



National Research Nuclear  
University MEPhI

# Catch "The Russian Spies" Challenge

Solved by students:



Volyanskiy Julian,  
Ivanova Nina,  
Khudoyarova Anastasia,  
Nikolaeva Elizabeth

# Discovery

## Data sources:

- BoardingData.csv
- FrequentFlyerForum-Profiles.json
- PointzAggregator-AirlinesData.xml
- Sirena-export-fixed.tab
- Skyteam\_Timetable.pdf
- SkyTeam-Exchange.yaml
- YourBoardingPassDotAero.zip



## Discovery

### Spy search criteria:



- Changes travel class frequently
  - Does not have a meal during the flight
- Travels without luggage
  - Travels frequently to different countries



# Data Preparation

File:

.csv

.tab

.xml

.json

.xlsx

.yaml

.pdf

Method or Tool:

remove empty line

LibreOffice Calc

Java

Python

Python parsers

Python

Adobe reader, Kotlin

.CSV

# Model planning

| ForumPersonallInformation |             |
|---------------------------|-------------|
| person_id                 | integer     |
| nickname                  | varchar(30) |
| passenger_document        | varchar(30) |
| sex                       | varchar(30) |
| first_name                | varchar(30) |
| last_name                 | varchar(30) |

| ForumAirport |             |
|--------------|-------------|
| person_id    | integer     |
| airport_id   | integer     |
| name         | varchar(30) |
| abbr         | varchar(30) |
| country      | varchar(30) |

| ForumPersonIdFlight |         |
|---------------------|---------|
| person_id           | integer |
| dep_airport_id      | integer |
| dest_airport_id     | integer |

| ForumPersonIdInfoFlight |             |
|-------------------------|-------------|
| person_id               | integer     |
| date                    | date        |
| codesh                  | varchar(30) |
| flight_number           | varchar(30) |

| ForumPersonIdLoyalty |             |
|----------------------|-------------|
| person_id            | integer     |
| type                 | varchar(30) |
| abbr                 | varchar(30) |
| loyalty_id           | integer     |

| SkyTeamExchange |             |
|-----------------|-------------|
| dep_date        | date        |
| flight_number   | varchar(30) |
| dep_code        | varchar(30) |
| status          | varchar(30) |
| dest_code       | varchar(30) |
| someinfo        | varchar(30) |
| travel_class    | varchar     |
| fare            | varchar(30) |

| BoardingData       |             |
|--------------------|-------------|
| first_name         | varchar(30) |
| second_name        | varchar(30) |
| last_name          | varchar(30) |
| sex                | varchar(30) |
| birth_date         | varchar(30) |
| passenger_document | varchar(30) |
| booking_code       | varchar(30) |
| ticket_number      | varchar(30) |
| baggage            | varchar(30) |
| dep_date           | date        |
| dep_time           | time        |
| flight_number      | varchar(30) |
| codesh             | varchar(30) |
| dest_city          | varchar(30) |

| YourBoardingPassDotAero |              |
|-------------------------|--------------|
| uid                     | integer      |
| sequence                | integer      |
| sex                     | varchar(30)  |
| first_name              | varchar(30)  |
| last_name               | varchar(30)  |
| Y_info                  | varchar(30)  |
| flight_number           | varchar(30)  |
| dep_city                | varchar(30)  |
| dest_city               | varchar(30)  |
| gate                    | varchar(30)  |
| dest_code               | varchar(30)  |
| dep_code                | varchar(30)  |
| dep_date                | date         |
| dep_time                | time         |
| operator_info           | varchar(30)  |
| info                    | varchar(100) |
| seat                    | varchar(30)  |
| PNR                     | varchar(30)  |
| ticket_number           | varchar(30)  |

| SirenaExportFixed |             |
|-------------------|-------------|
| PaxName           | varchar(60) |
| PaxBirthDate      | varchar(10) |
| DepartDate        | date        |
| DepartTime        | time        |
| ArrivalDate       | date        |
| ArrivalTime       | time        |
| FlightCodeSh      | varchar(10) |
| From_             | varchar(3)  |
| Dest              | varchar(3)  |
| Code_eTicket      | varchar(22) |
| TravelDoc         | varchar(11) |
| Seat              | varchar(3)  |
| Meal              | varchar(4)  |
| TrvClis_Fare      | char        |
| Baggage           | varchar(20) |
| PaxAdditionalInfo | varchar(20) |
| AdditionalInfo    | varchar(15) |
| AgentInfo         | varchar(50) |

