CS Interview Ntbk

July 4, 2021

1 Rain Lambek CS first interview dataset analysis

[228]:
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
```

2 Part 1: Preliminary investigaion and cleaning

[229]:
```
data = pd.read_csv("CS_data.csv")
data.head()
```

[229]:

	skingoal_healthy	skingoal_glowing	skingoal_clear	skingoal_hydrated	\
0	0.0	1.0	1.0	0.0	
1	1.0	0.0	1.0	1.0	
2	0.0	1.0	1.0	0.0	
3	1.0	0.0	1.0	1.0	
4	1.0	0.0	1.0	1.0	

	skingoal_protected	skingoal_balanced	skingoal_repaired	skingoal_calm	\
0	0.0	0.0	0.0	0.0	
1	0.0	0.0	0.0	1.0	
2	0.0	0.0	0.0	0.0	
3	0.0	0.0	0.0	1.0	
4	0.0	0.0	0.0	1.0	

	skingoal_bright	skingoal_rejuvenated	…	age	pregnant	vegan	\
0	1.0	0.0	…	age1824	pregnantno	veganno	
1	0.0	0.0	…	age1824	pregnantno	veganyes	
2	1.0	0.0	…	age1824	pregnantno	veganno	
3	0.0	0.0	…	age1824	pregnantno	veganyes	
4	0.0	0.0	…	age1824	pregnantno	veganyes	

	currentproducts_cleanser	currentproducts_serum	currentproducts_eyecream	\
0	1	1	1	
1	1	1	0	
2	1	1	1	

```
3                          1                          1                          0
4                          1                          1                          0

   currentproducts_dailymoisturizer currentproducts_spf  \
0                                 1                    0
1                                 1                    1
2                                 1                    0
3                                 1                    1
4                                 1                    1

  currentproducts_nightcream                  getmatched
0                          1          getmatched_cleanser
1                          0        getmatched_nightcream
2                          1        getmatched_nightcream
3                          0  getmatched_dailymoisturizer
4                          0             getmatched_serum

[5 rows x 34 columns]
```

[230]:
```
data.mean()
```

[230]:
```
skingoal_healthy                  0.8029
skingoal_glowing                  0.6297
skingoal_clear                    0.7725
skingoal_hydrated                 0.6926
skingoal_protected                0.0000
skingoal_balanced                 0.0000
skingoal_repaired                 0.0000
skingoal_calm                     0.4829
skingoal_bright                   0.4804
skingoal_rejuvenated              0.0000
skingoal_restored                 0.0000
skinconcern_agingskin             0.5052
skinconcern_acne                  0.5632
skinconcern_dryskin               0.4139
skinconcern_firmnesselasticity    0.3516
skinconcern_pigmentation          0.4427
skinconcern_sensitivity           0.3632
currentproducts_cleanser          0.8901
currentproducts_serum             0.5645
currentproducts_eyecream          0.3895
currentproducts_dailymoisturizer  0.8324
currentproducts_spf               0.5236
currentproducts_nightcream        0.4272
dtype: float64
```

Looks like there are some columns that have no data. They must have been newly added or the

data went missing somehow. Just from this it looks like acne is the biggest concern, healthy and clear skin are the biggest goals, and people probably buy cleanser and daily moisturizer to deal with concerns + reach goals.

[231]:
```
cdata = data.dropna(how='all') #remove the nan cells if there are any
cdata.head()
```

[231]:

	skingoal_healthy	skingoal_glowing	skingoal_clear	skingoal_hydrated	\
0	0.0	1.0	1.0	0.0	
1	1.0	0.0	1.0	1.0	
2	0.0	1.0	1.0	0.0	
3	1.0	0.0	1.0	1.0	
4	1.0	0.0	1.0	1.0	

	skingoal_protected	skingoal_balanced	skingoal_repaired	skingoal_calm	\
0	0.0	0.0	0.0	0.0	
1	0.0	0.0	0.0	1.0	
2	0.0	0.0	0.0	0.0	
3	0.0	0.0	0.0	1.0	
4	0.0	0.0	0.0	1.0	

	skingoal_bright	skingoal_rejuvenated	…	age	pregnant	vegan	\
0	1.0	0.0	…	age1824	pregnantno	veganno	
1	0.0	0.0	…	age1824	pregnantno	veganyes	
2	1.0	0.0	…	age1824	pregnantno	veganno	
3	0.0	0.0	…	age1824	pregnantno	veganyes	
4	0.0	0.0	…	age1824	pregnantno	veganyes	

	currentproducts_cleanser	currentproducts_serum	currentproducts_eyecream	\
0	1	1	1	
1	1	1	0	
2	1	1	1	
3	1	1	0	
4	1	1	0	

	currentproducts_dailymoisturizer	currentproducts_spf	\
0	1	0	
1	1	1	
2	1	0	
3	1	1	
4	1	1	

	currentproducts_nightcream	getmatched
0	1	getmatched_cleanser
1	0	getmatched_nightcream
2	1	getmatched_nightcream
3	0	getmatched_dailymoisturizer

```
4                          0              getmatched_serum

[5 rows x 34 columns]
```

[232]:
```
print(cdata['age'].head())
print(type(cdata['age'][0]))
# need to change age column so I can use it as a numerical value if desired␣
 ↪later on
```

```
0    age1824
1    age1824
2    age1824
3    age1824
4    age1824
Name: age, dtype: object
<class 'str'>
```

[233]:
```
def get_age(age):
    if type(age) == str:
        a = age.split('e')[1]
        a = (int(a[0:2])+int(a[2:4]))/2
        return a
    else:
        pass
        #print(age)
cdata['age'] = cdata['age'].apply(lambda x: get_age(x))
cdata.head(10)
```

[233]:

	skingoal_healthy	skingoal_glowing	skingoal_clear	skingoal_hydrated	\
0	0.0	1.0	1.0	0.0	
1	1.0	0.0	1.0	1.0	
2	0.0	1.0	1.0	0.0	
3	1.0	0.0	1.0	1.0	
4	1.0	0.0	1.0	1.0	
5	0.0	1.0	1.0	0.0	
6	1.0	0.0	1.0	1.0	
7	1.0	0.0	1.0	0.0	
8	1.0	0.0	1.0	0.0	
9	1.0	0.0	1.0	0.0	

	skingoal_protected	skingoal_balanced	skingoal_repaired	skingoal_calm	\
0	0.0	0.0	0.0	0.0	
1	0.0	0.0	0.0	1.0	
2	0.0	0.0	0.0	0.0	
3	0.0	0.0	0.0	1.0	
4	0.0	0.0	0.0	1.0	
5	0.0	0.0	0.0	0.0	

```
6                 0.0                 0.0                 0.0             1.0
7                 0.0                 0.0                 0.0             1.0
8                 0.0                 0.0                 0.0             1.0
9                 0.0                 0.0                 0.0             1.0

   skingoal_bright  skingoal_rejuvenated  …   age    pregnant     vegan  \
0              1.0                   0.0  …  21.0  pregnantno   veganno
1              0.0                   0.0  …  21.0  pregnantno  veganyes
2              1.0                   0.0  …  21.0  pregnantno   veganno
3              0.0                   0.0  …  21.0  pregnantno  veganyes
4              0.0                   0.0  …  21.0  pregnantno  veganyes
5              1.0                   0.0  …  21.0  pregnantno   veganno
6              0.0                   0.0  …  21.0  pregnantno  veganyes
7              0.0                   0.0  …  21.0  pregnantno  veganyes
8              0.0                   0.0  …  21.0  pregnantno  veganyes
9              0.0                   0.0  …  21.0  pregnantno  veganyes

   currentproducts_cleanser  currentproducts_serum  currentproducts_eyecream  \
0                         1                      1                         1
1                         1                      1                         0
2                         1                      1                         1
3                         1                      1                         0
4                         1                      1                         0
5                         1                      1                         1
6                         1                      1                         0
7                         1                      1                         0
8                         1                      1                         0
9                         1                      1                         0

   currentproducts_dailymoisturizer currentproducts_spf  \
0                                 1                    0
1                                 1                    1
2                                 1                    0
3                                 1                    1
4                                 1                    1
5                                 1                    0
6                                 1                    1
7                                 0                    1
8                                 0                    1
9                                 0                    1

  currentproducts_nightcream                  getmatched
0                          1          getmatched_cleanser
1                          0        getmatched_nightcream
2                          1        getmatched_nightcream
3                          0  getmatched_dailymoisturizer
4                          0             getmatched_serum
```

```
5                          1  getmatched_dailymoisturizer
6                          0          getmatched_cleanser
7                          1          getmatched_eyecream
8                          1             getmatched_serum
9                          1          getmatched_cleanser

