

# DATA CHALLENGE QUESTIONS 2024

# ACTIVITY IN THE NORTH EAST

## Question 1 – Revenue of Sport & Leisure Industry in the North East

The North East of England has a thriving sports community; home to the Great North Run, Newcastle UFC, Newcastle UWFC, Sunderland AFC, Sunderland AFC Women, Durham County Cricket Club and many more. The impact on the culture of the North East is undeniable, and we would like to explore the contribution that leisure activities provides to the North East's economy.

We would like you to:

- a) Estimate the revenue that sporting events and leisure activities generate for the North East.**
  - We have listed some sources below to get you started, but you may use any other source that is relevant.
- b) Forecast the revenue generated by 2030 based on current trends.**
- c) Present a short proposal on how to increase the revenue generated by sporting events and leisure activities. Support your proposal with evidence.**

Please clearly highlight the sources you have used and explain any assumptions made.

Relevant sources:

[UK statistics on the economic impact of sporting events](#)

---

[Sport England on the economic impact of sporting events](#)

---

[Newcastle council spending](#)

---

[Sunderland A.F.C accounts](#)

---

# ACTIVITY IN THE NORTH EAST

Data in this challenge is illustrative

## Question 2 – Analysis of half-marathon times

Each year in September, over 50,000 people take part in the Great North Run, travelling from all over the world to compete for a first-place finish or set a new personal best. With such intense competition comes the need to train and to ensure that the training is as efficient and effective as possible.

We have a dataset from a previous North East Half Marathon – Geordie Stride Half Marathon, which contains race times of participants as well as other information.

We need your help to solve the following problems:

**a) There are 14 variables available in this dataset. For each variable:**

- Carry out appropriate data quality checks and clean each variable
- Given how each variable is distributed, please provide some key statistics that you can use you summarise each variable
- Provide hypotheses for the distribution of each variable and the data cleaning steps you carried out

**b) Analyse the effect that each variable has on final race time:**

- As part of your analysis, please investigate the combined impacts of some variables
- Please justify the reasoning behind the variables you chose to analyse together
- Give a recommendation on ways in which runners can reduce their run time for the Great North Run. Remember to back up your proposal with evidence!

**c) Build a linear regression model to predict the race time of a participant X given the data provided. You may also use other types of models if you prefer.**

- Use the previous question to inform which variables to include in your model
- Provide an explanation of your model and the parameters it has
- Some variables you may want to consider are gender, training frequency and club membership.
- Use your model to predict this runner's finish time:

gender	trained_10_week	trained_im	has_trainer	cadence	age	bmi	n_marathons_run	bib_colour	VO2_max	heart_rate	shoe_size
female	2	4	1	155	17	19	0	yellow	12.75	6.31	4

**d) Assess how well your model performs.**

**Your submission should be clearly presented - a neat Excel or code base with a PowerPoint to explain your results and final model.**

# ACTIVITY IN THE NORTH EAST

Figures in this challenge are illustrative

## Question 3 – Leisure centres in the North East

You have been approached by our client, Active Life, who are looking to open a series of leisure centres in the North East. They need help deciding what sporting activities to offer at their gyms. In addition to the cost of hosting different activities, they are also interested in finding out which sports are the most popular in the North East.

Please help Active Life decide which of the following services to offer.

### 1. Please research:

- Which leisure activities are the most popular.
- How much Active Life could charge for each activity.

### 2. Given an initial yearly budget of £100,000, and land area of 120m x 70m, calculate the estimated profit that Active Life could make in the first year of business, and the activity plan that would maximise Active Life's profit.

- Use the below table as a cost summary for different activities Active Life could offer.
- Describe how Active Life could best optimise their space.
- Explain the potential effect of seasonality on the activities chosen, and how Active Life could best utilise this information.
- Explain any assumptions you have made throughout.

Service	Cost	Area	Capacity
Running coach	£20/hour salary		10 people / hour
Squash courts (%)	£5,000/year/court	10m x 6m per court	2 people / court
Driving range	£65,000 per year	100m x 25m	10 bays
Football pitch	£2,000 per month	90m x 40m	30 people / hour
Swimming pool	£20,000 per year	25m x 6m	20 people / hour