MEDICATION ERROR AND EFFECTS (Group-3)

Nishant Jain
Nikki wagholikar
Abhishek Choure
Oendree Das

Data Description:

This dataset talks about the medication error reports, submitted to the FDA, to support the post-marketing safety surveillance program for drug and therapeutic biologic products. The structure of the AERS database adheres to the international safety reporting guidance issued by the International Conference on Harmonisation (ICH E2B). Adverse events and medication errors are coded to terms in the Medical Dictionary for Regulatory Activities (MedDRA) terminology.

Data Files:

https://drive.google.com/file/d/1XiRsJBIZpb2sZ2caThtjj71nAzl3rGEq/view?usp=sharing

Technologies Used:

- IICS
- PowerBI
- Snowflake
- Python
- SQL

Data Dictionary:

1. **Demographics_data_2015.csv**

Column Name	Description
primaryid	Unique Id in the table
caseid	Unique Id for every case
Event Date	Date the adverse event occurred or began. (YYYYMMDD format). If a complete date is

	not available, a partial date is provided.
Manufacture Received Date	Date manufacturer first received initial information. In subsequent versions of a case, the latest manufacturer received date will be provided (YYYYMMDD format). If a complete date is not available, a partial date will be provided. See the NOTE on dates at the end
Init Fda Dt	Init Fda Dt
FDA Received Date	Date FDA received Case. In subsequent versions of a case, the latest manufacturer received date will be provided (YYYYMMDD format).
Report Type	Code for the type of report submitted. "EXP" means expedited (15 day), "PER" is periodic, and "DIR" is direct.
Manufacturer Number	Manufacturer's unique report identifier.
Manufacturer Filing Report	Coded name of manufacturer sending report; if not found, then verbatim name of organization
Patient's Age	Numeric value of patient's age at event
Patient's Age Unit	Unit abbreviation for patient's age. "DEC" is decade, "YR" is year, "MON" is month, "WK" is week, "DY" is day, and "HR" is hour.
Sex	Gender
Patient's Weight	Weight. Numeric value of patient's weight.
Patient's Weight Unit	Unit abbreviation for patient's weight. "KG" is kilograms, "LBS" is pounds, and "GMS" is grams.
Report Date	Date report was sent (YYYYMMDD format). If a complete date is not available, a partial date is provided.
Occupation Code	Abbreviation for the reporter's type of

	occupation in the latest version of a case. "MD" is physician, "PH" is pharmacist, "OT" is other health professional, "LW" is lawyer, and "CN" is consumer.
Reporter Country	Country
Occurrence Country	Country where it occurred
Serialid	Serialid

2. **Drug Information.csv**

Column Name	**Description**
Isr	Number that uniquely identifies an AERS report. Primary link field between data files.
Drug Sequence No.	Unique number for identifying a drug for a case. This field can be used as a primary link between different tables within this dataset.
Drug Role	Code for drug's reported role in event. "PS" is primary suspect drug, "SS" is secondary suspect drug, "C" is concomitant, and "I" is interacting.
Drug Name	Name of medicinal product. If a "Valid Trade Name" is populated for this case, then Drug Name = Valid Trade Name; if not, then Drug Name = "Verbatim" name, exactly as entered on the report. For the great majority of reports, there is a "Valid Trade Name."
Validated/Verbatim Name	"1" indicates that a validated trade name is used for the drug name and "2" indicates that a verbatim name is used.
Route	The route of drug administration.
Dose	Verbatim text for dose, frequency, and route, exactly as entered on report.

Rechallenge	Rechallenge code, indicating if reaction recurred when drug therapy was restarted. "Y" is positive rechallenge, "N" is negative rechallenge, "U" is unknown, and "D" does not apply.
SerialId	Serialid

3. **Drug Therapy Duration.csv**

Column Name	**Description**
ISR	Unique Id in the table
caseid	Unique Id for every case
Drug Sequence	Drug Sequence Number
Start Date	Date therapy was started (or re-started) for this drug (YYYYMMDD) If a complete date not available, a partial date is provided.
Start Date Unparsed	Date therapy was started (or re-started) for this drug (YYYYMMDD) If a complete date not available, a partial date is provided, in raw data format.
End Date	Date therapy was stopped for this drug. (YYYYMMDD) If a complete date not available, a partial date will be provided.
End Date Unparsed	Date therapy was stopped for this drug. (YYYYMMDD) If a complete date not available, a partial date will be provided, in raw data format.
Serialid	Serialid

4. **Event Terms.csv**

Column Name	Description
Isr(primaryid)	Number that uniquely identifies an AERS report. Primary link field between data files.
Caseid	Number that uniquely identifies the cases
Preferred Term(PT)	Preferred Term" level medical terminology describing the event, using the Medical Dictionary for Regulatory Activities (MedDRA). The order of the terms for a given event does not imply priority. In other words, the first term listed is not necessarily considered more significant than the last one listed.
Serialid	Serialid

5. **Patient Outcomes.csv**

Column Name	Description
primaryid	The number that uniquely identifies an AERS report. Primary link field between data files.
caseid	The patient outcome
outc_cod	Outcome code
Serialid	Serialid
outc_cod_description	Outcome Code Definition. DE, Death | LT, Life-Threatening | HO, Hospitalization - Initial or Prolonged | DS, Disability | CA, Congenital Anomaly | RI, Required Intervention to Prevent Permanent Impairment/Damage | OT, Other

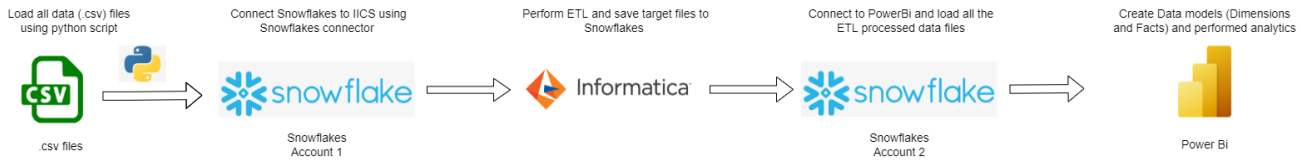
6. Preferred Term Indicators.csv

Column Name	Description
primaryid	primaryid
caseid	caseid
indi_pt	“Preferred Term" level medical terminology describing the indication for use, using the Medical Dictionary for Regulatory Activities (MedDRA). The order of the terms for a given event does not imply priority. In other words, the first term listed is not necessarily considered more significant than the last one listed.