| Report       |             |
|--------------|-------------|
| from_city    | varchar(50) |
| from_country | varchar(25) |
| from_airport | varchar(3)  |
| to_city      | varchar(25) |
| to_country   | varchar(25) |
| to_airport   | varchar(3)  |
| date_from    | date        |
| date_to      | date        |
| days         | varchar(50) |
| depTime      | time        |
| arrTime      | time        |
| flight       | varchar(10) |
| aircraft     | varchar(3)  |
| travelTime   | varchar(10) |

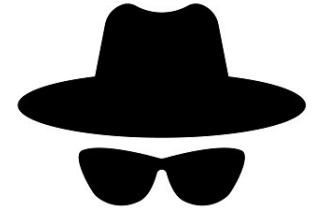
| AirlinesData    |             |
|-----------------|-------------|
| uid             | varchar(9)  |
| first_name      | varchar(25) |
| last_name       | varchar(25) |
| cards_type      | varchar(8)  |
| card_number     | varchar(12) |
| bonus_program   | varchar(30) |
| activities_type | varchar(8)  |
| activity_type   | varchar(6)  |
| code            | varchar(10) |
| date            | date        |
| departure       | varchar(3)  |
| arrival         | varchar(3)  |
| fare            | varchar(6)  |

## Result table

|                    |             |
|--------------------|-------------|
| id                 | serial      |
| first_name         | varchar(30) |
| last_name          | varchar(30) |
| birth_date         | date        |
| passenger_document | varchar(30) |
| travel_class       | varchar(30) |
| food_info          | varchar(30) |
| dep_city           | varchar(30) |
| dest_city          | varchar(30) |
| dep_date           | date        |
| baggage            | varchar(30) |



## Model Building: Preparation



- Estimates are developed for each passenger according to spy search criteria
- Total number of flights is counted
- Passenger route tracking algorithm is created: number of circle and collapsed routes is counted
- Minimum threshold of class turnover has been determined
- Passenger's meal requirement is estimated
- Passenger's baggage presence is determined

