

Homework-1-help

January 3, 2017

- Go to <http://www.github.com> and create an account.
- Send your Github username to dg2991@columbia.edu.
- Install Git - <https://github.com/blog/1510-installing-git-from-github-for-mac>. Make sure to install command line tools.
- When I have received your email, you should get a confirmation that you have been added to the repo.
- Click on this link: <https://classroom.github.com/assignment-invitations/11415026d0459793405d3c1ff95cc259>
- Follow the instructions to clone that repo to your local machine.
- You should type a command like:

```
$ git clone \url{https://github.com/Columbia-Intro-Data-Science/python-introduct
```

```
In [1]: import pandas as pd
df = pd.read_csv('https://archive.ics.uci.edu/ml/machine-learning-databases
df
```

```
Out[1]:
```

39	State-gov	77516	Bachelors	13	\
0	Self-emp-not-inc	83311	Bachelors	13	
1	Private	215646	HS-grad	9	
2	Private	234721	11th	7	
3	Private	338409	Bachelors	13	
4	Private	284582	Masters	14	
5	Private	160187	9th	5	
6	Self-emp-not-inc	209642	HS-grad	9	
7	Private	45781	Masters	14	
8	Private	159449	Bachelors	13	
9	Private	280464	Some-college	10	
10	State-gov	141297	Bachelors	13	
11	Private	122272	Bachelors	13	
12	Private	205019	Assoc-acdm	12	
13	Private	121772	Assoc-voc	11	
14	Private	245487	7th-8th	4	
15	Self-emp-not-inc	176756	HS-grad	9	
16	Private	186824	HS-grad	9	
17	Private	28887	11th	7	
18	Self-emp-not-inc	292175	Masters	14	

19	40	Private	193524	Doctorate	16
20	54	Private	302146	HS-grad	9
21	35	Federal-gov	76845	9th	5
22	43	Private	117037	11th	7
23	59	Private	109015	HS-grad	9
24	56	Local-gov	216851	Bachelors	13
25	19	Private	168294	HS-grad	9
26	54	?	180211	Some-college	10
27	39	Private	367260	HS-grad	9
28	49	Private	193366	HS-grad	9
29	23	Local-gov	190709	Assoc-acdm	12
...
32530	30	?	33811	Bachelors	13
32531	34	Private	204461	Doctorate	16
32532	54	Private	337992	Bachelors	13
32533	37	Private	179137	Some-college	10
32534	22	Private	325033	12th	8
32535	34	Private	160216	Bachelors	13
32536	30	Private	345898	HS-grad	9
32537	38	Private	139180	Bachelors	13
32538	71	?	287372	Doctorate	16
32539	45	State-gov	252208	HS-grad	9
32540	41	?	202822	HS-grad	9
32541	72	?	129912	HS-grad	9
32542	45	Local-gov	119199	Assoc-acdm	12
32543	31	Private	199655	Masters	14
32544	39	Local-gov	111499	Assoc-acdm	12
32545	37	Private	198216	Assoc-acdm	12
32546	43	Private	260761	HS-grad	9
32547	65	Self-emp-not-inc	99359	Prof-school	15
32548	43	State-gov	255835	Some-college	10
32549	43	Self-emp-not-inc	27242	Some-college	10
32550	32	Private	34066	10th	6
32551	43	Private	84661	Assoc-voc	11
32552	32	Private	116138	Masters	14
32553	53	Private	321865	Masters	14
32554	22	Private	310152	Some-college	10
32555	27	Private	257302	Assoc-acdm	12
32556	40	Private	154374	HS-grad	9
32557	58	Private	151910	HS-grad	9
32558	22	Private	201490	HS-grad	9
32559	52	Self-emp-inc	287927	HS-grad	9

	Never-married	Adm-clerical	Not-in-family	\
0	Married-civ-spouse	Exec-managerial	Husband	
1	Divorced	Handlers-cleaners	Not-in-family	
2	Married-civ-spouse	Handlers-cleaners	Husband	
3	Married-civ-spouse	Prof-specialty	Wife	

4	Married-civ-spouse	Exec-managerial	Wife
5	Married-spouse-absent	Other-service	Not-in-family
6	Married-civ-spouse	Exec-managerial	Husband
7	Never-married	Prof-specialty	Not-in-family
8	Married-civ-spouse	Exec-managerial	Husband
9	Married-civ-spouse	Exec-managerial	Husband
10	Married-civ-spouse	Prof-specialty	Husband
11	Never-married	Adm-clerical	Own-child
12	Never-married	Sales	Not-in-family
13	Married-civ-spouse	Craft-repair	Husband
14	Married-civ-spouse	Transport-moving	Husband
15	Never-married	Farming-fishing	Own-child
16	Never-married	Machine-op-inspct	Unmarried
17	Married-civ-spouse	Sales	Husband
18	Divorced	Exec-managerial	Unmarried
19	Married-civ-spouse	Prof-specialty	Husband
20	Separated	Other-service	Unmarried
21	Married-civ-spouse	Farming-fishing	Husband
22	Married-civ-spouse	Transport-moving	Husband
23	Divorced	Tech-support	Unmarried
24	Married-civ-spouse	Tech-support	Husband
25	Never-married	Craft-repair	Own-child
26	Married-civ-spouse	?	Husband
27	Divorced	Exec-managerial	Not-in-family
28	Married-civ-spouse	Craft-repair	Husband
29	Never-married	Protective-serv	Not-in-family
...
32530	Never-married	?	Not-in-family
32531	Married-civ-spouse	Prof-specialty	Husband
32532	Married-civ-spouse	Exec-managerial	Husband
32533	Divorced	Adm-clerical	Unmarried
32534	Never-married	Protective-serv	Own-child
32535	Never-married	Exec-managerial	Not-in-family
32536	Never-married	Craft-repair	Not-in-family
32537	Divorced	Prof-specialty	Unmarried
32538	Married-civ-spouse	?	Husband
32539	Separated	Adm-clerical	Own-child
32540	Separated	?	Not-in-family
32541	Married-civ-spouse	?	Husband
32542	Divorced	Prof-specialty	Unmarried
32543	Divorced	Other-service	Not-in-family
32544	Married-civ-spouse	Adm-clerical	Wife
32545	Divorced	Tech-support	Not-in-family
32546	Married-civ-spouse	Machine-op-inspct	Husband
32547	Never-married	Prof-specialty	Not-in-family
32548	Divorced	Adm-clerical	Other-relative
32549	Married-civ-spouse	Craft-repair	Husband
32550	Married-civ-spouse	Handlers-cleaners	Husband

