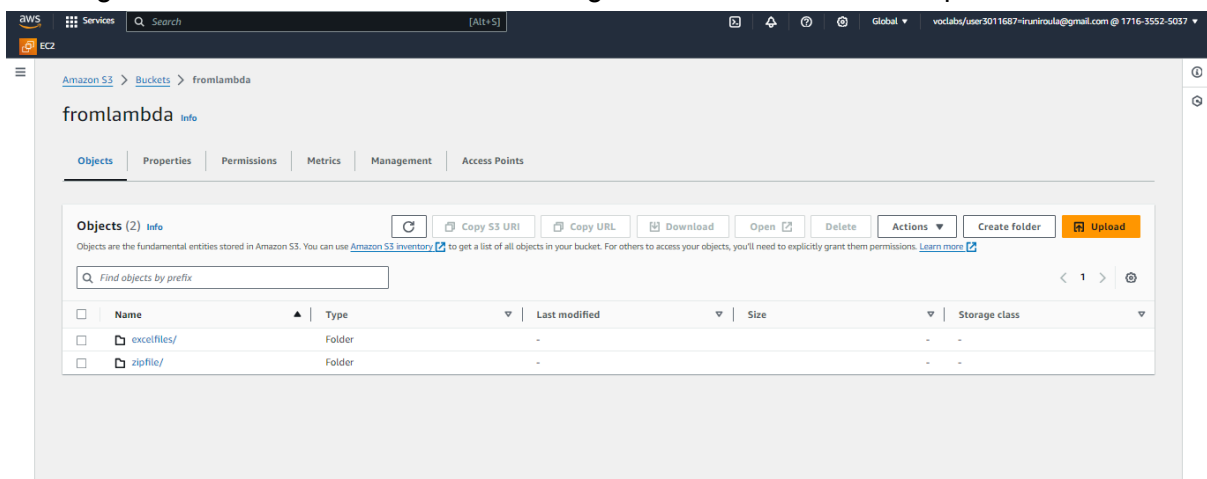


Task 2 : Python and Lambda

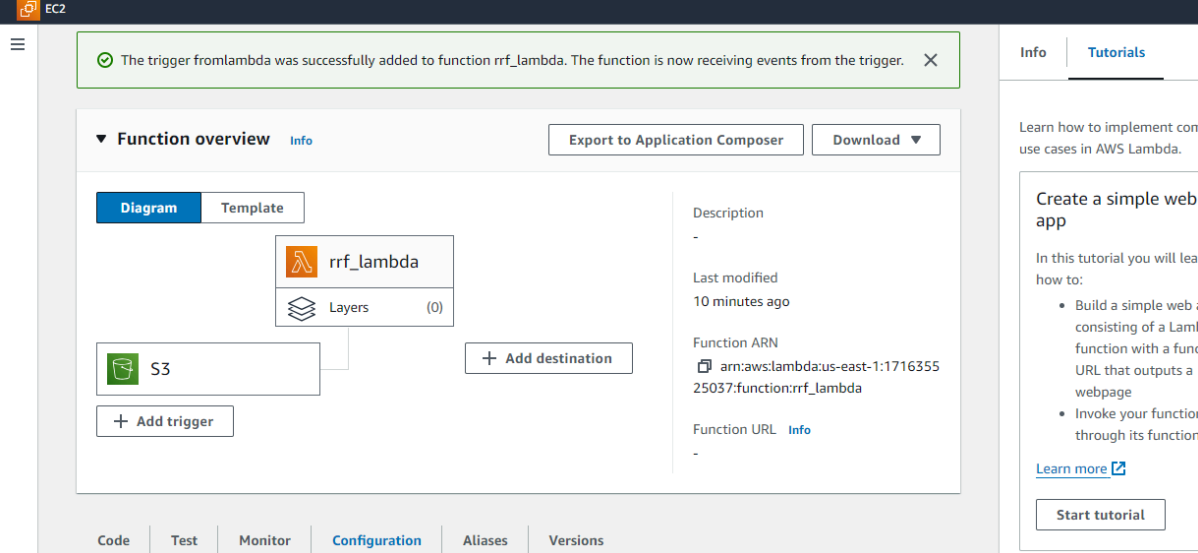
Work with `RXNORM` file,

1. Scrap the latest `RXNORM` file from NLM webpage
2. Download the latest `RXNORM` file with `api_key`
3. Create a log file for the downloaded file
4. Add header into each `fff` from `RXNORM.xlsx`
5. Add `CODE_SET` & `VERSION_MONTH` column with default values `RxNorm` and version month from downloaded filename
6. Convert dates into `YYYY-MM-DD`
7. Save files as `txt` delimited by comma(,)
8. Validate `row_count` between original and converted files

