

Data Analyst Case study

Sendy is the leading African Platform that enables businesses & individuals to do more. We are on a mission to empower people and businesses by making it easier to trade. We believe that everyone should be able to trade and thrive in Africa. No one should be left out of trading because the cost is too great or the tools are inaccessible. So we are building easy tools to empower businesses and people. Tools that shorten the distance between seeing an opportunity and making a living from it.

Case Study Tasks:

Please prepare a presentation (and csv for the second task) that includes the answers to the tasks described below. You'll find more information for the tasks attached.:

1. Using the attached dataset (train.csv) analyse the data and visualize the most important aspects using your preferred method. Furthermore, share three ideas on how to increase the % of Partners that accept. Document your steps where needed.
2. Predict the reaction of a partner rider to an order: is a rider most likely to ignore, decline or accept the dispatch they receive? Use train.csv to train your model and test.csv to predict the missing value ('target'). Please document your steps and method used. The csv "SubmissionExample" will help with the format.

Key success factors:

- Be precise and structure your answers in a clear manner. Don't beat around the bush.
- Presentation of the answers is key. Show us what you did.
- Showcase your creativity and have fun doing it!

About the Data

The dataset provided includes order dispatch details and rider metrics based on orders made on the Sendy platform. The challenge is to predict the reaction of a partner rider to an order: is a rider most likely to ignore, decline or accept the dispatch they receive?

The training dataset provided here is a subset of over 200,000 dispatches and only includes direct orders (i.e. Sendy “express” orders) with bikes in Nairobi. All data in this subset have been fully anonymized while preserving the distribution.

The objective of this challenge is to create a machine learning model that will predict whether a rider will accept, decline or ignore an order sent to them.

Files available for download:

- **Train.csv** - contains the target. This is the dataset that you will use to train your model.
- **Test.csv** - resembles Train.csv but without the target-related columns. This is the dataset on which you will apply your model to.
- **Riders.csv** - information on the riders available.

Definitions

Dispatch Data

- **ID** - Unique ID for each order request
- **order_id** - Unique number identifying the order
- **client_id** - Unique number identifying the customer on a platform
- **client_type** - Specifies the customer type (Business or Personal)
- **rider_id** - Unique number to uniquely identify the rider

- **rider_license_status** - Identifies riders who have a license to access restricted areas i.e. 0 (Cannot access a restricted area) and 1 (Can access a restricted area)
- **rider_carrier_type** - Identifies the box option that a rider currently has i.e. 0 (No Box option) and 1 (Box option)
- **rider_amount** - The earnings a partner would earn if they successfully complete an order.
- **order_license_status** - Identifies orders that require a pick-up or drop-off in a restricted area i.e. 0 (Restricted area) and 1 (Non-Restricted area)
- **order_carrier_type** - Identifies the box option the customer specified while placing their orders i.e. 0 (No box option), 1 (Box option), 2 (Any option)
- **vendor_type** – For this competition limited to bikes. However, in practice, Sendy's service extends to Vans and Trucks.
- **Pickup Latitude and Longitude (pickup_lat and pickup_long)** - Latitude and longitude of pick up location
- **Destination Latitude and Longitude (drop_off_lat and drop_off_long)** - Latitude and longitude of delivery location
- **Rider Latitude and Longitude (rider_lat and rider_long)** - Latitude and longitude of the Rider at the time of dispatch.
- **target** - The reaction of a rider in regards to a particular dispatch. Did a rider ignore (0), decline (1) or accept (2) a dispatch?

Dispatch times

- **dispatch_day** - Day of Month i.e. 1-31
- **dispatch_day_of_week** - Weekday (Monday = 1)
- **dispatch_time** - Time of day the dispatch was sent out to the riders