

Catch "The Russian Spies" Challenge

Участники команды:

Емельянов Станислав

Степанов Алексей

Тарасов Максим

Христолюбова Анна



Исходные данные

Набор документов в различных форматах, с данными об авиаперелетах и пассажирах

Основные данные о перелетах:

- Пункты вылета и посадки (Город, Аэропорт)
- Код Рейса
- Дата

Основные данные о пассажирах

- Имя
- Код карты лояльности

Форматы документов: csv, json, yaml, excel, xml, tab



Предобработка данных



Предобработка данных служит следующим **целям**:

- 1) Улучшение качества данных, путём устранения пропусков, дополнения из других источников
- 2) Подготовка данных к объединению, Приведение к одному формату представления данных



Предобработка данных

FrequentFlyerForum-Profiles.csv

- Иерархическая структура в виде массива из одного элемента, в котором лежит массив профилей пассажиров
- После перевода в табличный вид стало очевидно, что каждый из атрибутов может быть списком
- Т.к. поле “паспорт” у всех пустое, оно было удалено



Предобработка данных

FrequentFlyerForum-Profiles.csv

	Registered Flights	NickName	Travel Documents	Sex	Loyalty Programm	Real Name
0	{'Date': '2017-03-04', 'Codeshare': False, 'A...	FrequentFlyer0	{'Passports': None}}	Male	{'Status': 'Basic', 'programm': 'KE', 'Number...	{'Last Name': None, 'First Name': None}
1	{'Date': '2017-08-31', 'Codeshare': True, 'Ar...	FrequentFlyer2	{'Passports': None}}	Male	{'Status': 'Basic', 'programm': 'SU', 'Number...	{'Last Name': 'ZYKOV', 'First Name': 'NAZAR'}
2	{'Date': '2017-09-15', 'Codeshare': True, 'Ar...	FrequentFlyer4	{'Passports': None}}	None	{'Status': 'Elite', 'programm': 'SU', 'Number...	{'Last Name': None, 'First Name': None}
3	{'Date': '2017-09-18', 'Codeshare': True, 'Ar...	FrequentFlyer5	{'Passports': None}}	None	{'Status': 'Elite+', 'programm': 'FB', 'Numbe...	{'Last Name': None, 'First Name': None}
4	{'Date': '2017-08-15', 'Codeshare': False, 'A...	FrequentFlyer6	{'Passports': None}}	Male	{'Status': 'Basic', 'programm': 'SU', 'Number...	{'Last Name': 'D'YAKOV', 'First Name': 'RUSLAN'}



Обработанный:

	NickName	Arrival_Airport	Arrival_City	Arrival_Country	Departure_Airport	Departure_City	Departure_Country	Date	Codeshare	Flight	Sex	First Name	Last Name
0	0	PIT	Pittsburgh PA	United States of	BOS	Boston MA	United States of	2017-03-04	False	DL3377	Male	None	None
1	0	DTW	Detroit MI	United States of	GRB	Green Bay WI	United States of	2017-07-21	False	DL3630	Male	None	None
2	0	LGA	New York NY	United States of	PIT	Pittsburgh PA	United States of	2017-03-22	False	DL5361	Male	None	None
3	0	WAW	Warsaw	Poland	SVO	Moscow	Russian Federation	2017-01-04	True	SU3702	Male	None	None
4	0	ATL	Atlanta GA	United States of	XNA	Fayetteville AR	United States of	2017-08-08	False	DL3412	Male	None	None



Предобработка данных

Sirena-export-fixed.tab

Данный документ обладает столбцами фиксированной длины, был переведён в .csv при помощи excel

- Сбросили столбец “Seat”, т.к. не было уникальных значений
- Столбцы с информацией о карте лояльности Unnamed: 17 и Unnamed: 18 объединили в один и убрали символ #, чтобы был как остальные
- Имена сделали транслитом



