Challenge - Data Scientist – Product

Introduction

Congratulations for making it to the Challenge stage of the application process! The goal of this task is to objectively assess your technical prowess and your ability to solve business problems using data exploration, statistical analysis and potentially machine learning.

Please note that the topic, data and problem are reflective of cases we have solved in the past.

Deadline: Max 7 days

Deliverable:

1. **Model Endpoint / API–** Please provide us a model endpoint + pre-processing which we will use to score the model.
2. **Python / Jupyter Notebook –** Please provide a jupyter notebook with answers to the task questions

Evaluation:

We will evaluate your model on the accuracy on the evaluation dataset.

The Task

The datasets contain examples of our business problems. One of the largest areas we are concerned with is RTO (Return to Origin)

Utilizing this data, your business knowledge and potentially your interests, please accomplish the following tasks:

1. Please write a code and create a visualisation to answer the following question:
    a. How are the RTO customers different from Non RTO orders by
        i. Discount
        ii. Shipping Charges
        iii. Promo / coupons
2. Visualise RTO behaviour with respect to the value of transactions
3. Define a target metric for RTO
4. Using your logic from above, build a model (heuristic/statistical/ML) to RTO and non RTO orders
    a. Note that features which are directly correlated with your target metric could lead to overfitting
5. Now suppose we use your model to identify RTO users and implement some business actions to try to control RTO (commonly known as RTO reduction)

a. How would you set up a test/experiment to check whether we are actually reducing RTO?
b. What metrics and techniques would you use to assess the impact of the business action?

NB – Please execute your solves in Jupyter Notebook using the datasets provided in the 'data' folder as CSV files

---

The Datasets

1. codorders010119-310319 – DS
    a. contains orders from Jan'19 till Mar'19
2. codorders-010419-300619 – DS
    a. contains orders from Apr'19 till Jun'19
3. codorders-010719-300919 – DS
    a. contains orders from Jul'19 till Sep'19

Features / Variables

Variable Name	**Description**
Id	Order ID
Name	Name of the user
Sequence	Number of orders that the use has done from start
Item(s)	Number of different Items in the Order
Purchased On	Purchase Date
Payment Method	Method of Payment
Total Qty Ordered	total quantity in the order
Total Amount Payable	total amount
State	State
City	City
Status	Status of the order
Shipped Date	Shipped date
Delivered Date	Delivered date
Transporter Name	Transporter Name
AWB Status	AWB Status
Shipping Charges	Shipping Charges
Discount	Discount
Credits	credits used in the order
Promocode	Coupon Code / Promotion Code
Order Type	Order Type
Payment Status	Payment Status
Shipping Address Street	Delivery address
Pincode	Delivery Pincode
Address Modified	Whether address modified