Course 2 Module 5 Programming Assignment

Assignment is to ETL MIMIC data into the OMOP CONDITION_OCCURRENCE table

Detailed instructions with Slide Notes

Assignment is to ETL MIMIC data into the OMOP CONDITION_OCCURRENCE table

ETL Steps

- 1. Understand source/target data models
- 2. Profile source tables
- 3. Create ETL mappings
- 4. Write transformation code
- 5. Execute transformation
- 6. Perform data quality assessment
- 7. Package documentation

Step 1: Understand source/target data models

CONDITION_OCCURRENCE is the TARGET OMOP table.

Read the OMOP documentation about the type of data stored in CONDITION_OCCURRENCE and for three fields below that are in that table:

- person_id
- visit_occurrence_id
- condition source value

Definitions of each variable from OMOP documentation in NOTES on next page.



Table Details: condition occurrence



The definition of the variables from OMOP documentation: person_id (int) - A foreign key identifier to the Person who is experiencing the condition. The demographic details of that Person are stored in the PERSON table. – This will be same as to SUBJECT_ID in MIMIC data.

visit_occurrence_id (int) - A foreign key to the visit in the VISIT_OCCURRENCE table during which the Condition was determined (diagnosed). – This is equivalent to HADM_ID in MIMIC data.

condition_source_value (Char (50)) -The source code for the Condition as it appears in the source data. This code is mapped to a Standard Condition Concept in the Standardized Vocabularies and the original code is stored here for reference. This will be exactly same as to MIMIC data (ICD 9code).

Step 1: Understand source/target data models

CONDITION_OCCURRENCE is the TARGET OMOP table.

Select one or more MIMIC tables from the table screen shots on the next slides that you feel are most related to the three fields in CONDITION_OCCURRENCE.

Table Details: condition occurrence

Schema Details	Preview			
condition_occurrence	e_id	FLOAT	NULLABLE	int64
person_id		FLOAT	NULLABLE	int64
condition_concept_id	I	FLOAT	NULLABLE	int64
condition_start_date		STRING	NULLABLE	parse_date()
condition_start_datet	ime	STRING	NULLABLE	parse_datetime()
condition_end_date		STRING	NULLABLE	parse_date()
condition_end_dateti	me	STRING	NULLABLE	parse_datetime()
condition_type_conc	ept_id	FLOAT	NULLABLE	int64
stop_reason		STRING	NULLABLE	Describe this field
provider_id		FLOAT	NULLABLE	int64
visit_occurrence_id				
visit_occurrence_iu		FLOAT	NULLABLE	int64
visit_detail_id		FLOAT	NULLABLE	int64
	ue			
visit_detail_id		FLOAT	NULLABLE	int64
visit_detail_id condition_source_val	ncept_id	FLOAT	NULLABLE	int64 Describe this field
visit_detail_id condition_source_val condition_source_co	ncept_id	FLOAT STRING FLOAT	NULLABLE NULLABLE NULLABLE	int64 Describe this field int64

Table Details: ADMISSIONS Schema Details Preview



HAS_CHARTEVENTS_DATA

INTEGER

Table Details: CAREGIVERS

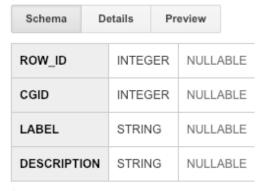


Table Details: CPTEVENTS

Schema Details	Preview	
ROW_ID	INTEGER	NU
SUBJECT_ID	INTEGER	NU
HADM_ID	INTEGER	NU
COSTCENTER	STRING	NU
CHARTDATE	DATETIME	NU
CPT_CD	STRING	NU
CPT_NUMBER	INTEGER	NU
CPT_SUFFIX	STRING	NU
TICKET_ID_SEQ	INTEGER	NU
SECTIONHEADER	STRING	NU
SUBSECTIONHEADER	STRING	NU
DESCRIPTION	STRING	NU

