

## Challenge 1

1-Write the SQL queries necessary to generate a list of the five restaurants that have the highest average number of visitors on holidays. The result table should also contain that average per restaurant.

Query:

```
Select
    Top 5
    store_id,
    Avg(1.0 * reserve_visitors) As [Avg Visitors]
From
    [dbo].[restaurants_visitors_cleaned] As R
Left Join
    [dbo].[date_info] As D
    On R.visit_date = D.calendar_date
Where
    D.holiday_flg = 1
Group By
    store_id
Order By
    Avg(1.0 * reserve_visitors) Desc
```

Query Result:

store_id	Avg Visitors
e89735e80d614a7e	8.500000
db80363d35f10926	6.904109
bcce1ea4350b7b72	5.500000
e053c561f32acc28	5.184615
bb09595bab7d5cfb	5.142857

2-Use SQL to discover which day of the week there are usually more visitors on average in restaurants.

Query:

```
Select
    Rank() Over(Order By Avg(1.0 * reserve_visitors) Desc) As Position,
    day_of_week,
    Avg(1.0 * reserve_visitors) As [Avg Visitors]
From
    [dbo].[restaurants_visitors_cleaned] As R
Inner Join
    [dbo].[date_info] As D
    On R.visit_date = D.calendar_date
Group By
    day_of_week
```

Query Result:

Position	day_of_week	Avg Visitors
1	Friday	4.311475
2	Wednesday	4.190618
3	Thursday	4.185873
4	Tuesday	4.108614
5	Monday	4.032590
6	Saturday	3.939040
7	Sunday	3.410092

Answer: Friday

3- How was the percentage of growth of the amount of visitors' week over week for the last four weeks of the data? You can solve this question using SQL or any other tool that you prefer. If you use other tools, please add your code or files.

```
Select
    Top 4 --Last four end/completed weeks
    DatePart(ISO_Week, calendar_date) As WeekNum,
    Sum(reserve_visitors) As [Current Week Visitors],
    Lag(Sum(reserve_visitors)) Over(Order By DatePart(ISO_Week, calendar_date)) As
[Previous Week Visitors],
    (Sum(1.0 * reserve_visitors) - Lag(Sum(1.0 * reserve_visitors)) Over(Order By
DatePart(ISO_Week, calendar_date))) / Lag(Sum(1.0 * reserve_visitors)) Over(Order By
DatePart(ISO_Week, calendar_date)) * 100 As [WoW % Growth]
From
    date_info As D
Left Join
    restaurants_visitors_cleaned As R
        On D.calendar_date = R.visit_date
Where
    DatePart(Year, calendar_date) = 2017
Group By
    DatePart(ISO_Week, calendar_date)
Having
    Count(Distinct calendar_date) = 7 --Displays completed weeks only
Order By
    DatePart(ISO_Week, calendar_date) Desc
```

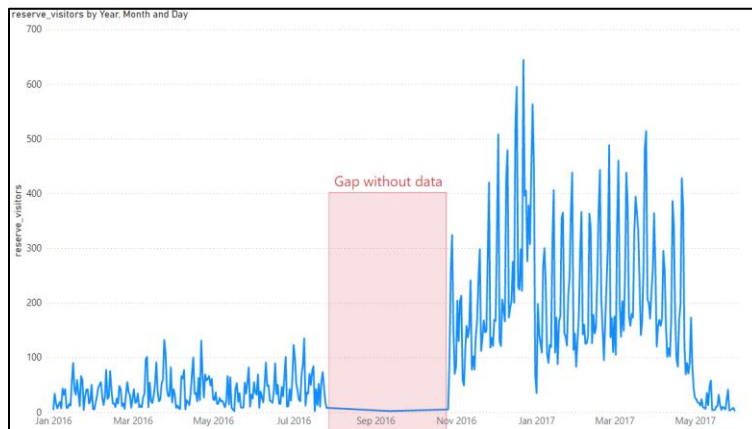
WeekNum	Current Week Visitors	Previous Week Visitors	WoW % Growth
21	63	78	-19.230700
20	78	170	-54.117600
19	170	130	30.769200
18	130	618	-78.964400

4- Forecast for the next six months, after the last date of the data, the sum of visitors of all the restaurants and validate the accuracy of your forecast.

