Data Scientist Task

A UK CEO of a FTSE 100 firm has recently got in touch with our team. In the wake of recent elections – e.g., in the UK, USA and France – and the rise in support for political parties on the right, she has been reading up about a growing ideological gap over time between men and women (link, or see pdf attached) that she thinks may be playing a role in these shifts.

The CEO wants to figure out how best to publicly position the company in the UK and track how this may change. She has called us to see if we can give her an evidence-based diagnosis right now to contribute to an updated strategy and set up a process to track any changes. She wants to understand whether there's **evidence of this ideological gender divide in the UK and if it's different across generations**. She's also particularly interested in anything we can tell her about **young people's attitudes and if we can challenge the idea they're a monolith** – something that would be useful for a new proposal she wants to make to the company's board and the tracking they need to do.

The CEO intends to run this project as a tracker, and as such she needs two things from you:

1. A short slide deck outlining where things stand now, looking at the gender divide and how this has changed over the generations.

2. Confirmation that you have a process in place to conduct a quick tracker survey that can track how key metrics change over time. This should be an automated function that can take new waves of data and perform the same analysis as a time series.

Data source:

- Focaldata dataset containing a large sample (n=30,000) collected in the runup to the UK's general election in July 2024 (see **Table 1** for variable descriptions).

- The dataset includes demographic variables plus responses to certain attitudinal questions. The attitude questions concern beliefs on certain *social* and *economic* issues. These questions are the same as those asked in the British Election Study over the past few decades, and are commonly used as indicators of liberal/authoritarian leaning, on the one hand, and economic leanings, on the other.

- The data was collected over a few months but the team is happy to treat it as a cross-sectional snapshot of public attitudes.

Output:

- Max. 5 page slide deck for the CEO. The CEO is time pressured, so the format should include a combination of key data viz. accompanied by concise insights.

- Code scripts used for data analysis (in python or R) that can be run and reviewed. These scripts should be functionalised so you can quickly run the next wave and output a battery of standard graphics. You should try splitting the data into smaller batches, labelling with a month and then run your functions as a proof of concept.

What we expect when we discuss your deck: We will want you to tell us your justification for any analytic approaches taken and how you built the functions that will track the data. This will include why particular modelling techniques were chosen and should include thinking about tradeoffs in rigour/complexity and pragmatism/simpleness of interpretation. This may also include gaps in the data or research that you think are important to answering the objective. Again, **this information shouldn't be in the slides**, but will be important in our review and discussion of your submission.

What we don't expect when we discuss your slides: No perfectly argued, full-blown thesis. The aim is simply to provide some robust guidance based on existing data. You can certainly propose hypotheses for what you find in the data – and can outline any guidance for future research. The tracking functions do not have to be fully commented or robust, but should be able to take a random sample of the data and run correctly.

Expected turnaround: 5 days

If you have any questions let us know!

Callum (callum@focaldata.com)

Matt (matt@focaldata.com)

Data:

A new global gender divide is emerging.pdf

uk_election_data_case_study_weighted.sav

Table 1.

Variable name	Description
Demographics	
respondent_id	Unique, anonymised ID number for each survey respondent
weight	Respondent weight to ensure national representativity (based on age, gender, region, ethnicity, education)
age	Age band
gender	Gender
education	Level of education (lower or higher than degree level)
ethnicity	Ethnicity
region	UK NUTS1 region name
cons_code	Westminster constituency code
cons_name	Westminster constituency name
household	Household tenure
religion	Religion
socgrad	Social grade classification
marital	Marital status
pv_2019	Previous vote: 2019 general election
pv_2024	Previous vote: 2024 general election. *Note: this was asked as voting intention prior to the election, but is treated here as actual 2024 vote.*
eu_ref	Vote in the 2016 EU "Brexit" Referendum
Attitudinal variables (responses on agree-disagree scale)	
econ_redist	"Government should redistribute from the better off to the less well off"
econ_bigbiz	"Big business take advantage of ordinary people"

econ_unfair	"Ordinary people do not get their fair share of the nation's wealth"
econ_onelaw	"There is one law for the rich, and another for the poor"
econ_exploit	"Management will always try to get the better of employees if it gets the chance"
soc_respect	"Young people today don't have enough respect for traditional British values"
soc_deathpen	"For some crimes, the death penalty is the most appropriate sentence"
soc_schools	"Schools should teach children to obey authority"
soc_censor	"Censorship of films and magazines is necessary to uphold moral standards"
soc_punish	"People who break the law should be given stiffer sentences"