National Research Nuclear University MEPhI

Catch "The Russian Spies" Challenge

Solved by students:

Volyanskiy Julian,

Ivanova Nina,

Khudoyarova Anastasia,

Nikolaeva Elizabeth

Discovery

Data sources:

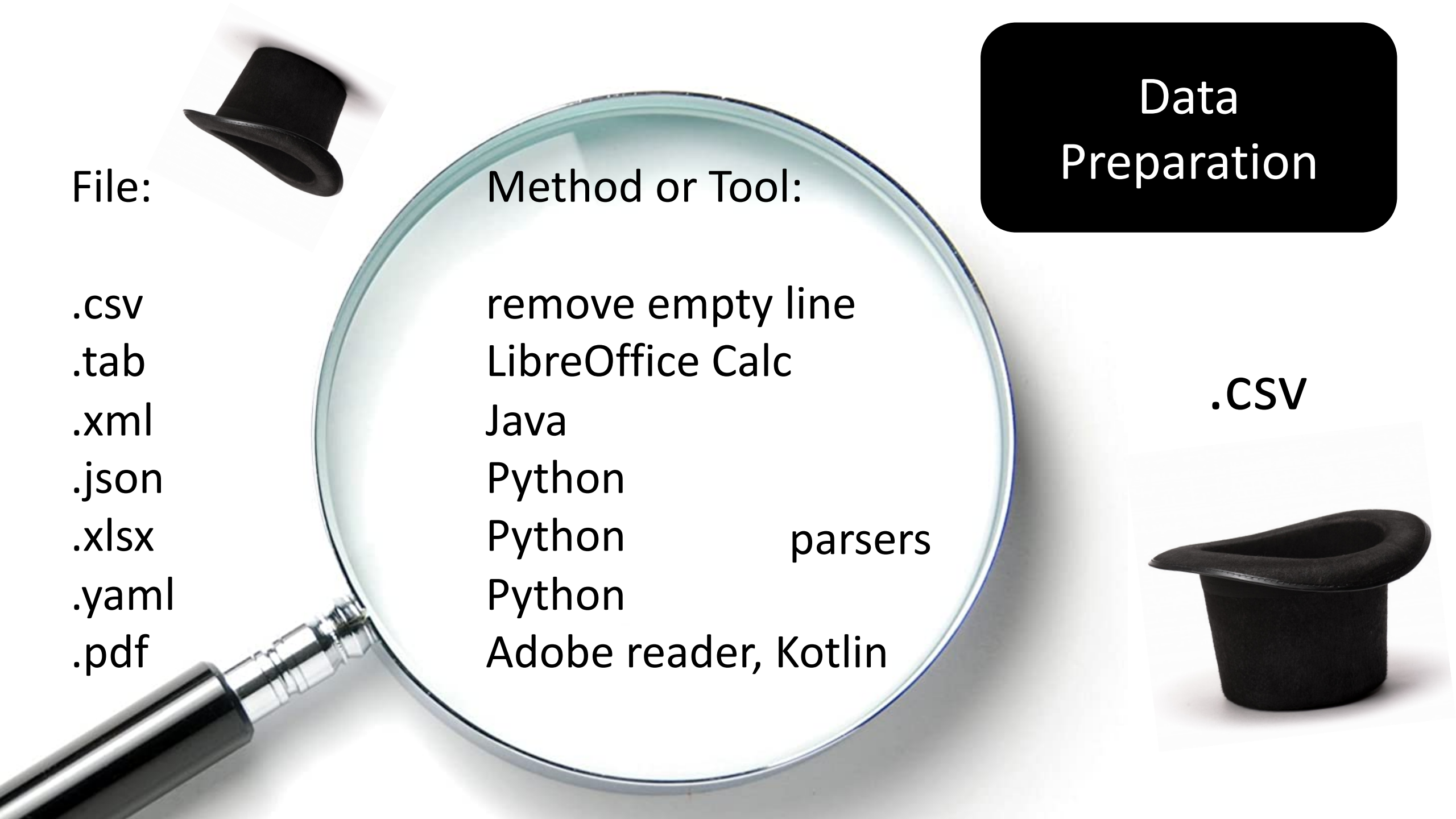
- BoardingData.csv
- FrequentFlyerForum-Profiles.json
- PointzAggregator-AirlinesData.xml
- Sirena-export-fixed.tab
- Skyteam_Timetable.pdf
- SkyTeam-Exchange.yaml
- YourBoardingPassDotAero.zip

