```
/*

Airline Passenger Satisfaction Data Cleaning and Exploratory Data Analysis with SQL Queries

Skills Use:

Platform: GCP BigQuery

*/
-----------------------------------------------------------------------------------------------------------------

-- DATA CLEANING

-- Preview the table

SELECT
    *
FROM
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfaction` LIMIT 100;


-- Remove the int64_field_0 column (was not in the original dataset); the new table is AirlinePassengerSatisfactionV2

SELECT
    *
    EXCEPT (int64_field_0)
FROM
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfaction`;

SELECT
    *
FROM
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2` LIMIT 100;


-- Check if id, the primary column, has duplicate values

SELECT
    COUNT(DISTINCT id) AS id_count,
    COUNT(*) AS total_row_count
FROM
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`;


-- Check if the columns contain null

SELECT
    COUNTIF(id IS NULL) AS id,
    COUNTIF(Gender IS NULL) AS Gender,
    COUNTIF(Customer_Type IS NULL) AS Customer_Type,
    COUNTIF(Age IS NULL) AS Age,
    COUNTIF(Type_of_Travel IS NULL) AS Type_of_Travel,
    COUNTIF(Class IS NULL) AS Class,
    COUNTIF(Flight_Distance IS NULL) AS Flight_Distance,
    COUNTIF(Inflight_wifi_service IS NULL) AS Inflight_wifi_service,
    COUNTIF(Departure_Arrival_time_convenient IS NULL) AS Departure_Arrival_time_convenient,
    COUNTIF(Ease_of_Online_booking IS NULL) AS Ease_of_Online_booking,
    COUNTIF(Gate_location IS NULL) AS Gate_location,
    COUNTIF(Food_and_drink IS NULL) AS Food_and_drink,
    COUNTIF(Online_boarding IS NULL) AS Online_boarding,
```

```
    COUNTIF(Seat_comfort IS NULL) AS Seat_comfort,
    COUNTIF(Inflight_entertainment IS NULL) AS Inflight_entertainment,
    COUNTIF(On_board_service IS NULL) AS On_board_service,
    COUNTIF(Leg_room_service IS NULL) AS Leg_room_service,
    COUNTIF(Baggage_handling IS NULL) AS Baggage_handling,
    COUNTIF(Checkin_service IS NULL) AS Checkin_service,
    COUNTIF(Inflight_service IS NULL) AS Inflight_service,
    COUNTIF(Cleanliness IS NULL) AS Cleanliness,
    COUNTIF(Departure_Delay_in_Minutes IS NULL) AS Departure_Delay_in_Minutes,
    COUNTIF(Arrival_Delay_in_Minutes IS NULL) AS Arrival_Delay_in_Minutes,
    COUNTIF(satisfaction IS NULL) AS satisfaction
FROM
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`;


SELECT
    *
FROM
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
WHERE
    Arrival_Delay_in_Minutes IS NULL;


-- Further check on the columns with numeric values to ensure values contained are within the expected range
-- 1 - 5 for columns derived from satisfaction survey and no unexpected values for other columns

SELECT
    MAX(Age) AS Max_Age,
    MIN(Age) AS Min_Age,
    MAX(Flight_Distance) AS Max_Flight_Distance,
    MIN(Flight_Distance) AS Min_Flight_Distance,
    MAX(Inflight_wifi_service) AS Max_Inflight_wifi,
    MIN(Inflight_wifi_service) AS Min_Inflight_wifi,
    MAX(Departure_Arrival_time_convenient) AS Max_Departure_Arrival_time_convenient,
    MIN(Departure_Arrival_time_convenient) AS Min_Departure_Arrival_time_convenient,
    MAX(Ease_of_Online_booking) AS Max_Ease_of_Online_booking,
    MIN(Ease_of_Online_booking) AS Min_Ease_of_Online_booking,
    MAX(Gate_location) AS Max_Gate_location,
    MIN(Gate_location) AS Min_Gate_location,
    MAX(Food_and_drink) AS Max_Food_and_drink,
    MIN(Food_and_drink) AS Min_Food_and_drink,
    MAX(Online_boarding) AS Max_Online_boarding,
    MIN(Online_boarding) AS Min_Online_boarding,
    MAX(Seat_comfort) AS Max_Seat_comfort,
    MIN(Seat_comfort) AS Min_Seat_comfort,
