## import pandas as pd

## 1. read\_csv

a. na\_values - additional strings to recognize as NaN. The Adult dataset uses "?" for unknown values.

b. skipinitialspace - if True, leading spaces after delimiters are removed (so " United-States" becomes "United-States").

```
df = pd.read csv('adult dataset.csv',
                na values='?',
                                   # treat '?' as missing data (NaN),
                index col=False, # don't pull out any column to be
the index
                skipinitialspace=True)
df
               workclass
                          fnlwgt
                                      education
                                                 educational-num
       age
                 Private
                           226802
        25
                                           11th
1
        38
                 Private
                                        HS-grad
                                                                9
                           89814
2
        28
               Local-gov
                          336951
                                     Assoc-acdm
                                                               12
3
        44
                 Private
                          160323
                                   Some-college
                                                               10
4
        18
                     NaN
                          103497
                                   Some-college
                                                               10
                                     Assoc-acdm
48837
        27
                 Private
                          257302
                                                               12
                           154374
                                        HS-grad
                                                                9
48838
        40
                 Private
48839
        58
                 Private
                           151910
                                        HS-grad
                                                                9
                                                                9
48840
        22
                 Private
                          201490
                                        HS-grad
        52
                                                                9
48841
            Self-emp-inc
                          287927
                                        HS-grad
           marital-status
                                   occupation relationship
                                                              race
gender
            Never-married
                           Machine-op-inspct
                                                 Own-child
                                                             Black
Male
       Married-civ-spouse
                              Farming-fishing
                                                   Husband
                                                             White
1
Male
                              Protective-serv
       Married-civ-spouse
                                                   Husband
                                                             White
Male
3
                                                   Husband Black
       Married-civ-spouse
                            Machine-op-inspct
Male
            Never-married
                                                 Own-child
                                                            White
                                          NaN
Female
. . .
       Married-civ-spouse
                                 Tech-support
                                                       Wife White
48837
Female
48838 Married-civ-spouse
                            Machine-op-inspct
                                                   Husband White
Male
                  Widowed
48839
                                 Adm-clerical
                                                 Unmarried White
Female
                                 Adm-clerical
                                                 Own-child White
48840
            Never-married
```

Male 48841 N Female	Married-civ-s	pouse Exec-	managerial	Wife W	hite
	capital-gain	capital-loss	hours-per-week	native-cou	ntry
income 0	0	Θ	40	United-Sta	ates
<=50K			-		
1 <=50K	0	0	50	United-St	ates
<=30K 2	0	0	40	United-Sta	ates
>50K					
3 >50K	7688	0	40	United-Sta	ates
4	0	0	30	United-St	ates
<=50K					
48837	0	0	38	United-St	ates
<=50K	0	0	40	Umitad Ct.	-+
48838 >50K	0	Θ	40	United-St	ates
48839	Θ	0	40	United-St	ates
<=50K 48840	0	0	20	United-St	2+06
40040 <=50K	в	8	20	United-31	aces
48841	15024	0	40	United-St	ates
>50K					
[48842 r	rows x 15 col	umns]			
df.isnu	ll(). <mark>sum</mark> ()				
age		0			
workclas	ss 2	799			
<pre>fnlwgt education</pre>	nn.	0 0			
	onal-num	0			
marital-		0			
occupati relation		809			
race	131111	0			
gender		0			
capital- capital-		0 0			
hours-pe	er-week	0			
native-d	country	857 0			
income					

```
#Dropped any rows with missing values
df.dropna(axis=0,inplace=True)
df
               workclass
                          fnlwgt
                                     education
                                                educational-num \
       age
0
        25
                          226802
                 Private
                                          11th
1
        38
                 Private
                          89814
                                       HS-grad
                                                              9
2
        28
                          336951
                                    Assoc-acdm
                                                             12
               Local-gov
3
        44
                 Private
                         160323
                                  Some-college
                                                             10
5
        34
                 Private
                          198693
                                          10th
                                                              6
48837
                          257302
                                                             12
        27
                 Private
                                    Assoc-acdm
                                       HS-grad
                                                              9
        40
                 Private
                          154374
48838
48839
        58
                                       HS-grad
                                                              9
                 Private
                          151910
                                                              9
48840
        22
                 Private
                          201490
                                       HS-grad
        52
48841
           Self-emp-inc
                          287927
                                       HS-grad
           marital-status
                                  occupation relationship
                                                              race
gender
0
            Never-married Machine-op-inspct
                                                  Own-child
                                                             Black
Male
1
       Married-civ-spouse
                             Farming-fishing
                                                    Husband
                                                             White
Male
                             Protective-serv
                                                    Husband White
2
       Married-civ-spouse
Male
                                                    Husband Black
       Married-civ-spouse Machine-op-inspct
Male
            Never-married
                               Other-service Not-in-family White
Male
. . .
48837
      Married-civ-spouse
                                Tech-support
                                                       Wife White
Female
48838 Married-civ-spouse Machine-op-inspct
                                                    Husband White
Male
48839
                  Widowed
                                Adm-clerical
                                                  Unmarried White
Female
48840
            Never-married
                                Adm-clerical
                                                  Own-child White
Male
48841 Married-civ-spouse Exec-managerial
                                                       Wife White
Female
       capital-gain capital-loss hours-per-week native-country
income
                                                   United-States
                                               40
<=50K
                                               50
                                                   United-States
1
<=50K
                                               40 United-States
2
>50K
```

