

D+A BUSINESS CASE

# Dataset Challenge

PepsiCo Confidential

## **Introductory Note:**

This case study is entirely fictitious and has been developed specifically as an assessment tool, so some conclusions may not be credible. You should ensure that your presentation is of a standard that is clear, commercially credible and suitable for a senior manager audience.

## **Objective:**

The objective of this problem is to determine the factors that can drive dealers to meet the minimum sales objectives

# Dataset description (1)

This dataset (deliveryDatasetChallenge.json) includes data from 29,389 truck routes, carried out by 29,389 employees, and the following variables:

- anonID: the route unique identifier
- birthdate: the employee's birthdate
- routeDate: the day the route was carried out (between 20/05/11 to 20/05/22).
- region: the world's area where the route was executed:
  - NA: North America
  - AMESA: Africa, Middle East and South Asia
  - APAC: Asia Pacific, Australia/New Zealand, and China
  - LATAM: Latin America
  - Europe
- gender: employee's self-determined gender (F - Female, M - Male, X - Non binary)
- areaWealthLevel: economical area development, compared to the whole country (Low, Mid or High)
- areaPopulation: population of the covered area, in thousands
- badWeather: bad weather conditions in the area, such as precipitations or strong wind
- weatherRestrictions: affectations in the area due to weather
- routeTotalDistance: the route distance covered (kms)

# Dataset description (2)

- numberOfShops: total shops we cover in the area
- marketShare: % of market share that the company has in the area in its categories.
- avgAreaBenefits: weekly economic benefit in the area (in thousand \$)
- timeFromAvg: time spent in the route, compared to the average (negative would mean that it took shorter than the average time)
- advertising: investment in POS (point of sales) material in stores (from 0 meaning no investment to 3 (high investment))
- employeeLYScore: rating the score from last year\* (from 1 to 5, being 5 the highest one). \*New employees have 3 by default.
- employeeTenure: time the employee has been in the company
  - 0: joined in the last 12 months
  - 1: From 1 to 3 years
  - 2: From 3 to 10 years
  - 3: More than 10 years
- employeePrevComps: number of companies the employee worked previously developing the same function (3 means 3 or more).
- success: the dealer has distributed at least the expected value (1) or less (0).

# Assignment

You are requested to deliver the following:

## NOTEBOOK

A Python 3 notebook used to solve a classification problem in which you have to build a model to guess if the dealer could meet the sale objective.

Please make sure it is as readable as possible, and be prepared to having to explain it to the assessors if necessary.

## CLASSIFICATION

The success classification for the NA IDs: a .csv including only anonID and success columns.

anonID	success
AAWQ8310	1
ADMR9131	1
FLWM1930	0
XKSO3719	1

## VIDEO

A video explaining briefly the code implementation and the modelling part, as well as the results obtained.

They must include at least a ROC/AUC measures and a confusion matrix for the most proper threshold.

Please feel free to highlight interesting findings or suggestions regarding exploratory data analysis if this could be relevant.

# VIDEO REQUIREMENTS

- It is highly recommended to use a presentation tool such as PowerPoint.
- The video must show the screen and a voice over the presentation. It is not allowed to record your face.
- You can use software to film the screen, such as OBS (<https://obsproject.com/>).
- The formats can be .mp4, .mov, .wmv and .avi, and it must be uploaded to a file transfer platform due to size restrictions. Alternatively, they can be uploaded in a video streaming platform.
- It cannot include any PepsiCo references.
- This video must be shorter than 6 minutes.

# Recommendations

## DURATION

The estimated time to complete the test: 3-4 hours

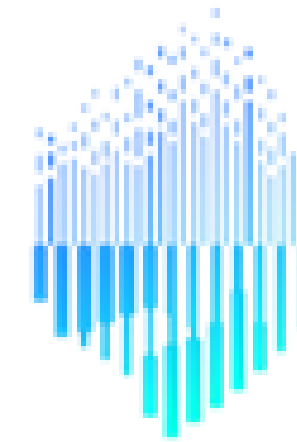
## DEADLINE

You will have 3 days to complete the test, including the day that is sent

## DELIVERY

Please send your Business Case to your Talent Acquisition point of contact

**THANK YOU**



DATA + ANALYTICS  
PEPSICO