
Peek: Take Home Assignment Analysis of Covid-19 on Business

Presented by Jeremy Caothien

Questions Today

→ Question 1

How did the start of the pandemic impact our partners' business?

→ Question 2

What happened to customer booking and cancellation behavior after the start of the pandemic?

→ Question 3

Create a proposal that explains if there were any other major pandemic events contributing to a change in bookings or cancellation rates.

→ Question 4

Look at the SQL Query and describe what the `transaction_sequence` and `context_sequence` fields are.

Tools I used:

- **Google BigQuery**
Free trial, uploaded take-home CSV and got to querying away.
- **Google Data Studio (Visualization)**
Integration with BigQuery makes it easy to visualize the query to spot trends.
- **Google Search**
Great to do research and make sure I know what I'm talking about! Helpful for Question 3.

But First, Assumptions

- **total_cancellations:**
Because cancellation rate is defined as number of cancellations/bookings, multiplying $\text{cancellation_rate} \times \text{total_bookings} = \text{total_cancellations}$
- **avg_canceled_ahead_days & total_booked_ahead_days:**
Respective of the transaction type, $\frac{\text{total_canceled_ahead_days}}{\text{total_cancellations}}$ $\frac{\text{total_booked_ahead_days}}{\text{total_bookings}}$
- **applying absolute value for negative booked_ahead & canceled_ahead values:**
These are 95 records with negative values, and I couldn't see any world where negative values made sense for these contexts.

```

##base query
select
transaction_date,
--cancellations
round(total_bookings*cancellation_rate,0) as total_cancellations,
abs(total_canceled_ahead_days) as total_canceled_ahead_days,
round(abs(total_canceled_ahead_days)/(total_bookings*cancellation_rate),1) as
avg_canceled_ahead_days,
--bookings
total_bookings,
abs(total_booked_ahead_days) as total_booked_ahead_days,
round(abs(total_booked_ahead_days)/total_bookings,1) as avg_booked_ahead_days,
--
cancellation_rate,
seven_day_rolling_avg

from peektakehome.takehome

order by 1

```

```

##by year-month query
select
format_date("%Y-%m %B",transaction_date) as year_month,
--cancellations
sum(round(total_bookings*cancellation_rate,0)) as monthly_total_cancellations,
sum(abs(total_canceled_ahead_days)) as monthly_total_canceled_ahead_days,
round(sum(abs(total_canceled_ahead_days))/sum((total_bookings*cancellation_rate)),1) as
monthly_avg_canceled_ahead_days,
--bookings
sum(total_bookings) as monthly_total_bookings,
sum(abs(total_booked_ahead_days)) as monthly_total_booked_ahead_days,
round(sum(abs(total_booked_ahead_days))/sum(total_bookings),1) as monthly_avg_booked_ahead_days,
--
avg(cancellation_rate) as monthly_avg_cancellation_rate,
avg(seven_day_rolling_avg) as monthly_avg_seven_day_rolling_avg

from peektakehome.takehome
group by 1
order by 1

```

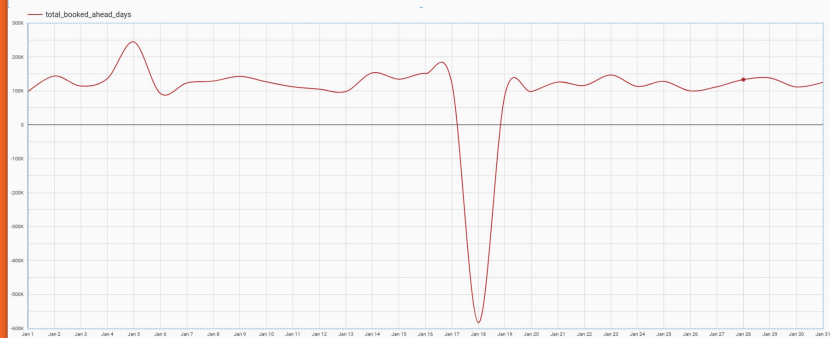
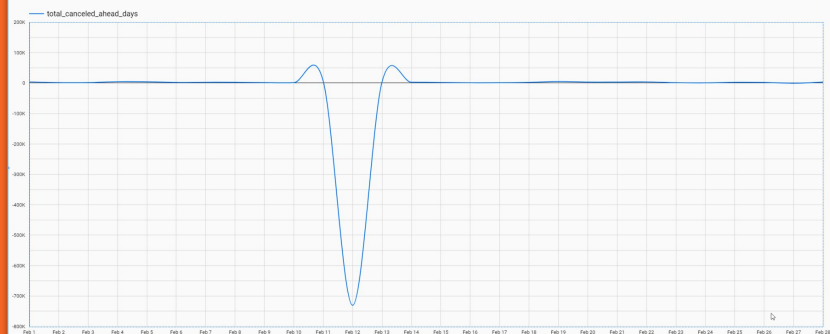
Query

Overall, I didn't think there was a need to do anything fancy (CTEs, subqueries, self-joins).

I was interested in looking at both daily trends and monthly trends.

I included an absolute value function to turn negative values into positive, more on the reasoning on the next slide.

transaction_date	total_cancellations	total_canceled_ahead_days	avg_canceled_ahead_days	total_bookings	total_booked_ahead_days
2019-01-18	283.0	2583	9.1	6957	-583153
2019-01-30	191.0	-643	-3.4	5226	111566
2019-02-12	307.0	-730461	-2379.4	6077	154477



The Case for Absolute Value for Negatives

During my analysis, I noticed there were 95 negative values for `total_canceled_ahead_days` and `total_booked_ahead_days`, sometimes egregiously negative.

