors making this a volatile business.

❖ Our goal here is to empower the management team with some insights from the provided data, which will be benefited for the company to build better business and generate more revenue.
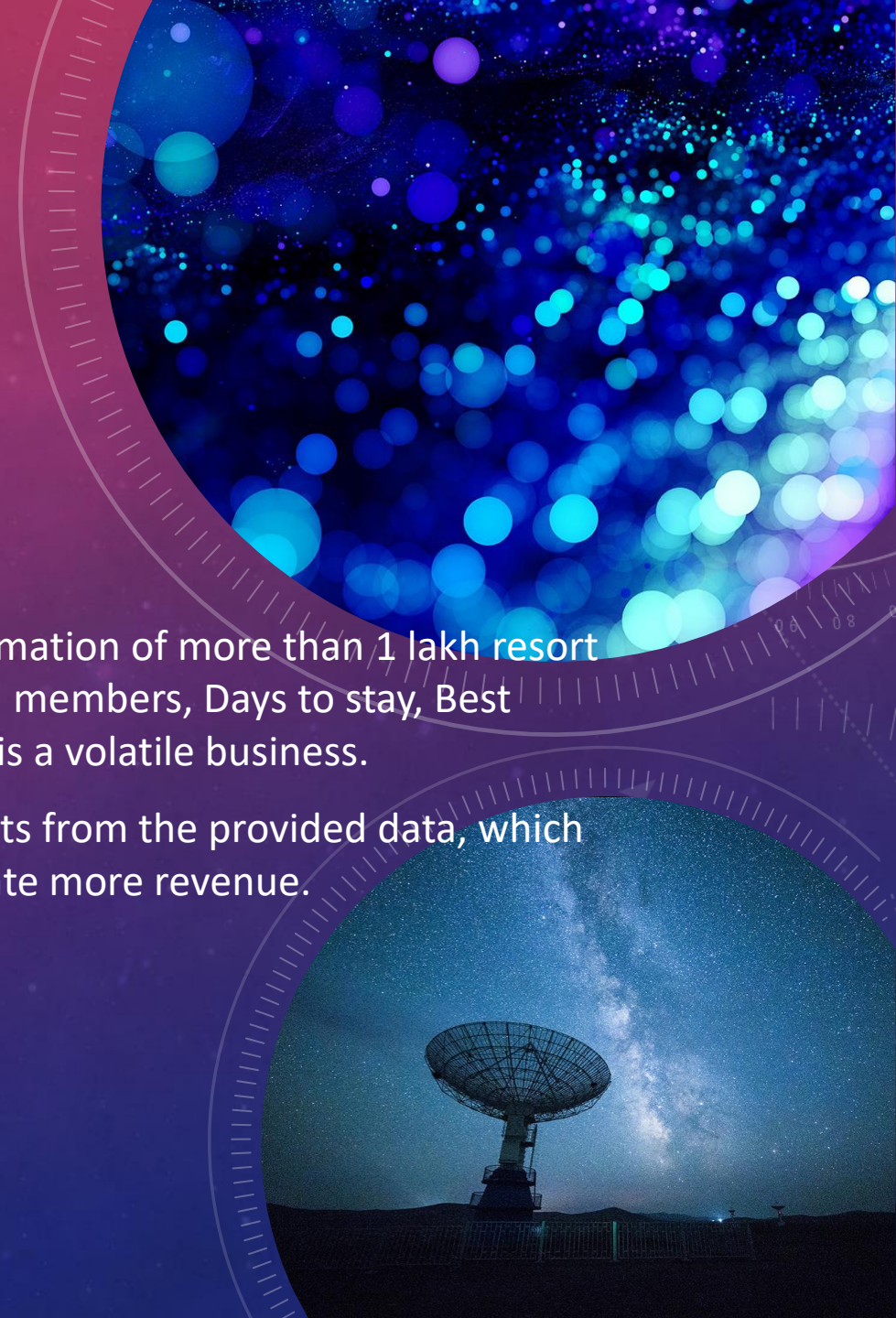
CONTENTS

Data Cleaning
- Data Exploration and understanding
- Data wrangling logics with missing value handling
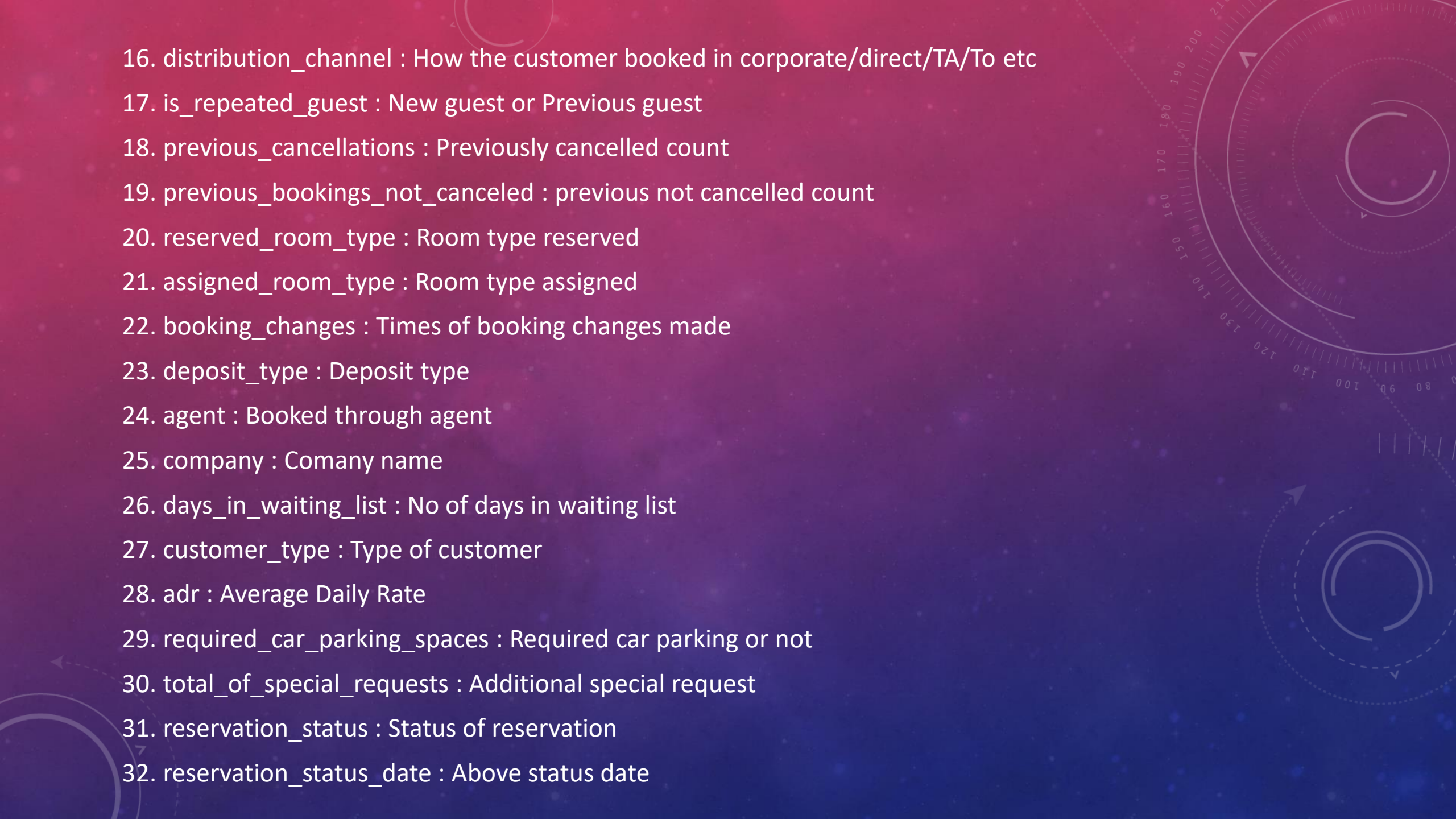
Data visualization
- Data visualization
- finding out insights