3		7688	0		40	United-St	ates	
>50K 5		0	0		30	United-St	atos	
<=50K		U	U		30	Ollit Ced-20	aces	
48837 <=50K		0	0		38	United-St	ates	
48838 >50K		0	0		40	United-St	ates	
48839 <=50K		0	0		40	United-St	ates	
48840		0	0		20	United-St	ates	
<=50K 48841 >50K		15024	0		40	United-St	ates	
[45222	rows	x 15 columns]						
df.dro df	p_dup	licates()						
0 1 2 3 5	age 25 38 28 44 34	workclass Private Private Local-gov Private Private	fnlwgt 226802 89814 336951 160323 198693	education 11th HS-grad Assoc-acdm Some-college 10th		ucational-	num \ 7 9 12 10 6	
48837 48838 48839 48840 48841	27 40 58 22 52	Private Private Private Private Private Self-emp-inc	257302 154374 151910 201490 287927	Assoc-acdm HS-grad HS-grad HS-grad HS-grad			12 9 9 9	
		marital-status		occupation	rela	ationship	race	
gender 0 Male	\	Never-married	Machin	e-op-inspct	(	Own-child	Black	
1 Male	Marr	ied-civ-spouse	Farm	ing-fishing		Husband	White	
2 Male	Marr	ied-civ-spouse	Prot	ective-serv		Husband	White	
3 Male	Marr	ied-civ-spouse	Machin	e-op-inspct		Husband	Black	
5 Male		Never-married	0t	her-service	Not-:	in-family	White	
48837	Marr	ied-civ-spouse	Т	ech-support		Wife	White	

Female					
48838	Married-civ-s	pouse Machine	e-op-inspct	Husband	White
Male					
48839	W10	dowed Ad	lm-clerical	Unmarried	White
Female 48840	Never-ma	rried Ad	lm-clerical	Own-child	White
Male	Never-IIIa	iileu Au	IIII-CTELICAT	OWII-CIII CU	wiitte
48841	Married-civ-s	nouse Exec-	managerial	Wife	White
Female		pouse 2,100	mariager za c		200
	capital-gain	capital-loss	hours-per-week	native-cour	ntry
income	_	_			
0	0	0	40	United-Sta	ates
<=50K 1	0	0	50	United-Sta	2+00
=50K	U	U	30	ourrea-2rg	ates
2	0	0	40	United-Sta	ates
>50K	O .	O .	40	Onition St	1003
3	7688	0	40	United-Sta	ates
>50K					
5	0	Θ	30	United-Sta	ates
<=50K					
48837	0	0	38	United-Sta	2+05
40037 <=50K	в	в	30	Ollitted-219	ates
48838	Θ	Θ	40	United-Sta	ates
>50K	0	0	40	Onition St	1005
48839	0	0	40	United-Sta	ates
<=50K					
48840	0	0	20	United-Sta	ates
<=50K					
48841	15024	0	40	United-Sta	ates
>50K					
[45222	rows x 15 colu	umns 1			
[47222	IOM2 V ID COLL	uiii13 j			

A dictionary comprehension that builds a dictionary where: each key is a unique country name from the 'native-country' column each value is a subset DataFrame containing just rows for that country sub - a DataFrame (just the rows where 'native-country' == country') country: a string {country: sub for country, sub in ...} You're looping over each (country, sub\_df) pair from the groupby You're storing them as key-value pairs: {country: sub\_df}.

