Extraction process :

Contact collection

Function call to create a dataframe with the specified properties.

```
apiK = "pat-na1-3c7b0af9-bb66-40e7-a256-ce4c5eb27e81"
properties = ["allowed_to_collect", "firstname", "lastname","raw_email", "country","phone", "technical_test___create_date","industry","address","hs_object_id"]
limit = 200
DataFrame_Contacts = get_contacts(apiK,limit,properties)
```

✓ 19.1s

```
Loop #0
Loop #1
Loop #2
Loop #3
Loop #4
Loop #5
Loop #6
Loop #7
Loop #8
Loop #9
Loop #10
Loop #11
Loop #12
Loop #13
Loop #14
Loop #15
Loop #16
Loop #17
```

Requests of 200 contacts until obtaining more than 6,000 data contacts.

Result Extraction Contacts into a Data frame and Exporting as a .CSV

```
DataFrame_Contacts.to_csv('Technical_Test_OTF_Extraction.csv', index=False)
DataFrame_Contacts
```

✓ 0.1s

	address	allowed_to_collect	country	firstname	hs_object_id	industry	lastname	phone	raw_email	technical_test___create_date
0	Blackpool Rue, 6576	true	Waterford	Zoe	416102	Poultry and fish	Owen	0-774-386-624	Zoe <zoe_owen450104633@acrit.org> Contact Info.	2021-07-13
1	Parkfield Avenue, 5340	true	Ireland	Zara	413403	Fruit and vegetables	Rodwell	6-777-367-783	Zara <zara_rodwell1398442854@nickia.com> Conta...	2021-01-09
2	Abourne Lane, 876	true	Ireland	Zara	417951	Milling	Freeburn	5-618-556-540	Zara <zara_freeburn1593147546@gmail.com> Conta...	2021-08-30
3	Chester Crossroad, 7070	true	Dublin	Winnie	419852	Dairy products	Walter	1-161-604-327	Winnie <winnie_walter538064895@sheye.org> Cont...	2021-02-10
4	Tilloch Crossroad, 8332	true	Dublin	Zoe	425352	Meat	Owen	5-645-416-200	Zoe <zoe_owen1652446013@bungar.biz> Contact Info.	2021-11-02
...	...	...	...	...	...	...	...	...	...	...
6195	Elba Alley, 9380	true	Cork	Aiden	460366	Animal feeds	Clifford	8-731-470-025	Aiden <aiden_clifford794001861@atink.com> Cont...	2021-08-26
6196	Thomas Doyle Alley, 1931	true	Plymouth	Aiden	461382	Poultry and fish	Eyres	6-576-550-331	Aiden <aiden_eyres544152863@ubusive.com> Conta...	2021-05-30
6197	Parkfield Alley, 7350	true	Dublin	Aiden	462207	Dairy products	Eddison	5-250-504-442	Aiden <aiden_eddison1108996147@infotech44.tech...	2021-03-16
6198	Chelsea Manor Boulevard, 3151	true	England	Aiden	463002	Milling	Hope	1-121-762-622	Aiden <aiden_hope1270128896@kideod.biz> Contac...	2021-01-06
6199	Callcott Way, 1611	true	Dublin	Aiden	460472	Meat	Clifford	8-788-240-104	Aiden <aiden_clifford608122767@mafthy.com> Con...	2021-12-19

6200 rows × 10 columns

Creation of a dataframe to store the collected information and enable data transformation.

Transformation process :

Call the function to Country Recognition

```
DataFrame_Contacts['City/Country'] = DataFrame_Contacts['country'].apply(Country_Recognition)
print(DataFrame_Contacts['City/Country'])
```

✓ 6.3s

```
0            (Waterford, Ireland)
1                 (None, Ireland)
2                 (None, Ireland)
3                (Dublin, Ireland)
4                (Dublin, Ireland)
                    ...
6195              (Cork, Ireland)
6196    (Plymouth, United Kingdom)
6197             (Dublin, Ireland)
6198     (England, United Kingdom)
6199             (Dublin, Ireland)
Name: City/Country, Length: 6200, dtype: object
```

Creation of a property with the city and country data recognized by the created function.