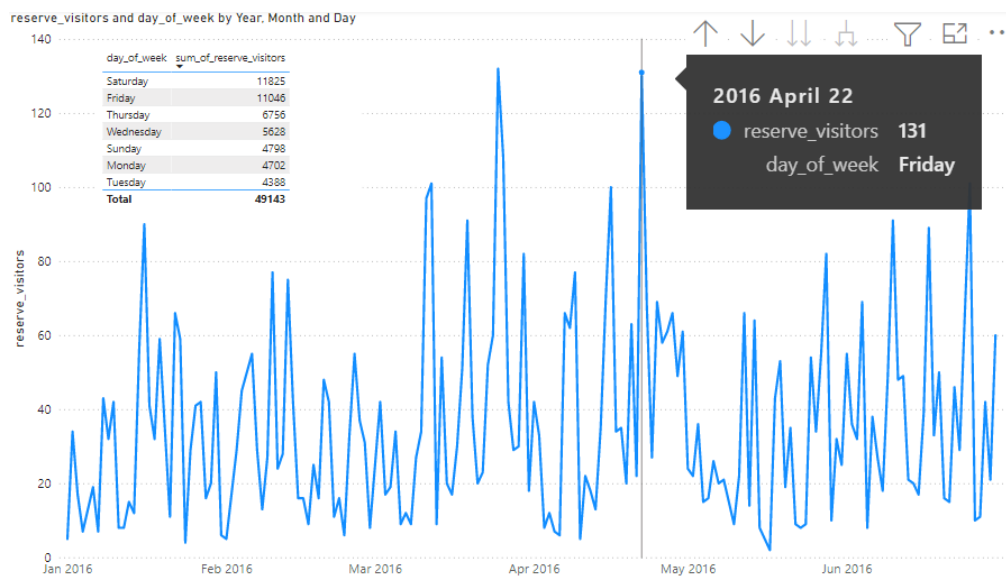
**I was not able to make the prediction.** But here I documented the process:

First, I did an exploration on the available data, and I found the picture up to July 2016 is completely different than from November 2016 onwards which has large fluctuations. Also, there is a gap with missing data, something seems to have changed drastically the behavior of the number of visitors.

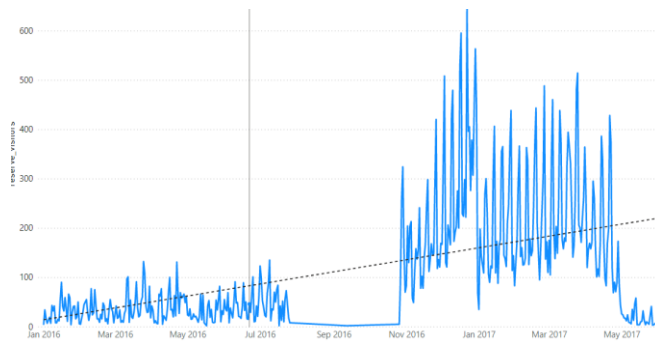
#### 1) Missing data from 7/26/2016 to 10/26/2016



#### 2) Seasonality is present, the spikes in the graph usually corresponds to Saturday or Friday.



### 3) Data shows an uptrend

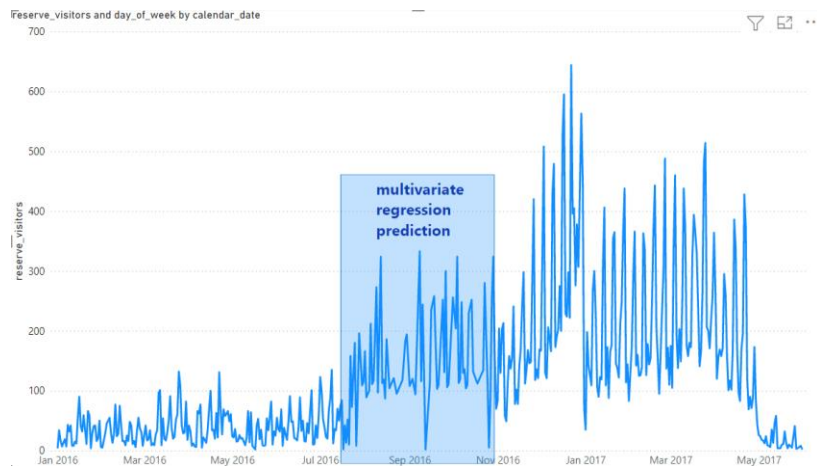


### 4) The trend has a significant drop in May 2017

Based on that, first, I predicted the missing values between 7/26/2016 and 10/26/2016 using multivariate regression. My dependent variables were the date, weekday and week number.