Output - 'United-States': DataFrame of US records, 'India': DataFrame of Indian records, 'Mexico': DataFrame of Mexican records,

```
\begin{array}{lll} country\_groups = \{country: \ sub \ for \ country, \ sub \ in \ df.groupby(\ 'native-country')\} \end{array}
```

```
for country, group in list(country groups.items())[:5]:
   print(f"Country: {country}")
   print(group.head(1)) # or group.head() for more rows
   print("-" * 30)
Country: Cambodia
age workclass fnlwgt education educational-num \
417 37 State-gov 67083 Some-college 10
marital-status occupation relationship
417 Married-civ-spouse Prof-specialty Other-relative Asian-Pac-
Islander
   gender capital-gain capital-loss hours-per-week native-country
income
417 Male 0 0 40 Cambodia
<=50K
          _____
Country: Canada
   age workclass fnlwgt education educational-num
marital-status \
868 35 Federal-gov 207973 Doctorate 16 Married-
civ-spouse
    occupation relationship race gender capital-gain capital-
loss \
868 Prof-specialty Husband White Male 0
    hours-per-week native-country income
868 55 Canada <=50K
Country: China
    age workclass fnlwgt education educational-num marital-
status \
945 38 Private 216319 Doctorate 16 Married-civ-
spouse
occupation relationship race gender capital-
945 Prof-specialty Husband Asian-Pac-Islander Male
capital-loss hours-per-week native-country income
945 0
945
               40 China >50K
Country: Columbia
age workclass fnlwgt education educational-num marital-
status \
```

```
385 47 Private 200808 11th
                                            7 Married-civ-
spouse
      occupation relationship race gender capital-gain capital-
loss \
385 Other-service Husband White Male
    hours-per-week native-country income
385
   35 Columbia <=50K
Country: Cuba
    age workclass fnlwgt education educational-num marital-
status \
586 36 Private 300333 11th 7 Married-civ-
spouse
        occupation relationship race gender capital-gain
capital-loss \
586 Exec-managerial Husband White Male
    hours-per-week native-country income
586
             40 Cuba <=50K
gender groups = {gender:sub for gender, sub in df.groupby('gender')}
for gender, group in list(gender groups.items())[:5]:
   print(f"gender: {gender}")
   print(group.head(1)) # or group.head() for more rows
   print("-" * 30)
gender: Female
  age workclass fnlwgt education educational-num marital-status
8 24 Private 369667 Some-college
                                             10 Never-married
     occupation relationship race gender capital-gain capital-
loss \
8 Other-service Unmarried White Female
  hours-per-week native-country income
8
   40 United-States <=50K
gender: Male
  age workclass fnlwgt education educational-num marital-status \
0 25 Private 226802 11th 7 Never-married
```

```
occupation relationship race gender capital-gain
capital-loss \
0 Machine-op-inspct Own-child Black Male
hours-per-week native-country income
0 40 United-States <=50K
race groups = {race: sub for race, sub in df.groupby('race')}
for race, group in list(race groups.items())[:5]:
   print(f"Race: {race}")
   print(group.head(1)) # or group.head() for more rows
   print("-" * 30)
Race: Amer-Indian-Eskimo
   age workclass fnlwgt education educational-num marital-status \
94 34 Private 198751 Masters 14 Never-married
    occupation relationship
                                race gender capital-
gain \
94 Other-service Not-in-family Amer-Indian-Eskimo Male
capital-loss hours-per-week native-country income
94 0 40 United-States <=50K
Race: Asian-Pac-Islander
   age workclass fnlwgt education educational-num marital-
status \
141 18 Private 262118 Some-college 10 Never-
married
    occupation relationship race gender capital-
gain \
141 Adm-clerical Own-child Asian-Pac-Islander Female
capital-loss hours-per-week native-country income
141 0 22 Germany <=50K
Race: Black
age workclass fnlwgt education educational-num marital-status \
0 25 Private 226802 11th 7 Never-married
  occupation relationship race gender capital-gain
capital-loss \
0 Machine-op-inspct Own-child Black Male 0
0
```

```
hours-per-week native-country income
0 40 United-States <=50K
Race: Other
  age workclass fnlwgt education educational-num marital-
status \
74 46 Private 269034 Some-college 10 Married-civ-
spouse
   occupation relationship race gender capital-gain capital-
74 Craft-repair Husband Other Male 0
hours-per-week native-country income
74 40 Dominican-Republic <=50K
             . . . . . . . . . . . . . . . . . .
Race: White
   age workclass fnlwgt education educational-num marital-
1 38 Private 89814 HS-grad 9 Married-civ-
spouse
        occupation relationship race gender capital-gain capital-
loss \
1 Farming-fishing Husband White Male
   hours-per-week native-country income
1 50 United-States <=50K</pre>
#Creating data subsets from the dictionary
india df = country groups['India']
usa df = country groups['United-States']
female df = gender groups['Female']
male df = gender groups['Male']
india df.head()
     age workclass fnlwgt education educational-num \
36 Private 116358 Masters 14
398

      564
      26
      Private
      341672
      Some-college

      1101
      55
      Private
      176219
      Bachelors

      1556
      33
      State-gov
      150688
      Doctorate

                                                              10
                                                             13
                                                             16
1619 24 Private 109414 Some-college
                                                    10
marital-status occupation relationship \
398 Married-civ-spouse Exec-managerial Husband
564 Never-married Adm-clerical Other-relative
```

```
1101
     Married-civ-spouse
                          Machine-op-inspct
                                                     Husband
1556
      Married-civ-spouse
                             Prof-specialty
                                                     Husband
1619
           Never-married
                                       Sales
                                             Other-relative
                    race gender
                                 capital-gain capital-loss
                                                              hours-
per-week \
398
      Asian-Pac-Islander
                           Male
                                                           0
45
564
      Asian-Pac-Islander
                           Male
                                                           0
60
1101 Asian-Pac-Islander
                           Male
                                                           0
40
1556 Asian-Pac-Islander
                           Male
                                                           0
50
1619 Asian-Pac-Islander
                           Male
                                                           0
20
     native-country income
398
              India
                     <=50K
              India
564
                     <=50K
1101
              India
                     <=50K
1556
              India
                      >50K
1619
              India
                     <=50K
```

