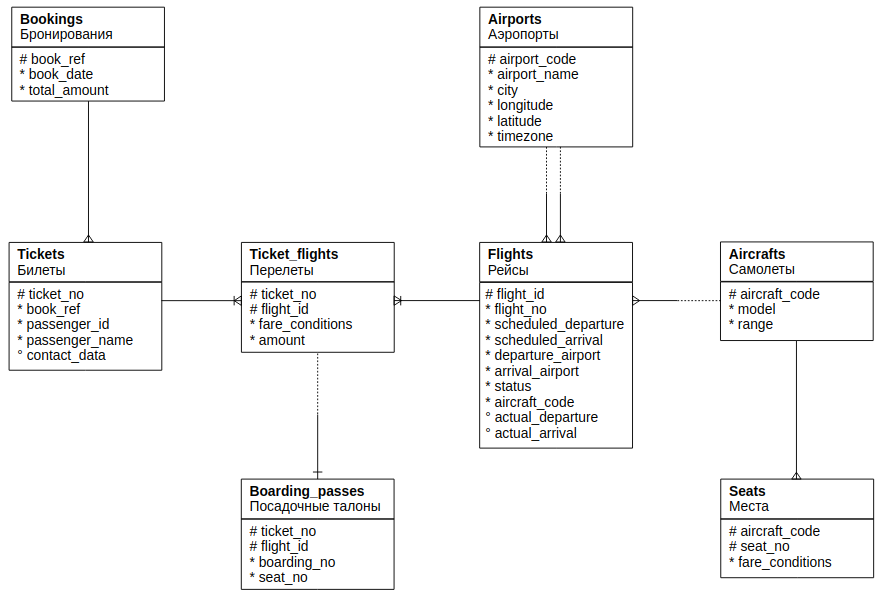
**Проект 4 «Авиарейсы без потерь»**

Цель проекта: извлечь из базы данных информацию, позволяющую проанализировать рентабельность рейсов из Анапы в зимнее время 2017 г., и дать рекомендации по выявленным низкодоходным рейсам.

Схема данных **BOOKINGS** содержит следующие таблицы:



Задание 4.1

База данных содержит список аэропортов практически всех крупных городов России. В большинстве городов есть только один аэропорт. Исключение составляет:

**Moscow, Ulyanovsk**

Код:

**Select**

**a.city, count(a.airport\_code)**

**from**

**dst\_project.AIRPORTS as a**

**group by**

**a.city**

**order by**

**count(a.airport\_code) desc**

Задание 4.2

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

**6**

**select**

**count(distinct f.status)**

**from**

**dst\_project.FLIGHTS as f**

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

**58**

**Select**

**f.status, count(f.flight\_id)**

**from**

**dst\_project.FLIGHTS as f**

**group by**

**f.status**

\*\*Т.к. заранее известно статусов всего 6, посмотрели сразу на все статусы, выявили самые распространенные, для ответа на вопрос взяли количество рейсов со статусом **Departed.**

**Вопрос 3.** Места определяют схему салона каждой модели. Сколько мест имеет самолет модели 773 (*Boeing 777-300*)?

**402**

**Select**

**c.model, count(s.seat\_no)**

**from**

**dst\_project.aircrafts as c**

**join dst\_project.seats as s on c.aircraft\_code=s.aircraft\_code**

**where**

**c.aircraft\_code='773'**

**group**

**by c.model**

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

**74227**

**select**

**count(f.flight\_id)**

**from**

**dst\_project.FLIGHTS as f**

**where**

**f.actual\_arrival between '2017-04-01' and '2017-09-01' and f.status ='Arrived' and f.status !='Cancelled'**

Здесь и далее *состоявшийся рейс* означает, что он не отменён, и самолёт прибыл в пункт назначения.