Dates will be converted to integer using Epoch timestamp and the weekday will start from 0=Monday to 6=Sunday.

The result just doesn't satisfy me because the prediction is saying that we had at least 100 visitors each day in that period.

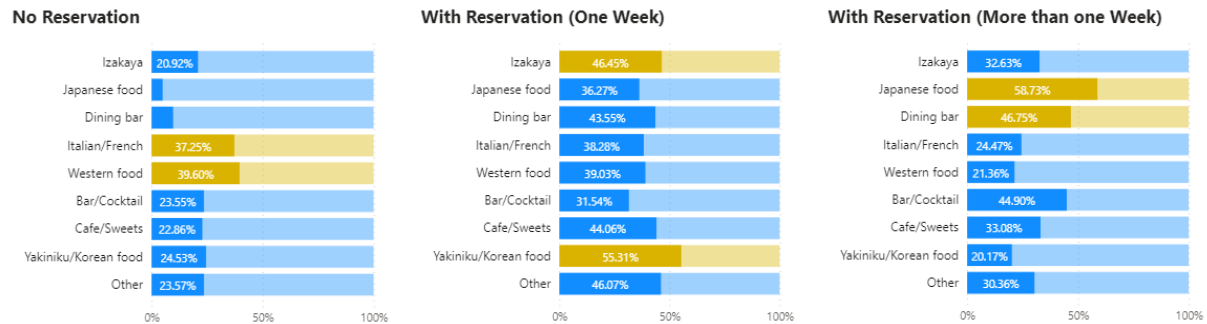


Then I tried to fit the data into ARIMA model, but I just couldn't find the optimal parameters to fit the model.

5-Based on the data and your ideas, plan strategies to double the total restaurant visitors in six months.

Based on data:

- Visitors at Western, Italian & French food restaurants tend to go without reservation.
- Asian food and Dining bar's visitors are more likely to make a reservation.



- Wednesday, Thursday, and Friday are the days with most visitors

day	Average of reserve_visitors
Friday	4.31
Wednesday	4.19
Thursday	4.19
Tuesday	4.11
Monday	4.03
Saturday	3.94
Sunday	3.41
<b>Total</b>	<b>4.04</b>

Ideas and strategies:

1. Send offers to customers who made reservations for 4+ people:
  - a. If the reservation was a holiday offer a discount for groups of 4+ people in the same or another restaurant.
  - b. If the reservation was not a holiday, send offers and discounts for groups periodically the next months.
2. Celebrate other countries holidays. In example: 25<sup>th</sup> April is holiday in Italy, so Italian restaurants could offer discounts or culture to incentive people to try other types of food.
3. Send recommendations to customers:
  - a. If they usually go to a certain type of food restaurants, recommend other restaurants with the same type of food that they don't know yet. If they are more open to try new flavors, incentive them to have discounts right after

*6-Imagine that these restaurants are in your city (and not in Japan), what other data would you want to join in order of get more insights to increase the visitors?*

Products consumed

Bill amount

If the user is local/native/tourist

Age

Gender

Weather data

Restaurant's visitors' capacity

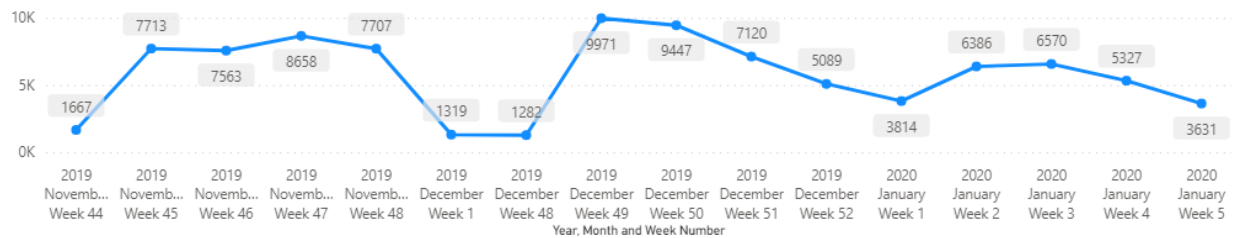
Data about cancelled reservations or if the customer didn't go.