Table Details: D_CPT

Schema	Details	Pr	eview
ROW_ID			INTE
CATEGORY	1		INTE
SECTIONR	ANGE		STRII
SECTIONH	STRII		
SUBSECTION	ONRANGE		STRII
SUBSECTION	ONHEADER		STRII
CODESUFF	FIX		STRI
MINCODEI	NSUBSECTION	ON	INTE
MAXCODE	INSUBSECTI	ON	INTE

Table Details: D_ICD_PROCEDURES

Schema	ema Details Preview		eview	
ROW_ID	INTEG	ER	NULLABLE	Describe this
ICD9_CODE	STRIN	G	NULLABLE	Describe this
SHORT_TITLE	STRIN	STRING NULLABLE		Describe this
LONG_TITLE	STRIN	G	NULLABLE	Describe this

Table Details: D_ICD_DIAGNOSES

Schema D	etails	Pr	eview	
ROW_ID	INTEG	ER	NULLABLE	Describe t
ICD9_CODE	STRIN	G	NULLABLE	Describe t
SHORT_TITLE	STRIN	G	NULLABLE	Describe t
LONG_TITLE	STRIN	G	NULLABLE	Describe t

Use these screen captures (and next slide) to select one or more MIMIC tables that contain data for **OMOP CONDITION_OCCURRENCE table**

Table Details: ICUSTAYS

Schema	Details	s Preview		
ROW_ID		INTEGER		NL
SUBJECT_	ID	INT	EGER	NL
HADM_ID		INT	EGER	NL
ICUSTAY_II	o	INT	EGER	NL
DBSOURCE	E	STRING		NL
FIRST_CAREUNIT		STRING		NL
LAST_CAR	EUNIT	STRING		NL
FIRST_WAF	RDID	INTEGER		NL
LAST_WAR	DID	INTEGER		NL
INTIME		DATETIME		NL
OUTTIME		DA	TETIME	NL
LOS		FL	TAC	NL

Table Details: DIAGNOSES_ICD

Schema	D	Details	Р	review	
ROW_ID		INTEGE	R	NULLABLE	Describe tl
SUBJECT_	ID	INTEGE	R	NULLABLE	Describe tl
HADM_ID		INTEGE	R	NULLABLE	Describe tl
SEQ_NUM		INTEGE	R	NULLABLE	Describe tl
ICD9_CODI	E	STRING	9	NULLABLE	Describe ti

Table Details: DRGCODES

Schema	Details	Previe	w
ROW_ID	IN	ITEGER	NULLAB
SUBJECT_ID	IN	ITEGER	NULLAB
HADM_ID	IN	ITEGER	NULLAB
DRG_TYPE	S	TRING	NULLAB
DRG_CODE	s	TRING	NULLAB
DESCRIPTION	N S	TRING	NULLAB
DRG_SEVERI	TY IN	ITEGER	NULLAB
DRG_MORTA	LITY	ITEGER	NULLAB

Table Details: D LABITEMS

Schema [Details P	review
ROW_ID	INTEGER	NULLABLE
ITEMID	INTEGER	NULLABLE
LABEL	STRING	NULLABLE
FLUID	STRING	NULLABLE
CATEGORY	STRING	NULLABLE
LØINC_CODE	STRING	NULLABLE

Table Details: ICUSTAYS

Schema Details	S	Preview	
ROW_ID	INT	EGER	NL
SUBJECT_ID	INT	EGER	NL
HADM_ID	INT	EGER	NL
ICUSTAY_ID	INTEGER		NL
DBSOURCE	STRING		NL
FIRST_CAREUNIT	STRING		NL
LAST_CAREUNIT	STRING		NL
FIRST_WARDID	INT	EGER	NL
LAST_WARDID	INT	EGER	NL
INTIME	DA	ГЕТІМЕ	NL
OUTTIME	DATETIME N		NL
LOS	FLO	DAT	NL

Table Details: LABEVENTS

Schema	Details P	review
ROW_ID	INTEGER	NULLABLE
SUBJECT_ID	INTEGER	NULLABLE
HADM_ID	INTEGER	NULLABLE
ITEMID	INTEGER	NULLABLE
CHARTTIME	DATETIME	NULLABLE
VALUE	STRING	NULLABLE
VALUENUM	FLOAT	NULLABLE
VALUEUOM	STRING	NULLABLE
FLAG	STRING	NULLABLE

