

## 02\_\_crosswalk

May 26, 2024

### 1 Crosswalk from onet Data to Swiss Jobs

onet soc to soc crosswalk: <https://www.onetcenter.org/crosswalks.html>

soc to isco08 crosswalk: [https://www.bls.gov/soc/ISCO\\_SOC\\_Crosswalk.xls](https://www.bls.gov/soc/ISCO_SOC_Crosswalk.xls)

ch isco 19: [https://www.i14y.admin.ch/en/catalog/datasets/HCL\\_CH\\_ISCO\\_19\\_PROF\\_1\\_2/distributions](https://www.i14y.admin.ch/en/catalog/datasets/HCL_CH_ISCO_19_PROF_1_2/distributions)

#### 1.1 Content

1. Crosswalk onet Osborne and Frey Imports
2. Crosswalks
3. safe to db

```
[1]: # imports
import pandas as pd
from db import get_database
```

## Import Crosswalk files, data from Osborne & Frey and ONET data

```
[2]: # import onet crosswalk file
onet_file = "files/crosswalk_ONETSOC_SOC.xlsx"
sheet = "ONETSOC_SOC"

onetsoc_soc = pd.read_excel(io=onet_file, sheet_name=sheet , skiprows=3,
                             dtype=str)

# rename cols
onetsoc_soc = onetsoc_soc.rename(columns={'onet-SOC 2019 Code': 'onetsoc19'})
onetsoc_soc = onetsoc_soc.rename(columns={'2018 SOC Code': '2010soc'})

# List of columns to drop
cols_to_drop = ['onet-SOC 2019 Title', '2018 SOC Title']

# Drop specified columns
onetsoc_soc = onetsoc_soc.drop(columns=cols_to_drop)
onetsoc_soc = onetsoc_soc.dropna()
print(onetsoc_soc.shape[0])

onetsoc_soc.head(2)
```

1016

```
[2]:      onetsoc19  2010soc  
0  11-1011.00  11-1011  
1  11-1011.03  11-1011
```

```
[3]: # import ISCO crosswalk file  
file_name = "files/crosswalk_ISCO_SOC.xls"  
sheet = "ISCO_SOC"  
  
isco_soc = pd.read_excel(io=file_name, sheet_name=sheet , skiprows=6, dtype=str)  
  
# rename cols  
isco_soc = isco_soc.rename(columns={'ISCO-08 Code': 'isco08'})  
isco_soc = isco_soc.rename(columns={'2010 SOC Code': '2010soc'})  
  
# List of columns to drop  
cols_to_drop = ['ISCO-08 Title EN', 'part', '2010 SOC Title', "Comment 8/17/11"]  
  
# Drop specified columns  
isco_soc = isco_soc.drop(columns=cols_to_drop)  
isco_soc = isco_soc.dropna()  
print(isco_soc.shape[0])  
  
isco_soc.head(2)
```

1125

```
[3]:      isco08  2010soc  
0    0110  55-1011  
1    0110  55-1012
```

```
[4]: # import CH crosswalk file  
file_name = "files/crosswalk_CH_ISCO_19.xlsx"  
sheet = "CH_ISCO_19"  
  
ch_isco = pd.read_excel(io=file_name, sheet_name=sheet, skiprows=0, dtype=str)  
  
# rename cols  
ch_isco = ch_isco.rename(columns={'Berufsgattungen': 'isco08'})  
  
# List of columns to drop  
cols_to_drop = ["Berufshauptgruppen", 'Berufsgruppen', 'Berufsuntergruppen',  
↳ "Berufsarten", "Code"]  
  
# Drop specified columns  
ch_isco = ch_isco.drop(columns=cols_to_drop)  
ch_isco = ch_isco.dropna()
```

```
print(ch_isco.shape[0])

ch_isco.head(2)
```