Call the function to Found Emails

```
DataFrame_Contacts['Email'] = DataFrame_Contacts['raw_email'].apply(Found_Emails)
DataFrame_Contacts['Email']
```

✓ 0.0s

```
0                    zoe_owen450104633@acrit.org
1               zara_rodwell1398442854@nickia.com
2               zara_freeburn1593147546@gmail.com
3                winnie_walter538064895@sheye.org
4                   zoe_owen1652446013@bungar.biz
                          ...
6195            aiden_clifford794001861@atink.com
6196             aiden_eyres544152863@ubusive.com
6197    aiden_eddison1108996147@infotech44.tech
6198              aiden_hope1270128896@kideod.biz
6199           aiden_clifford608122767@mafthy.com
Name: Email, Length: 6200, dtype: object
```

Creation of a property to store the extracted Email.

Call the function to Fix Phone Numbers

```
DataFrame_Contacts['Phone_Number'] = Fix_Phone_Numbers(DataFrame_Contacts['City/Country'],DataFrame_Contacts['phone'])
DataFrame_Contacts['Phone_Number']
```

✓ 0.0s

```
0        (+353) 7 7438 6624
1       (+353) 67 7736 7783
2       (+353) 56 1855 6540
3       (+353) 11 6160 4327
4       (+353) 56 4541 6200
              ...
6195    (+353) 87 3147 0025
6196     (+44) 65 7655 0331
6197    (+353) 52 5050 4442
6198     (+44) 11 2176 2622
6199    (+353) 87 8824 0104
Name: Phone_Number, Length: 6200, dtype: object
```

Creation of a property to store the phone number in the required format and based on the country.

Clean Dtaframe, Result of Transformation and Exporting as a .CSV

```
DataFrame_Contacts.drop(columns=["allowed_to_collect","country","raw_email","phone"], inplace=True)
DataFrame_Contacts.to_csv('Technical_Test_OTF_Transformation.csv', index=False)
DataFrame_Contacts
```

✓ 0.1s

	address	firstname	hs_object_id	industry	lastname	technical_test__create_date	City/Country	Email	Phone_Number
0	Blackpool Rue, 6576	Zoe	416102	Poultry and fish	Owen	2021-07-13	(Waterford, Ireland)	zoe_owen450104633@acrit.org	(+353) 7 7438 6624
1	Parkfield Avenue, 5340	Zara	413403	Fruit and vegetables	Rodwell	2021-01-09	(None, Ireland)	zara_rodwell1398442854@nickia.com	(+353) 67 7736 7783
2	Abourne Lane, 876	Zara	417951	Milling	Freeburn	2021-08-30	(None, Ireland)	zara_freeburn1593147546@gmail.com	(+353) 56 1855 6540
3	Chester Crossroad, 7070	Winnie	419852	Dairy products	Walter	2021-02-10	(Dublin, Ireland)	winnie_walter538064895@sheye.org	(+353) 11 6160 4327
4	Tilloch Crossroad, 8332	Zoe	425352	Meat	Owen	2021-11-02	(Dublin, Ireland)	zoe_owen1652446013@bungar.biz	(+353) 56 4541 6200
...	...	...	...	...	...	...	...	...	...
6195	Elba Alley, 9380	Aiden	460366	Animal feeds	Clifford	2021-08-26	(Cork, Ireland)	aiden_clifford794001861@atink.com	(+353) 87 3147 0025
6196	Thomas Doyle Alley, 1931	Aiden	461382	Poultry and fish	Eyres	2021-05-30	(Plymouth, United Kingdom)	aiden_eyres544152863@ubusive.com	(+44) 65 7655 0331
6197	Parkfield Alley, 7350	Aiden	462207	Dairy products	Eddison	2021-03-16	(Dublin, Ireland)	aiden_eddison1108996147@infotech44.tech	(+353) 52 5050 4442
6198	Chelsea Manor Boulevard, 3151	Aiden	463002	Milling	Hope	2021-01-06	(England, United Kingdom)	aiden_hope1270128896@kideod.biz	(+44) 11 2176 2622
6199	Callcott Way, 1611	Aiden	460472	Meat	Clifford	2021-12-19	(Dublin, Ireland)	aiden_clifford608122767@mafthy.com	(+353) 87 8824 0104