## Another way to create subsets

This method is inefficient as we will have to type 40 line for 40 subsets

```
df us
           = df[df['native-country']=='United-States']
df india
           = df[df['native-country']=='India']
df male
           = df[df['gender']=='Male']
df female = df[df['gender']=='Female']
           = df[df['race']=='White']
df white
           = df[df['race']=='Black']
df black
df india.head()
      age workclass fnlwgt
                                 education educational-num
398
       36
             Private 116358
                                   Masters
                                                          14
       26
                                                          10
564
             Private 341672
                              Some-college
1101
       55
             Private
                     176219
                                 Bachelors
                                                          13
1556
       33
           State-gov
                      150688
                                 Doctorate
                                                          16
1619
       24
             Private 109414
                              Some-college
                                                          10
          marital-status
                                 occupation
                                                relationship
                            Exec-managerial
398
      Married-civ-spouse
                                                     Husband
564
                               Adm-clerical
                                             Other-relative
           Never-married
1101
      Married-civ-spouse
                          Machine-op-inspct
                                                     Husband
1556
      Married-civ-spouse
                             Prof-specialty
                                                     Husband
```

1619	Never-married		Sales	Other-relative	
200.14		gender	capital-gai	n capital-loss	hours-
per-we 398 45	eek \ Asian-Pac-Islander	Male	(	9 0	
564 60	Asian-Pac-Islander	Male		9 0	
1101 40	Asian-Pac-Islander	Male		0	
1556 50	Asian-Pac-Islander	Male		9 0	
1619 20	Asian-Pac-Islander	Male		9 0	
398 564 1101 1556 1619	native-country incom- India <=50 India <=50 India <=50 India >50 India <=50	Κ Κ Κ			
	tical concatenation d = pd.concat([india d				
0 1 2 3 4	age workclass 36 Private 26 Private 55 Private 33 State-gov 24 Private	fnlwgt 116358 341672 176219 150688 109414	education Maste Some-college Bachelo Doctorage Some-college	rs ge rs te	num \ 14 10 13 16 10
41434 41435 41436 41437	27 Private 40 Private 58 Private 22 Private	257302 154374 151910 201490	Assoc-aco HS-gra HS-gra HS-gra	ad ad	12 9 9
41438	52 Self-emp-inc	287927	HS-gr		9
0 1 2 3 4	marital-status Married-civ-spouse Never-married Married-civ-spouse Married-civ-spouse Never-married	<i>A</i> Machir	occupation c-managerial Adm-clerical ne-op-inspct of-specialty Sales	relationship Husband Other-relative Husband Husband Other-relative	
41434 41435 41436 41437	Married-civ-spouse Married-civ-spouse Widowed Never-married	Machir <i>H</i>	 Tech-support ne-op-inspct Adm-clerical Adm-clerical	Wife Wife Husband Unmarried Own-child	