7. Report Sources.csv

Column Name	Description
primary_id	Primary id
caseid	caseid
rpsr_code	Report Source Code Definition. FGN, Foreign | SDY, Study | LIT, Literature | CSM, Consumer | HP, Health Professional | UF, User Facility | CR, Company Representative | DT, Distributor | OTH, Other
serialid	serialid

Architecture diagram



SQL

- Creating schema for all the tables

```
create or replace TABLE MOCK_PROJECT_DB.PUBLIC."Demorgaphics_data_2015" (
    "primaryid" NUMBER(38,0) NOT NULL,
    "caseid" NUMBER(38,0),
    "i_f_code" VARCHAR(40),
    "event_dt" TIMESTAMP_NTZ(9),
    "mfr_dt" TIMESTAMP_NTZ(9),
    "init_fda_dt" TIMESTAMP_NTZ(9),
    "fda_dt" TIMESTAMP_NTZ(9),
    "rept_cod" VARCHAR(40),
    "mfr_num" VARCHAR(255),
    "mfr_sndr" VARCHAR(100),
    "age" NUMBER(38,0),
    "age_cod" VARCHAR(10),
    "age_grp" VARCHAR(10),
    "sex" VARCHAR(10),
    "wt" NUMBER(38,0),
    "wt_cod" VARCHAR(10),
    "rept_dt" TIMESTAMP_NTZ(9),
    "occp_cod" VARCHAR(10),
    "reporter_country" VARCHAR(40),
    "occr_country" VARCHAR(10),
    "serialid" NUMBER(38,0),
    primary key ("primaryid")
);


create or replace TABLE MOCK_PROJECT_DB.PUBLIC."DrugInformation_data_2015"
(
    "primaryid" NUMBER(38,0),
    "caseid" NUMBER(38,0),
    "drug_seq" NUMBER(38,0),
    "role_cod" VARCHAR(10),
    "drugname" VARCHAR(16777216),
    "prod_ai" VARCHAR(16777216),
    "val_vbm" NUMBER(38,0),
    "route" VARCHAR(255),
    "dose_vbm" VARCHAR(16777216),
```

```
    "cum_dose_chr" NUMBER(38,10),
    "cum_dose_unit" VARCHAR(255),
    "rechal" VARCHAR(255),
    "dose_amt" VARCHAR(16777216),
    "dose_unit" VARCHAR(255),
    "dose_form" VARCHAR(255),
    "dose_freq" VARCHAR(255),
    "serialid" NUMBER(38,0)
);



create or replace TABLE
MOCK_PROJECT_DB.PUBLIC."DrugTherapyDuration_data_2015" (
    "primaryid" NUMBER(38,0),
    "caseid" NUMBER(38,0),
    "dsg_drug_seq" NUMBER(38,0),
    "start_dt" TIMESTAMP_NTZ(9),
    "start_dt_unparsed" NUMBER(38,0),
    "end_dt" TIMESTAMP_NTZ(9),
    "end_dt_unparsed" NUMBER(38,0),
    "dur_cod" VARCHAR(40),
    "serialid" NUMBER(38,0)
);



create or replace TABLE MOCK_PROJECT_DB.PUBLIC."EventTerms_data_2015" (
    "primaryid" NUMBER(38,0),
    "caseid" NUMBER(38,0),
    "pt" VARCHAR(4000),
    "drug_rec_act" VARCHAR(4000),
    "serialid" NUMBER(38,0)
);



create or replace TABLE MOCK_PROJECT_DB.PUBLIC."PatientOutcomes_data_2015"
(
```

```
    "primaryid" NUMBER(38,0),
    "caseid" NUMBER(38,0),
    "outc_cod" VARCHAR(10),
    "outc_cod_description" VARCHAR(255),
    "serialid" NUMBER(38,0)
);



create or replace TABLE
MOCK_PROJECT_DB.PUBLIC."PreferredTermIndicators_data_2015" (
    "primaryid" NUMBER(38,0),
    "caseid" NUMBER(38,0),
    "indi_drug_seq" NUMBER(38,0),
    "indi_pt" VARCHAR(4000),
    "serialid" NUMBER(38,0)
);


create or replace TABLE MOCK_PROJECT_DB.PUBLIC."ReportSources_data_2015" (
    "primary_id" NUMBER(38,0),
    "caseid" NUMBER(38,0),
    "rpsr_code" VARCHAR(10),
    "serialid" NUMBER(38,0)
```

ETL

- Configured snowflake connector in IICS to fetch the tables which were uploaded using the below .py script

Load.py

```
#Importing the required packages for all your data framing needs.
from tokenize import String
import pandas as pd
# The Snowflake Connector library.
import snowflake.connector as snow
from snowflake.connector.pandas_tools import write_pandas
```

```
conn = snow.connect(user="<USERNAME>",
password="<PASSWORD>",
account="<ACCOUNT NAME> eg - <in---.uae-north.azure>",
# (the prefix in your snowflake space... for example,
# company.snowflakecomputing.com would just be "company" as the ACCOUNT
name)
warehouse="<WAREHOUSE NAME>",
database="<DATABASE NAME>",
schema="<SCHEMA NAME>")

# Create a cursor object.
cur = conn.cursor()

## Phase II: Upload from the Exported Data File.
# Let's import a new dataframe so that we can test this.
original = r"<LOCAL FILEPATH>" # <- Replace with your path.
delimiter = "," # Replace if you're using a different delimiter.

# Get it as a pandas dataframe.
total = pd.read_csv(original, sep = delimiter, low_memory=False)

# Actually write to the table in snowflake.
write_pandas(conn, total, "<TABLE NAME | SNOWFLAKE>")

# (Optionally, you can check to see if what you loaded is identical
# to what you have in your pandas dataframe. Perhaps... a topic for a
future
# blog post.)

## Phase III: Turn off the warehouse.
# Create a cursor object.
cur = conn.cursor()

sql = "ALTER WAREHOUSE MOCK_PROJECT SUSPEND"
cur.execute(sql)

# Close your cursor and your connection.
cur.close()
conn.close()
```

Data warehouse

Source Tables:

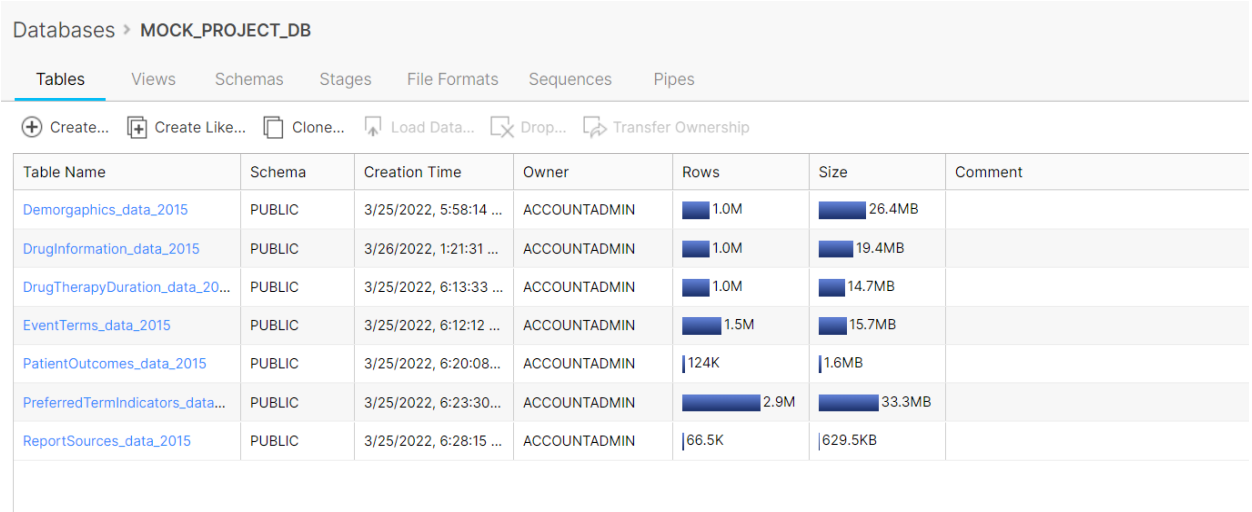
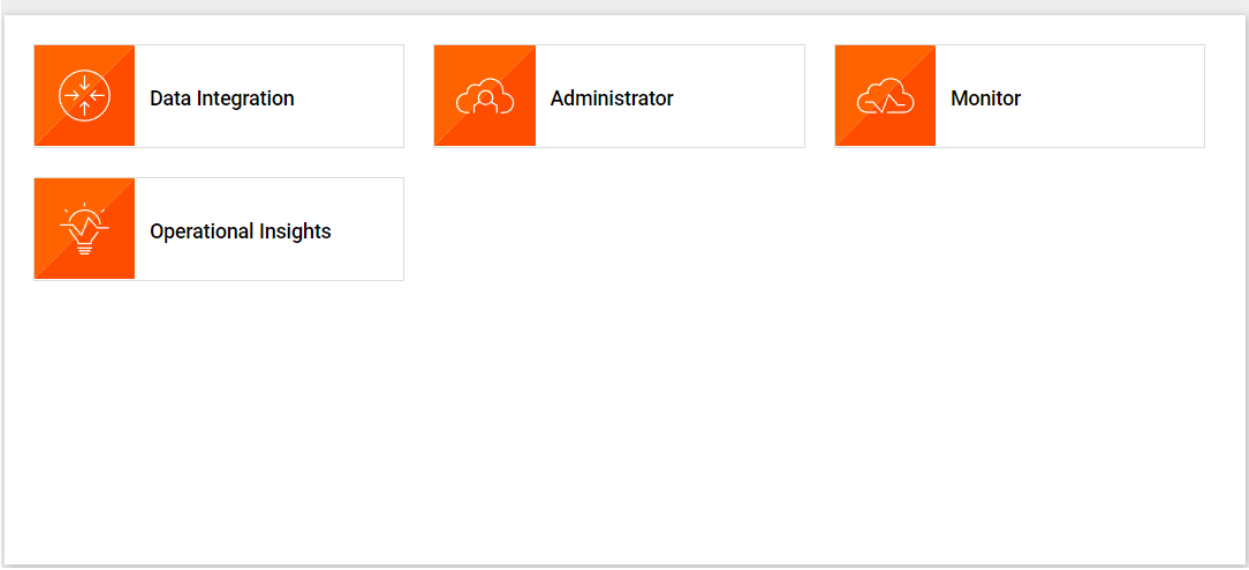


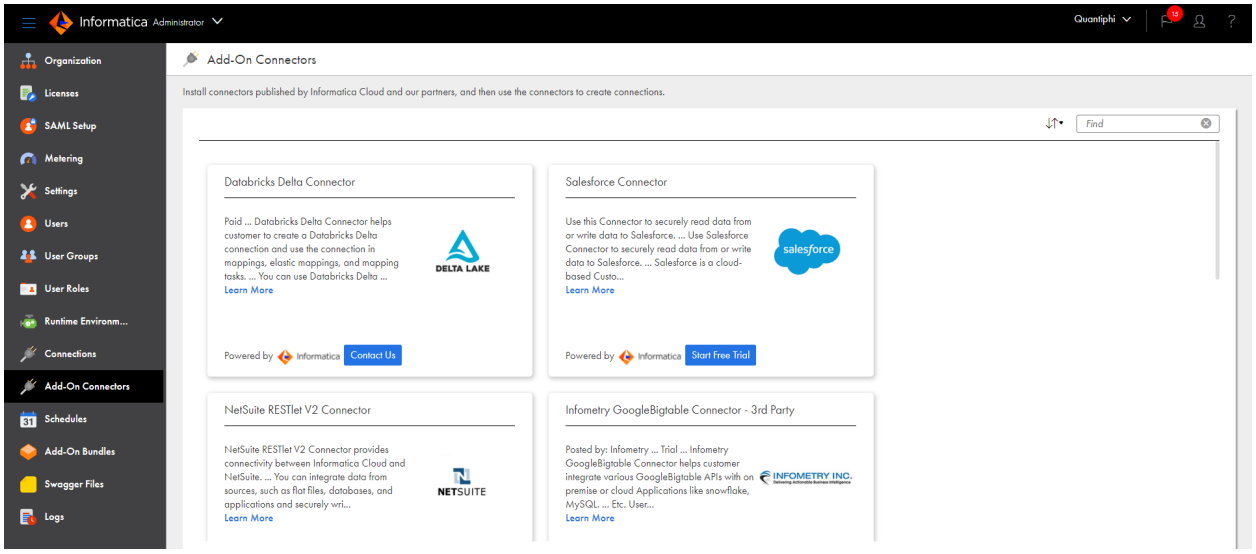
Table Name	Schema	Creation Time	Owner	Rows	Size	Comment
Demorgaphics_data_2015	PUBLIC	3/25/2022, 5:58:14 ...	ACCOUNTADMIN	1.0M	26.4MB	
DrugInformation_data_2015	PUBLIC	3/26/2022, 1:21:31 ...	ACCOUNTADMIN	1.0M	19.4MB	
DrugTherapyDuration_data_20...	PUBLIC	3/25/2022, 6:13:33 ...	ACCOUNTADMIN	1.0M	14.7MB	
EventTerms_data_2015	PUBLIC	3/25/2022, 6:12:12 ...	ACCOUNTADMIN	1.5M	15.7MB	
PatientOutcomes_data_2015	PUBLIC	3/25/2022, 6:20:08...	ACCOUNTADMIN	124K	1.6MB	
PreferredTermIndicators_data...	PUBLIC	3/25/2022, 6:23:30...	ACCOUNTADMIN	2.9M	33.3MB	
ReportSources_data_2015	PUBLIC	3/25/2022, 6:28:15 ...	ACCOUNTADMIN	66.5K	629.5KB	

IICS:

Step 1: Go to Admin Panel

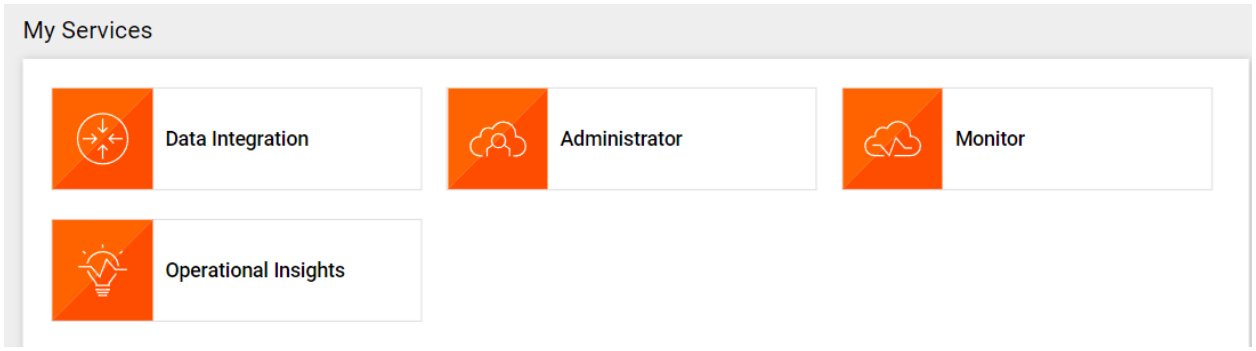


Step 2: Go to Add-On Connectors

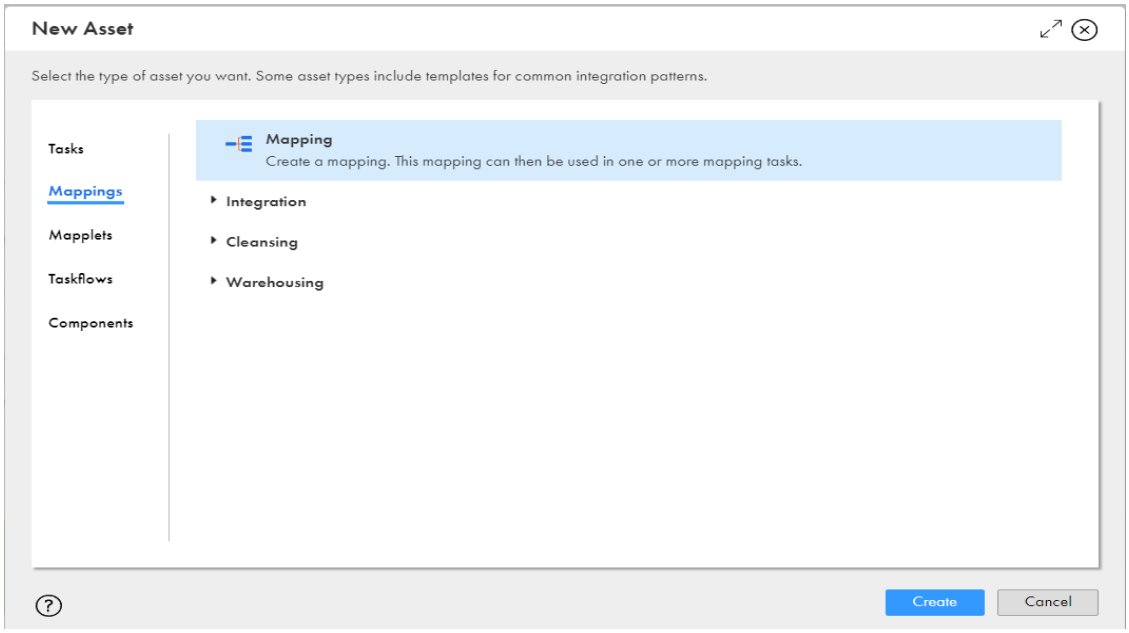


Step 3: Search for Snowflake connector and activate it.

