# Business Case: sales prediction

## Context

As Data Scientist, you have been hired by a video game company to answer two critical questions:

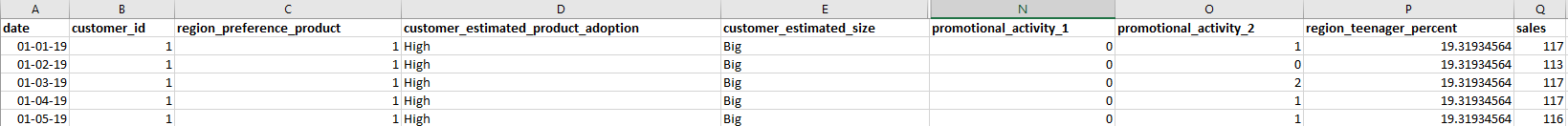
1. Is it possible to accurately forecast the sales of its video game for each of its customers (local video games stores)?
2. Given a certain marketing budget, how can we best distribute marketing activities to maximize sales?

## Dataset

To help you answer these questions, the company has provided you with a .csv file containing its video game sales for the period of January 2019 to February 2020. Other relevant information about its customers and competitors are also present in the file.

The dataset has the following columns:

* **date**: year-month of the sales
* **customer\_id**: unique identifier of the customer
* **region\_id**: unique identifier of the region in which the customer is located
* **region\_preference\_product**: Boolean (0/1) indicating whether the video game was more sold on average in this region than in the other regions; the value –1 means this information is unknown
* **customer\_estimated\_product\_adoptio**n: estimated adoption of the product by the customer
* **customer\_estimated\_size**: estimated size of the customer’s store
* **region\_preference\_competitor\_[x]**: Boolean (0/1) indicating whether the competitor’s video game [x] sold more on average in this region than in the other regions; the value –1 means this information is unknown
* **promotional\_activity\_[x**]: numeric indicating the number of promotional activities of type [x] that the company had with the customer over the month
* **region\_teenager\_percent**: the percentage of the region’s population that corresponds to teenagers
* **sales**: the total number of video games sold to the customer over the month



## Your mission

*Design, implement and evaluate an algorithm that predicts the video game’s sales.*

Here are some guidelines:

* Build the model so that it can be used to derive the optimal marketing strategy
* The ability to explain the predictions of the model is key to convince the company’s stakeholders
* Provide a high-level description of the algorithm
* Provide an implementation of the algorithm in Python

*Propose a method to explain how one can derive the optimal marketing campaign for March 2020, given a fixed budget, from the predictive model.*

Here are some guidelines:

* Provide a high-level description of the methodology (can be pseudo-code) and the techniques that you would use to answer this question

## Evaluation

Your presentation of the use case will consist of two parts. Please be prepared to give a 10-to-15-minute oral presentation.

## Deliverables

1. The code you created to tackle this challenge.
2. A document to describe your approach including how you solved the problem, why you chose that methodology, and what you could accomplish given more time.
3. 10 to 15min presentation