I'm interpreting a negative value to mean that the booking took place in the past, so how is it possible for someone to retroactively cancel or book an activity?

Suppose this was true, It that means for February 12, 2019 on average, cancellations are happening for events that are 6.5 years in the past.

These may be outliers that I can't explain with current context, but I know negative values are impossible here. Therefore, I will use `abs(total_canceled_ahead_days)` to turn these negative into positives. These dips are surrounded by two peaks, which suggest to me that is a *possibility* that it spiked.

Question 1

How did the start of the pandemic impact our partners' business?

The WHO declared Covid-19 as a pandemic on [March 11, 2020](#)

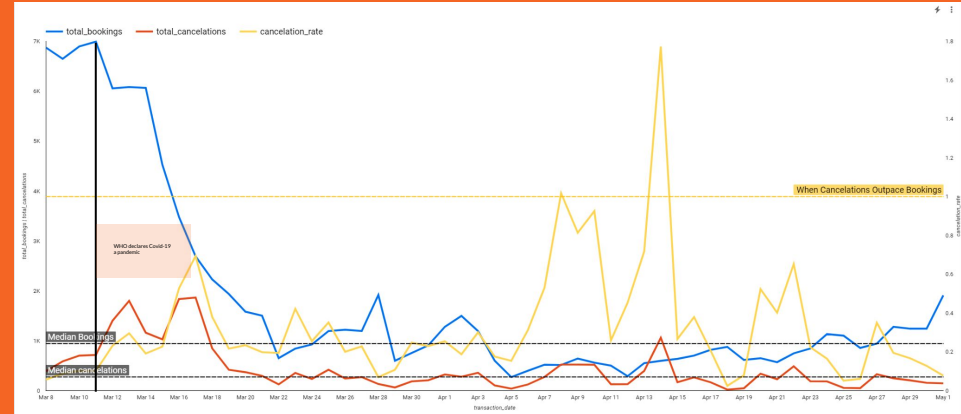
Soon after this announcement, bookings plummeted and cancellations skyrocketed.

Naturally, partner revenue goes down because

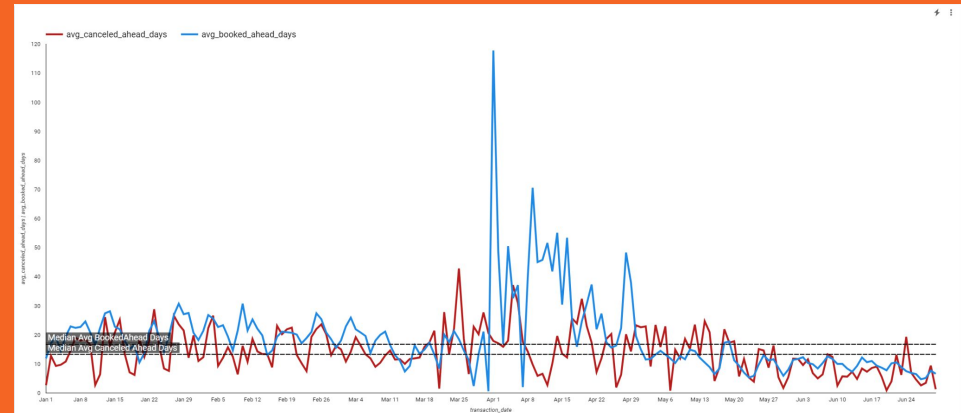
1. Fewer bookings means less demand, therefore less cash flow
2. Higher cancellations is missed revenue

Beyond the revenue angle, there is also an element of increased costs

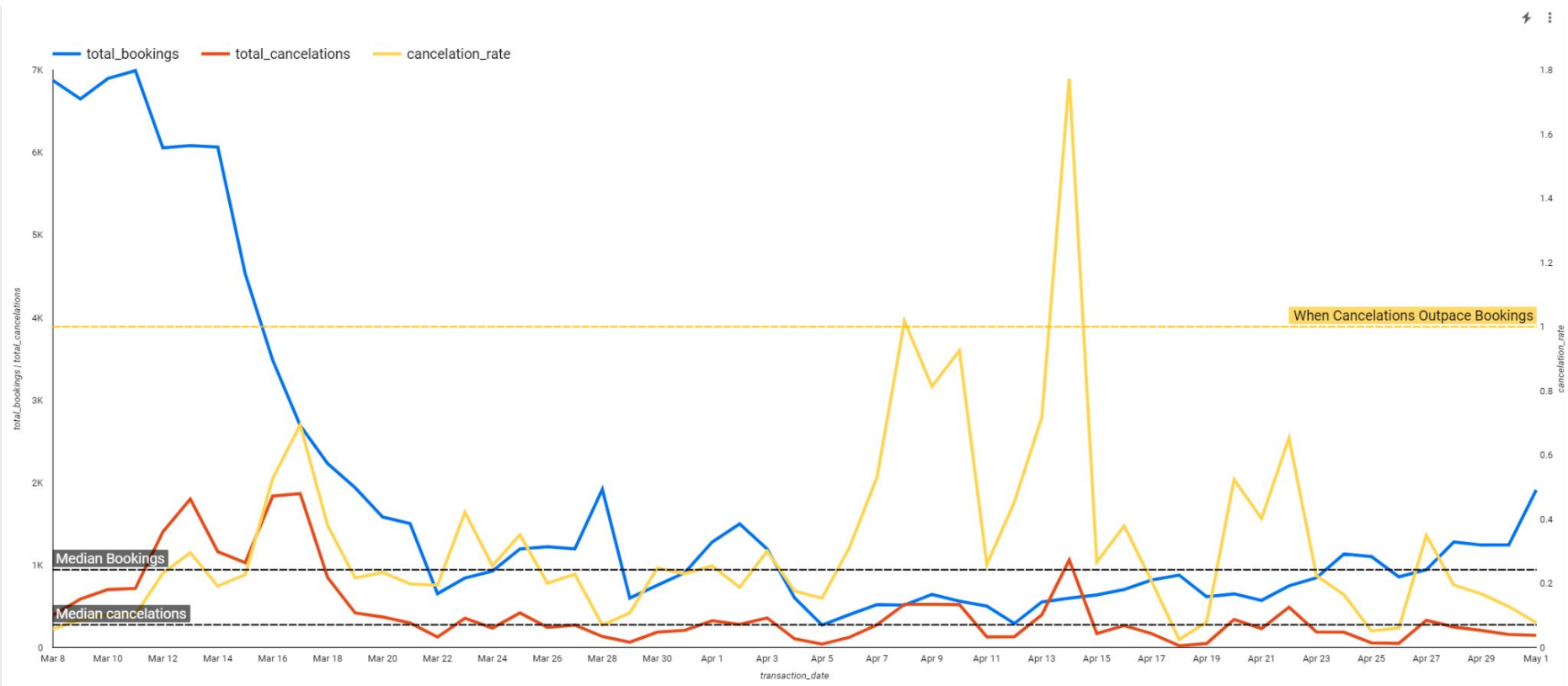
1. Before the pandemic, the lead time between booking and cancellations was significantly lower, which makes accommodations more predictable for providers.
2. Some buyers were optimistic to buy bookings way in advance, hoping to time the end of the pandemic. This provided some **revenue for the providers**.
3. However, as time went on, cancellations were **shorter noticed**, which means that providers may have overstocked on equipment, and **now have excess preparation that they are eating the costs for, and they are losing revenue due to refunds**.



