



Проект 4

Авиарейсы без потерь.

Целиков Сергей

Группа DST-PRO-11

Структура датасета:

- flight_id** - Идентификатор рейса. Характеризует конкретный рейс. Уникальное поле
- **flight_no** - Номер рейса. Определяет авиакомпанию. Необходим для дальнейшей возможной группировки
- **departure_airport** – IATA код аэропорта вылета. Информационное поле. Для данного датасета одинаковое для всех строк. Можно удалить при постобработке.
- **departure_airport_name** – Название аэропорта вылета. Информационное поле. Для данного датасета одинаковое для всех строк. Можно удалить при постобработке.
- **departure_city** – Название города вылета. Информационное поле. Для данного датасета одинаковое для всех строк. Можно удалить при постобработке
- **departure_latitude** – Географическая широта аэропорта вылета. Может использоваться в дальнейшем для определения дальности перелета
- **departure_longitude** – Географическая долгота аэропорта вылета. Может использоваться в дальнейшем для определения дальности перелета
- **arrival_airport** - IATA код аэропорта прилета. Информационное поле. Может использоваться для группировки данных при постобработке.
- **arrival_airport_name** - Название аэропорта прилета. Информационное поле. Может понадобиться для связи с внешними данными или его можно удалить при постобработке.
- **arrival_city** – Название города прилета. Информационное поле. Может понадобиться для связи с внешними данными или его можно удалить при постобработке

Структура датасета (продолжение):

arrival_latitude - Географическая широта аэропорта прилета. Может использоваться в дальнейшем для определения дальности перелета

- **arrival_longitude** - Географическая долгота аэропорта прилета. Может использоваться в дальнейшем для определения дальности перелета
- **scheduled_departure** – Плановое время вылета. Может использоваться в постобработке.
- **scheduled_arrival** – Плановое время прилета. Может использоваться в постобработке.
- **actual_departure** – Реальное время вылета рейса. Может использоваться в постобработке.
- **actual_arrival** – Реальное время прилета рейса. Может использоваться в постобработке
- **dow** – День недели (0 – воскресенье, 6 – суббота). Может использоваться для оценки прибыльности рейсов по дням недели
- **scheduled_len** – Плановая длительность рейса в минутах. Может использоваться для оценки прибыльности рейсов
- **actual_len** – Реальная длительность рейса в минутах. Может использоваться для оценки прибыльности рейсов
- **departure_delay** – Задержка отправления рейса в минутах относительно плановой. Может использоваться для оценки прибыльности рейсов
- **arrival_delay** – Задержка прибытия рейса в минутах относительно плановой. Может использоваться для оценки прибыльности рейсов

Структура датасета (продолжение):

aircraft_code – Код модели воздушного судна. Может использоваться для группировки при анализе

- **model** – Название модели воздушного судна. Информационное поле. Может понадобиться для связи с внешними данными
- **range** – Максимальная дальность полета воздушного судна в км. Может понадобиться для оценки рентабельности использования данного типа воздушного судна на рейсе.
- **economy_seats** – Общее количество мест эконом-класса для данного судна. Необходимо для анализа рентабельности рейса.
- **business_seats** – Общее количество мест бизнес-класса для данного судна. Необходимо для анализа рентабельности рейса.
- **economy_seats_sold** - Количество проданных мест эконом-класса для данного рейса. Необходимо для анализа рентабельности рейса.
- **economy_revenue** – Выручка от продажи мест эконом-класса для данного рейса. Необходимо для анализа рентабельности рейса.
- **business_seats_sold** - Количество проданных мест бизнес-класса для данного рейса. Необходимо для анализа рентабельности рейса.
- **business_revenue** – Выручка от продажи мест бизнес-класса для данного рейса. Необходимо для анализа рентабельности рейса.

Данные, которые можно добавить

При проведении исследования базы данных можно было бы получить выборку из таблицы билетов и полетов (Ticket_flights) в привязке к таблицам посадочных талонов (Boarding_passes), бронирований (Bookings) и перелетов (Flights). Это позволило бы получить список всех билетов и бронирований для исследуемых рейсов.