582

```
[4]: isco08                                     Name_de
3   0110      Offiziere in regulären Streitkräften
7   0210  Unteroffiziere in regulären Streitkräften
```

```
[5]: # import data from onet from mongodb
dbname = get_database()

collection_name = dbname["onet"]

data = collection_name.find()
onet = pd.DataFrame(list(data))
onet.head(2)
```

```
[5]:      _id      code display \
0  6620df94dde8b852bac8e7de  27-2011.00    long
1  6620df94dde8b852bac8e7dd  13-2011.00    long

      occupation \
0  {'code': '27-2011.00', 'title': 'Actors', 'tag...
1  {'code': '13-2011.00', 'title': 'Accountants a...

      tasks \
0  {'task': [{'id': 7646, 'green': False, 'relate...
1  {'task': [{'id': 21505, 'green': False, 'relat...

      technology_skills \
0  {'category': [{'related': 'https://services.on...
1  {'see_all': {'href': 'https://services.onetcen...

      tools_used \
0  {'category': [{'related': 'https://services.on...
1  {'category': [{'related': 'https://services.on...

      tools_technology \
0  {'tools': {'category': [{'related': 'https://s...
1  {'tools': {'category': [{'related': 'https://s...

      knowledge \
0  {'element': [{'id': '2.C.7.c', 'related': 'htt...
1  {'element': [{'id': '2.C.1.c', 'related': 'htt...

      skills ... \
```

```

0 {'element': [{'id': '2.A.1.a', 'related': 'htt... ...
1 {'element': [{'id': '2.A.1.b', 'related': 'htt... ...

                                work_activities \
0 {'element': [{'id': '4.A.4.a.4', 'related': 'h...
1 {'element': [{'id': '4.A.1.a.1', 'related': 'h...

                                detailed_work_activities \
0 {'activity': [{'id': '4.A.4.a.8.I01.D04', 'rel...
1 {'activity': [{'id': '4.A.3.b.6.I01.D02', 'rel...

                                work_context \
0 {'element': [{'id': '4.C.1.b.1.e', 'related': ...
1 {'element': [{'id': '4.C.1.a.2.h', 'related': ...

                                job_zone \
0 {'value': 2, 'title': 'Job Zone Two: Some Prep...
1 {'value': 4, 'title': 'Job Zone Four: Consider...

                                education \
0 {'level_required': {'category': [{'name': 'Les...
1                                     NaN

                                interests \
0 {'element': [{'id': '1.B.1.c', 'related': 'htt...
1 {'element': [{'id': '1.B.1.f', 'related': 'htt...

                                work_styles \
0 {'element': [{'id': '1.C.3.a', 'related': 'htt...
1 {'element': [{'id': '1.C.5.b', 'related': 'htt...

                                work_values \
0 {'element': [{'id': '1.B.2.d', 'related': 'htt...
1 {'element': [{'id': '1.B.2.a', 'related': 'htt...

                                related_occupations \
0 {'occupation': [{'href': 'https://services.one...
1 {'occupation': [{'href': 'https://services.one...

                                additional_information
0 {'source': [{'url': 'https://www.actorsequity...
1 {'source': [{'url': 'https://www.aacsb.edu/', ...

[2 rows x 21 columns]

```

```

[6]: # import Source Frey / Osborne
      fo = pd.read_csv("files/F0.csv")

```

```
[7]: # Convert 'code' column to string
onet['code'] = onet['code'].astype(str)
fo["code"] = fo["code"].astype(str)

# Rename 'code' column to 'onetsoc19'
onet = onet.rename(columns={'code': 'onetsoc19'})
fo = fo.rename(columns={'code': '2010soc'})

print(onet.shape[0])
```

