SQL codes we used to extract data from Mimic-iii

Query to extract static variables

```
WITH trauma_patients AS (
    SELECT DISTINCT
        hadm_id,
        subject_id
    FROM
        `physionet-data.mimiciii_clinical.diagnoses_icd`
    WHERE
        icd9_code BETWEEN '800' AND '959'  -- Trauma ICD-9 code range
)

SELECT
    tp.subject_id,
    tp.hadm_id,
    p.gender AS sex,
    MAX(CASE WHEN ce.itemid = 226707 THEN ce.valuenum END) AS height,  -- Item ID
for height
    MAX(CASE WHEN ce.itemid = 763 THEN ce.valuenum END) AS weight      -- Item ID
for weight
FROM
    trauma_patients tp
JOIN `physionet-data.mimiciii_clinical.patients` p ON tp.subject_id = p.subject_id
LEFT JOIN `physionet-data.mimiciii_clinical.chartevents` ce ON tp.subject_id =
ce.subject_id
GROUP BY
    tp.subject_id, tp.hadm_id, p.gender;
```

Query to extract dynamic variables Group 1

```
WITH trauma_patients AS (
    SELECT DISTINCT
        hadm_id,
        subject_id
    FROM
        `physionet-data.mimiciii_clinical.diagnoses_icd`
    WHERE
        icd9_code BETWEEN '800' AND '959'  -- Trauma ICD-9 code range
)

SELECT
    tp.subject_id,
    tp.hadm_id,
    MAX(CASE WHEN ce.itemid = 220045 THEN ce.valuenum END) AS heart_rate,
-- Heart Rate
    MAX(CASE WHEN ce.itemid = 220179 THEN ce.valuenum END) AS systolic_bp,
```

```
-- Systolic BP
    MAX(CASE WHEN ce.itemid = 220180 THEN ce.valuenum END) AS diastolic_bp,
-- Diastolic BP
    MAX(CASE WHEN ce.itemid = 220181 THEN ce.valuenum END) AS mean_bp,
-- Mean BP
    MAX(CASE WHEN ce.itemid = 220210 THEN ce.valuenum END) AS respiratory_rate,
-- Respiratory Rate
    MAX(CASE WHEN ce.itemid = 223761 THEN ce.valuenum END) AS temperature,
-- Temperature
    MAX(CASE WHEN ce.itemid = 220277 THEN ce.valuenum END) AS oxygen_saturation,
-- Oxygen Saturation
    MAX(CASE WHEN ce.itemid = 220621 THEN ce.valuenum END) AS glucose,
-- Glucose Levels
    MAX(CASE WHEN ce.itemid = 198 THEN ce.valuenum END) AS gcs
-- Glasgow Coma Score
FROM
    trauma_patients tp
LEFT JOIN `physionet-data.mimiciii_clinical.chartevents` ce ON tp.hadm_id =
ce.hadm_id
GROUP BY
    tp.subject_id, tp.hadm_id;
```

Query to extract dynamic variables Group 2

```
WITH trauma_patients AS (
    SELECT DISTINCT
        hadm_id,
        subject_id
    FROM
        `physionet-data.mimiciii_clinical.diagnoses_icd`
    WHERE
        icd9_code BETWEEN '800' AND '959'  -- Trauma ICD-9 code range
)

SELECT
    tp.subject_id,
    tp.hadm_id,
    MAX(CASE WHEN le.itemid = 50868 THEN le.valuenum END) AS anion_gap,
-- Anion Gap
    MAX(CASE WHEN le.itemid = 50862 THEN le.valuenum END) AS albumin,
-- Albumin Levels
    MAX(CASE WHEN le.itemid = 50882 THEN le.valuenum END) AS bicarbonate,
-- Bicarbonate Levels
    MAX(CASE WHEN le.itemid = 50885 THEN le.valuenum END) AS bilirubin,
-- Bilirubin Levels
    MAX(CASE WHEN le.itemid = 50893 THEN le.valuenum END) AS calcium,
-- Calcium Levels
    MAX(CASE WHEN le.itemid = 50912 THEN le.valuenum END) AS creatinine,
-- Creatinine Levels
    MAX(CASE WHEN le.itemid = 50902 THEN le.valuenum END) AS chloride,
-- Chloride Levels
```

```
    MAX(CASE WHEN le.itemid = 50813 THEN le.valuenum END) AS lactate,
-- Lactate Levels
    MAX(CASE WHEN le.itemid = 51265 THEN le.valuenum END) AS platelet,
-- Platelet Levels
    MAX(CASE WHEN le.itemid = 50971 THEN le.valuenum END) AS potassium,
-- Potassium Levels
    MAX(CASE WHEN le.itemid = 50983 THEN le.valuenum END) AS sodium,
-- Sodium Levels
    MAX(CASE WHEN le.itemid = 51222 THEN le.valuenum END) AS prothrombin_time,
-- Prothrombin Time
    MAX(CASE WHEN le.itemid = 51275 THEN le.valuenum END) AS inr,
-- International Normalized Ratio (INR)
    MAX(CASE WHEN le.itemid = 51221 THEN le.valuenum END) AS hematocrit,
-- Hematocrit
    MAX(CASE WHEN le.itemid = 51222 THEN le.valuenum END) AS hemoglobin,
-- Hemoglobin
    MAX(CASE WHEN le.itemid = 51006 THEN le.valuenum END) AS bun,
-- Blood Urea Nitrogen (BUN)
    MAX(CASE WHEN le.itemid = 51301 THEN le.valuenum END) AS wbc
-- White Blood Cell Count
FROM
    trauma_patients tp
LEFT JOIN `physionet-data.mimiciii_clinical.labevents` le ON tp.hadm_id =
le.hadm_id
GROUP BY
    tp.subject_id, tp.hadm_id;
```