Creating S3 bucket “fromlambda” and creating folders “excelfiles” and “zipfile”



Creating Lambda function and adding S3 bucket as trigger

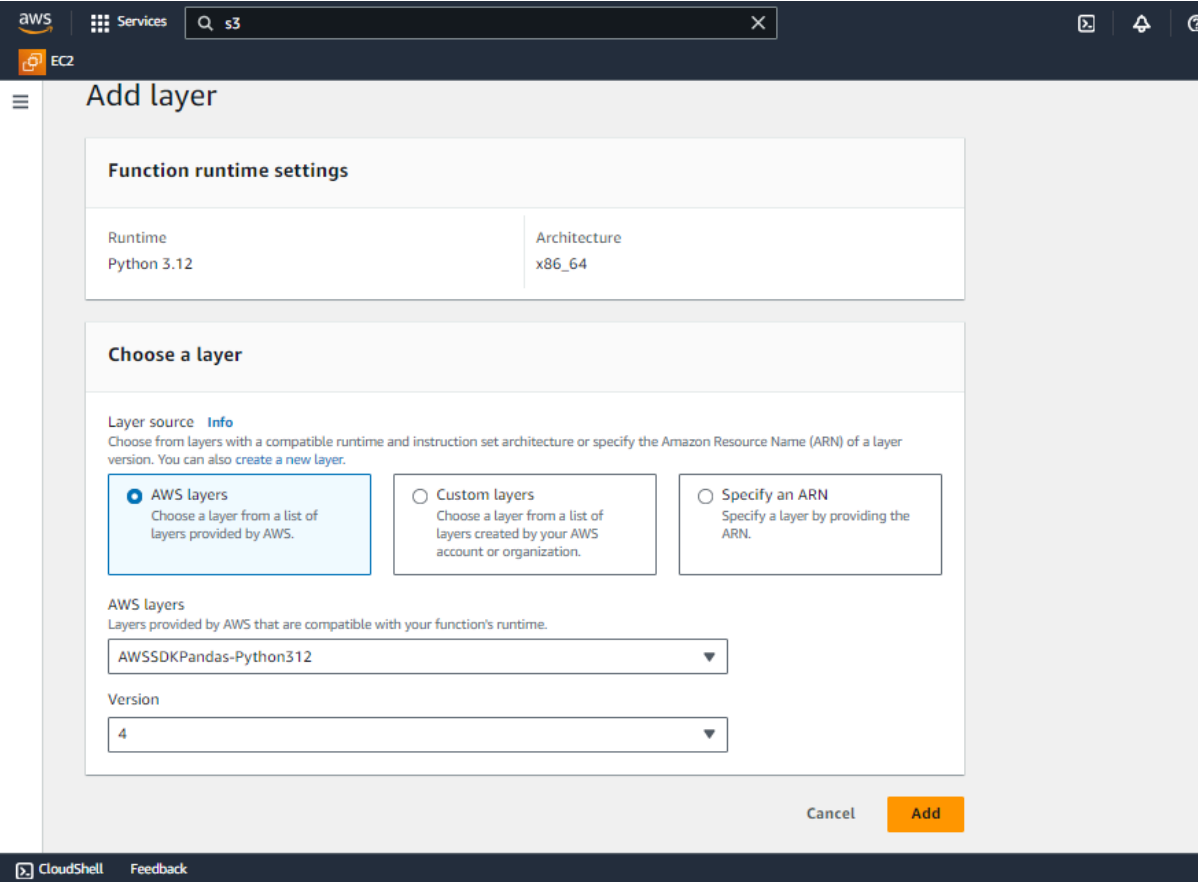


The screenshot shows the AWS Lambda console interface. At the top, a green notification bar states: "The trigger from lambda was successfully added to function rrf_lambda. The function is now receiving events from the trigger." Below this, the "Function overview" section is displayed. It includes a "Diagram" tab showing the function "rrf_lambda" with an S3 bucket as a trigger. The "Layers" section shows "(0)" layers. The "Function details" section on the right provides the following information:

- Description: -
- Last modified: 10 minutes ago
- Function ARN: `arn:aws:lambda:us-east-1:171635525037:function:rrf_lambda`
- Function URL: [Info](#)

At the bottom of the console, there are tabs for "Code", "Test", "Monitor", "Configuration", "Aliases", and "Versions". On the right side, there is a "Tutorials" section with the heading "Create a simple web app" and a "Start tutorial" button.

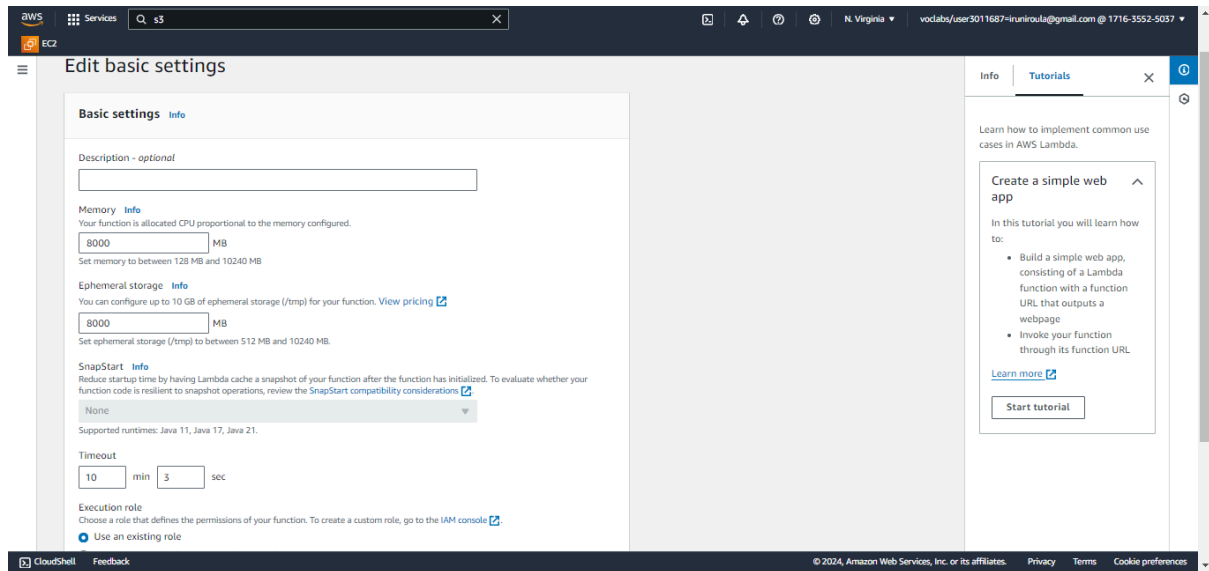
Adding pandas layer to lambda function



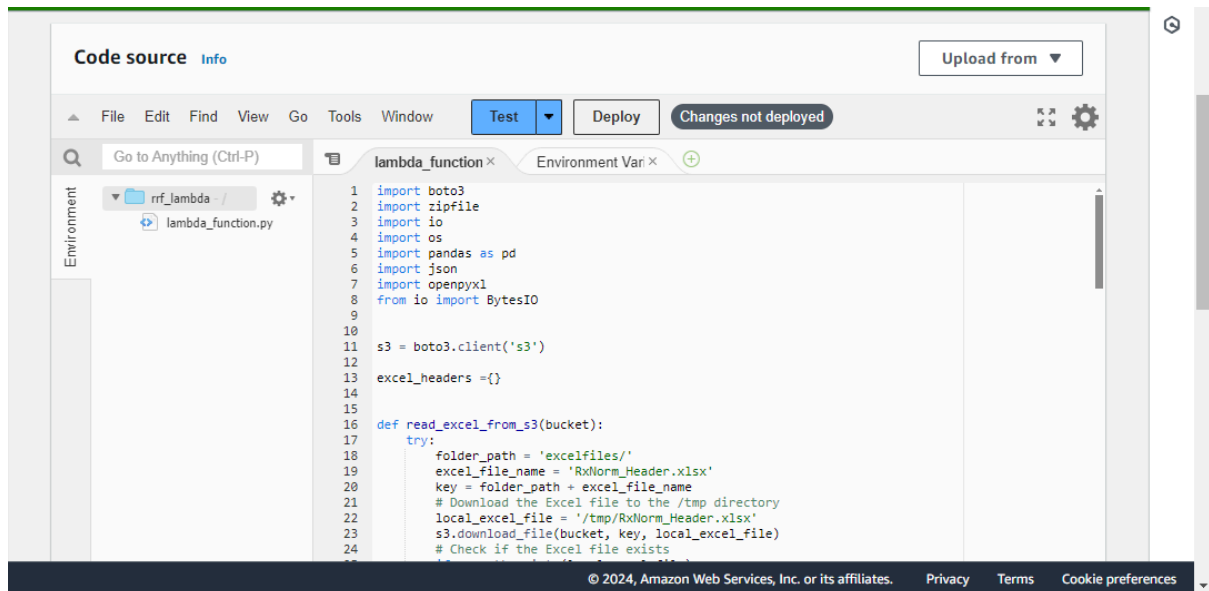
The screenshot shows the "Add layer" dialog in the AWS Lambda console. The dialog is titled "Add layer" and has a search bar at the top with "s3" entered. The "Function runtime settings" section shows the runtime as "Python 3.12" and the architecture as "x86_64". The "Choose a layer" section shows the "AWS layers" option selected. Below this, the "AWS layers" section shows a list of layers provided by AWS that are compatible with the function's runtime. The layer "AWSSDKPandas-Python312" is selected, and the version "4" is chosen.

At the bottom of the dialog, there are "Cancel" and "Add" buttons.