1016

```
[8]: onet.head(2)
```

```
[8]:      _id  onetsoc19 display \
0  6620df94dde8b852bac8e7de  27-2011.00    long
1  6620df94dde8b852bac8e7dd  13-2011.00    long

      occupation \
0  {'code': '27-2011.00', 'title': 'Actors', 'tag...
1  {'code': '13-2011.00', 'title': 'Accountants a...

      tasks \
0  {'task': [{'id': 7646, 'green': False, 'relate...
1  {'task': [{'id': 21505, 'green': False, 'relat...

      technology_skills \
0  {'category': [{'related': 'https://services.on...
1  {'see_all': {'href': 'https://services.onetcen...

      tools_used \
0  {'category': [{'related': 'https://services.on...
1  {'category': [{'related': 'https://services.on...

      tools_technology \
0  {'tools': {'category': [{'related': 'https://s...
1  {'tools': {'category': [{'related': 'https://s...

      knowledge \
0  {'element': [{'id': '2.C.7.c', 'related': 'htt...
1  {'element': [{'id': '2.C.1.c', 'related': 'htt...

      skills ... \
0  {'element': [{'id': '2.A.1.a', 'related': 'htt... ...
1  {'element': [{'id': '2.A.1.b', 'related': 'htt... ...

      work_activities \
0  {'element': [{'id': '4.A.4.a.4', 'related': 'h...
```

```

1 {'element': [{'id': '4.A.1.a.1', 'related': 'h...

                                detailed_work_activities \
0 {'activity': [{'id': '4.A.4.a.8.I01.D04', 'rel...
1 {'activity': [{'id': '4.A.3.b.6.I01.D02', 'rel...

                                work_context \
0 {'element': [{'id': '4.C.1.b.1.e', 'related': ...
1 {'element': [{'id': '4.C.1.a.2.h', 'related': ...

                                job_zone \
0 {'value': 2, 'title': 'Job Zone Two: Some Prep...
1 {'value': 4, 'title': 'Job Zone Four: Consider...

                                education \
0 {'level_required': {'category': [{'name': 'Les...
1                                     NaN

                                interests \
0 {'element': [{'id': '1.B.1.c', 'related': 'htt...
1 {'element': [{'id': '1.B.1.f', 'related': 'htt...

                                work_styles \
0 {'element': [{'id': '1.C.3.a', 'related': 'htt...
1 {'element': [{'id': '1.C.5.b', 'related': 'htt...

                                work_values \
0 {'element': [{'id': '1.B.2.d', 'related': 'htt...
1 {'element': [{'id': '1.B.2.a', 'related': 'htt...

                                related_occupations \
0 {'occupation': [{'href': 'https://services.one...
1 {'occupation': [{'href': 'https://services.one...

                                additional_information
0 {'source': [{'url': 'https://www.actorsequity...
1 {'source': [{'url': 'https://www.aacsb.edu/', ...

[2 rows x 21 columns]

```

## Crosswalks

```

[10]: # crosswalk onet soc
joined = pd.merge(onet, onetsoc_soc, left_on='onetsoc19', right_on="onetsoc19",
                  how='left')
joined = joined.dropna()

```

```
# print amount of cols
print(joined.shape[0])
```

851

```
[11]: # crosswalk soc
joined = pd.merge(joined, isco_soc, left_on='2010soc', right_on="2010soc",
↳how='left')
fo_joined = pd.merge(fo, isco_soc, left_on='2010soc', right_on="2010soc",
↳how='left')

# print amount of cols
print(joined.shape[0])
```

1065

```
[14]: # crosswalk isco08
joined = pd.merge(joined, ch_isco, left_on='isco08', right_on="isco08",
↳how='left')
fo_joined = pd.merge(fo_joined, ch_isco, left_on='isco08', right_on="isco08",
↳how='left')

# print amount of cols
print(joined.shape[0])
print(fo_joined.shape[0])
```

1065

803

```
[17]: # drop na values
joined = joined.dropna()
joined = joined.drop(columns=["_id"])
fo_joined = fo_joined.dropna()
print(joined.shape[0])
```

