

# data\_preparation

June 12, 2023

## 0.1 Import necessary libraries

```
[40]: pip install dbfread
```

Requirement already satisfied: dbfread in  
c:\users\nichola.ondiek\anaconda3\lib\site-packages (2.0.7)  
Note: you may need to restart the kernel to use updated packages.

```
[41]: import os
import glob
import numpy as np
import pandas as pd
import shutil
from dbfread import DBF
from pandas import DataFrame
from shutil import copyfile

pd.set_option('display.max_columns', None)
```

```
[ ]:
```

## 0.2 Read files

```
[42]: df = pd.read_csv(r'ds_salaries.csv')
```

```
[43]: df.head()
```

```
[43]: Unnamed: 0  work_year  experience_level  employment_type  \
0           0         2020                MI              FT
1           1         2020                SE              FT
2           2         2020                SE              FT
3           3         2020                MI              FT
4           4         2020                SE              FT

           job_title  salary  salary_currency  salary_in_usd  \
0      Data Scientist   70000              EUR         79833
1  Machine Learning Scientist  260000              USD        260000
```

2	Big Data Engineer	85000	GBP	109024
3	Product Data Analyst	20000	USD	20000
4	Machine Learning Engineer	150000	USD	150000

	employee_residence	remote_ratio	company_location	company_size
0	DE	0	DE	L
1	JP	0	JP	S
2	GB	50	GB	M
3	HN	0	HN	S
4	US	50	US	L

```
[44]: df.shape
```

```
[44]: (607, 12)
```

```
[45]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 607 entries, 0 to 606
Data columns (total 12 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Unnamed: 0            607 non-null   int64
1   work_year             607 non-null   int64
2   experience_level      607 non-null   object
3   employment_type       607 non-null   object
4   job_title             607 non-null   object
5   salary                607 non-null   int64
6   salary_currency       607 non-null   object
7   salary_in_usd         607 non-null   int64
8   employee_residence    607 non-null   object
9   remote_ratio          607 non-null   int64
10  company_location      607 non-null   object
11  company_size          607 non-null   object
dtypes: int64(5), object(7)
memory usage: 57.0+ KB
```

```
[46]: df.describe()
```

	Unnamed: 0	work_year	salary	salary_in_usd	remote_ratio
count	607.000000	607.000000	6.070000e+02	607.000000	607.000000
mean	303.000000	2021.405272	3.240001e+05	112297.869852	70.92257
std	175.370085	0.692133	1.544357e+06	70957.259411	40.70913
min	0.000000	2020.000000	4.000000e+03	2859.000000	0.000000
25%	151.500000	2021.000000	7.000000e+04	62726.000000	50.000000
50%	303.000000	2022.000000	1.150000e+05	101570.000000	100.000000
75%	454.500000	2022.000000	1.650000e+05	150000.000000	100.000000
max	606.000000	2022.000000	3.040000e+07	600000.000000	100.000000

```
[ ]:
```

### 0.3 Filter data - one condition

```
[47]: df[df['work_year'] == 2020]
```

```
[47]:
```

	Unnamed: 0	work_year	experience_level	employment_type	\
0	0	2020	MI	FT	
1	1	2020	SE	FT	
2	2	2020	SE	FT	
3	3	2020	MI	FT	
4	4	2020	SE	FT	
..	...	...	...	...	
67	67	2020	SE	FT	
68	68	2020	EN	FT	
69	69	2020	SE	FT	
70	70	2020	MI	FT	
71	71	2020	MI	FT	

  

	job_title	salary	salary_currency	salary_in_usd	\
0	Data Scientist	70000	EUR	79833	
1	Machine Learning Scientist	260000	USD	260000	
2	Big Data Engineer	85000	GBP	109024	
3	Product Data Analyst	20000	USD	20000	
4	Machine Learning Engineer	150000	USD	150000	
..	...	...	...	...	
67	Data Science Manager	190200	USD	190200	
68	Data Scientist	105000	USD	105000	
69	Data Scientist	80000	EUR	91237	
70	Data Scientist	55000	EUR	62726	
71	Data Scientist	37000	EUR	42197	