[10 rows x 34 columns]
```

3 Part 2: Age, gender, and ethnicity distributions

[99]:
```
cdata['age'].value_counts()
```

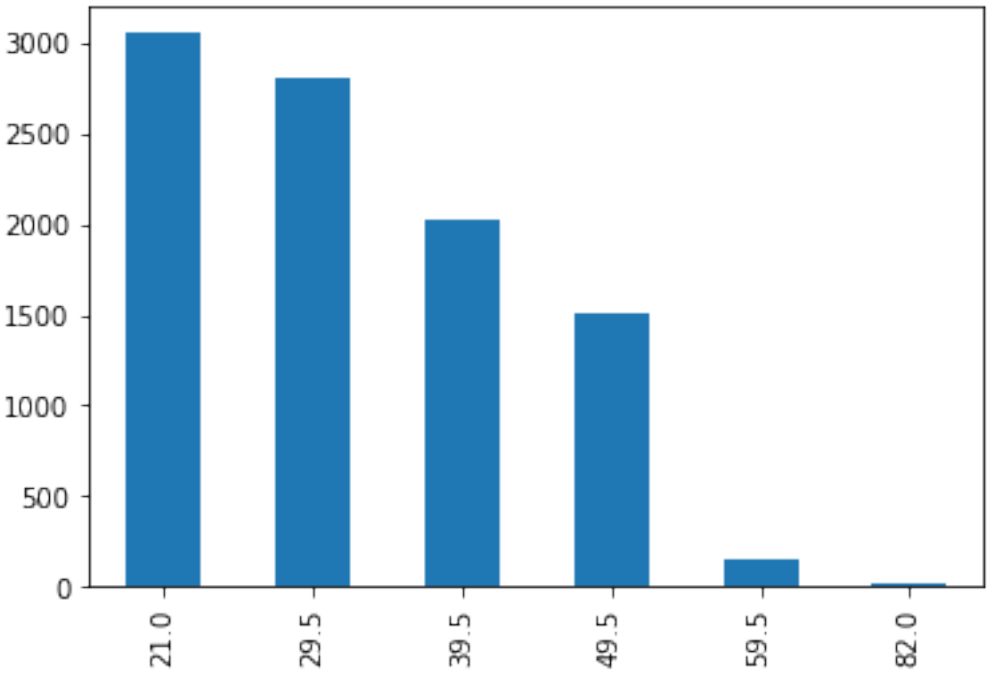
[99]:
```
21.0    3048
29.5    2798
39.5    2016
49.5    1501
59.5     149
82.0      18
Name: age, dtype: int64
```

[100]:
```
cdata['age'].mean()
```

[100]:
```
32.61511017838405
```

[101]:
```
cdata.age.value_counts().plot(kind = 'bar')
```

[101]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7e162efcd0>
```



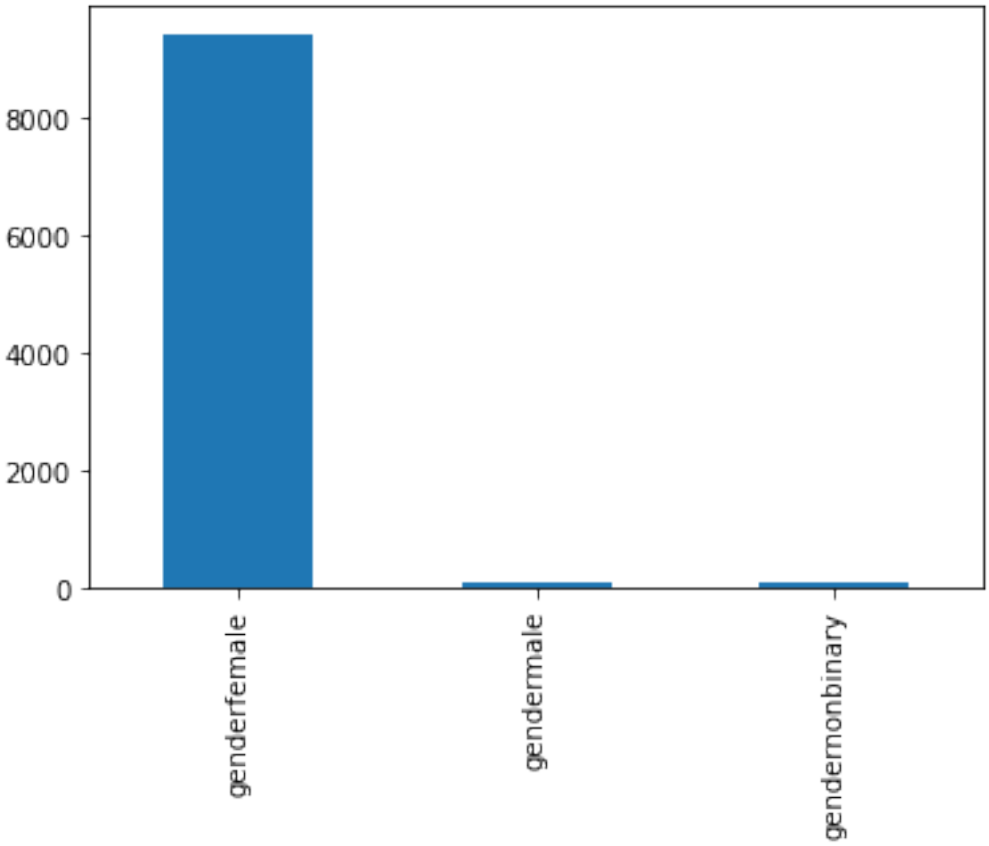
Looks kind of like an Exponential distribution. As age increases the counts drop fast. The majority of the customers are ages 18-44 with the mean being somewhere around 33 (can't get exact since age is submitted as a range)

[102]:
```
print(cdata['gender'].value_counts())
```

```
genderfemale       9418
gendermale          104
gendernonbinary      91
Name: gender, dtype: int64
```

[103]:
```
cdata.gender.value_counts().plot(kind = 'bar')
```

[103]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7e15dc1390>
```



Not surprising here. Almost all of the customers are women, and there are almost more non-binary customers than men despite the fact that non-binary people consist of around 0.2% of the population! The low amount of masculine customers is certainly a possible target for future customers.

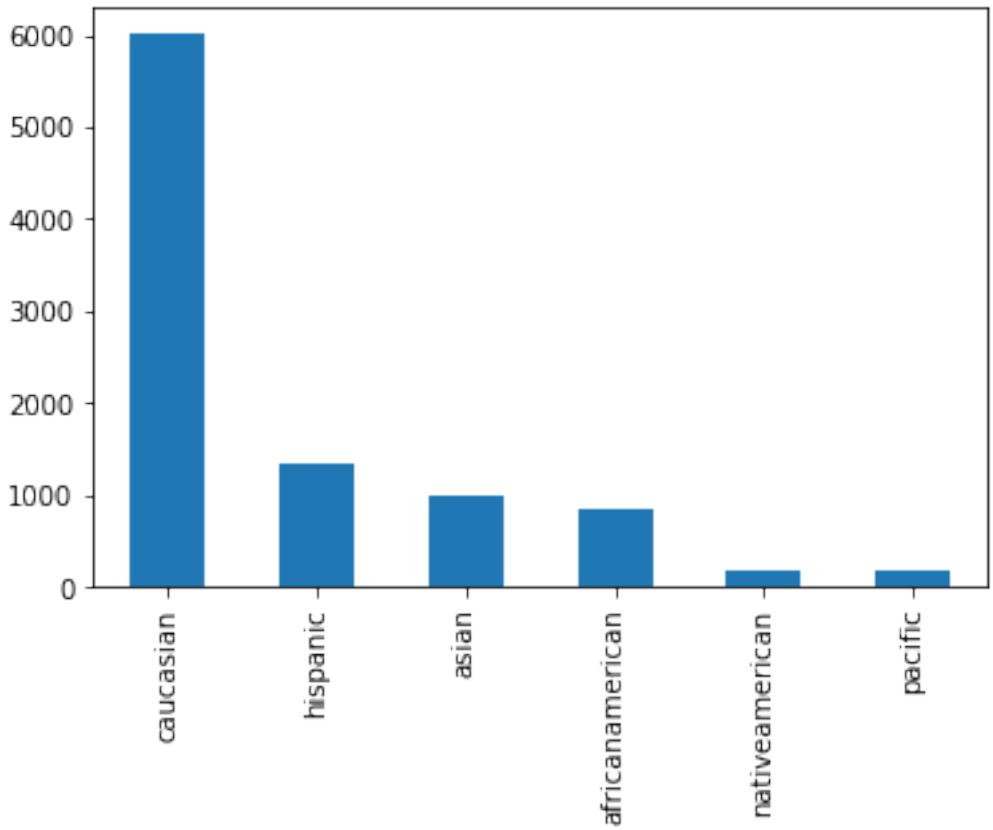
[234]:
```
def get_eth(eth):
    if type(eth) == str:
        a = eth.split('y')[1]
        return a
    else:
        pass
        #print(eth)
cdata['ethnicity'] = cdata['ethnicity'].apply(lambda x: get_eth(x))
cdata['ethnicity'].head(10)
```

[234]:
```
0          caucasian
1    africanamerican
2          caucasian
3    africanamerican
4    africanamerican
```

