SELECT

COUNT(id\_reservation)

FROM

fct\_bookings;

#70772

SELECT

COUNT(id\_reservation)

FROM

fct\_cancellations;

#36502

SELECT

\*

FROM

fct\_bookings

ORDER BY ds DESC;

SELECT

\*

FROM

fct\_bookings

ORDER BY ds ASC;

#2017-04-12 to 2018-11-30

SELECT

DATEDIFF(max(ds),min(ds)) AS 'daydiff'

FROM

fct\_bookings;

# 587 days in between in this table

SELECT

\*

FROM

fct\_cancellations

ORDER BY ds DESC;

SELECT

\*

FROM

fct\_cancellations

ORDER BY ds ASC;

#from 2017-01-01 to 2018-11-30

SELECT

DATEDIFF(max(ds),min(ds)) AS 'daydiff'

FROM

fct\_cancellations;

# 698 days in between in this table

SELECT

COUNT(b.id\_reservation)

FROM

fct\_bookings b

LEFT JOIN

fct\_cancellations c ON b.id\_reservation = c.id\_reservation;

#70772, which is equal to the rows of fct\_bookings

SELECT

COUNT(b.id\_reservation)

FROM

fct\_bookings b

LEFT JOIN

fct\_cancellations c ON b.id\_reservation = c.id\_reservation

where c.ds is not null;

#512 out of 70772 is cancelled

SELECT

\*

FROM

fct\_bookings b

INNER JOIN

dim\_entities e ON b.id\_entity = e.id\_entity

INNER JOIN

dim\_listings l ON b.id\_listing = l.id\_listing

LEFT JOIN

fct\_cancellations c ON b.id\_reservation = c.id\_reservation;

#There are a total of 65931 reservations.

#Here I got the 'joins' which the following project will base on.

SELECT

e.id\_entity,

COUNT(b.id\_reservation) AS 'number\_order\_entity',

ROUND(COUNT(b.id\_reservation) / (SELECT

COUNT(b.id\_reservation)

FROM

fct\_bookings b

INNER JOIN

dim\_entities e ON b.id\_entity = e.id\_entity

INNER JOIN

dim\_listings l ON b.id\_listing = l.id\_listing

LEFT JOIN

fct\_cancellations c ON b.id\_reservation = c.id\_reservation) \* 100,

2) AS 'Percentage(%)'

FROM

fct\_bookings b

INNER JOIN

dim\_entities e ON b.id\_entity = e.id\_entity

INNER JOIN

dim\_listings l ON b.id\_listing = l.id\_listing

LEFT JOIN

fct\_cancellations c ON b.id\_reservation = c.id\_reservation

GROUP BY e.id\_entity

ORDER BY number\_order\_entity DESC;

/\*There are a total of 141 entities and top five in number of bookings are 52, B98, 8BE, 267 and 53,

which count for 57% of bookings.\*/

SELECT

l.dim\_region,

l.dim\_country,

COUNT(l.dim\_country) AS 'number\_country\_listing',

ROUND(COUNT(l.dim\_country) / (SELECT

COUNT(l.id\_listing)

FROM

fct\_bookings b

INNER JOIN

dim\_entities e ON b.id\_entity = e.id\_entity

INNER JOIN

dim\_listings l ON b.id\_listing = l.id\_listing

LEFT JOIN

fct\_cancellations c ON b.id\_reservation = c.id\_reservation) \* 100,

1) AS 'Percentage(%)'

FROM

fct\_bookings b

INNER JOIN

dim\_entities e ON b.id\_entity = e.id\_entity

INNER JOIN

dim\_listings l ON b.id\_listing = l.id\_listing

LEFT JOIN

fct\_cancellations c ON b.id\_reservation = c.id\_reservation

GROUP BY l.dim\_region,l.dim\_country

ORDER BY count(l.dim\_region) DESC;

/\*There are a total of 147 counties where listings are located in.