  

	employee_residence	remote_ratio	company_location	company_size
0	DE	0	DE	L
1	JP	0	JP	S
2	GB	50	GB	M
3	HN	0	HN	S
4	US	50	US	L
..	...	...	...	...
67	US	100	US	M
68	US	100	US	S
69	AT	0	AT	S
70	FR	50	LU	S
71	FR	50	FR	S

```
[72 rows x 12 columns]
```

```
[ ]:
```

## 0.4 Filter data - multiple conditions

```
[69]: df[(df['EDCODE'] == 990) & (df['PAYROLLNUM'] == 20150070367)]
```

```
[69]: Empty DataFrame
Columns: [SOURCE, PAYROLLNUM, SALUTE, SURNAME, FIRSTNAME, OTHERNAMES, EDCODE,
AMOUNT, TAXABLE, BALANCE, AGENTCODE, AGENTTYPE, REF1, REF2, REF3, REFACCOUNT,
IDNUM, TAXPIN, REGNUM, GENDER, BLOODGRP, MARITAL, ETHNICCODE, RELIGION, EDUPEAK,
HCONTYCODE, HSCOUTCODE, HCWARDCODE, HVILAGCODE, SPECIONEED, PAYGROUP, DESIGCODE,
TITLECODE, JOBGROUP, SCALEPOINT, INCRMONTH, LINCRCODE, BIRTHDATE, DATEHIRED,
DATEOFPOST, ENGAGECODE, PENSONCODE, ENDDATE, VOTECODE, WCOUTCYCODE, SCOUTYCODE,
CWARDCODE, VILAGECODE, STATONCODE, PAYMODE, BANKCODE, BRANCHCODE, ACCOUNTNUM,
PAYBOOK, SUBVCODE, VHEADCODE, SUBHCODE, HOUSECODE, DEFERRED, PAYDATE, ADSTATUS,
FLEXIBO, WHOEDITED, DATEEDITED, TIMEEDITED, EDITCOUNT, ROWNUM, Full-Name]
Index: []
```

```
[ ]:
```

```
[70]: df[(df['EDCODE'] == 990) | (df['PAYROLLNUM'] == 1977029253)]
```

```
[70]:
```

	SOURCE	PAYROLLNUM	SALUTE	SURNAME	FIRSTNAME	OTHERNAMES	\
0	PYR1807c.csv	1977029253	Mr	Shambe	Mohamed	Mahmoud	
1	PYR1807c.csv	1977029253	Mr	Shambe	Mohamed	Mahmoud	
2	PYR1807c.csv	1977029253	Mr	Shambe	Mohamed	Mahmoud	
3	PYR1807c.csv	1977029253	Mr	Shambe	Mohamed	Mahmoud	
4	PYR1807c.csv	1977029253	Mr	Shambe	Mohamed	Mahmoud	
...	...	...	...	...	...	...	
136656	PYR1906c.csv	20150076065	Mr	Mwashangi	Austine	Gona	
138528	PYR1906c.csv	20170017937	Miss	Mbuthia	Anastacia	Wangari	
138529	PYR1906c.csv	20170017937	Miss	Mbuthia	Anastacia	Wangari	
138748	PYR1906c.csv	20170021431	Mr	Murungi	Erick	Munene	
138876	PYR1906c.csv	20170021511	Mr	Lagat	Felix	Kipngetich	

  

	EDCODE	AMOUNT	TAXABLE	BALANCE	AGENTCODE	AGENTTYPE	REF1	\
0	801	61840.00	1	0.00	0	0	0	
1	802	13000.00	1	0.00	0	0	60	
2	805	14650.00	1	0.00	0	0	0	
3	813	8000.00	1	0.00	0	0	0	
4	877	20000.00	1	0.00	0	0	0	
...	...	...	...	...	...	...	...	
136656	990	4500.00	0	191000.00	0	0	33	
138528	990	0.00	0	3865.25	0	0	0	
138529	990	3865.25	0	0.00	0	0	0	
138748	990	7229.60	0	0.00	0	0	0	
138876	990	34402.85	0	0.00	0	0	0	