```
5          caucasian
6    africanamerican
7          caucasian
8          caucasian
9          caucasian
Name: ethnicity, dtype: object
```

[106]:
```
cdata.ethnicity.value_counts().plot(kind = 'bar')
```

[106]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7e1622ef50>
```



Mostly caucasin, as expected. Hispanic population is a little bit below the 18.5% US average, Asian is a good amount above the US average so that market might be saturated, and everything else is as expected

4 Part 3: Are skin goals correlated to skin concerns

[110]:
```
cdata["skingoal_rejuvenated"].corr(cdata["skingoal_rejuvenated"])
```

[110]:
```
nan
```

[117]:
```
# check if there is some bad data that could be causing the nan
from pandas.api.types import is_numeric_dtype
is_numeric_dtype(cdata["skingoal_rejuvenated"])
```

[117]: True

[123]:
```
# check variance to see if we are dividing by 0, and we sure are!
cdata["skingoal_rejuvenated"].var()
```

[123]: 0.0

[126]:
```
cdata.corr()
```

[126]:

	skingoal_healthy	skingoal_glowing	\
skingoal_healthy	1.000000	0.122407	
skingoal_glowing	0.122407	1.000000	
skingoal_clear	0.171254	0.062029	
skingoal_hydrated	0.152060	0.250750	
skingoal_protected	NaN	NaN	
skingoal_balanced	NaN	NaN	
skingoal_repaired	NaN	NaN	
skingoal_calm	0.163388	0.037786	
skingoal_bright	0.148360	0.353118	
skingoal_rejuvenated	NaN	NaN	
skingoal_restored	NaN	NaN	
skinconcern_agingskin	0.083336	0.114632	
skinconcern_acne	0.053757	-0.076186	
skinconcern_dryskin	-0.009290	0.073442	
skinconcern_firmnesselasticity	0.080553	0.134876	
skinconcern_pigmentation	0.009900	0.058502	
skinconcern_sensitivity	0.057427	-0.061607	
age	-0.000487	0.011253	
currentproducts_cleanser	0.096758	0.062266	
currentproducts_serum	0.018571	0.100384	
currentproducts_eyecream	0.046758	0.115648	
currentproducts_dailymoisturizer	0.083898	0.089474	
currentproducts_spf	0.050339	0.067541	
currentproducts_nightcream	0.052347	0.074066	

	skingoal_clear	skingoal_hydrated	\
skingoal_healthy	0.171254	0.152060	
skingoal_glowing	0.062029	0.250750	
skingoal_clear	1.000000	0.027226	
skingoal_hydrated	0.027226	1.000000	
skingoal_protected	NaN	NaN	
skingoal_balanced	NaN	NaN	
skingoal_repaired	NaN	NaN	

skingoal_calm	0.200298	0.174537
skingoal_bright	0.071096	0.249316
skingoal_rejuvenated	NaN	NaN
skingoal_restored	NaN	NaN
skinconcern_agingskin	-0.219311	0.087988
skinconcern_acne	0.495018	-0.158930
skinconcern_dryskin	-0.055879	0.327080
skinconcern_firmnesselasticity	-0.210383	0.118840
skinconcern_pigmentation	0.062020	-0.064201
skinconcern_sensitivity	0.091404	0.043927
age	-0.354481	0.057895
currentproducts_cleanser	0.124300	0.074259
currentproducts_serum	-0.014319	0.056068
currentproducts_eyecream	-0.102182	0.097031
currentproducts_dailymoisturizer	0.056654	0.129855
currentproducts_spf	0.011076	0.075307
currentproducts_nightcream	-0.072390	0.094726

	skingoal_protected	skingoal_balanced	\
skingoal_healthy	NaN	NaN	
skingoal_glowing	NaN	NaN	
skingoal_clear	NaN	NaN	
skingoal_hydrated	NaN	NaN	
skingoal_protected	NaN	NaN	
skingoal_balanced	NaN	NaN	
skingoal_repaired	NaN	NaN	
skingoal_calm	NaN	NaN	
skingoal_bright	NaN	NaN	
skingoal_rejuvenated	NaN	NaN	
skingoal_restored	NaN	NaN	
skinconcern_agingskin	NaN	NaN	
skinconcern_acne	NaN	NaN	
skinconcern_dryskin	NaN	NaN	
skinconcern_firmnesselasticity	NaN	NaN	
skinconcern_pigmentation	NaN	NaN	
skinconcern_sensitivity	NaN	NaN	
age	NaN	NaN	
currentproducts_cleanser	NaN	NaN	
currentproducts_serum	NaN	NaN	
currentproducts_eyecream	NaN	NaN	
currentproducts_dailymoisturizer	NaN	NaN	
currentproducts_spf	NaN	NaN	
currentproducts_nightcream	NaN	NaN	

	skingoal_repaired	skingoal_calm	\
skingoal_healthy	NaN	0.163388	
skingoal_glowing	NaN	0.037786	

skingoal_clear	NaN	0.200298
skingoal_hydrated	NaN	0.174537
skingoal_protected	NaN	NaN
skingoal_balanced	NaN	NaN
skingoal_repaired	NaN	NaN
skingoal_calm	NaN	1.000000
skingoal_bright	NaN	0.210745
skingoal_rejuvenated	NaN	NaN
skingoal_restored	NaN	NaN
skinconcern_agingskin	NaN	-0.030264
skinconcern_acne	NaN	0.064679
skinconcern_dryskin	NaN	0.002550
skinconcern_firmnesselasticity	NaN	-0.062397
skinconcern_pigmentation	NaN	-0.002336
skinconcern_sensitivity	NaN	0.253040
age	NaN	-0.006918
currentproducts_cleanser	NaN	0.085550
currentproducts_serum	NaN	0.051674
currentproducts_eyecream	NaN	0.034925
currentproducts_dailymoisturizer	NaN	0.061796
currentproducts_spf	NaN	0.035475
currentproducts_nightcream	NaN	0.064748

