Pro Services Data Scientist - Technical test

Intro

For this task, we would like you to train a model (or more than 1 if you deem it appropriate) to describe a team's playing style, based on StatsBomb's open data.
You may approach this as a regression, clustering, classification, or dimensionality reduction task, depending on which you think is better suited to the task.

You can also use whatever statistical or machine leaning methodology you think is appropriate, and you should present your results in whatever way you feel is digestible.

The output of your model can be one or more numeric fields, classification/cluster labels, or similar, but please provide a description of what these fields mean, if appropriate, to facilitate their interpretation.

Deliverables

Please submit any code you write for this task along with the output from your models, along with answers to the questions below to `dinesh.vatvani@statsbomb.com`

The code can be submitted in any format you prefer (Notebook, library, Script, etc.)

Questions:

1. What other approaches could you have chosen to tackle this task and why did you settle on the approach you chose?
2. What are the flaws and limitations of your approach?
3. What insights or metrics could you derive based on the outputs of your model?
4. Given infinite time, how might you improve on your approach?
5. What changes would you make to your implementation before you considered it production-ready?

We will judge candidates' responses on:

- The robustness and validity of their approach from a statistical standpoint.
- The quality their code in terms of maintainability and clarity.

Please spend no more than 10 hours on this task. We also expect you to run out of time before you're completely happy, so don't panic! We care much more about seeing how you think about and work on a real problem, and how you write code for other people's consumption than we do about you necessarily finishing the task and delivering useful results.

Resources:

You may find the following resources helpful:

StatsBomb open data: A respository with StatsBomb open data
StatsBombPy: A Library to facilitate fetching StatsBomb data