Data Science

Problem 1: Predictive shopping list for Monoprix

Monoprix is a french retailer whose activity is primarily focused on cities. In order to reward its loyal customers, Monoprix started to deploy new innovative services. The goal is to provide new ways for the customer to interact with Monoprix, and help him in his daily life.

In this situation has Monoprix decided to launch a new vocal based experience, using smart speakers like Google Home. The purpose of the service is to build shopping lists by simply talking to the smart speaker. For example, you can ask « Remind me to buy egg » and the speaker will add eggs to the shopping list.

Monoprix wants this service to be intelligent, and asked us to create an algorithm that could learn from the customers habits and suggest them the products they might have forgotten to add on the list.

Therefore, Monoprix provides you with 10 years of purchase history, for the customers of their loyalty program. Also, you have at your disposal the full product catalog.

A. **First step**: Recommendation engine
   Question: How can you use Monoprix's data to build the recommendation algorithm ?

B. **Second step**: Natural Language
   Question: During the project, we realize that the product catalog is really dirty because products are wrongly named. Which solutions can you propound to correct products' names?

Problem 2: Measuring the effect of a marketing campaign

A Pharmaceutical company is trying to measure how much their last marketing campaign on a specific product has helped increase its sales and asks for your advice on how to do it. Previewing this, the Brand/Product manager already built a marketing campaign set up in which he separated one control region (specific region where he didn't roll out the campaign) from the others.

A. Assume the product has been on market in the last 2 years with a stable demand**. Explain a model you would advise the company to use and its main assumptions.**

B. Assume now that the product is new, so that the campaign was a launching one. **In this scenario is it possible to measure the effect of the campaign on sales? If yes, what model would you suggest and why?**

Problem 3: Regression Analysis

A supermarket company has a new internal policy to not discriminate **significantly** salary according to the location of their employees. They gathered the data from all of their employees and want you to verify if they are already following the new policy.

Before answering the questions below take a look at the annexed dataset: (**1stPhase-SelectiveProcess-Data Science-Data Base.csv**)

a) Question: Describe how can you use the supermarket data to verify if employees from different locations have significantly different salaries ? (Include here how you are going to treat the variables before feeding into the model)

b) Question: Implement the approach you described in python or r