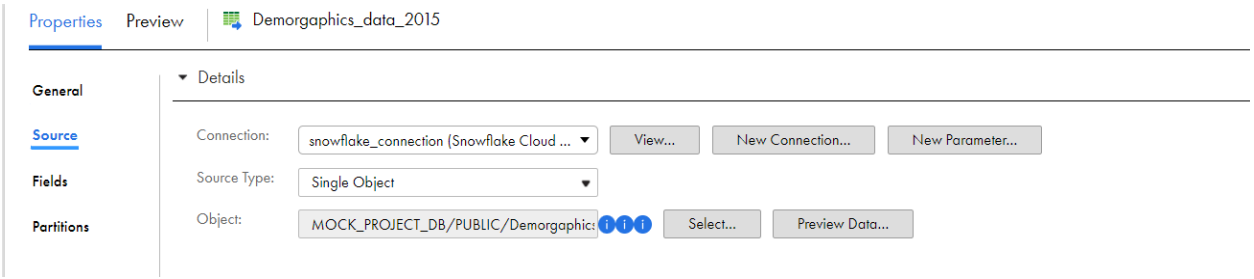
Step 4: Go to Data Integration



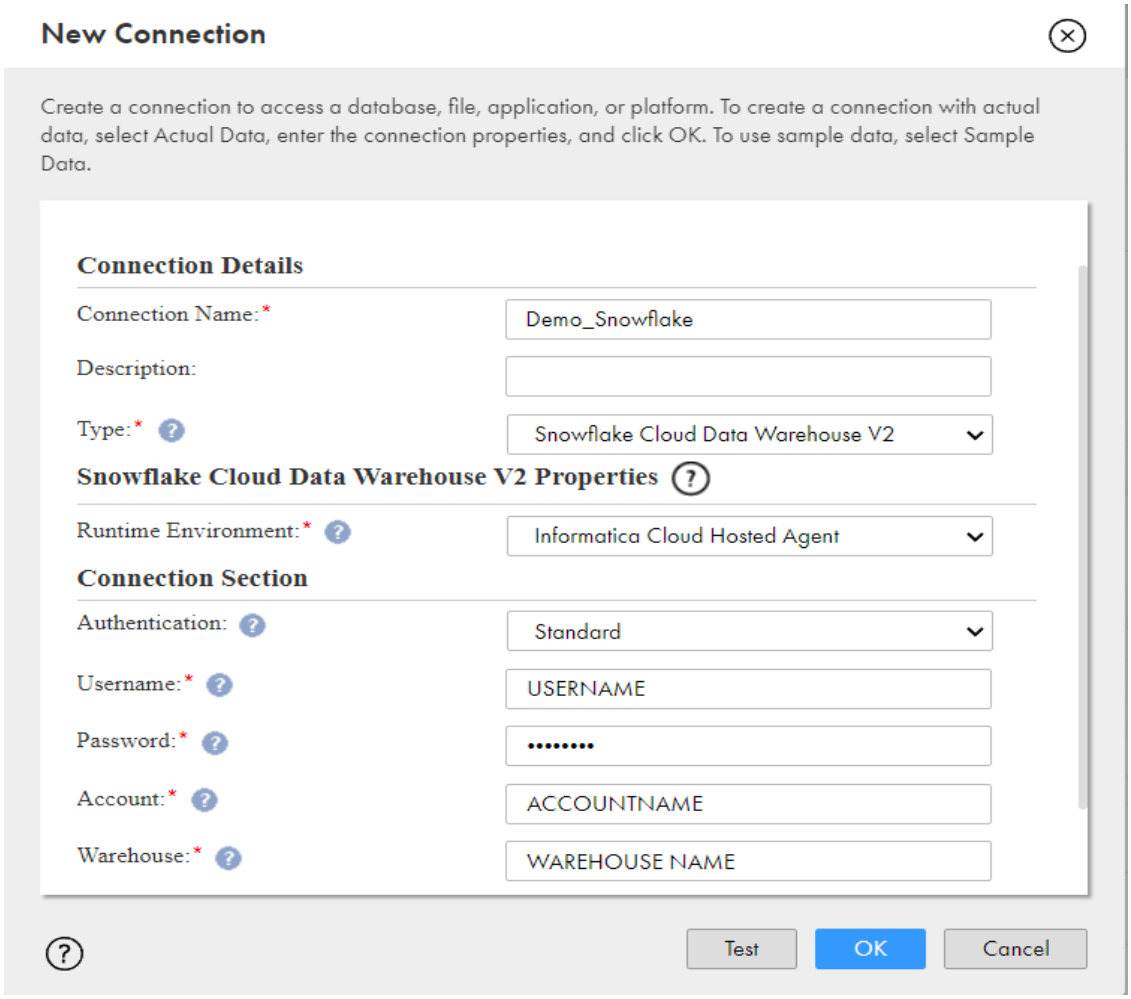
Step 5: Create a new mapping



Step 6: Source

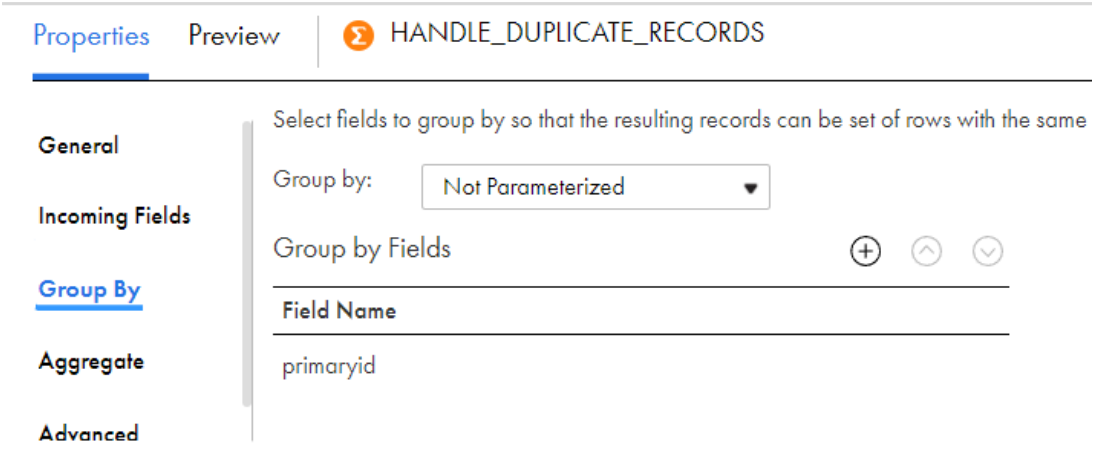


Step 7: Configure connection (if not exists) create a new Connection, test it and save.



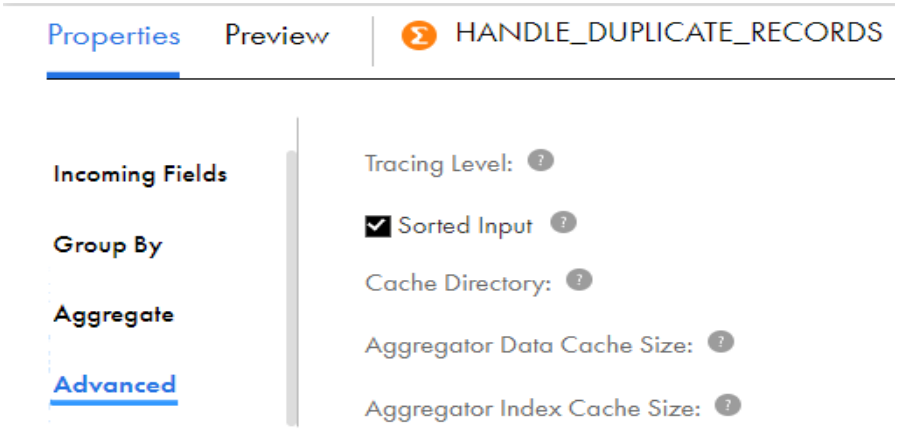
Step 8: Create Mapping

- **Aggregator - Remove Duplicate Records**

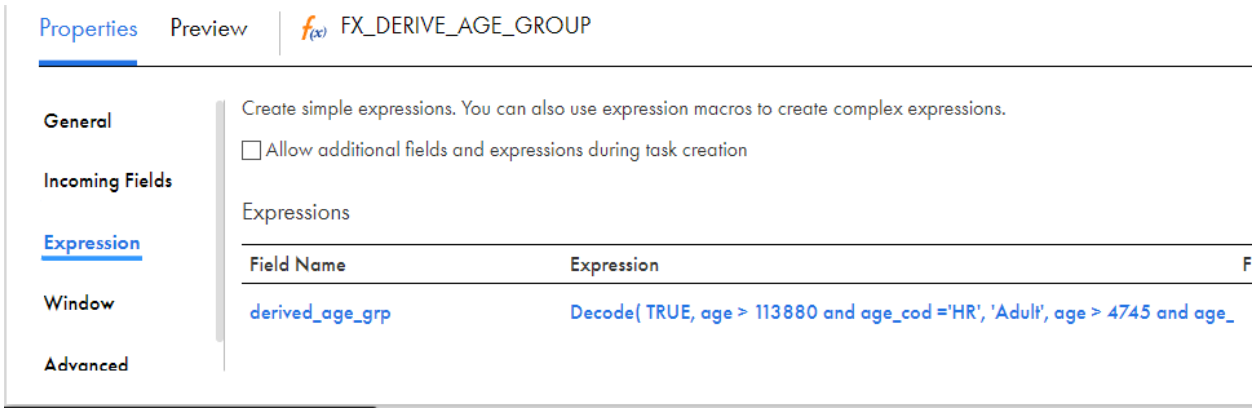


(a) Select a column on whose basis you want to delete duplicate records.

(b) Check the Sorted Input

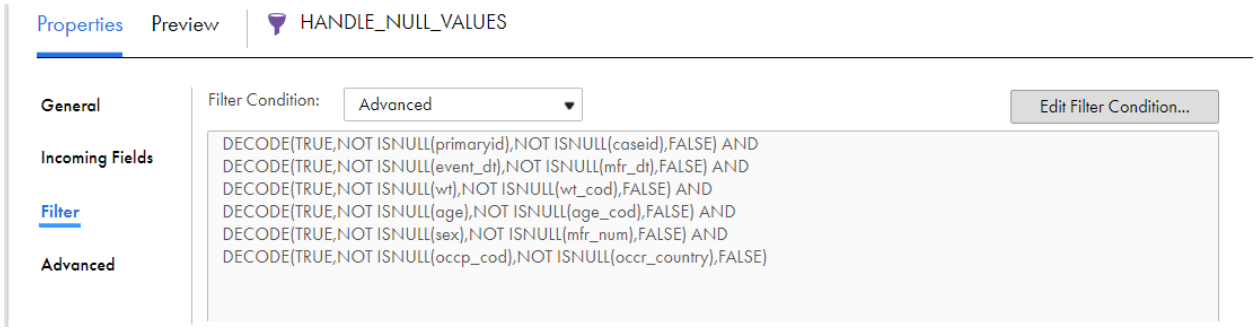


- **(OPTIONAL) Expression - Derive age group from the age column**

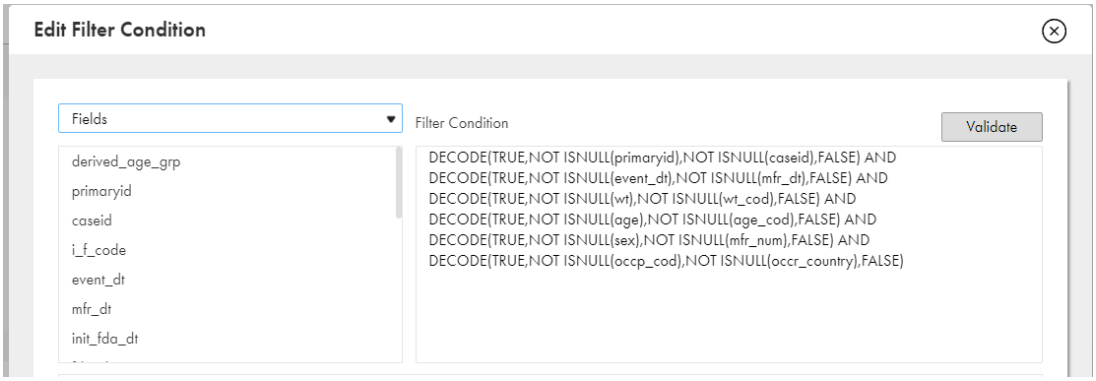


(a) Add a new expression.