Задание 4.3

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

**437**

**select**

**count(\*)**

**from**

**dst\_project.FLIGHTS as f**

**where**

**f.status ='Cancelled'**

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

*Boeing:*



*Sukhoi Superjet:*



*Airbus:*



**select**

**count(c.aircraft\_code)**

**from**

**dst\_project.aircrafts as c**

**where**

**c.model like 'Airbus%' -- or c.model like 'Sukhoi Superjet%' ...**

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

Europe

Australia

**Europe, Asia**

Asia

**select case**

**when a.timezone like 'Eur%' then 'Europe'**

**when a.timezone like 'Asi%' then 'Asia'**

**when a.timezone like 'Austr%' then 'Australia'**

**else 'others'**

**end**

**as timezone\_part, count(\*)**

**from**

**dst\_project.AIRPORTS as a**

**group by 1**

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

**157571**

**select**

**f.flight\_id, (f.actual\_arrival-f.scheduled\_arrival) as arrival\_long**

**from**

**dst\_project.FLIGHTS as f**

**where**

**f.actual\_arrival is not null**

**order by 2**

**limit 1**

Задание 4.4

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

**14.08.2016**

**select**

**f.scheduled\_departure**

**from**

**dst\_project.FLIGHTS as f**

**order by 1**

**limit 1**

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

**530**

**select**

**date\_part('hour',f.scheduled\_arrival-f.scheduled\_departure)\*60+date\_part('minute', f.scheduled\_arrival-f.scheduled\_departure) as scheduled\_fl\_tyme**

**from**

**dst\_project.FLIGHTS as f**

**order by 1 desc**

**limit 1**

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

DME - UUS

DME - AAQ

DME - PCS

SVO - UUS

**select**

**f.departure\_airport, f.arrival\_airport, date\_part('hour',f.scheduled\_arrival-f.scheduled\_departure)\*60+date\_part('minute', f.scheduled\_arrival-f.scheduled\_departure) as scheduled\_fl\_tyme**

**from**

**dst\_project.FLIGHTS as f**

**order by 3 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(date\_part('hour',f.actual\_arrival-f.actual\_departure)\*60+date\_part('minute', f.scheduled\_arrival-f.scheduled\_departure)) as avg\_fly**

**from**

**dst\_project.FLIGHTS as f**

Задание 4.5

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

**Economy**

**select**

**s.fare\_conditions, count(s.seat\_no), s.aircraft\_code**

**from**

**dst\_project.seats as s**

**where**

**s.aircraft\_code like '%SU9%'**

**group**

**by s.fare\_conditions, s.aircraft\_code**

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

**3400**

**select**

**min(b.total\_amount)**

**from**

**dst\_project.BOOKINGS as b**

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

**2A**

**select**

**bp.seat\_no**

**from**

**dst\_project.boarding\_passes as bp**

**join dst\_project.tickets as t on bp.ticket\_no=t.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?

**486**

**select**

**count(\*)**

**from**

**dst\_project.FLIGHTS as f**

**join dst\_project.airports as a on f.arrival\_airport=a.airport\_code**

**where**

**date\_part('year',f.actual\_arrival)=2017**

**and f.status='Arrived'**

**and f.arrival\_airport in**

**(**

**select**

**a.airport\_code**

**from**

**dst\_project.airports as a**

**where**

**a.city like 'Anapa'**

**)**

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

**127**

**select**

**count(distinct f.scheduled\_departure)**

**from**

**dst\_project.FLIGHTS as f**

**join dst\_project.airports as a on f.departure\_airport=a.airport\_code**

**where**

**date\_part('year',f.actual\_departure)=2017**

**and (date\_part('month', f.actual\_departure)=1 or date\_part('month', f.actual\_departure)=2 or date\_part('month', f.actual\_departure)=12)**

