

1. АНАЛИЗ. ИЗУЧАЕМ ЗАКОНОМЕРНОСТИ В ДАННЫХ

Задание 4.1

```
SELECT a.city
FROM dst_project.airports a
GROUP BY a.city
HAVING count(a.city) > 1
```

Задание 4.2

Вопрос 1. Таблица рейсов содержит всю информацию о прошлых, текущих и запланированных рейсах. Сколько всего статусов для рейсов определено в таблице?

```
SELECT count(DISTINCT f.status)
FROM dst_project.flights f
```

Вопрос 2. Какое количество самолетов находятся в воздухе на момент среза в базе (статус рейса «самолёт уже вылетел и находится в воздухе»).

```
SELECT count(f.flight_id)
FROM dst_project.flights f
WHERE f.status = 'Departed'
```

Вопрос 3. Места определяют схему салона каждой модели. Сколько мест имеет самолет модели (*Boeing 777-300*)?

```
SELECT count(s.seat_no)
FROM dst_project.seats s
WHERE s.aircraft_code = '773'
```

Вопрос 4. Сколько состоявшихся (фактических) рейсов было совершено между 1 апреля 2017 года и 1 сентября 2017 года?

```
SELECT count(f.flight_id)
FROM dst_project.flights f
WHERE f.actual_arrival BETWEEN '2017-04-01' AND '2017-09-01'
AND f.status = 'Arrived'
```

Задание 4.3

Вопрос 1. Сколько всего рейсов было отменено по данным базы?

```
SELECT count(f.flight_id)
```

```
FROM dst_project.flights f
WHERE f.status = 'Cancelled'
```

Вопрос 2. Сколько самолетов моделей типа *Boeing*, *Sukhoi Superjet*, *Airbus* находится в базе авиаперевозок?

Boeing:

```
SELECT count(a.aircraft_code)
FROM dst_project.aircrafts a
WHERE a.model LIKE 'Boeing%'
```

Sukhoi Superjet:

```
SELECT count(a.aircraft_code)
FROM dst_project.aircrafts a
WHERE a.model LIKE 'Sukhoi%'
```

Airbus:

```
SELECT count(a.aircraft_code)
FROM dst_project.aircrafts a
WHERE a.model LIKE 'Airbus%'
```

Вопрос 3. В какой части (частях) света находится больше аэропортов?

```
SELECT SPLIT_PART (a.timezone, '/', 1) AS continent,
       count(a.airport_code) AS airport_count
FROM dst_project.airports a
GROUP BY continent
```

Вопрос 4. У какого рейса была самая большая задержка прибытия за все время сбора данных? Введите *id* рейса (*flight_id*).

```
SELECT f.flight_id
FROM dst_project.flights f
WHERE (f.actual_arrival - f.scheduled_arrival) =
      (SELECT max(f.actual_arrival - f.scheduled_arrival)
       FROM dst_project.flights f)
```

Задание 4.4

Вопрос 1. Когда был запланирован самый первый вылет, сохраненный в базе данных?

```
SELECT min(f.scheduled_departure)
FROM dst_project.flights f
```

Вопрос 2. Сколько минут составляет запланированное время полета в самом длительном рейсе?

```
SELECT max((extract(epoch
                  FROM (f.scheduled_arrival)) - extract(epoch
                  FROM (f.scheduled_departure))))/60) AS
max_duration_min
FROM dst_project.flights f
```

Вопрос 3. Между какими аэропортами пролегает самый длительный по времени запланированный рейс?

```
SELECT departure_airport,
       arrival_airport
FROM dst_project.flights
ORDER BY scheduled_arrival - scheduled_departure DESC
LIMIT 1
```

Вопрос 4. Сколько составляет средняя дальность полета среди всех самолетов в минутах? Секунды округляются в меньшую сторону (отбрасываются до минут).

```
SELECT avg((extract(epoch
                  FROM (f.scheduled_arrival)) - extract(epoch
                  FROM (f.scheduled_departure))))/60) AS avg_duration
FROM dst_project.flights f
```

Задание 4.5

Вопрос 1. Мест какого класса у *SU9* больше всего?

```
SELECT s.fare_conditions,
       count(s.seat_no) AS seat_count
FROM dst_project.seats s
WHERE s.aircraft_code = 'SU9'
GROUP BY s.fare_conditions
ORDER BY seat_count desc
LIMIT 1
```

Вопрос 2. Какую самую минимальную стоимость составило бронирование за всю историю?

```
SELECT min(b.total_amount) AS min_price
FROM dst_project.bookings b
```

Вопрос 3. Какой номер места был у пассажира с id = 4313 788533?

```
SELECT b.seat_no
FROM dst_project.tickets t
JOIN dst_project.boarding_passes b ON t.ticket_no = b.ticket_no
WHERE t.passenger_id = '4313 788533'
```

Задание 5.1

Вопрос 1. Анапа — курортный город на юге России. Сколько рейсов прибыло в Анапу за 2017 год?

```
SELECT count(f.flight_id) AS flights_num_2017
FROM dst_project.flights f
WHERE
  f.arrival_airport =
    (SELECT a.airport_code AS code
     FROM dst_project.airports AS a
     WHERE a.city = 'Anapa')
  AND date_trunc ('year', f.actual_arrival) = '2017/01/01'
```

Вопрос 2. Сколько рейсов из Анапы вылетело зимой 2017 года?