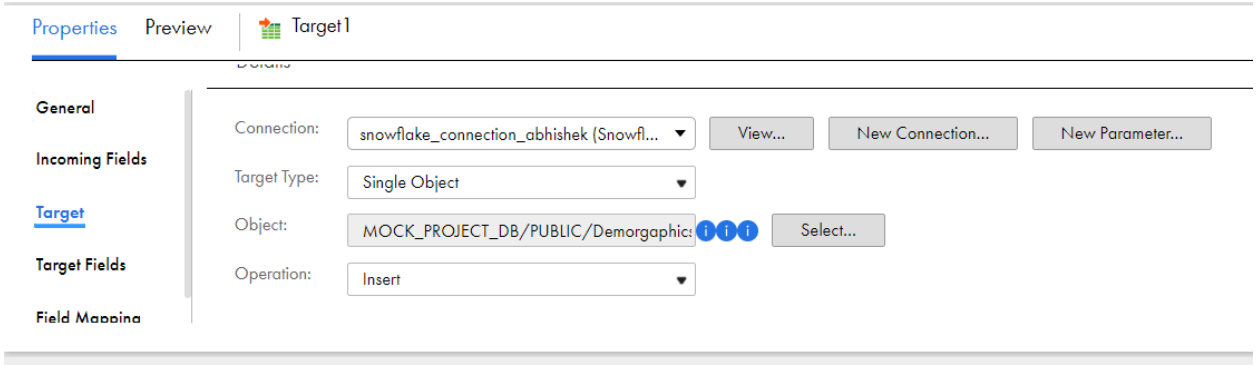
- **Filter - Remove records with NULL values.**



(a) Select Advanced from the dropdown and use Decode to write multiple conditions and validate it.

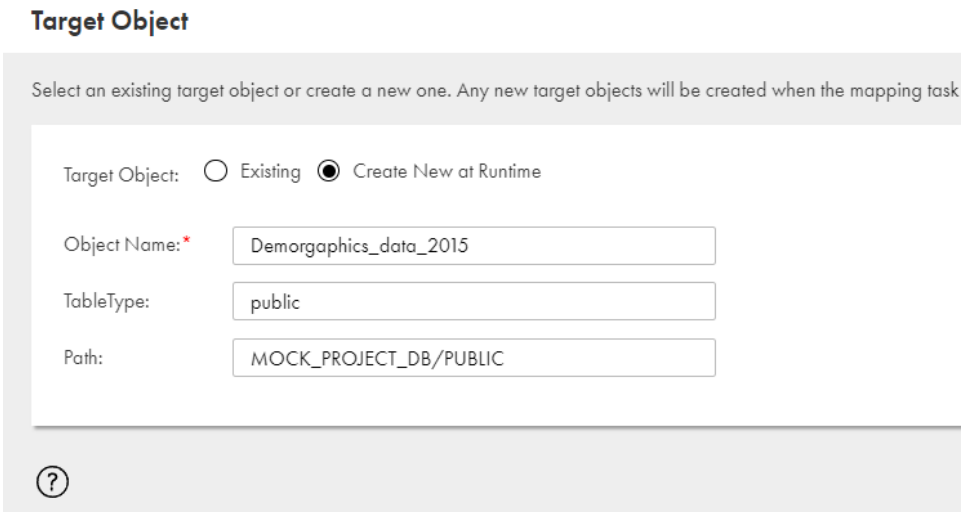


- **Target**

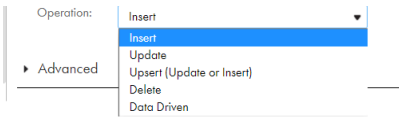


(a) Repeat Step 7 if the target connection is not created.

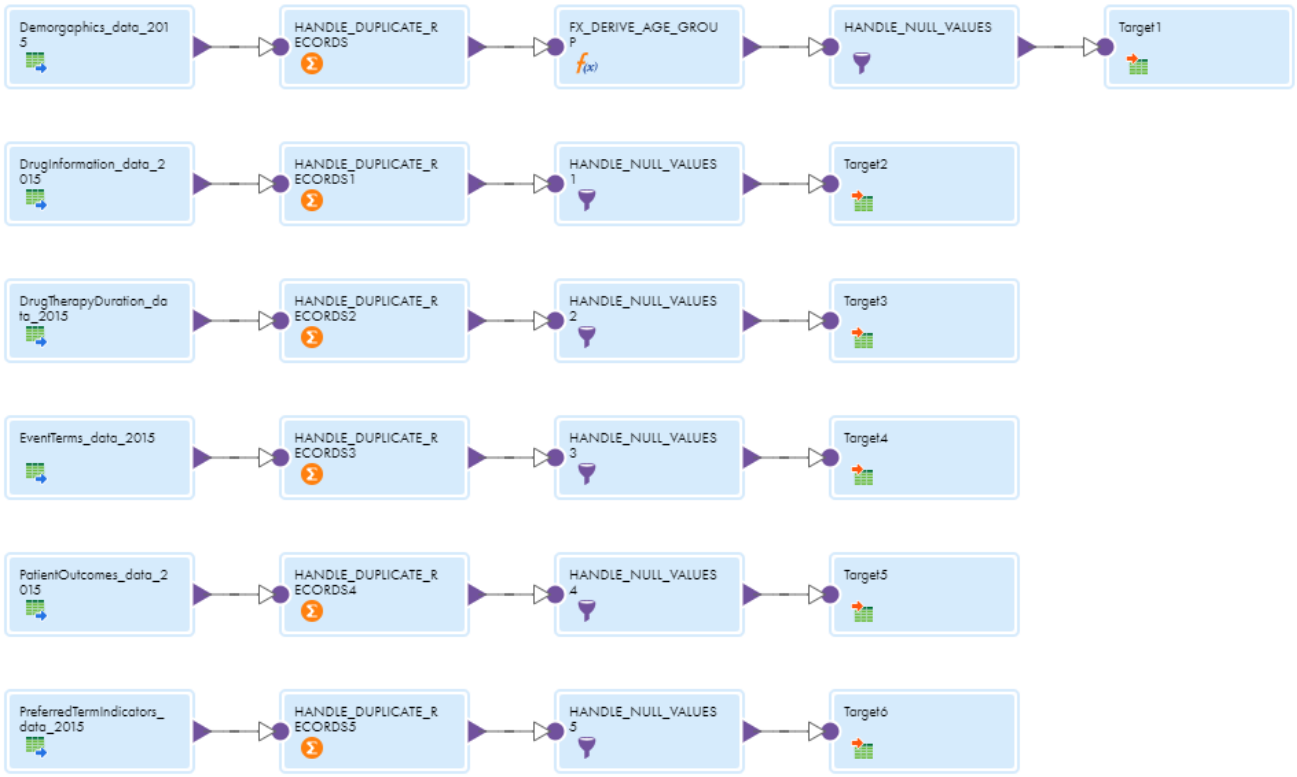
(b) Select Object or Create a new at runtime



