

Principles of Data Engineering

Project 2.2

QUERIES:

1) In how many trips, buses departed with between 25% and 75% of their full capacity, in terms of sold tickets, in the years 2016 and 2017? (Santiago)

-- Query 1

-- Generates a table with columns trip_id, capacity, quantity

```
SELECT      tr.trip_id, bu.capacity, count(*) as quantity
FROM        tickets ti join
            trips tr      on tr.trip_id=ti.trip_id join
            buses bu      on bu.bus_id=tr.bus_id
WHERE       year(tr.date) = 2016 or year(tr.date) = 2017
GROUP BY    tr.trip_id, bu.capacity
HAVING      count(*)>=0.25*bu.capacity and count(*)<=0.75*bu.capacity
ORDER BY    quantity desc
```

-- A total of 12,237 trips had buses departing with between 25% and 75% of their full capacity, in terms of sold tickets, in the years 2016 and 2017.

2) List the top 10 trips that generated the most revenue (SUM of final_price) in the first three months during weekdays (Monday to Friday) in the year 2016? (Santiago)

--Query 2

```
SELECT      top 10 sum(ti.final_price) max_revenue, tr.trip_id
FROM        tickets ti join
            trips tr on ti.trip_id=tr.trip_id
WHERE       year(tr.date)=2016
            and DATEPART (month, tr.date) in (1,2,3)
            and DATEPART (weekday, tr.date) in (2,3,4,5,6)
GROUP BY    tr.trip_id
ORDER BY    max_revenue desc
```

/*This code gives as a result a table with the top 10 trips that generated the most revenue in the first 3 months of 2016.

The table presents the total revenue in the 3 months, and the trip id of that trip.*/

-- The trip_ids for the top 10 trips with the most revenue are 346, 1734, 369, 922, 1066, 204, 984, 931, 1337, and 600.

3) Which routes of trips generated the top 10% most and the top 10% least revenue (SUM of final_price) in 2017? (Kellie Ann)

-- Query 3

-- Routes of trips generated the top 10% most and the top 10% least revenue (SUM of final_price) in 2017

-- Generates a table to include the route-id, total-revenue, revenue-category

WITH route_revenue_2017 AS (

```

SELECT
    r.route_id,
    SUM(tk.final_price) AS total_revenue
FROM routes r
JOIN trips t ON r.route_id = t.route_id
JOIN tickets tk ON tk.trip_id = t.trip_id
WHERE YEAR(tk.purchase_date) = 2017
GROUP BY r.route_id
),
ranked_routes AS (
    SELECT
        route_id,
        total_revenue,
        PERCENT_RANK() OVER (ORDER BY total_revenue) AS revenue_percentile
    FROM route_revenue_2017
)
SELECT
    route_id,
    total_revenue,
    CASE
        WHEN revenue_percentile >= 0.9 THEN 'Top 10%'
        WHEN revenue_percentile <= 0.1 THEN 'Bottom 10%'
    END AS revenue_category
FROM ranked_routes
WHERE revenue_percentile >= 0.9 OR revenue_percentile <= 0.1;
-- The route_id's that generated the bottom 10% are 114, 7, 79, 129, 62, 53, 28, 47, 18, 27,
94, 42, 89, 41
-- and the bottom 10% are 96, 29, 44, 133, 132, 50, 116, 64, 45, 126, 97, 137, 131

```

4) List the amount of money discounted from the total revenue (SUM of final_price) and the number of tickets that received a discount in the year 2017 to the customers that were either below 23 years old or above 65 years old at the moment they purchase the ticket.
(Antonieta)

```

-- Query 4
-- This generates a table with the columns as total_discounted_amount and
tickets_with_discount.
SELECT
    SUM(t.final_price * d.percentage / 100.0) AS total_discounted_amount,
    COUNT(*) AS tickets_with_discount
FROM tickets t
JOIN customers c ON t.customer_id = c.customer_id
JOIN discounts d ON t.discount_id = d.discount_id
WHERE
    DATEPART(year, t.purchase_date) = 2017
    AND (
        DATEDIFF(year, c.birth_date, t.purchase_date) < 23
        OR DATEDIFF(year, c.birth_date, t.purchase_date) > 65
    );
-- The answer for the total_discounted amount is 125.0105 and the tickets_with_discount is
7895.