    MAX(Inflight_entertainment) AS Max_Inflight_entertainment,
    MIN(Inflight_entertainment) AS Min_Inflight_entertainment,
    MAX(On_board_service) AS Max_On_board_service,
    MIN(On_board_service) AS Min_On_board_service,
    MAX(Leg_room_service) AS Max_Leg_room_service,
    MIN(Leg_room_service) AS Min_Leg_room_service,
    MAX(Baggage_handling) AS Max_Baggage_handling,
    MIN(Baggage_handling) AS Min_Baggage_handling,
    MAX(Checkin_service) AS Max_Checkin_service,
    MIN(Checkin_service) AS Min_Checkin_service,
    MAX(Inflight_service) AS Max_Inflight_service,
    MIN(Inflight_service) AS Min_Inflight_service,
    MAX(Cleanliness) AS Max_Cleanliness,
    MIN(Cleanliness) AS Min_Cleanliness,
    MAX(Departure_Delay_in_Minutes) AS Max_Departure_Delay_in_Minutes,
    MIN(Departure_Delay_in_Minutes) AS Min_Departure_Delay_in_Minutes,
```

```
    MAX(Arrival_Delay_in_Minutes) AS Max_Arrival_Delay_in_Minutes,
    MIN(Arrival_Delay_in_Minutes) AS Min_Arrival_Delay_in_Minutes,
FROM
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`;


-- From the brief, 0 in the colums meant the flyer did not answer the question; effectively 0 meant null

UPDATE
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
SET
    Inflight_wifi_service = NULL
WHERE
    Inflight_wifi_service = 0;

UPDATE
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
SET
    Departure_Arrival_time_convenient = NULL
WHERE
    Departure_Arrival_time_convenient = 0;



UPDATE
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
SET
    Ease_of_Online_booking = NULL
WHERE
    Ease_of_Online_booking = 0;

UPDATE
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
SET
    Gate_location = NULL
WHERE
    Gate_location = 0;

UPDATE
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
SET
    Food_and_drink = NULL
WHERE
    Food_and_drink = 0;

UPDATE
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
SET
    Online_boarding = NULL
WHERE
    Online_boarding = 0;

UPDATE
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
SET
    Seat_comfort = NULL
WHERE
    Seat_comfort = 0;

UPDATE
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
SET
```

```
    Inflight_entertainment = NULL
WHERE
    Inflight_entertainment = 0;

UPDATE
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
SET
    On_board_service = NULL
WHERE
    On_board_service = 0;

UPDATE
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
SET
    Leg_room_service = NULL
WHERE
    Leg_room_service = 0;

UPDATE
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
SET
    Baggage_handling = NULL
WHERE
    Baggage_handling = 0;

UPDATE
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
SET
    Checkin_service = NULL
WHERE
    Checkin_service = 0;

UPDATE
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
SET
    Inflight_service = NULL
WHERE
    Inflight_service = 0;

UPDATE
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
SET
    Cleanliness = NULL
WHERE
    Cleanliness = 0;


-- Check the variables contained in the string columns

SELECT
    Gender AS Gender,
    Customer_Type AS Customer_Type,
    Type_of_Travel AS Type_of_Travel,
    Class AS Class,
    satisfaction AS satisfaction
FROM
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
GROUP BY Gender, Customer_Type, Type_of_Travel, Class, satisfaction
ORDER BY Gender, Customer_Type, Type_of_Travel, Class, satisfaction;
```

```sql
-----------------------------------------------------------------------------------------------------------------

-- EXPLORATORY DATA ANALYSIS WITH SQL QUERIES


-- Breakdown of flyers by gender

SELECT
    Gender AS Gender,
    COUNT(id) AS Number,