Top five are US, GB, DE, FR and AU and they account for 70% of listings.\*/

#Scenario 1

SELECT

b.ds\_checkin AS date,

COUNT(b.id\_reservation) AS 'Total Bookings',

SUM(b.m\_booking\_value) AS 'Total Booking Value',

SUM(b.m\_booking\_value) / SUM(b.m\_nights\_booked) AS 'Price per Night',

SUM(b.m\_nights\_booked) / COUNT(b.id\_reservation) AS 'Nights per Booking',

SUM(b.m\_booking\_value) / SUM((b.m\_guests) \* (b.m\_nights\_booked)) AS 'Budget per Person per Night',

COUNT(c.id\_reservation) AS 'Total Cancellations',

ROUND(COUNT(c.id\_reservation) / COUNT(b.id\_reservation) \* 100,

1) AS 'Cancellation Rate(%)'

FROM

fct\_bookings b

INNER JOIN

dim\_entities e ON b.id\_entity = e.id\_entity

INNER JOIN

dim\_listings l ON b.id\_listing = l.id\_listing

LEFT JOIN

fct\_cancellations c ON b.id\_reservation = c.id\_reservation

WHERE

YEAR(b.ds\_checkin) IN (2017 , 2018)

AND l.dim\_country IN ('US')

GROUP BY b.ds\_checkin

ORDER BY SUM(b.m\_booking\_value) DESC;

/\*2017-11-06 623355

2018-09-25 559418

I exported the csv file 'Booking Values in Spikes and Normal Days' and see the spikes more clearly in Tableau\*/

#I researched these dates, but they are not national holiday or event days.So I guess the bump may be caused by specific group of clients

#First, I tried to find out what entities reserve rooms on these two days.

SELECT

e.id\_entity

FROM

dim\_entities e

WHERE

e.id\_entity IN (SELECT DISTINCT

b.id\_entity

FROM

fct\_bookings b

INNER JOIN

dim\_entities e ON b.id\_entity = e.id\_entity

INNER JOIN

dim\_listings l ON b.id\_listing = l.id\_listing

LEFT JOIN

fct\_cancellations c ON b.id\_reservation = c.id\_reservation

WHERE

DATE(b.ds\_checkin) IN ('2017-11-06' , '2018-09-25')

AND l.dim\_country IN ('US'));

#23 entities

SELECT

b.id\_entity,

SUM(b.m\_booking\_value) AS 'total booking value',

round(SUM(b.m\_booking\_value) \* 100/ (SELECT

SUM(b.m\_booking\_value)

FROM

fct\_bookings b

INNER JOIN

dim\_entities e ON b.id\_entity = e.id\_entity

INNER JOIN

dim\_listings l ON b.id\_listing = l.id\_listing

LEFT JOIN

fct\_cancellations c ON b.id\_reservation = c.id\_reservation

WHERE

DATE(b.ds\_checkin) IN ('2017-11-06' , '2018-09-25')

AND l.dim\_country IN ('US')),1) AS 'percentage(%)'

FROM

fct\_bookings b

INNER JOIN

dim\_entities e ON b.id\_entity = e.id\_entity

INNER JOIN

dim\_listings l ON b.id\_listing = l.id\_listing

LEFT JOIN

fct\_cancellations c ON b.id\_reservation = c.id\_reservation

WHERE

DATE(b.ds\_checkin) IN ('2017-11-06' , '2018-09-25')

AND l.dim\_country IN ('US')

GROUP BY b.id\_entity

ORDER BY SUM(b.m\_booking\_value) DESC;

/\*Entity 'F6' reserve 77.2% out of all 23 entities

no 1: 77.2% F6

no 2: 7.4% 52

no 3: 3.1% 53

no 4: 2.7% B98

export 'Entities' total booking value in spike days \*/

/\* try to get to know where the listing is located\*/

SELECT

l.id\_listing, l.dim\_region, l.dim\_state, l.dim\_market

FROM

fct\_bookings b

INNER JOIN

dim\_entities e ON b.id\_entity = e.id\_entity

INNER JOIN

dim\_listings l ON b.id\_listing = l.id\_listing

WHERE

e.id\_entity IN ('F6' , '52', '53', 'B98')

AND DATE(b.ds\_checkin) IN ('2017-11-06' , '2018-09-25')