Вариант 1: дек 2016, янв 2017, фев 2017

Ответ: 16930

```
SELECT count(f.flight_id) AS winter_flights_num
FROM dst_project.flights f
WHERE
  f.departure_airport =
    (SELECT a.airport_code AS code
     FROM dst_project.airports AS a
     WHERE a.city = 'Anapa')
  AND -- блок ниже выводит данные за январь и февраль 2017
    (
      date_trunc ('year', f.actual_departure) = '2017/01/01'
      AND date_trunc ('month', f.actual_departure) in ('2017/01/01', '2017/02/01')
    )
  OR -- блок ниже выводит данные за декабрь 2016
```

```
(
  date_trunc('year', f.actual_departure) = '2016/01/01'
  AND date_trunc('month', f.actual_departure) = '2016/12/01'
)
AND
  f.status = 'Arrived'
```

Вариант 2: дек 2017, янв 2017, фев 2017

Ответ: 127

! Этот ответ приняла платформа

```
SELECT count(f.flight_id) AS winter_flights_num
FROM dst_project.flights f
WHERE
  f.departure_airport =
    (SELECT a.airport_code AS code
     FROM dst_project.airports AS a
     WHERE a.city = 'Anapa')
  AND -- блок ниже выводит данные за январь, февраль, декабрь 2017
    date_trunc('month', f.actual_departure) in ('2017-01-01',
                                                '2017-02-01', '2017-12-01')
  AND
    f.status = 'Arrived'
```

Интересно: в декабре 2017 количество перелетов – 0.

Вопрос 3. Посчитайте количество отмененных рейсов из Анапы за все время.

```
SELECT count(*)
FROM dst_project.flights AS f
WHERE f.departure_airport = 'AAQ'
      AND status = 'Cancelled'
```

Вопрос 4. Сколько рейсов из Анапы не летают в Москву?

```
SELECT count(*)
FROM dst_project.flights AS f
WHERE f.departure_airport = 'AAQ'
      AND f.arrival_airport not in
        (SELECT DISTINCT a.airport_code
         FROM dst_project.airports a
         WHERE a.city = 'Moscow' )
```

Вопрос 5. Какая модель самолета летящего на рейсах из Анапы имеет больше всего мест?

```
WITH q AS
  (SELECT f.aircraft_code,
         count(*)
   FROM dst_project.flights AS f
  WHERE f.departure_airport = 'AAQ'
  GROUP BY f.aircraft_code
  ORDER BY count(*) DESC
  LIMIT 1)

SELECT a.model
FROM dst_project.aircrafts a
JOIN q ON a.aircraft_code = q.aircraft_code
```

2. ДАТАСЕТ. ФОРМИРУЕМ ЗАПРОС ДЛЯ АНАЛИТИКИ.

Заметка: зимний период можно считать по-разному:

- Декабрь 2016, январь и февраль 2017
- Декабрь, январь, февраль 2017.

Выбрала первый вариант, т.к. у второго варианта аномальное количество перелетов в декабре (0).

Запрос для формирования датасета.

```
WITH flight_info AS
  (SELECT f.flight_id,
         f.flight_no,
         f.departure_airport,
         f.arrival_airport,
         (extract(epoch
                   FROM (f.scheduled_arrival::TIMESTAMP)) - extract(epoch
                               FROM (f.scheduled_departure::TIMESTAMP)))/60 AS
flight_duration_min, -- плановая длительность полета в минутах
         (extract(epoch
                   FROM (f.actual_arrival::TIMESTAMP)) - extract(epoch
                               FROM (f.scheduled_arrival::TIMESTAMP)))/60 AS
flight_delay_min, -- реальное опоздание рейса в минутах
         f.aircraft_code,
         to_char(f.scheduled_departure, 'HH24:MI') AS flight_time, -- достает из даты вылета
часы и минуты
```

```

        to_char(f.scheduled_departure, 'D') AS flight_dow, -- достаем день недели из даты
вылета (от 1 до 7)
        extract(MONTH
        FROM f.scheduled_departure) AS month_2017 -- для удобства достаем номер
месяца для анализа
FROM dst_project.flights AS f
WHERE f.departure_airport = 'AAQ'
AND (date_trunc('month', f.scheduled_departure) IN ('2017-01-01',
        '2017-02-01',
        '2016-12-01')) -- берем период декабрь 2016, янв и фев 2017
AND status not in ('Cancelled')), -- в анализе исключаем рейсы, которые были отменены

```

```

tickets_info AS
(SELECT t_f.flight_id,
        count(t_f.ticket_no) AS passenger_count,
        sum(t_f.amount) AS flight_income
FROM dst_project.ticket_flights AS t_f
GROUP BY t_f.flight_id),

```

```

aircraft_info AS
(SELECT a.aircraft_code,
        a.model,
        count(*) AS seats_count
FROM dst_project.aircrafts a
JOIN dst_project.seats s ON a.aircraft_code = s.aircraft_code
GROUP BY a.aircraft_code)

```

```

SELECT
        f.flight_id,
        f.flight_no,
        f.departure_airport,
        f.arrival_airport,
        f.month_2017,
        f.flight_time,
        f.flight_dow,
        f.flight_duration_min,
        f.flight_delay_min,
        a.model,
        a.seats_count,
        t.passenger_count,
        1.0*t.passenger_count/a.seats_count AS occupancy, -- считаем коэффициент
заполняемости рейса
        t.flight_income
FROM flight_info f

```

```
LEFT JOIN tickets_info t ON f.flight_id = t.flight_id
JOIN aircraft_info a ON a.aircraft_code = f.aircraft_code
ORDER BY f.flight_no, f.flight_dow, t.flight_income
```