	REF2	REF3	REFACCOUNT	IDNUM	TAXPIN	REGNUM	GENDER	\
0	0	0	NaN	158828	A001572986G	0	M	
1	4	0	NaN	158828	A001572986G	0	M	
2	0	0	NaN	158828	A001572986G	0	M	
3	0	0	NaN	158828	A001572986G	0	M	
4	0	0	NaN	158828	A001572986G	0	M	
...	...	...	...	...	...	...	...	
136656	0	0	OVERPAYMENT	25971490	A009327679N	0	M	
138528	0	0	OVERPAYMENT REC	29559181	A010810769X	0	F	
138529	0	0	OVERPAYMENT REC	29559181	A010810769X	0	F	
138748	0	0	OVERPAYMENT REC	29115847	A009938337P	0	M	
138876	0	0	OVERPAYMENT REC	29836409	A009707875T	0	M	

	BLOODGRP	MARITAL	ETHNICCODE	RELIGION	EDUPEAK	HCONTYCODE	\
0	NaN	M	1	0	5	0	
1	NaN	M	1	0	5	0	
2	NaN	M	1	0	5	0	
3	NaN	M	1	0	5	0	
4	NaN	M	1	0	5	0	
...	...	...	...	...	...	...	
136656	NaN	M	16	0	14	0	
138528	NaN	S	4	0	6	0	
138529	NaN	S	4	0	6	0	
138748	NaN	S	6	0	6	0	
138876	NaN	S	23	0	6	0	

	HSCOUTCODE	HWARDCODE	HVILAGCODE	SPECIONEED	PAYGROUP	DESIGCODE	\
0	0	0	0	0	AA	C16AM	
1	0	0	0	0	AA	C16AM	
2	0	0	0	0	AA	C16AM	
3	0	0	0	0	AA	C16AM	
4	0	0	0	0	AA	C16AM	
...	...	...	...	...	...	...	
136656	5	105	1	0	BA	T04BD	
138528	22	105	105	0	BA	T04BD	
138529	22	105	105	0	BA	T04BD	
138748	22	105	105	0	BA	F10AH	
138876	22	105	105	0	BA	F13BH	

	TITLECODE	JOBGROUP	SCALEPOINT	INCRMONTH	LINCRDATE	BIRTHDATE	\
0	NaN	M	7	1	1/1/2017	9/13/1959	
1	NaN	M	7	1	1/1/2017	9/13/1959	
2	NaN	M	7	1	1/1/2017	9/13/1959	
3	NaN	M	7	1	1/1/2017	9/13/1959	
4	NaN	M	7	1	1/1/2017	9/13/1959	
...	...	...	...	...	...	...	

136656	NaN	D	4	11	11/1/2018	9/26/1986
138528	NaN	D	1	2	2/1/2019	9/21/1988
138529	NaN	D	1	2	2/1/2019	9/21/1988
138748	NaN	H	3	4	4/1/2019	7/17/1991
138876	NaN	H	3	4	4/1/2019	8/28/1992

	DATEHIRED	DATEOFPOST	ENGAGECODE	PENSONCODE	ENDDATE	VOTECODE	\
0	6/20/1977	1/1/2014	3	1	NaN	30501	
1	6/20/1977	1/1/2014	3	1	NaN	30501	
2	6/20/1977	1/1/2014	3	1	NaN	30501	
3	6/20/1977	1/1/2014	3	1	NaN	30501	
4	6/20/1977	1/1/2014	3	1	NaN	30501	
...	...	...	...	...	...	...	
136656	11/1/2015	11/1/2015	3	16	NaN	30501	
138528	1/4/2017	2/1/2019	5	0	1/31/2020	30501	
138529	1/4/2017	2/1/2019	5	0	1/31/2020	30501	
138748	4/3/2017	4/3/2017	3	16	NaN	30501	
138876	4/3/2017	4/3/2017	3	16	NaN	30501	

	WCOUTYCODE	SCOUTYCODE	CWARDCODE	VILAGECODE	STATONCODE	PAYMODE	\
0	5	22	105	1	8149	B	
1	5	22	105	1	8149	B	
2	5	22	105	1	8149	B	
3	5	22	105	1	8149	B	
4	5	22	105	1	8149	B	
...	...	...	...	...	...	...	
136656	5	22	105	1	1	B	
138528	5	22	105	1	1	B	
138529	5	22	105	1	1	B	
138748	5	22	106	1	7	B	
138876	5	22	105	1	1	B	