    ROUND(COUNT(id)/103904 *100, 2) AS GenderPercent
FROM
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
GROUP BY
    Gender;


-- Breakdown of flyers by type of customer

SELECT
    Customer_Type AS TypeOfCustomer,
    COUNT(id) AS Number,
    ROUND(COUNT(id)/103904 *100, 2) AS CustomerTypePercent
FROM
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
GROUP BY
    Customer_Type;


-- Breakdown of flyers by travel type

SELECT
    Type_of_Travel AS TravelType,
    COUNT(id) AS Number,
    ROUND(COUNT(id)/103904 *100, 2) AS TravelTypePercent
FROM
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
GROUP BY
    Type_of_Travel;


-- Breakdown of flyers by class type

SELECT
    Class AS ClassType,
    COUNT(id) AS Number,
    ROUND(COUNT(id)/103904 *100, 1) AS ClassTypePercent
FROM
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
GROUP BY
    Class;


-- Breakdown of flyers by identified satisfaction level

SELECT
    satisfaction AS Satisfaction,
    COUNT(id) AS Number,
    ROUND(COUNT(id)/103904 *100, 1) AS SatisfactionPercent
FROM
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
```

```sql
GROUP BY
    satisfaction;



-- Determine whether demographic variables (gender and age) correlate with identified satisfaction level

SELECT
    satisfaction AS Satisfaction,
    Gender AS Gender,
    COUNT(id) AS Number
FROM
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
GROUP BY
    satisfaction, Gender
ORDER BY
    Gender;



SELECT
    satisfaction AS Satisfaction,
    MAX(Age) AS Max_Age,
    MIN(Age) AS Min_Age,
    ROUND(AVG(Age), 2) AS Avg_Age,
    COUNT(id) AS Number
FROM
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
GROUP BY
    satisfaction;



-- Determine whether identified satisfaction level correlates with any of the column derived from the satisfaction survey

SELECT
    satisfaction AS Satisfaction,
    ROUND(AVG(Inflight_wifi_service), 2) AS Avg_Inflight_wifi,
    ROUND(AVG(Departure_Arrival_time_convenient), 2) AS Avg_Departure_Arrival_time_convenient,
    ROUND(AVG(Ease_of_Online_booking), 2) AS Avg_Ease_of_Online_booking,
    ROUND(AVG(Gate_location), 2) AS Avg_Gate_location,
    ROUND(AVG(Food_and_drink), 2) AS Avg_Food_and_drink,
    ROUND(AVG(Online_boarding), 2) AS Avg_Online_boarding,
    ROUND(AVG(Seat_comfort), 2) AS Avg_Seat_comfort,
    ROUND(AVG(Inflight_entertainment), 2) AS Avg_Inflight_entertainment,
    ROUND(AVG(On_board_service), 2) AS Avg_On_board_service,
    ROUND(AVG(Leg_room_service), 2) AS Avg_Leg_room_service,
    ROUND(AVG(Baggage_handling), 2) AS Avg_Baggage_handling,
    ROUND(AVG(Checkin_service), 2) AS Avg_Checkin_service,
    ROUND(AVG(Inflight_service) , 2)AS Avg_Inflight_service,
    ROUND(AVG(Cleanliness), 2) AS Avg_Cleanliness,
FROM
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
GROUP BY
    satisfaction;



-- Determine whether identified satisfaction level correlates with flight delays

SELECT
    satisfaction AS Satisfaction,
    ROUND(AVG(Departure_Delay_in_Minutes), 2) AS Avg_Departure_Delay_in_Minutes,
```

```
    ROUND(AVG(Arrival_Delay_in_Minutes), 2) AS Avg_Arrival_Delay_in_Minutes,
FROM
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
GROUP BY
    satisfaction;


-- Determine whether identified satisfaction level correlates with type of customer

SELECT
    satisfaction AS Satisfaction,
    Customer_Type AS TypeOfCustomer,
    COUNT(id) AS Number
FROM
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
GROUP BY
    satisfaction, Customer_Type
ORDER BY
    Customer_Type;


-- Determine whether identified satisfaction level correlates with travel type

SELECT
    satisfaction AS Satisfaction,
