



**СБЕРБАНК
ТЕХНОЛОГИИ**

Обнаружение аномалий в данных авиасообщения

**Команда:
Бабушкина Татьяна
Дарбинян Оганес
Малышевский Сергей
Никонов Иван**

Были сформулированы 5 гипотез

Реализованы программные инструменты для проверки данных гипотез

В результате:

- по одной гипотезы не найдено прецедентов
- по остальным гипотезам сформированы списки подозрительных пассажиров, для которых требуются дополнительные проверки из других источников

Краткие итоги:

- Программный продукт позволил определить подозрительных пассажиров
- Полученный функционал будет в дальнейшем полезен для для обнаружения подозрительной активности в авиасообщении

Постановка задачи

Цель:

предложить свой начальный продукт для поиска подозрительных пассажиров авиатранспорта

Этапы решения:

1. Получить наиболее объемный датасет, полученный с помощью объединения исходных выгрузок
2. Сформулировать гипотезы по определению аномальных данных
3. Предложить модель проверки полученных гипотез
4. Выявить подозрительных пассажиров

Входные данные

| Данные | Содержание |
|--------------------|---|
| PointzAggregator | Данные о перемещения пассажиров внутри аэропорта |
| SkyTeam | Информация о перелетах авиакомпании из SkyTeam |
| Boarding | Данные о пассажирах |
| SkyTeamExchange | Шина обмена информацией для группы компаний альянса |
| YourBoardingPass | Таблица с информацией из посадочных талонов |
| BoardingData | Информация с системы посадки на самолеты |
| FrequentFlyerForum | Информация с форума на сайте авиакомпании |

Проверенные гипотезы

1. Пассажир с билетами в разные места, оформленные на одно время
2. Пассажиры, летающие в паре, не являющиеся родственниками или друзьями
3. Пассажир, летающий преимущественно в один и тот же город
4. Пассажир, который не задерживается в месте прилета надолго
5. Пассажир, у которых маршрут перелетов не целостен

Другие варианты гипотез

1. Пассажир летает с чужими документами
2. Пассажир с аномально большим количеством накопленных миль
3. Обратить внимание на международно значимые даты, посмотреть выделяется ли группа людей, которая всегда пользуется авиаперелетом в эти дни
4. Пассажир едет в одну сторону по одному паспорту, обратно по другому
5. Пассажир с аномально большим количеством накопленных миль

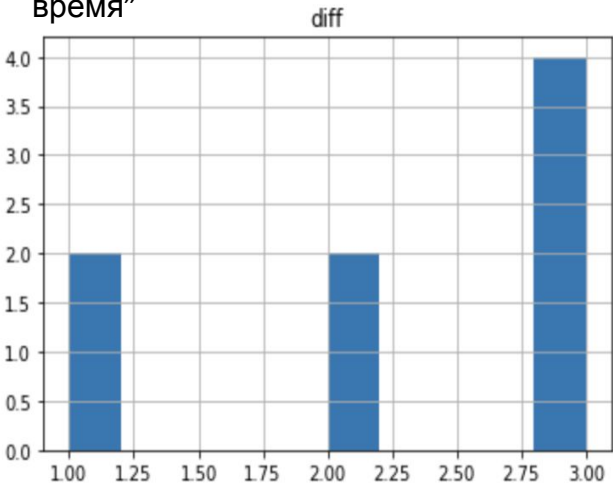
Пассажир с билетами в разные места, оформленные на одно время

В данной гипотезе предполагается, что человек перемещается, используя дубликат реально существующих документов. Возможно данная техника могла бы использоваться, чтобы у человека было алиби и по документам он находился в другом месте. Соответственно необходимо было определить людей находившихся одновременно в разных местах.