41438	Married-civ-spouse	Exec-	managerial	Wife	
	race	gender	capital-gain	capital-loss	hours-
per-we	ek \ Asian-Pac-Islander	Male	0	0	
45 1	Asian-Pac-Islander	Male	0	0	
60	Asian-Pac-Islander	Male	0	0	
40	Asian-Pac-Islander	Male	0	0	
50 4	Asian-Pac-Islander	Male	9	0	
20					
41434 38	White	Female	0	0	
41435 40	White	Male	0	0	
41436 40	White	Female	0	0	
41437 20	White	Male	0	0	
41438 40	White	Female	15024	0	
0 1 2 3 4  41434 41435 41436 41437 41438	native-country incom India <=50 India <=50 India <=50 India >50 India <=50 India <=50 India <=50 India <=50 United-States <=50 United-States <=50 United-States <=50 United-States <=50 United-States >50 India <=50 India <	9K 9K 9K 9K 9K 9K 9K 9K			
sort =	df.sort_values(by=	hours-pe	r-week',axis= <mark>0</mark>	ascending= <mark>Fal</mark>	se)
20722 31741 2078 34884	age workcla 43 Self-emp-i 37 Priva 33 State-o 33 Self-emp-not-i	inc 2867 ate 2411 gov 1627	50 Prof-schoo 74 Bacheloo 05 Some-colle	ol rs ge	l-num \ 15 13 10 11

41004	2 5 1	102204	•		1.1
41994 3		te 183304	Assoc-voc		11
27100 7		 no 71676 Con			10
37190 7	•		ne-college 11th		10
9076 3					7
16470 5			Doctorate		16
41359 7			10th		6
27732 2	7 Priva	te 147951	HS-grad		9
	marital-status	occupat	ion ro	lationchin	raco
gender \	maritat-Status	occupat	.1011 16	lationship	race
	rried-civ-spouse	Prof-specia	1+1/	Husband	Black
Male	Titeu-Civ-spouse	riui-specia	істу	Husballu	DLack
	rried civ chauce	Drof specie	1+1	Husband	White
	rried-civ-spouse	Prof-specia	псту	пиѕрапи	wiiite
Male	Divonced	Othon com	, i	Unmarried	\d\b i + 0
2078	Divorced	Other-serv	rce	unillar i Leu	White
Female	Diversed	0+6		المسمسمة مما	م الله المالية
34884	Divorced	Other-ser	rice	Unmarried	White
Female		Tueseeset me		ام محمل میں ا	م الله المالية
	rried-civ-spouse	Transport-mov	ring	Husband	White
Male					
27100	Widowad	Adm alam	anl Nat	in family	\d\b i + 0
37190	Widowed	Adm-cleri	.cat Not	-in-family	White
Female	Divonced	Cormina fich	ina	Unmanniad	\d\b i + 0
9076	Divorced	Farming-fish	iriig	Unmarried	White
Male	nnied siv spause	Drof crosis	1+.,	Huchand	\d\b i + 0
	rried-civ-spouse	Prof-specia	itty	Husband	White
Male 41359	Divorced	Other-serv	rico Not	-in-family	White
Female	DIVOICEU	Other-Serv	TCE NOT	- III- I alli I Cy	MILLE
27732	Never-married	Machine-op-ins	nct Otho	r-relative	White
Male	Never-marrieu	Machine-op-ins	spec othe	1-letative	WIITLE
ria Ce					
ca	pital-gain capit	al-loss hours-	ner-week	native-coun	trv
income	pricat gain capit	at toss nours	per week	nacive coun	cı y
20722	0	0	99	United-Sta	tes
>50K	V	U	33	Onitica Sta	CCS
31741	0	0	99	United-Sta	tes
>50K	V	U	33	Onitica Sta	CCS
2078	0	0	99	United-Sta	tes
>50K	V	U	33	Onitica Sta	CCS
34884	0	0	99	United-Sta	tec
<=50K	U	U	33	Onitica Sta	CCS
41994	0	0	99	United-Sta	tes
>50K	· ·	•	33	JIII CO JCO	
> 50K					
37190	0	1944	1	United-Sta	tes
<=50K	•		_	J 200 5 20	