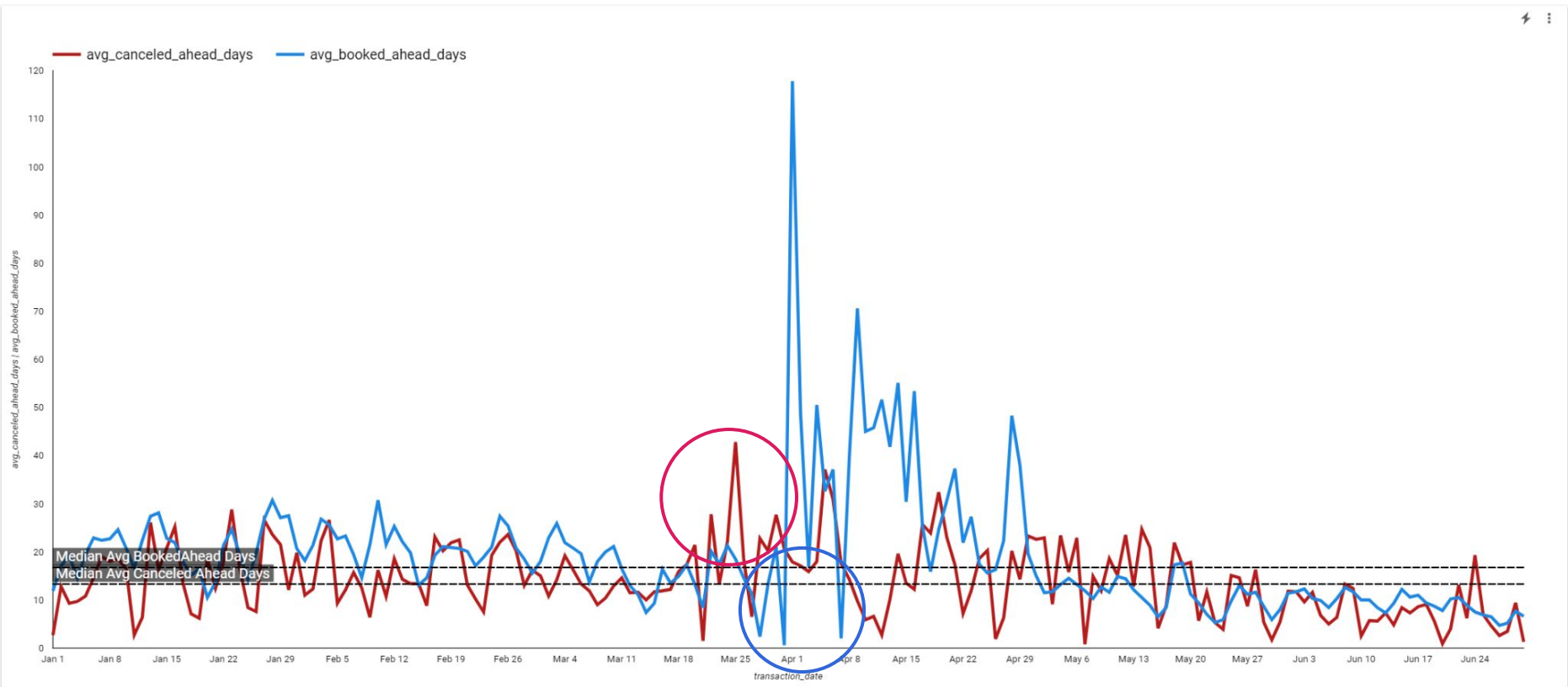
Bookings vs Cancellations
March 8, 2020 through May 1, 2020



Avg Booking Days Out vs Avg Cancellations Days Out
January 1, 2020 through June 30, 2020



Bookings vs Cancellations
March 8, 2020 through May 1, 2020



Avg Booking Days Out vs Avg Cancellations Days Out
January 1, 2020 through June 30, 2020

Question 1

How did the start of the pandemic impact our partners' business?