Предобработка данных

Sirena-export-fixed.tab исходный DataFrame после excel

PaxName	PaxBirthDa	teDepartDate	DepartTime	ArrivalDate	ArrivalTime	FlightCod	eShFrom	Dest	Code	e-Ticket	Trav	eIDoc	Seat
ОЗЕРОВ ИЛЬДАР ДАНИИЛОВИЧ	1999-05-15	30.05.2017	0:05	30.05.2017	8:05	SU1306NO	SVO	OVB	ZBQSPY7360415302044672	9375	53270	NaN	
КОЛОСОВ САМИР ТАМЕРЛАНОВИЧ	NaT	27.12.2017	2:15	27.12.2017	4:40	SU1323NO	MMK	SVO	MHPBBX7398421117936516	2244	645520	NaN	
ИГНАТОВА СНЕЖАНА КОНСТАНТИНОВНА	NaT	19.09.2017	6:40	19.09.2017	7:45	SU1481NO	KJA	SVO	REDLVB5174973140468001	8115	961316	NaN	
ЖАРОВ ПЛАТОН АЛЬБЕРТОВИЧ	1999-02-05	18.03.2017	22:10	19.03.2017	1:05	SU1180NO	SVO	VOG	NSJNGQ5274206497242737	98 6	865148	NaN	
НИКОЛЬСКИЙ НИКОЛАЙ ИГОРЕВИЧ	1990-12-26	18.03.2017	22:10	19.03.2017	1:05	SU1180NO	SVO	VOG	VWNYGF6247422701565929	4396	926588	NaN	

Meal	Trv	ClsFare Baggage	PaxAdditionalInfo	Unnamed: 17	Unnamed: 18	AgentInfo	DepartDateTime	ArrivalDateTime
NaN	J	JGRPKN0PC	S	FF#SU	38116280.0	Go2See	2017-05-30 00:05:00	2017-05-30 08:05:00
KSML	Y	YRSTUQ	9	FF#FB	284903754.0	NaN	2017-12-27 02:15:00	2017-12-27 04:40:00
NaN	Y	YSTNJL	F	NaN	NaN	KupiBilet	2017-09-19 06:40:00	2017-09-19 07:45:00
NaN	J	JFLXLS0PC	#	FF#FB	884556993.0	Travelgenio	2017-03-18 22:10:00	2017-03-19 01:05:00
NaN	Y	YFLXPG	NaN	FF#SU	183142068.0	OZON.travel	2017-03-18 22:10:00	2017-03-19 01:05:00



Предобработка данных

Sirena-export-fixed.tab результат

	PaxName	PaxBirthDa	teDepartDate	DepartTime	ArrivalDate	ArrivalTime	FlightCod	eShFrom	Dest	Code	e-Ticket	Trav	eIDoc	Meal	Trv	ClsFare Baggage
0	IL'DAR OZEROV	1999-05-15	30.05.2017	0:05	30.05.2017	8:05	SU1306NO	SVO	OVB	ZBQSPY7360415302044672	9375	53270	NaN	J	JGRPGN0PC	
1	SAMIR KOLOSOV	NaT	27.12.2017	2:15	27.12.2017	4:40	SU1323NO	MMK	SVO	MHPBBX7398421117936516	2244	645520	KSML	Y	YRSTUQ	
2	SNEZHANA IGNATOVA	NaT	19.09.2017	6:40	19.09.2017	7:45	SU1481NO	KJA	SVO	REDLVB5174973140468001	8115	961316	NaN	Y	YSTNJL	
3	PLATON ZHAROV	1999-02-05	18.03.2017	22:10	19.03.2017	1:05	SU1180NO	SVO	VOG	NSJNGQ5274206497242737	98 6	865148	NaN	J	JFLXLS0PC	
4	NIKOLAJ NIKOL'SKIJ	1990-12-26	18.03.2017	22:10	19.03.2017	1:05	SU1180NO	SVO	VOG	VWNYGF6247422701565929	4396	926588	NaN	Y	YFLXPG	

PaxAdditionalInfo	AgentInfo	DepartDateTime	ArrivalDateTime	card
S	Go2See	2017-05-30 00:05:00	2017-05-30 08:05:00	FF 38116280
9	NaN	2017-12-27 02:15:00	2017-12-27 04:40:00	FF 284903754
F	KupiBilet	2017-09-19 06:40:00	2017-09-19 07:45:00	None
#	Travelgenio	2017-03-18 22:10:00	2017-03-19 01:05:00	FF 884556993
NaN	OZON.travel	2017-03-18 22:10:00	2017-03-19 01:05:00	FF 183142068