Use these screen captures (and previous slide) to select one or more MIMIC tables that contain data for OMOP CONDITION_OCCURRENCE table

Table Details: NOTEEVENTS

Schema	D	etails	Pre	view	
ROW_ID		INTEG	ER	NULL	ABLE
SUBJECT_	ID	INTEG	ER	NULL	ABLE
HADM_ID		INTEG	ER	NULL	.ABLE
CHARTDAT	CHARTDATE		DATETIME		.ABLE
CHARTTIME		DATETIME		NULL	ABLE
STORETIM	STORETIME		DATETIME		ABLE
CATEGORY	,	STRING		NULL	ABLE
DESCRIPTI	ON	STRIN	STRING		ABLE
CGID		INTEGER		NULL	ABLE
ISERROR		STRIN	IG	NULL	ABLE
TEXT		STRIN	IG	NULL	ABLE

Table Details: PATIENTS

Schema	Details		Prev	/iew		
ROW_ID	ROW_ID		INTEGER		NULLABLE	
SUBJECT_ID		INTEGER		NULLABLE		
GENDER		STRIN	1G	NUL	LABLE	
DOB		DATETIME		NUL	LABLE	
DOD		DATETIME		NUL	LABLE	
DOD_HOSP		DATE	TIME	NUL	LABLE	
DOD_SSN		DATETIME		NUL	LABLE	
EXPIRE_FLA	\G	INTEG	SER	NUL	LABLE	

Table Details: PRESCRIPTIONS

Preview

Details

Schema

ROW_ID	INTEGER	NULLABLE
SUBJECT_ID	INTEGER	NULLABLE
HADM_ID	INTEGER	NULLABLE
ICUSTAY_ID	INTEGER	NULLABLE
STARTDATE	DATETIME	NULLABLE
ENDDATE	DATETIME	NULLABLE
DRUG_TYPE	STRING	NULLABLE
DRUG	STRING	NULLABLE
DRUG_NAME_POE	STRING	NULLABLE
DRUG_NAME_GENERIC	STRING	NULLABLE
FORMULARY_DRUG_CD	STRING	NULLABLE
GSN	STRING	NULLABLE
NDC	STRING	NULLABLE
PROD_STRENGTH	STRING	NULLABLE
DOSE_VAL_RX	STRING	NULLABLE
DOSE_UNIT_RX	STRING	NULLABLE
FORM_VAL_DISP	STRING	NULLABLE
FORM_UNIT_DISP	STRING	NULLABLE
ROUTE	STRING	NULLABLE

Table Details: PROCEDURES_ICD

Schema	Details	Preview		
ROW_ID	INTEGE	R NULLA	BLE	Describe this
SUBJECT_ID	INTEGE	R NULLA	BLE	Describe this
HADM_ID	INTEGE	R NULLA	BLE	Describe this
SEQ_NUM	INTEGE	R NULLA	BLE	Describe this
ICD9_CODE	STRING	NULLA	BLE	Describe this

Table Details: TRANSFERS

Preview

Schema Details

Schema	Details		Preview		
ROW_ID		INT	EGER	Ν	IULLABLE
SUBJECT_	ID	INT	EGER	Ν	IULLABLE
HADM_ID		INT	EGER	Ν	IULLABLE
ICUSTAY_I	D	INT	EGER	Ν	IULLABLE
DBSOURC	E	STF	RING	Ν	IULLABLE
EVENTTYP	PE	STF	RING	Ν	IULLABLE
PREV_CAF	REUNIT	STF	RING	Ν	IULLABLE
CURR_CAI	REUNIT	STF	RING	Ν	IULLABLE
PREV_WAI	RDID	INT	EGER	Ν	IULLABLE
CURR_WA	RDID	INT	EGER	Ν	IULLABLE
INTIME		DAT	ГЕТІМЕ	Ν	IULLABLE
OUTTIME		DAT	ГЕТІМЕ	Ν	IULLABLE
LOS		FLC	DAT	Ν	IULLABLE

Step 1: Understand source/target data models



Table Details: D_ICD_DIAGNOSES

Schema	De	etails	Pre	eview		
ROW_ID		INTEG	ER	NULLABL	Е	Describe th
ICD9_CODE	•	STRING		NULLABLE		Describe th
SHORT_TIT	LE	STRIN	G	NULLABL	Е	Describe th
LONG_TITL	.E	STRIN	G	NULLABL	Е	Describe th

This one will be then joined to the target table so that ICD9 codes will be readable language.

