CASE STUDY – *Data Scientist*

Welcome to the *EMP Sales Challenge*.
Your application made us curious to know more about your skills and we would like to take you to the next step.
Here is your chance to convince us that you are the right person to join the *Data, Business Intelligence and Analytics* team at EMP.
We wish you good luck and hope you can enjoy the quest!

SUBMISSION DEADLINE:
10 days after receiving the challenge

HOW TO SUBMIT:
please send us your code and results in any format you prefer (slides, jupyter notebook, visualizations, text document, etc.) via email.

THE CHALLENGE

Background
At EMP we deal daily with a great amount of data from different sources (from web, shop, inventory, purchasing, logistics, finance, etc.), produced by the customer journey.
In this challenge, we will focus on the sales data, produced by customers visiting and purchasing items from our web shop (emp.de in Germany).

Tasks

1. Describe the dataset and eventual anomalies you find.
2. Which patterns do you find in the purchasing behavior of the customers?
3. What are the categories and genres which customers are mostly interested in?
4. Split customers in different groups based on their purchasing behavior.
    a. Justify your choice for your adopted method(s) and model(s).
    b. Describe the defined customer groups. What are the features which are driving the differentiation amongst the different groups?
    c. Give suggestions on how the business should treat these clusters differently.
5. *(optional)* Assuming that the '*Category_Reporting'* tells you the category of all the items in that order, predict:
    a. The number of items per category which will be ordered on a monthly basis for the rest of May 2021.
    b. The number of returns for the rest of May 2021.
6. *(optional)* As, at this point in your analysis, you are the dataset expert, suggest any ideas (initiatives, further analyses) you might have in mind which can be helpful for the business.

Note: *you will need to make assumptions for completing this task. Also, keep in mind that there is no 'right' or 'wrong'.*

DATA DESCRIPTION

Column	Description
CustomerHash	CustomerID
OrderHash	OrderID
Date_Order	Date the order was placed
Customer_Age	Customer Age
Gender	Customer Gender
Account_Age	Indicates how old is the customer account (in floor-rounded years)
AffinityProductGender	Gender preference with Items/Products
Category_Reporting	Category of the "orderStarter" Item. That is, the Category of the first product in the basket that was actually bought
Genre	Genre of the "orderStarter" Item. That is, the Genre of the first product in the basket that was actually bought
ClientType	Device Type
PartnerProgram1	Usage of PartnerProgram 1
PartnerProgram2	Usage of PartnerProgram 2
FirstOrder	Indicates if the order is the very first order of a customer
Marketing_Channel	The attributed marketing channel
Pieces_Ordered	Number of ordered items
Pieces_Outbound	Number of shipped items
Pieces_Returns	Number of returned items
Pieces_Fulfilled	Number of items actually fulfilled (after returns)
OrderValue	The actual cost amount of the order
Revenue_Goods	Fulfilled OrderValue after returns
Delivery_Value	OrderValue of items delivered
Return_Value	OrderValue of items returned
Discount_Total	Sum of all discounts

OrderProfit	Profit per customer order