Алгоритм:

- 1) В датасете **BoardingData** формируется новое поле, содержащее значение “Номер документа + дата + время”
- 2) Данные группируются и определяются случаи, когда у пассажира больше одного вылета за день
- 3) Вычисляется минимальная разница во времени вылета за день
- 4) Полученные значения сортируются от меньшего к большему
- 5) Проверяются случаи с малой разницей

Результат алгоритма при использовании в качестве значения уникального поля сочетание “Фамилия + Имя + Отчество + дата + время”



| Passenger_and_date | diff |
|---------------------------------------|------|
| GLADKOVA REGINA M. 2017-02-14 | 1 |
| PANTELEEV GRIGORII A. 2017-01-01 | 1 |
| NOVIKOVA NATASHA V. 2017-03-02 | 2 |
| BASOVA VITALINA A. 2017-01-18 | 2 |
| BORODINA IANA A. 2017-03-08 | 3 |
| BULATOVA LILIIA D. 2017-01-02 | 3 |
| MUKHINA NATALIIA SAVELEVNA 2017-01-09 | 3 |
| ZAVIALOVA OLGA I. 2017-01-23 | 3 |

Результат: Случаи перемещения пассажиров по дубликатам паспорта в данной выборке не обнаружены

Пассажиры, летающие в паре, не являющиеся родственниками или друзьями

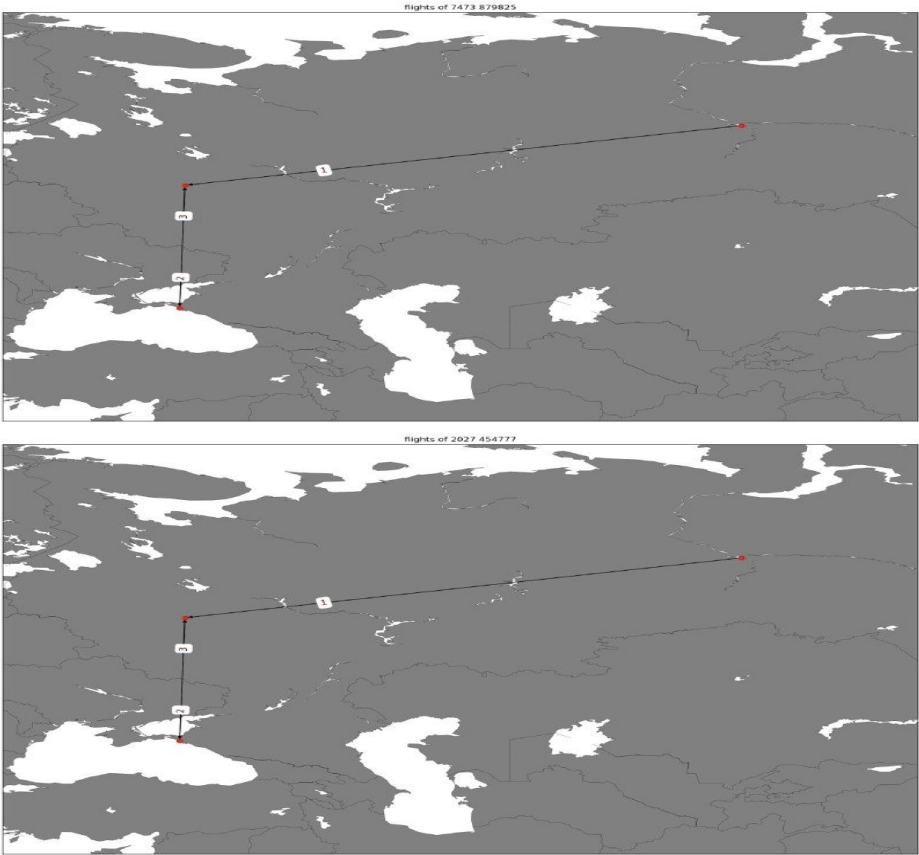
Предполагается, двое мужчин, не имеющих близких родственных связей, перемещаются вместе. При этом, для конспирации, они иногда летят в одно место разными рейсами.