```

5) Estimate the monthly average ratio of sold tickets of registered passengers over not registered passengers in the year 2017 and only for trips departing after 11 AM on the weekends (Saturday and Sunday). (Barbara)

```
-- Query 5
-- This query makes a table with a column for Month, MonthNumber,
Ratio_Refistered_To_NonRegistered
SELECT
    DATENAME(MONTH, t.date) AS [Month],
    MONTH(t.date) AS MonthNumber,
    /* Avoid division by zero with NULLIF(...) */
    1.0 * SUM(CASE WHEN tk.customer_id IS NOT NULL THEN 1 ELSE 0 END)
    / NULLIF(SUM(CASE WHEN tk.customer_id IS NULL THEN 1 ELSE 0 END), 0)
    AS Ratio_Registered_To_NonRegistered
FROM trips AS t
JOIN routes AS r
    ON t.route_id = r.route_id
JOIN weekdays AS w
    ON r.weekday_id = w.weekday_id
JOIN tickets AS tk
    ON t.trip_id = tk.trip_id
WHERE
    YEAR(t.date) = 2017
    AND w.name IN ('Saturday','Sunday')
    AND t.start_time_actual > '11:00:00'
GROUP BY
    DATENAME(MONTH, t.date),
    MONTH(t.date)
ORDER BY
    MonthNumber;
-- The monthly average ratio of sold tickets of registered passengers over not registered
passengers in the year 2017 and only for trips departing after 11 AM on the weekends are:
-- January 1.234, February 1.259, March 1.378, April 1.285, May 1.174, June 1.309, July
1.371, August 1.157, September 1.410, October 1.322, November 1.338, December 1.385
```

6) What are the most demanded routes in terms of number of tickets sold, for routes departing between 2:00 pm and 8:00 pm for each weekday in the months of June, July and August in the year 2016? (Ethan)

```
-- Query 6
-- The columns from the table are route_id, weekday_name, and tickets_sold
WITH RouteRanks AS (
    SELECT
        r.route_id,
        w.name AS weekday_name,
        COUNT(tk.ticket_id) AS tickets_sold,
        ROW_NUMBER() OVER (PARTITION BY w.name ORDER BY COUNT(tk.ticket_id)
        DESC) AS rank
```

```

FROM trips t
JOIN tickets tk ON t.trip_id = tk.trip_id
JOIN routes r ON t.route_id = r.route_id
JOIN weekdays w ON r.weekday_id = w.weekday_id
WHERE
    YEAR(t.date) = 2016
    AND MONTH(t.date) IN (6, 7, 8)
    AND CAST(t.start_time_actual AS TIME) BETWEEN '14:00:00' AND '20:00:00'
    AND w.name IN ('Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday')
    GROUP BY r.route_id, w.name
)
SELECT route_id, weekday_name, tickets_sold
FROM RouteRanks
WHERE rank = 1
ORDER BY weekday_name;
-- The routes in highest demand during the weekdays between 2-8 pm in the months of
June, July, and August are 118, 77, 106, 87, and 96.

```

7) For routes departing from Tampa to any destination, for each weekday, what is the most demanded hour in terms of generated revenue (SUM of final_price)? **(Santiago)**

```

-- Query 7
WITH DailyRevenue AS (
    SELECT
        r.weekday_id,
        w.name AS weekday_name,
        r.scheduled_start_time,
        SUM(t.final_price) AS Generated_Revenue,
        ROW_NUMBER() OVER (
            PARTITION BY r.weekday_id
            ORDER BY SUM(t.final_price) DESC
        ) AS rn
    FROM
        dbo.routes r
        JOIN dbo.trips tr ON r.route_id = tr.route_id
        JOIN dbo.tickets t ON t.trip_id = tr.trip_id
        JOIN dbo.weekdays w ON r.weekday_id = w.weekday_id
    WHERE
        r.city_state_id_origin = 1 AND (w.weekday_id=1 or w.weekday_id=2 or w.weekday_id=3
or w.weekday_id=4 or w.weekday_id=5)
    GROUP BY
        r.weekday_id,
        w.name,
        r.scheduled_start_time
)
SELECT weekday_name, scheduled_start_time, Generated_Revenue
FROM
    DailyRevenue
WHERE

```

```
rn = 1
ORDER BY
weekday_id;
```

--This answer will provide a table with the day of the week, the hour which is more in demand, and the generated revenue in that hour for all the trips going out from Tampa.