	skingoal_bright	skingoal_rejuvenated	…	\
skingoal_healthy	0.148360	NaN	…	
skingoal_glowing	0.353118	NaN	…	
skingoal_clear	0.071096	NaN	…	
skingoal_hydrated	0.249316	NaN	…	
skingoal_protected	NaN	NaN	…	
skingoal_balanced	NaN	NaN	…	
skingoal_repaired	NaN	NaN	…	
skingoal_calm	0.210745	NaN	…	
skingoal_bright	1.000000	NaN	…	
skingoal_rejuvenated	NaN	NaN	…	
skingoal_restored	NaN	NaN	…	
skinconcern_agingskin	0.150932	NaN	…	
skinconcern_acne	-0.089834	NaN	…	
skinconcern_dryskin	-0.020065	NaN	…	
skinconcern_firmnesselasticity	0.145425	NaN	…	
skinconcern_pigmentation	0.145175	NaN	…	
skinconcern_sensitivity	-0.075668	NaN	…	
age	0.082824	NaN	…	
currentproducts_cleanser	0.058210	NaN	…	
currentproducts_serum	0.134483	NaN	…	
currentproducts_eyecream	0.142774	NaN	…	
currentproducts_dailymoisturizer	0.081533	NaN	…	
currentproducts_spf	0.088016	NaN	…	

```
currentproducts_nightcream                0.099428                    NaN  …

                                  skinconcern_firmnesselasticity  \
skingoal_healthy                                        0.080553
skingoal_glowing                                        0.134876
skingoal_clear                                         -0.210383
skingoal_hydrated                                       0.118840
skingoal_protected                                           NaN
skingoal_balanced                                            NaN
skingoal_repaired                                            NaN
skingoal_calm                                          -0.062397
skingoal_bright                                         0.145425
skingoal_rejuvenated                                         NaN
skingoal_restored                                            NaN
skinconcern_agingskin                                   0.346728
skinconcern_acne                                       -0.404193
skinconcern_dryskin                                    -0.154048
skinconcern_firmnesselasticity                          1.000000
skinconcern_pigmentation                               -0.099313
skinconcern_sensitivity                                -0.302671
age                                                     0.379316
currentproducts_cleanser                                0.031746
currentproducts_serum                                   0.125124
currentproducts_eyecream                                0.222271
currentproducts_dailymoisturizer                        0.050623
currentproducts_spf                                     0.044460
currentproducts_nightcream                              0.176114

                                  skinconcern_pigmentation  \
skingoal_healthy                                  0.009900
skingoal_glowing                                  0.058502
skingoal_clear                                    0.062020
skingoal_hydrated                                -0.064201
skingoal_protected                                     NaN
skingoal_balanced                                      NaN
skingoal_repaired                                      NaN
skingoal_calm                                    -0.002336
skingoal_bright                                   0.145175
skingoal_rejuvenated                                   NaN
skingoal_restored                                      NaN
skinconcern_agingskin                            -0.096449
skinconcern_acne                                 -0.003364
skinconcern_dryskin                              -0.249889
skinconcern_firmnesselasticity                   -0.099313
skinconcern_pigmentation                          1.000000
skinconcern_sensitivity                          -0.188753
age                                              -0.058604
```

```
currentproducts_cleanser                          0.094319
currentproducts_serum                             0.083222
currentproducts_eyecream                         -0.016232
currentproducts_dailymoisturizer                  0.058194
currentproducts_spf                               0.118924
currentproducts_nightcream                        0.027995

                                  skinconcern_sensitivity       age  \
skingoal_healthy                                 0.057427 -0.000487
skingoal_glowing                                -0.061607  0.011253
skingoal_clear                                   0.091404 -0.354481
skingoal_hydrated                                0.043927  0.057895
skingoal_protected                                    NaN       NaN
skingoal_balanced                                     NaN       NaN
skingoal_repaired                                     NaN       NaN
skingoal_calm                                    0.253040 -0.006918
skingoal_bright                                 -0.075668  0.082824
skingoal_rejuvenated                                  NaN       NaN
skingoal_restored                                     NaN       NaN
skinconcern_agingskin                           -0.273609  0.524293
skinconcern_acne                                 0.059723 -0.429208
skinconcern_dryskin                              0.028588 -0.110689
skinconcern_firmnesselasticity                  -0.302671  0.379316
skinconcern_pigmentation                        -0.188753 -0.058604
skinconcern_sensitivity                          1.000000 -0.152979
age                                             -0.152979  1.000000
currentproducts_cleanser                         0.032681 -0.014911
currentproducts_serum                           -0.060920  0.127247
currentproducts_eyecream                        -0.075758  0.279668
currentproducts_dailymoisturizer                 0.053289  0.000535
currentproducts_spf                             -0.034021  0.013176
currentproducts_nightcream                      -0.079273  0.236250

                                  currentproducts_cleanser  \
skingoal_healthy                                  0.096758
skingoal_glowing                                  0.062266
skingoal_clear                                    0.124300
skingoal_hydrated                                 0.074259
skingoal_protected                                     NaN
skingoal_balanced                                      NaN
skingoal_repaired                                      NaN
skingoal_calm                                     0.085550
skingoal_bright                                   0.058210
skingoal_rejuvenated                                   NaN
skingoal_restored                                      NaN
skinconcern_agingskin                             0.083271
skinconcern_acne                                  0.083774
```

```
skinconcern_dryskin                              0.007060
skinconcern_firmnesselasticity                   0.031746
skinconcern_pigmentation                         0.094319
skinconcern_sensitivity                          0.032681
age                                             -0.014911
currentproducts_cleanser                         1.000000
currentproducts_serum                            0.289784
currentproducts_eyecream                         0.195429
currentproducts_dailymoisturizer                 0.448384
currentproducts_spf                              0.267231
currentproducts_nightcream                       0.219430

                                 currentproducts_serum  \
skingoal_healthy                              0.018571
skingoal_glowing                              0.100384
skingoal_clear                               -0.014319
skingoal_hydrated                             0.056068
skingoal_protected                                 NaN
skingoal_balanced                                  NaN
skingoal_repaired                                  NaN
skingoal_calm                                 0.051674
skingoal_bright                               0.134483
skingoal_rejuvenated                               NaN
skingoal_restored                                  NaN
skinconcern_agingskin                         0.162223
skinconcern_acne                             -0.066388
skinconcern_dryskin                          -0.047282
skinconcern_firmnesselasticity                0.125124
skinconcern_pigmentation                      0.083222
skinconcern_sensitivity                      -0.060920
age                                           0.127247
currentproducts_cleanser                      0.289784
currentproducts_serum                         1.000000
currentproducts_eyecream                      0.393443
currentproducts_dailymoisturizer              0.315938
currentproducts_spf                           0.360724
currentproducts_nightcream                    0.379769

                                 currentproducts_eyecream  \
skingoal_healthy                                 0.046758
skingoal_glowing                                 0.115648
skingoal_clear                                  -0.102182
skingoal_hydrated                                0.097031
skingoal_protected                                    NaN
skingoal_balanced                                     NaN
skingoal_repaired                                     NaN
skingoal_calm                                    0.034925
```

skingoal_bright	0.142774
skingoal_rejuvenated	NaN
skingoal_restored	NaN
skinconcern_agingskin	0.302801
skinconcern_acne	-0.192119
skinconcern_dryskin	-0.029620
skinconcern_firmnesselasticity	0.222271
skinconcern_pigmentation	-0.016232
skinconcern_sensitivity	-0.075758
age	0.279668
currentproducts_cleanser	0.195429
currentproducts_serum	0.393443
currentproducts_eyecream	1.000000
currentproducts_dailymoisturizer	0.229388
currentproducts_spf	0.262610
currentproducts_nightcream	0.433237

	currentproducts_dailymoisturizer \
skingoal_healthy	0.083898
skingoal_glowing	0.089474
skingoal_clear	0.056654
skingoal_hydrated	0.129855
skingoal_protected	NaN
skingoal_balanced	NaN
skingoal_repaired	NaN
skingoal_calm	0.061796
skingoal_bright	0.081533
skingoal_rejuvenated	NaN
skingoal_restored	NaN
skinconcern_agingskin	0.102661
skinconcern_acne	0.002657
skinconcern_dryskin	0.064521
skinconcern_firmnesselasticity	0.050623
skinconcern_pigmentation	0.058194
skinconcern_sensitivity	0.053289
age	0.000535
currentproducts_cleanser	0.448384
currentproducts_serum	0.315938
currentproducts_eyecream	0.229388
currentproducts_dailymoisturizer	1.000000
currentproducts_spf	0.302097
currentproducts_nightcream	0.238674

	currentproducts_spf \
skingoal_healthy	0.050339
skingoal_glowing	0.067541
skingoal_clear	0.011076

skingoal_hydrated	0.075307
skingoal_protected	NaN
skingoal_balanced	NaN
skingoal_repaired	NaN
skingoal_calm	0.035475
skingoal_bright	0.088016
skingoal_rejuvenated	NaN
skingoal_restored	NaN
skinconcern_agingskin	0.104431
skinconcern_acne	-0.045986
skinconcern_dryskin	-0.025684
skinconcern_firmnesselasticity	0.044460
skinconcern_pigmentation	0.118924
skinconcern_sensitivity	-0.034021
age	0.013176
currentproducts_cleanser	0.267231
currentproducts_serum	0.360724
currentproducts_eyecream	0.262610
currentproducts_dailymoisturizer	0.302097
currentproducts_spf	1.000000
currentproducts_nightcream	0.295952

	currentproducts_nightcream
skingoal_healthy	0.052347
skingoal_glowing	0.074066
skingoal_clear	-0.072390
skingoal_hydrated	0.094726
skingoal_protected	NaN
skingoal_balanced	NaN
skingoal_repaired	NaN
skingoal_calm	0.064748
skingoal_bright	0.099428
skingoal_rejuvenated	NaN
skingoal_restored	NaN
skinconcern_agingskin	0.235638
skinconcern_acne	-0.146724
skinconcern_dryskin	-0.023880
skinconcern_firmnesselasticity	0.176114
skinconcern_pigmentation	0.027995
skinconcern_sensitivity	-0.079273
age	0.236250
currentproducts_cleanser	0.219430
currentproducts_serum	0.379769
currentproducts_eyecream	0.433237
currentproducts_dailymoisturizer	0.238674
currentproducts_spf	0.295952
currentproducts_nightcream	1.000000

```
[24 rows x 24 columns]
```

Acne/clear and dryskin/hydrated seem to be correalted, as expected, but there isn't any other probable correlation

5 Part 4: Skin concerns by age

For question 3 I need to group by unique age (group), and look at the counts and mean of each concern. For question 4 the process is he same except with ethnicity instead of age.