32551	Married-civ-spouse	Sales	Husband
32552	Never-married	Tech-support	Not-in-family
32553	Married-civ-spouse	Exec-managerial	Husband
32554	Never-married	Protective-serv	Not-in-family
32555	Married-civ-spouse	Tech-support	Wife
32556	Married-civ-spouse	Machine-op-inspct	Husband
32557	Widowed	Adm-clerical	Unmarried
32558	Never-married	Adm-clerical	Own-child
32559	Married-civ-spouse	Exec-managerial	Wife

		White	Male	2174	0	40	United-States	<=5
0		White	Male	0	0	13	United-States	<=5
1		White	Male	0	0	40	United-States	<=5
2		Black	Male	0	0	40	United-States	<=5
3		Black	Female	0	0	40	Cuba	<=5
4		White	Female	0	0	40	United-States	<=5
5		Black	Female	0	0	16	Jamaica	<=5
6		White	Male	0	0	45	United-States	>5
7		White	Female	14084	0	50	United-States	>5
8		White	Male	5178	0	40	United-States	>5
9		Black	Male	0	0	80	United-States	>5
10	Asian-Pac-Islander		Male	0	0	40	India	>5
11		White	Female	0	0	30	United-States	<=5
12		Black	Male	0	0	50	United-States	<=5
13	Asian-Pac-Islander		Male	0	0	40	?	>5
14	Amer-Indian-Eskimo		Male	0	0	45	Mexico	<=5
15		White	Male	0	0	35	United-States	<=5
16		White	Male	0	0	40	United-States	<=5
17		White	Male	0	0	50	United-States	<=5
18		White	Female	0	0	45	United-States	>5
19		White	Male	0	0	60	United-States	>5
20		Black	Female	0	0	20	United-States	<=5
21		Black	Male	0	0	40	United-States	<=5
22		White	Male	0	2042	40	United-States	<=5
23		White	Female	0	0	40	United-States	<=5
24		White	Male	0	0	40	United-States	>5
25		White	Male	0	0	40	United-States	<=5
26	Asian-Pac-Islander		Male	0	0	60	South	>5
27		White	Male	0	0	80	United-States	<=5
28		White	Male	0	0	40	United-States	<=5
29		White	Male	0	0	52	United-States	<=5
...	
32530	Asian-Pac-Islander		Female	0	0	99	United-States	<=5
32531		White	Male	0	0	60	United-States	>5
32532	Asian-Pac-Islander		Male	0	0	50	Japan	>5
32533		White	Female	0	0	39	United-States	<=5
32534		Black	Male	0	0	35	United-States	<=5
32535		White	Female	0	0	55	United-States	>5

32536		Black	Male	0	0	46	United-States	<=5
32537		Black	Female	15020	0	45	United-States	>5
32538		White	Male	0	0	10	United-States	>5
32539		White	Female	0	0	40	United-States	<=5
32540		Black	Female	0	0	32	United-States	<=5
32541		White	Male	0	0	25	United-States	<=5
32542		White	Female	0	0	48	United-States	<=5
32543		Other	Female	0	0	30	United-States	<=5
32544		White	Female	0	0	20	United-States	>5
32545		White	Female	0	0	40	United-States	<=5
32546		White	Male	0	0	40	Mexico	<=5
32547		White	Male	1086	0	60	United-States	<=5
32548		White	Female	0	0	40	United-States	<=5
32549		White	Male	0	0	50	United-States	<=5
32550	Amer-Indian-Eskimo		Male	0	0	40	United-States	<=5
32551		White	Male	0	0	45	United-States	<=5
32552	Asian-Pac-Islander		Male	0	0	11	Taiwan	<=5
32553		White	Male	0	0	40	United-States	>5
32554		White	Male	0	0	40	United-States	<=5
32555		White	Female	0	0	38	United-States	<=5
32556		White	Male	0	0	40	United-States	>5
32557		White	Female	0	0	40	United-States	<=5
32558		White	Male	0	0	20	United-States	<=5
32559		White	Female	15024	0	40	United-States	>5

[32560 rows x 15 columns]

Problem: Show percentage of people based on gender, sex and country that make above \$50k.