Shasa Business Case

Part 1: Data Wrangling & Analytics

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 files (use them wisely):

▪The file ‘store\_sales.txt’ has the next columns:

-id: Has the id of a store and it is unique per store.

-sales\_count: Has the count of sales that the store of the row had at the date and time of the column sales\_datetime.

-sales\_datetime: Has the date and the hour of the day.

▪The file ‘date\_info.txt’ has the next 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.

-holiday\_flg: Has a variable that value 1 if the calendar day is a holiday and 0 if is not.

▪The file ‘store\_info.txt’ has the next columns:

-store\_id: Same id that in the ´stores\_sales.csv’ file.

-store\_type: Weather (type) of the store in the row.

-zone\_name: City or geological area of the store.

-latitude: Latitude coordinate of the zone.

-longitude: Longitude coordinate of the zone.

Please consider the following points:

1. You may create a presentation to expose the main ideas of your answers and the solution path you used to get to them (try to be concise and very visual during your presentation, graphic in each case that you can)
2. Send us the answer and code you used in any kind of text file. Feel free to solve with any programming language that you feel more comfortable if not specified on question

Extra: If you prefer, you can structure your work on *GitHub* and share the link (this replaces part b)

Challenge 1

1. Write the SQL queries necessary to generate a list of the five stores that have the highest average number of sales on holidays. The result table should also contain that average per store.

2. Use SQL to discover which day of the week there are usually more sales on average in stores.

3. How was the percentage of growth of the amount of sales 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.

4. What do you think the sales amount will be for the next three months after the last date of the data?

5. What other data would you want to join in order of get more insights to increase the sales?

Part 2: User Analysis

To solve this challenge, we share to you two files (use them wisely):

▪The file ‘File\_order\_info.txt’ has the next columns:

Order\_date: Has the date when the order was created

User\_id: ID of the user that made the order

Order\_id: ID of the order

Payment\_type: Type of payment (Card, cash or other)

Sale\_Amount: Has the amount of the sale

▪The file ‘File\_user\_info.txt’ has the next columns:

User\_id: ID of the user

First order date: Has the date of the first order made by the user

Last order date: Has the date of the last order made by the user

First order\_id: ID of the first order

Last order\_id: ID of the last order

Challenge 2

1. How many active users (Active: Active this week that did have an order last week or before that) and new users (New: Active this week that didn’t have an order before) do we have for each week of November 2019 to February 2020?

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?

3. What’s the average sales amount by type of user (Active, new, reengaged) for each week of November 2019 to February 2020?

4. On your preferred tool (Excel, Python, R, etc.) create charts for each of your results and give your opinion/recommendations regarding to the different type of users.

Good luck!