8) Knowing that at least 10% of the capacity of the full passenger capacity of the buses, in relation to the number of tickets sold, is necessary to justify the departure of a bus on a trip. The managers would like to know which trips should not have departed either the years 2016 or 2017? (Kellie Ann)

-- Query 8

-- The table in SQL has the columns trip_id, date, bus_id, tickets_sold, capacity, and minimum_required_passengers.

```
WITH trip_ticket_counts AS (
    SELECT
        t.trip_id,
        t.date,
        t.bus_id,
        COUNT(tk.ticket_id) AS tickets_sold
    FROM trips t
    LEFT JOIN tickets tk ON tk.trip_id = t.trip_id
    WHERE YEAR(t.date) IN (2016, 2017)
    GROUP BY t.trip_id, t.date, t.bus_id
),
trip_with_capacity AS (
    SELECT
        tc.trip_id,
        tc.date,
        tc.bus_id,
        tc.tickets_sold,
        b.capacity,
        CEILING(b.capacity * 0.10) AS minimum_required_passengers
    FROM trip_ticket_counts tc
    JOIN buses b ON tc.bus_id = b.bus_id
)
SELECT
    trip_id,
    date,
    bus_id,
    tickets_sold,
    capacity,
    minimum_required_passengers
FROM trip_with_capacity
WHERE tickets_sold < minimum_required_passengers;
-- The trips that should not have left include 743, 1210, 1513, 2806, 3246, 7394, 14215, and 14403.
```

9) Which weeks (from 1 to 52) of the year 2016 sold between 10% and 30% of the full capacity of tickets on the weekends? (Antonieta)

-- Query 9

```
WITH weekend_trips AS (
    SELECT
        DATEPART(WEEK, tr.date) AS week_number,
        tr.trip_id,
        b.capacity,
        COUNT(t.ticket_id) AS tickets_sold
    FROM trips tr
    JOIN tickets t ON t.trip_id = tr.trip_id
    JOIN buses b ON tr.bus_id = b.bus_id
    WHERE
        YEAR(tr.date) = 2016
        AND DATENAME(WEEKDAY, tr.date) IN ('Saturday', 'Sunday')
    GROUP BY DATEPART(WEEK, tr.date), tr.trip_id, b.capacity
),
weekly_summary AS (
    SELECT
        week_number,
        SUM(tickets_sold) AS total_tickets,
        SUM(capacity) AS total_capacity,
        CAST(SUM(tickets_sold) AS FLOAT) / NULLIF(SUM(capacity), 0) AS load_ratio
    FROM weekend_trips
    GROUP BY week_number
)
SELECT
    week_number,
    load_ratio
FROM weekly_summary
WHERE load_ratio BETWEEN 0.10 AND 0.30
ORDER BY week_number;
```

--In 2016, **48 out of 52 weeks** had weekend ticket sales that fell between **10% and 30%** of the total available bus capacity. The qualifying weeks include weeks **2 through 52**, with the exception of weeks **1, 9, and 47**, which did not meet the criteria. **Week 53** was excluded from the analysis, as the question specifically focused on weeks **1 through 52**.

10) List the five employees that have sold the most tickets without a discount in the last three months of the year 2017 and the five employees that have generated the most revenue (SUM of final_price) selling tickets with a discount for weekdays in the year 2017? (Barbara)

-- Query 10

-- PART A

-- Top 5 employees with the most tickets sold WITHOUT discount in
-- the last three months of 2017 (Oct 1–Dec 31, 2017)

```
SELECT TOP 5
    e.employee_id,
    e.first_name,
    e.last_name,
    COUNT(tk.ticket_id) AS TicketsSold
FROM tickets AS tk
JOIN employees AS e
```

```

        ON tk.employee_id = e.employee_id
WHERE
    tk.discount_id IS NULL
    AND tk.purchase_date >= '2017-10-01'
    AND tk.purchase_date <= '2017-12-31'
GROUP BY
    e.employee_id,
    e.first_name,
    e.last_name
ORDER BY
    TicketsSold DESC;

-- PART B
-- Top 5 employees with the most revenue from DISCOUNTED tickets
-- on WEEKDAYS in 2017

```

```

SELECT TOP 5
    e.employee_id,
    e.first_name,
    e.last_name,
    SUM(tk.final_price) AS TotalRevenue
FROM tickets AS tk
JOIN employees AS e
    ON tk.employee_id = e.employee_id
JOIN trips AS t
    ON tk.trip_id = t.trip_id
JOIN routes AS r
    ON t.route_id = r.route_id
JOIN weekdays AS w
    ON r.weekday_id = w.weekday_id
WHERE
    tk.discount_id IS NOT NULL
    AND w.name NOT IN ('Saturday','Sunday')
    AND YEAR(tk.purchase_date) = 2017
GROUP BY
    e.employee_id,
    e.first_name,
    e.last_name
ORDER BY
    TotalRevenue DESC;