(c) Choose operation as INSERT if creating a new at runtime or update if object already exists



Step 9 : Save and Run the Mapping



Target Tables:

Databases › MOCK_PROJECT_DB

Tables | Views | Schemas | Stages | File Formats | Sequences | Pipes

Create... | Create Like... | Clone... | Load Data... | Drop... | Transfer Ownership

Table Name	Schema	Creation Time	Owner	Rows	Size	Comment
Demorgaphics_data_2015	PUBLIC	3/29/2022, 5:09:35...	ACCOUNTADMIN	83.8K	3.0MB	
DrugInformation_data_2015	PUBLIC	3/29/2022, 6:33:36...	ACCOUNTADMIN	86.8K	2.8MB	
DrugTherapyDuration_data_20...	PUBLIC	3:26:51 PM	ACCOUNTADMIN	114.7K	1.9MB	
EventTerms_data_2015	PUBLIC	3/29/2022, 5:36:51 ...	ACCOUNTADMIN	530.9K	8.1MB	
PatientOutcomes_data_2015	PUBLIC	3/29/2022, 5:36:37...	ACCOUNTADMIN	109.0K	1.4MB	
PreferredTermIndicators_data...	PUBLIC	3/29/2022, 5:40:38...	ACCOUNTADMIN	1.2M	13.4MB	
ReportSources_data_2015	PUBLIC	3/29/2022, 5:45:26...	ACCOUNTADMIN	44.0K	412KB	

Transformations used:

- Source
- Filter
- Aggregator
- Function Expression
- Target

Data modeling

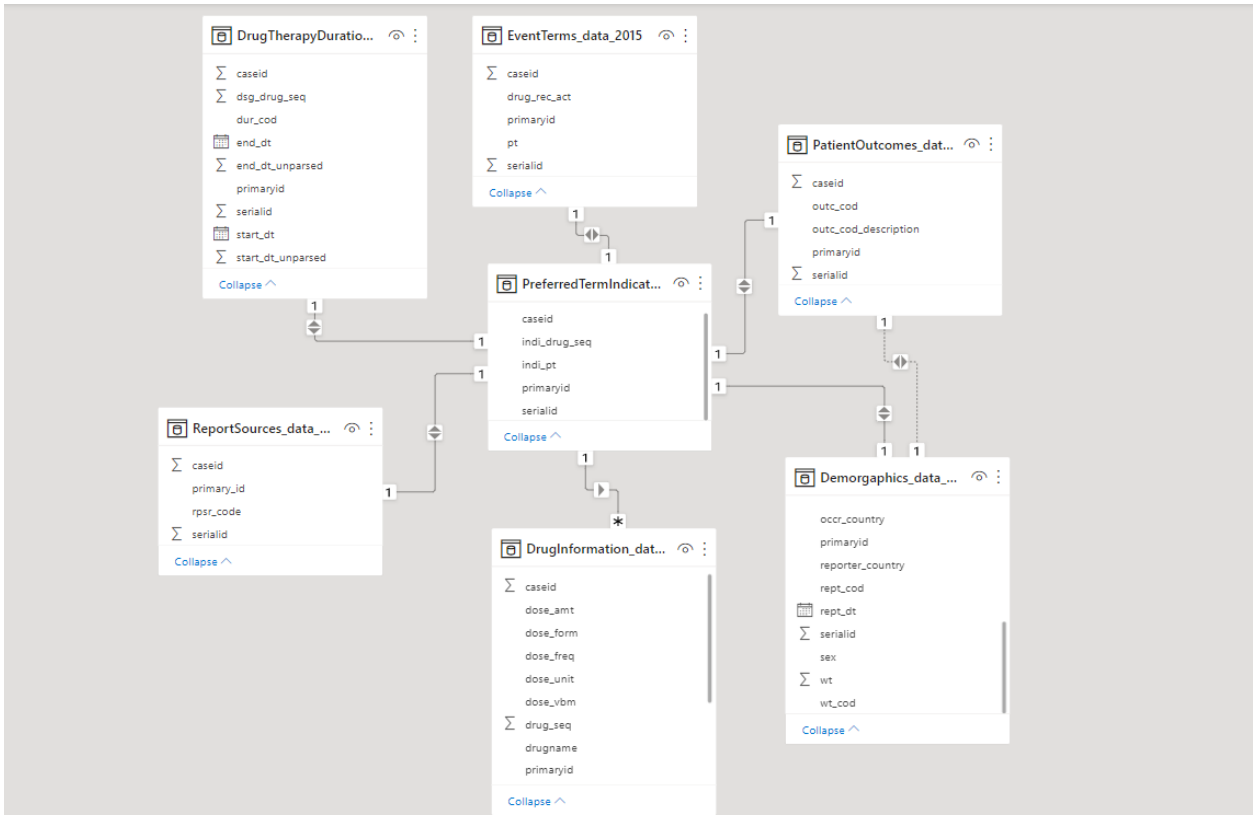
- Identify dimensions and facts and perform data modeling.

  Dimension Tables :

  1. Demorgaphics_data_2015
  2. DrugInformation_data_2015
  3. ReportSources_data_2015

  Fact Tables :

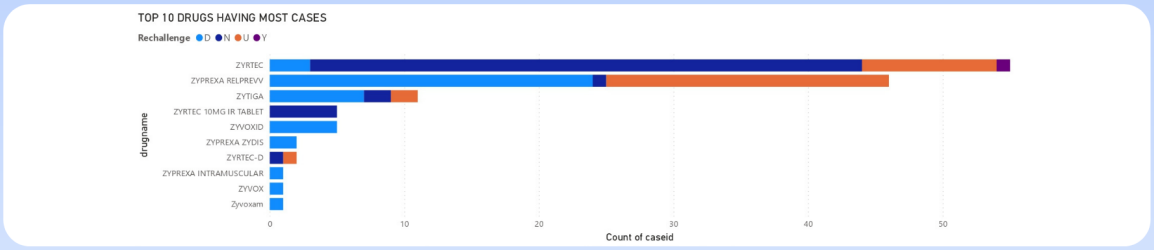
  1. DrugTherapyDuration_data_2015
  2. EventTerms_data_2015
  3. PatientOutcomes_data_2015
  4. PreferredTermIndicator_data_2015