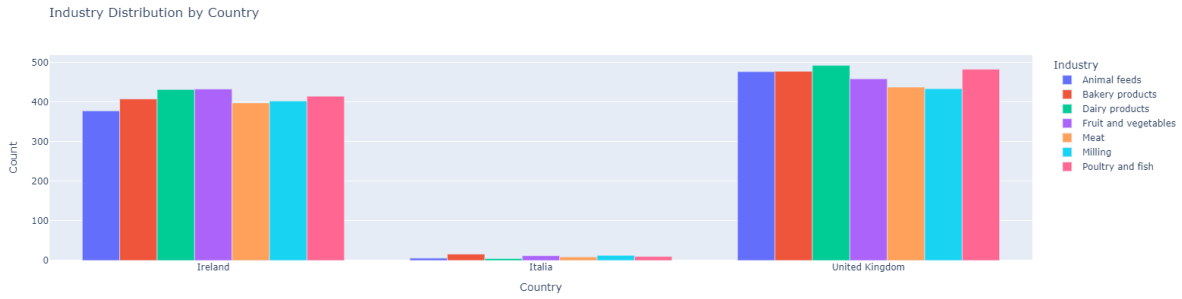
6200 rows × 9 columns

Dataframe with the result of the Transformation process.

Interactive graphic:

Creation of an interactive bar chart

```
DataFrame_Contacts['country'] = DataFrame_Contacts['City/Country'].apply(lambda x: x[1])

DataFrame_Count = DataFrame_Contacts.groupby(['industry', 'country']).size().reset_index(name='count')

Values_Industry = DataFrame_Contacts['industry'].unique().tolist()

Interactive_Chart = px.bar(DataFrame_Count,
                           x='country',
                           y='count',
                           color='industry',
                           barmode='group',
                           title='Industry Distribution by Country',
                           labels={'count': 'Count', 'country': 'Country', 'industry': 'Industry'})

Interactive_Chart.show()

DataFrame_Contacts.drop(columns=["country"], inplace=True)
```



Graphic of industry types by country.

Algorithm to duplicate management:

Call the algorithm to duplicate management

```
DataFrame_Contacts = Duplicate_Management(DataFrame_Contacts)
DataFrame_Contacts
```

✓ 2.4s

	address	firstname	hs_object_id	industry	lastname	technical_test__create_date	City/Country	Email	Phone_Number
0	Tilloch Crossroad, 8332	Zoe	425352	;Meat;Poultry and fish	Owen	2021-11-02	(Dublin, Ireland)	zoe_owen1652446013@bungar.biz	(+353) 56 4541 6200
1	Becklow Way, 8143	Zara	432351	Fruit and vegetables	Rodwell	2021-04-28	(Dublin, Ireland)	zara_rodwell1270511699@zorer.org	(+353) 33 8625 2475
2	Fieldstone Grove, 9968	Zara	428601	;Meat;Milling	Freeburn	2021-10-11	(Limerick, Ireland)	zara_freeburn2028884073@gembat.biz	(+353) 62 2415 6216
3	Collins Lane, 3994	Winnie	437452	Dairy products	Walter	2021-10-10	(None, Ireland)	winnie_walter1068546463@naiker.biz	(+353) 25 6760 4285
4	Jackson Hill, 547	Zara	427351	;Animal feeds;Meat	Oldfield	2021-11-12	(London, United Kingdom)	zara_oldfield587137840@ubusive.com	(+44) 76 1486 6578
...	...	...	...	...	...	...	...	...	...
2844	Fairfield Drive, 7609	Aiden	454193	;Animal feeds;Dairy products	Eddison	2021-04-12	(None, Ireland)	aiden_eddison148873831@atink.com	(+353) 20 0330 1221
2845	Buttonwood Tunnel, 9765	Aiden	454630	;Animal feeds;Milling	Cooper	2021-11-24	(Limerick, Ireland)	aiden_cooper489314766@ovock.tech	(+353) 33 7667 0122
2846	Belgrave Rue, 6292	Aiden	458519	;Bakery products;Milling	Hope	2021-04-11	(London, United Kingdom)	aiden_hope1577155349@dionrab.com	(+44) 63 1678 4584
2847	Chantry Alley, 6477	Aiden	460175	;Animal feeds;Poultry and fish	Eyres	2021-09-25	(Plymouth, United Kingdom)	aiden_eyres868665927@sheye.org	(+44) 14 0317 3282
2848	Callcott Way, 1611	Aiden	460472	;Meat;Animal feeds	Clifford	2021-12-19	(Dublin, Ireland)	aiden_clifford608122767@mafthy.com	(+353) 87 8824 0104

2849 rows × 9 columns

Dataframe with unique values by full name and email address.

Load process:

Determination and assignment of values for the industry property with the Multiple checkboxes data type in HubSpot.

Loading processed contacts into the personal HubSpot account.