```

--the five employees that have sold the most tickets without a discount in the last three months of the year 2017: Wren, Korie, Sandra, Denise, Robena
-- And the five employees that have generated the most revenue (SUM of final_price) selling tickets with a discount for weekdays in the year 2017 are: Robena, Mord, Korie, Bryon, Ardeen

11) What is the most demanded cabin type in terms of sold tickets without a discount in the first three months of the year 2017 for registered customers? (Ethan)

```

-- Query 11
-- Yields a table with columns cabin_type and tickets_sold
SELECT TOP 1

```

```

        cb.name AS cabin_type,
        COUNT(tk.ticket_id) AS tickets_sold
FROM tickets tk
JOIN cabin_types cb ON tk.cabin_type_id = cb.cabin_type_id
WHERE
    tk.customer_id IS NOT NULL
    AND tk.discount_id IS NULL
    AND YEAR(tk.purchase_date) = 2017
    AND MONTH(tk.purchase_date) IN (1, 2, 3)
GROUP BY cb.name
ORDER BY tickets_sold DESC;
-- The most demanded cabin type is second class.

```

12) What is the purchase location in which most tickets for non-registered customers were sold for weekdays only in the first two quarters of the year 2016? (Santiago)

```

-- Query 12
SELECT
    weekday_id, weekday_name, purchase_location, total_tickets
FROM (
    SELECT
        w.weekday_id AS weekday_id,
        w.name AS weekday_name,
        l.name AS purchase_location,
        COUNT(*) AS total_tickets,
        ROW_NUMBER() OVER (
            PARTITION BY w.name
            ORDER BY COUNT(*) DESC
        ) AS rn
    FROM
        dbo.tickets t
        JOIN dbo.trips tr ON t.trip_id = tr.trip_id
        JOIN dbo.routes r ON tr.route_id = r.route_id
        JOIN dbo.weekdays w ON r.weekday_id = w.weekday_id
        JOIN dbo.locations l ON t.purchase_location_id = l.location_id
    WHERE
        t.customer_id IS NULL AND
        t.purchase_date BETWEEN '2016-01-01' AND '2016-06-30' AND
        (w.weekday_id=1 or w.weekday_id=2 or w.weekday_id=3 or w.weekday_id=4
or w.weekday_id=5)
    GROUP BY
        w.name, l.name, w.weekday_id
) AS ranked
WHERE rn = 1
ORDER BY weekday_id;

```

/*This answer will provide a table with the day of the week, the purchase location that sold the most amount of tickets in that day, and the amount of tickets that were sold. This amount will be for the first 2 quarters of the year 2016. The amount of tickets sold is the sum of all the tickets sold in a specific location for every monday (for example) for the first semester of the year.*/

13) From all the trips in the last two quarters of the year 2017, which trips departed with more tickets sold to registered customers than to non-registered customers? (Kellie Ann)

-- Query 13

-- This table includes the trips departed with more tickets sold to registered customers than to non-registered customers.

-- Generates a table with the columns being trip_id, registered_count, non-registered_count

-WITH ticket_classification AS (

SELECT

t.trip_id,

CASE

WHEN tk.customer_id IS NOT NULL THEN 'registered'

ELSE 'non_registered'

END AS customer_type

FROM trips t

JOIN tickets tk ON t.trip_id = tk.trip_id

WHERE

YEAR(t.date) = 2017 AND

MONTH(t.date) BETWEEN 7 AND 12 -- Q3 and Q4

),

ticket_counts AS (

SELECT

trip_id,

SUM(CASE WHEN customer_type = 'registered' THEN 1 ELSE 0 END) AS

registered_count,

SUM(CASE WHEN customer_type = 'non_registered' THEN 1 ELSE 0 END) AS

non_registered_count

FROM ticket_classification

GROUP BY trip_id

)

SELECT

trip_id,

registered_count,

non_registered_count

FROM ticket_counts

WHERE registered_count > non_registered_count;

-- The table yields 2,154 results, therefore there were 2,154 trips that departed with more tickets sold to registered customers than to non-registered customers.