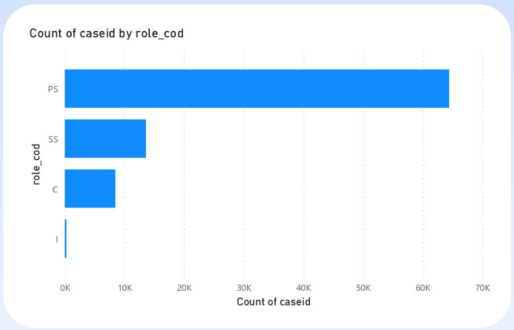


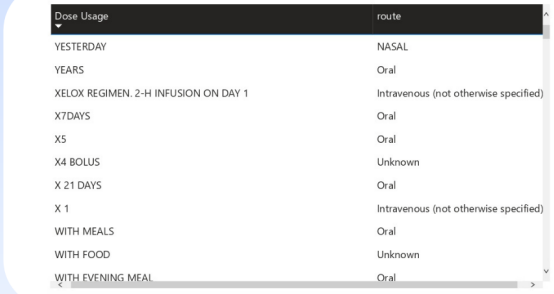
Dashboarding and story building

- **Dashboarding using PowerBI | Snowflake DWH**

DASHBOARD 1

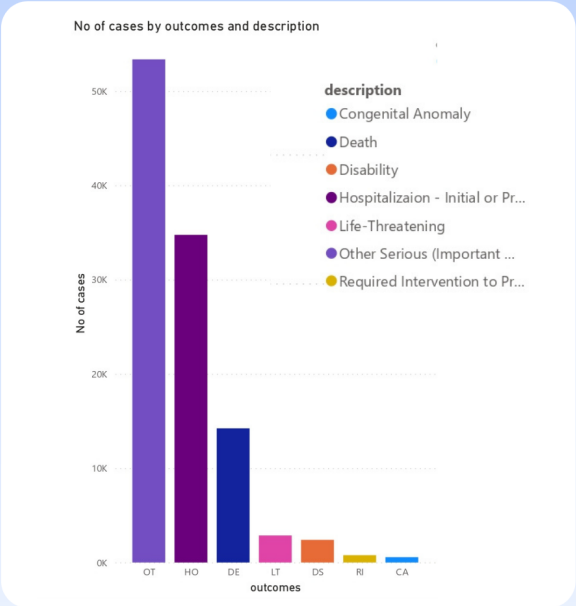


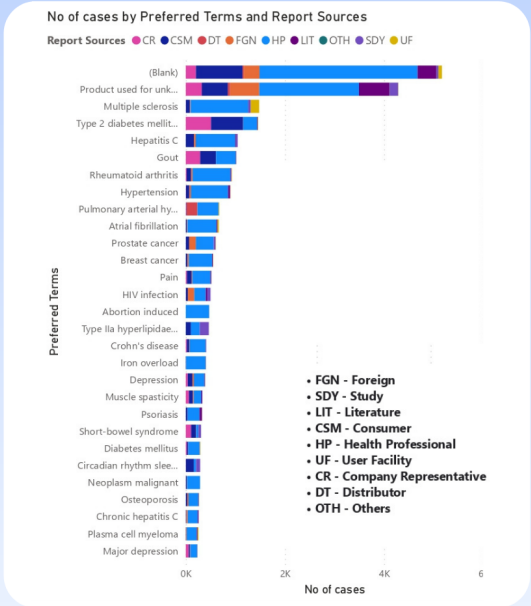




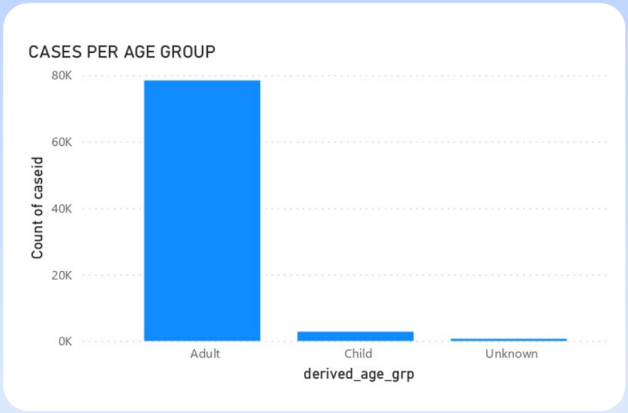
Dose Usage	route
YESTERDAY	NASAL
YEARS	Oral
XELOX REGIMEN. 2-H INFUSION ON DAY 1	Intravenous (not otherwise specified)
X7DAYS	Oral
X5	Oral
X4 BOLUS	Unknown
X 21 DAYS	Oral
X 1	Intravenous (not otherwise specified)
WITH MEALS	Oral
WITH FOOD	Unknown
WITH EVENING MEAL	Oral

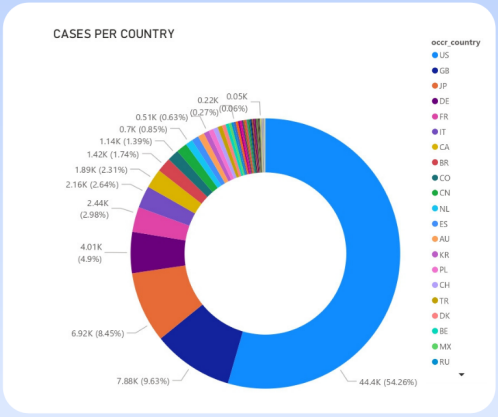
DASHBOARD 2





DASHBOARD 3





81.83K

No of cases

DASHBOARD 4

31K

Max of Duration (no of days)

1

Min of Duration(no of days)

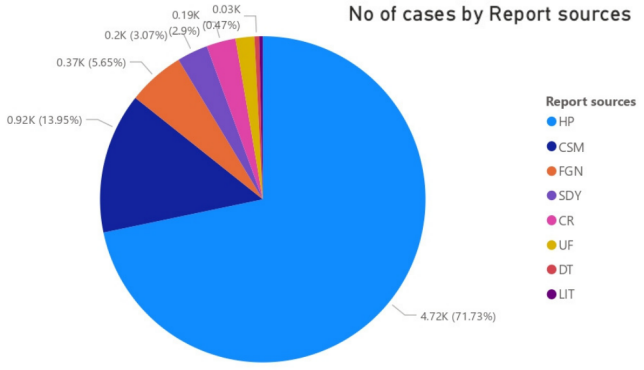
290.84

Average of Duration(no of days)

8

Total no of report sources

Year	Quarter	Month	Day	Year	Quarter	Month	Day	Duration
1930	Qtr 3	September	3	2015	Qtr 4	October	31	311
1942	Qtr 3	September	30	2015	Qtr 3	September	19	266
1950	Qtr 1	January	1	2010	Qtr 1	January	1	219
1954	Qtr 1	January	1	2010	Qtr 1	January	1	204
1956	Qtr 1	February	15	2015	Qtr 1	February	19	215
1959	Qtr 1	January	1	2012	Qtr 1	January	1	193
1960	Qtr 1	January	1	2010	Qtr 1	January	1	182
1960	Qtr 1	January	1	2014	Qtr 1	January	1	197
1969	Qtr 2	June	23	2015	Qtr 2	April	10	167
1970	Qtr 1	January	1	2012	Qtr 1	January	1	153
1970	Qtr 3	July	9	2013	Qtr 3	July	9	157
1972	Qtr 1	January	1	1999	Qtr 1	January	1	98
1972	Qtr 2	May	1	1989	Qtr 4	December	12	64
1973	Qtr 1	January	1	2014	Qtr 2	April	1	150
Total								269662



References

- https://onlinehelp.informatica.com/IICS/prod/admin/en/index.htm#page/aa-admin-whats-new/Serverless_runtime_environments.html
- https://docs.snowflake.com/en/
- https://stackoverflow.com/
- Questions and Answers on FDA's Adverse Event Reporting System (FAERS) | FDA
- Qlik Sense (fda.gov)