

Technical Proof Data Scientist

Throughout this test several of your skills will be evaluated as a data scientist to solve a business problem. Based on the initial context and the information provided, it is expected that you apply your knowledge in descriptive, prescriptive and predictive analytics in order to present a business case.

Context

One of the company's subsidiaries is in need of profiling its commercial performance in order to generate growth estimates for the upcoming periods, so they can generate budgets and plan their roadmap. This particular subsidiary is located in a country where stability is not necessarily maintained in the industry sector, whether due to political or social issues, so considering this risk might be important for such projections.

The commercial team limits its analysis to the following information components; nevertheless, they consider there might useful information in the rest that could potentiate their analysis giving them more insights into their current situation.

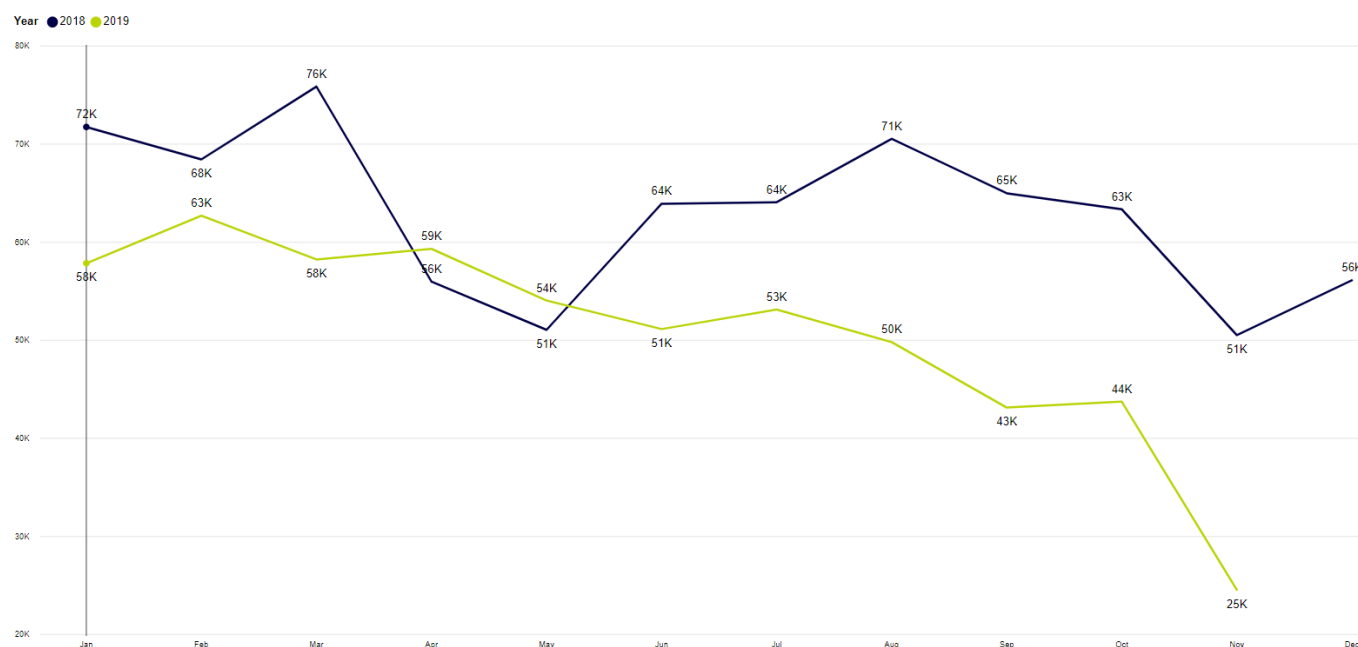
Data description

The database, in a `csv` file format, has, among its multiple features, the following:

Column name	Type	Description
No.Factura	Character	Order code
Pagador	Character	Payer code
Resp.Pago	Character	Payer name
Cod.Solic.	Character	Requester code
Solicitante	Character	Requester name
Cód.Vend	Character	Commercial advisor code
Vendedor	Character	Commercial advisor name
CondiciónPago	Character	Niveles de plazo de pago de crédito por venta
Pto.exped./depto.entradaMcía	Character	Distribution place
Descripción.1	Character	Product type
Descripción.2	Character	Product segment
Material	Character	Material code
Denominación	Character	Common name
UMB	Character	Measurement unit

Column name	Type	Description
Negocio	Character	Business segment
FechaFactCont	Date	Order date
Grupo de porte mat.	Character	Product porte
Estado	Character	State
Región	Character	Region
Condado	Character	County
Cant.Base	Digit	Quantity
NetoDólares	Digit	Incomes

The minimal database must be formed by sub-selecting the **type of product** `cement` , which generates a demand distribution similar to



Challenge

You are requested to present a business case in English where you resolve most of the business requirements below, your limit time for the final presentation will be **15** minutes and **2** hours for development. The analysis procedure and presentation, as well as the visualization, can be built on any format of your choice; although we strongly recommend you to build your analysis on `Python`, `R` or `MATLAB`. The option of combining results coming from different sources is open.

- ☐ Descriptive analysis of the data
 - ☐ Exploratory data analysis for demand and income variables
 - ☐ Time-series analysis of the demand and income variables
 - ☐ Outliers identification on the dataset, possible reasons or causes behind them
 - ☐ Reasons behind your outliers removal method
 - ☐ Impact of the dataset condition onto the construction of analytics models

- ☐ Predictive analysis based on the data
 - ☐ Sales demand forecast - 12 months
 - ☐ Reasons behind your method selection
 - ☐ Accuracy kpi

Further consider the following concerns:

- ☐ What strategy could be implemented if the data size triples in a 1 month window?
- ☐ How will this method be sustained over time? How would you refine it?
- ☐ How to build an architecture for your solution that runs automatically?