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.

Respuesta con MySQL:

select A.store\_id,A.zone\_name,avg(B.sales\_count) as Venta from store\_info A, sales B,date\_info C where A.store\_id=B.id and date(B.sales\_datetime)= C.calendar\_date and C.holiday\_flg=1 group by A.store\_id,A.zone\_name order by avg(B.sales\_count) DESC limit 6;

Comparación de resultados con otras herramientas:

|  |  |
| --- | --- |
| MySQL | Interfaz de usuario gráfica, Texto  Descripción generada automáticamente con confianza media |
| Python | Texto  Descripción generada automáticamente |
| Power BI | Tabla  Descripción generada automáticamente |

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

Select A.day\_of\_week,avg(B.sales\_count) as Max\_ventaprom from date\_info A, sales B where A.calendar\_date=date(sales\_datetime) group by A.day\_of\_week order by avg(B.sales\_count) desc limit 1;

Comparación de resultados con otras herramientas:

|  |  |  |
| --- | --- | --- |
| MySQL | Python | Power BI |
| Imagen de la pantalla de un celular con letras  Descripción generada automáticamente con confianza baja | Texto  Descripción generada automáticamente | Interfaz de usuario gráfica, Aplicación  Descripción generada automáticamente |

1. 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.

select s.\*

, ifnull(round((s.ventas - n.ventas)/n.ventas \* 10000)/100, 0) as growth\_rate from (

SELECT year(sales\_datetime) as yearr,week(sales\_datetime) as Week,sum(sales\_count) ventas from sales

where year(sales\_datetime)=2017

group by yearr,Week

order by year(sales\_datetime), week(sales\_datetime) ) s left join (

SELECT year(sales\_datetime) as yearr,week(sales\_datetime) as Week,sum(sales\_count) ventas from sales

where year(sales\_datetime)=2017

group by yearr,Week

) n on n.Week = s.Week - 1 and n.yearr = s.yearr;

|  |  |
| --- | --- |
| MySQL | Python |
| Interfaz de usuario gráfica, Aplicación, Teams  Descripción generada automáticamente | Texto  Descripción generada automáticamente |

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

Las tiendas “HOT & COLD” representan las tiendas con mayor promedio de ventas durante el 2016

Durante meses de donde el clima es frío o lluvioso se destacó el promedio de ventas en las tiendas de “COLD” durante 2016

En 2017 fue más notable el promedio de ventas en las tiendas “HOT”

Se estima que el mes de abril de 2017 sea el mes con mayor promedio de ventas en el trimestre indicado

El comportamiento de ventas (promedio o monto) se mantendrá estable durante los últimos dos meses del año 2017 para las tiendas “HOT & COLD”

El comportamiento de ventas (promedio o monto) disminuirá para las tiendas “COLD”

El comportamiento de ventas (promedio o monto) disminuirá para las tiendas “HOT”

Utilizando un modelo de regression lineal se hizo un análisis tomando a las tiendas de Estado de México, cuyo clima es frio. Los resultados de crecimento de ventas es el siguiente:

Imagen que contiene Texto

Descripción generada automáticamente

Los resultados de crecimento de ventas para tiendas HOT es el siguiente:

Imagen de la pantalla de un computador

Descripción generada automáticamente con confianza baja

Los resultados de crecimento de ventas para tiendas HOT&COLD es el siguiente:

Texto

Descripción generada automáticamente

Los resultados del modelo de regression lineal coinciden con la hipótesis de que las ventas disminuirán para tiendas COLD y se mantendrán en cierto grado de estabilidad para tiendas HOT&COLD, sin embargo, contrario a las premisas dadas, se estima que las tiendas HOT incrementen sus ventas.

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

Incluir al análisis un registro de días inhábiles, feriados/festivos en México, ya que algunas celebraciones pueden marcar alguna tendencia temporal de moda a nivel nacional, regional o local. Un análisis socioeconómico en las zonas donde se localizan las tiendas pueden ayudar a identificar otros patrones de comportamiendo de ventas que pueden ser utilidad, por ejemplo, en una zona donde el nivel de vida está por encima del promedio en México o en una zona de alta capacidad turística, los clientes pueden tener preferencia por la calidad o la moda, mientras que en un caso contrario, las ventas estarían limitadas en cantidad y precio.

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) 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?

SELECT year(last\_complete\_time) as Year,week(last\_complete\_time) as Week, count(distinct(user\_id)) Activos from user\_info where yearweek(first\_complete\_time)=yearweek(last\_complete\_time - INTERVAL 1 WEEK) and date(last\_complete\_time)>=’2019-11-01’ and date(last\_complete\_time)<=’2020-02-01’ group by year(last\_complete\_time),week(last\_complete\_time);

Comparación entre MySQL y Python para obtener activos

|  |  |
| --- | --- |
| MySQL | Python |
| Imagen que contiene foto, sostener, tabla, hombre  Descripción generada automáticamente | Texto  Descripción generada automáticamente con confianza media |

Usando MySQL para obtener nuevos:

SELECT year(A.last\_complete\_time) as Year,week(A.last\_complete\_time) as Week, count(distinct(A.user\_id)) Nuevos from user\_info A, order\_info B where date(A.first\_complete\_time)=date(A.last\_complete\_time) and date(A.last\_complete\_time)>=’2019-11-01’ and date(A.last\_complete\_time)<=’2020-02-01’ and date(A.last\_complete\_time)=(B.order\_date) group by year(A.last\_complete\_time),week(A.last\_complete\_time);

Comparación entre MySQL y Python para obtener nuevos

|  |  |
| --- | --- |
| MySQL | Python |
| Un conjunto de letras blancas en un fondo blanco  Descripción generada automáticamente con confianza media | Texto  Descripción generada automáticamente con confianza baja |

SELECT year(A.last\_complete\_time) as Year,week(A.last\_complete\_time) as Week, count(distinct(A.user\_id)) Nuevos from user\_info A, order\_info B where date(A.first\_complete\_time)=date(A.last\_complete\_time) and date(A.last\_complete\_time)>=’2019-11-01’ and date(A.last\_complete\_time)<=’2020-02-01’ and date(A.last\_complete\_time)=(B.order\_date) group by year(A.last\_complete\_time),week(A.last\_complete\_time);

1. 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?

select year(A.last\_complete\_time) as Year,week(A.last\_complete\_time) as Week ,count(distinct(A.user\_id)) from user\_info A, order\_info B where yearweek(A.first\_complete\_time)=yearweek(last\_complete\_time) and yearweek(A.last\_complete\_time)>yearweek(B.order\_date) and date(A.last\_complete\_time)>=’2019-11-01’ and date(A.last\_complete\_time)<=’2020-02-01’ group by year(A.last\_complete\_time),week(A.last\_complete\_time);

Comparación entre Mysql y Python

|  |  |
| --- | --- |
| Power BI | Tabla  Descripción generada automáticamente |
| Python | Imagen que contiene Texto  Descripción generada automáticamente |

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

|  |  |
| --- | --- |
| Activos | Interfaz de usuario gráfica, Texto  Descripción generada automáticamente con confianza media |
| Nuevos | Interfaz de usuario gráfica  Descripción generada automáticamente con confianza media |
| Reenngaged | Texto  Descripción generada automáticamente |

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.

Carpeta donde se localizan los archivos necesarios:

S:\BI\Dulce Odete Jardon

Good luck!