Q1: Create a summary of type of drugs and their total amount used by ethnicity. Report the top usage in each ethnicity group. You may have to make certain assumptions in calculating their total amount.

SQL Query

```
SELECT ethnicity, drug, drug_occurences
FROM (
    SELECT
        a.ethnicity,
        p.drug,
        COUNT(*) AS drug_occurences,
        ROW_NUMBER() OVER (PARTITION BY a.ethnicity ORDER BY COUNT(*) DESC) AS rank
    FROM PRESCRIPTIONS p
    JOIN ADMISSIONS a ON p.hadm_id = a.hadm_id
    GROUP BY a.ethnicity, p.drug
) sub
WHERE rank = 1
ORDER BY ethnicity;
```

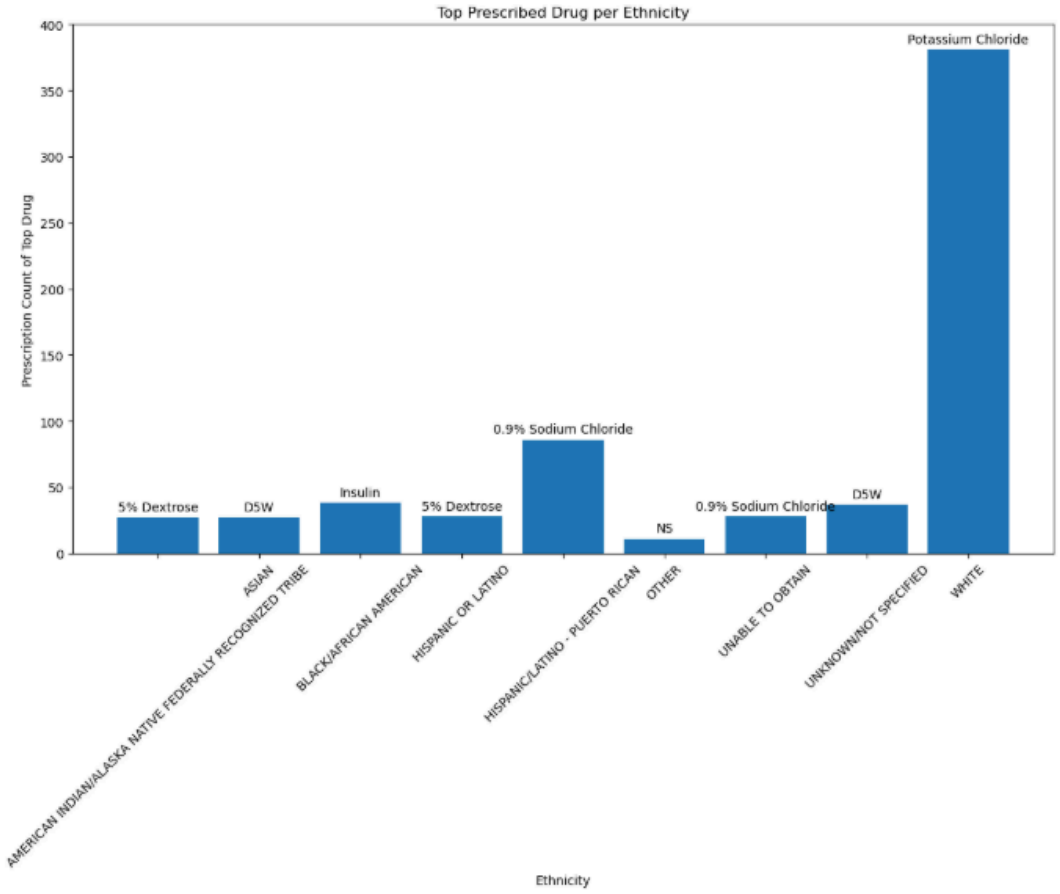
Query Explanation

In my query, I joined the PRESCRIPTIONS and ADMISSIONS tables to obtain information on the ethnicity of the individual of each prescription. I joined on HADM_ID because I wanted to obtain information based on admission-level data. From here I counted the number of times each drug was prescribed for each ethnicity and gave rankings of the drugs within each ethnicity based on the number of occurrences for each drug. Finally, I filtered the query to only show the results with rank 1 and ordered the results alphabetically by ethnicity.

