Data Scientist Interview - MBM.

Part – 1:

Objective:

The purpose of this challenge is to test your ability to **code**, **present analysis** and **write a scalable solution** in code **repository** which can be shared with us to better gauge your ability to work as a Data Scientist with us.

The primary objectives of this challenge are as follows:

1. Model Training
2. Model Validation and Testing
3. Model Prediction

Challenge:

The attached d**ata1** file is extracted from a live CRM/ Marketing platform comprising an anonymous 10K records. This feature/training set is **sampled from** a much larger training set for Machine Learning model training. To solve this challenge you are free to use any **Python packages from available ML toolkits.**

This is unlabelled or unclassified data using which you are expected to provide an approach to solving a particular problem using a **combination of Clustering and Classification**. The data extract comprises of 3 main categories of features:

1. Account Features
2. Personal ID Features
3. Redemption Activity

The objective of this task is to **classify** **consumers** that might be bad i.e. **fraudulent**. From **all the features** of training data you are expected to perform the **Clustering** and subsequently **Classification** to **visualize the** **Consumer group** and **classify** **the** **potentially bad consumers.**

**A summary** of technical challenge tasks are as follows:

1. Clustering
2. Classification
3. Model **metrics** and performance measures
4. What, Why and How of the work presented

Part – 2:

Objective:

The purpose of this challenge is to test your ability to do **research skills** and do **regression/time series analysis**.

The primary objectives of this challenge are as follows:

1. Model Training
2. Model Validation and Testing

Challenge:

The attached data2 file is extracted from **historical media spend by month across different channels**. To solve this challenge, you are free to use **any R packages from available ML toolkits.**

This is a time series data analysis which you are expected to provide an approach to solving a marketing mixed model**.** **Marketing mix modelling (MMM)** is statistical analysis such **as multivariate regressions on sales and marketing time series data** to estimate the impact of various marketing tactics (marketing mix) on sales. The data extract comprises these main categories of features:

* **Media Spend by Media Channel**
* **Price Categories of Different Product Variants**
* **Target** **Sales (Y Variable)**

From all the features of data you are expected to perform time series regression to calculate **the contributions** of each media channel and the sales **return on ad spend (****ROAS).**

A summary of technical challenge tasks are as follows:

1. Data Exploration and Visualization
2. Data Preparation
3. **Calculate Adstock Affect:** **the prolonged or lagged effect** of advertising on consumer purchase behaviour
4. **Decay Effect** if present
5. Model Training
6. Model metrics and performance measures

Summary:

Please submit the work by Monday. **Please send us the completed code repository with GIT repository** and if possible presentation slides for the findings.