The WHO declared Covid-19 as a pandemic on [March 11, 2020](#)

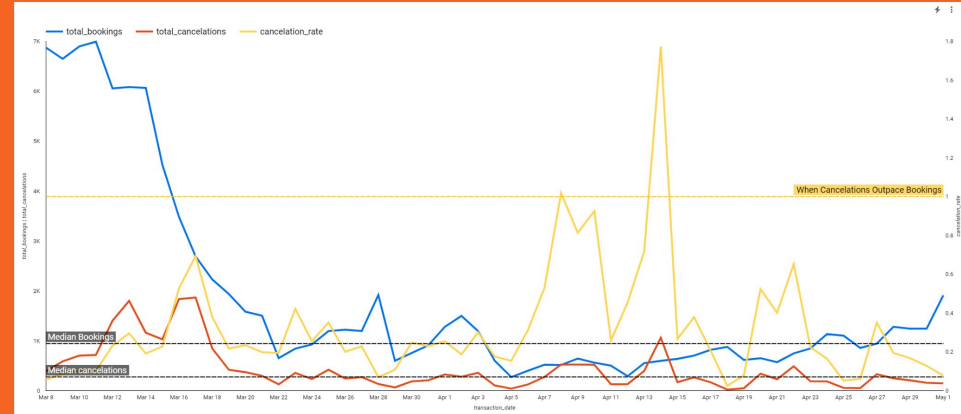
Soon after this announcement, bookings plummeted and cancellations skyrocketed.

Naturally, partner revenue goes down because

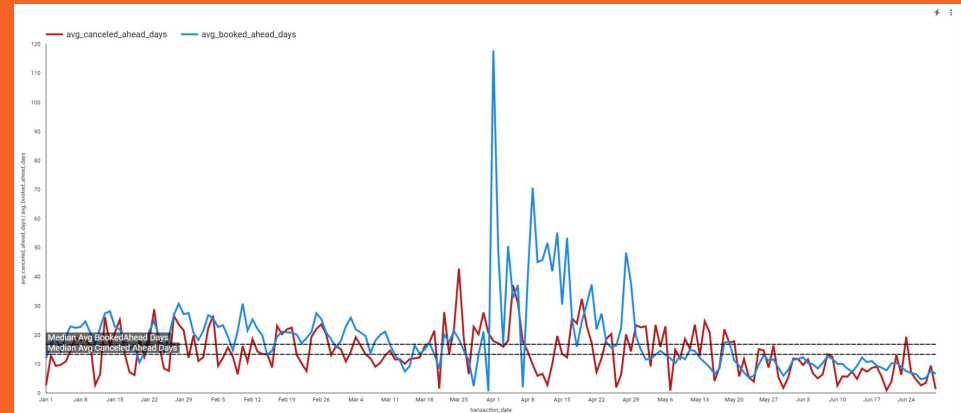
1. Fewer bookings means less demand, therefore less cash flow
2. Higher cancellations is missed revenue

Beyond the revenue angle, there is also an element of increased costs

1. Before the pandemic, the variance in lead time between booking and cancellations was significantly lower, which makes accommodations more predictable for providers.
2. Some buyers were optimistic to buy bookings way in advance, hoping to time the end of the pandemic. This provided some **revenue for the providers**.
3. However, as time went on, cancellations were **shorter noticed**, which means that providers may have overstocked on equipment, and **now have excess preparation that they are eating the costs for, and they are losing revenue due to refunds**.



Bookings vs Cancellations
March 8, 2020 through May 1, 2020



Avg Booking Days Out vs Avg Cancellations Days Out
January 1, 2020 through June 30, 2020

Question 2

What happened to customer booking and cancellation behavior after the start of the pandemic?

The WHO declared Covid-19 as a pandemic on [March 11, 2020](#)

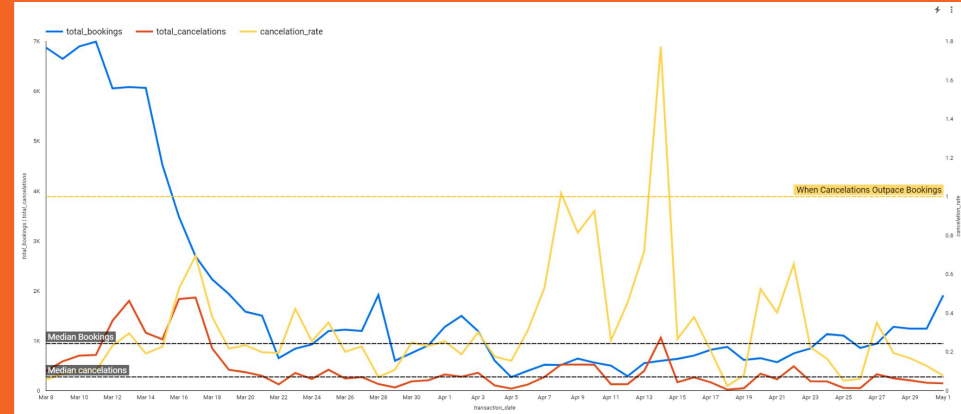
Soon after this announcement, bookings plummeted and cancellations skyrocketed.