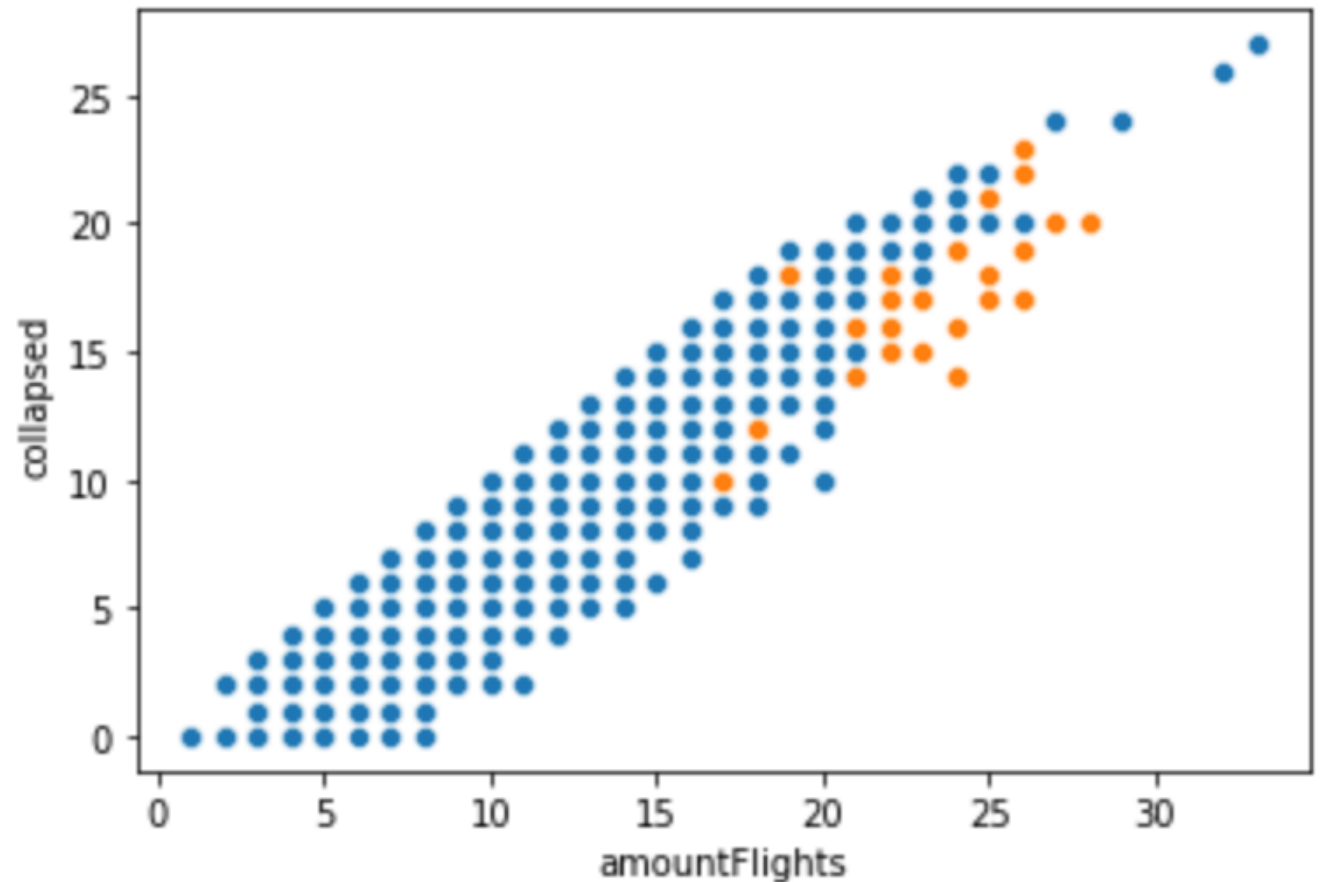


# Model Building: Analysis



IsolationForest algorithm is used for anomaly detection

