Задача 3: джоин датафреймов

Есть два файла:

1. Референс файл RATIO9.zip в архиве - его нужно разархивировать.
   Headers для RATIO9 файла имеет следующие поля:
   - index - уникальный индекс;
   - start_date
   - end_date
   - ratio_key - составной ключ для мапинга на конкретный Facility за определенный период;
     Пример: 1998F101061430 где 1998 - год выписки (discharge_date), F - флаг ID: facility_id или medicare_id (значения F или M), 101061430 - значение facility_id или medicare_id в зависимости от флага;
   - ratio_1 - ratio_37 - множитель для вычисления cost (Общие затраты на пациента);
   - dratio_1 - dratio_37 - Direct Ratio - множитель для вычисления Direct Cost (прямые затраты на пациента);
   - valid_flag - значение ниже или равное 3 указывает на невалидность записи (не учитываем данные записи).
2. Инпут файл hospitlEncounter.csv
   Для данной задачи потребуются следующие поля:
   - record_identifier - уникальный идентификатор записи;
   - facility_id - уникальный идентификатор для мапинга;
   - fac_medicare_id - уникальный идентификатор для мапинга;
   - discharge_date - год выписки;
   - patient_id - уникальный идентификатор пациента.

Задача:

- вычитать оба файла в датафрейм
- организовать джойн этих двух файлов по следующим правилам:
  - (HE) FacilityId = (Ratio9) Facility Id и
  - (HE) discharge date between (Ratio9) start and end date
  - if no matches: (HE) discharge date - 1 year between (Ratio9) start and end date
  - if no matches: (HE) discharge date - 2 year between (Ratio9) start and end date
  - if no matches: (HE) discharge date - 3 year between (Ratio9) start and end date
- если до сих пор не совпадает:
  - (HE) MedicareId = (Ratio9) MedicareId и
  - (HE) discharge date between (Ratio9) start and end date
  - if no matches: (HE) discharge date - 1 year between (Ratio9) start and end date
  - if no matches: (HE) discharge date - 2 year between (Ratio9) start and end date
  - if no matches: (HE) discharge date - 3 year between (Ratio9) start and end date

В итоге должны получить ровно столько записей, сколько было в hospitlEncounter-файле + наилучшее совпадение из референс файла для каждой записи. Результирующий датафрейм должен иметь следующий вид:

```
+-----------------+-----------+---------------+-------------+----------+-
-----+-------------+----------+----------+----------+--------------------+
|record_identifier|facility_id|fac_medicare_id|discharge_date|patient_id|
index|     ratio_key|start_date|  end_date|valid_flag|              ratios|
+-----------------+-----------+---------------+-------------+----------+-
-----+-------------+----------+----------+----------+--------------------+
|              !@#|  110010030|           null|    2010-03-02|
1|211095|2010F110010030|2009-07-01|2010-06-30|         5|[[0.699936,
0.430...|
|              001|  110010030|           null|    2010-03-02|
1|211095|2010F110010030|2009-07-01|2010-06-30|         5|[[0.699936,
0.430...|
|              002|  101013510|           null|    2012-01-29|
158|246874|2012F101013510|2011-10-01|2012-09-30|         5|[[1.439838,
0.875...|
|              003|  101010330|           null|    2012-03-14|
159|246791|2012F101010330|2012-01-01|2012-12-31|         5|[[0.372632,
0.221...|
|              004|  101082590|           null|    2012-01-04|
160|247522|2012F101082590|2011-07-01|2012-06-30|         5|[[0.69322,
0.4319...|
+-----------------+-----------+---------------+-------------+----------+-
-----+-------------+----------+----------+----------+--------------------+

root
 |-- record_identifier: string (nullable = true)
 |-- facility_id: string (nullable = true)
 |-- fac_medicare_id: string (nullable = true)
 |-- discharge_date: date (nullable = true)
 |-- patient_id: string (nullable = true)
 |-- index: integer (nullable = true)
 |-- ratio_key: string (nullable = true)
 |-- start_date: date (nullable = true)
 |-- end_date: date (nullable = true)
 |-- valid_flag: integer (nullable = true)
 |-- ratios: array (nullable = true)
 |    |-- element: struct (containsNull = true)
 |    |    |-- ratio: double (nullable = false)
 |    |    |-- dratio: double (nullable = false)
```