Bookings

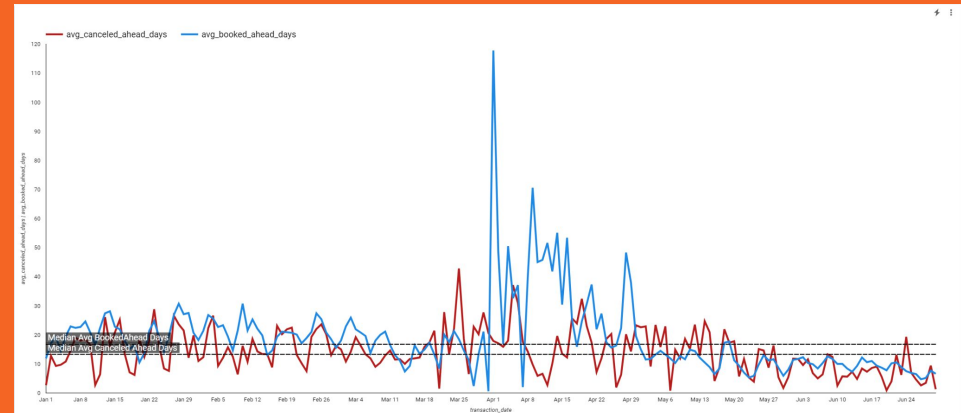
1. Bookings as a whole dramatically went down at the beginning of the pandemic.
2. However, of the bookings that were purchased, buyers were initially very optimistic that they can purchase experiences once Covid-19 was over. Some booking even as far as 4 months in advance, with multiple instances of booking at least 1 month out in advance.

Cancellations

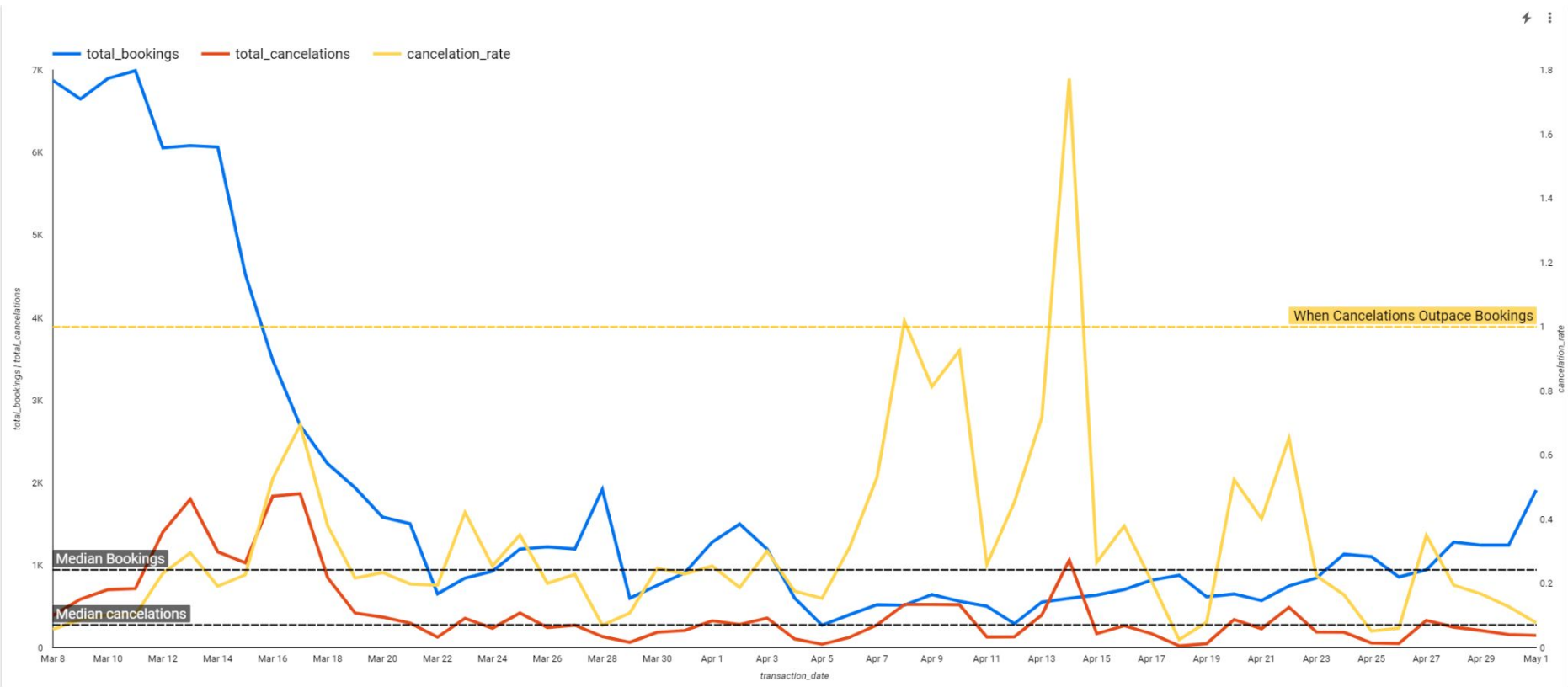
1. At the start of the pandemic there was a spike in cancellations, which tapers off to around median behavior. Perhaps the majority of people were optimistic that they didn't need to cancel their plans quite yet. However, around the 1-month anniversary of lockdowns, cancellations outpaced bookings before tapering back to median levels.
2. In the early days of the Pandemic, cancellations were leaving plenty of days ahead to cancel, above median levels.
3. However, as time went on, cancellations became increasingly short-noticed as people were probably realizing Covid-19 wasn't going anywhere, and they needed to cancel their plans closer to their activity date.