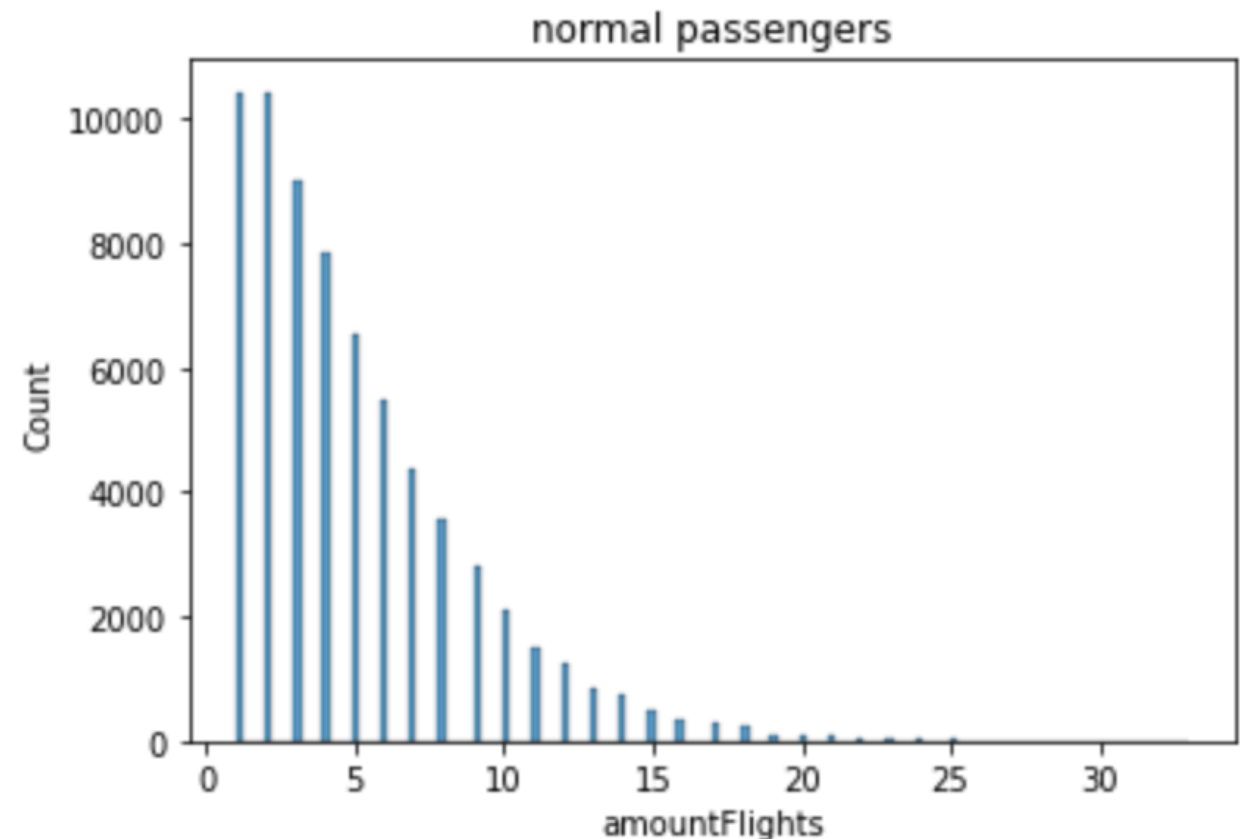
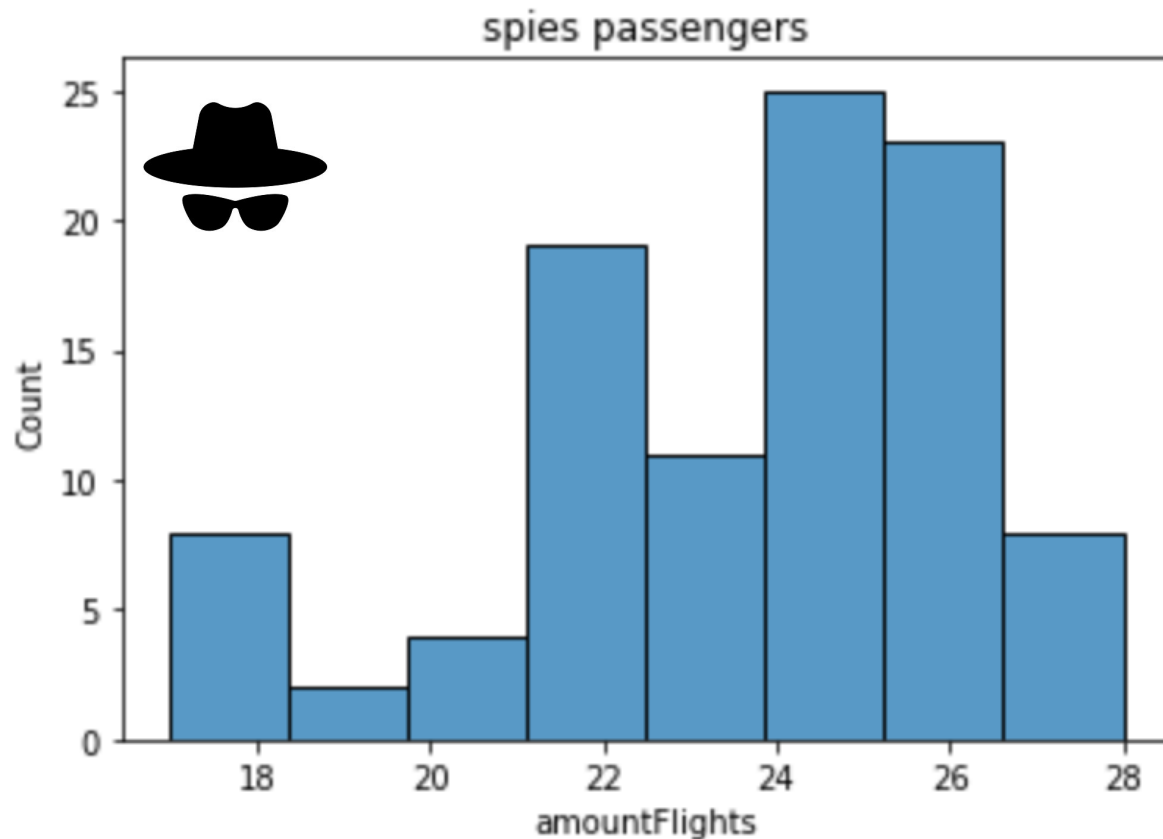
Pairwise distribution  
of the number of  
**suspicious** and **normal** routes:



# Model Building: Analysis



- Normal and spies passengers distributions are determined:

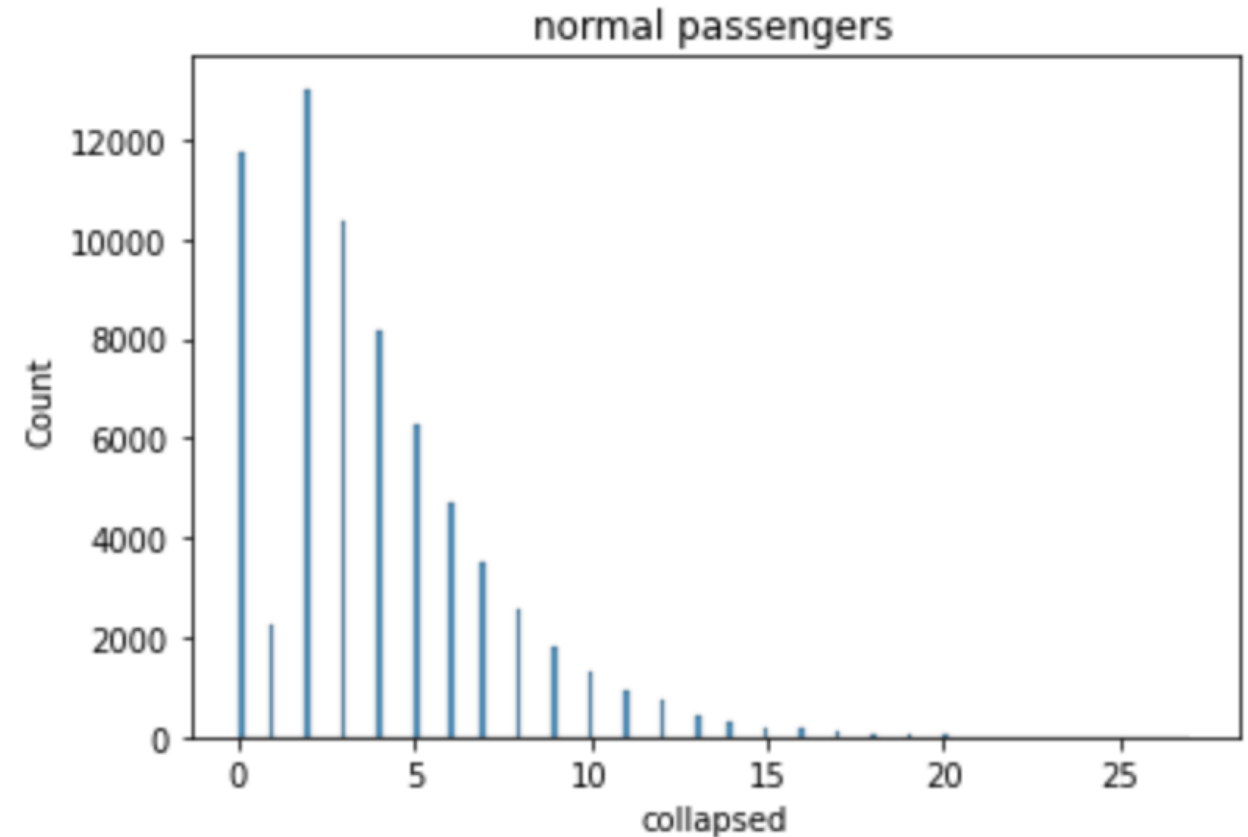
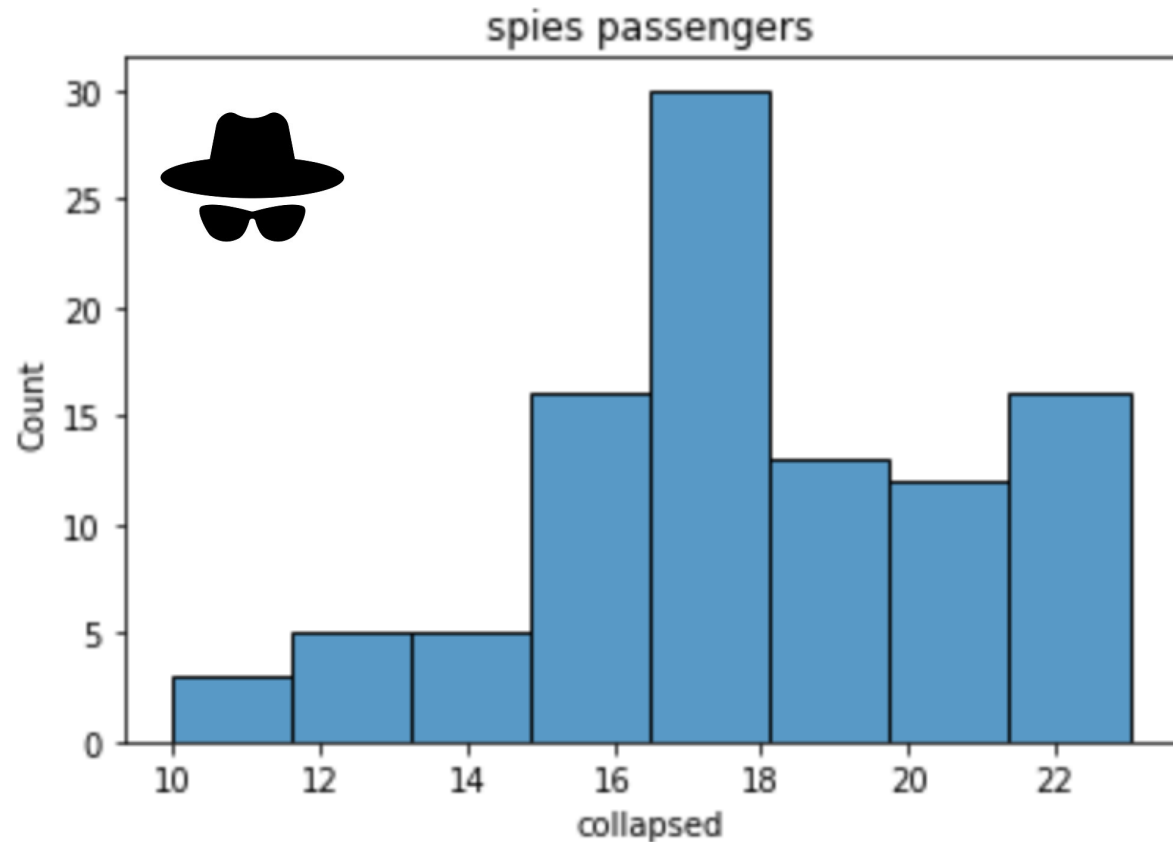