AND l.dim\_country IN ('US')

GROUP BY l.id\_listing;

/\*found: most are in SF, CA, North America\*/

SELECT

l.dim\_state,

COUNT(l.dim\_state),

COUNT(l.dim\_state) \* 100 / (SELECT

COUNT(l.dim\_state)

FROM

fct\_bookings b

INNER JOIN

dim\_entities e ON b.id\_entity = e.id\_entity

INNER JOIN

dim\_listings l ON b.id\_listing = l.id\_listing

WHERE

e.id\_entity IN ('F6' , '52', '53', 'B98')

AND DATE(b.ds\_checkin) IN ('2017-11-06' , '2018-09-25')

AND l.dim\_country IN ('US')) AS Percentage

FROM

fct\_bookings b

INNER JOIN

dim\_entities e ON b.id\_entity = e.id\_entity

INNER JOIN

dim\_listings l ON b.id\_listing = l.id\_listing

WHERE

e.id\_entity IN ('F6' , '52', '53', 'B98')

AND DATE(b.ds\_checkin) IN ('2017-11-06' , '2018-09-25')

AND l.dim\_country IN ('US')

GROUP BY l.dim\_state

ORDER BY COUNT(l.dim\_state) DESC;

#export Listings by state.csv

SELECT

l.dim\_market,

COUNT(l.dim\_market),

COUNT(l.dim\_market) \* 100 / (SELECT

COUNT(l.dim\_market)

FROM

fct\_bookings b

INNER JOIN

dim\_entities e ON b.id\_entity = e.id\_entity

INNER JOIN

dim\_listings l ON b.id\_listing = l.id\_listing

WHERE

e.id\_entity IN ('F6' , '52', '53', 'B98')

AND DATE(b.ds\_checkin) IN ('2017-11-06' , '2018-09-25')

AND l.dim\_country IN ('US')) AS Percentage

FROM

fct\_bookings b

INNER JOIN

dim\_entities e ON b.id\_entity = e.id\_entity

INNER JOIN

dim\_listings l ON b.id\_listing = l.id\_listing

WHERE

e.id\_entity IN ('F6' , '52', '53', 'B98')

AND DATE(b.ds\_checkin) IN ('2017-11-06' , '2018-09-25')

AND l.dim\_country IN ('US')

GROUP BY l.dim\_market

ORDER BY COUNT(l.dim\_market) DESC;

#export Listings by market.csv

#The root causes to two main spikes are an annual event in San Francisco, CA and reservation

#Then I will do customer behavior study.

SELECT

l.id\_listing,

l.dim\_room\_type,

l.dim\_listing\_tier,

l.dim\_cancellation\_policy,

l.dim\_person\_capacity,

l.dim\_is\_active,

l.dim\_bedrooms

FROM

fct\_bookings b

INNER JOIN

dim\_entities e ON b.id\_entity = e.id\_entity

INNER JOIN

dim\_listings l ON b.id\_listing = l.id\_listing

WHERE

e.id\_entity IN (SELECT

e.id\_entity

FROM

fct\_bookings b

INNER JOIN

dim\_entities e ON b.id\_entity = e.id\_entity

INNER JOIN

dim\_listings l ON b.id\_listing = l.id\_listing

WHERE

DATE(b.ds\_checkin) IN ('2017-11-06' , '2018-09-25')

AND l.dim\_country IN ('US'))

AND DATE(b.ds\_checkin) IN ('2017-11-06' , '2018-09-25')

AND l.dim\_country IN ('US')

GROUP BY l.id\_listing;

#export Listing preference.csv

SELECT

b.ds\_checkin AS date,

COUNT(b.id\_reservation) AS 'Total Bookings',

SUM(b.m\_nights\_booked) / COUNT(b.id\_reservation) AS 'Nights per Booking',

SUM(b.m\_booking\_value) AS 'Total Booking Value',

SUM(b.m\_booking\_value) / SUM(b.m\_nights\_booked) AS 'Price per Night',

SUM(b.m\_nights\_booked \* l.dim\_bedrooms) AS 'demand-room',

SUM(b.m\_booking\_value) / SUM(b.m\_nights\_booked \* l.dim\_bedrooms) AS 'Price per room per night',