[153]:
```
aggData = cdata.groupby('age')
aggData = aggData.agg(['sum','mean'])
aggData.head()
```

[153]:

	skingoal_healthy		skingoal_glowing		skingoal_clear	\
	sum	mean	sum	mean	sum	
age						
21.0	2455.0	0.805446	1993.0	0.653871	2826.0	
29.5	2299.0	0.821658	1680.0	0.600429	2363.0	
39.5	1683.0	0.834821	1319.0	0.654266	1392.0	
49.5	1189.0	0.792139	990.0	0.659560	786.0	
59.5	121.0	0.812081	104.0	0.697987	74.0	

	skingoal_hydrated		skingoal_protected		…	\
	mean	sum	mean	sum	mean	…
age						…
21.0	0.927165	2064.0	0.677165	0.0	0.0	…
29.5	0.844532	1962.0	0.701215	0.0	0.0	…
39.5	0.690476	1379.0	0.684028	0.0	0.0	…
49.5	0.523651	1162.0	0.774151	0.0	0.0	…
59.5	0.496644	110.0	0.738255	0.0	0.0	…

	currentproducts_serum		currentproducts_eyecream		\
	sum	mean	sum	mean	
age					
21.0	1458	0.478346	665	0.218176	
29.5	1750	0.625447	1131	0.404217	
39.5	1332	0.660714	1069	0.530258	
49.5	991	0.660227	945	0.629580	
59.5	96	0.644295	74	0.496644	

	currentproducts_dailymoisturizer		currentproducts_spf		\
	sum	mean	sum	mean	
age					
21.0	2628	0.862205	1563	0.512795	
29.5	2485	0.888134	1621	0.579342	

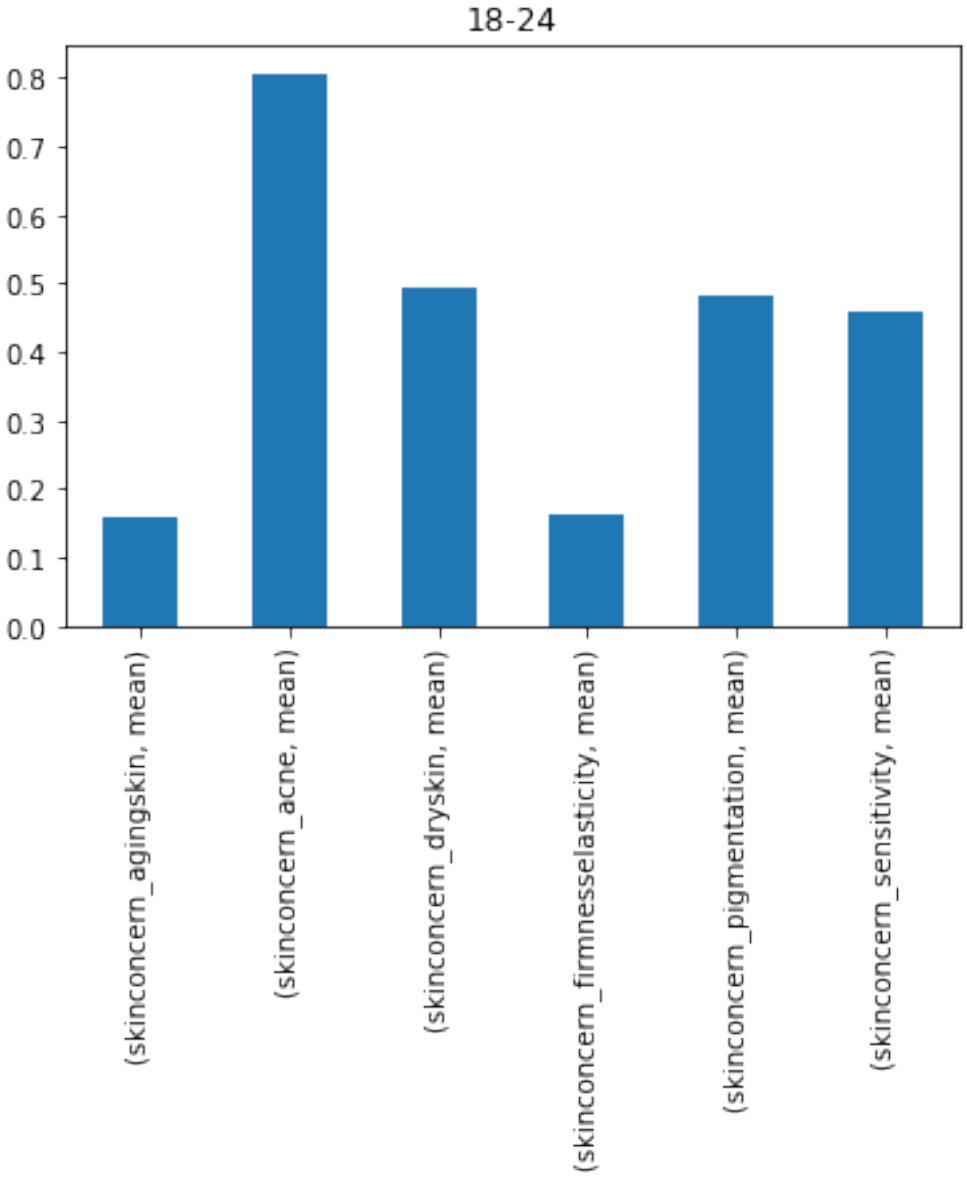
```
39.5                          1720  0.853175                1133  0.562004
49.5                          1325  0.882745                 836  0.556962
59.5                           133  0.892617                  62  0.416107

     currentproducts_nightcream
                            sum      mean
age
21.0                        824  0.270341
29.5                       1358  0.485347
39.5                       1037  0.514385
49.5                        949  0.632245
59.5                         89  0.597315

[5 rows x 46 columns]
```

