Mind AI
https://about.mind.ai/careers

Cinema Screening Rules Challenge

OVERVIEW

It's the year 2040. There is a cinema located in Myeongdong, Seoul, which has been operating for 15 years and is sold out every day. It opens 5 days a week and has two screening rooms. The first one shows 3 movies a day, and the second one shows 2 screenings a day. It only shows movies from the year 2024 and earlier, yet the cinema is sold out every day. Not much is known about how the cinema decides which movie to screen on which day in which screening room, only that the screening schedule is unique every week. Your task is to use your skills in machine learning and data mining to find out the rules the operators use to decide the screening schedule.

WHAT IS KNOWN

The pool of movies is known and specified in the file `movies.csv`.

The weekly screening schedules from the past 15 years are also known and documented in the file `past_screenings.csv`.

Two rules are already known:

1. Movies of the same day and room must be sorted by year in ascending order.
2. Movies of the same name must not be shown more than 1 time every 5 days.

All rules that should be discovered follow the same pattern as the second rule:

Movies `<movie_filter?>` `<characteristics?>` `<limit_function>` `<limit>` time(s) every `<frequency>` day(s).

movie_filter

A `<movie_filter>` is optional in the rule and can be:

1. A **filter expression**: An atomic condition such as `genre=Drama`, `director=Ingmar Bergman`, or `year=1968`.
2. A **boolean expression**: An expression that combines one or more filter expressions using boolean operators (`and`, `or`, `not`). Boolean expressions can be nested and combined to form more complex conditions.

All column names can be used as keys for the filter expression, while the values of the filter expressions should be one of the values of the column.

Boolean Operators

- **and**: Combines two boolean expressions or filter expressions, and evaluates to `true` if both expressions are `true`.
- **or**: Combines two boolean expressions or filter expressions, and evaluates to `true` if at least one of the expressions is `true`.
- **not**: Applies to a single boolean or filter expression, and inverts its value (`true` becomes `false`, and vice versa).

Examples

1. Movies of the same name must not be shown more than 1 time every 5 days.
2. Movies `where (genre=Drama) and (director=Ingmar Bergman)` of the same name must not be shown more than 1 time every 5 days.
3. Movies `where not(genre=Drama)` of the same name must not be shown more than 1 time every 5 days.
4. Movies `where (((genre=Drama) and (director=Ingmar Bergman)) or (year=1968))` of the same name must not be shown more than 1 time every 5 days.
5. Movies `where not(((genre=Drama) or (director=Ingmar Bergman)))` of the same name must not be shown more than 1 time every 5 days.

characteristics

Rules can optionally be applied to clusters of movies with shared characteristics. These shared characteristics are defined by the column names in the movie dataset. If a cluster is defined by a single characteristic, it is written as `of the same characteristic_a.` If a cluster is defined by multiple characteristics, they are written as `of the same characteristic_a and characteristic_b`.

Examples

1. Movies must not be shown more than 1 time every 5 days.

2. Movies `of the same genre` must not be shown more than 1 time every 5 days.
3. Movies `of the same genre and director` must not be shown more than 1 time every 5 days.
4. Movies `of the same genre and director and year` must not be shown more than 1 time every 5 days.

limit_function

The limit function is either `must not be shown more than` **or** `must be shown at least`.

limit and frequency

Both limit and frequency are integer values.

WHAT IS UNKNOWN

It is unknown which rules or how many rules the cinema applies to schedule the screenings; it could be 2 rules, 200 rules, or anything in between. The complexity of these rules is also unknown.

EXPECTED SOLUTION AND FORMAT

The expected result is a set of sentences that follow the above described pattern:

Movies `<movie_filter?>` `<characteristics?>` `<limit_function>` `<limit>` time(s)? every `<frequency>` day(s)?.

Example of expected solution

- Movies `where (genre=Drama) and (director=Ingmar Bergman)` must not be shown more than 1 time every 2 days.
- Movies `of the same genre and director and year` must not be shown more than 2 times every 1 day.

EVALUATION CRITERIA

Your solution will be assessed based on its F1 Score.