SUM((b.m\_guests) \* (b.m\_nights\_booked)) as 'demand-guest',

SUM(b.m\_booking\_value) / SUM((b.m\_guests) \* (b.m\_nights\_booked)) AS 'Price per Person per Night',

AVG(DATEDIFF(b.ds\_checkin, b.ds)) AS 'Reservation beforehand',

COUNT(c.id\_reservation) AS 'Total Cancellations',

AVG(DATEDIFF(b.ds\_checkin, c.ds)) AS 'Cancellation beforehand',

ROUND(COUNT(c.id\_reservation) / COUNT(b.id\_reservation) \* 100,

1) AS 'Cancellation Rate(%)'

FROM

fct\_bookings b

INNER JOIN

dim\_entities e ON b.id\_entity = e.id\_entity

INNER JOIN

dim\_listings l ON b.id\_listing = l.id\_listing

LEFT JOIN

fct\_cancellations c ON b.id\_reservation = c.id\_reservation

WHERE

b.ds\_checkin IN ('2017-11-06' , '2018-09-25')

AND l.dim\_country IN ('US')

AND l.dim\_bedrooms IS NOT NULL

GROUP BY b.ds\_checkin

ORDER BY SUM(b.m\_booking\_value) DESC;

#export Reservation behavior for spikes.csv

SELECT

b.ds\_checkin AS date,

COUNT(b.id\_reservation) AS 'Total Bookings',

SUM(b.m\_nights\_booked) / COUNT(b.id\_reservation) AS 'Nights per Booking',

SUM(b.m\_booking\_value) AS 'Total Booking Value',

SUM(b.m\_booking\_value) / SUM(b.m\_nights\_booked) AS 'Price per Night',

SUM(b.m\_nights\_booked \* l.dim\_bedrooms) AS 'demand-room',

SUM(b.m\_booking\_value) / SUM(b.m\_nights\_booked \* l.dim\_bedrooms) AS 'Price per room per night',

SUM((b.m\_guests) \* (b.m\_nights\_booked)) as 'demand-guest',

SUM(b.m\_booking\_value) / SUM((b.m\_guests) \* (b.m\_nights\_booked)) AS 'Price per Person per Night',

AVG(DATEDIFF(b.ds\_checkin, b.ds)) AS 'Reservation beforehand',

COUNT(c.id\_reservation) AS 'Total Cancellations',

AVG(DATEDIFF(b.ds\_checkin, c.ds)) AS 'Cancellation beforehand',

ROUND(COUNT(c.id\_reservation) / COUNT(b.id\_reservation) \* 100,

1) AS 'Cancellation Rate(%)'

FROM

fct\_bookings b

INNER JOIN

dim\_entities e ON b.id\_entity = e.id\_entity

INNER JOIN

dim\_listings l ON b.id\_listing = l.id\_listing

LEFT JOIN

fct\_cancellations c ON b.id\_reservation = c.id\_reservation

WHERE

YEAR(b.ds\_checkin) IN ('2017' , '2018')

AND l.dim\_country IN ('US')

AND l.dim\_bedrooms IS NOT NULL

GROUP BY b.ds\_checkin

ORDER BY SUM(b.m\_booking\_value) DESC;

/\*Reservation behavior for all.csv\*/

/\*As shown below, I tried to figure out the financial impact related to spike days\*/

select SUM(b.m\_booking\_value) AS 'Total Booking Value'

FROM

fct\_bookings b

INNER JOIN

dim\_entities e ON b.id\_entity = e.id\_entity

INNER JOIN

dim\_listings l ON b.id\_listing = l.id\_listing

LEFT JOIN

fct\_cancellations c ON b.id\_reservation = c.id\_reservation

WHERE

YEAR(b.ds\_checkin) IN ('2017' , '2018')

AND l.dim\_country IN ('US');

select SUM(b.m\_booking\_value) AS 'Total Booking Value'