846

```
[17]:      onetsoc19 display      occupation \
0  27-2011.00    long {'code': '27-2011.00', 'title': 'Actors', 'tag...
1  23-1021.00    long {'code': '23-1021.00', 'title': 'Administrativ...

      tasks \
0 {'task': [{'id': 7646, 'green': False, 'relate...
1 {'task': [{'id': 7627, 'green': False, 'relate...

      technology_skills \
0 {'category': [{'related': 'https://services.on...
1 {'category': [{'related': 'https://services.on...
```

```

                                tools_used \
0  {'category': [{'related': 'https://services.on...
1  {'category': [{'related': 'https://services.on...

                                tools_technology \
0  {'tools': {'category': [{'related': 'https://s...
1  {'tools': {'category': [{'related': 'https://s...

                                knowledge \
0  {'element': [{'id': '2.C.7.c', 'related': 'htt...
1  {'element': [{'id': '2.C.8.b', 'related': 'htt...

                                skills \
0  {'element': [{'id': '2.A.1.a', 'related': 'htt...
1  {'element': [{'id': '2.A.1.b', 'related': 'htt...

                                abilities ... \
0  {'element': [{'id': '1.A.1.a.3', 'related': 'h... ...
1  {'element': [{'id': '1.A.1.b.5', 'related': 'h... ...

                                job_zone \
0  {'value': 2, 'title': 'Job Zone Two: Some Prep...
1  {'value': 5, 'title': 'Job Zone Five: Extensiv...

                                education \
0  {'level_required': {'category': [{'name': 'Les...
1  {'level_required': {'category': [{'name': 'Doc...

                                interests \
0  {'element': [{'id': '1.B.1.c', 'related': 'htt...
1  {'element': [{'id': '1.B.1.f', 'related': 'htt...

                                work_styles \
0  {'element': [{'id': '1.C.3.a', 'related': 'htt...
1  {'element': [{'id': '1.C.5.c', 'related': 'htt...

                                work_values \
0  {'element': [{'id': '1.B.2.d', 'related': 'htt...
1  {'element': [{'id': '1.B.2.a', 'related': 'htt...

                                related_occupations \
0  {'occupation': [{'href': 'https://services.one...
1  {'occupation': [{'href': 'https://services.one...

                                additional_information 2010soc isco08 \
0  {'source': [{'url': 'https://www.actorsequity... 27-2011 2655
1  {'source': [{'url': 'https://www.americanbar.o... 23-1021 2612

```



```

        Name_de
0  Schauspieler
1      Richter

[2 rows x 23 columns]

```

```

[19]: # drop further cols
joined = joined.drop(columns=["onetsoc19", "2010soc"])
fo_joined = fo_joined.drop(columns=["2010soc", "Occupation", "Name_de", "Rank"])
fo_joined = fo_joined.rename(columns={'Label': '
↳ 'fo_computerisation', "Probability": "fo_probability"})

```

```

### safe to db

```

```

[22]: # save to mongodb
dbname = get_database()

collection_name = dbname["joined"]
fo_col = dbname["fo_joined"]

# Insert the dictionaries into the MongoDB collection
collection_name.insert_many(joined.to_dict("records"))
fo_col.insert_many(fo_joined.to_dict("records"))

```

```

[22]: InsertManyResult([ObjectId('6649fba78a508636be6bfd83'),
ObjectId('6649fba78a508636be6bfd84'), ObjectId('6649fba78a508636be6bfd85'),
ObjectId('6649fba78a508636be6bfd86'), ObjectId('6649fba78a508636be6bfd87'),
ObjectId('6649fba78a508636be6bfd88'), ObjectId('6649fba78a508636be6bfd89'),
ObjectId('6649fba78a508636be6bfd8a'), ObjectId('6649fba78a508636be6bfd8b'),
ObjectId('6649fba78a508636be6bfd8c'), ObjectId('6649fba78a508636be6bfd8d'),
ObjectId('6649fba78a508636be6bfd8e'), ObjectId('6649fba78a508636be6bfd8f'),
ObjectId('6649fba78a508636be6bfd90'), ObjectId('6649fba78a508636be6bfd91'),
ObjectId('6649fba78a508636be6bfd92'), ObjectId('6649fba78a508636be6bfd93'),
ObjectId('6649fba78a508636be6bfd94'), ObjectId('6649fba78a508636be6bfd95'),
ObjectId('6649fba78a508636be6bfd96'), ObjectId('6649fba78a508636be6bfd97'),
ObjectId('6649fba78a508636be6bfd98'), ObjectId('6649fba78a508636be6bfd99'),
ObjectId('6649fba78a508636be6bfd9a'), ObjectId('6649fba78a508636be6bfd9b'),
ObjectId('6649fba78a508636be6bfd9c'), ObjectId('6649fba78a508636be6bfd9d'),
ObjectId('6649fba78a508636be6bfd9e'), ObjectId('6649fba78a508636be6bfd9f'),
ObjectId('6649fba78a508636be6bfda0'), ObjectId('6649fba78a508636be6bfda1'),
ObjectId('6649fba78a508636be6bfda2'), ObjectId('6649fba78a508636be6bfda3'),
ObjectId('6649fba78a508636be6bfda4'), ObjectId('6649fba78a508636be6bfda5'),
ObjectId('6649fba78a508636be6bfda6'), ObjectId('6649fba78a508636be6bfda7'),
ObjectId('6649fba78a508636be6bfda8'), ObjectId('6649fba78a508636be6bfda9'),
ObjectId('6649fba78a508636be6bfdaa'), ObjectId('6649fba78a508636be6bfdad'),
ObjectId('6649fba78a508636be6bfdac'), ObjectId('6649fba78a508636be6bfdad'),
ObjectId('6649fba78a508636be6bfdae'), ObjectId('6649fba78a508636be6bfdaf'),

