

Part I, 1 (all SQL queries can be found in hw2.ipynb)

- Assumptions:
 - treat PRESCRIPTIONS.DRUG as the drug type
 - Use $\text{DOSE_VAL_RX} \times \text{DAYS}$ or count prescriptions as a proxy

ethnicity varchar	drug varchar	total_prescriptions int64
AMERICAN INDIAN/ALASKA NATIVE FEDERALLY RECOGNIZED TRIBE	5% Dextrose	27
ASIAN	D5W	27
BLACK/AFRICAN AMERICAN	Insulin	38
HISPANIC OR LATINO	5% Dextrose	28
HISPANIC/LATINO – PUERTO RICAN	0.9% Sodium Chloride	86
OTHER	NS	11
UNABLE TO OBTAIN	0.9% Sodium Chloride	28
UNKNOWN/NOT SPECIFIED	D5W	37
WHITE	Potassium Chloride	381

- Explanation of the query:
 - JOIN admissions to access ETHNICITY
 - GROUP BY ethnicity, drug to get counts
 - ROW_NUMBER... PARTITION BY ethnicity finds the top drug per group
 - QUALIFY filters to only the top 1 per ethnicity
- Summary:
 - Multiple ethnicities' top prescription is 5% Dextrose, including American Indian/Alaska native federally recognized tribe, Asian, Hispanic or Latino, and some ethnicities that were not collected.
 - 0.9% Sodium Chloride is also a top prescription for many ethnicities, including Hispanic/Latino - Puerto Rican and some ethnicities that were not collected.
 - Top prescription for white is potassium chloride, and top prescription for black/African American is insulin.
 - Ethnic groups differ in their top drug types for different clinical needs and healthcare access (IV fluid, comorbidity, etc.)

Part I, 2

- Explanation of the query:
 - Calculate patient age at admission
 - Categorize patients into age groups
 - Join with procedures, count the frequency
 - Use row number to obtain the top 3 per group
- Summary:
 - age group 20-49's and age group 50-79's top procedures are Venous cath NEC, Entral infus nutrit sub, and Insert endotracheal tube
 - age group ≤ 19 's top procedures are Venous cath NEC, Skin closure NEC, Spinal tap

- age group >80's top procedures are Venous cath NEC, Packed cell transfusion, Insert endotracheal tube
- venous cath NEC is the most popular procedure because it might be a pre-step for delivering fluids and medication by inserting a catheter to a vein

age_group varchar	procedure_name varchar	procedure_count int64	rank int64
20-49	Venous cath NEC	11	1
20-49	Entral infus nutrit sub	11	2
20-49	Cont inv mec ven 96+ hrs	9	3
50-79	Venous cath NEC	185	1
50-79	Entral infus nutrit sub	170	2
50-79	Insert endotracheal tube	51	3
<=19	Venous cath NEC	3	1
<=19	Skin closure NEC	2	2
<=19	Vertebral fx repair	1	3
>=80	Venous cath NEC	22	1
>=80	Packed cell transfusion	16	2
>=80	Insert endotracheal tube	9	3
12 rows		4 columns	

Part I, 3

- Explanation:
 - Calculate ICU length of stay in days
 - Join with gender and ethnicity
 - Compute average for each subgroup
- Summary:
 - Female patients overall had longer ICU stays than males in several ethnic categories. Notably Black/African American females had the highest average ICU stay at 11.2 days. Hispanic or Latino females followed at 7.5 days, with White females at 5.2 days.
 - For males, the highest average was for American Indian/Alaska Native males at 11.3 days. Males with Unable to Obtain ethnicity info had an even higher average at 13.4 days, suggesting this group may represent emergency cases where background info was unavailable. Other groups such as Asian males also had

relatively high ICU stays (7.1 days), whereas White males and Black males had shorter stays (3.2 and 3.0 days, respectively).

gender varchar	ethnicity varchar	avg_icu_days double
F	ASIAN	0.7
F	BLACK/AFRICAN AMERICAN	11.2
F	HISPANIC OR LATINO	7.5
F	OTHER	1.3
F	UNKNOWN/NOT SPECIFIED	5.5
F	WHITE	5.2
M	AMERICAN INDIAN/ALASKA NATIVE FEDERALLY RECOGNIZED TRIBE	11.3
M	ASIAN	7.1
M	BLACK/AFRICAN AMERICAN	3.0
M	HISPANIC/LATINO – PUERTO RICAN	3.2
M	OTHER	0.1
M	UNABLE TO OBTAIN	13.4
M	UNKNOWN/NOT SPECIFIED	2.1
M	WHITE	3.2
14 rows		3 columns

Part II

No copies of the AWS credentials file is stored on any publicly accessible location, nor is the file in any way shared with anyone outside of DATA_ENG 300 (Spring 2025).

- Linda Liu

- Code for data table creation, data upload, and data extraction/analysis can be found in hw2.ipynb

Q1 result

AMERICAN INDIAN/ALASKA NATIVE FEDERALLY RECOGNIZED TRIBE: 5% Dextrose (27 prescriptions)
WHITE: Potassium Chloride (381 prescriptions)
HISPANIC/LATINO – PUERTO RICAN: 0.9% Sodium Chloride (86 prescriptions)
BLACK/AFRICAN AMERICAN: Insulin (38 prescriptions)
UNKNOWN/NOT SPECIFIED: D5W (37 prescriptions)
HISPANIC OR LATINO: 5% Dextrose (28 prescriptions)
OTHER: NS (11 prescriptions)
ASIAN: D5W (27 prescriptions)
UNABLE TO OBTAIN: 0.9% Sodium Chloride (28 prescriptions)

Q2 result

Age group 50–79:

Venous catheterization, not elsewhere classified – 26 procedures
Enteral infusion of concentrated nutritional substances – 22 procedures
Transfusion of packed cells – 13 procedures

Age group 20–49:

Venous catheterization, not elsewhere classified – 9 procedures
Enteral infusion of concentrated nutritional substances – 7 procedures
Percutaneous abdominal drainage – 6 procedures

Age group 80+:

Venous catheterization, not elsewhere classified – 19 procedures
Transfusion of packed cells – 13 procedures
Insertion of endotracheal tube – 8 procedures

Age group 0–19:

Venous catheterization, not elsewhere classified – 2 procedures
Other skeletal traction – 1 procedures
Percutaneous [endoscopic] gastrostomy [PEG] – 1 procedures

Q3 result

Average ICU stay by gender:

M: 3.38 days

F: 5.69 days

Average ICU stay by ethnicity:

WHITE: 4.22 days

UNKNOWN/NOT SPECIFIED: 5.19 days

OTHER: 0.93 days

BLACK/AFRICAN AMERICAN: 8.28 days

AMERICAN INDIAN/ALASKA NATIVE FEDERALLY RECOGNIZED TRIBE: 11.34 days

HISPANIC/LATINO – PUERTO RICAN: 3.24 days

HISPANIC OR LATINO: 6.48 days

ASIAN: 3.89 days

UNABLE TO OBTAIN: 13.36 days

Verification:

- Q1, 2, and 3 results are similar to those in Part I, indicating correct extraction and analysis in Cassandra.