14) What is the location type with most sold tickets to registered customers in the first two months of the years 2016 and 2017? (Antonieta)

-- Query 14

-- TEMP TESTING VERSION to show results

SELECT TOP 3

l.location_type_id,

COUNT(*) AS tickets_sold

FROM tickets t

JOIN customers c ON t.customer_id = c.customer_id

JOIN locations l ON t.purchase_location_id = l.location_id

WHERE

DATEPART(year, t.purchase_date) IN (2016, 2017)

```

    AND DATEPART(month, t.purchase_date) IN (1, 2)
GROUP BY l.location_type_id
ORDER BY COUNT(*) DESC;
-- The location type with the most sold tickets in January and February of 2016 and 2017
is:location_type_id = 1 With a total of 8,334 tickets sold

```

15) What is the hour of the day in each weekday (excluding weekends) where the most tickets departing after 11 AM were sold in 2017 to non-registered customers? (Barbara)

```

-- Query 15
WITH HourlySales AS
(
    SELECT
        w.name AS WeekdayName,
        DATEPART(HOUR, t.start_time_actual) AS DepartureHour,
        COUNT(*) AS TicketsSold,
        ROW_NUMBER() OVER (
            PARTITION BY w.name
            ORDER BY COUNT(*) DESC
        ) AS rn
    FROM tickets AS tk
    JOIN trips AS t
    ON tk.trip_id = t.trip_id
    JOIN routes AS r
    ON t.route_id = r.route_id
    JOIN weekdays AS w
    ON r.weekday_id = w.weekday_id
    WHERE
        YEAR(t.date) = 2017
        AND w.name NOT IN ('Saturday','Sunday')
        AND t.start_time_actual > '11:00:00'
        AND tk.customer_id IS NULL
    GROUP BY
        w.name,
        DATEPART(HOUR, t.start_time_actual)
)
SELECT
    WeekdayName,
    DepartureHour,
    TicketsSold
FROM HourlySales
WHERE rn = 1
ORDER BY WeekdayName;
-- The hour of the day in each weekday (excluding weekends) where the most tickets
departing after 11 AM were sold in 2017 to non-registered customers are:
-- Monday 18, Thursday 16, Tuesday 16, Wednesday 14, Friday 14

```

16) What are the two discounts that were applied the most to non-registered customers in even months (February, April, etc.) in the year 2017 and what is the amount of revenue discounted for these three discounts? (Ethan)

```

-- Query 16
-- The columns produced are discount_id, discount_name, discount_count, total_discount
WITH non_registered_discounts AS (
    SELECT
        d.discount_id,
        d.name AS discount_name,
        COUNT(*) AS discount_count,
        SUM(t.final_price * d.percentage / 100.0) AS total_discount
    FROM tickets t
    JOIN discounts d ON t.discount_id = d.discount_id
    LEFT JOIN customers c ON t.customer_id = c.customer_id
    WHERE
        c.customer_id IS NULL -- non-registered
        AND YEAR(t.purchase_date) = 2017
        AND MONTH(t.purchase_date) % 2 = 0 -- even months
    GROUP BY d.discount_id, d.name
)
SELECT TOP 2 *
FROM non_registered_discounts
ORDER BY discount_count DESC;
-- The discount_ids were 2 and 4.

```

17) What is the hour of the day in which most tickets are purchased by non-registered customers and which are not finally boarded by the customer (missed trip)? (Kellie Ann)

```

-- Query 17
SELECT
    DATEPART(HOUR, purchase_time) AS purchase_hour,
    COUNT(*) AS missed_tickets_count
FROM tickets
WHERE
    customer_id IS NULL AND
    (boarding_date IS NULL OR boarding_time IS NULL)
GROUP BY DATEPART(HOUR, purchase_time)
ORDER BY missed_tickets_count DESC;
-- The hour that the most tickets are purchased by non-registered customers and which are
not finally boarded by the customer is the 12th hour.

```

18) What are the three registered customers that bought the most tickets departing after 11 AM in the first four months of the year 2016 and what is the revenue generated (SUM of final_price) for these customers? (Antonieta)

```

-- Query 18
SELECT TOP 3
    c.customer_id,
    c.first_name,
    c.last_name,
    COUNT(*) AS tickets_bought,
    SUM(t.final_price) AS total_revenue
FROM tickets t
JOIN customers c ON t.customer_id = c.customer_id

```

```

JOIN trips tr ON t.trip_id = tr.trip_id
WHERE
    DATEPART(year, t.boarding_date) = 2016
    AND DATEPART(month, t.boarding_date) BETWEEN 1 AND 4
    AND CAST(tr.start_time_actual AS time) > '11:00:00'
GROUP BY
    c.customer_id, c.first_name, c.last_name
ORDER BY
    tickets_bought DESC;
--The three customers who bought the most tickets for trips departing after 11:00 AM during
January to April 2016 are: Jobie, Cele, and Ric

