2/19/24:

- uploaded Fitbit Fitness Tracker Data dataset to BigQuery
    - **dailyActivities_merged.csv**
        - **schema**: Id (INTEGER), ActivityDate (DATE), TotalSteps (INTEGER), TotalDistance (FLOAT), TrackerDistance (FLOAT), LoggedActivitiesDistance (FLOAT), VeryActiveDistance (FLOAT), ModeratelyActiveDistance (FLOAT), LightActiveDistance (FLOAT), SedentaryActiveDistance (FLOAT), VeryActiveMinutes (INTEGER), FairlyActiveMinutes (INTEGER), LightlyActiveMinutes (INTEGER), SedentaryMinutes (INTEGER), Calories (INTEGER)
        - changed ActivityDate to day_of_week (Monday, Tuesday, etc.) using FORMAT_DATE(); summed all minute schema and saved them as total_VAM, total_FAM, total_LAM, total_SM → this table is now dailyActivity_clean
        - added all active minutes (VAM, FAM, LAM) and saved them as total_active_minutes, ordered descending → this table is now dailyActivity_total
        - verified no entries were lost by querying sum(VeryActiveMinutes) from dailyActivities_merged and sum(total_VAM) from dailyActivity_clean
    - **dailyCalories_merged.csv**
        - **schema**: Id (INTEGER), ActivityDay (DATE), Calories (INTEGER)
        - changed ActivityDay to day_of_week (Monday, Tuesday, etc.) using FORMAT_DATE(); summed Calories and saved it as total_calories → this table is totalCalories_by_day
    - **hourlyCalories_merged.csv**
        - **schema**: Id (NUMERIC), ActivityHour (STRING), Calories (NUMERIC)
            - schema were not automatically detected, ActivityHour is meant to be of type DATETIME but had to be input as STRING: further cleaning had to be done to isolate hourly times from date for analysis
        - took time portion of ActivityHour string and saved it as ActivityTime using SUBSTRING() and removed leading and trailing spaces using TRIM(); summed Calories as total_calories_by_hour → this table is hourlyCalories_clean
        - summed total_calories_by_hour as agg_total_calories; grouped by ActivityTime; ordered by agg_total_calories descending → this table is hourlyCalories_agg
        - verified no entries were lost in cleaning process by querying sum(Calories) from hourlyCalories_merged and sum(agg_total_calories)

from hourlyCalories_agg

- uploaded Fitness Consumer Survey Data dataset to BigQuery
    - **survey 605.csv**

2/20/24

- uploaded Fitbit Fitness Tracker Data to BigQuery
    - **sleepDay_merged.csv**
        - **schema:** Id (INTEGER), SleepDay (STRING), TotalSleepRecords (INTEGER), TotalMinutesAsleep (INTEGER), TotalTimeInBed (INTEGER)
            - schema again not automatically detected, SleepDay is meant to be of type DATETIME but was input as STRING
        - summed TotalMinutesAsleep as agg_sleep_minutes, summed TotalTimeInBed as agg_time_bed, divided two sums as sleep_proportion, grouped by Id → saved as sleepDay_agg
        - changed SleepDay from STRING type into DATE using FORMAT_DATE(), PARSE_DATE(), and SUBSTRING()
- queried Id, sum(Calories) as total_calories from **dailyCalories_merged.csv**, grouped by Id → saved as totalCalories_by_user
- joined sleepDay_agg with totalCalories_by_user → saved as sleep_calories
- queried Id, sleep_proportion, total_calories from sleep_calories → saved as SleepVsCalories
- Joined SleepVsCalories with dailyActivity_by_user → saved as SleepVsActivity_Minutes
- queried Id, sleep_proportion, total_VAM+total_LAM+total_FAM as total_active_minutes, total_SM, and total_calories → saved as SleepVsActivity

2/21/24

- further cleaned **hourlyCalories_merged.csv**
    - split ActivityHour into ActivityDate and ActivityHour using TRIM(SUBSTRING())
    - converted ActivityDate and ActivityHour from STRING data type to DATE and TIME respectively using PARSE_DATE() and PARSE_TIME() → saved as hourlyCalories_clean_time
    - changed ActivityDate to day_of_week using FORMAT_DATE(), summed Calories as total_calories_by_hour → saved as totalCalories_by_dayHour
- uploaded Fitbit Fitness Tracker Data to BigQuery
    - **minuteSleep.csv**

- **schema:** Id (INTEGER), date (STRING), value (INTEGER), logId (INTEGER)
    - schema not automatically detected, date is meant to be of type TIMESTAMP but was input as STRING
- converted date to TIMESTAMP using PARSE_TIMESTAMP()