Bajaj Finserv Health Data Engineering Programming Challenge (Qualifier 2)

(21st May 2023)

Introduction

This is a two-step challenge and those who complete both the steps will be consider for **next round**.

You will have a **Json file** to prepare the dataset which can be downloaded from the link mentioned below in **Step 1**.

This challenge consists of 2 steps. You need to do a few tasks over the dataset. Task are mentioned below:

1. JSON parsing with nested columns
2. Data transformation and aggregations
3. Validations
4. Graph Plot
5. Insights

SUBMIT THE COMPLETE CODE IN A GIT REPO AND DEATILS TO BE UPDATED IN BELOW MICROSOFT FORM

Form Link: https://forms.office.com/r/2WBTv8FnLv

Challenge

DOWNLOAD THE JSON FILE FORM HERE LINK

STEP 1

a.

Read the json file and select the columns mentioned below :

1. appointmentId
2. phoneNumber
3. firstName (from patientDetails)
4. lastName (from patientDetails)
5. gender (from patientDetails)*
6. birthDate (from patientDetails)**
7. medicines (from consultationData)

* Transform gender column data as mentioned below

'M' to male

'F' to female

null/Anything else to others

** Rename birthDate column as DOB

b.

Create a derived column fullName from fisrtName and lastName separated by a " " (space).

* if firstName is 'abc' and lastName is 'xyz' then fullName must be 'abc xyz'

c.

Add a column isValidMobile of boolean which contains the values true or false against each value from phoneNumber column If a phone number is a valid Indian phone number.

true : indicates number is valid

false : indicates number is not valid

* Design a logic to check whether a phone number is valid or not

* General rules to check -

1. A valid number can have '+91' or '91' as prefix, a valid number must lie be between 6000000000-9999999999

eg : +919876787687 is valid

919877475896 is valid

+913454768688 is invalid

9876787687 is valid

3454768688 is invalid

57769666 is invalid

d.

Add a column phoneNumberHash which contains hash against only valid number and null in case of invalid from phoneNumber column.

* Use SHA256 hashing algorithm to hash the mobile number

* Number +919876776576 and 9876776576 should have same hash

e.

Add a column Age which will contain the age in integer of the person, use DOB column to calulate the age.

* if DOB is 2000-01-01T00:00:00.000Z then Age should be 23

* if DOB is 2000-06-01T00:00:00.000Z then Age should be 22

* if DOB is null then age should be null

f.

Add below aggrigated columns against unique appointmentId---

Add a column noOfMedicines which will contain the number of medicines prescribed against each appointmentId.

Add a column noOfActiveMedicines which will contain the number of active medicines prescribed against each appointmentId.

Add a column noOfInActiveMedicines which will contain the number of inactive medicines prescribed against each appointmentId.

* If the value of IsActive Column is true , that means medicine is active

* If the value of IsActive Column is false , that means medicine is inactive

g.

Add a column medicineNames which should contain the name of all active medicines separated by a character "," (comma) against each appointmentId.

STEP 2

Your final dataframe should have these columns, Export this dataframe in a csv file with no index and use'~' separator:-

1. appointmentId
2. fullName
3. phoneNumber
4. isValidMobile
5. phoneNumberHash
6. gender

7. DOB

8. Age

9. noOfMedicines

10. noOfActiveMedicines

11. noOfInActiveMedicines

12. MedicineNames

h.

1. Export the aggregated data in json file format

Keys:

1. Age

2. gender

3. validPhoneNumbers (no of valid mobile numbers)

4. appointments (no of appointments)

5. medicines (no of medicines prescribed)

6. activeMedicines (no of active medicines prescribed)

2. Plot a pie chart for number of appointments against gender

* TAKE THE SCREENSHOT OF CHART AND ATTACH IT IN REPO.

Evaluation Criteria

Evaluation will be based on both Python code . BAJAJ Health team will decide cut-offs based on the successful entries received. Code quality, standards, error handling etc. will be considered during evaluation.

DO NOT FORWARD THIS DOCUMENT