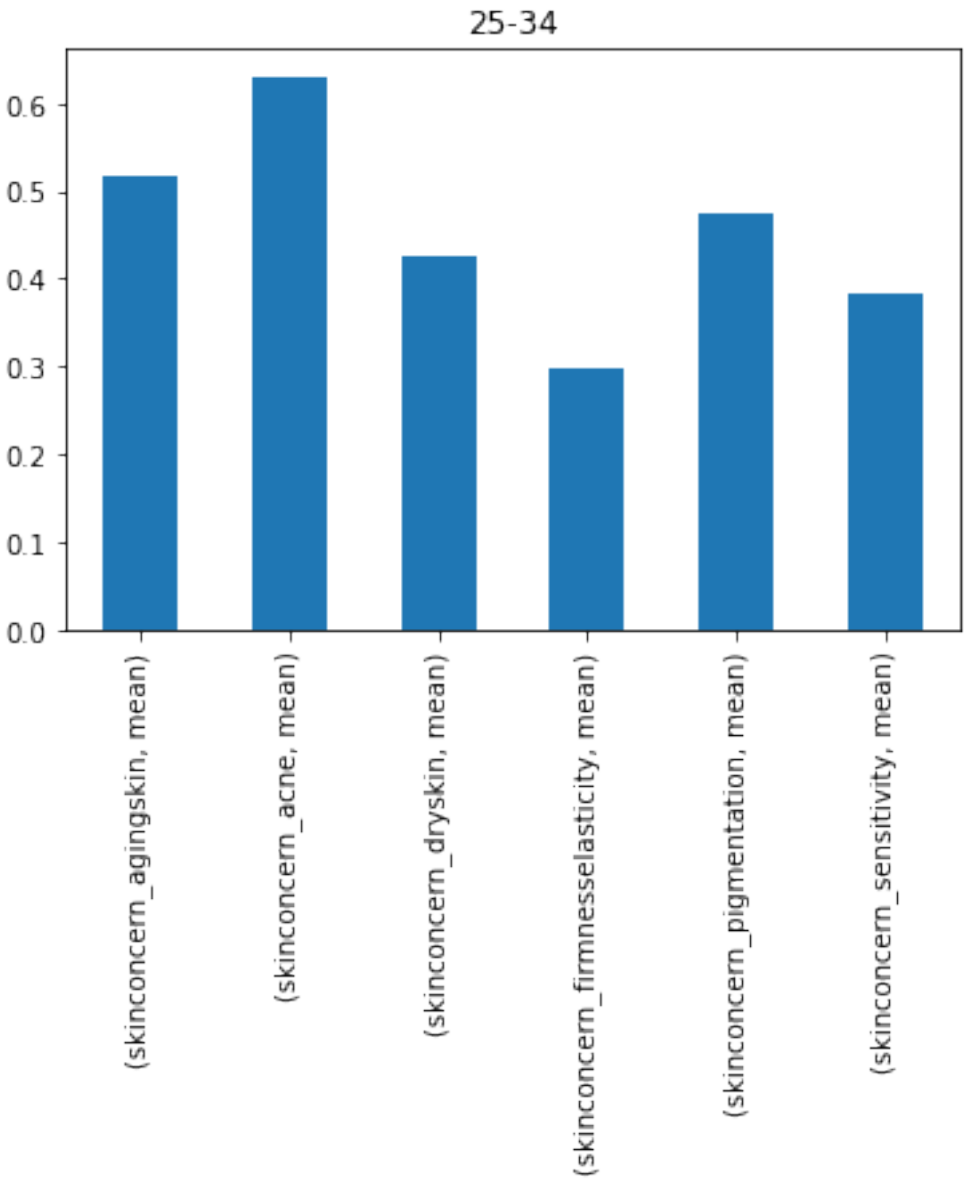
[154]:
```
aggData.iloc[:,22:34]
```

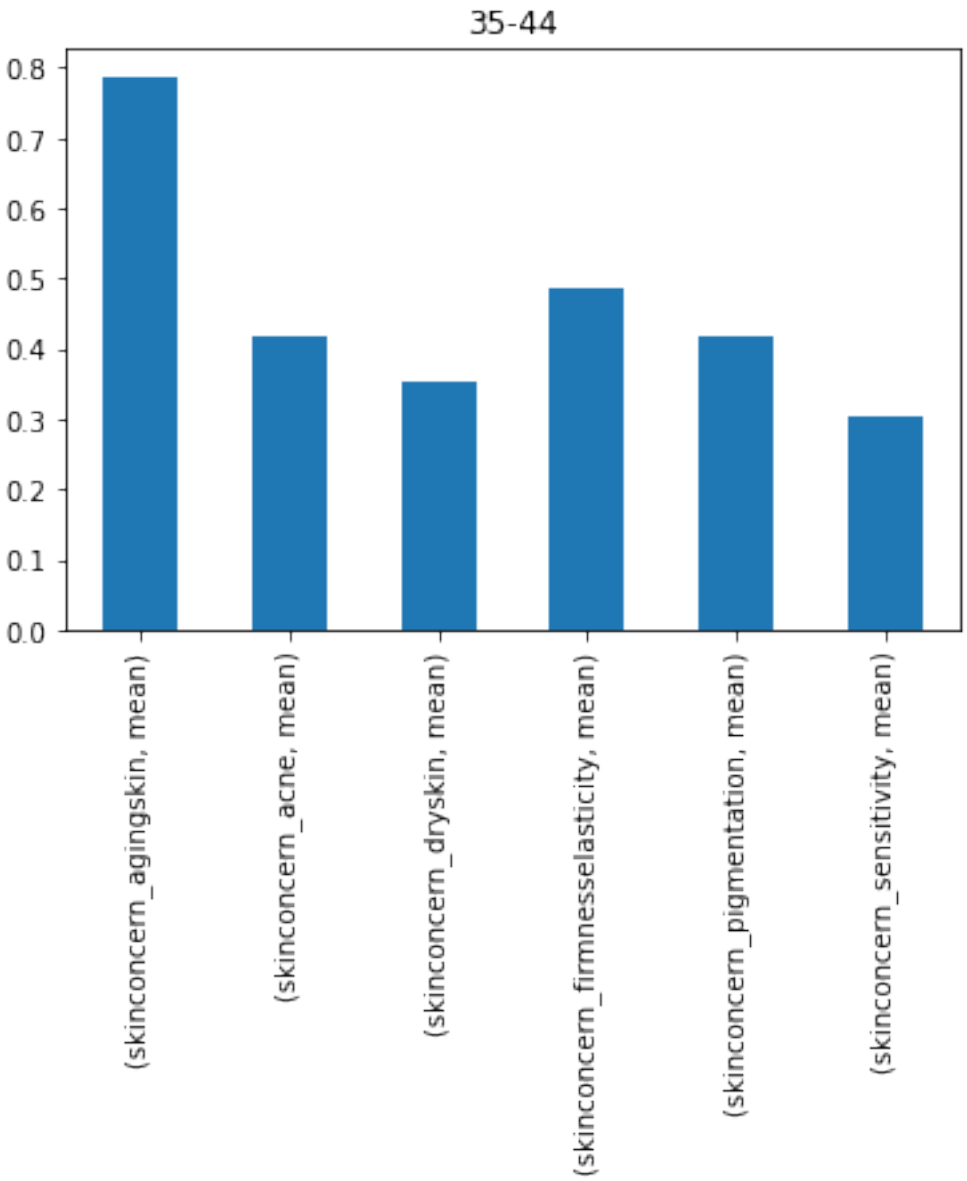
[154]:
```
     skinconcern_agingskin           skinconcern_acne            \
                       sum      mean              sum      mean
age
21.0                  489.0  0.160433           2459.0  0.806759
29.5                 1447.0  0.517155           1764.0  0.630450
39.5                 1588.0  0.787698            845.0  0.419147
49.5                 1298.0  0.864757            314.0  0.209194
59.5                  133.0  0.892617             35.0  0.234899
82.0                   15.0  0.833333              2.0  0.111111

     skinconcern_dryskin           skinconcern_firmnesselasticity            \
                     sum      mean                            sum      mean
age
21.0              1505.0  0.493766                          499.0  0.163714
29.5              1189.0  0.424946                          830.0  0.296640
39.5               709.0  0.351687                          982.0  0.487103
49.5               541.0  0.360426                         1005.0  0.669554
59.5                50.0  0.335570                          101.0  0.677852
82.0                 4.0  0.222222                           14.0  0.777778

     skinconcern_pigmentation           skinconcern_sensitivity
                          sum      mean                     sum      mean
age
21.0                   1466.0  0.480971                  1394.0  0.457349
29.5                   1328.0  0.474625                  1074.0  0.383846
39.5                    842.0  0.417659                   614.0  0.304563
49.5                    618.0  0.411726                   397.0  0.264490
59.5                     57.0  0.382550                    30.0  0.201342
82.0                      8.0  0.444444                     4.0  0.222222
```

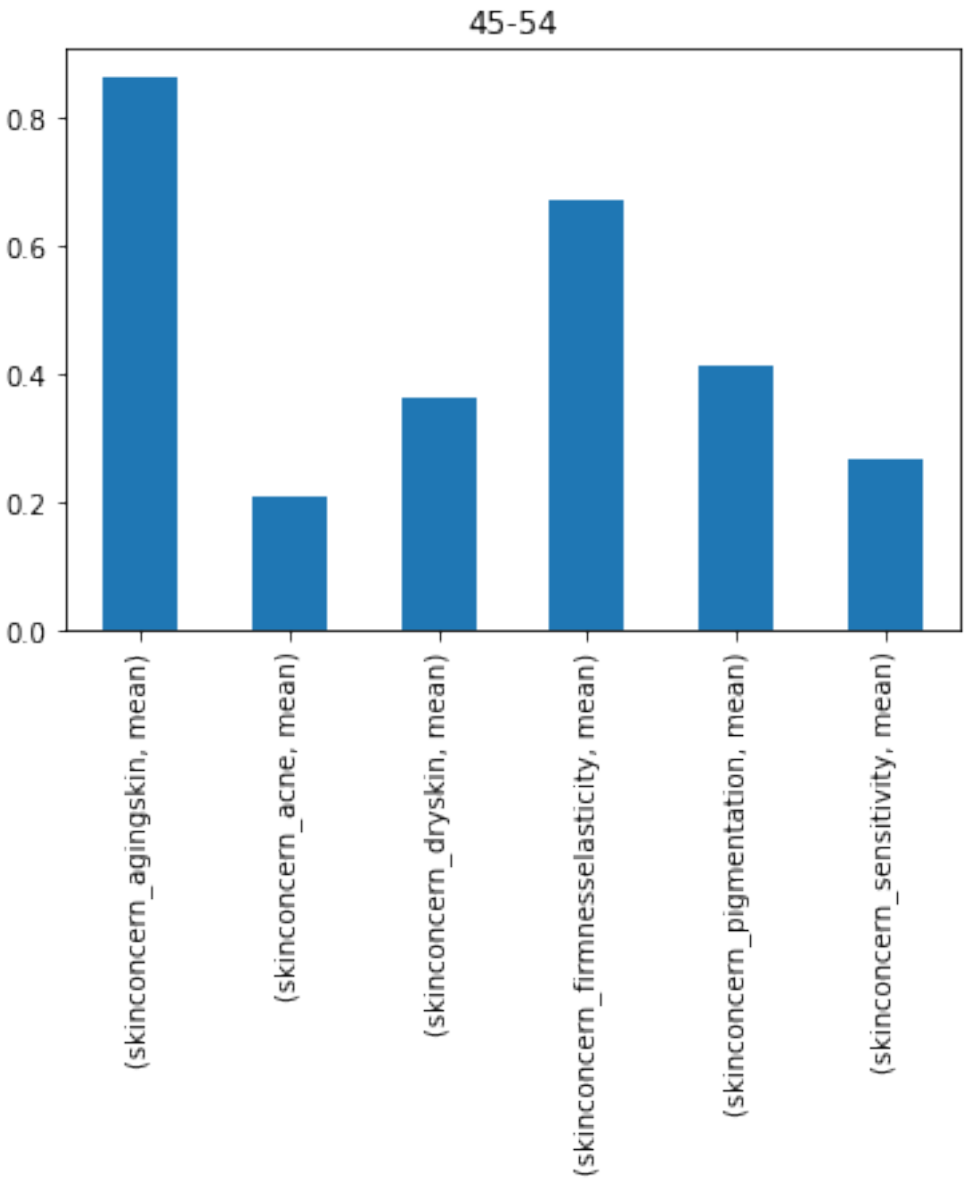
[213]:
```
aggData = cdata.groupby('age')
aggData = aggData.agg(['mean'])
```

