Palantir Data Engineer Challenge

The purpose of this challenge is for you to demonstrate your analytic and creative skillset. In particular, we'll be assessing your submission on your ability to:

- **Analyse data to derive insights**: can you shape and transform data to generate answers to questions?
- **Perform self-directed hypothesis research**: where your results appear unintuitive, can you move beyond the numbers and conduct your own research to determine whether the results are valid, and reference external sources of information where appropriate?
- **Identify and account for data quality issues**: no dataset is perfect, with irregularities often hiding below the surface. Can you identify these anomalies, and ensure they don't affect the correctness of your analysis?
- **Communicate results clearly and at an appropriate level of detail**: is your narrative convincing enough that a non-expert believes your analysis to be reliable and correct?
- **Demonstrate technical expertise with data management tooling**: are you comfortable operating in a technical environment?

You can use any tool of your choice to complete the challenge. Along with your solutions, you should include evidence of the methodology you've used to solve the problems (e.g. Excel spreadsheets, Python notebooks, or other code files).

You can complete the challenge in your own time, although we ask that you send us your results within a week of receiving the challenge, or let us know if you need more time. When complete, submit your solutions and supporting material to your Palantir recruiting POC.

We're just as interested in gaining insight into your problem-solving approach as whether your numerical answers are correct. Please ensure that the supporting material you submit with your solutions is presented in a clear manner, and feel free to include any additional commentary that you feel is relevant for us to better understand your problem-solving approach.

Finally, please submit your own original work, and do not share your answers or the questions with others. Once you complete the challenge we'll review your submission and follow up with you to discuss next steps.

Questions

All questions that follow refer to population statistics for the United Kingdom. Your Palantir recruiting POC will have supplied this raw population data to you along with these questions. You should use this data as your primary source to answer all questions that follow.

1. Inspect the structure and content of the dataset referenced above. Are there any oddities that could cause someone problems when using the data to answer questions? How would you deal with them?
2. Which geography contained the smallest total population in each year from 2013 to 2016?
3. In this question, the female-to-male ratio is defined as [# of all females in geography] divided by [# of all males in geography], aggregated across all age groups. For example, in 2016, Fylde had a total of 39,838 females and 38,152 males, giving a female-to-male ratio of ~1.044.
    - Which geography had the highest female-to-male ratio in 2013, and what was the ratio?
    - Comparing each geography's female-to-male ratio measurements in 2013 vs 2016, which geography's ratio changed the most? Which changed the least?
4. This question relates to the age distribution of the total UK population.
    - Plot a visualisation of the 2016 age distribution, split out by sex. What anomalies do you see? Describe possible factors that could account for each observed anomaly.
    - The UK's population is getting older over time, but according to this dataset that is not the case everywhere across the UK. Investigate which geographies have a particularly high or low proportion of over-65s, and which show the largest change in proportion of over-65s between 2013 and 2016. Are your numerical results realistic? What factors could explain the outliers?