```

19) What are the top 3 routes on weekends (Saturday, Sunday) that registered the highest number of delayed minutes in the year 2017? **(Barbara)**

```

-- Query 19
SELECT TOP 3
    r.route_id,
    SUM(
        CASE
            WHEN DATEDIFF(MINUTE, r.scheduled_start_time, t.start_time_actual) > 0
            THEN DATEDIFF(MINUTE, r.scheduled_start_time, t.start_time_actual)
            ELSE 0
        END
    ) AS TotalDelayedMinutes
FROM trips AS t
JOIN routes AS r
    ON t.route_id = r.route_id
JOIN weekdays AS w
    ON r.weekday_id = w.weekday_id
WHERE
    YEAR(t.date) = 2017
    AND w.name IN ('Saturday','Sunday')
GROUP BY
    r.route_id
ORDER BY
    TotalDelayedMinutes DESC;
-- The top 3 routes on weekends that registered the highest number of delayed minutes in
the year 2017 are: 57, 134, 69

```

20) What are the top 3 routes in the first half of the year 2017 that registered the lowest number of missed trips by registered customers? **(Ethan)**

```

-- Query 20
SELECT TOP 3
    r.route_id,
    COUNT(*) AS missed_trips
FROM tickets t
JOIN trips tr ON t.trip_id = tr.trip_id
JOIN routes r ON tr.route_id = r.route_id
WHERE

```

```
t.customer_id IS NOT NULL
AND YEAR(tr.date) = 2017
AND MONTH(tr.date) BETWEEN 1 AND 6
AND (t.boarding_date IS NULL OR t.boarding_time IS NULL)
GROUP BY r.route_id
ORDER BY missed_trips ASC;
-- Result:
-- The top 3 routes with the fewest missed trips by registered customers in the first half of
2017 are:
-- Route 132 → 6 missed trips
-- Route 21 → 6 missed trips
-- Route 85 → 6 missed trips
```

1) One SQL file with all codes and comments inserted

2) One PDF document with solutions shown. This may include showing tables and answers to the questions.

Question 1

There were 12,237 trips in 2016 and 2017 that the buses departed with an occupancy rate of between 25% and 75% of their capacity.

	trip_id	capacity	quantity
1	13471	32	16
2	12229	32	16
3	7194	32	16
4	13284	32	16
5	11459	32	15
6	10978	32	15
7	11676	32	15
8	12864	32	15
9	922	32	15
10	6106	32	15
11	12547	32	15
12	4117	32	15
13	894	32	15
14	8022	32	15
15	11199	32	15
16	9161	32	15
17	2824	32	15
18	12585	32	15
19	14334	32	15
20	346	32	15
21	10090	32	15
22	8822	32	15
23	3034	32	15
24	10342	32	15
25	13587	32	15
26	12465	32	15
27	9176	32	15
28	2312	32	15
29	11707	32	15

ortation 00:00:00 12,237 rows

Question 2

For the list of the top 10 trips that generated most revenue from between the first three months of the year 2016, i.e., weekdays (Monday through Friday). We used the SUM of the final_price to get the total revenue for each trip. The top revenue-generating trip was trip_id 346 with \$135.50, then trip_id 1734 with \$132.00, and trip_id 369 with \$127.00.

	max_revenue	trip_id
1	135.50	346
2	132.00	1734
3	127.00	369
4	125.50	922
5	125.00	1066
6	124.00	204
7	123.50	894
8	122.50	931
9	121.00	1337
10	121.00	600

Question 3

The routes of trips that generated the top 10% of the most revenue included routes; 114, 7, 79, 129, 62, 53, 28, 47, 18, 27, 94, 42, 89, 41. The routes of trips that generated the top 10% of the most revenue included routes;the bottom 10% are 96, 29, 44, 133, 132, 50, 116, 64, 45, 126, 97, 137, 131.

	route_id	total_revenue	revenue_category
1	114	3510.00	Bottom 10%
2	7	3650.00	Bottom 10%
3	79	3656.50	Bottom 10%
4	129	3662.00	Bottom 10%
5	62	3670.50	Bottom 10%
6	53	3692.50	Bottom 10%
7	28	3711.00	Bottom 10%
8	47	3720.50	Bottom 10%
9	18	3723.50	Bottom 10%
10	27	3725.00	Bottom 10%
11	94	3786.50	Bottom 10%
12	42	3798.00	Bottom 10%
13	89	3806.00	Bottom 10%
14	41	3834.00	Bottom 10%
15	96	4410.00	Top 10%
16	29	4410.00	Top 10%
17	44	4417.00	Top 10%
18	133	4422.50	Top 10%
19	132	4449.00	Top 10%
20	50	4453.50	Top 10%
21	116	4455.50	Top 10%
22	64	4456.00	Top 10%
23	45	4471.50	Top 10%
24	126	4485.50	Top 10%
25	97	4488.50	Top 10%
26	137	4544.00	Top 10%
27	131	4745.00	Top 10%