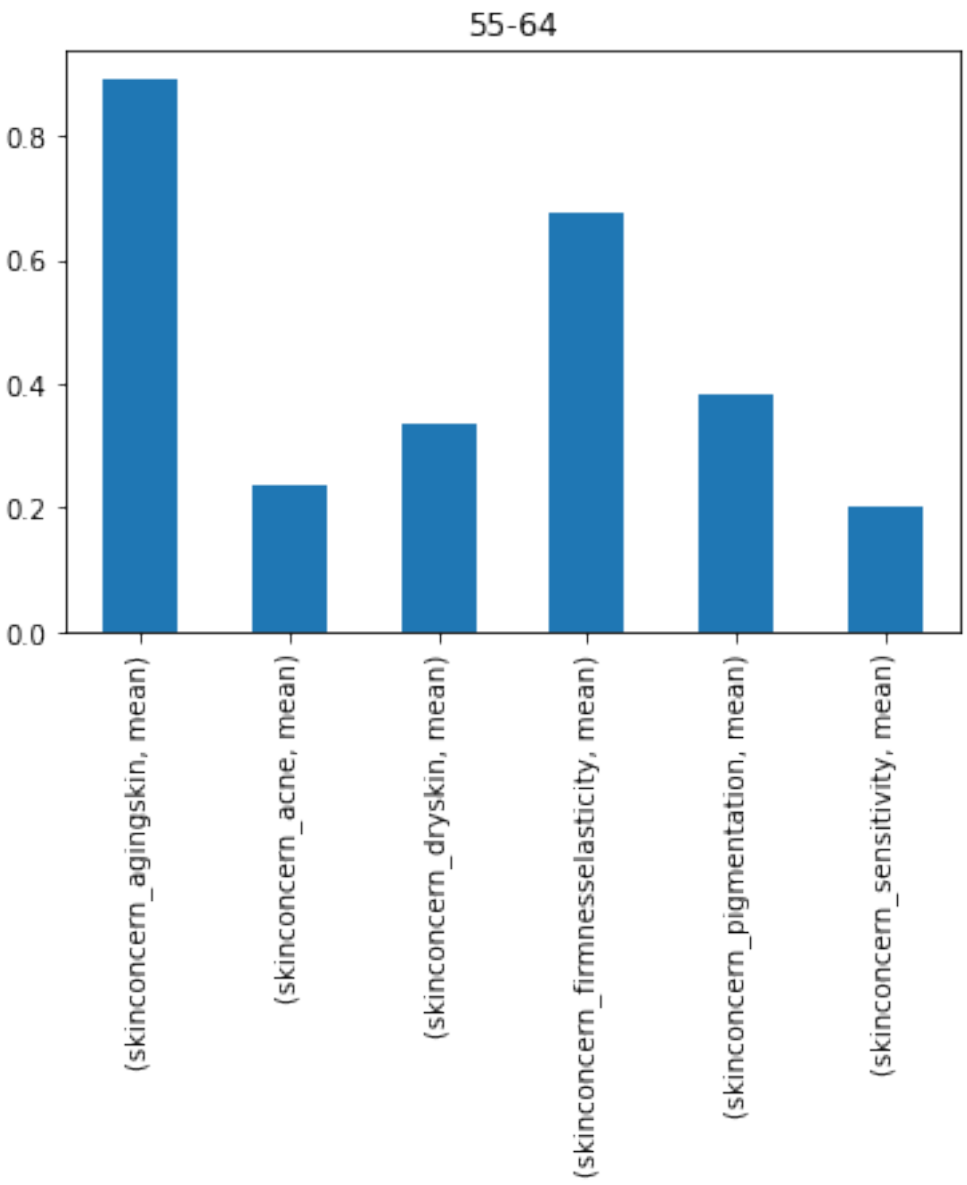
[212]:
```
ages = ['18-24','25-34','35-44','45-54','55-64','65-99']
for ii in range(0,6):
    aggData.iloc[ii,11:17].plot(subplots = True,kind = 'bar')
    plt.title(ages[ii])
    plt.show()
```

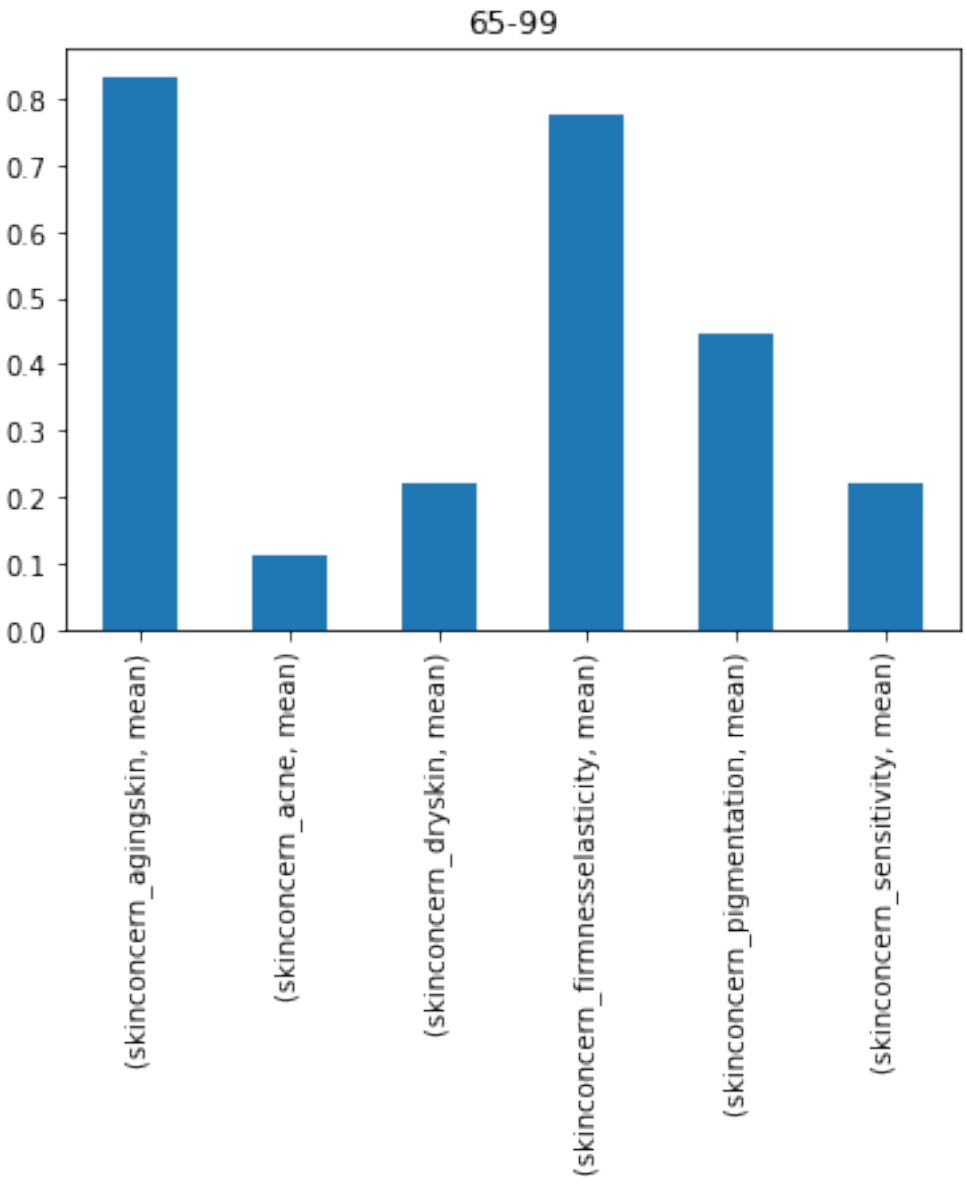



25-34
0.6
0.5
0.4
0.3
0.2
0.1
0.0
(skinconcern_agingskin, mean)
(skinconcern_acne, mean)
(skinconcern_dryskin, mean)
(skinconcern_firmnesselasticity, mean)
(skinconcern_pigmentation, mean)
(skinconcern_sensitivity, mean)


35-44
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
(skinconcern_agingskin, mean)
(skinconcern_acne, mean)
(skinconcern_dryskin, mean)
(skinconcern_firmnesselasticity, mean)
(skinconcern_pigmentation, mean)
(skinconcern_sensitivity, mean)


45-54
0.8
0.6
0.4
0.2
0.0
(skinconcern_agingskin, mean)
(skinconcern_acne, mean)
(skinconcern_dryskin, mean)
(skinconcern_firmnesselasticity, mean)
(skinconcern_pigmentation, mean)
(skinconcern_sensitivity, mean)


55-64
0.0
0.2
0.4
0.6
0.8
(skinconcern_agingskin, mean)
(skinconcern_acne, mean)
(skinconcern_dryskin, mean)
(skinconcern_firmnesselasticity, mean)
(skinconcern_pigmentation, mean)
(skinconcern_sensitivity, mean)



18-24: Acne, dry skin, pigmenation, and sensitivty are all concerns here (in descending order)

25-34: Everything seems to be a concern here with acne being the highest and firmness elascity being the lowest

35-44: Starting from 35 aging skin is the biggest concern. In 35-44 everything is high enough that I think the group is concerned with sensitivty being the lowest.

45-54: This group seems to be mostly concerned about aging skin and firmness elascity, but the other categories are still returning high numbers!

55-64: This group is almost identical to the 45-54 group when it comes to concerns.

65-99: This group follows the same distribution as the 45-54 and 55-64 groups but with even more

emphasis on firmness elascity, but aging skin is still the biggest concern.

