Didi Business Intelligence Challenge

Part1 - Technical Skills

Q1: This challenge is an opportunity for you to show us a little bit of the great talent that we know you have. To solve this challenge, we share to you three csv files (use them wisely):

The file 'restaurants_visitors.csv' has the next five columns:
-id: Has the id of a restaurant in Japan and it is unique per restaurant.
-reserve_visitors: Has the amount of visitors that the restaurant of the row had at the date and time of the column visit_datetime.
-visit_datetime: Has the date and the hour of the day.
-visit_date: Has the date extracted from the column visit_datetime.
-reserve_datetime: Hour of the reservation.

The file 'date_info.csv' has the next three columns:
-calendar_date: Has the calendar date
-day_of_week: Has the day of the week that corresponds with the calendar date of the row. Format: ()
-holiday_flg: Has a dummy variable that value 1 if the calendar day is a holiday in Japan and 0 if is not.

The file 'store_info.csv' has the next five columns:
-store_id: Same id that in the ´restaurants_visitors.csv' file.
-genre_name: Genre (type) of the restaurant in the row.
-area_name: City or geological area of the restaurant.
-latitude: Latitude coordinate of the restaurant.
-longitude: Longitude coordinate of the restaurant.

Challenge:

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.

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

3- How was the percentage of growth of the amount of visitors week over week for the last four weeks of the data? Use SQL too.

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. You can solve this question using the tool that you prefer.

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

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?

7- How many channels can you think of downloading a DiDi Rides APP and how will you estimate the quality and cost of each channel?

8- We want to build up a model to predict “Possible Churn Users” for DiDi Rides APP (e.g.: no trips in the past 4 weeks). Please list all features that you can think about and the data mining or machine learning model or other methods you may use for this case.

You can present your answers in any way you want (pdf, ppt, visualization tool, etc.), some questions demand a code and others no. You must send us the mandatory SQL codes of your questions in any kind of text file. When apply, you can add your code, visualization, files, exploratory analysis, GitHub page or any tool and explanation that you consider useful to understand your ideas.