Предобработка данных

SkyTeam-Exchange.csv

Данный файл представлял собой отображение где в качестве ключей были даты, а в качестве значения - отображения где в качестве ключей код рейса в данный день, а в качестве значения совокупность данных о полёте и код карты лояльности в виде отображения

- Конвертация из yaml в json утилитой в linux, далее парсинг пакетом json после был разбор

Исходный

```
1 q['2017-01-01']['AF1145'] # пример записи

{'FF': {'FB 520518073': {'CLASS': 'Y', 'FARE': 'YRSTFN'}},
 'FROM': 'SVO',
 'STATUS': 'LANDED',
 'TO': 'CDG'}
```



Промежуточная таблица после парсинга JSON

card	date	flight	from	to	status	class	fare
FB 520518073	2017-01-01	AF1145	SVO	CDG	LANDED	Y	YRSTFN
FB 18427809	2017-01-01	AF1844	SVO	CDG	LANDED	Y	YGRPZR
KE 993927664	2017-01-01	AF4855	SVO	CDG	LANDED	Y	YRSTJQ
SU 789402704	2017-01-01	AF4901	SVO	CDG	LANDED	Y	YGRPAT
KE 352734424	2017-01-01	AZ545	SVO	FCO	LANDED	Y	YFLXKO

*Были только коды карт, но не было имен имена взяли из другого источника, где совпадали коды карт
PointzAggregator-AirlinesData.csv

card	date	flight	from	to	status	class	fare	first_name	last_name
FB 520518073	2017-01-01	AF1145	SVO	CDG	LANDED	Y	YRSTFN	MELANIA	AFANASEVA
FB 18427809	2017-01-01	AF1844	SVO	CDG	LANDED	Y	YGRPZR	SOFIA	MASLOVA
KE 993927664	2017-01-01	AF4855	SVO	CDG	LANDED	Y	YRSTJQ	NaN	NaN
SU 789402704	2017-01-01	AF4901	SVO	CDG	LANDED	Y	YGRPAT	BORIS	MALININ
KE 352734424	2017-01-01	AZ545	SVO	FCO	LANDED	Y	YFLXKO	FEDOR	ROGOV



Предобработка данных

YourBoardingPassDotAero.zip

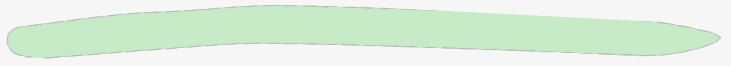
Данный архив содержит в себе файлы в формате excel

- Составили единый фрейм строки составлялись путём получения значений ячеек пакетом `orenpuxl`
- Столбцы `seat` и `gate` не имели значений (Nan)

sequence	gender	name	fare	flightNumber	cityFrom	cityTo	airportFrom	airportTo	date	time	share	PNR	ticket	BoardDateTime
29	MRS	MADINA SHIRYAEVA	Y	CZ6625	ZHENGZHOU	XINZHOU	CGO	WUT	2017-09-22	14:50	Operated by Some Other Airline	OQQOHE	2669364900304337	2017-09-22 14:50:00
75	MRS	ALINA SUKHANOVA	Y	SV1481	RIYADH	QAISUMAH	RUH	AQI	2017-09-22	05:10	Operated by Some Other Airline	ZLMHOE	9161977196434544	2017-09-22 05:10:00
42	MR	LEV GONCHAROV	Y	SU2314	MOSCOW	BERLIN	SVO	SXF	2017-09-22	14:30	Operated by Some Other Airline	KYWCAH	3154693033418596	2017-09-22 14:30:00
100	MRS	KULAGINA AMINA	Y	SU1175	ASTRAKHAN	MOSCOW	ASF	SVO	2017-09-22	19:50	Operated by Some Other Airline	QMQCLS	5632976945423178	2017-09-22 19:50:00
6	MR	OLEYNIKOV VLADIMIR	Y	AZ715	ATHENS	ROME	ATH	FCO	2017-09-22	13:15	Operated by Some Other Airline	DOUZUE	8399241805063994	2017-09-22 13:15:00

Предобработка данных

Skyteam_Timetable.pdf



Данный файл содержит в себе информацию о рейсах

- Распарсили, используя пакет tabula.py
- Данный справочник относится к периоду 2018-2019, хотя данные в остальных наборах за 2017-2018 год, вследствие чего в дальнейшем не использовался



