# **Context and Setup**

## Kick-start terminal instructions

*# Create challenge folder*

mkdir -p ~/code/WLaCoutur/04-Decision-Science/01-Project-Setup/data-context-and-setup && cd $\_

*# Download challenge*

curl -s -H "Authorization: Token oDBkU2uqVbPdWJmtDXwiVtJW, User=WLaCoutur" "https://kitt.lewagon.com/camps/1917/challenges/setup\_script?path=04-Decision-Science%2F01-Project-Setup%2F01-Context-and-Setup" | bash

*# Open challenge folder in your text editor*

code .

ℹ️ Kitt no longer picks up Github pushes two weeks after camp has finished.

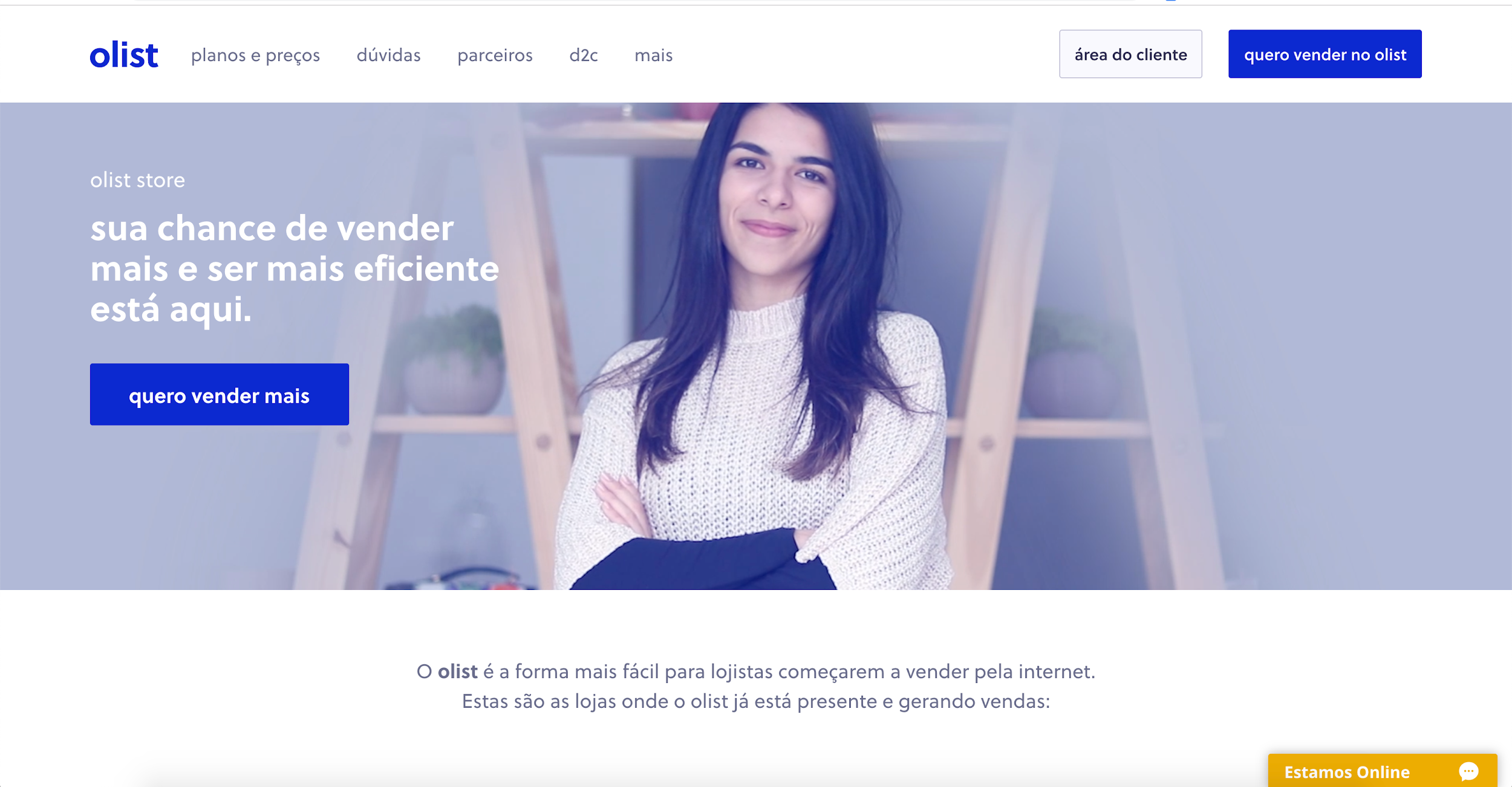
**Report an issue**

## **Objectives of the module**

We will analyze a dataset provided by an e-commerce marketplace called Olist to answer the CEO’s question:

How could Olist increase its profit?

