Your client, European Union, wants to run a segmentation analysis for 30 countries based on the employment in different industries. The client is not sure which variables to include / exclude, and how to think about different steps in the analysis. The client needs your help in understanding its member countries and how they differ from each other along long “meaningful” dimensions.

Your responses to the following questions must be based on appropriate data analysis using Python coding. You are flexible to use coding and the steps within coding, wherever you feel it is necessary (remember: you are the one running the analysis and specific steps within it to help your client. The client does not know/cannot advice you what/how/ where to write codes and perform other types of analysis

Data: EUEmployment.csv

1. **(20 points)**
   1. Which variables would you suggest to include (and exclude) in the cluster (or segmentation) analysis? And Why? (**5 points**)
   2. Based on the variables included in the analysis from step 1a, perform a clustering analysis and identify three possible clustering solutions (e.g., solution 1: 2 clusters, solution 2: 3 clusters, solution 3: 4 clusters, etc.) (**10 points**)
   3. Which of the clustering solutions (which you identified in 1b), would you propose to the client as the final recommendation? And why? (**2 points**)
   4. Based on your proposed clustering solution in 1c, write a comparative analysis on how countries in a segment are different from those in other segments. And what implications it has for the client? (**3 points**)

Q2. Using the same dataset (EUEmployment.csv), please run a principal component analysis (PCA) and extract two and three PCA components. Interpret and provide clear explanations of the results. Which of the solution would you propose to the client? **(15 points)**

Q3. Consider a scenario of running Clustering and PCA analysis on a dataset which includes variables *weight* (in kg) with range 45-110 and *height* (in inches) with range 55 to 65. The range of *weight* is much wider than that of *height.* Does it create any challenges for Clustering and/or PCA analysis? If yes, what are those challenges and how would you overcome them? If no, why not? **(10 points)**

Q4. Suppose you are advising a school in Dehradun to update its academic and operation policy. The school wants to know more about the experience of students taking online classes. They have not done any research on this topic yet. They want to understand students' in-depth perceptions and attitudes toward online tasks and why the respondents feel that way. They also want to know how students approach their online studies on a day-to-day basis, the order in which they complete all of their activities, and what works well and does not work well.

1. Would you recommend quantitative or qualitative research? Why? **(5 points)**
2. If you believe qualitative research is a better choice, which specific methodology (e.g., focus group, 1:1 interview, etc.) would you recommend to meet the objectives mentioned above? **(7 points)**
3. For the chosen methodology in ‘b’, please write an “Introduction” (Max 300-400 words) which you need to provide to participants to help them understand the nature of your discussion with them **(3 points)**

Q5. Please read the following article and address the questions below

**How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did**

<https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/?sh=3f445efb6668>

Suppose you are a business consultant at a successful firm (not Target). You have been invited to make a presentation on YOUR VIEWS regarding fair analytical business practices, ethics in business analytics, sharing and using customer data for profit, and privacy in today’s world. The presentation is going to be at a conference in downtown Toronto in September 2022 with diverse groups of audiences attending it. When thinking about the presentation, you are conflicted between the following arguments

* Target did nothing wrong by using and analyzing customer level data to send customized messages related to promotions / coupons on products customers typically buy or most likely to buy. Since no illegal action was taken it is a fair game in business to use data to the best of your advantages.
* Target should not have done what it did by sending coupons and promotions on maternity products to customers who did not reveal the information about their pregnancy voluntarily. This is a clear case of invading customer privacy and unethical business practices. Though Target was in compliance of laws and regulations, i.e., in legal compliance, it did violate social, moral, ethical codes of conducting business.

Based on the above description and the story mentioned in the article,

1. What are YOUR VIEWS/ FINAL set of arguments (~500 words) on fair research and data collection practices, ethics in business, sharing and using customer data for profit, and privacy in today’s world? You can write a balanced set of arguments OR arguments favouring / opposing to what Target did. **(5 points)**
2. Also, would you accept a job offer from Target to work on assignments similar to what Pole did? **(5 points)**

**Article**

**How Target Figured Out a Teen Girl Was Pregnant Before Her Father Did**

Every time you go shopping, you share intimate details about your consumption patterns with retailers. And many of those retailers are studying those details to figure out what you like, what you need, and which coupons are most likely to make you happy. Target, for example, has figured out how to data-mine its way into your womb, to figure out whether you have a baby on the way long before you need to start buying diapers. Charles Duhigg outlines in the New York Times how Target tries to hook parents-to-be at that crucial moment before they turn into rampant -- and loyal -- buyers of all things pastel, plastic, and miniature. He talked to Target statistician Andrew Pole -- before Target freaked out and cut off all communications -- about the clues to a customer's impending bundle of joy. Target assigns every customer a Guest ID number, tied to their credit card, name, or email address that becomes a bucket that stores a history of everything they've bought and any demographic information Target has collected from them or bought from other sources. Using that, Pole looked at historical buying data for all the ladies who had signed up for Target baby registries in the past. From the NYT:

[Pole] ran test after test, analyzing the data, and before long some useful patterns emerged. Lotions, for example. Lots of people buy lotion, but one of Pole’s colleagues noticed that women on the baby registry were buying larger quantities of unscented lotion around the beginning of their second trimester. Another analyst noted that sometime in the first 20 weeks, pregnant women loaded up on supplements like calcium, magnesium and zinc. Many shoppers purchase soap and cotton balls, but when someone suddenly starts buying lots of scent-free soap and extra-big bags of cotton balls, in addition to hand sanitizers and washcloths, it signals they could be getting close to their delivery date. Or have a rather nasty infection...As Pole’s computers crawled through the data, he was able to identify about 25 products that, when analyzed together, allowed him to assign each shopper a “pregnancy prediction” score. More important, he could also estimate her due date to within a small window, so Target could send coupons timed to very specific stages of her pregnancy.

One Target employee I spoke to provide a hypothetical example. Take a fictional Target shopper named Jenny Ward, who is 23, lives in Atlanta and in March bought cocoa-butter lotion, a purse large enough to double as a diaper bag, zinc and magnesium supplements and a bright blue rug. There’s, say, an 87 percent chance that she’s pregnant and that her delivery date is sometime in late August. And perhaps that it's a boy based on the color of that rug? So Target started sending coupons for baby items to customers according to their pregnancy scores. Duhigg shares an anecdote -- so good that it sounds made up -- that conveys how eerily accurate the targeting is. An angry man went into a Target outside of Minneapolis, demanding to talk to a manager:

“My daughter got this in the mail!” he said. “She’s still in high school, and you’re sending her coupons for baby clothes and cribs? Are you trying to encourage her to get pregnant?” The manager didn’t have any idea what the man was talking about. He looked at the mailer. Sure enough, it was addressed to the man’s daughter and contained advertisements for maternity clothing, nursery furniture and pictures of smiling infants. The manager apologized and then called a few days later to apologize again. (Nice customer service, Target.) On the phone, though, the father was somewhat abashed. “I had a talk with my daughter,” he said. “It turns out there’s been some activities in my house I haven’t been completely aware of. She’s due in August. I owe you an apology.”

What Target discovered fairly quickly is that it creeped people out that the company knew about their pregnancies in advance. “If we send someone a catalog and say, ‘Congratulations on your first child!’ and they’ve never told us they’re pregnant, that’s going to make some people uncomfortable,” Pole told me. “We are very conservative about compliance with all privacy laws. But even if you’re following the law, you can do things where people get queasy.” Bold is mine. That's a quote for our times.

So Target got sneakier about sending the coupons. The company can create personalized booklets; instead of sending people with high pregnancy scores books o' coupons solely for diapers, rattles, strollers, and the "Go the F\*\*\* to Sleep" book, they more subtly spread them about: “Then we started mixing in all these ads for things we knew pregnant women would never buy, so the baby ads looked random. We’d put an ad for a lawn mower next to diapers. We’d put a coupon for wine glasses next to infant clothes. That way, it looked like all the products were chosen by chance. “And we found out that as long as a pregnant woman thinks she hasn’t been spied on, she’ll use the coupons. She just assumes that everyone else on her block got the same mailer for diapers and cribs. As long as we don’t spook her, it works.”

So the Target philosophy towards expecting parents is similar to the first date philosophy? Even if you've fully stalked the person on Facebook and Google beforehand, pretend like you know less than you do so as not to creep the person out. Duhigg suggests that Target's gangbusters revenue growth -- $44 billion in 2002, when Pole was hired, to $67 billion in 2010 -- is attributable to Pole's helping the retail giant corner the baby-on-board market, citing company president Gregg Steinhafel boasting to investors about the company’s “heightened focus on items and categories that appeal to specific guest segments such as mom and baby.”