Предобработка данных

Приведение к единой структуре для объединения в один фрейм


Далее все таблицы приводились к одному формату, с одинаковыми названиями столбцов, и отсекая столбцы не входящие в множество столбцов будущего единого фрейма

- имя – name только Имя Фамилия
- дата полёта – date
- номер самолёта – flight_number
- fare
- аэропорт вылета
- аэропорт посадки
- город вылета
- город посадки

name	date	flight_number	city_from	airport_from	city_to	airport_to	fare
SAVELII RUSANOV	2017-03-22	SU1369	NaN	NaN	Moscow	NaN	NaN
LEV ISAEV	2017-03-18	SU1180	NaN	NaN	Volgograd	NaN	NaN
NIKOLAI NIKOLSKII	2017-03-18	SU1180	NaN	NaN	Volgograd	NaN	NaN
ANATOLII SHILOV	2017-03-18	SU1180	NaN	NaN	Volgograd	NaN	NaN
MIROSLAVA SEMENOVA	2017-03-12	SU6284	NaN	NaN	Moscow	NaN	NaN



Выявление гипотез

- 
- 1) Предположили, что шпионы обязательно должны были побывать в Москве
 - 2) Предположили, что шпионы должны часто летать
 - 3) Предположили, если шпион за кем-то следит, у него будут совпадать полёты с другими людьми
 - 4) Даты перемещений совпадают с необычными событиями в мире



Подтверждение гипотез

1) Предположили, что наши шпионы обязательно должны были побывать в Москве

Сделали фрейм со всеми полетами, кто хоть когда-нибудь был в Москве

```
1 from_msc_df = res_df[(res_df.city_from == 'Moscow') | (res_df.airport_from.isin(['SVO', 'VKO', 'DME']))]  
2 to_msc_df = res_df[(res_df.city_to == 'Moscow') | (res_df.airport_to.isin(['SVO', 'VKO', 'DME']))]
```

```
1 msc_df = pd.concat([from_msc_df, to_msc_df])  
2 msc_visitors_df = res_df[(res_df.name.isin(from_msc_df.name) & res_df.name.isin(to_msc_df.name))]
```

```
1 msc_visitors_df.head()
```

	name	date	flight_number	city_from	airport_from	city_to	airport_to	fare
0	SAVELII RUSANOV	2017-03-22	SU1369	NaN	NaN	Moscow	NaN	NaN
1	LEV ISAEV	2017-03-18	SU1180	NaN	NaN	Volgograd	NaN	NaN
2	NIKOLAI NIKOLSKII	2017-03-18	SU1180	NaN	NaN	Volgograd	NaN	NaN
4	MIROSLAVA SEMENOVA	2017-03-12	SU6284	NaN	NaN	Moscow	NaN	NaN
5	ARTEM KAPUSTIN	2017-03-12	SU6284	NaN	NaN	Moscow	NaN	NaN



Подтверждение гипотез



2) Предположили, что наши шпионы должны часто летать

```
1 temp_df = msc_visitors_df.groupby('name').count().sort_values(by='date')
2
3 filt_msc_visitors_df = msc_visitors_df[msc_visitors_df.name.isin(temp_df[temp_df['date'] > 50].index)]
4 len(filt_msc_visitors_df)
```

572883



Подтверждение гипотез

3) Предположили, если шпион за кем-то следит, у него будут совпадать полёты с другими людьми

```
1 def get_cross(record):
2     temp = for_check.apply(lambda x: len(record & x))
3     temp = temp[temp > 2] # минимальное количество совместных полётов
4     temp = temp.reset_index()
5     temp = temp.agg(tuple, axis=1)
6     return list(temp)

1 temp = suspicious_statistic.set_index('name')[['date', 'airport_from', 'airport_to']].agg(frozenset, axis=1)
2 for_check = temp.groupby('name').agg(set)

1 super_df = for_check.apply(get_cross)

1 # Отсеивание пересечение с самим собой
2 super_filtered_df = super_df[super_df.apply(lambda x: False if len(x) == 1 else True)]
3
4 super_filtered_df = super_filtered_df.apply(lambda x: sorted(x, key = lambda a: a[1], reverse=True))
```

