

B2W Challenge

B2W is the largest, most beloved and fastest growing digital company in Latin America. We have built the only digital ecosystem in Latin America that connects people, products, services and businesses. B2W's portfolio consists of four leading brands in different segments of e-commerce, the fastest growing Marketplace in Latin America and a platform of services in the technology, distribution and finance sectors.

In this highly competitive Market, we need to constantly find ways to adjust/optimize the relationship between Price and Demand to keep offering the best customer experience while also achieving the commercial targets.

We want to give you a small taste of the pricing challenge we face every day and the opportunity to show us how, with your analytical skills, you could help us on making better pricing decisions. The attached file test_prc.zip holds two files under CSV format, where the first line is the header, commas separate the columns and periods (".") separate decimal values. The files, "sales.csv" and "comp_prices.csv" once unzipped, contain 8.2M and 1.7M, respectively, with 351.092 and 50.114 lines each. Please find the structure of the archives below.

sales.csv

PROD_ID	DATE_ORDER	QTY_ORDER	REVENUE
P1	2015-07-01	1.0	23.45
P2	2015-05-23	3.0	65.39

comp_prices.csv

PROD_ID	DATE_EXTRACTION	COMPETITOR	COMPETITOR_PRICE	PAY_TYPE
P1	2015-07-01 08:10:38	C1	25.99	1
P1	2015-05-23 08:10:38	C2	27.99	1

The sales.csv file contains **transactional information** where each line represents a sale. The comp_prices.csv file contains **monitoring data of competitors' prices**. We have data available for 6 competitors, C1 to C6, which are monitored twice per day. The information below describes the data in each column:

PROD_ID: Product ID. We provide data for 9 different products, P1 to P9;

DATE_ORDER: Sales Date, under YYYY-MM-DD format;

QTY_ORDER: Quantity Sold;

REVENUE: Sale revenue. There can be variations of the price for the same product, depending on the sales channel or discounts, which are applied to the base price;

DATE_EXTRACTION: Date and Time of the extraction of the competitors' price, under YYYY-MM-DD HH:MM:SS format;

COMPETITOR: Competitors' ID (C1 to C6);

COMPETITOR_PRICE: Competitors' price per product, which can depend on the payment method;

PAY_TYPE: Payment Method (1=deferred payment, 2=immediate payment).

Deliverables:

- 1) Models for Demand Forecasting: The main objective is to create a model to predict the quantity sold for each product given a prescribed price. Along with the statistical model, we need metrics, relationships and descriptions of these data in order to understand the sales behavior. What does the data tell us? How are the different data sources related? Is there a particular competitor that seems more important?
- 2) Presentation of the results: we want to know what were the steps and your strategy (approach to the problem) during the analysis, even if these may seem wrong. The process you went through and the reasoning behind it, is as important as the solutions you found. For this, please prepare a clear and objective presentation to explain both your methodology and your results. In case you are selected for the interview, you will need to make a 20-minute (max) presentation.

Final Instructions:

Your project will be evaluated based on the methodology used in the modeling as well as the clarity of the presentation to communicate the knowledge developed. Some potential differentiating factors:

- Show an understanding of SQL;
- Use Version Control (Git for example);
- Show methods for clustering; The model should be done in R or Python, shared with us in the following way:
- Separate the material in two files: one named Scripts and the other Analysis;

- In the Scripts folder, please include the Python or R scripts you developed (the clarity of the content and of the documentation also counts);
- In the Analysis folder, please include a PDF document with your presentation.

GOOD LUCK!