	BANKCODE	BRANCHCODE	ACCOUNTNUM	PAYBOOK	SUBVCODE	VHEADCODE	\
0	1	173	1104755645	3	888	149	
1	1	173	1104755645	3	888	149	
2	1	173	1104755645	3	888	149	
3	1	173	1104755645	3	888	149	
4	1	173	1104755645	3	888	149	
...	...	...	...	...	...	...	
136656	1	173	1110207492	5	301	301	
138528	1	173	1208028332	7	304	304	
138529	1	173	1208028332	7	304	304	
138748	1	173	1202346219	1	702	702	
138876	1	173	1197610634	1	702	702	

	SUBHCODE	HOUSECODE	DEFERRED	PAYDATE	ADSTATUS	FLEXIBO	\
0	335	0	0	7/26/2018	0	14901	

1	335	0	0	7/26/2018	0	14901
2	335	0	0	7/26/2018	0	14901
3	335	0	0	7/26/2018	0	14901
4	335	0	0	7/26/2018	0	14901
...	...	...	...	...	...	...
136656	301	0	0	6/11/2019	-79	0
138528	304	0	0	6/11/2019	-98	0
138529	304	0	1	6/11/2019	-98	0
138748	702	0	0	6/11/2019	-98	0
138876	702	0	0	6/11/2019	-98	0

	WHOEDITED	DATEEDITED	TIMEEDITED	EDITCOUNT	ROWNUM	\
0	20140004388	6/6/2018	17:54:47	13	1	
1	20140004388	6/6/2018	17:54:47	13	2	
2	20140004388	6/6/2018	17:54:47	13	3	
3	20140004388	6/6/2018	17:54:47	13	4	
4	20140004388	6/6/2018	17:54:47	13	5	
...	...	...	...	...	...	...
136656	20100004224	1/17/2017	11:09:13	1	7802	
138528	20100004224	6/7/2019	9:24:19	1	9674	
138529	20100004224	6/7/2019	9:24:19	1	9675	
138748	20100004224	6/7/2019	9:27:08	1	9894	
138876	20100004224	6/7/2019	9:28:03	1	10022	

	Full-Name
0	Mohamed Mahmoud Shambe
1	Mohamed Mahmoud Shambe
2	Mohamed Mahmoud Shambe
3	Mohamed Mahmoud Shambe
4	Mohamed Mahmoud Shambe
...	...
136656	Austine Gona Mwashangi
138528	Anastacia Wangari Mbuthia
138529	Anastacia Wangari Mbuthia
138748	Erick Munene Murungi
138876	Felix Kipngetich Lagat

[178 rows x 68 columns]

[ ]:

## 0.5 Select specific columns from data

```
[ ]: df.loc[:5,['BranchID', 'Client ID', 'Gender ID', 'Account ID', 'Legacy Account_ID']]
df.iloc[:5,[1,2,3,4,5]]
```

```
[ ]:
```

## 0.6 Identify unique values in data

```
[49]: df.work_year.unique()
```

```
[49]: array([2020, 2021, 2022], dtype=int64)
```

```
[ ]:
```

```
[50]: df.value_counts('salary_currency')
```

```
[50]: salary_currency
```

```
USD      398
EUR       95
GBP       44
INR       27
CAD       18
JPY        3
PLN        3
TRY        3
CNY        2
DKK        2
BRL        2
HUF        2
MXN        2
SGD        2
AUD        2
CHF        1
CLP        1
dtype: int64
```

```
[ ]:
```

## 0.7 Highlight duplicated values in data

```
[51]: df[df.duplicated(['salary_currency'])] # Can take mutiple column arguments in_
      ↪ duplicated
```

```
[51]:
```

	Unnamed: 0	work_year	experience_level	employment_type	\
3	3	2020	MI	FT	
4	4	2020	SE	FT	
5	5	2020	EN	FT	
6	6	2020	SE	FT	
8	8	2020	MI	FT	
..	...	...	...	...	
602	602	2022	SE	FT	