Discovery

Spy search criteria:

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

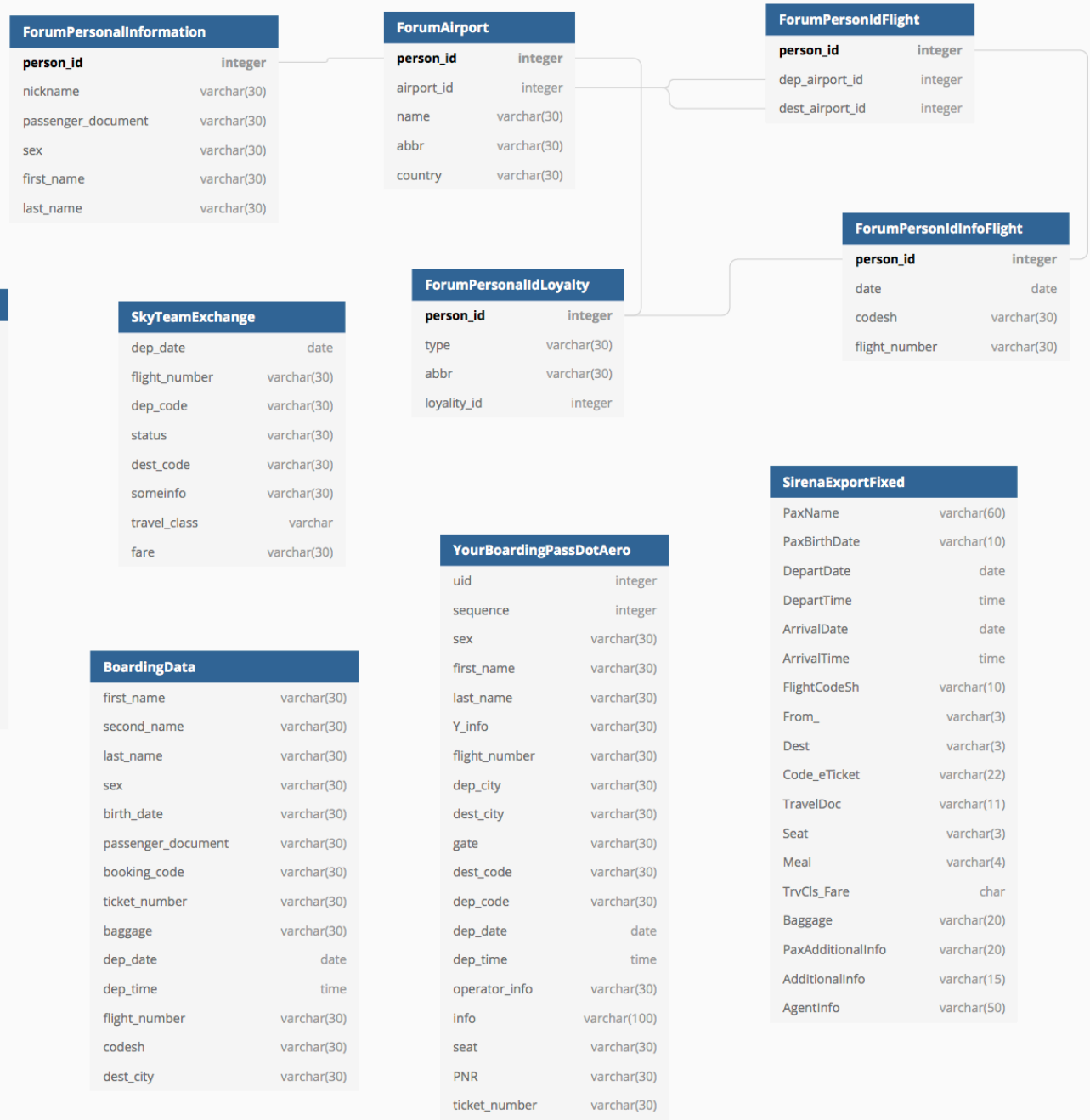
Data Preparation



File:	Method or Tool:	
.csv	remove empty line	
.tab	LibreOffice Calc	
.xml	Java	
.json	Python	
.xlsx	Python	parsers
.yaml	Python	
.pdf	Adobe reader, Kotlin	

.csv

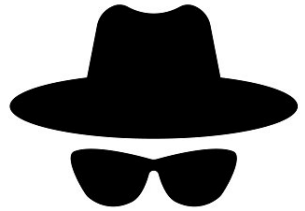
Model planning



Result table

id	serial
first_name	varchar(30)
last_name	varchar(30)
birth_date	varchar(30)
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