Summary of Findings

I found that top used drug is Potassium Chloride for Whites, 5% Dextrose for American Indian/Alaska Native Federally Recognized Tribes and Hispanic/Latino, D5W for Asians, Insulin for Black/African American, and 0.9% Sodium Chloride for Hispanic/Latino-Puerto Rican. Some other commonly used drugs for unspecified individuals are 0.9% Sodium Chloride and D5W. The total counts for the usage of the respective top drug for each ethnicity was greatest for White with a count of 381 and lowest for American Indian/Alaska Native Federally Recognized Tribes and Asians with a count of 27, excluding the individuals with unknown ethnicities.

ethnicity varchar	drug varchar	drug_occurences 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


Top Prescribed Drug per Ethnicity
Prescription Count of Top Drug
Ethnicity
400
350
300
250
200
150
100
50
0
5% Dextrose
D5W
Insulin
5% Dextrose
0.9% Sodium Chloride
NS
0.9% Sodium Chloride
D5W
Potassium Chloride
AMERICAN INDIAN/ALASKA NATIVE FEDERALLY RECOGNIZED TRIBE
ASIAN
BLACK/AFRICAN AMERICAN
HISPANIC OR LATINO
HISPANIC/LATINO - PUERTO RICAN
OTHER
UNABLE TO OBTAIN
UNKNOWN/NOT SPECIFIED
WHITE

Q2: Create a summary of procedures performed on patients by age groups (<=19, 20-49, 50-79, >80). Report the top three procedures, along with the name of the procedures, performed in each age group.

SQL Query

```
WITH ranked_procedures AS (
    SELECT
        CASE
            WHEN age <= 19 THEN '0-19'
            WHEN age BETWEEN 20 AND 49 THEN '20-49'
            WHEN age BETWEEN 50 AND 79 THEN '50-79'
            ELSE '80+'
        END AS age_group,
        d.short_title AS procedure_name,
        COUNT(*) AS procedure_count,
        ROW_NUMBER() OVER (
            PARTITION BY
                CASE
                    WHEN age <= 19 THEN '0-19'
                    WHEN age BETWEEN 20 AND 49 THEN '20-49'
                    WHEN age BETWEEN 50 AND 79 THEN '50-79'
                    ELSE '80+'
                END
            ORDER BY COUNT(*) DESC
        ) AS rank_in_group
    FROM (
        SELECT
            p.subject_id, p.hadm_id, p.icd9_code, p.seq_num,
            DATEDIFF('year', pat.dob, a.admittime) AS age
        FROM PROCEDURES_ICD p
        JOIN ADMISSIONS a ON p.hadm_id = a.hadm_id
        JOIN PATIENTS pat ON p.subject_id = pat.subject_id
    ) AS procedures_with_age
    JOIN D_ICD_PROCEDURES d
    ON procedures_with_age.icd9_code = d.icd9_code
    GROUP BY age_group, procedure_name
)
SELECT age_group, procedure_name, procedure_count
FROM ranked_procedures
WHERE rank_in_group <= 3
ORDER BY age_group, rank_in_group;
```

Query Explanation

In my query, I joined the PROCEDURES_ICD table with the ADMISSIONS and PATIENTS table on the icd9_code, subject_id, or hadm_code to pull all the necessary information. To calculate the age of each patient at the time of their admission to the hospital I used the difference between their date of birth and the admission date. From the joined tables, the information I selected to show were the age groups, procedure, and count of procedures for that age group. To create the age groups, I used a CASE statement to classify the patients into the four age groups of <=19, 20-49, 50-79, and >80. From here, I grouped the data by the age group and

procedure name and then ordered the data by age group and procedure count in descending order. I also ranked all the rows of data within each age group and only returned the top three for each group..

Summary of Findings
The table below summarizes my findings on the top three procedures for each age group along with the count. For the age group from 0 to 19, the top three procedures are Venous cath NEC, Vertebral fx repair, Interruption vena cava. For the age group from 20 to 49, the top three procedures are Venous cath NEC, Enthral infus nutrit sub, and Cont inv mec ven 96+ hrs. For the age group from 50-70, the top three procedures are Venous cath NEC, Entral infus nutrit sub, and packed cell transfusion. For the 80+ age group, the top three procedures are Venous cath NEC, packed cell transfution, and insert endotracheal tube.

We can see that venous cath NEC is the top procedure for all four age groups. Entral infus nutrit sub is the second top procedure for the 20-49 and 50-79 age groups and Packed cell transfusion is one of the top three procedures for the 50-79 and 80+ age groups. The other top procedures varied across the age groups.

age_group varchar	procedure_name varchar	procedure_count int64
0-19	Venous cath NEC	2
0-19	Vertebral fx repair	1
0-19	Interruption vena cava	1
20-49	Venous cath NEC	9
20-49	Entral infus nutrit sub	7
20-49	Insert endotracheal tube	6
50-79	Venous cath NEC	25
50-79	Entral infus nutrit sub	22
50-79	Packed cell transfusion	13
80+	Venous cath NEC	20
80+	Packed cell transfusion	13
80+	Insert endotracheal tube	8

Q3: How long do patients stay in the ICU? Is there a difference in the ICU length of stay among gender or ethnicity?