9076	0 0	1 Unit	ed-States
<=50K	0	1 11.24	and Challes
16470	0 0	1 Unit	ed-States
>50K 41359	0 0	1 Unit	ed-States
<=50K	0	1 01110	.eu-States
27732	0 0	1 Unit	ed-States
<=50K			
[45222 rows x 15	columns]		
transpose = df.T			
transpose			
	0	1	2
\	25	20	
age 28	25	38	
workclass	Private	Private	
Local-gov	ritvate	ritvace	
fnlwgt	226802	89814	
336951		3332.	
education	11th	HS-grad	Assoc-
acdm			
educational-num	7	9	
12	N		
marital-status	Never-married	Married-civ-spouse	Married-civ-
spouse occupation	Machine-op-inspct	Farming-fishing	Protective-
serv	Hachime-op-inspec	r ar ming-rishing	110fective-
relationship	Own-child	Husband	
Husband			
Husballa			
race	Black	White	
race White			
race White gender	Black Male	White Male	
race White gender Male	Male	Male	
race White gender Male capital-gain			
race White gender Male capital-gain 0	Male 0	Male 0	
race White gender Male capital-gain O capital-loss	Male	Male	
race White gender Male capital-gain 0 capital-loss	Male 0 0	Male 0	
race White gender Male capital-gain O capital-loss	Male 0	Male 0	
race White gender Male capital-gain 0 capital-loss 0 hours-per-week 40 native-country	Male 0 0	Male 0	United-
race White gender Male capital-gain 0 capital-loss 0 hours-per-week 40 native-country States	Male 0 40 United-States	Male 0 0 50 United-States	United-
race White gender Male capital-gain 0 capital-loss 0 hours-per-week 40 native-country States income	Male 0 0 40	Male 0 0 50	United-
race White gender Male capital-gain 0 capital-loss 0 hours-per-week 40 native-country States	Male 0 40 United-States	Male 0 0 50 United-States	United-
race White gender Male capital-gain 0 capital-loss 0 hours-per-week 40 native-country States income	Male 0 40 United-States <=50K	Male 0 0 50 United-States <=50K	
race White gender Male capital-gain 0 capital-loss 0 hours-per-week 40 native-country States income >50K	Male 0 40 United-States	Male 0 0 50 United-States	United- 7
race White gender Male capital-gain 0 capital-loss 0 hours-per-week 40 native-country States income	Male 0 40 United-States <=50K	Male 0 0 50 United-States <=50K	

workclass	Pri	vate	Private	Self-emp-not-inc
fnlwgt	16	0323	198693	104626
education	Some-col	lege	10th	Prof-school
educational-num		10	6	15
marital-status	Married-civ-sp	ouse	Never-married	Married-civ-spouse
occupation	Machine-op-in	spct	Other-service	Prof-specialty
relationship	Hus	band	Not-in-family	Husband
race	В	lack	White	White
gender		Male	Male	Male
capital-gain		7688	Θ	3103
capital-loss		0	0	0
hours-per-week		40	30	32
native-country	United-St	ates	United-States	United-States
income		>50K	<=50K	>50K
				10
\	8		9	10
age	24		55	65
workclass	Private		Private	Private
fnlwgt	369667		104996	184454
education	Some-college		7th-8th	HS-grad
educational-num	10		4	9
marital-status	Never-married	Marr	ied-civ-spouse	Married-civ-spouse
occupation	Other-service		Craft-repair	Machine-op-inspct
relationship	Unmarried		Husband	Husband
race	White		White	White
gender	Female		Male	Male