Approximate of the meal duration

## Challenge 2:

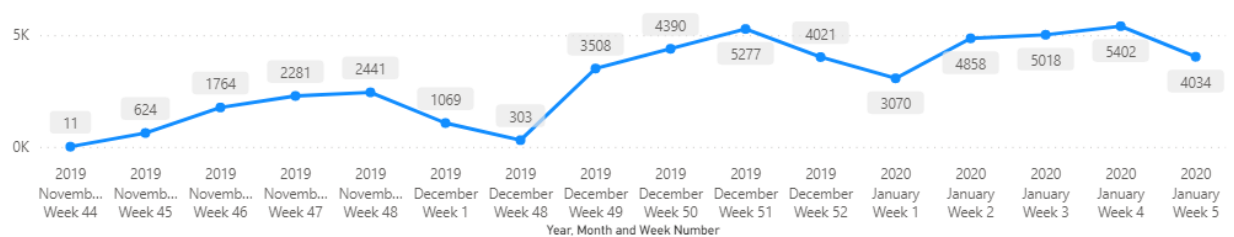
- 1- How many active users and new users do we have for each week of November 2019 to February 2020?

### New & Active Users: 87890



- 2- How many reengaged users do we have (Reengaged: active this week that didn't have an order last week but they did before that) for each week of November 2019 to February 2020?

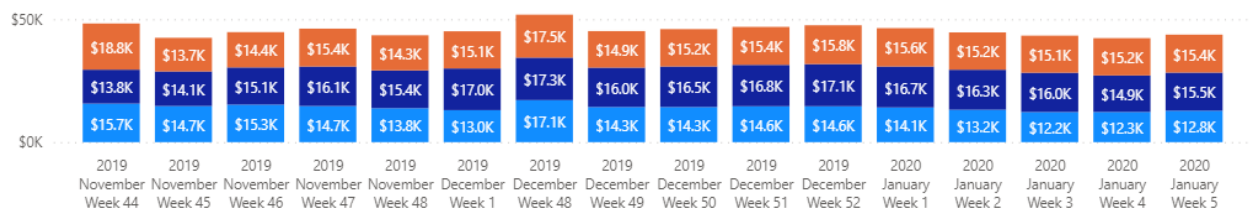
### Reengaged Users: 26853



- 3- What's the average GMV by type of user (Active, new, reengaged) for each week of November 2019 to February 2020?

### Average gmv by type of user

type of user ● Active ● New ● Reengaged



- 5- Based on the charts give your opinion/recommendations regarding to the different type of users.
  - Almost 90% of users sign up in the period nov-feb
  - Even though active users has the minimum average gmv of all three type of users, their income represents 2/3 of total gmv.
  - The number of reengaged users is increasing, would be interesting to investigate the cause.
  - Only 5% of users kept active during the period, as a recommendation: offering a monthly/yearly subscription to reengaged and new users could increase the percentage of active users.



Query Used for challenge 2:

```
Create view vw_user_orders_challenge2
As
With cte_user_orders(calendar_date, user_id, WeekNum, YearWeek, order_id, payment_type,
gmV, previous_order, days_after_prev_order)As(
Select
    calendar_date,
    0.user_id,
    DatePart(ISO_Week, calendar_date) As WeekNum,
    Concat(DatePart(Year, calendar_date), Right('0' + DateName(Week, calendar_date),
2)) As YearWeek,
    order_id,
    payment_type,
    gmV,
    Lag(0.order_date) Over(Partition By 0.user_id Order By 0.order_date) As
Previous_order,
    DateDiff(Day, Lag(0.order_date) Over(Partition By 0.user_id Order By
0.order_date), 0.order_date) As days_after_prev_order
From
    Dates As D
Left Join
    order_info As O
        On D.calendar_date = O.order_date
Inner Join
    user_info As U
        On O.user_id = U.user_id
)
Select
    C.*,
    U.first_order,
    O.order_date As first_order_date,
    DateDiff(ISO_Week, O.order_date, C.calendar_date) As weeks_since_first_order, --
this column will prevent overwrite new users as active users if they make two+ orders the
first week they joined
    DatePart(ISO_Week, order_date) first_order_weeknum
From
    cte_user_orders As C
Inner Join
    user_info As U
        On C.user_id = U.user_id
Inner Join
    order_info As O
        On O.order_id = U.first_order
```