# Model Building: Analysis



- Normal and spies passengers distributions are determined amongst **collapsed** routes:



# Model Building: Results

List of caught spies:

|       | FirstName | LastName   | amountFlights | travleClass | foodInfo | circle | collapsed | baggage |
|-------|-----------|------------|---------------|-------------|----------|--------|-----------|---------|
| 518   | DMITRII   | ANTONOV    | 18            | 0           | 1        | 2      | 12        | 1       |
| 970   | ROBERT    | BELOUSOV   | 23            | 1           | 1        | 3      | 15        | 0       |
| 1423  | MADINA    | KOPYLOVA   | 25            | 1           | 1        | 3      | 17        | 0       |
| 3372  | IRINA     | DAVYDOVA   | 21            | 0           | 1        | 2      | 16        | 0       |
| 3687  | EMILIIA   | ISAEVA     | 26            | 0           | 1        | 1      | 23        | 0       |
| ...   | ...       | ...        | ...           | ...         | ...      | ...    | ...       | ...     |
| 62729 | TIMOFEI   | ROMANOV    | 21            | 1           | 1        | 2      | 16        | 1       |
| 64263 | NAZAR     | KUZNETSOV  | 26            | 0           | 1        | 1      | 22        | 0       |
| 66081 | TAMERLAN  | MIKHEEV    | 23            | 1           | 1        | 2      | 17        | 0       |
| 67481 | KAROLINA  | KRIUCHKOVA | 22            | 0           | 1        | 2      | 17        | 0       |
| 68435 | RINAT     | GLADKOV    | 22            | 0           | 1        | 2      | 18        | 0       |

100 rows × 8 columns