Conclusion
- How the insights will be helpful for stakeholders
- Conclusion

DATA UNDERSTANDING

- The provided data consists of 32 columns and 119390 rows.

- It is hotel booking data of whole year 2015 date and month wise.

- It has 4 float type, 16 integer type and 12 object type columns.

- Data columns are as follows:-

1. hotel : Type of hotel i.e. City or Resort
2. is_canceled : Booking cancelled or not
3. lead_time : Booking date and staying date duration
4. arrival_date_year : arrival year
5. arrival_date_month : arrival month
6. arrival_date_week_number : arrival week number
7. arrival_date_day_of_month : arrival day
8. stays_in_weekend_nights : Weekend nights stay
9. stays_in_week_nights : Week nights stay
10. adults : No of adult persons
11. children : No of children
12. babies : No of babies
13. meal : Kind of meal preferred
14. country : Guests are from which Country
15. market_segment : Customer type segment

16. distribution_channel : How the customer booked in corporate/direct/TA/To etc

17. is_repeated_guest : New guest or Previous guest

18. previous_cancellations : Previously cancelled count

19. previous_bookings_not_canceled : previous not cancelled count

20. reserved_room_type : Room type reserved

21. assigned_room_type : Room type assigned

22. booking_changes : Times of booking changes made

23. deposit_type : Deposit type

24. agent : Booked through agent

25. company : Comany name

26. days_in_waiting_list : No of days in waiting list

27. customer_type : Type of customer

28. adr : Average Daily Rate

29. required_car_parking_spaces : Required car parking or not

30. total_of_special_requests : Additional special request

31. reservation_status : Status of reservation

32. reservation_status_date : Above status date

DATA WRANGLING AND HANDLING NULL VALUES

❖ There are 31994 rows having duplicate values.

❖ They will be removed from data set.

```python
# Write your code to make your dataset analysis ready.
# 1. Removing Duplicates
df.drop_duplicates(inplace = True)
df
```

❖ Two new columns will be required for better insight.

❖ Two new columns are total days and total guests.

```python
# 3. Create two new columns for total person stayed and total days stay
df['total_days'] = df['stays_in_week_nights'] + df['stays_in_weekend_nights']
df['total_guests'] = df['babies'] + df['children'] + df['adults']
```

❖ Converting columns to proper datatypes.

❖ Float type rows to integer type.

❖ There are four columns having Null values

❖ We will replace them with zero

```python
# 2. Filling null values with 0
df.fillna(0,inplace=True)
df.isna().sum()
```

❖ If the total stay is zero then it not a valid row.

❖ So we remove those rows from data set.

```python
# 4. Check for valid stay
df.drop(df[df['total_days'] == 0].index,inplace=True)
df[df['total_days'] == 0]
```

❖ We will also change the reservation date to datetime.

❖ Now data is clean and ready to be explored.

```python
# 5. Changing the datatype to appropriate
df[['children', 'company', 'agent','total_guests']] = df[['children', 'company', 'agent','total_guests']].astype('int64')
from datetime import datetime
from datetime import date
df['reservation_status_date'] = df['reservation_status_date'].apply(lambda x : datetime.strptime(x,'%Y-%m-%d'))
```
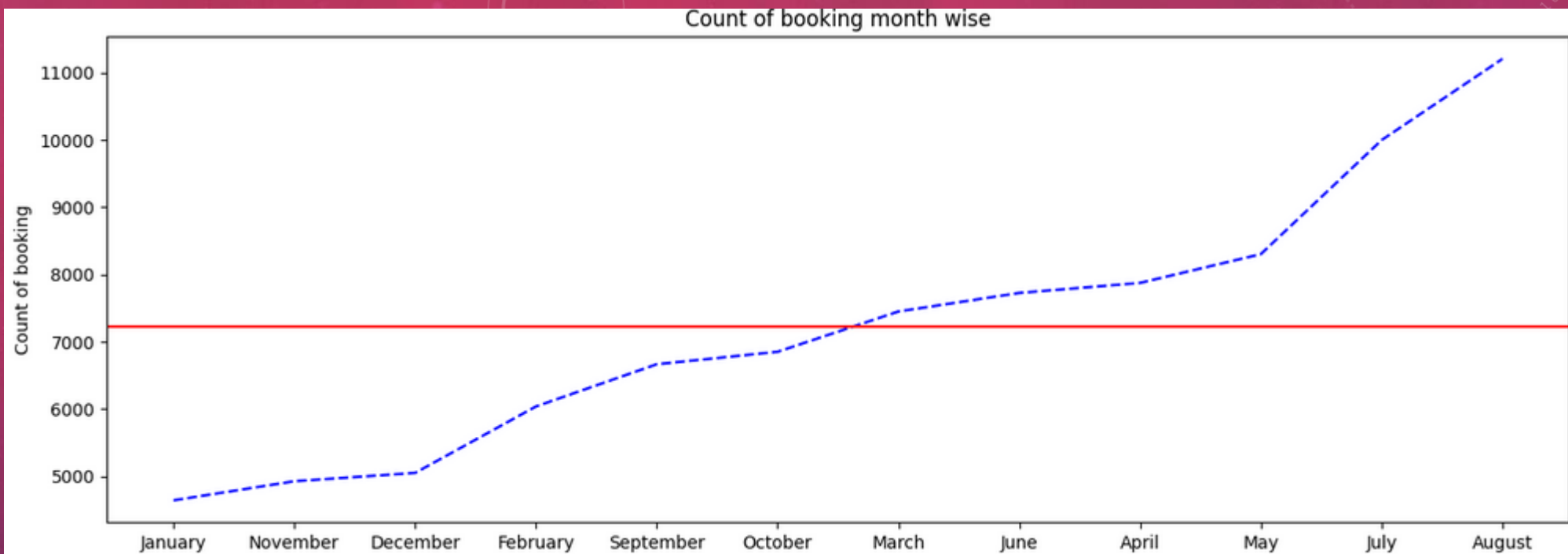
Count of booking month wise