Paste one or more MIMIC table(s) from the previous two slides that contain data for ETL into OMOP CONDITION_OCCURRENCE here!

Table Details: condition occurrence



Step 2: Profile source table or tables

Using the White Rabbit profiling data from the 100 patient MIMIC database provided in the Assessment to comment on the distribution of the SUBJECT_ID field from one of the MIMIC tables selected in Step 1

- DIAGNOSES_ICD
- ID#41976 had 266 entries in Diagnoses_ID then follows with a longtail with most patients having more than one diagnosis.
- ICD9_CODE 4019 (hyptertension), 42731 (atrial fibrillation), 5849(acute kidney failure) were the top 3 diagnoses.

Step 3: Create ETL mappings

Table Details: DIAGNOSES_ICD

Schema	Details	Preview	
ROW_ID	INTEGER	NULLABLE	Describe (
SUBJECT_ID	INTEGER	NULLABLE	Describe tl
HADM_ID	INTEGER	NULLABLE	Describe tl
SEQ_NUM	INTEGER	NULLABLE	Describe tl
ICD9_CODE	STRING	NULLABLE	Describe tl

person_id (int) - A foreign key identifier to the Person who is experiencing the condition.— This will be same as to SUBJECT_ID in MIMIC data as this describes unique key identifier to the person and it is integers.

visit_occurrence_id (int) - A foreign key to the visit in the VISIT_OCCURRENCE table during which the Condition was determined (diagnosed). – This is equivalent to HADM_ID in MIMIC data as this is the unique admission id and it is integers.

condition_source_value (Char (50)) -The source code for the Condition as it appears in the source data. -This will be exactly same as to MIMIC data (ICD9_code) as this is what describes the diagnosis (condition) experienced by the patient in MIMIC data.

Table Details: condition_occurrence

Table Details: condition_occurrence



Step 4: Write transformation code

```
WITH person1 as (select distinct mp.subject id as person id,
                 mp.subject id as person source value from mimic3 demo.PATIENTS mp),
  person2 as (select distinct p1.person id, p1.person source value
          ,mp.GENDER as gender source value
          .CASE mp.GENDER
          WHEN 'F' then 8532
          WHEN 'M' then 8507
          ELSE 0 END as gender concept id
        from person1 p1 join mimic3 demo.PATIENTS mp on p1.person id = mp.subject id).
  person3 as (select distinct p2.person id, p2.person source value
                ,p2.gender source_value, p2.gender_concept_id
                ,extract(year from mp.dob) as year of birth
                ,extract(month from mp.dob) as month of birth
                ,extract(day from mp.dob) as day of birth
                ,dob as birth datetime
         from person2 p2 join mimic3 demo.PATIENTS mp on p2.person id = mp.subject id),
  person4 as (select distinct p3.person id, p3.person source value
                ,p3.gender source value,p3.gender concept id
             ,p3.year of birth, p3.month of birth, p3.day of birth, p3.birth datetime
            ,ma.ethnicity as race source value
             ,case ma.ethnicity
                                                           when 'WHITE' then 8527
                                                           when 'BLACK/AFRICAN AMERICAN' then 8516
                                                           when 'ASIAN' then 8515
                                                           when 'HISPANIC/LATINO-PUERTO RICAN' then 44814653
                                                           when 'HISPANIC OR LATINO' then 44814653
                                                           when 'UNKNOWN/NOT SPECIFIED' then 44814653
                                                           when 'OTHER' then 44814653
                                                           when 'AMERICAN INDIAN/ALASKA NATIVE FEDERALLY RECOGNIZED TRIBE' then 8657
                                                           when 'UNABLE TO OBTAIN' then 44814650
                                                           else 0 end as race concept id
           from person3 p3 join mimic3_demo.ADMISSIONS ma on p3.person_id = ma.subject_id),
  person as (select distinct p4.person id, p4.person source value
            ,p4.gender source value,p4.gender concept id
            ,p4.year of birth, p4.month of birth, p4.day of birth, p4.birth datetime
            ,p4.race source value, p4.race concept id
            ,ethnicity as ethnicity_source_value
            ,case ma.ethnicity
                                                           when 'WHITE' then 38003564
                                                           when 'BLACK/AFRICAN AMERICAN' then 38003564
                                                           when 'ASIAN' then 38003564
                                                           when 'HISPANIC/LATINO-PUERTO RICAN' then 38003563
                                                           when 'HISPANIC OR LATINO' then 38003563
                                                           when 'UNKNOWN/NOT SPECIFIED' then 44814653
                                                           when 'OTHER' then 38003564
                                                           when 'AMERICAN INDIAN/ALASKA NATIVE FEDERALLY RECOGNIZED TRIBE' then 38003564
                                                           when 'UNABLE TO OBTAIN' then 44814650
                                                           else 0 end as ethnicity concept id
```