1 super_filtered_df

name	
ADEL AGEEV	[(ADEL AGEEV, 45), (ADEL' AGEEV, 25)]
ADEL DIAKONOV	[(ADEL DIAKONOV, 52), (ADEL' D'YAKONOV, 21)]
ADEL KOMAROV	[(ADEL KOMAROV, 66), (ADEL' KOMAROV, 18), (OKS...
ADEL KORSHUNOV	[(ADEL KORSHUNOV, 63), (ADEL' KORSHUNOV, 17)]
ADEL PETROV	[(ADEL PETROV, 48), (ADEL' PETROV, 24)]
	...
ZLATA KRAVTSOVA	[(ZLATA KRAVTSOVA, 60), (INNA SHISHKINA, 3), (...]
ZLATA MALYSHEVA	[(ZLATA MALYSHEVA, 44), (POLINA PANFILOVA, 3)]
ZLATA NAUMOVA	[(ZLATA NAUMOVA, 39), (KIRILL BOGOMOLOV, 3)]
ZLATA SURKOVA	[(ZLATA SURKOVA, 46), (SVETLANA DEMIDOVA, 4), ...]
ZLATA ZHURAVLEVA	[(ZLATA ZHURAVLEVA, 68), (ARIANA SVIRIDOVA, 3)]
Length: 1045, dtype: object	



Подтверждение гипотез

Полученное количество совместных полетов:

	A	B	C	D	E	F	G	H	I	J
1	name	0								
2	ADEL KOMAROV	[('ADEL KOMAROV', 66), ('ADEL' KOMAROV", 18), ('OKSANA BARANOVA', 3)]								
3	ADEL SNEGIREV	[('ADEL SNEGIREV', 57), ('KRISTINA MOISEEVA', 3)]								
4	ADELINA ALEKSANDROVA	[('ADELINA ALEKSANDROVA', 45), ('ALBINA PLOTNIKOVA', 3)]								
5	ADELINA DUBININA	[('ADELINA DUBININA', 45), ('ALENA ORLOVA', 3)]								
6	ADELINA KALUGINA	[('ADELINA KALUGINA', 66), ('NATASHA DOROFEEVA', 3)]								
7	ADELINA LAPINA	[('ADELINA LAPINA', 43), ('DANIIL ANTIPOV', 3), ('VASILISA MITROFANOVA', 3)]								
8	ADELINA MESHKOVA	[('ADELINA MESHKOVA', 40), ('NADEZHDA KALUGINA', 3)]								
9	ADELINA SELIVANOVA	[('ADELINA SELIVANOVA', 56), ('FEDOR MURATOV', 3)]								
10	ADELINA SOTNIKOVA	[('ADELINA SOTNIKOVA', 49), ('ANTON KALMYKOV', 3), ('EKATERINA AGAFONOVA', 3)]								
11	AGATA KORNILOVA	[('AGATA KORNILOVA', 41), ('ANGELINA MESHKOVA', 3)]								
12	AGATA LARIONOVA	[('AGATA LARIONOVA', 54), ('ROSTISLAV BYCHKOV', 3)]								
13	AGATA MOSKVINA	[('AGATA MOSKVINA', 41), ('RADMIR ZHDANOV', 3)]								
14	AGATA POTAPOVA	[('AGATA POTAPOVA', 41), ('VSEVOLOD SHVETSOV', 3)]								
15	AGATA RUMYANTSEVA	[('AGATA RUMYANTSEVA', 44), ('DENIS TCAREV', 3)]								
16	AIDAR OZEROV	[('AIDAR OZEROV', 53), ('ZARINA DAVYDOVA', 3)]								
17	ALBERT KUKUSHKIN	[('ALBERT KUKUSHKIN', 52), ('GALINA AVDEEVA', 3)]								
18	ALBINA KONSTANTINOVA	[('ALBINA KONSTANTINOVA', 52), ('VLAD DRUZHININ', 3)]								
19	ALBINA PANKOVA	[('ALBINA PANKOVA', 44), ('AL'BINA PANKOVA", 15), ('GALINA KORCHAGINA', 3)]								
20	ALBINA PLOTNIKOVA	[('ALBINA PLOTNIKOVA', 47), ('ADELINA ALEKSANDROVA', 3), ('VALENTINA KORCHAGINA', 3)]								
21	ALBINA ZUBKOVA	[('ALBINA ZUBKOVA', 53), ('PETR SAMSONOV', 3)]								
22	ALEKSANDR KULIKOV	[('ALEKSANDR KULIKOV', 66), ('LEONID NIKITIN', 4), ('VSEVOLOD SEROV', 3)]								
23	ALEKSANDR PLOTNIKOV	[('ALEKSANDR PLOTNIKOV', 58), ('MARINA VOLKOVA', 3)]								
24	ALEKSANDRA CHIZHOVA	[('ALEKSANDRA CHIZHOVA', 48), ('PAVEL GORBACHEV', 3)]								





Спасибо за внимание!