Алгоритм:

- 1) Список пассажиров фильтруется по признаку “мужчины”
- 2) Составляется список перелетов(рейс + дата + время) для каждого человека
- 3) Производится поиск людей, у которых отличаются фамилии и есть несколько совпадающих перелетов
- 4) Производится поиск совпадающих пунктов назначения и перелетов в разные даты для данных людей

| PassengerLastName | PassengerFirstName | PassengerSecondName | PassengerDocument | FlightNumber | FlightDate | FlightTime | Destination |
|-------------------|--------------------|---------------------|-------------------|--------------|------------|------------|-------------|
| GORELOV | LEV | S. | 7473 879825 | SU1145 | 2017-01-10 | 08:00 | Moscow |
| LEONOV | ANDREI | MAKSIMOVICH | 2027 454777 | SU1145 | 2017-01-10 | 08:00 | Moscow |
| LEONOV | ANDREI | MAKSIMOVICH | 2027 454777 | SU1383 | 2017-01-04 | 06:40 | Moscow |
| GORELOV | LEV | SAVVOVICH | 7473 879825 | SU1383 | 2017-01-04 | 06:40 | Moscow |

| PassengerLastName | PassengerFirstName | PassengerSecondName | PassengerDocument | FlightNumber | FlightDate | FlightTime | Destination |
|-------------------|--------------------|---------------------|-------------------|--------------|------------|------------|-------------|
| LEONOV | ANDREI | MAKSIMOVICH | 2027 454777 | SU1144 | 2017-01-08 | 17:15 | Anapa |
| GORELOV | LEV | SAVVOVICH | 7473 879825 | SU1144 | 2017-01-05 | 17:15 | Anapa |



Результат: Найден один случай соответствующий данным условиям. Два пассажира прилетели вместе из Ханты-Мансийска в Москву. Потом в разные дни улетели в Анапу. И вернулись одним рейсом в Москву.

Пассажир, летающий в одно и тоже место

В данной гипотезе предполагается, что есть пассажиры часто летающие в определенные места, при этом это неизвестный курорт или время прилетов нерегулярно и, скорее всего, это не его место работы.

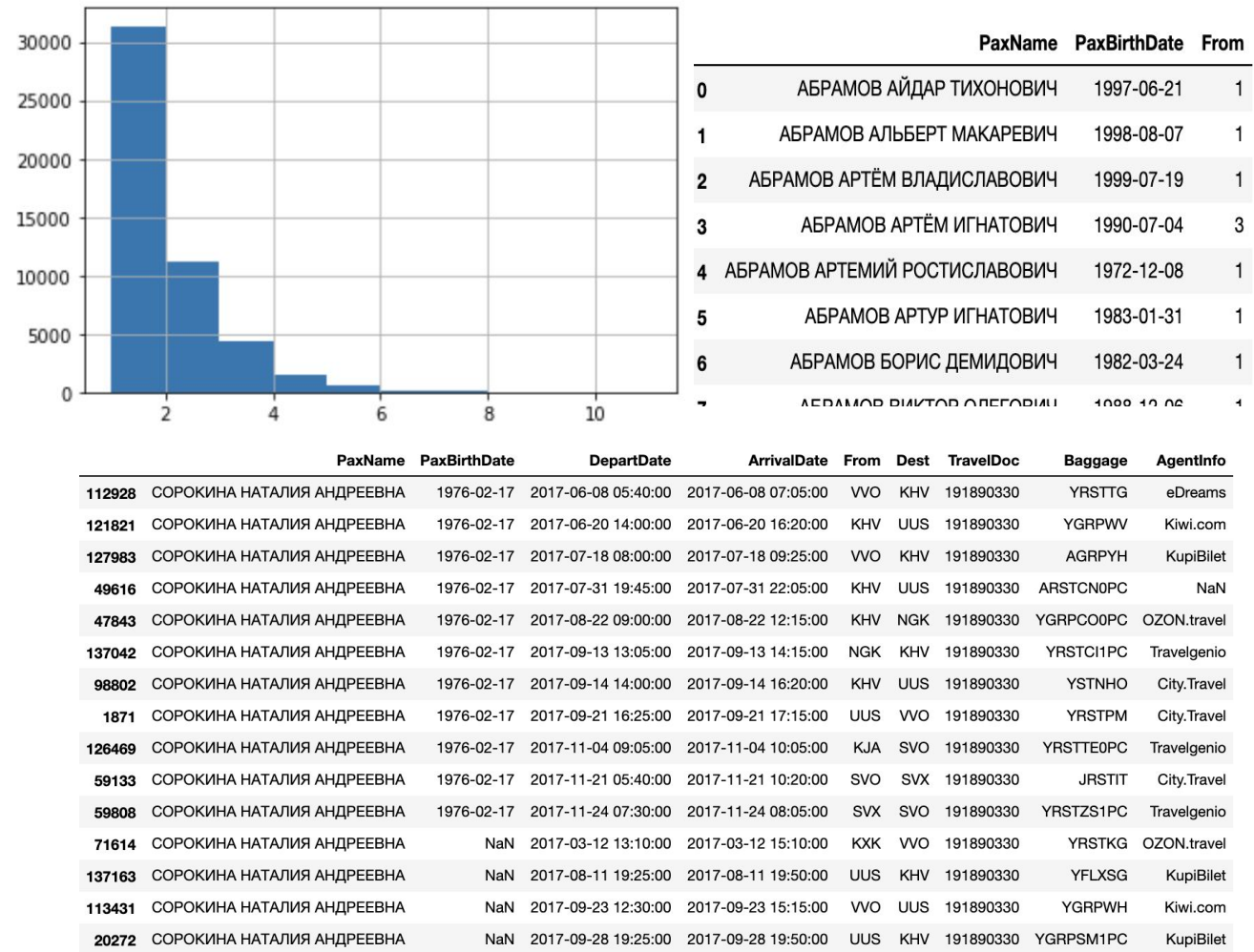
Гистограмма количества перелетов пассажиров
Как правило, у каждого пассажира зафиксировано не более 4 полетов