Code for merging static and dynamic variables

```
# Load the CSV files
static_df = pd.read_csv('path/to/static_variables.csv')
dynamic_group1_df = pd.read_csv('path/to/dynamic_variables_group1.csv')
dynamic_group2_df = pd.read_csv('path/to/dynamic_variables_group2.csv')

# Merge the datasets on subject_id and hadm_id
merged_dynamic_df = pd.merge(dynamic_group1_df, dynamic_group2_df, on=
['subject_id', 'hadm_id'], how='inner')
final_df = pd.merge(static_df, merged_dynamic_df, on=['subject_id', 'hadm_id'],
how='inner')

# Save the merged dataset as a new CSV
final_df.to_csv('path/to/final_merged_variables.csv', index=False)

# Display the merged dataset
final_df.head()
```

Query to extract Dob

```
WITH trauma_patients AS (
    SELECT DISTINCT
        subject_id
    FROM
        `physionet-data.mimiciii_clinical.diagnoses_icd`
    WHERE
        icd9_code BETWEEN '800' AND '959'  -- Trauma ICD-9 code range
)

SELECT
    tp.subject_id,
    p.dob  -- Date of birth
FROM
    trauma_patients tp
LEFT JOIN `physionet-data.mimiciii_clinical.patients` p ON tp.subject_id =
p.subject_id;
```

Merging DOB with current dataset

```
# Load your existing merged dataset (static and dynamic variables)
merged_df = pd.read_csv('path/to/your/existing_merged_dataset.csv')

# Load the dob data
dob_df = pd.read_csv('path/to/dob_data.csv')

# Merge the datasets on 'subject_id'
final_merged_df = pd.merge(merged_df, dob_df, on='subject_id', how='left')

# Save the final merged dataset with dob
final_merged_df.to_csv('path/to/final_merged_with_dob.csv', index=False)

# Display the first few rows of the final merged dataset
print(final_merged_df.head())
```

Query to extract admittime

```
SELECT subject_id, hadm_id, admittime
FROM `physionet-data.mimiciii_clinical.admissions`
WHERE subject_id IN (
    SELECT DISTINCT subject_id
    FROM `your_project_name.your_dataset_name.your_existing_table` -- replace with
your dataset name and table
)
```

Code for merging admittime with current dataset

```
import pandas as pd

# Load the existing dataset
final_merged_df = pd.read_csv('/content/final_merged_variables_new.csv')

# Load the extracted admittime data
admit_time_df = pd.read_csv('/content/admit_time_data.csv')  # Update with the
correct path if needed

# Merge the datasets on 'subject_id' and 'hadm_id'
final_merged_with_admittime_df = pd.merge(final_merged_df, admit_time_df, on=
['subject_id', 'hadm_id'], how='left')

# Check the merged dataset
print(final_merged_with_admittime_df.head())

# Save the merged dataset with admittime
final_merged_with_admittime_df.to_csv('/content/final_merged_with_admittime.csv',
index=False)
```

Query to extract dnr patients

```
SELECT hadm_id
FROM physionet-data.mimiciii_notes.noteevents
WHERE LOWER(text) LIKE '%do not resuscitate%' OR LOWER(text) LIKE '%dnr%'
```