Приложение



Ссылка на коллаб:

https://colab.research.google.com/drive/1bG_kScDENG4qx3a5eAnmgqvb28kaAZIk?usp=sharing



Outro



Содержание:

- 1) Вопрос транслитерации
- 2) Количество людей побывавших в Москве



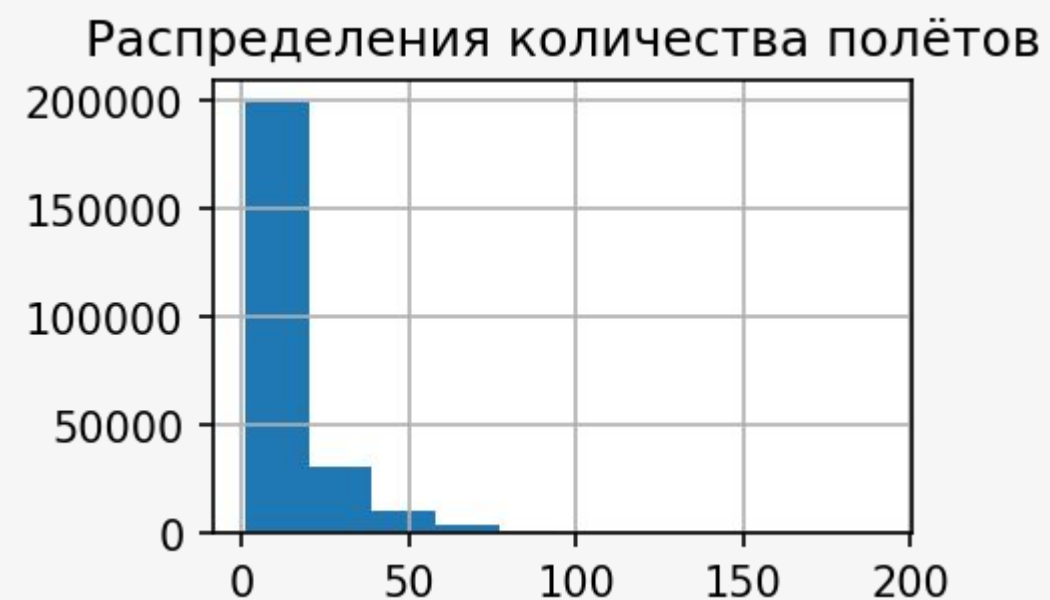
Outro

Количество людей побывавших в Москве 1/2

```
1 temp_df = msc_visitors_df.groupby('name').count().sort_values(by='date')
2
3 filt_msc_visitors_df = msc_visitors_df[msc_visitors_df.name.isin(temp_df[temp_df['date'] > 50].index)]
4 len(filt_msc_visitors_df)
```

572883

Не было понятно



Решили отсеять по порогу 50, на основании данного распределения, руководствуясь интуицией



Outro

Количество людей побывавших в Москве 2/2

```
[ ] 1 temp_df = msc_visitors_df.groupby('name').count().sort_values(by='date')
    2
    3 filt_msc_visitors_df = msc_visitors_df[msc_visitors_df.name.isin(temp_df[temp_df['date'] > 50].index)]
    4 print(f'Количество полётов часто летающих людей: {len(filt_msc_visitors_df)}')
```

Количество полётов часто летающих людей: 558615

```
▶ 1 suspicious_people = set(filt_msc_visitors_df.name.unique())
   2 print(f'Количество часто летающих: {len(suspicious_people)}')
```

↗ Количество часто летающих: 7835

```
▶ 1 suspicious_people
```

```
↳ {'SEMEN KOROLKOV',
   'EVA KALACHEVA',
   'ALBERT EGOROV',
   'GERMAN KOLOSOV',
```