В целом, мало перелетов, чтобы из них дополнительно фильтровать по особенностям места прилета

Наиболее популярный пассажир (11 перелетов) - Сорокина Наталья Андреевна
По журналу событий ее путь выглядит адекватным, всегда есть информация по багажу и агенству, в котором был куплен билет, номер документа всегда проставлен и одинаков

Результат:
Наиболее активные пассажиры не вызывают подозрений

*(с большим числом перелетов)



Пассажир, который не задерживается в месте прилета надолго

Предполагается, что есть пассажиры, прилетающие в определенное место не задерживаются там надолго и через несколько часов

(меньше суток) возвращаются обратно или следуют в новый пункт назначения.

Алгоритм:

- 1) Список пассажиров (**PaxName + PaxBirthDate** - считается одним пассажиром)
- 2) Сортируем журнал событий по пассажиру и дате отправления
- 3) Добавляем столбец со следующим временем перелета (отправление)
- 4) Считаем длительность перехода (время между текущей датой прилета и следующей дата отправления)
- 5) Выводим список пассажиров с минимальным временем перехода
- 6) Отдельно смотрим на полный журнал событий по перелетам конкретных пассажиров из списка выше

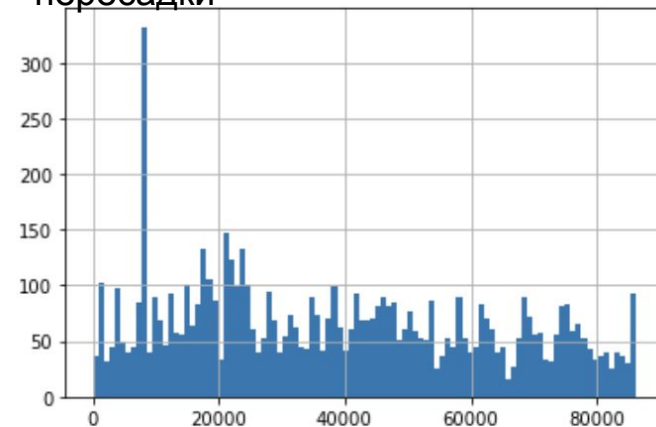
Интересный пассажир: Моргунова Снежана Юрьевна