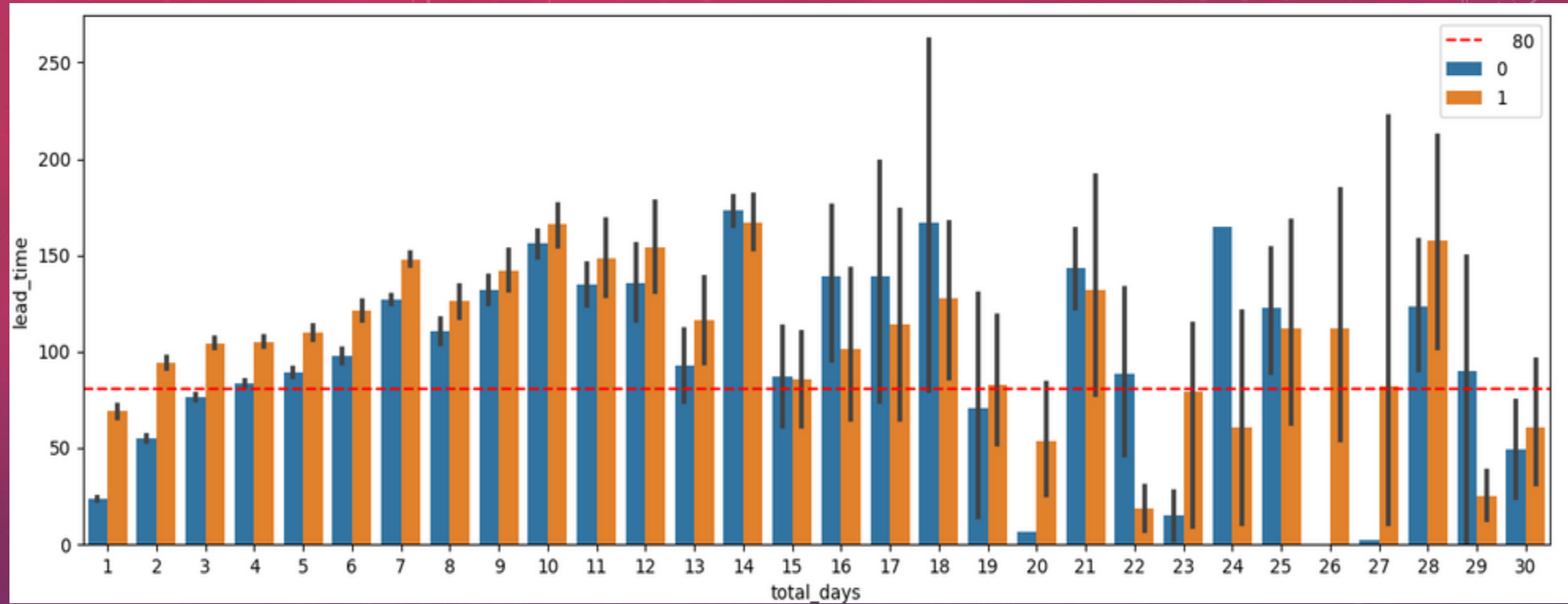## Conclusion:

❖ It's a graph of month wise booking counts.
❖ Here we can see the average line is 7229. We can find that more number of people come in months starting from march till august.
❖ So more staff is required for these months and less staff is required for other months. Accordingly team has to be managed so that proper work distribution will be managed.
❖ Also we can have some offers in the lower booking time frame to attract more people.
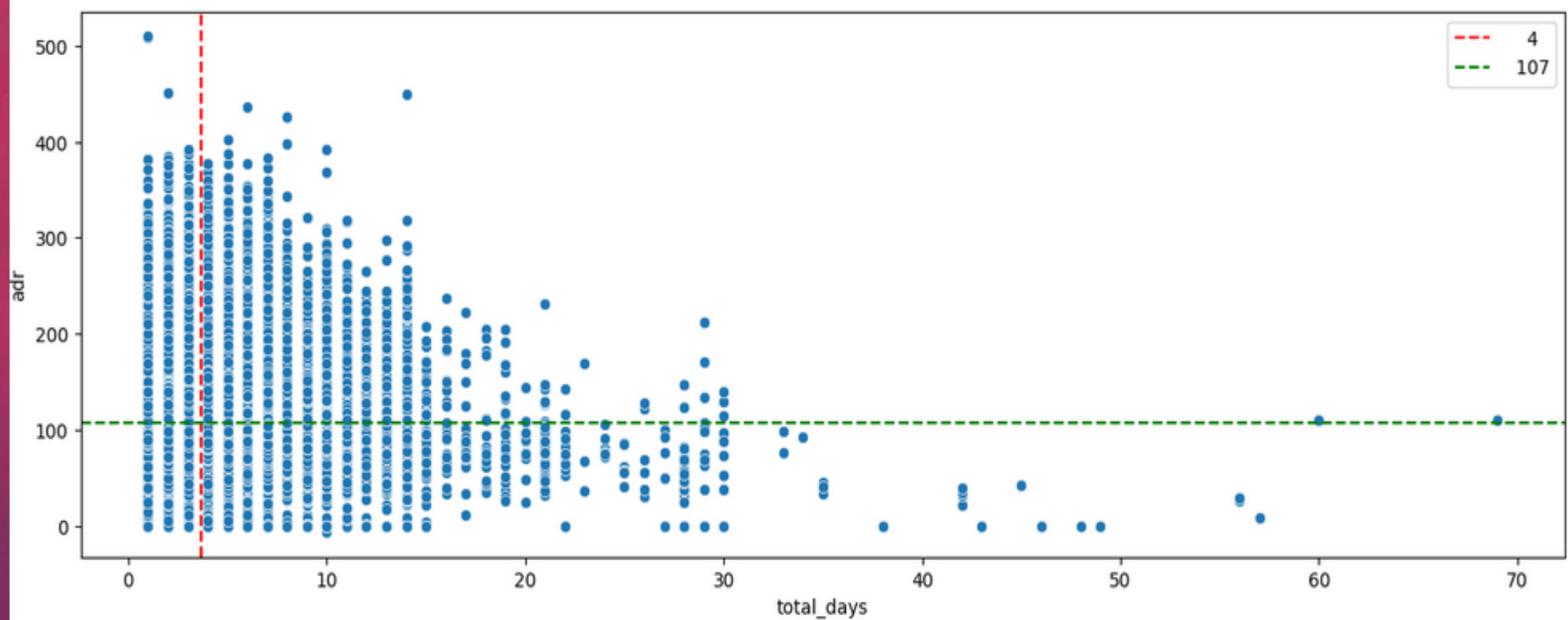
Conclusion:

❖ It's a graph of average adr to the total day stay.

❖ Here we can see the average adr is 107 and average days of stay is 4 days.

❖ Adr is less for more than two weeks of stay. But adr is quite same for 1st week and slightly less for 2nd week stay.

❖ Having offer to stay more is a better option for customers.