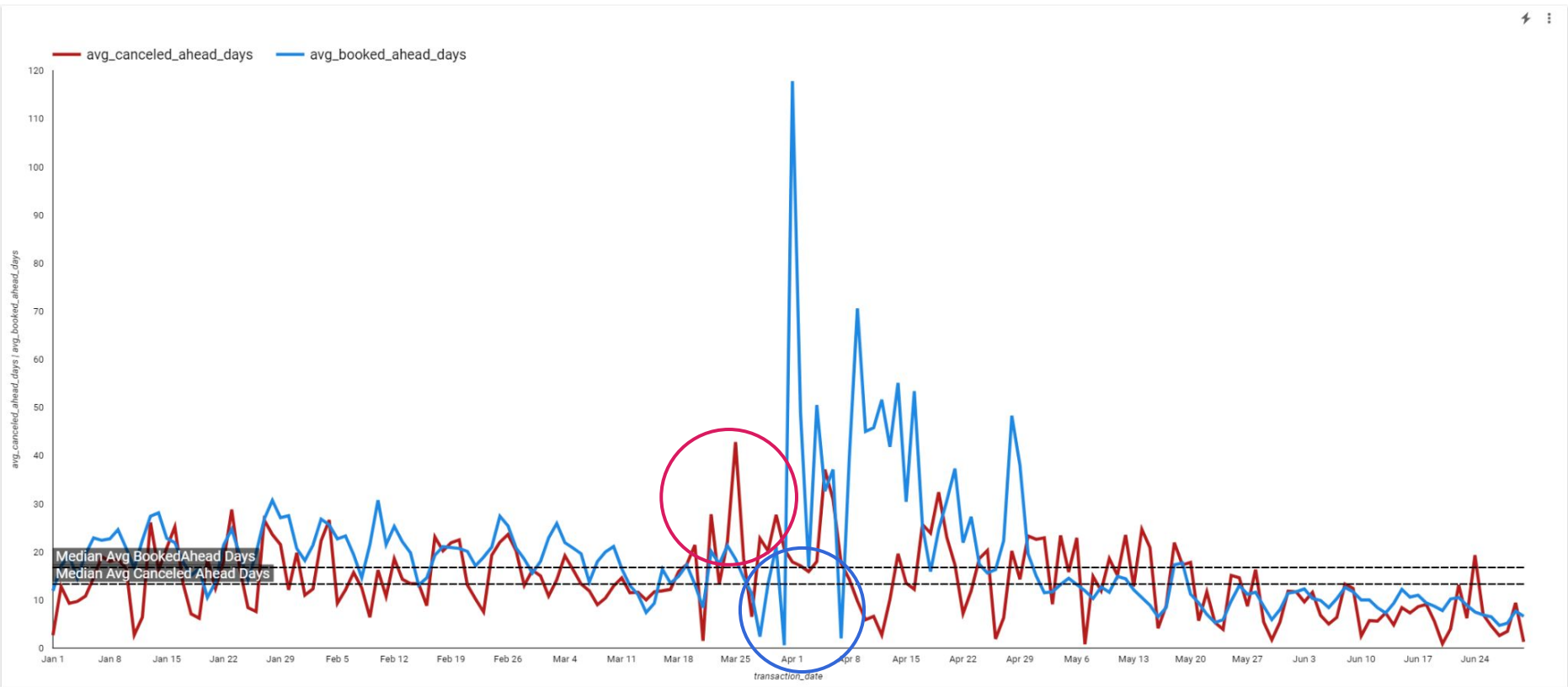
Bookings vs Cancellations
March 8, 2020 through May 1, 2020



Avg Booking Days Out vs Avg Cancellations Days Out
January 1, 2020 through June 30, 2020



Bookings vs Cancellations
March 8, 2020 through May 1, 2020



Avg Booking Days Out vs Avg Cancellations Days Out
January 1, 2020 through June 30, 2020

Question 2

What happened to customer booking and cancellation behavior after the start of the pandemic?

The WHO declared Covid-19 as a pandemic on [March 11, 2020](#)

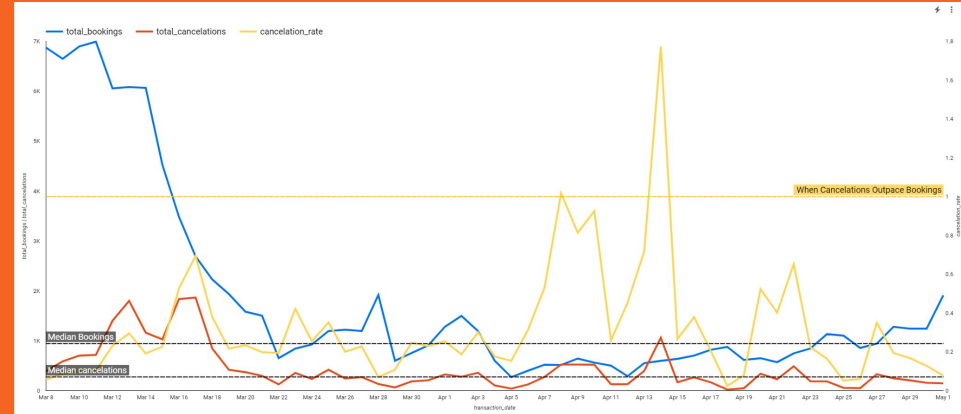
Soon after this announcement, bookings plummeted and cancellations skyrocketed.