Question 4

	total_discounted_amount	tickets_with_discount
1	125.010500000	7895

Question 5

	Month	MonthNumber	Ratio_Registered_To_NonRegistered
1	January	1	1.233707865168
2	February	2	1.258793969849
3	March	3	1.377717391304
4	April	4	1.284836065573
5	May	5	1.174563591022
6	June	6	1.309462915601
7	July	7	1.370762711864
8	August	8	1.157004830917
9	September	9	1.410126582278
10	October	10	1.322274881516
11	November	11	1.337662337662
12	December	12	1.384946236559

Question 6

The following routes are the most in-demand for each day of the week between 2:00 PM and 8:00 PM during June, July, and August of 2016. Each route had the highest number of tickets sold on that weekday:

Route 77 is the most in-demand route on Monday between 2 PM and 8 PM with 132 tickets sold. Route 87 is the most in-demand route on Tuesday between 2 PM and 8 PM with 129 tickets sold. Route 96 is the most in-demand route on Wednesday between 2 PM and 8 PM with 140 tickets sold. Route 106 is the most in-demand route on Thursday between 2 PM and 8 PM with 131 tickets sold. Route 118 is the most in-demand route on Friday between 2 PM and 8 PM with 134 tickets sold.

Here is the output table from SQL:

	route_id	weekday_name	tickets_sold
1	118	Friday	134
2	77	Monday	132
3	106	Thursday	131
4	87	Tuesday	129
5	97	Wednesday	140

Question 7

For routes that depart from Tampa we checked each weekday to find out what hour of the day was the most profitable, utilizing the SUM of final_price. The result was that the most profitable times of departure varied on weekdays. For example, Monday at 10:00 AM brought \$8,820.50, while Friday at 12:00 PM brought the most revenue of all with \$8,893.00. The other rush hours were on Tuesday at 12:00 PM, Wednesday at 8:00 PM, and Thursday at 4:00 AM. These are representative of the changing passenger behavior and demand during the week.

	weekday_name	scheduled_start_time	Generated_Revenue
1	Monday	10:00:00.0000000	8820.50
2	Tuesday	12:00:00.0000000	8643.50
3	Wednesday	20:00:00.0000000	8766.00
4	Thursday	04:00:00.0000000	8803.50
5	Friday	12:00:00.0000000	8893.00

Question 8

	trip_id	date	bus_id	tickets_sold	capacity	minimum_required_passengers
1	743	2016-02-03	2	3	32	4
2	1210	2016-02-26	1	3	32	4
3	1513	2016-03-12	7	3	32	4
4	2806	2016-05-16	7	3	32	4
5	3246	2016-06-07	7	3	32	4
6	7394	2016-12-31	3	3	32	4
7	14215	2017-12-07	8	3	32	4
8	14403	2017-12-17	2	3	32	4

Question 9

	week_number	load_ratio			
1	2	0.28828125	18	20	0.28125
2	3	0.27109375	19	21	0.2953125
3	4	0.2859375	20	22	0.2890625
4	5	0.29453125	21	23	0.29921875
5	6	0.2859375	22	24	0.28828125
6	7	0.2921875	23	25	0.27421875
7	8	0.2859375	24	26	0.29609375
8	10	0.28125	25	27	0.28359375
9	11	0.29140625	26	28	0.2953125
10	12	0.29765625	27	29	0.27734375
11	13	0.29296875	28	30	0.28828125
12	14	0.2921875	29	31	0.29296875
13	15	0.284375	30	32	0.28984375
14	16	0.27734375	31	33	0.2828125
15	17	0.28203125	32	34	0.296875
16	18	0.2875	33	36	0.2984375
17	19	0.28046875	34	37	0.290625
35	38	0.29296875			
36	39	0.2859375			
37	40	0.29609375			
38	41	0.29296875			
39	42	0.2828125			
40	43	0.2796875			
41	44	0.28984375			
42	45	0.29375			
43	46	0.2734375			
44	48	0.27734375			
45	49	0.28203125			
46	50	0.27109375			
47	51	0.28046875			
48	52	0.28046875			
49	53	0.28046875			

In 2016, 48 out of 52 weeks had weekend ticket sales that fell between 10% and 30% of the total available bus capacity. The qualifying weeks include weeks 2 through 52, with the exception of weeks 1, 9, and 47, which did not meet the criteria. Week 53 was excluded from the analysis, as the question specifically focused on weeks 1 through 52.

