ARCTIC SHORES
Data-Driven People Insights

Python/Pandas Exercise

Introduction

This programming exercise is designed to take 60-120 minutes. The aim is to demonstrate programming approach as a priority over finishing it!

Please link to a git repository containing your work and make commits as you progress through this exercise.

Task Overview

You will create a small Python 3 app designed to process a small sample of input data. You should use the following language version and packages for this project:

- Python 3.10
- Pandas
- pytest

We encourage you to additionally use any tooling, code formatters or commit hooks that will assist you in creating the best possible submission.

Background Context

We are releasing a new psychometric level – “CourageCards”. In this assessment, the following happens:

- A deck of red and green cards is shuffled.
- A candidate picks the top card.
- If it is green they get one point.
- They can then choose to pick another card or bank their points.
- If they bank their points, they keep their points and the deck is shuffled.
- If they pick another green card, they get another point.
- If they pick another red card, they lose all their unbanked points and the deck is shuffled.
- They will do this a maximum of 45 times.

Task

We have the first set of data from our new level, in the included test_level_data.json. The `event` column refers to the action taken by the candidate, and can contain the following strings:

`start` - The level start time
`shuffle_cards` - The deck is shuffled.
`green_card` - The candidate selected a green card and gets one point
`red_card` - The candidate selected a red card and lost their unbanked points
`banked` - The candidate opted to bank their points, preventing their loss in the instance of a red card being selected.
`end` - The level end time

So a “round of play” would follow the following sequence:

- shuffle_cards
- green_card * n
- red_card or banked

You will create a CLI tool which takes test_level_data.json as its input, and returns a csv. The CSV should contain the following information:

- The total time spent on the level
- The mean number of green cards selected across all rounds of play
- The total points the candidate has received