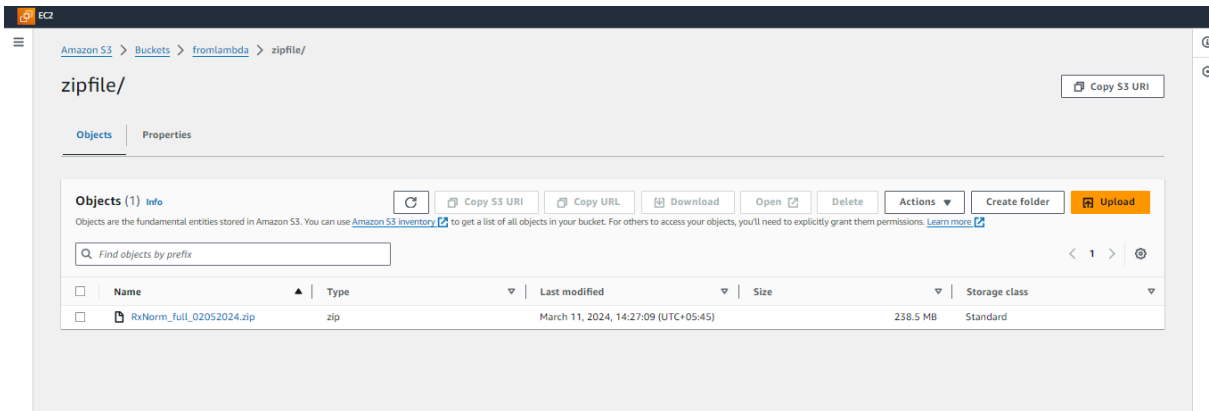
Changing lambda configuration



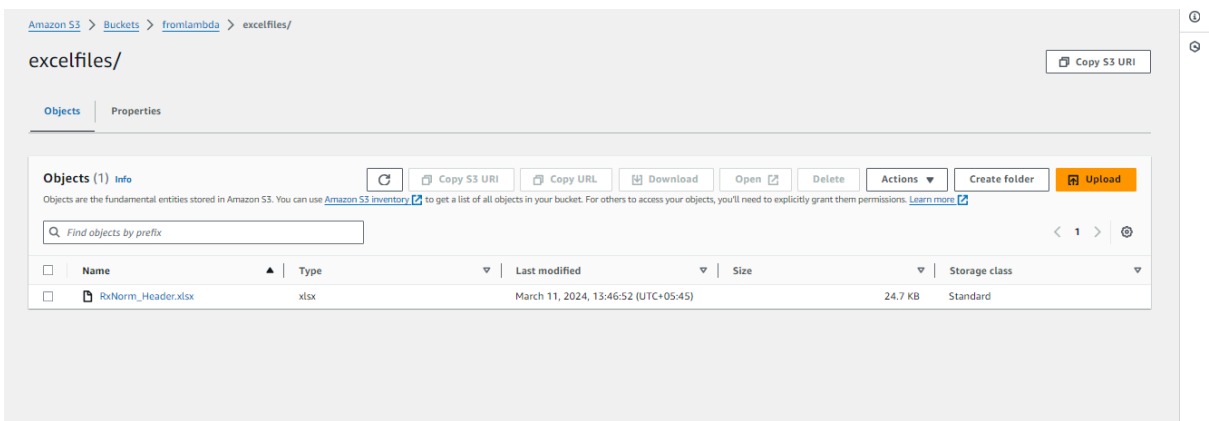
Adding code to execute in lambda



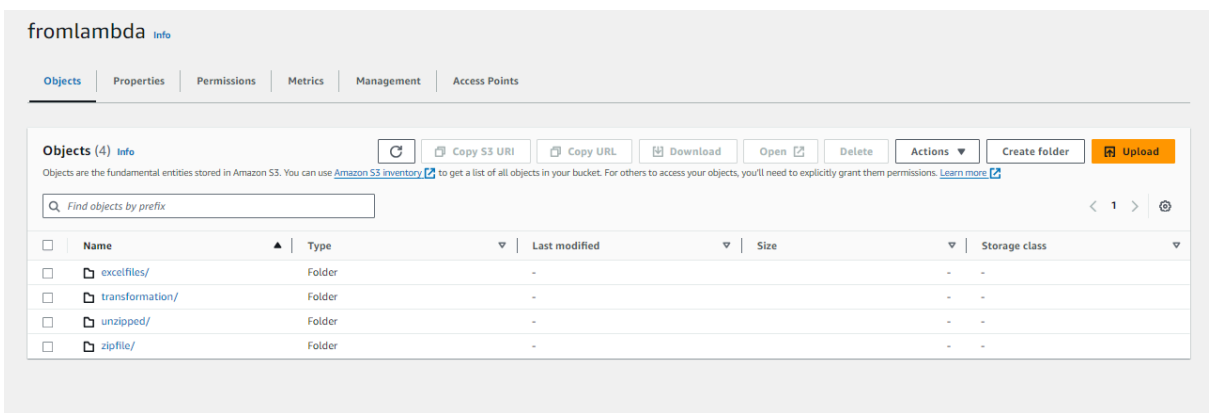
Uploading zip file in the designated folder



Uploading excel file in the designated folder



unzipped/ and transformation/ folders were created after lambda was triggered.



Contents of unzipped folder

EC2

Amazon S3 > Buckets > fromlambda > unzipped/

unzipped/

Copy S3 URI

Objects | Properties

Objects (9) Info

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 Inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Find objects by prefix

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	RXNATOMARCHIVE.RRF	RRF	March 11, 2024, 14:29:59 (UTC+05:45)	71.4 MB	Standard
<input type="checkbox"/>	RXNCONSO.RRF	RRF	March 11, 2024, 14:30:11 (UTC+05:45)	118.6 MB	Standard
<input type="checkbox"/>	RXNCUL.RRF	RRF	March 11, 2024, 14:30:12 (UTC+05:45)	1.7 MB	Standard
<input type="checkbox"/>	RXNCUICHANGES.RRF	RRF	March 11, 2024, 14:30:12 (UTC+05:45)	14.9 KB	Standard
<input type="checkbox"/>	RXNDOC.RRF	RRF	March 11, 2024, 14:30:12 (UTC+05:45)	214.2 KB	Standard
<input type="checkbox"/>	RXNREL.RRF	RRF	March 11, 2024, 14:31:12 (UTC+05:45)	484.4 MB	Standard
<input type="checkbox"/>	RXNSAB.RRF	RRF	March 11, 2024, 14:31:18 (UTC+05:45)	9.8 KB	Standard
<input type="checkbox"/>	RXNSAT.RRF	RRF	March 11, 2024, 14:32:07 (UTC+05:45)	498.7 MB	Standard
<input type="checkbox"/>	RXNSTY.RRF	RRF	March 11, 2024, 14:32:14 (UTC+05:45)	18.4 MB	Standard