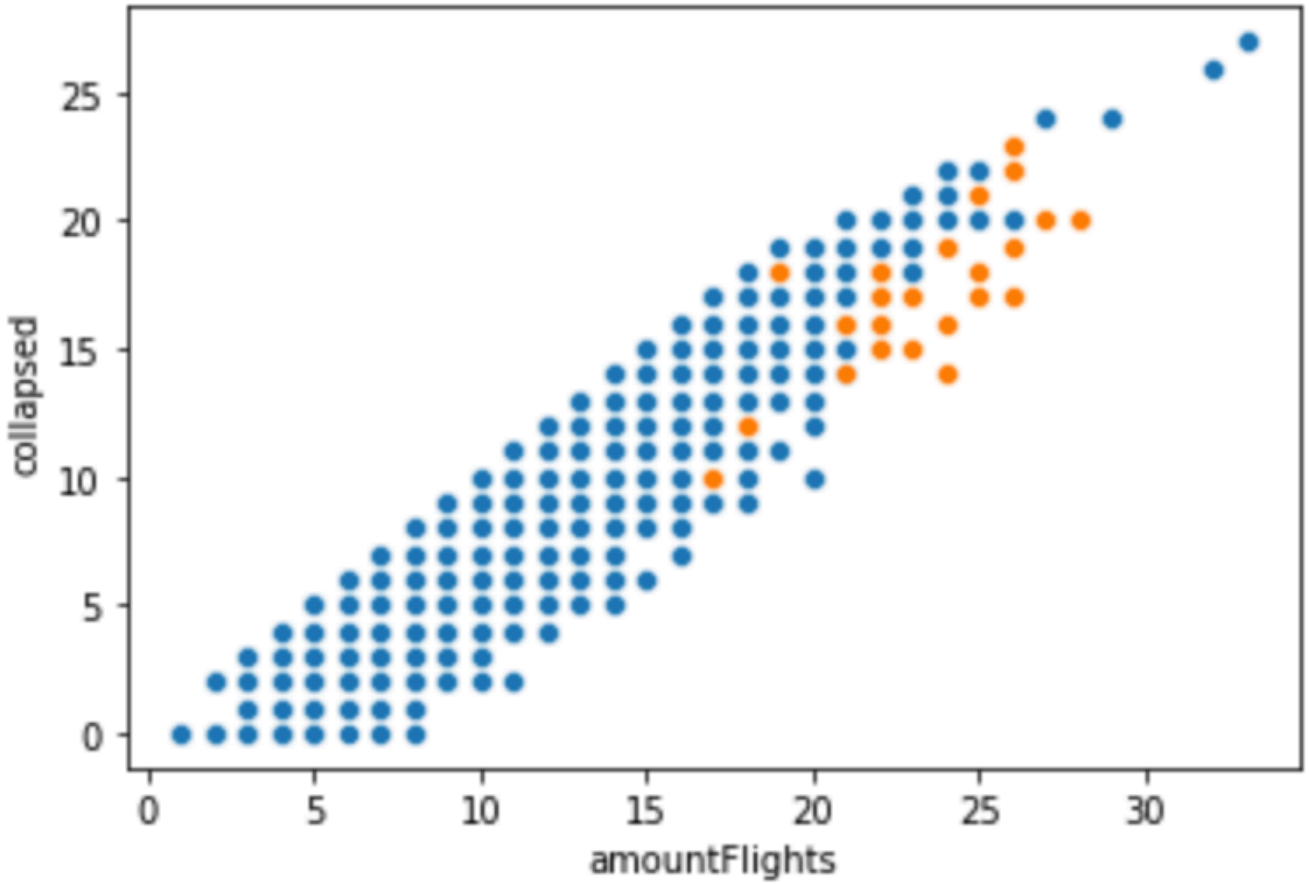
according to spy search criteria

- 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

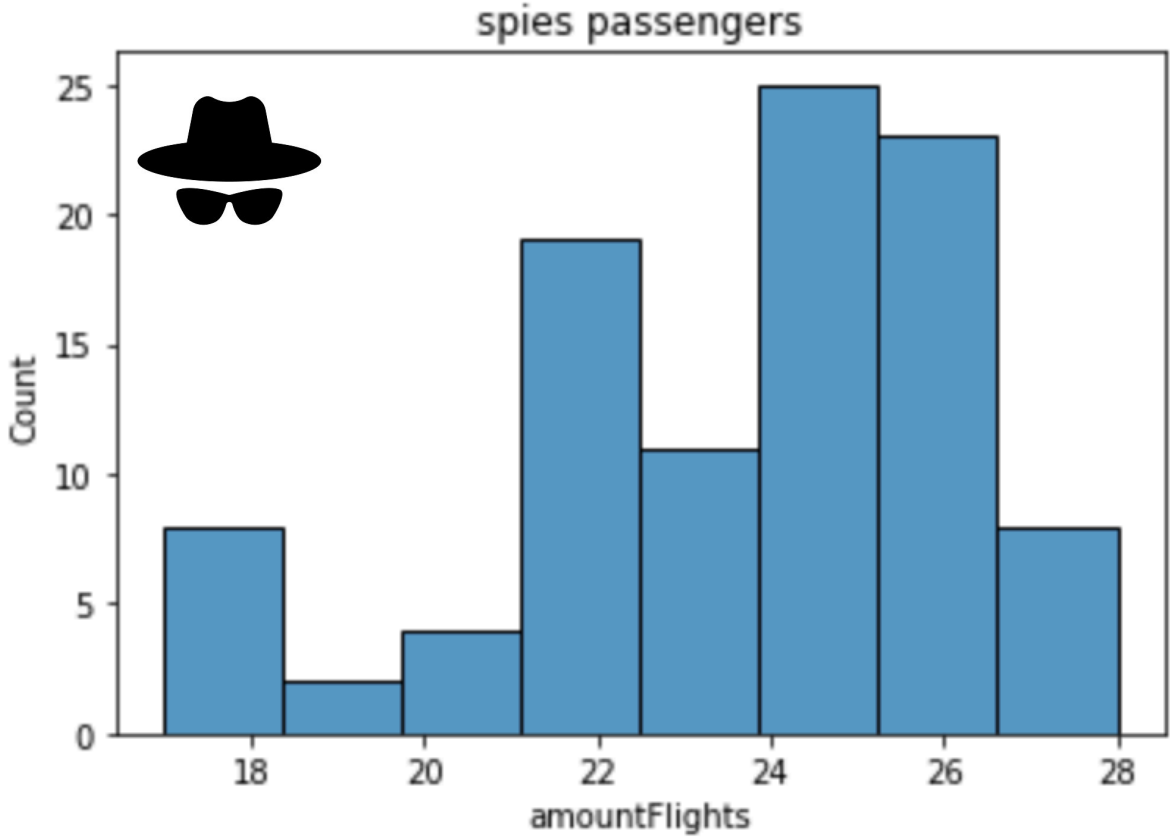