**and f.departure\_airport in**

**(**

**select**

**a.airport\_code**

**from**

**dst\_project.airports as a**

**where**

**a.city like 'Anapa'**

**)**

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

**1**

**select**

**count(\*)**

**from**

**dst\_project.FLIGHTS as f**

**join dst\_project.airports as a on f.arrival\_airport=a.airport\_code**

**where**

**f.status='Cancelled' and f.departure\_airport in**

**(**

**select**

**a.airport\_code**

**from**

**dst\_project.airports as a**

**where**

**a.city like 'Anapa'**

**)**

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

**453**

**Select**

**count(f.flight\_no)**

**from**

**dst\_project.FLIGHTS as f**

**where**

**f.departure\_airport='AAQ'**

**and (f.arrival\_airport !='SVO' and f.arrival\_airport !='VKO' and f.arrival\_airport !='DME')**

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

**Boeing 737-300**

**with cs as**

**(**

**select**

**s.aircraft\_code, count(s.seat\_no) as seat\_count**

**from**

**dst\_project.seats as s**

**group by**

**s.aircraft\_code**

**)**

**select**

**c.model, cs.seat\_count**

**from**

**dst\_project.FLIGHTS as f**

**join cs on f.aircraft\_code=cs.aircraft\_code**

**join dst\_project.aircrafts as c on f.aircraft\_code=c.aircraft\_code**

**where**

**f.departure\_airport='AAQ'**

**group by c.model, cs.seat\_count**

**Выгрузка итогового датасета**

**WITH**

**table\_city\_dep --finding departure city**

**AS (**

**SELECT**

**f.flight\_id,f.departure\_airport, a.city as departure\_city**

**FROM**

**dst\_project.flights f**

**JOIN dst\_project.AIRPORTS as a on f.departure\_airport=a.airport\_code**

**),**

**table\_city\_arr --finding arrival city**

**AS (**

**SELECT**

**f.flight\_id,f.arrival\_airport, a.city as arrival\_city**

**FROM**

**dst\_project.flights f**

**JOIN dst\_project.AIRPORTS as a on f.arrival\_airport=a.airport\_code**

**),**

**table\_fl\_duration --finding flight duration**

**AS (**

**SELECT**

**f.flight\_id, date\_part('hour',f.actual\_arrival-f.actual\_departure)\*60+date\_part('minute', f.actual\_arrival-f.actual\_departure) AS flight\_duration\_minute**