603	603	2022	SE	FT
604	604	2022	SE	FT
605	605	2022	SE	FT
606	606	2022	MI	FT

	job_title	salary	salary_currency	salary_in_usd	\
3	Product Data Analyst	20000	USD	20000	
4	Machine Learning Engineer	150000	USD	150000	
5	Data Analyst	72000	USD	72000	
6	Lead Data Scientist	190000	USD	190000	
8	Business Data Analyst	135000	USD	135000	
..	...	...	...	...	
602	Data Engineer	154000	USD	154000	
603	Data Engineer	126000	USD	126000	
604	Data Analyst	129000	USD	129000	
605	Data Analyst	150000	USD	150000	
606	AI Scientist	200000	USD	200000	

	employee_residence	remote_ratio	company_location	company_size
3	HN	0	HN	S
4	US	50	US	L
5	US	100	US	L
6	US	100	US	S
8	US	100	US	L
..	...	...	...	...
602	US	100	US	M
603	US	100	US	M
604	US	0	US	M
605	US	100	US	M
606	IN	100	US	L

[590 rows x 12 columns]

## 0.8 Split file by column values to multiple csvs

```
[53]: for i, x in df.groupby('work_year'):
        x.to_csv(r"C:\Users\nichola.ondiek\Python\New folder\{}.csv".format(i),
        index=False)
```

```
[ ]:
```

## 0.9 Split file to multiple Excel sheets

```
[54]: column_name = 'salary_currency'

unique_values = df[column_name].unique()
```

```
writer = pd.ExcelWriter(r'C:\Users\nichola.ondiek\Python\New folder\output.
↳xlsx')

for unique_value in unique_values:
    frame = df[df['salary_currency'] == unique_value]
    frame.to_excel(writer, sheet_name=unique_value)
writer.save()
```

[ ]:

## 0.10 Missing values by percentage

```
[55]: df.isna().mean().round(4) * 100
```

```
[55]: Unnamed: 0          0.0
work_year             0.0
experience_level       0.0
employment_type        0.0
job_title              0.0
salary                0.0
salary_currency        0.0
salary_in_usd          0.0
employee_residence     0.0
remote_ratio           0.0
company_location       0.0
company_size           0.0
dtype: float64
```

[ ]:

## 0.11 Create folders from filename automatically

```
[56]: path = r'C:\Users\nichola.ondiek\Python\New folder'

for file_path in glob.glob(os.path.join(path, '*.csv')):
    new_dir = file_path.rsplit('.', 1)[0]
    try:
        os.mkdir(os.path.join(path, new_dir))
    except WindowsError:
        # Handle the case where the target dir already exist.
        pass
    shutil.move(file_path, os.path.join(new_dir, os.path.basename(file_path)))
```

[ ]:

## 0.12 Pivoting

```
[62]: df = pd.read_csv(r'C:\Users\nichola.ondiek\Desktop\Data science\IPPD_
↳Audit\Nichola\Payroll\30501.csv', low_memory=False)
df['Full-Name'] = df['FIRSTNAME'].map(str) + ' ' + df['OTHERNAMES'].map(str) +
↳' ' + df['SURNAME'].map(str)
pivot = pd.pivot_table(df, columns='EDCODE', index=['Full-Name', 'PAYROLLNUM',
↳'IDNUM'], values=['AMOUNT'], aggfunc='sum')
pivot
```

```
[62]:
```

			AMOUNT			
EDCODE			801	802	803	\
Full-Name	PAYROLLNUM	IDNUM				
Aarif Abdulaziz Ibrahim	20170105692	11649215	296690.00	38400.0	NaN	
Abalhassan Mohamed Hamza	20160099276	30512613	219240.00	32400.0	NaN	
Abarufa Dido Abarufa	20140028744	27872042	778710.00	184800.0	103200.0	
Abbas Salim Abdalla	1989080516	5355642	661720.00	156000.0	NaN	
Abdalla Ahmed Ali	1982121379	159770	NaN	NaN	NaN	
...			...	...	...	
Zubeir Salim Zubeir	20160137413	24847613	597420.00	156000.0	NaN	
Zuhura Nyangweso Athman	20170021575	30605984	49620.00	6400.0	NaN	
Zulekha Hamisi Mohamed	20190028561	25124521	39697.35	6840.0	NaN	
Zulfa Mohamed Yusuf	20190028678	32840708	42788.00	6840.0	NaN	
Zulfa Musa Hassan	20140028655	22807060	353780.00	50400.0	22600.0	