CloudShell Feedback

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Contents of transformation folder

EC2

Amazon S3 > Buckets > fromlambda > transformation/

transformation/

Copy S3 URI

Objects | Properties

Objects (9) Info

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 Inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Find objects by prefix

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	RXNATOMARCHIVE.c	c	March 11, 2024, 14:29:58 (UTC+05:45)	69.6 MB	Standard
<input type="checkbox"/>	RXNCONSO.c	c	March 11, 2024, 14:30:10 (UTC+05:45)	138.5 MB	Standard
<input type="checkbox"/>	RXNCUL.c	c	March 11, 2024, 14:30:12 (UTC+05:45)	2.1 MB	Standard
<input type="checkbox"/>	RXNCUICHANGES.c	c	March 11, 2024, 14:30:12 (UTC+05:45)	17.9 KB	Standard
<input type="checkbox"/>	RXNDOC.c	c	March 11, 2024, 14:30:12 (UTC+05:45)	271.7 KB	Standard
<input type="checkbox"/>	RXNREL.c	c	March 11, 2024, 14:31:04 (UTC+05:45)	645.2 MB	Standard
<input type="checkbox"/>	RXNSAB.c	c	March 11, 2024, 14:31:18 (UTC+05:45)	10.3 KB	Standard
<input type="checkbox"/>	RXNSAT.c	c	March 11, 2024, 14:31:59 (UTC+05:45)	622.4 MB	Standard
<input type="checkbox"/>	RXNSTY.c	c	March 11, 2024, 14:32:14 (UTC+05:45)	26.1 MB	Standard

CloudShell Feedback

© 2024, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

Comma separated values form RXNATOMARCHIVE.c

Status

✔ Successfully returned 50 records in 3857 ms

Bytes returned: 6945 B

RawFormatted

1RXAUI,AUI,STR,ARCHIVE_TIMESTAMP,CREATED_TIMESTAMP,UPDATED_TIMESTAMP,CODE,IS_BRAND,LAT, LAST_RELEASED,SAUI,VSAB,RXCUI,SAB,TTY,MERGED_TO,RXCUI,Code Set,Version Month

2947,A10335796,Mesna,2020-04-27,2005-03-10,2020-04-27,44,,ENG,2020-04-06,,RXNORM_19AB_200406F,44,RXNORM,IN,44,RxNorm,2024-02-05

31424,A10334758,beta-Alanine,2020-04-27,2005-03-10,2020-04-27,61,,ENG,2020-04-06,,RXNORM_19AB_200406F,61,RXNORM,IN,61,RxNorm,2024-02-05

41684,A10334529,4-Aminobenzoic Acid,2020-04-27,2005-03-10,2020-04-27,74,,ENG,2020-04-06,,RXNORM_19AB_200406F,74,RXNORM,IN,74,RxNorm,2024-02-05

52192,A16791816,Eicosapentaenoic Acid,2020-04-27,2005-03-10,2020-04-27,90,,ENG,2020-04-06,,RXNORM_19AB_200406F,90,RXNORM,PIN,90,RxNorm,2024-02-05

62265,A10334531,5-Hydroxytryptophan,2020-04-27,2005-03-10,2020-11-06,94,,ENG,2020-04-06,,RXNORM_19AB_200406F,94,RXNORM,IN,94,RxNorm,2024-02-05

72311,A16793037,Ticlopidine Hydrochloride,2020-04-27,2005-03-10,2020-04-27,97,,ENG,2020-04-06,,RXNORM_19AB_200406F,97,RXNORM,PIN,97,RxNorm,2024-02-05

82332,A10334533,6-Aminocaproic Acid,2020-04-27,2005-03-10,2020-04-27,99,,ENG,2020-04-06,,RXNORM_19AB_200406F,99,RXNORM,IN,99,RxNorm,2024-02-05