Conclusion:

❖ It's a graph of total days stay wrt lead time over cancelled or not.

❖ Here cancellation is more if the total days stay is two weeks and less there after and more till the 4th week stay.

❖ Lead time is lower in one or two days stay. So charge more on those days.

❖ It has been seen that more people are staying for 7,14,21 and 28 days. So its good to have offers in those days so that people stay till the end of that weekend.

❖ It has also been found that lead time has nothing to do with the cancellation.

Day wise adr vs total persons
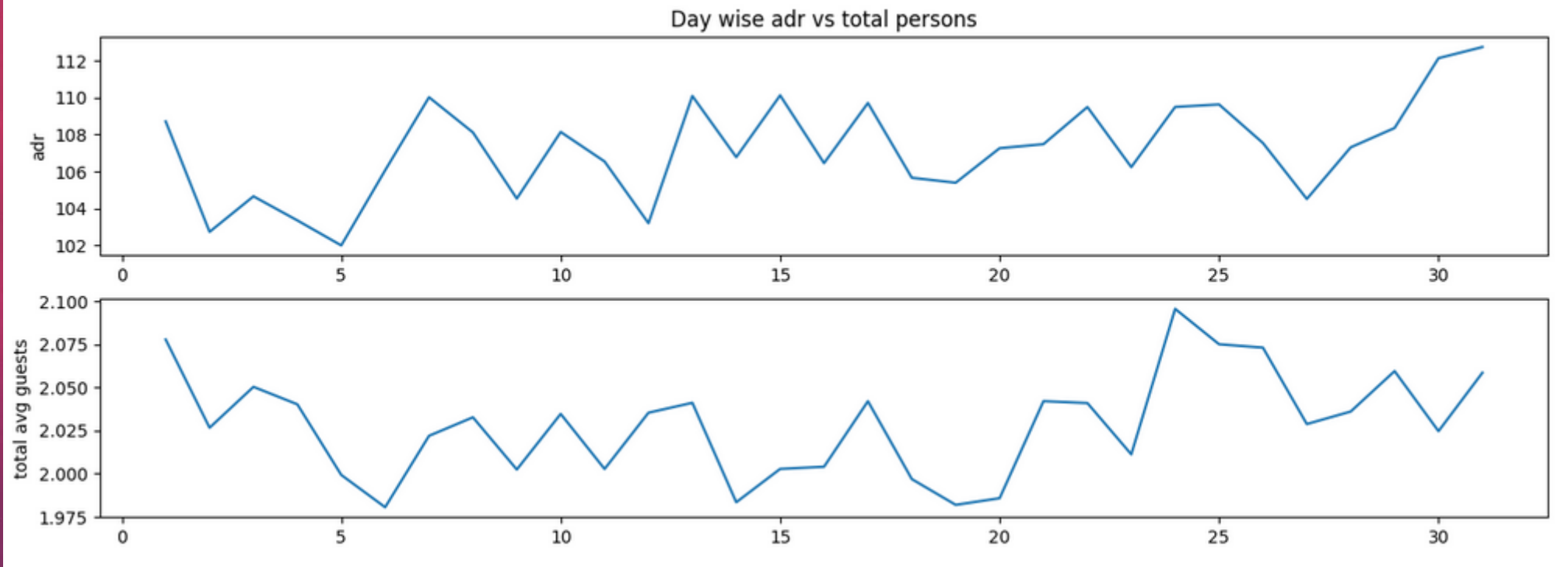
Conclusion:

❖ It is a graph showing day wise average adr and average total persons.
❖ More adr at the 30th and 31st days. Also adr is more at the weekend days.
❖ Averge person is more in 1st and last week of the month.
❖ So its better to charge more in 1st and last week of the month.
❖ Also adr should be kept high for non week end days.
❖ Average person is lowest in 6th day of the month.
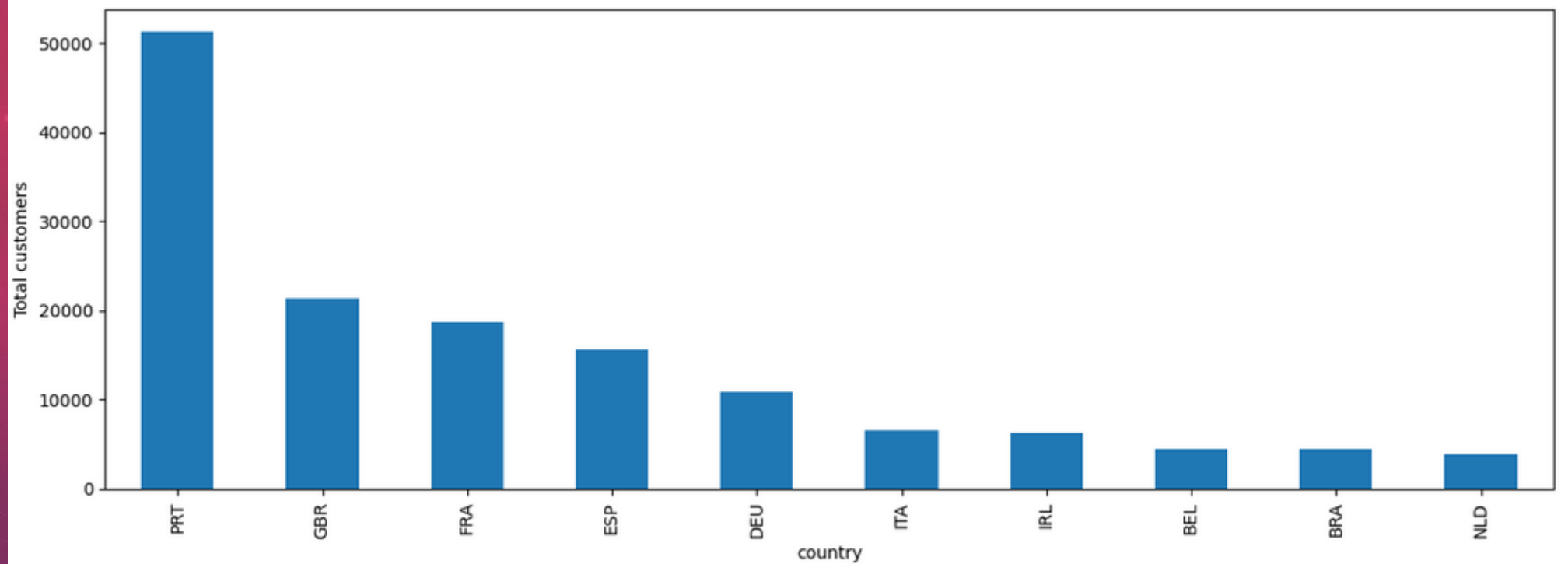
Conclusion:

- ❖ It's a graph of total customers to the country they are from.
- ❖ It's found that most people are from Portuguese.
- ❖ So the management team has to identify the reason behind their holiday and strengthen accordingly. So those people keep coming throughout the year.
- ❖ Also keep the customers more happy is the key principle in business.
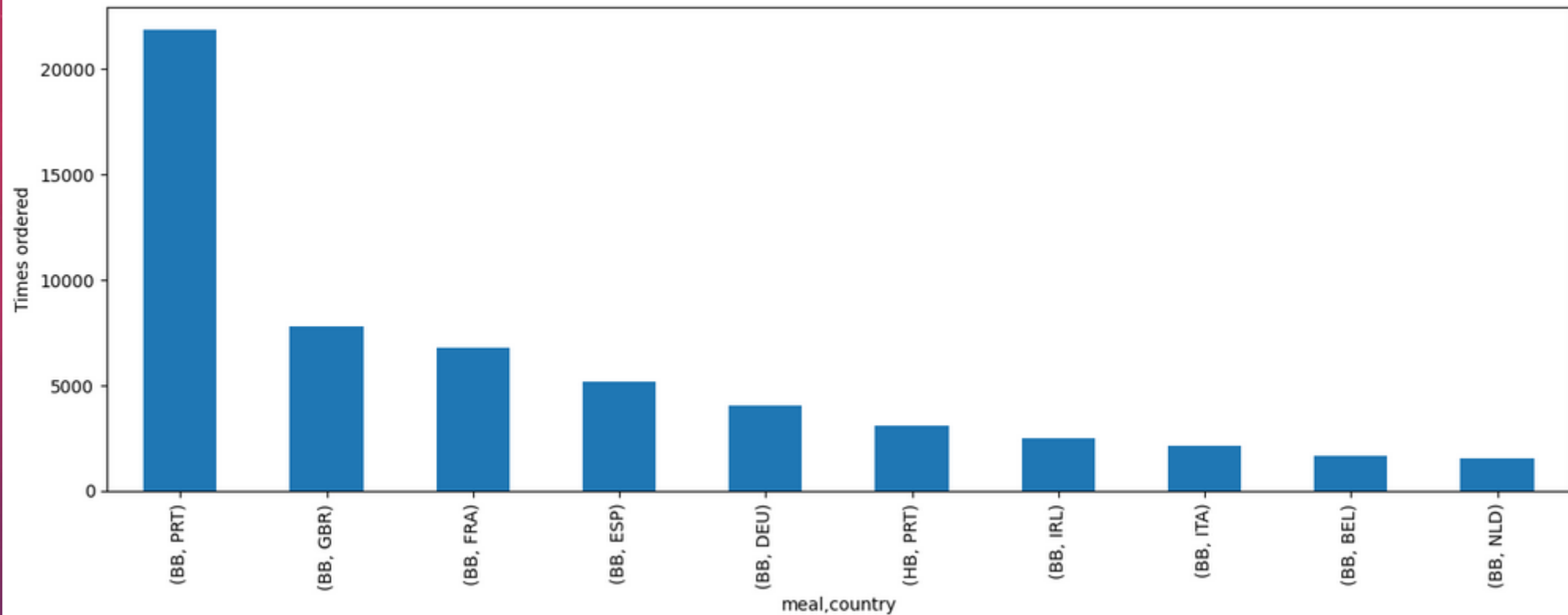
Conclusion:

❖ It's a graph of total meal ordered to the type and coutry it was preferred.
❖ Its found that mostly Breakfast is preferred by many people coming the hotel. And Its good to have those kind of foods only for them.
❖ The second most preferred meal type is HB i.e half board (breakfast and evening meal.
❖ Inventory can be supplied by those kind of meals only.
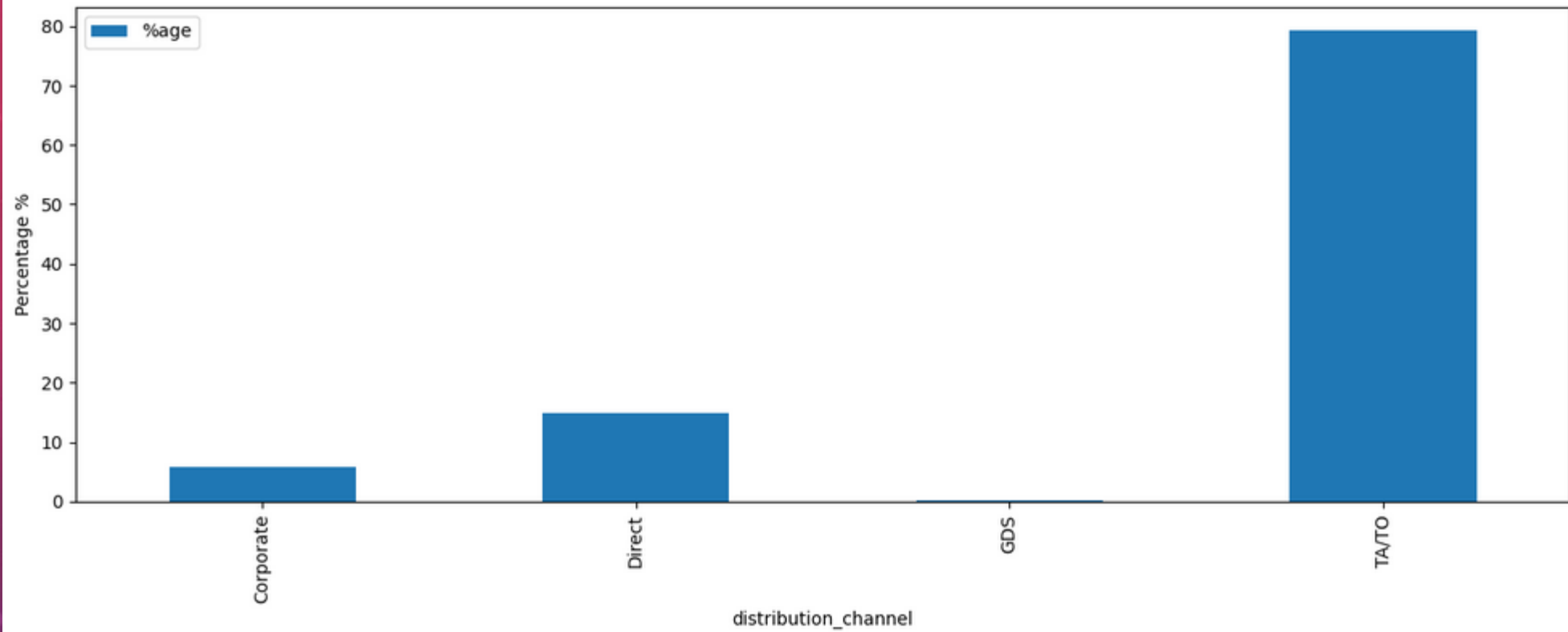
Conclusion:

- ❖ It's a graph showing the distribution channel to the %age of guests.
- ❖ It can be seen that most people booked by the travel agent/travel offices.
- ❖ More that 80% booked from a travel agent/office.
- ❖ So agent will be preferred by most people and direct and corporate is very less.
- ❖ Management team has to improve this kind of bookings. They can advertise more in corporate and in direct type of booking.
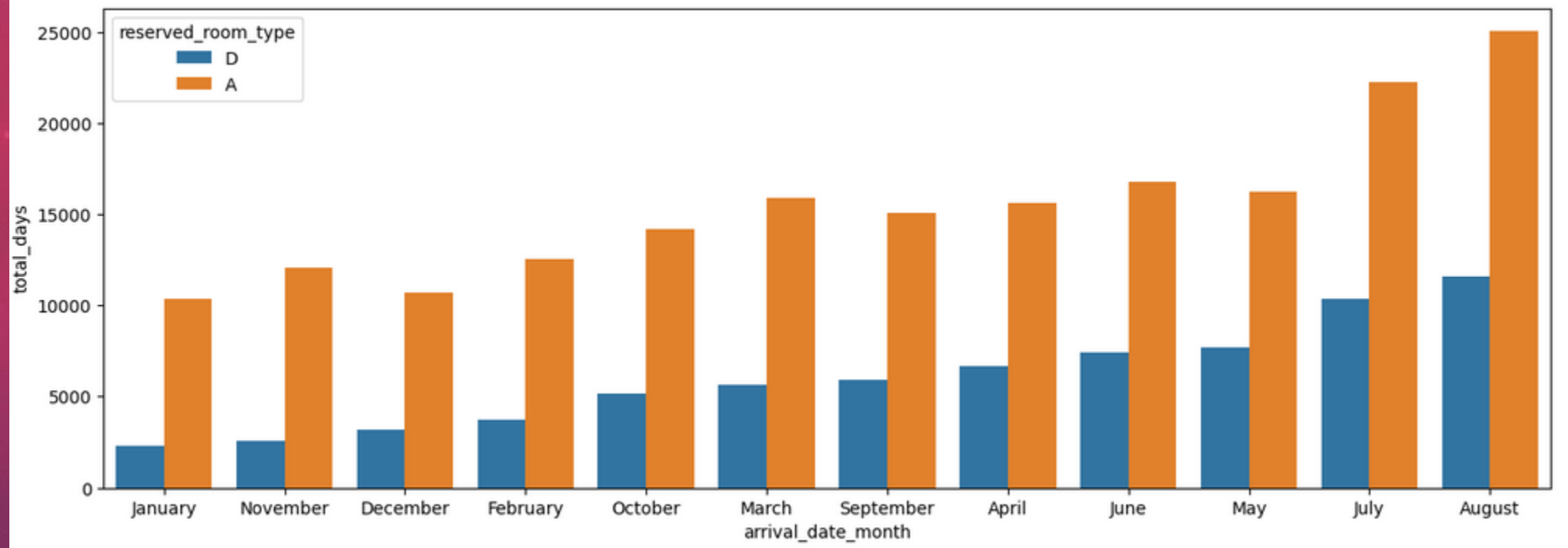
Conclusion:

- ❖ It's a graph showing the distribution channel to the average adr.
- ❖ SO it can be concluded that average adr is more in city hotels.
- ❖ So mostly city hotels are preferred by customers.
- ❖ Also the GDS has higher adr for city hotels as it is a global system.
- ❖ For TA/TO, which is the major booking group, the adr is highest due the more number of booking in city hotels than in resorts.
- ❖ Resort adr is lower due to less booking in resorts than city hotels.
- ❖ So resort booking needs to be improvised.
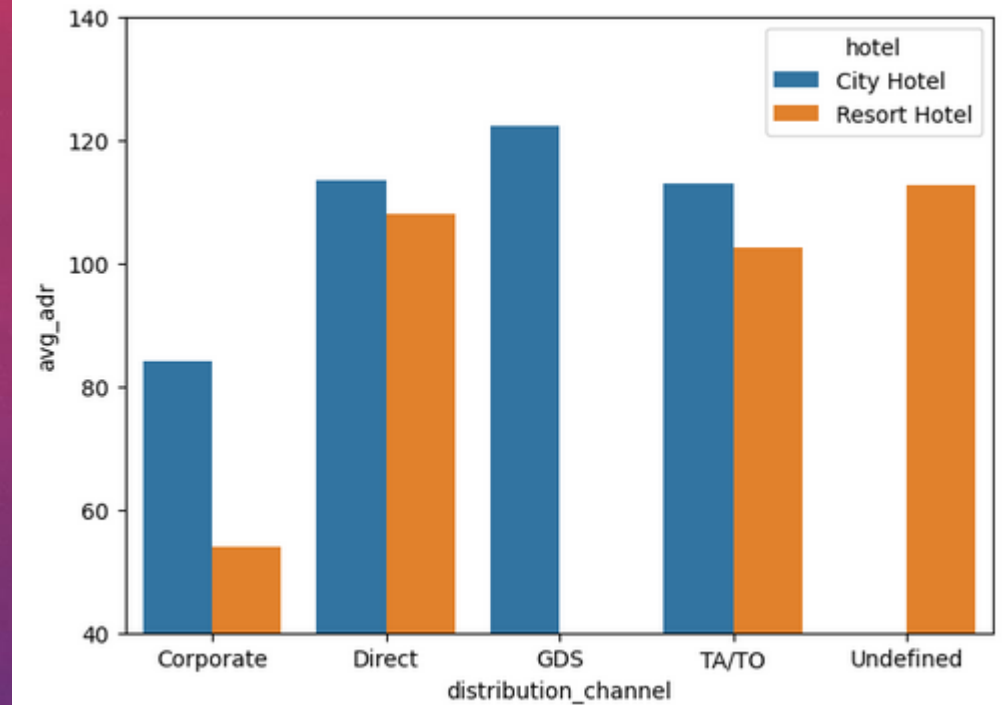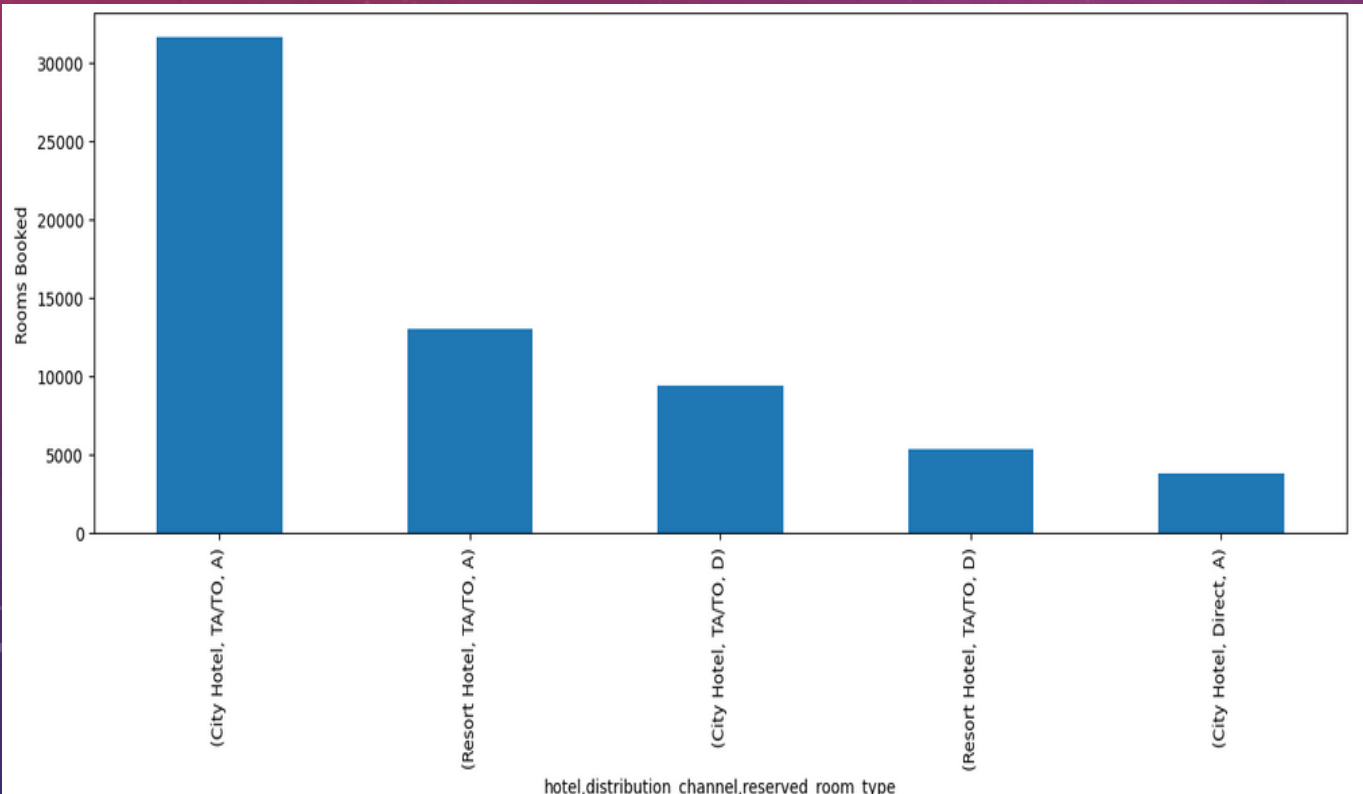
Conclusion:

- ❖ It's a graph showing the most preferred room type w.r.t. hotel type/distribution channel.
- ❖ It has been found that 'A' type room is booked most in both city and resorts.
- ❖ Also type 'D' is the 2nd most booked room type.
- ❖ The management team can reduce other type of rooms or they equalize the other room varieties to make all room types preferred by customers.
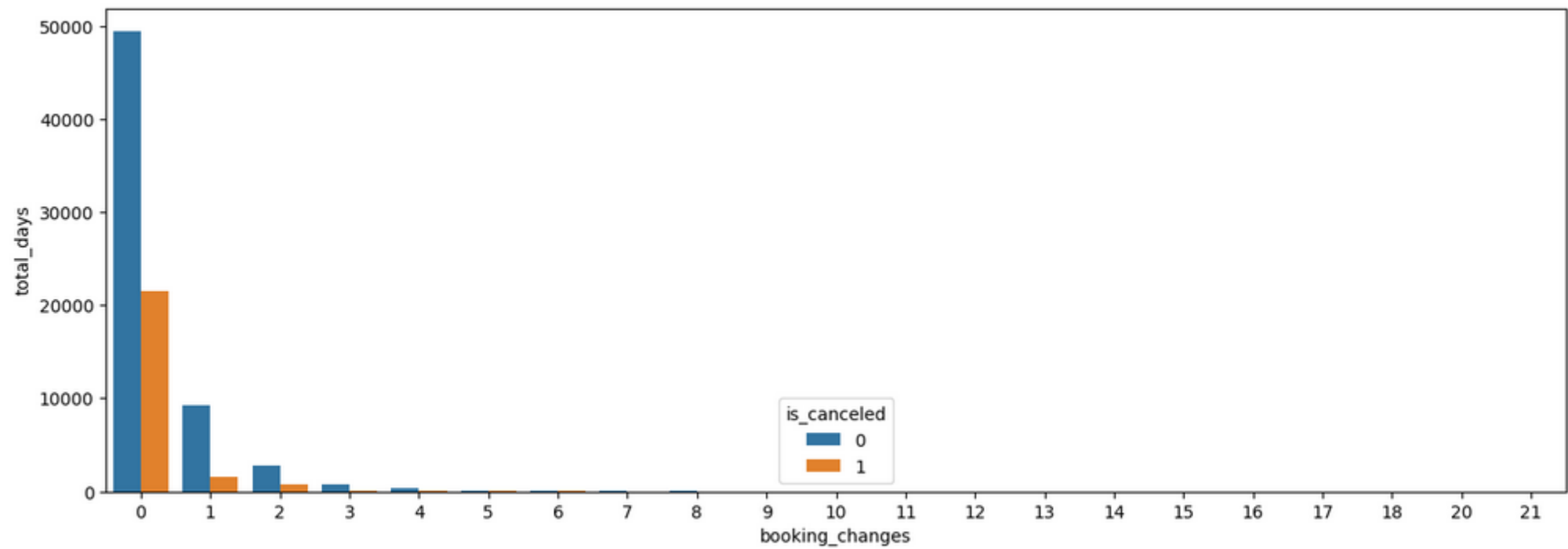
Conclusion:

- ❖ It's a graph showing the month wise booking of room types A and D.
- ❖ Its found out that A type room is preferred by the customer throughout the year and less in December and January and more in July and August.
- ❖ Also D type room is preferred less in the same ways.
- ❖ This is because of the booking counts.
- ❖ So more booking more will the counts of room types.
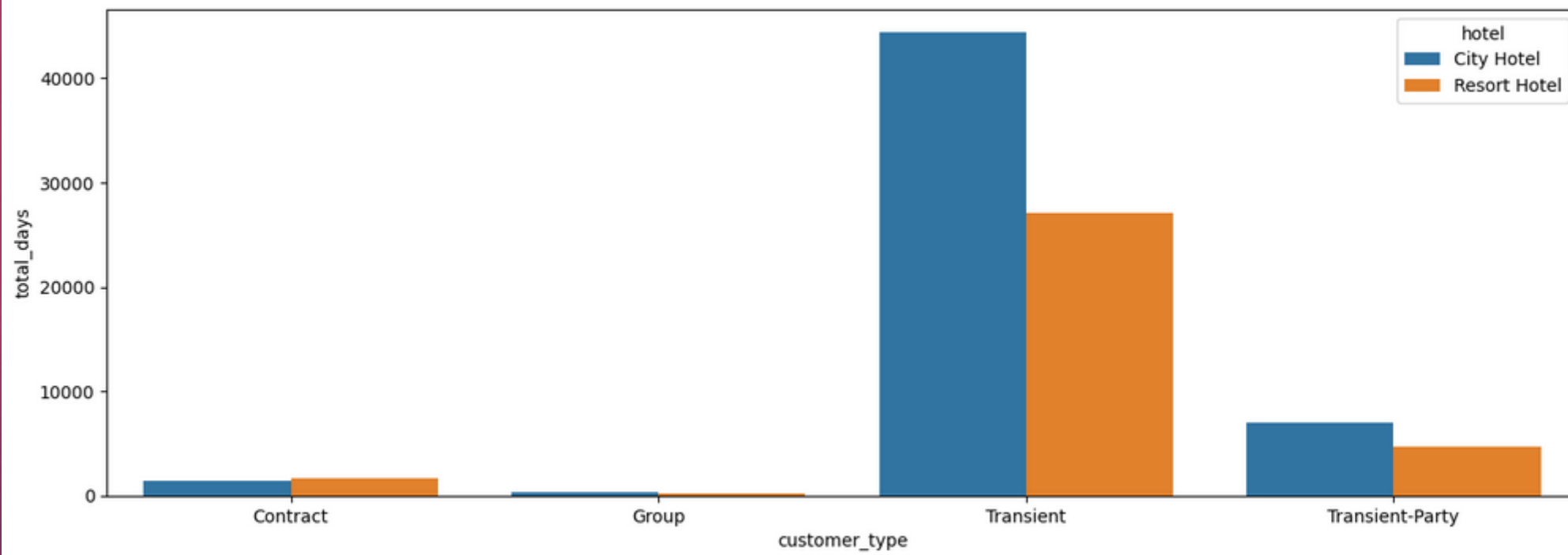
Conclusion:

- ❖ It's a graph showing the customer types over the stays in hotels.
- ❖ It has been found that more people falls under Transient type of customers.
- ❖ Also city hotels are preferred by them than resorts.
- ❖ It's a graph to know about the customer types.
- ❖ For transient customers charges can be increased than other type of customers.
- ❖ Also management team should increase the other type of customers by giving some type of offers to them.

Conclusion:

❖ It's a graph showing cancellation wrt the booking changes.
❖ So the booking changes is very less.
❖ So for betterment it should be capped at 2 times/3 times.
❖ Also booking changes will have very less effect on cancellation.
❖ SO its better to study which things been changes frequently and study the changing factors has a greater significance overall.
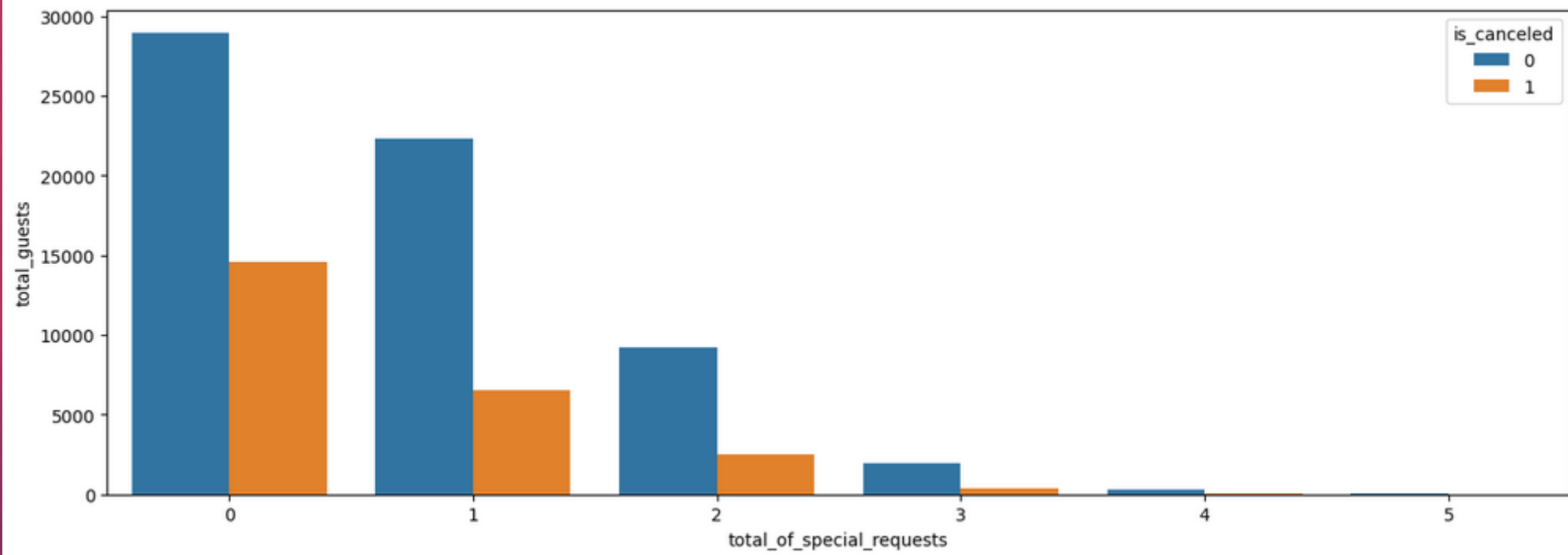
Conclusion:

❖ It's a graph showing cancellation wrt the number of special requests.
❖ So cancellation rate is reducing if there is a special request.
❖ It could be strategy to freely give any special request which will be preferred by most persons to every 100th customers.
❖ On top of it giving special request options to the customers will definitely reduce the cancellation rate.

Conclusion:

- ❖ It's a correlation heatmap of all required numerical columns.
- ❖ It is showing two significant correlations.
- ❖ Total guests and adr is 39% positively correlated and total days and lead time are 32% positively correlated.
- ❖ Nothing else has been found out from this graph as other columns are not correlated to each other.
- ❖ More adr for more number of guests.
- ❖ More number of days are likely to have more lead time.

Conclusion Of the EDA

- ✓ More hotel crew staff should be hired for temporary basis from march to august so as to manage the pool of customers.
- ✓ Reduce the adr to for stays more than 5-6 Days. Also It would be better to have offers for full week stay.
- ✓ Lead time is low for one or two day booking. So charge more for less days.
- ✓ Charge more on first and last week.
- ✓ The management team can plan and do the cultural festive of the countries to attract more people from the countries. Also they can fill up their inventories with stuff preferred by those people.
- ✓ They can set the default meal plan as BB. Also the material required or the inventory can be set to BB meal plan.
- ✓ Its good to have resorts than city hotels in terms of adr.
- ✓ More number of such rooms to be constructed for better customer attraction.
- ✓ Room type A should be constructed over D type. Also we can get the ration between these room types.
- ✓ Room not assigned as per requirement doesn't impact the cancellation.
- ✓ Advertisement should be done keeping the transient person into account so that more transient customers come for exciting offers. Also city hotels should be formed in terms of transient people comfort.
- ✓ Booking changes should be limited to two times only for betterment.
- ✓ It could be strategy to freely give any special request which will be preferred by most persons to every 100th customers.

THANK YOU !

SAMBIT PANI