  

			805	806	807	808	812	
EDCODE								\
Full-Name	PAYROLLNUM	IDNUM						
Aarif Abdulaziz Ibrahim	20170105692	11649215	79200.00	NaN	NaN	NaN	NaN	
Abalhassan Mohamed Hamza	20160099276	30512613	51600.00	NaN	NaN	NaN	NaN	
Abarufa Dido Abarufa	20140028744	27872042	205200.00	NaN	NaN	NaN	NaN	
Abbas Salim Abdalla	1989080516	5355642	175800.00	NaN	NaN	NaN	NaN	
Abdalla Ahmed Ali	1982121379	159770	540000.00	NaN	NaN	NaN	NaN	
...			...	..	..	..	..	
Zubeir Salim Zubeir	20160137413	24847613	175800.00	NaN	NaN	NaN	NaN	
Zuhura Nyangweso Athman	20170021575	30605984	13200.00	NaN	NaN	NaN	NaN	
Zulekha Hamisi Mohamed	20190028561	25124521	9373.35	NaN	NaN	NaN	NaN	
Zulfa Mohamed Yusuf	20190028678	32840708	10893.35	NaN	NaN	NaN	NaN	
Zulfa Musa Hassan	20140028655	22807060	100000.00	NaN	NaN	NaN	NaN	

  

			813	817	819	
EDCODE						\
Full-Name	PAYROLLNUM	IDNUM				
Aarif Abdulaziz Ibrahim	20170105692	11649215	48000.0	27000.0	NaN	
Abalhassan Mohamed Hamza	20160099276	30512613	36000.0	NaN	NaN	
Abarufa Dido Abarufa	20140028744	27872042	96000.0	NaN	NaN	
Abbas Salim Abdalla	1989080516	5355642	96000.0	NaN	24000.0	

Abdalla Ahmed Ali	1982121379	159770	240000.0	NaN	NaN
...			...	...	...
Zubeir Salim Zubeir	20160137413	24847613	96000.0	NaN	240000.0
Zuhura Nyangweso Athman	20170021575	30605984	8000.0	NaN	7700.0
Zulekha Hamisi Mohamed	20190028561	25124521	7600.0	NaN	NaN
Zulfa Mohamed Yusuf	20190028678	32840708	7600.0	NaN	NaN
Zulfa Musa Hassan	20140028655	22807060	48000.0	NaN	NaN

EDCODE			820	831	833	835
Full-Name	PAYROLLNUM	IDNUM				
Aarif Abdulaziz Ibrahim	20170105692	11649215	NaN	NaN	NaN	NaN
Abalhassan Mohamed Hamza	20160099276	30512613	NaN	NaN	NaN	NaN
Abarufa Dido Abarufa	20140028744	27872042	NaN	NaN	NaN	119720.0
Abbas Salim Abdalla	1989080516	5355642	NaN	NaN	NaN	NaN
Abdalla Ahmed Ali	1982121379	159770	NaN	NaN	NaN	NaN
...			...	...	...	...
Zubeir Salim Zubeir	20160137413	24847613	NaN	NaN	228000.0	NaN
Zuhura Nyangweso Athman	20170021575	30605984	NaN	NaN	NaN	NaN
Zulekha Hamisi Mohamed	20190028561	25124521	NaN	NaN	NaN	NaN
Zulfa Mohamed Yusuf	20190028678	32840708	NaN	NaN	NaN	NaN
Zulfa Musa Hassan	20140028655	22807060	NaN	NaN	NaN	NaN