Подобная выборка могла бы дать следующую дополнительную информацию:

- За сколько дней до вылета осуществляется бронирование для конкретных пассажиров (является ли полет запланированным или спонтанным).
- Количество человек в бронировании (летит ли человек один или с семьёй/друзьями/коллегами)
- Сравнив проданные билеты и посадочные талоны, можно посчитать пассажиров, которые выкупили билеты, но не явились на посадку. Скорее всего, для этих пассажиров будут произведены возвраты средств и это уменьшит выручку от продажи билетов на рейс.
- Можно также провести из базы выборку обратных рейсов (с прилетом в город Анапа) в указанный интервал времени (+- один месяц), оценить также экономику этих рейсов и посчитать количество пассажиров, воспользовавшихся рейсами в обе стороны.
- Можно посчитать пассажиров, которые летают часто. Возможно, они пользовались данным направлением несколько раз за указанный период? Или за более широкий период времени? Какие другие направления они предпочитают?
- По базе посадочных талонов можно провести анализ предпочитаемых посадочных мест для разных пассажиров.
- Существуют ли альтернативные (стыковочные) варианты перелетов между городами. Какая их длительность/стоимость?

К сожалению, данный анализ практически невозможен из-за низкой производительности сервера, что делает затруднительным получение выборки в разумный период времени.



Данные, которые можно добавить из ВНЕШНИХ ИСТОЧНИКОВ

- Расстояние между аэропортами (можно посчитать используя географические координаты)
- Стоимость горючего для каждого типа самолета и его потребление на час полета.
- Аэропортовые сборы с авиакомпаний и пассажиров
- Стоимость лизинга данного типа самолета.
- Количество членов экипажа, расходы авиакомпании в ФОТ для экипажей.
- Стоимость проживания экипажей в гостиницах (если обратный рейс приходится на следующие сутки)
- Возможность выбора альтернативных типов воздушных судов для рейса.

Оценка прибыльности рейса

Для оценки прибыльности рейса можно использовать следующие данные, непосредственно присутствующие в датасете:

- **economy_revenue + business_revenue** - сумма двух этих показателей дает понимание дохода, полученного от продажи билетов на данный рейс.
- **economy_seats, business_seats, economy_seats_sold, business_seats_sold** – данные показатели позволяют посчитать загрузженность рейса, то есть соотношение полученного дохода (и количества проданных билетов) к максимально возможному доходу.
- Также для оценки дохода можно использовать прочие данные, отсутствующие в датасете: доход от перевозки негабаритного и дополнительного багажа, от предоставления дополнительных платных услуг.
- Для определения расходов на рейс необходимо суммировать стоимость всех расходных факторов, таких как стоимость потраченного на рейс горючего, аэропортовые сборы, лизинговые платежи за эксплуатацию воздушного судна, фонд оплаты труда персонала, операционные расходы и т. д. Часть из этих данных (стоимость горючего, сборы) достаточно просто соотнести с конкретным рейсом. Остальные, такие как ФОТ и лизинговые платежи, необходимо распределять на все рейсы авиакомпании.
- Вычитая расходы на рейс из полученных доходов, можно оценить прибыльность рейса.
- Большая часть из этих данных недоступна, что не дает возможности корректно оценить прибыльность рейсов.

ВЫВОДЫ

Анализ полученного датасета предполагалось провести на уровне Advanced. К сожалению, в подобном анализе нет большого смысла ввиду недостаточности информации в выборке и технической невозможности проведения выборок по базе билетов.

- Нет информации о рейсах в Новокузнецк (количество и стоимость проданных билетов) - их приходится исключать из анализа.
- Исследовать прибыль невозможно, так как нет данных о расходах. Предполагается отталкиваться от стоимости топлива. Но это не единственный пункт расходов, а лишь один из многих. Тем более, что нет информации о расходах топлива на конкретном рейсе, а расстояние или длительность рейса дают очень неточную характеристику по потреблению топлива. Стоимость топлива меняется динамически во времени и зависит от цен на нефть, аэропорта заправки, сервисной компании и т. д. Расход топлива сильно зависит от погодных условий (направление и сила ветра). Расход топлива не равномерен во время рейса и (в случае коротких рейсов, а это наш вариант) маневрирование на земле, взлет и посадка вносят очень сильный вклад в общий расход топлива.
- В нашем случае, когда есть всего два рейса с приблизительно одинаковой загруженностью, имело бы смысл исследовать расписание рейсов с возможной его оптимизацией. Но здесь опять недостаточно информации для проведения анализа. Исходя из того, что Анапа - маленький курортный город, логично предположить, что большинство пассажиров - это не местные жители, а гости из Новокузнецка, Белгорода, Москвы. Таким образом, нужно в первую очередь анализировать прямые рейсы из этих городов, а не обратные (вывозные).

Исходя из вышеизложенного, проведение детального анализа рентабельности рейсов не представляется возможным. Необходимо обогащение датасета большим количеством внешних данных для получения адекватной оценки рентабельности рейсов.