Стало понятно

Если человек хоть раз был в Москве, летел туда или от туда, то в фрейм попадают все его рейсы



Intro

Транслитерация

```
1 super_filtered_df

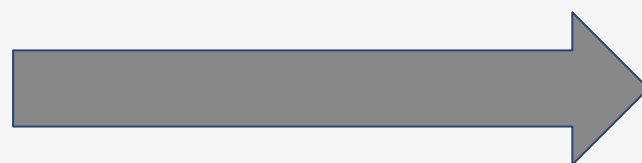
name
ADEL DIAKONOV      [(ADEL DIAKONOV, 52), (ADEL DYAKONOV, 20)]
ADEL DYAKONOV      [(ADEL DYAKONOV, 64), (ADEL DIAKONOV, 20)]
ADEL KOMAROV       [(ADEL KOMAROV, 92), (OKSANA BARANOVA, 3)]
ADEL KONSTANTINOV  [(ADEL KONSTANTINOV, 48), (MARSEL BELYAEV, 4)]
ADEL KOSHELEV      [(ADEL KOSHELEV, 60), (ROMAN FADEEV, 3)]
...
```

Два человека, с одинаковым именем, но разным написанием, способом транслитерации для устранения:

- 1) Использовали пакет `iuliiia`, с поддержкой различных стандартов транслитерации
- 2) Находили в итоговом наборе данных людей с одинаковым именем, с небольшой разницей в написании, при помощи расстояния Левенштейна
- 3) Полёты дописывались в первую запись, вторая удалялась

name	flights
ADEL DIAKONOV	[(ADEL DIAKONOV, 116)]
ADEL KOMAROV	[(ADEL KOMAROV, 92), (OKSANA BARANOVA, 3)]
ADEL KONSTANTINOV	[(ADEL KONSTANTINOV, 48), (MARSEL BELYAEV, 4)]
ADEL KOSHELEV	[(ADEL KOSHELEV, 60), (ROMAN FADEEV, 3)]
ADEL RUSAKOV	[(ADEL RUSAKOV, 72), (ANGELINA LOBANOVA, 3)]
...	...
ZARINA SOKOLOVA	[(ZARINA SOKOLOVA, 42), (RINAT SIZOV, 4), (BOG...)]
ZARINA TIMOFEEVA	[(ZARINA TIMOFEEVA, 53), (RADMIR VINOGRADOV, 3)]
ZLATA EFIMOVA	[(ZLATA EFIMOVA, 90), (RODION GROMOV, 3)]
ZLATA LOGINOVA	[(ZLATA LOGINOVA, 67), (MADINA RUDAKOVA, 3)]
ZLATA ZOTOVA	[(ZLATA ZOTOVA, 70), (FILIPP GLUSHKOV, 3)]

Итоговый фрейм после удаления дубликатов с похожим написанием



Отсечение тех, у кого после удаления копии не осталось совместных полётов с другими людьми

	name	flights
0	ADEL KOMAROV	[(ADEL KOMAROV, 92), (OKSANA BARANOVA, 3)]
1	ADEL KONSTANTINOV	[(ADEL KONSTANTINOV, 48), (MARSEL BELYAEV, ...)]
2	ADEL KOSHELEV	[(ADEL KOSHELEV, 60), (ROMAN FADEEV, 3)]
3	ADEL RUSAKOV	[(ADEL RUSAKOV, 72), (ANGELINA LOBANOVA, 3)]
4	ADEL SNEGIREV	[(ADEL SNEGIREV, 70), (KRISTINA MOISEEVA, ...)]
...
911	ZARINA SOKOLOVA	[(ZARINA SOKOLOVA, 42), (RINAT SIZOV, 4), ...]
912	ZARINA TIMOFEEVA	[(ZARINA TIMOFEEVA, 53), (RADMIR VINOGRADOV, ...)]
913	ZLATA EFIMOVA	[(ZLATA EFIMOVA, 90), (RODION GROMOV, 3)]
914	ZLATA LOGINOVA	[(ZLATA LOGINOVA, 67), (MADINA RUDAKOVA, 3)]
915	ZLATA ZOTOVA	[(ZLATA ZOTOVA, 70), (FILIPP GLUSHKOV, 3)]

Итоговый фрейм после удаления летавших в одиночку