Bookings

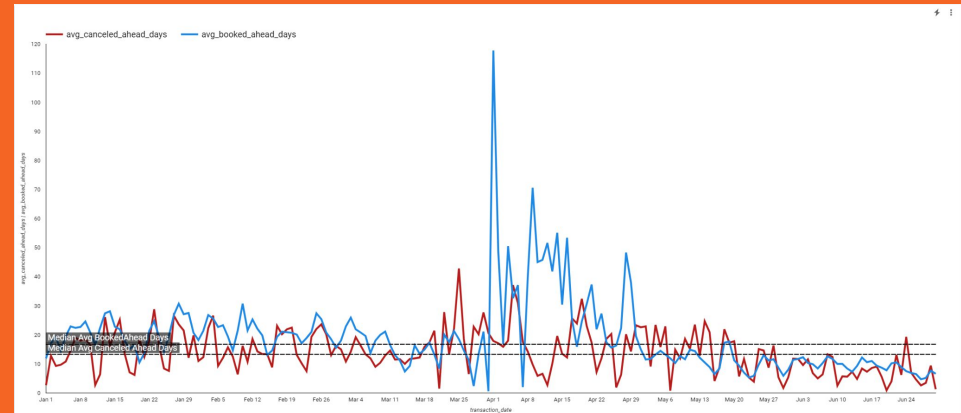
1. Bookings as a whole dramatically went down at the beginning of the pandemic.
2. However, of the bookings that were purchased, buyers were initially very optimistic that they can purchase experiences once Covid-19 was over. Some booking even as far as 4 months in advance, with multiple instances of booking at least 1 month out in advance.

Cancellations

1. At the start of the pandemic there was a spike in cancellations, which tapers off to around median behavior. Perhaps the majority of people were optimistic that they didn't need to cancel their plans quite yet. However, around the 1-month anniversary of lockdowns, cancellations outpaced bookings before tapering back to median levels.
2. In the early days of the Pandemic, cancellations were leaving plenty of days ahead to cancel, above median levels.
3. However, as time went on, cancellations became increasingly short-noticed as people were probably realizing Covid-19 wasn't going anywhere, and they needed to cancel their plans closer to their activity date.



Bookings vs Cancellations
March 8, 2020 through May 1, 2020



Avg Booking Days Out vs Avg Cancellations Days Out
January 1, 2020 through June 30, 2020

Questions 3

Explain the other pandemic events contributing to change in bookings or cancellations rates

→ Covid-19 Second Wave

The second wave of Covid started in [August 2020](#) and lasted through [February 2021](#)

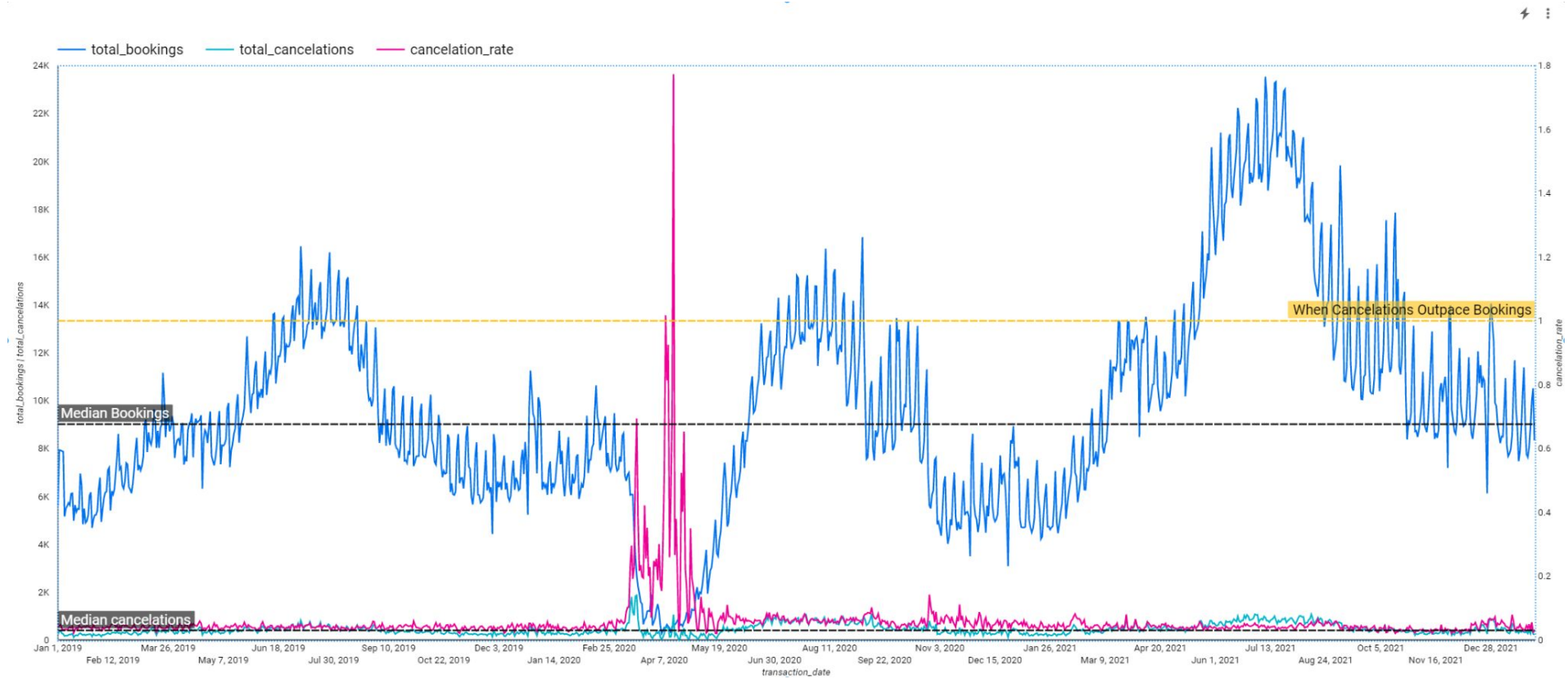
→ US hits 100MM Vaccine Milestone

[As more Americans became vaccinated](#), people were increasingly optimistic that they can return to a sense of normalcy, and are more interested in booking events.