IsolationForest algorithm is used for anomaly detection

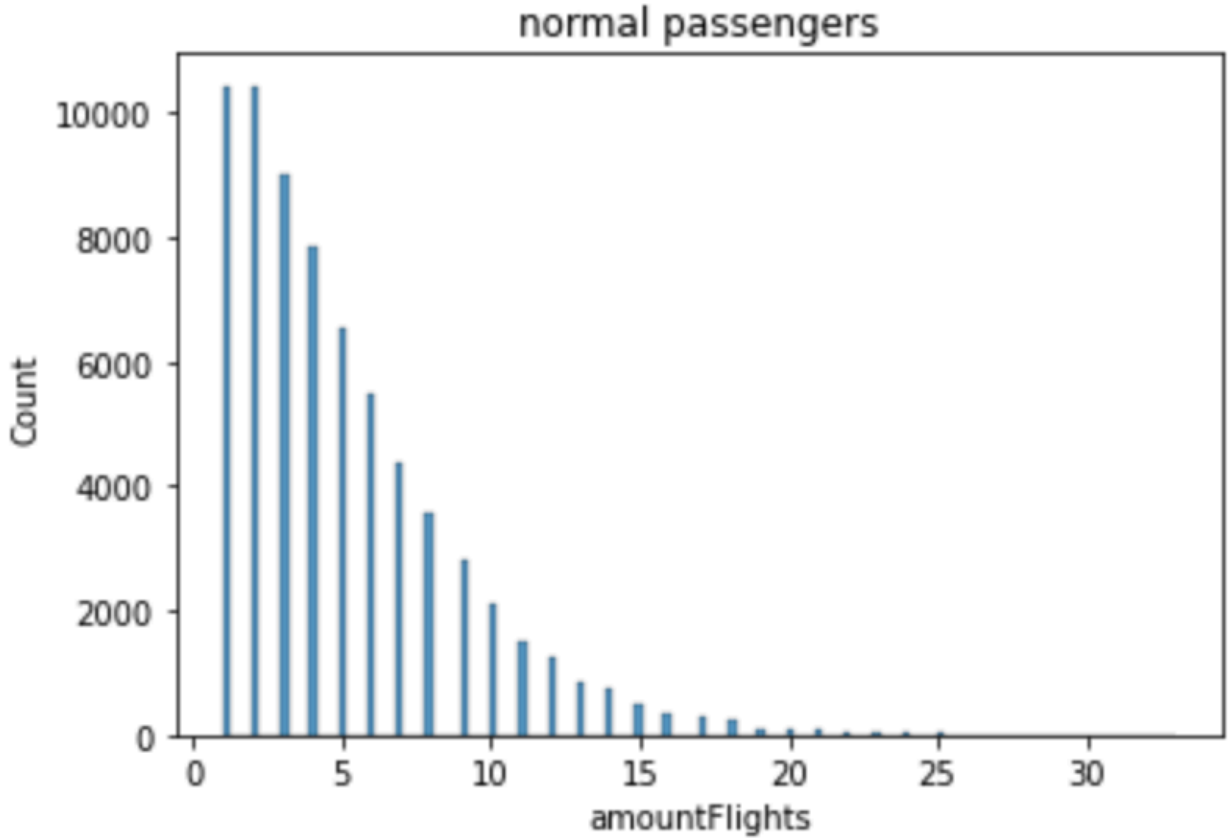
Pairwise distribution
of the number of
suspicious and normal routes:



Model Building: Analysis

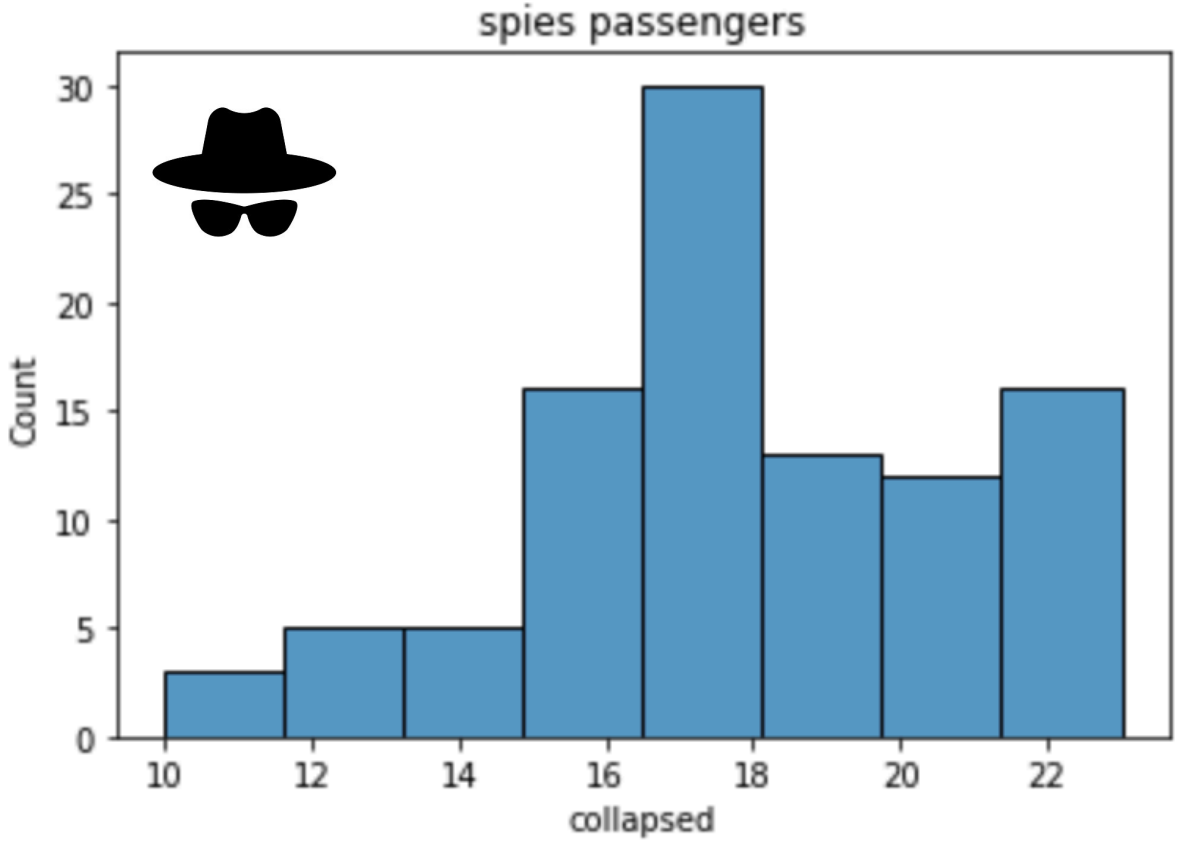
- Normal and spies passengers distributions are determined:

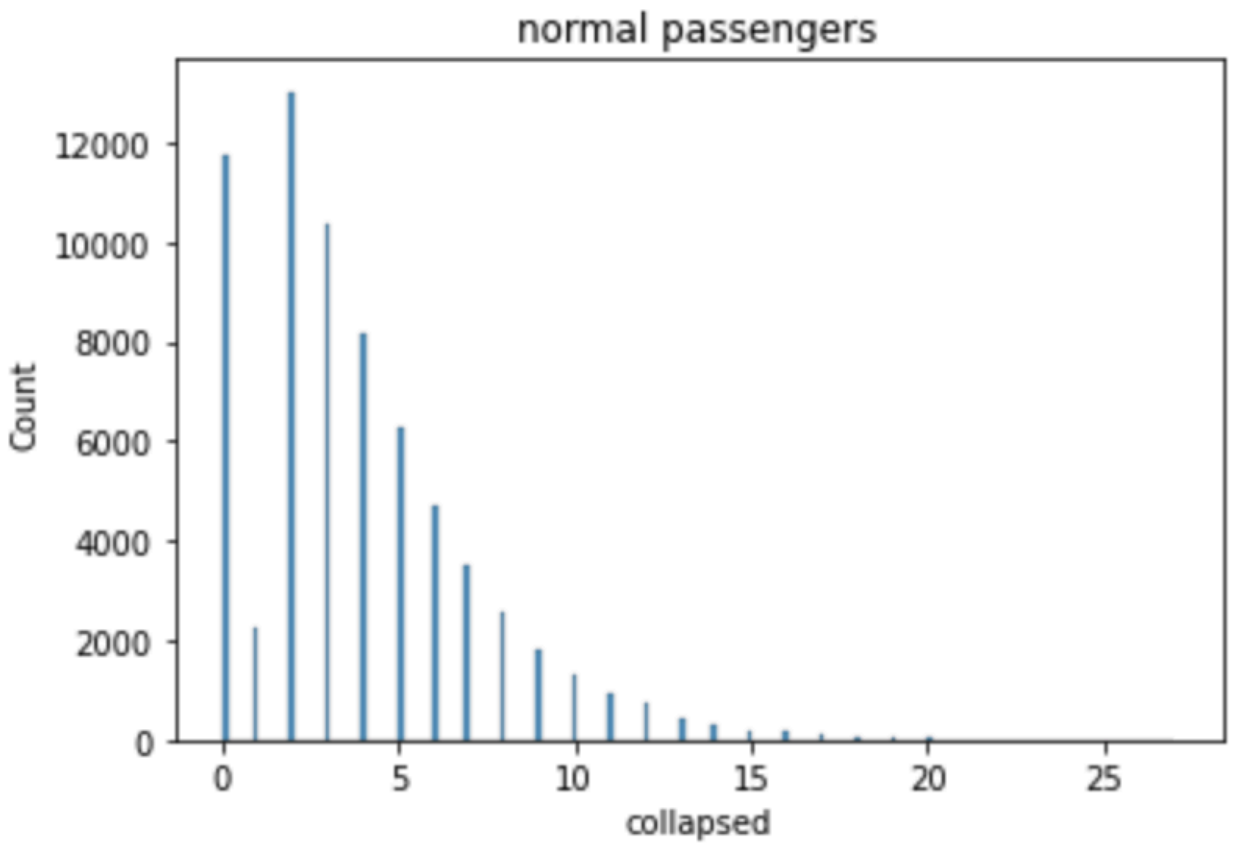




Model Building: Analysis

- Normal and spies passengers distributions are determined amongst collapsed routs:





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
1719	ULIANA	BUROVA	22	1	1	2	18	0
3372	IRINA	DAVYDOVA	21	0	1	2	16	0
...	...	...	...	...	...	...	...	...
64973	EMIL	ULIANOV	24	1	1	4	14	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
68640	ROBERT	BELOUSOV	23	1	1	3	15	0

100 rows × 8 columns