from person4 p4 join mimic3 demo.ADMISSIONS ma on p4.person id = ma.subject id)

select * from person

Paste the SQL statements that transform data from one or more MIMIC tables into the three OMOP CONDITION_OCCURRENCE fields (patient-id, visit_occurrence_id, condition_source_value) into the Coursera Submission Site

Transformation code shown here is from the Course 2 videos showing transformation of MIMIC PATIENTS to OMOP PERSON

Step 5: Execute transformation code

Execute the ETL code from Step 4 but do not submit the output table.

Use the output table for Step 6.

There is no submission for this Step.

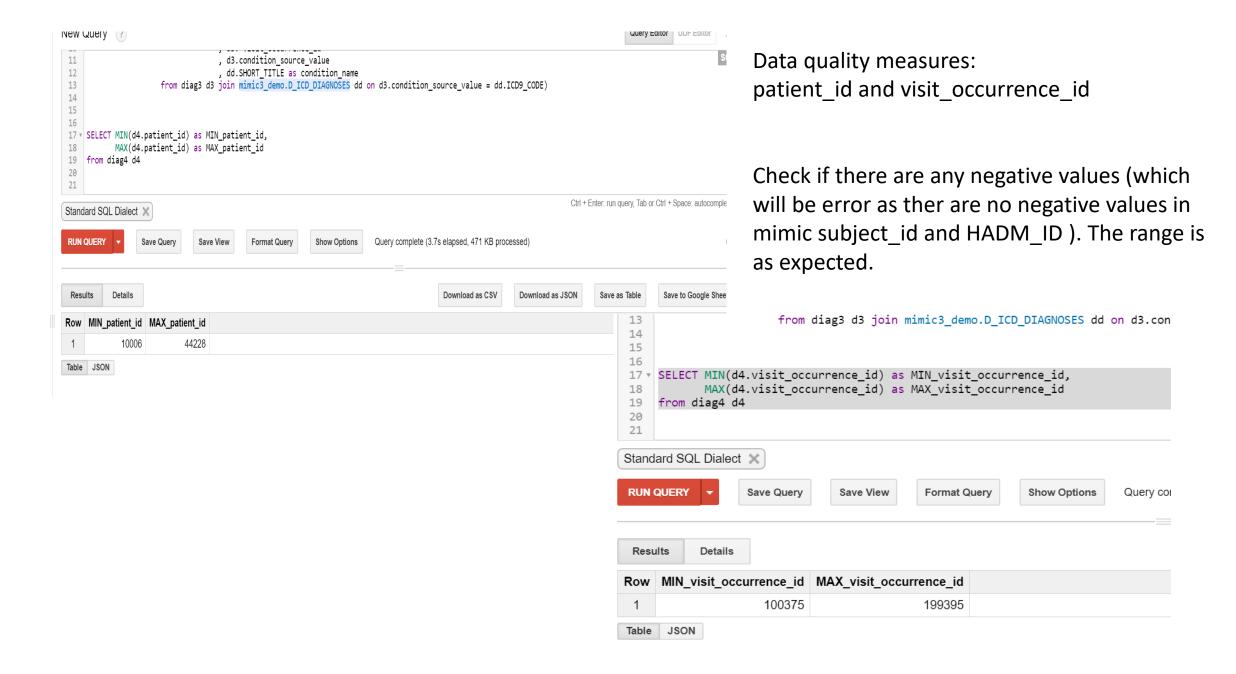
Step 6: Perform data quality assessment

