

Data Challenge

2021-01-11

Our Objective

One of our hypothetical customers produces consumer goods and asked for our help. They provided us with some datasets and our objective is to derive insights that can add value to their business. Due to confidentiality reasons we don't know their sector and products.

Datasets

We have been provided with data that is available at [The National Bureau of Economic Research](#). The files contain the following information:

1. Consumer Expenditure Surveys for 1996-2000 containing expenditures, income, wealth and basic household demographics;
2. Consumer Price Indices for 1990-2000 with prices per quarter.

There is a data dictionary file that describes the data variables, schema and meanings (`data_dictionary.pdf`).

The 2nd dataset was processed so that the variables are consistent across time.

Your Task

Use regression, classification and/or clustering methods to derive insights that can support our mission. If you want you can use more than one of these approaches.

Execute your approach in Python using a Jupiter Notebook. Send us the notebook source code and any file that was generated during this task.

In the report discuss your methodology and justify your choices for the feature selection and for the algorithms. Share the insights that you obtained from the model(s) and frame them into a data-driven story that helps our client. We are interested in your reasoning, curiosity, interpretation, imagination, algorithmic approach; storytelling-, technical-, reporting- and communication-skills.

Instructions

- Once you receive the assignment you have 72 hours to return it.
- Please do not spend more than 2 or 3h in this data challenge.
- We recommend that you start by defining the quick wins and how the reader will perceive/evaluate your report. Then plan accordingly.
- Email your code and report to the person that sent you the technical test.