SQL Query

```
SELECT
    a.ethnicity,
    p.gender,
    AVG(i.los) AS avg_los,
FROM ICUSTAYS i
JOIN ADMISSIONS a ON i.hadm_id = a.hadm_id
JOIN PATIENTS p ON i.subject_id = p.subject_id
WHERE i.los IS NOT NULL
GROUP BY a.ethnicity, p.gender
ORDER BY a.ethnicity, p.gender;
```

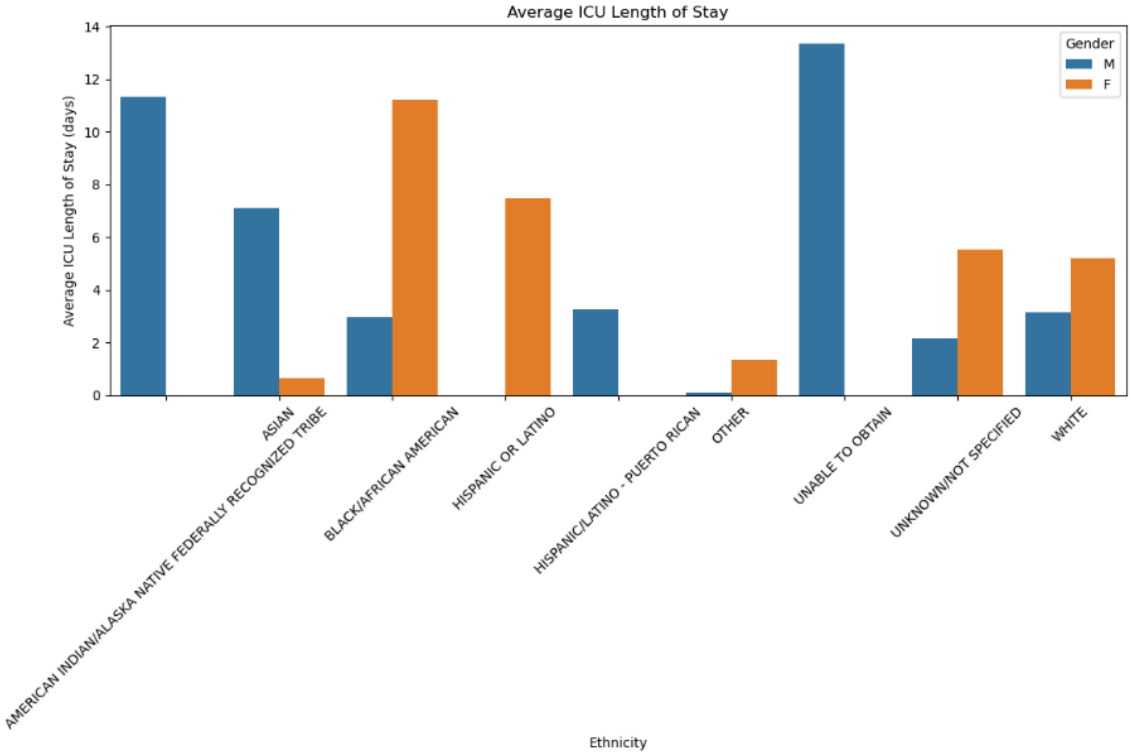
Query Explanation

In my query, I joined the ICUSTAYS and ADMISSIONS tables through hadm_id then joined ICUSTAYS with PATIENTS through subject_id. From here, I was able to select the ethnicity, gender, and length of stay information. Because I needed to aggregate the length of stay (los) information, I decided to show the average length of stay for each ethnicity and gender group. From here, I ordered the data by ethnicity then by gender.

Summary of Findings

The group with the highest average length of stay was male American Indian/Alaska native federally recognized tribes of around 11.33 days followed by female Black/African Americans (11.20 days) and female Hispanic/Latino (7.46 days). Overall, there is a pretty significant difference in the average length of stay across ethnicities and genders. The full results for the average length of stay in the ICU per ethnicity and gender can be seen below.

ethnicity varchar	gender varchar	avg_los double
AMERICAN INDIAN/ALASKA NATIVE FEDERALLY RECOGNIZED TRIBE	M	11.33715
ASIAN	F	0.6628
ASIAN	M	7.1173
BLACK/AFRICAN AMERICAN	F	11.201225
BLACK/AFRICAN AMERICAN	M	2.9772666666666665
HISPANIC OR LATINO	F	7.459633333333334
HISPANIC/LATINO - PUERTO RICAN	M	3.2430666666666665
OTHER	F	1.33615
OTHER	M	0.1059
UNABLE TO OBTAIN	M	13.357
UNKNOWN/NOT SPECIFIED	F	5.543011111111111
UNKNOWN/NOT SPECIFIED	M	2.1454500000000003
WHITE	F	5.195874999999999
WHITE	M	3.153883333333333


Average ICU Length of Stay
Average ICU Length of Stay (days)
Gender
M
F
0
2
4
6
8
10
12
14
AMERICAN INDIAN/ALASKA NATIVE FEDERALLY RECOGNIZED TRIBE
ASIAN
BLACK/AFRICAN AMERICAN
HISPANIC OR LATINO
HISPANIC/LATINO - PUERTO RICAN
OTHER
UNABLE TO OBTAIN
UNKNOWN/NOT SPECIFIED
WHITE
Ethnicity