Не везде проставлена дата рождения, но судя по номеру документа - здесь 2 пассажира. Последняя строчка относится к первым двум, тогда есть информация как пассажир приехал дважды в SVO за небольшой период времени из разных мест (между этими перелетами нет событий). Так же странно, что перелет длится 25 минут из REN (Оренбург) в SVO.

| PaxName | PaxBirthDate | DepartDate | ArrivalDate | From | Dest | TravelDoc | Baggage | AgentInfo |
|---------------------------|--------------|---------------------|---------------------|------|------|------------|-----------|-------------|
| МОРГУНОВА СНЕЖАНА ЮРЬЕВНА | 1970-07-03 | 2017-01-13 13:20:00 | 2017-01-13 14:30:00 | SLY | SVO | 8277862962 | YGRPGC0PC | eDreams |
| МОРГУНОВА СНЕЖАНА ЮРЬЕВНА | 1970-07-03 | 2017-06-29 13:45:00 | 2017-06-29 16:05:00 | SVO | AAQ | 8277862962 | ASTNAL | Kiwi.com |
| МОРГУНОВА СНЕЖАНА ЮРЬЕВНА | 1987-08-30 | 2017-01-13 13:10:00 | 2017-01-13 15:10:00 | SVO | ROV | 8597401686 | YGRPQS | trip.ru |
| МОРГУНОВА СНЕЖАНА ЮРЬЕВНА | 1987-08-30 | 2017-01-19 21:00:00 | 2017-01-20 00:05:00 | ROV | SVO | 8597401686 | YRSTYF1PC | OZON.travel |
| МОРГУНОВА СНЕЖАНА ЮРЬЕВНА | NaN | 2017-01-12 12:20:00 | 2017-01-12 12:55:00 | RTW | SVO | 8597401686 | JFLXZG | Aerobilet |
| МОРГУНОВА СНЕЖАНА ЮРЬЕВНА | NaN | 2017-06-18 19:50:00 | 2017-06-18 20:15:00 | REN | SVO | 8277862962 | YFLXUJ | Kiwi.com |

**Результат: сложно отследить маленькие переходы между перелетами с пересадками
+1 подозрительный пассажир**

Гистограмма длительности пересадки



Пассажир, у которых маршрут перелетов не целостен

В данную категорию можно отнести 2 случая:

1. Пассажир использовал другой вид транспорта, чтобы добраться в другой город и улетал оттуда
2. Пассажир использовал фальшивый документ и его перелет не вошел в историю профиля лояльности

Для нахождения таких подозрительных личностей использовался следующий сценарий:

- В датасетах **FrequentFlyerForum**, **SkyTeamExchange** и **PointzAggregator** отбирались пассажиры по уникальным никнеймам или id.
- Т.к. **PointzAggregator** небольшой, а в **SkyTeamExchange** нет какого-то uid. Они были объединены и записи с пустыми данными были убраны. Логика объединения: данные по uid были сгруппированы по номерам карт лояльности и после, по этим номерам датасеты были объединены.
- У каждого пассажира сортировались полеты по дате и проверялось условие, что город отлета тот же, что и город прилета в предыдущем полете.
- Для большей информативности данные переводятся в картинки с маршрутом пассажира

В датасете **FrequentFlyerForum** не было найдено пассажиров с “плохим” маршрутом, а в объединенном датасете **PointzAggregator** и **SkyTeamExchange** было найдено 158 “подозрительных” личностей. Конечно же это не означает, что данные пассажиры являются шпионами.

Пример подозрительного маршрута.

На этом рисунке можно заметить, что преимущественно маршруты выходят из одного города.

Изначально нет каких-то подозрений, пассажир вылетел из одного города и вернулся оттуда же.

Начиная с 5 рейса пассажир вылетает из одного города, но не возвращается. и так 3 вылета.

flights of 376792514



Программный продукт позволил определить подозрительных пассажиров

Эта информация будет в дальнейшем полезна для поиска необходимых людей (для обнаружения подозрительной активности при перелете?)