    Type_of_Travel AS TravelType,
    COUNT(id) AS Number
FROM
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
GROUP BY
    satisfaction, Type_of_Travel
ORDER BY
    Type_of_Travel;


-- Determine which travel type is preferred by each customer type

SELECT
    Customer_Type AS TypeOfCustomer,
    Type_of_Travel AS TravelType,
    COUNT(id) AS Number
FROM
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
GROUP BY
    Customer_Type, Type_of_Travel
ORDER BY
    Customer_Type, Type_of_Travel;


-- Satisfaction level breakdown by type of travel and customer type

SELECT
    Customer_Type AS TypeOfCustomer,
    Type_of_Travel AS TravelType,
    satisfaction AS Satisfaction,
    COUNT(id) AS Number
FROM
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
GROUP BY
    Customer_Type, Type_of_Travel, satisfaction
ORDER BY
```

```
    Customer_Type, Type_of_Travel, satisfaction;


-- Determine whether identified satisfaction level correlates with class type

SELECT
    satisfaction AS Satisfaction,
    Class AS ClassType,
    COUNT(id) AS Number
FROM
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
GROUP BY
    satisfaction, Class
ORDER BY
    Class;


-- Determine which class is preferred by each customer type

SELECT
    Customer_Type AS TypeOfCustomer,
    Class AS ClassType,
    COUNT(id) AS Number
FROM
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
GROUP BY
    Customer_Type, Class
ORDER BY
    Customer_Type, Class;


-- Satisfaction level breakdown by class type and customer type

SELECT
    Customer_Type AS TypeOfCustomer,
    Class AS ClassType,
    satisfaction AS Satisfaction,
    COUNT(id) AS Number
FROM
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
GROUP BY
    Customer_Type, Class, satisfaction
ORDER BY
    Customer_Type, Class, satisfaction;


-- Satisfaction level by class type and customer type in relation to the 14 measures of satisfaction from the survey

SELECT
    Customer_Type AS TypeOfCustomer,
    Class AS ClassType,
    satisfaction AS Satisfaction,
    COUNT(id) AS Number,
    ROUND(AVG(Inflight_wifi_service), 2) AS Avg_Inflight_wifi,
    ROUND(AVG(Departure_Arrival_time_convenient), 2) AS Avg_Departure_Arrival_time_convenient,
    ROUND(AVG(Ease_of_Online_booking), 2) AS Avg_Ease_of_Online_booking,
    ROUND(AVG(Gate_location), 2) AS Avg_Gate_location,
    ROUND(AVG(Food_and_drink), 2) AS Avg_Food_and_drink,
    ROUND(AVG(Online_boarding), 2) AS Avg_Online_boarding,
    ROUND(AVG(Seat_comfort), 2) AS Avg_Seat_comfort,
    ROUND(AVG(Inflight_entertainment), 2) AS Avg_Inflight_entertainment,
```

```
    ROUND(AVG(On_board_service), 2) AS Avg_On_board_service,
    ROUND(AVG(Leg_room_service), 2) AS Avg_Leg_room_service,
    ROUND(AVG(Baggage_handling), 2) AS Avg_Baggage_handling,
    ROUND(AVG(Checkin_service), 2) AS Avg_Checkin_service,
    ROUND(AVG(Inflight_service) , 2)AS Avg_Inflight_service,
    ROUND(AVG(Cleanliness), 2) AS Avg_Cleanliness
FROM
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
GROUP BY
    Customer_Type, Class, satisfaction
ORDER BY
    Customer_Type, Class, satisfaction;


-- Impact of flight delays on satisfaction by customer type and class type

SELECT
    Customer_Type AS TypeOfCustomer,
    Class AS ClassType,
    satisfaction AS Satisfaction,
    COUNT(id) AS Number,
    ROUND(AVG(Departure_Delay_in_Minutes), 2) AS Avg_Departure_Delay_in_Minutes,
    ROUND(AVG(Arrival_Delay_in_Minutes), 2) AS Avg_Arrival_Delay_in_Minutes,
FROM
    `causal-bison-323215.AirlinePassengerSatisfaction.AirlinePassengerSatisfactionV2`
GROUP BY
    Customer_Type, Class, satisfaction
ORDER BY
    Customer_Type, Class, satisfaction;
```