

Data Qualities Issues:

Script for data loading from json to tabular form.

```
/*-----
Receipts json data dump into tabular format using below SQL SERVER script.
-----*/
DECLARE @JSON3 VARCHAR(MAX)
SELECT @JSON3=BulkColumn
FROM OPENROWSET (BULK 'C:\Users\priya\Downloads\Fetch-Rewards\receipts.json', SINGLE_CLOB) AS i

DROP TABLE IF EXISTS #Receipts
SELECT DISTINCT
    JSON_Value (c.value, '$._id.$oid') AS id
    ,JSON_Value (c.value, '$.bonusPointsEarned') AS bonusPointsEarned
    ,JSON_Value (c.value, '$.bonusPointsEarnedReason') AS bonusPointsEarnedReason
    ,CONVERT(INT,JSON_Value (c.value, '$.purchasedItemCount')) AS purchasedItemCount
    ,DATEADD(S, CONVERT(INT,LEFT(JSON_Value (c.value, '$.createDate.$date"', 10))), '1970-01-01') AS createDate
    ,DATEADD(S, CONVERT(INT,LEFT(JSON_Value (c.value, '$.dateScanned.$date"', 10))), '1970-01-01') AS dateScanned
    ,DATEADD(S, CONVERT(INT,LEFT(JSON_Value (c.value, '$.finishedDate.$date"', 10))), '1970-01-01') AS finishedDate
    ,DATEADD(S, CONVERT(INT,LEFT(JSON_Value (c.value, '$.modifyDate.$date"', 10))), '1970-01-01') AS modifyDate
    ,DATEADD(S, CONVERT(INT,LEFT(JSON_Value (c.value, '$.pointsAwardedDate.$date"', 10))), '1970-01-01') AS pointsAwardedDate
    ,DATEADD(S, CONVERT(INT,LEFT(JSON_Value (c.value, '$.purchaseDate.$date"', 10))), '1970-01-01') AS purchaseDate
    ,JSON_Value (c.value, '$.pointsEarned') AS pointsEarned
    ,JSON_Value (p.value, '$.barcode') AS r_barcode
    ,JSON_Value (p.value, '$.description') AS r_description
    ,JSON_Value (p.value, '$.finalPrice') AS r_finalPrice
    ,JSON_Value (p.value, '$.itemPrice') AS r_itemPrice
    ,JSON_Value (p.value, '$.needsFetchReview') AS r_needsFetchReview
    ,JSON_Value (p.value, '$.needsFetchReviewReason') AS r_needsFetchReviewReason
    ,CONVERT(INT,JSON_Value (p.value, '$.partnerItemId')) AS r_partnerItemId
    ,JSON_Value (p.value, '$.pointsNotAwardedReason') AS r_pointsNotAwardedReason
    ,JSON_Value (p.value, '$.pointsPayerId') AS r_pointsPayerId
    ,JSON_Value (p.value, '$.preventTargetGapPoints') AS r_preventTargetGapPoints
    ,JSON_Value (p.value, '$.quantityPurchased') AS r_quantityPurchased
    ,JSON_Value (p.value, '$.rewardsGroup') AS r_rewardsGroup
    ,JSON_Value (p.value, '$.rewardsGroup') AS r_rewardsGroup
    ,JSON_Value (p.value, '$.rewardsProductPartnerId') AS r_rewardsProductPartnerId
    ,JSON_Value (p.value, '$.userFlaggedBarcode') AS r_userFlaggedBarcode
    ,JSON_Value (p.value, '$.userFlaggedDescription') AS r_userFlaggedDescription
    ,JSON_Value (p.value, '$.userFlaggedNewItem') AS r_userFlaggedNewItem
    ,JSON_Value (p.value, '$.userFlaggedPrice') AS r_userFlaggedPrice
    ,JSON_Value (p.value, '$.userFlaggedQuantity') AS r_userFlaggedQuantity
    ,JSON_Value (p.value, '$.pointsEarned') AS r_pointsEarned
    ,JSON_Value (p.value, '$.brandCode') AS brandCode
    ,JSON_Value (c.value, '$.rewardsReceiptStatus') AS rewardsReceiptStatus
    ,CONVERT(MONEY,JSON_Value (c.value, '$.totalSpent')) AS totalSpent
    ,JSON_Value (c.value, '$.userId') AS userId
INTO #Receipts

FROM OPENJSON (@JSON3, '$.Receipts') as c
CROSS APPLY OPENJSON (c.value, '$.rewardsReceiptItemList') as p
```

- Fields like finalPrice have data type NVARCHAR(String value) and I believe this field should be of type Money/Currency.

After executing above script run below script and you will see the below output.

EXEC tempdb.dbo.sp_help @objname = N'#Receipts';

3 %

Results Messages

1 #Receipts dbo user table 2021-09-01 18:58:33.880

	Column_name	Type	Computed	Length	Prec	Scale	Nullable	TrimTrailingBlanks	FixedLenNullInSource	Collation
5	createDate	datetime	no	8			yes	(n/a)	(n/a)	NULL
6	dateScanned	datetime	no	8			yes	(n/a)	(n/a)	NULL
7	finishedDate	datetime	no	8			yes	(n/a)	(n/a)	NULL
8	modifyDate	datetime	no	8			yes	(n/a)	(n/a)	NULL
9	pointsAwardedDate	datetime	no	8			yes	(n/a)	(n/a)	NULL
10	purchaseDate	datetime	no	8			yes	(n/a)	(n/a)	NULL
11	pointsEarned	nvarchar	no	8000			yes	(n/a)	(n/a)	SQL_Latin1_General_CP1_CI_AS
12	r_barcode	nvarchar	no	8000			yes	(n/a)	(n/a)	SQL_Latin1_General_CP1_CI_AS

- We have two fields in Receipts data source purchasedItemCount (total number of items purchased in one receipt) and quantityPurchased (individual quantity of item purchased in same receipt). Based on my analysis, the count should match for both these fields in a way that purchasedItemCount equals to sum of all quantityPurchased field for the same receipt. i.e., purchasedItemCount=SUM(quantityPurchased). But there are cases where counts of these fields are not matching.

Example: The count for one of the receipt purchasedItemCount is 2 and sum of quantityPurchased is 4.

SELECT purchasedItemCount,r_quantityPurchased,* FROM #Receipts WHERE id='5ff1e1d20a7214ada1000561'

/*If you run above select query will give you records around 6,941

3 %

Results Messages

	purchasedItemCount	r_quantityPurchased	id	bonusPointsEarned	bonusPointsEarnedReason	purchasedItemCount	createDate	dateScanned
1	2	3	5ff1e1d20a7214ada1000561	5	All-receipts receipt bonus	2	2021-01-03 15:25:06.000	2021-01-03 1
2	2	1	5ff1e1d20a7214ada1000561	5	All-receipts receipt bonus	2	2021-01-03 15:25:06.000	2021-01-03 1