Define, implement, execute one or more data quality measures. Submit final DQ measure and an explanation why you created your measure(s).

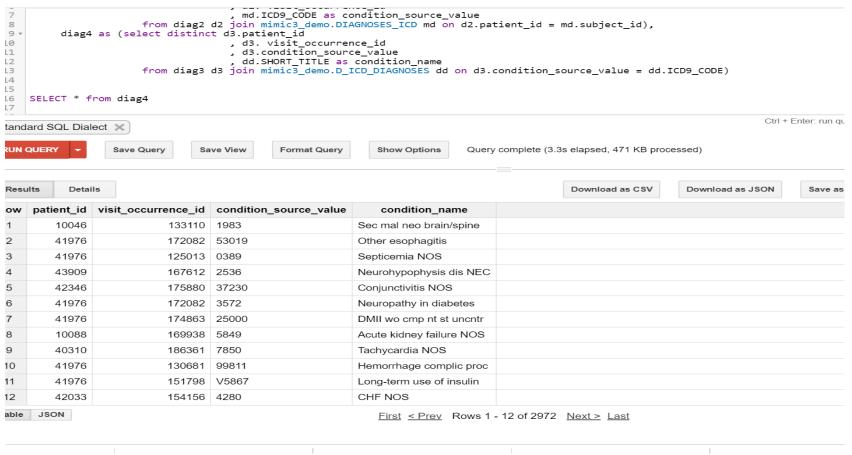
	Row	MIN_YOB	MAX_YOB	AVG_YOB
-	1	1844	2181	2061.569999999999

Example output table from the Course 2 videos showing transformation of MIMIC PATIENTS to OMOP PERSON

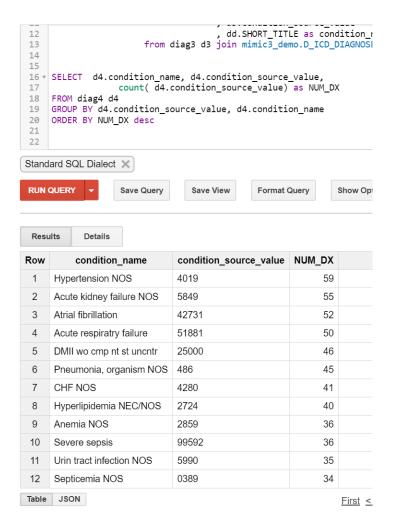
OK to paste link to a GoogleDoc instead.

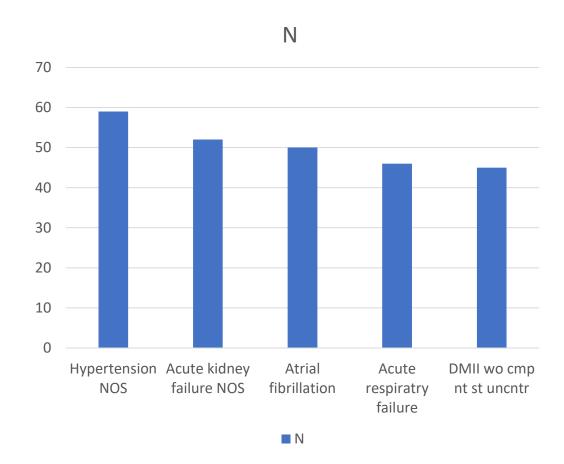


Check if the ICD9_codes from condition_source_value still matches with the labels from MIMICS.



Checking the Top 5 Diagnoses





Step 7: Package documentation

 Congratulations! The materials in the previous slides constitute a complete ETL package.

There is no submission for this Step.