Target was none too happy about Duhigg's plans to write this story. They refused to let him go to Target headquarters. When he flew out anyway, he discovered he was on a list of prohibited visitors. I think most readers of the excellent piece will find it both unsettling and unsurprising. With all the talk these days about the data grab most companies are engaged in, Target's collection and analysis seem as expected as its customers' babies. But with their analysis moving into areas as sensitive as pregnancy, and so accurately, who knows how else they might start profiling Target shoppers? The store's bulls-eye logo may now send a little shiver of fear down the closely-watched spines of some, though I can promise you that Target is not the only store doing this. Those people chilled by stores' tracking and profiling them may want to consider going the way of the common criminal -- and paying for far more of their purchases in cash.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Employment in different sectors by  Country | Group | Agriculture  (AGR) | Mining  (MIN) | Manufacturing  (MAG) | Power & water supply  (PS) | Construction  (CON) | Services  (SER) | Finance  (FIN) | Social & personal services  (SPS) | Transport & communication  (TC) |
| Belgium | EU | 2.6 | 0.2 | 20.8 | 0.8 | 6.3 | 16.9 | 8.7 | 36.9 | 6.8 |
| Denmark | EU | 5.6 | 0.1 | 20.4 | 0.7 | 6.4 | 14.5 | 9.1 | 36.3 | 7 |
| France | EU | 5.1 | 0.3 | 20.2 | 0.9 | 7.1 | 16.7 | 10.2 | 33.1 | 6.4 |
| Germany | EU | 3.2 | 0.7 | 24.8 | 1 | 9.4 | 17.2 | 9.6 | 28.4 | 5.6 |
| Greece | EU | 22.2 | 0.5 | 19.2 | 1 | 6.8 | 18.2 | 5.3 | 19.8 | 6.9 |
| Ireland | EU | 13.8 | 0.6 | 19.8 | 1.2 | 7.1 | 17.8 | 8.4 | 25.5 | 5.8 |
| Italy | EU | 8.4 | 1.1 | 21.9 | 0 | 9.1 | 21.6 | 4.6 | 28 | 5.3 |
| Luxembourg | EU | 3.3 | 0.1 | 19.6 | 0.7 | 9.9 | 21.2 | 8.7 | 29.6 | 6.8 |
| Netherlands | EU | 4.2 | 0.1 | 19.2 | 0.7 | 0.6 | 18.5 | 11.5 | 38.3 | 6.8 |
| Portugal | EU | 11.5 | 0.5 | 23.6 | 0.7 | 8.2 | 19.8 | 6.3 | 24.6 | 4.8 |
| Spain | EU | 9.9 | 0.5 | 21.1 | 0.6 | 9.5 | 20.1 | 5.9 | 26.7 | 5.8 |
| UK | EU | 2.2 | 0.7 | 21.3 | 1.2 | 7 | 20.2 | 12.4 | 28.4 | 6.5 |
| Austria | EFTA | 7.4 | 0.3 | 26.9 | 1.2 | 8.5 | 19.1 | 6.7 | 23.3 | 6.4 |
| Finland | EFTA | 8.5 | 0.2 | 19.3 | 1.2 | 6.8 | 14.6 | 8.6 | 33.2 | 7.5 |
| Iceland | EFTA | 10.5 | 0 | 18.7 | 0.9 | 10 | 14.5 | 8 | 30.7 | 6.7 |
| Norway | EFTA | 5.8 | 1.1 | 14.6 | 1.1 | 6.5 | 17.6 | 7.6 | 37.5 | 8.1 |
| Sweden | EFTA | 3.2 | 0.3 | 19 | 0.8 | 6.4 | 14.2 | 9.4 | 39.5 | 7.2 |
| Switzerland | EFTA | 5.6 | 0 | 24.7 | 0 | 9.2 | 20.5 | 10.7 | 23.1 | 6.2 |
| Albania | Eastern | 55.5 | 19.4 | 0 | 0 | 3.4 | 3.3 | 15.3 | 0 | 3 |
| Bulgaria | Eastern | 19 | 0 | 35 | 0 | 6.7 | 9.4 | 1.5 | 20.9 | 7.5 |
| Czech/Slovakia | Eastern | 12.8 | 37.3 | 0 | 0 | 8.4 | 10.2 | 1.6 | 22.9 | 6.9 |
| Hungary | Eastern | 15.3 | 28.9 | 0 | 0 | 6.4 | 13.3 | 0 | 27.3 | 8.8 |
| Poland | Eastern | 23.6 | 3.9 | 24.1 | 0.9 | 6.3 | 10.3 | 1.3 | 24.5 | 5.2 |
| Romania | Eastern | 22 | 2.6 | 37.9 | 2 | 5.8 | 6.9 | 0.6 | 15.3 | 6.8 |
| USSRF | Eastern | 18.5 | 0 | 28.8 | 0 | 10.2 | 7.9 | 0.6 | 25.6 | 8.4 |
| Yugoslavia | Eastern | 5 | 2.2 | 38.7 | 2.2 | 8.1 | 13.8 | 3.1 | 19.1 | 7.8 |
| Cyprus | Other | 13.5 | 0.3 | 19 | 0.5 | 9.1 | 23.7 | 6.7 | 21.2 | 6 |
| Gibraltar | Other | 0 | 0 | 6.8 | 2 | 16.9 | 24.5 | 10.8 | 34 | 5 |
| Malta | Other | 2.6 | 0.6 | 27.9 | 1.5 | 4.6 | 10.2 | 3.9 | 41.6 | 7.2 |
| Turkey | Other | 44.8 | 0.9 | 15.3 | 0.2 | 5.2 | 12.4 | 2.4 | 14.5 | 4.4 |