EDCODE			845	847	868
Full-Name	PAYROLLNUM	IDNUM			
Aarif Abdulaziz Ibrahim	20170105692	11649215	8000.0	NaN	NaN
Abalhassan Mohamed Hamza	20160099276	30512613	8000.0	NaN	NaN
Abarufa Dido Abarufa	20140028744	27872042	12000.0	NaN	NaN
Abbas Salim Abdalla	1989080516	5355642	12000.0	NaN	NaN
Abdalla Ahmed Ali	1982121379	159770	20000.0	48000.0	NaN
...			...	...	...
Zubeir Salim Zubeir	20160137413	24847613	12000.0	NaN	864000.0
Zuhura Nyangweso Athman	20170021575	30605984	NaN	NaN	NaN
Zulekha Hamisi Mohamed	20190028561	25124521	NaN	NaN	NaN
Zulfa Mohamed Yusuf	20190028678	32840708	NaN	NaN	NaN
Zulfa Musa Hassan	20140028655	22807060	8000.0	NaN	NaN

EDCODE			869	872	876
Full-Name	PAYROLLNUM	IDNUM			
Aarif Abdulaziz Ibrahim	20170105692	11649215	NaN	NaN	NaN
Abalhassan Mohamed Hamza	20160099276	30512613	NaN	NaN	NaN
Abarufa Dido Abarufa	20140028744	27872042	NaN	NaN	NaN
Abbas Salim Abdalla	1989080516	5355642	360000.0	NaN	NaN
Abdalla Ahmed Ali	1982121379	159770	NaN	2880000.0	NaN
...			...	...	...

Zubeir Salim Zubeir	20160137413	24847613	480000.0	NaN	NaN
Zuhura Nyangweso Athman	20170021575	30605984	50000.0	NaN	40000.0
Zulekha Hamisi Mohamed	20190028561	25124521	NaN	NaN	NaN
Zulfa Mohamed Yusuf	20190028678	32840708	NaN	NaN	NaN
Zulfa Musa Hassan	20140028655	22807060	NaN	NaN	NaN

EDCODE 877 901 902 \

Full-Name	PAYROLLNUM	IDNUM			
Aarif Abdulaziz Ibrahim	20170105692	11649215	NaN	NaN	NaN
Abalhassan Mohamed Hamza	20160099276	30512613	NaN	NaN	NaN
Abarufa Dido Abarufa	20140028744	27872042	NaN	NaN	125605.2
Abbas Salim Abdalla	1989080516	5355642	240000.0	NaN	NaN
Abdalla Ahmed Ali	1982121379	159770	NaN	NaN	NaN

...	...	..	...		
Zubeir Salim Zubeir	20160137413	24847613	NaN	NaN	88010.4
Zuhura Nyangweso Athman	20170021575	30605984	NaN	NaN	6322.4
Zulekha Hamisi Mohamed	20190028561	25124521	NaN	NaN	NaN
Zulfa Mohamed Yusuf	20190028678	32840708	NaN	NaN	NaN
Zulfa Musa Hassan	20140028655	22807060	NaN	NaN	47853.6

EDCODE 903 904 907 912 913 \

Full-Name	PAYROLLNUM	IDNUM					
Aarif Abdulaziz Ibrahim	20170105692	11649215	NaN	NaN	199971.0	NaN	NaN
Abalhassan Mohamed Hamza	20160099276	30512613	NaN	NaN	13594.0	NaN	NaN
Abarufa Dido Abarufa	20140028744	27872042	NaN	NaN	406128.0	NaN	NaN
Abbas Salim Abdalla	1989080516	5355642	13234.4	NaN	455904.0	NaN	NaN
Abdalla Ahmed Ali	1982121379	159770	NaN	NaN	355267.0	NaN	NaN

...	...	..	...	..	..		
Zubeir Salim Zubeir	20160137413	24847613	NaN	NaN	917744.0	NaN	NaN
Zuhura Nyangweso Athman	20170021575	30605984	NaN	NaN	NaN	NaN	NaN
Zulekha Hamisi Mohamed	20190028561	25124521	NaN	NaN	NaN	NaN	NaN
Zulfa Mohamed Yusuf	20190028678	32840708	NaN	NaN	NaN	NaN	NaN
Zulfa Musa Hassan	20140028655	22807060	NaN	NaN	NaN	NaN	NaN