**FROM**

**dst\_project.flights f**

**),**

**table\_all\_seat --count all seats for flight**

**AS (**

**SELECT**

**f.flight\_id, count(s.seat\_no) as all\_seats**

**FROM**

**dst\_project.flights f**

**LEFT JOIN dst\_project.seats as s on f.aircraft\_code=s.aircraft\_code**

**GROUP BY f.flight\_id**

**),**

**table\_eco\_seat --count economy seats for flight**

**AS (**

**SELECT**

**f.flight\_id, count(s.seat\_no) as eco\_seats**

**FROM**

**dst\_project.flights f**

**LEFT JOIN dst\_project.seats as s on f.aircraft\_code=s.aircraft\_code**

**WHERE s.fare\_conditions='Economy'**

**GROUP BY f.flight\_id**

**),**

**table\_com\_seat --count Comfort seats for flight**

**AS (**

**SELECT**

**f.flight\_id, count(s.seat\_no) as com\_seats**

**FROM**

**dst\_project.flights f**

**LEFT JOIN dst\_project.seats as s on f.aircraft\_code=s.aircraft\_code**

**WHERE s.fare\_conditions='Comfort'**

**GROUP BY f.flight\_id**

**),**

**table\_bus\_seat --count Business seats for flight**

**AS (**

**SELECT**

**f.flight\_id, count(s.seat\_no) as bus\_seats**

**FROM**

**dst\_project.flights f**

**LEFT JOIN dst\_project.seats as s on f.aircraft\_code=s.aircraft\_code**

**WHERE s.fare\_conditions='Business'**

**GROUP BY f.flight\_id**

**),**

**table\_ticket\_all --count tickets and amount for flight**

**AS (**

**SELECT**

**tf.flight\_id, count(tf.ticket\_no) as actual\_tickets\_count, sum(tf.amount) flight\_tickets\_amount**

**FROM**

**dst\_project.ticket\_flights as tf**

**GROUP BY tf.flight\_id**

**),**

**table\_ticket\_eco --count Economy tickets**

**AS (**

**SELECT**

**tf.flight\_id, count(tf.ticket\_no) as actual\_tickets\_eco, sum(tf.amount) eco\_tickets\_amount**

**FROM**

**dst\_project.ticket\_flights as tf**

**WHERE tf.fare\_conditions='Economy'**

**GROUP BY tf.flight\_id**

**),**

**table\_ticket\_com --count Comfort tickets**

**AS (**

**SELECT**

**tf.flight\_id, count(tf.ticket\_no) as actual\_tickets\_com, sum(tf.amount) com\_tickets\_amount**

**FROM**

**dst\_project.ticket\_flights as tf**

**WHERE tf.fare\_conditions='Comfort'**

**GROUP BY tf.flight\_id**

**),**

**table\_ticket\_bus --count Business tickets**

**AS (**

**SELECT**

**tf.flight\_id, count(tf.ticket\_no) as actual\_tickets\_bus, sum(tf.amount) bus\_tickets\_amount**

**FROM**

**dst\_project.ticket\_flights as tf**

**WHERE tf.fare\_conditions='Business'**

**GROUP BY tf.flight\_id**

**)**

**SELECT**

**f.flight\_id, table\_city\_dep.departure\_city, table\_city\_arr.arrival\_city,**

**f.actual\_departure, table\_fl\_duration.flight\_duration\_minute, c.model,**

**table\_all\_seat.all\_seats, table\_eco\_seat.eco\_seats, table\_com\_seat.com\_seats, table\_bus\_seat.bus\_seats,**

**table\_ticket\_all.actual\_tickets\_count, table\_ticket\_all.flight\_tickets\_amount,**

**table\_ticket\_eco.actual\_tickets\_eco, table\_ticket\_eco.eco\_tickets\_amount,**

**table\_ticket\_com.actual\_tickets\_com, table\_ticket\_com.com\_tickets\_amount,**

**table\_ticket\_bus.actual\_tickets\_bus, table\_ticket\_bus.bus\_tickets\_amount,**

**table\_ticket\_all.actual\_tickets\_count/table\_all\_seat.all\_seats as filled\_seats,**

**table\_ticket\_eco.actual\_tickets\_eco/table\_eco\_seat.eco\_seats as fill\_eco\_seats,**

**table\_ticket\_com.actual\_tickets\_com/table\_com\_seat.com\_seats as fill\_com\_seats,**

**table\_ticket\_bus.actual\_tickets\_bus/table\_bus\_seat.bus\_seats as fill\_bus\_seats**

**FROM dst\_project.flights f**

**LEFT JOIN table\_city\_dep on f.flight\_id=table\_city\_dep.flight\_id**

**LEFT JOIN table\_city\_arr on f.flight\_id=table\_city\_arr.flight\_id**

**LEFT JOIN table\_fl\_duration on f.flight\_id=table\_fl\_duration.flight\_id**

**LEFT JOIN dst\_project.aircrafts as c on f.aircraft\_code=c.aircraft\_code**

**LEFT JOIN table\_all\_seat on f.flight\_id=table\_all\_seat.flight\_id**

**LEFT JOIN table\_eco\_seat on f.flight\_id=table\_eco\_seat.flight\_id**

**LEFT JOIN table\_com\_seat on f.flight\_id=table\_com\_seat.flight\_id**

**LEFT JOIN table\_bus\_seat on f.flight\_id=table\_bus\_seat.flight\_id**

**LEFT JOIN table\_ticket\_all on f.flight\_id=table\_ticket\_all.flight\_id**

**LEFT JOIN table\_ticket\_eco on f.flight\_id=table\_ticket\_eco.flight\_id**

**LEFT JOIN table\_ticket\_com on f.flight\_id=table\_ticket\_com.flight\_id**

**LEFT JOIN table\_ticket\_bus on f.flight\_id=table\_ticket\_bus.flight\_id**

**WHERE f.departure\_airport = 'AAQ'**

**AND (date\_trunc('month', f.scheduled\_departure) in ('2017-01-01', '2017-02-01', '2016-12-01'))**

**AND f.status not in ('Cancelled')**