92453,A10334534,6-Mercaptopurine,2010-10-21,2005-03-10,2010-10-21,103,,ENG,2010-10-04,,RXNORM_10AA_101004F,103,RXNORM,IN,103,RxNorm,2024-02-05

102663,A10336065,Oxyquinoline,2020-04-27,2005-03-10,2020-04-27,110,,ENG,2020-04-06,,RXNORM_19AB_200406F,110,RXNORM,IN,110,RxNorm,2024-02-05

114330,A10334539,Acebutolol,2020-04-27,2005-03-10,2020-04-27,149,,ENG,2020-04-06,,RXNORM_19AB_200406F,149,RXNORM,IN,149,RxNorm,2024-02-05

124414,A10334540,Acenocoumarol,2020-04-27,2005-03-10,2020-04-27,154,,ENG,2020-04-06,,RXNORM_19AB_200406F,154,RXNORM,IN,154,RxNorm,2024-02-05

134458,A10334542,Acepromazine,2020-04-27,2005-03-10,2020-04-27,155,,ENG,2020-04-06,,RXNORM_19AB_200406F,155,RXNORM,IN,155,RxNorm,2024-02-05

144565,A10334544,Acetanilide,2006-11-15,2005-03-10,2006-11-15,162,,ENG,0000-00-00,,RXNORM_06AC_061012F,162,RXNORM,IN,162,RxNorm,2024-02-05

Formatted view of the file contents

Status

✔ Successfully returned 50 records in 3857 ms

Bytes returned: 6945 B

RawFormatted

< 1 >

RXAUI	AUI	STR	ARCHIVE_TIMESTAMP	CREATED_TIMESTAMP	UPDATED_TIME
947	A10335796	Mesna	2020-04-27	2005-03-10	2020-04-27
1424	A10334758	beta-Alanine	2020-04-27	2005-03-10	2020-04-27
1684	A10334529	4-Aminobenzoic Acid	2020-04-27	2005-03-10	2020-04-27
2192	A16791816	Eicosapentaenoic Acid	2020-04-27	2005-03-10	2020-04-27
2265	A10334531	5-Hydroxytryptophan	2020-04-27	2005-03-10	2020-11-06
2311	A16793037	Ticlopidine Hydrochloride	2020-04-27	2005-03-10	2020-04-27
2332	A10334533	6-Aminocaproic Acid	2020-04-27	2005-03-10	2020-04-27

Validating row counts before and after

▶	2024-03-11T14:30:02.466+05:45	Row count before transformation: 1133065
▶	2024-03-11T14:30:02.466+05:45	Version month: 2024-02-05
▶	2024-03-11T14:30:02.475+05:45	Row count of RXNCONSO after transformation: 1133065

▶	2024-03-11T14:30:11.008+05:45	Row count before transformation: 153
▶	2024-03-11T14:30:11.008+05:45	Version month: 2024-02-05
▶	2024-03-11T14:30:11.009+05:45	Row count of RXNCUICHANGES after transformation: 153

▶	2024-03-11T14:30:11.136+05:45	Row count before transformation: 30046
▶	2024-03-11T14:30:11.137+05:45	Version month: 2024-02-05
▶	2024-03-11T14:30:11.138+05:45	Row count of RXNCUI after transformation: 30046

▶	2024-03-11T14:30:11.387+05:45	Row count before transformation: 3445
▶	2024-03-11T14:30:11.388+05:45	Version month: 2024-02-05
▶	2024-03-11T14:30:11.388+05:45	Row count of RXNDOC after transformation: 3445

▶	2024-03-11T14:30:24.547+05:45	Row count before transformation: 7154306
▶	2024-03-11T14:30:24.547+05:45	Version month: 2024-02-05
▶	2024-03-11T14:30:24.613+05:45	Row count of RXNREL after transformation: 7154306

▶	2024-03-11T14:31:17.319+05:45	Row count before transformation: 13
▶	2024-03-11T14:31:17.325+05:45	Version month: 2024-02-05
▶	2024-03-11T14:31:17.325+05:45	Row count of RXNSAB after transformation: 13

▶	2024-03-11T14:32:12.579+05:45	Row count before transformation: 461874
▶	2024-03-11T14:32:12.579+05:45	Version month: 2024-02-05
▶	2024-03-11T14:32:12.583+05:45	Row count of RXNSTY after transformation: 461874