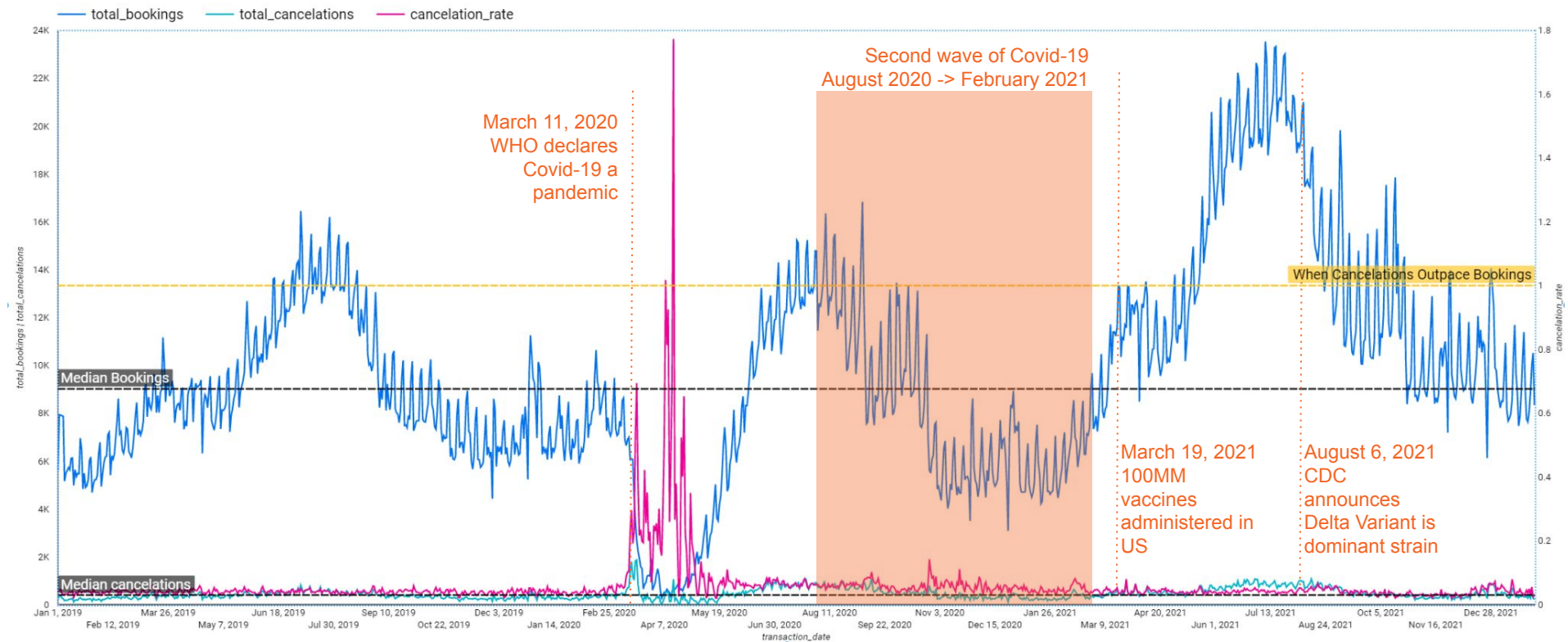
→ Delta variant becomes dominant strain in the US

[The delta variant had its notoriety for being twice as contagious as previous strains of Covid-19.](#)

This reputation definitely made people more weary of experiences.



January 1, 2019 Through January 31, 2022. Bookings vs Cancelations



Question 4

Please describe what the transaction_sequence and context_sequence fields are

```
SELECT
purchase_display_gid AS booking_id,
purchase_datetime AS transaction_datetime,
activity_date,
row_number() OVER (PARTITION BY purchase_display_gid ORDER BY purchase_datetime) AS
transaction_sequence,
row_number() OVER (PARTITION BY purchase_display_gid, context ORDER BY
purchase_datetime) AS context_sequence,
context,
purchase_status,
FROM `peek-services.abacus_read_only.transaction_records`
WHERE purchase_datetime >= '2019-01-01' AND purchase_datetime <= '2022-01-31'
```

Transaction_Sequence: Transaction Sequence groups all purchase_display_gids together, then orders the rows based on the purchase_datetime. It assigns a row number in chronological order based on the purchase_date for each activity_id. Once the function has gone through all purchase_datetimes for a specific purchase_display_gid, the row_number is reset to 1.

It would look something like this:

activity_name	Purchase_display_gid	purchase_datetime	transaction_sequence
"Kayak with Jeremy"	5730	2022-01-01 12:00:00PM	1
"Kayak with Jeremy"	5730	2022-01-01 :12:00:01 PM	2
"Surfing with Turtles"	6675	2022-03-04 :05:32:41 PM	1
"Blind-folded Poke Tasting"	2901	2022-02-15 :12:00:01 AM	1

Context_Sequence: Context_Sequence groups all purchase_display_gids together, within the gid, groups similar contexts (charge or refunds) together. Then, it assigns a row number for all contexts for a gid (starting with 1) based on chronological order of the purchase_datetime. Once exhausted, row_number is reset to 1. It resets to one when either the context changes or purchase_display_gid changes.

It would look something like this:

Purchase_display_gid	context	purchase_datetime	transaction_sequence
5730	charge	2022-01-01 12:00:00 PM	1
5730	charge	2022-01-05 :3:41:01 PM	2
5730	refund	2022-03-04 :05:32:41 PM	1
2901	refund	2022-02-15 :12:00:01 AM	1