FROM

fct\_bookings b

INNER JOIN

dim\_entities e ON b.id\_entity = e.id\_entity

INNER JOIN

dim\_listings l ON b.id\_listing = l.id\_listing

LEFT JOIN

fct\_cancellations c ON b.id\_reservation = c.id\_reservation

WHERE

b.ds\_checkin IN ('2017-11-06' , '2018-09-25')

AND l.dim\_country IN ('US');

select 1182773/29844148 \* 100 as 'Revenue Percentage';

#3.96% of all US market

select SUM(b.m\_booking\_value) AS 'Total Booking Value'

FROM

fct\_bookings b

INNER JOIN

dim\_entities e ON b.id\_entity = e.id\_entity

INNER JOIN

dim\_listings l ON b.id\_listing = l.id\_listing

LEFT JOIN

fct\_cancellations c ON b.id\_reservation = c.id\_reservation

WHERE

YEAR(b.ds\_checkin) IN ('2017' , '2018');

select 1182773/50590429 \* 100 as 'Revenue Percentage';

#2.34% of global market

SELECT

e.\*

FROM

fct\_bookings b

INNER JOIN

dim\_entities e ON b.id\_entity = e.id\_entity

INNER JOIN

dim\_listings l ON b.id\_listing = l.id\_listing

WHERE

e.id\_entity IN (SELECT

e.id\_entity

FROM

fct\_bookings b

INNER JOIN

dim\_entities e ON b.id\_entity = e.id\_entity

INNER JOIN

dim\_listings l ON b.id\_listing = l.id\_listing

WHERE

DATE(b.ds\_checkin) IN ('2017-11-06' , '2018-09-25')

AND l.dim\_country IN ('US'))

AND DATE(b.ds\_checkin) IN ('2017-11-06' , '2018-09-25')

AND l.dim\_country IN ('US')

group by e.id\_entity;

#export Entity info.csv

#scenario 2

SELECT

b.id\_entity

FROM

fct\_bookings b

INNER JOIN

dim\_entities e ON b.id\_entity = e.id\_entity

INNER JOIN

dim\_listings l ON b.id\_listing = l.id\_listing

LEFT JOIN

fct\_cancellations c ON b.id\_reservation = c.id\_reservation

WHERE

YEAR(b.ds) IN ('2018')

GROUP BY b.id\_entity

ORDER BY SUM(b.m\_booking\_value) DESC

LIMIT 10;

SELECT

b.ds\_checkin AS date,

COUNT(b.id\_reservation) AS 'Total Bookings',

SUM(b.m\_nights\_booked) / COUNT(b.id\_reservation) AS 'Nights per Booking',

SUM(b.m\_booking\_value) AS 'Total Booking Value',

SUM(b.m\_booking\_value) / SUM(b.m\_nights\_booked) AS 'Price per Night',

SUM(b.m\_nights\_booked \* l.dim\_bedrooms) AS 'demand-room',

SUM(b.m\_booking\_value) / SUM(b.m\_nights\_booked \* l.dim\_bedrooms) AS 'Price per room per night',

AVG(DATEDIFF(b.ds\_checkin, b.ds)) AS 'Reservation beforehand',

COUNT(c.id\_reservation) AS 'Total Cancellations',

AVG(DATEDIFF(b.ds\_checkin, c.ds)) AS 'Cancellation beforehand',

ROUND(COUNT(c.id\_reservation) / COUNT(b.id\_reservation) \* 100,

1) AS 'Cancellation Rate(%)'

FROM

fct\_bookings b

INNER JOIN

dim\_entities e ON b.id\_entity = e.id\_entity

INNER JOIN

dim\_listings l ON b.id\_listing = l.id\_listing

LEFT JOIN

fct\_cancellations c ON b.id\_reservation = c.id\_reservation

WHERE

b.id\_entity IN ('52' , 'B98',

'53',

'267',

'5693C',

'8BE',

'42E88',

'F6',

'183F',

'3312')

AND YEAR(b.ds\_checkin) IN (2017 , 2018)

GROUP BY b.ds\_checkin

order by b.ds\_checkin;

#export Top 10 entities