EDCODE 917 921 923 928 929 \

Full-Name	PAYROLLNUM	IDNUM					
Aarif Abdulaziz Ibrahim	20170105692	11649215	NaN	NaN	NaN	NaN	NaN
Abalhassan Mohamed Hamza	20160099276	30512613	NaN	NaN	NaN	NaN	NaN
Abarufa Dido Abarufa	20140028744	27872042	NaN	NaN	384900.0	NaN	NaN
Abbas Salim Abdalla	1989080516	5355642	NaN	19525.8	NaN	NaN	NaN
Abdalla Ahmed Ali	1982121379	159770	NaN	NaN	NaN	NaN	NaN

...	..	...	...	..	..		
Zubeir Salim Zubeir	20160137413	24847613	NaN	NaN	NaN	NaN	NaN
Zuhura Nyangweso Athman	20170021575	30605984	NaN	NaN	NaN	NaN	NaN

Zulekha Hamisi Mohamed	20190028561	25124521	NaN	NaN	NaN	NaN	NaN
Zulfa Mohamed Yusuf	20190028678	32840708	NaN	NaN	NaN	NaN	NaN
Zulfa Musa Hassan	20140028655	22807060	NaN	NaN	NaN	NaN	NaN

EDCODE			933	935	936
Full-Name	PAYROLLNUM	IDNUM			
Aarif Abdulaziz Ibrahim	20170105692	11649215	NaN	64436.05	11850.0
Abalhassan Mohamed Hamza	20160099276	30512613	NaN	30842.20	10300.0
Abarufa Dido Abarufa	20140028744	27872042	2400.0	323363.25	20400.0
Abbas Salim Abdalla	1989080516	5355642	NaN	0.00	20400.0
Abdalla Ahmed Ali	1982121379	159770	NaN	1030275.60	20400.0
...			...	...	...
Zubeir Salim Zubeir	20160137413	24847613	2400.0	739518.55	20400.0
Zuhura Nyangweso Athman	20170021575	30605984	400.0	35771.90	3000.0
Zulekha Hamisi Mohamed	20190028561	25124521	NaN	3576.40	2200.0
Zulfa Mohamed Yusuf	20190028678	32840708	NaN	4450.00	2200.0
Zulfa Musa Hassan	20140028655	22807060	2400.0	73453.40	13600.0

EDCODE			937	938	939	940
Full-Name	PAYROLLNUM	IDNUM				
Aarif Abdulaziz Ibrahim	20170105692	11649215	NaN	NaN	NaN	NaN
Abalhassan Mohamed Hamza	20160099276	30512613	NaN	NaN	NaN	NaN
Abarufa Dido Abarufa	20140028744	27872042	NaN	NaN	NaN	NaN
Abbas Salim Abdalla	1989080516	5355642	12249.55	50004.0	60000.0	NaN
Abdalla Ahmed Ali	1982121379	159770	NaN	NaN	NaN	NaN
...			...	...	...	..
Zubeir Salim Zubeir	20160137413	24847613	NaN	NaN	NaN	NaN
Zuhura Nyangweso Athman	20170021575	30605984	NaN	NaN	NaN	NaN
Zulekha Hamisi Mohamed	20190028561	25124521	NaN	NaN	NaN	NaN
Zulfa Mohamed Yusuf	20190028678	32840708	NaN	NaN	NaN	NaN
Zulfa Musa Hassan	20140028655	22807060	NaN	NaN	NaN	NaN

EDCODE			942	945	946	949	956	957	990
Full-Name	PAYROLLNUM	IDNUM							
Aarif Abdulaziz Ibrahim	20170105692	11649215	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Abalhassan Mohamed Hamza	20160099276	30512613	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Abarufa Dido Abarufa	20140028744	27872042	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Abbas Salim Abdalla	1989080516	5355642	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Abdalla Ahmed Ali	1982121379	159770	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...			..	..	..	..	..	..	..
Zubeir Salim Zubeir	20160137413	24847613	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Zuhura Nyangweso Athman	20170021575	30605984	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Zulekha Hamisi Mohamed	20190028561	25124521	NaN	NaN	NaN	NaN	NaN	NaN	NaN
Zulfa Mohamed Yusuf	20190028678	32840708	NaN	NaN	NaN	NaN	NaN	NaN	NaN

```
Zulfa Musa Hassan      20140028655 22807060 NaN NaN NaN NaN NaN NaN NaN
```

```
[1182 rows x 48 columns]
```

```
[ ]:
```