## **About Olist 🇧🇷**

****

Olist is a leading e-commerce service that connects merchants to main marketplaces in Brazil. They provide a wide range of offers including inventory management, dealing with reviews and customer contacts to logistic services.

Olist charges sellers a monthly fee. This fee is progressive with the volume of orders.

Here are the seller and customer workflows:

Seller:

* Seller joins Olist
* Seller uploads products catalogue
* Seller gets notified when a product is sold
* Seller hands over an item to the logistic carrier

👉 Note that multiple sellers can be involved in one customer order!

Customer:

* Browses products on the marketplace
* Purchases products from Olist.store
* Gets an expected date for delivery
* Receives the order
* Leaves a review about the order

👉 A review can be left as soon as the order is sent, meaning that a customer can leave a review for an order he did not receive yet!

## **Dataset**

The dataset consists of ~100k orders from 2016 and 2018 that were made on the Olist store, available as csv files on Le Wagon S3 bucket (❗️the datasets available on Kaggle may be slightly different).

✅ Download the 9 datasets compressed in the olist.zip file, unzip it and store them in your ~/code/WLaCoutur/04-Decision-Science/01-Project-Setup/data-context-and-setup/data/csv folder:

curl https://wagon-public-datasets.s3.amazonaws.com/olist/olist.zip **>** ~/code/WLaCoutur/04-Decision-Science/01-Project-Setup/data-context-and-setup/data/csv/olist.zip

unzip -d ~/code/WLaCoutur/04-Decision-Science/01-Project-Setup/data-context-and-setup/data/csv/ ~/code/WLaCoutur/04-Decision-Science/01-Project-Setup/data-context-and-setup/data/csv/olist.zip

rm ~/code/WLaCoutur/04-Decision-Science/01-Project-Setup/data-context-and-setup/data/csv/olist.zip

Check you have the 9 datasets on your machine:

ls ~/code/WLaCoutur/04-Decision-Science/01-Project-Setup/data-context-and-setup/data/csv

## **Setup**

### **1 - Project Structure**

Go to your local ~/code/WLaCoutur/04-Decision-Science folder.  
This will be your project structure for the week.

.

├── 01-Project-Setup

│ │ *# Your whole code logic and data, this is your "package"*

│ ├── data-context-and-setup

│ │ ├── data *# your data source (git ignored)*

│ │ │ ├── csv

│ │ │ │ ├── olist\_customers\_dataset.csv

│ │ │ │ ├── olist\_orders\_dataset.csv

│ │ │ │ └── ...

│ │ │ └── README.md *# database documentation*

│ │ │

│ │ └── olist *# your data-processing logic*

│ │ ├── data.py

│ │ ├── product.py

│ │ ├── seller.py

│ │ ├── utils.py

│ │ └── \_\_init\_\_.py *# turns the olist folder into a "package"*

│ │

│ │ *# Your notebooks & analyses, day-by-day, challenge-by-challenge*

│ ├── data-data-preparation

│ └── data-exploratory-analysis

├── 02-Statistical-Inference

│ └── ...

├── 03-Linear-Regression

│ └── ...

└── 04-Logistic-Regression

└── ...

### **2 - Edit the PYTHONPATH**

Add olist path to your PYTHONPATH.

This will allow you to easily import modules defined in olist in your notebooks throughout the week.

Open your terminal and navigate to your home directory by running:

cd

Now you’ll need to open your .zshrc file. As you might have noticed the file starts with a dot which means it’s a hidden file. To be able to see this file in your terminal you’ll need to run the command below, the flag -a will allow you to see hidden files:

ls -a

Next lets open the file using your text editor:

code .zshrc

Now in your terminal run:

cd ~/code/WLaCoutur/04-Decision-Science/01-Project-Setup/data-context-and-setup **&&** echo "export PYTHONPATH=\"$(pwd):\$PYTHONPATH\""

👉 Copy the resulting output line from your terminal and paste it at the bottom of your ~/.zshrc file. Don’t forget to save and restart all your terminal windows to take this change into account.

### **🔥 Check your setup**

Go to your home folder and run an ipython session:

cd

ipython

Then type the following to check that the setup phase from the previous exercise worked:

**from** olist.data **import** Olist

**Olist**().**ping**()

*# => pong*

If you get something else than pong, raise a ticket to get some help from a TA. You might have a problem with the $PYTHONPATH.

## **Push your code on GitHub**

From your 04-Decision-Science/01-Project-Setup/data-context-and-setup directory, commit and push your code:

cd ~/code/WLaCoutur/04-Decision-Science/01-Project-Setup/data-context-and-setup

git add .

git commit -m 'kick off olist challenge'

git push origin master