Question 10

	employee_id	first_name	last_name	TicketsSold
1	3	Wren	Grogona	69
2	9	Korie	Dugmore	65
3	20	Sandra	Heigl	65
4	14	Denise	Schruyer	64
5	6	Robena	Krolle	63

	employee_id	first_name	last_name	TotalRevenue
1	6	Robena	Krolle	4299.50
2	15	Mord	Belding	4246.00
3	9	Korie	Dugmore	4173.00
4	18	Bryon	Rudwell	4152.50
5	4	Ardeen	Cronchey	4152.00

Question 11

In the first three months of 2017, among registered customers who purchased tickets without using a discount, the most in-demand cabin type was "Second Class" with 1,175 tickets sold.

Here is the output table from SQL:

	cabin_type	tickets_sold
1	Second Class	1175

Question 12

The largest number of tickets sold to non-registered customers during weekdays in the first half of 2016 is at the Orlando FL Central Bus Station. This emerges from the data given for Wednesday, where 328 tickets were sold in that location, outpacing ticket volumes for any other weekday and purchasing location combination. While Tampa FL Central Bus Station maintained high sales throughout the rest of the weekdays, it failed to beat the peak observed in Orlando on Wednesday. This is an indication that Orlando witnessed a very high mid-week demand from non-registered customers during this period.

	weekday_id	weekday_name	purchase_location	total_tickets
1	1	Monday	Tampa FL Central Bus Station	305
2	2	Tuesday	Tampa FL Central Bus Station	305
3	3	Wednesday	Orlando FL Central Bus Station	328
4	4	Thursday	Tampa FL Central Bus Station	301
5	5	Friday	Tampa FL Central Bus Station	296

Question 13

	trip_id	registered_count	non_registered_count
1	12071	4	2
2	14324	6	1
3	11384	6	4
4	11453	5	4
5	12117	6	4
6	12758	4	3
7	11576	6	4
8	12904	6	4
9	14155	8	3
10	11599	6	4
11	12263	6	5
12	12927	7	3
13	13468	7	3
14	14132	5	4
15	12286	6	5
16	13783	7	2
17	14278	5	3
18	14447	6	3
19	13096	10	1
20	13637	6	3
21	13760	6	2
22	14178	4	3
23	14424	7	2
24	12409	6	5
25	13737	7	2

Query executed successfully. LAPTOP-7BPT2TFC\SQLEXPRESS ... LAPTOP-7BPT2TFC\Kellie... Bus transportation 00:00:00 2,154 rows

Question 14

	location_type_id	tickets_sold
1	1	8834
2	2	3535

The location type with the most sold tickets in January and February of 2016 and 2017 is: location_type_id = 1 with a total of 8,334 tickets sold

Question 15

	WeekdayName	DepartureHour	TicketsSold
1	Friday	14	466
2	Monday	18	442
3	Thursday	16	438
4	Tuesday	16	457
5	Wednesday	14	466

Question 16

For non-registered customers in the even-numbered months of 2017, the two most commonly applied discounts were:

1. "Walmart Customers Discount" (discount_id 2), applied 3,768 times, resulting in \$48.04 of total discounts.
2. "Amazon Prime Customers Discount" (discount_id 4), applied 3,675 times, resulting in \$68.91 of total discounts.

Here is the output table from SQL:

	discount_id	discount_name	discount_count	total_discount
1	2	Walmart Customers Discount	3768	48.042000000
2	4	Amazon Prime Customers Discount	3675	68.906250000

Question 17

	purchase_hour	missed_tickets_count
1	12	302
2	7	300
3	3	298
4	11	297
5	15	296
6	10	290
7	8	290
8	18	287
9	4	286
10	2	283
11	20	278
12	13	277
13	9	276
14	6	274
15	14	273
16	17	272
17	19	265
18	16	265
19	5	256
20	21	243
21	1	57

Question 18

	customer_id	first_name	last_name	tickets_bought	total_revenue
1	2739	Jobie	Jenno	15	128.50
2	4806	Cele	Jinda	15	124.50
3	2655	Ric	Tuley	13	106.50

The three customers who bought the most tickets for trips departing after 11:00 AM during January to April 2016 are: Jobie, Cele, and Ric

Question 19

	route_id	TotalDelayedMinutes
1	57	127
2	134	124
3	69	123

Question 20

In the first half of 2017, the following three routes had the fewest missed trips (defined as trips by registered customers where the boarding date or time was not recorded):

Route 132 had 6 missed trips.

Route 21 had 6 missed trips.

Route 85 had 6 missed trips.

Here is the output table from SQL:

	route_id	missed_trips
1	132	6
2	21	6
3	85	6