```



[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```
ObjectId('6649fba78a508636be6bffea'), ObjectId('6649fba78a508636be6bfef5'),  
ObjectId('6649fba78a508636be6bffeb'), ObjectId('6649fba78a508636be6bffec'), ObjectId('6649fba78a508636be6bffed'),  
ObjectId('6649fba78a508636be6bffee'), ObjectId('6649fba78a508636be6bffef'),  
ObjectId('6649fba78a508636be6bfffb'), ObjectId('6649fba78a508636be6bfffc'), ObjectId('6649fba78a508636be6bfffd'),  
ObjectId('6649fba78a508636be6bfffe'), ObjectId('6649fba78a508636be6bffff'),  
ObjectId('6649fba78a508636be6c0000'), ObjectId('6649fba78a508636be6c0001'),  
ObjectId('6649fba78a508636be6c0002'), ObjectId('6649fba78a508636be6c0003'),  
ObjectId('6649fba78a508636be6c0004'), ObjectId('6649fba78a508636be6c0005'),  
ObjectId('6649fba78a508636be6c0006'), ObjectId('6649fba78a508636be6c0007'),  
ObjectId('6649fba78a508636be6c0008'), ObjectId('6649fba78a508636be6c0009'),  
ObjectId('6649fba78a508636be6c000a'), ObjectId('6649fba78a508636be6c000b'),  
ObjectId('6649fba78a508636be6c000c'), ObjectId('6649fba78a508636be6c000d'),  
ObjectId('6649fba78a508636be6c000e'), ObjectId('6649fba78a508636be6c000f'),  
ObjectId('6649fba78a508636be6c0010'), ObjectId('6649fba78a508636be6c0011'),  
ObjectId('6649fba78a508636be6c0012'), ObjectId('6649fba78a508636be6c0013'),  
ObjectId('6649fba78a508636be6c0014'), ObjectId('6649fba78a508636be6c0015'),  
ObjectId('6649fba78a508636be6c0016'), ObjectId('6649fba78a508636be6c0017'),  
ObjectId('6649fba78a508636be6c0018'), ObjectId('6649fba78a508636be6c0019'),  
ObjectId('6649fba78a508636be6c001a'), ObjectId('6649fba78a508636be6c001b'),  
ObjectId('6649fba78a508636be6c001c'), ObjectId('6649fba78a508636be6c001d'),  
ObjectId('6649fba78a508636be6c001e'), ObjectId('6649fba78a508636be6c001f'),  
ObjectId('6649fba78a508636be6c0020'), ObjectId('6649fba78a508636be6c0021'),  
ObjectId('6649fba78a508636be6c0022'), ObjectId('6649fba78a508636be6c0023'),  
ObjectId('6649fba78a508636be6c0024'), ObjectId('6649fba78a508636be6c0025'),  
ObjectId('6649fba78a508636be6c0026'), ObjectId('6649fba78a508636be6c0027')],  
acknowledged=True)
```