capital-gain	0	0	6418
capital-loss	0	0	0
hours-per-week	40	10	40
native-country	United-States	United-States Un	ited-States
	501/	F.01/	501/
income	<=50K	<=50K	>50K
age workclass fnlwgt education educational-num marital-status occupation relationship race gender capital-gain capital-loss hours-per-week native-country	11 36 Federal-gov 212465 Bachelors 13 Married-civ-spouse Adm-clerical Husband White Male 0 0 40 United-States	Married-civ-spous Handlers-cleaner Husban Amer-Indian-Eskim Mal	2 e 6 h 6 e s d o e 0 0
income	<=50K	<=50	K
	48833	48834	
48835 \	42	22	
age 53	43	32	
workclass	Private	Private	
Private fnlwgt	84661	116138	
321865	Acces ves	Masters	
education Masters	Assoc-voc	Masters	
educational-num 14	11	14	
marital-status spouse	Married-civ-spouse	Never-married Ma	rried-civ-
occupation	Sales	Tech-support	Exec-
managerial relationship Husband	Husband	Not-in-family	
race	White	Asian-Pac-Islander	
White gender Male	Male	Male	

capital-gain 0		0	0
capital-loss		0	0
0			
hours-per-week 40		45	11
native-country States	United-Stat	es T	aiwan United-
income >50K	<=5	0K	<=50K
	48836	488	37
48838 \			
age	22		27
40 workclass	Private	Priva	+0
Private	ritvate	FIIVa	te
fnlwgt	310152	2573	02
154374			
education	Some-college	Assoc-ac	dm HS-
grad	10		10
educational-num 9	10		12
marital-status	Never-married	Married-civ-spou	se Married-civ-
spouse	NCVCI - IIIdi i ica	riai i ica-civ-spou	3C Harrica-Civ-
occupation	Protective-serv	Tech-suppo	rt Machine-op-
inspct			
relationship	Not-in-family	Wi	fe
Husband	White	Whi	+ 0
race White	wiite	WIIT	te
gender	Male	Fema	le
Male			
capital-gain 0	0		0
capital-loss 0	0		0
hours-per-week 40	40		38
native-country States	United-States	United-Stat	es United-
income	<=50K	<=5	0K
>50K			
age workclass fnlwgt	48839 58 Private 151910	48840 22 Private 201490	48841 52 Self-emp-inc 287927
education	HS-grad	HS-grad	HS-grad

```
educational-num
                       Widowed
marital-status
                                Never-married
                                               Married-civ-spouse
occupation
                  Adm-clerical
                                 Adm-clerical
                                                  Exec-managerial
relationship
                     Unmarried
                                    Own-child
                                                              Wife
race
                         White
                                        White
                                                             White
                        Female
                                         Male
                                                            Female
gender
                                                             15024
capital-gain
                             0
                                            0
capital-loss
                             0
                                            0
hours-per-week
                            40
                                           20
                                                                40
native-country
                 United-States United-States
                                                    United-States
                         <=50K
                                        <=50K
                                                             >50K
income
[15 rows x 45222 columns]
melted = pd.melt(df,
id_vars=['age','gender','race','occupation'],value_vars=['capital-
gain','capital-loss'],var_name='capital_type' ,value_name='amount')
melted.head()
                             occupation
                                         capital type
   age gender
                race
                                                       amount
0
    25
         Male Black Machine-op-inspct
                                         capital-gain
1
    38
         Male White
                        Farming-fishing
                                                             0
                                         capital-gain
2
    28
         Male White
                        Protective-serv
                                                             0
                                         capital-gain
3
    44
         Male Black Machine-op-inspct
                                         capital-gain
                                                         7688
    34
         Male White
                          Other-service capital-gain
                                                             0
# Convert 'amount' to numeric, forcing errors to NaN
melted['amount'] = pd.to numeric(melted['amount'], errors='coerce')
# Now safely create the pivot table
casted = pd.pivot table(
    melted.
    index=['age', 'gender', 'race', 'occupation'],
    columns=['capital type'],
    values='amount',
    aggfunc='mean' # or sum, count, etc.
).reset index()
casted.head()
capital_type age gender
                                         race
                                                       occupation \
                   Female Amer-Indian-Eskimo
                                                    Adm-clerical
0
               17
1
               17
                   Female
                           Amer-Indian-Eskimo
                                                    Other-service
2
                   Female Asian-Pac-Islander
               17
                                                    Adm-clerical
3
               17
                   Female
                                        Black
                                                    Adm-clerical
4
               17
                  Female
                                        Black Handlers-cleaners
capital type capital-gain capital-loss
0
                       0.0
                                     0.0
1
                     527.5
                                     0.0
2
                       0.0
                                     0.0
```