6 Part 5: Skin concerns by ethnicity

[155]:
```
aggDataEth = cdata.groupby('ethnicity')
aggDataEth = aggDataEth.agg(['sum','mean'])
aggDataEth.iloc[:,22:34]
```

[155]:

	skinconcern_agingskin		skinconcern_acne	
	sum	mean	sum	mean
ethnicity				
africanamerican	230.0	0.269321	624.0	0.730679
asian	436.0	0.439074	564.0	0.567976
caucasian	3529.0	0.586993	3235.0	0.538090
hispanic	624.0	0.469526	786.0	0.591422
nativeamerican	67.0	0.372222	121.0	0.672222
pacific	76.0	0.424581	104.0	0.581006

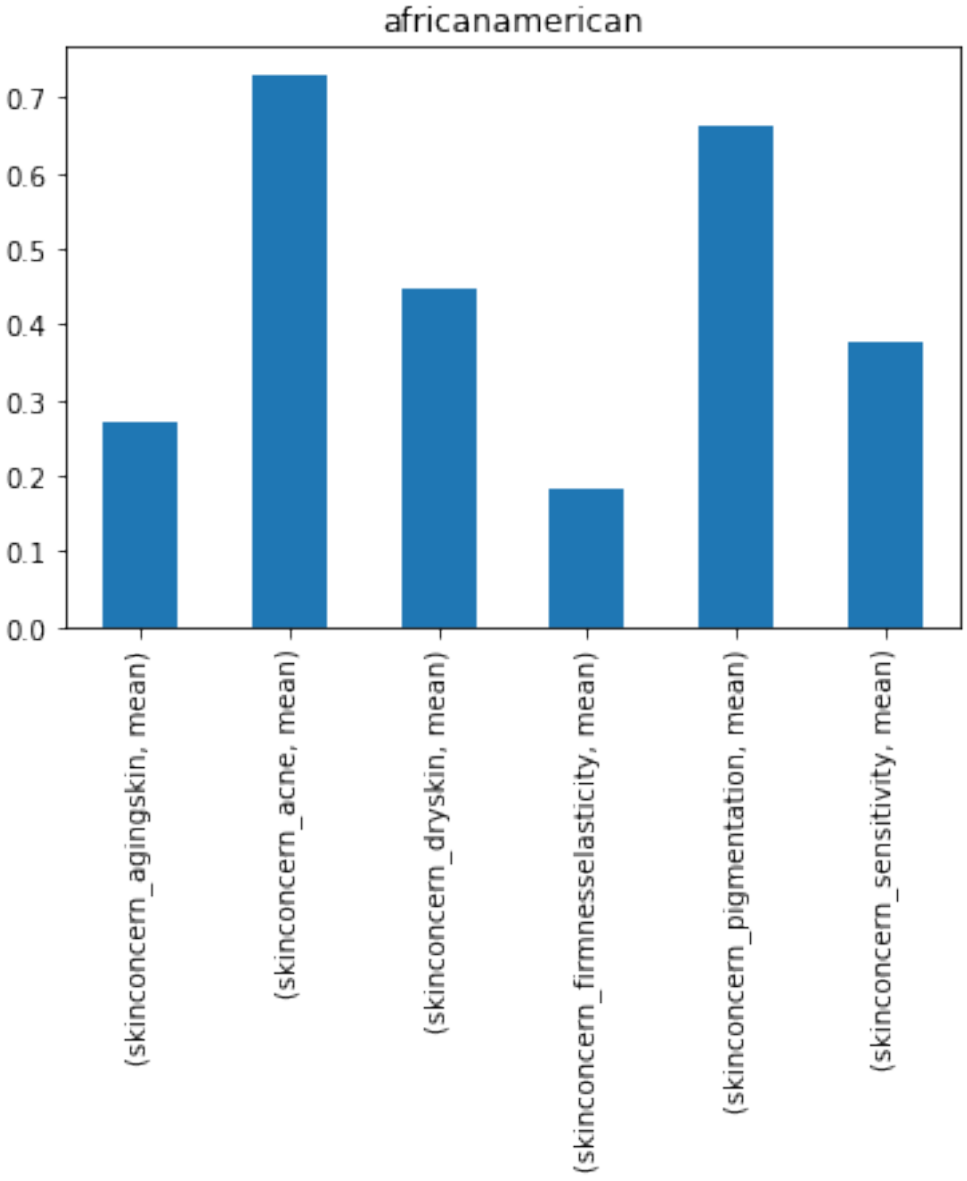
	skinconcern_dryskin		skinconcern_firmnesselasticity
	sum	mean	sum
ethnicity			
africanamerican	381.0	0.446136	155.0
asian	438.0	0.441088	309.0
caucasian	2454.0	0.408184	2360.0
hispanic	588.0	0.442438	505.0
nativeamerican	70.0	0.388889	44.0
pacific	79.0	0.441341	61.0

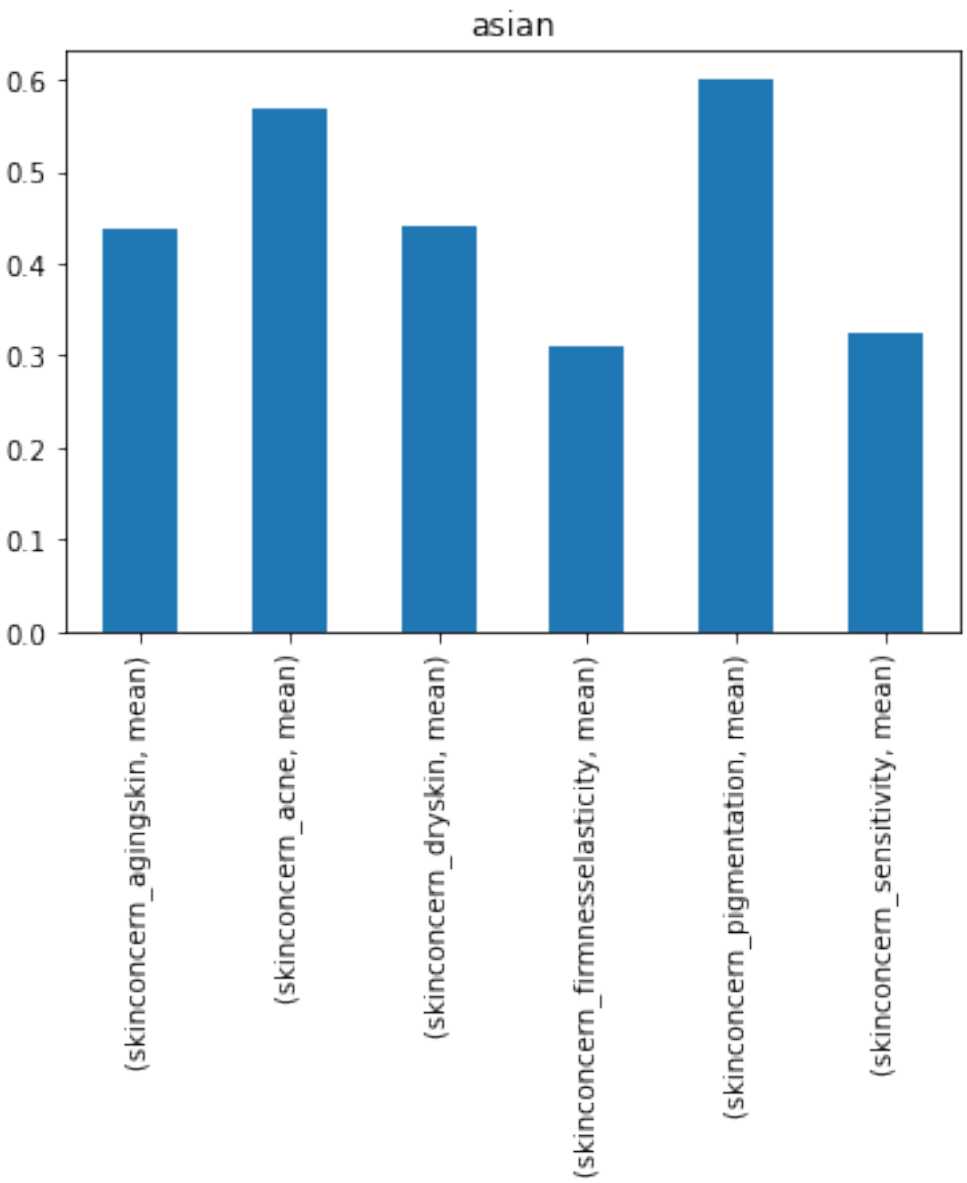
	skinconcern_pigmentation		
	mean	sum	mean
ethnicity			
africanamerican	0.181499	567.0	0.663934
asian	0.311178	597.0	0.601208
caucasian	0.392548	2305.0	0.383400
hispanic	0.379985	678.0	0.510158
nativeamerican	0.244444	81.0	0.450000
pacific	0.340782	91.0	0.508380

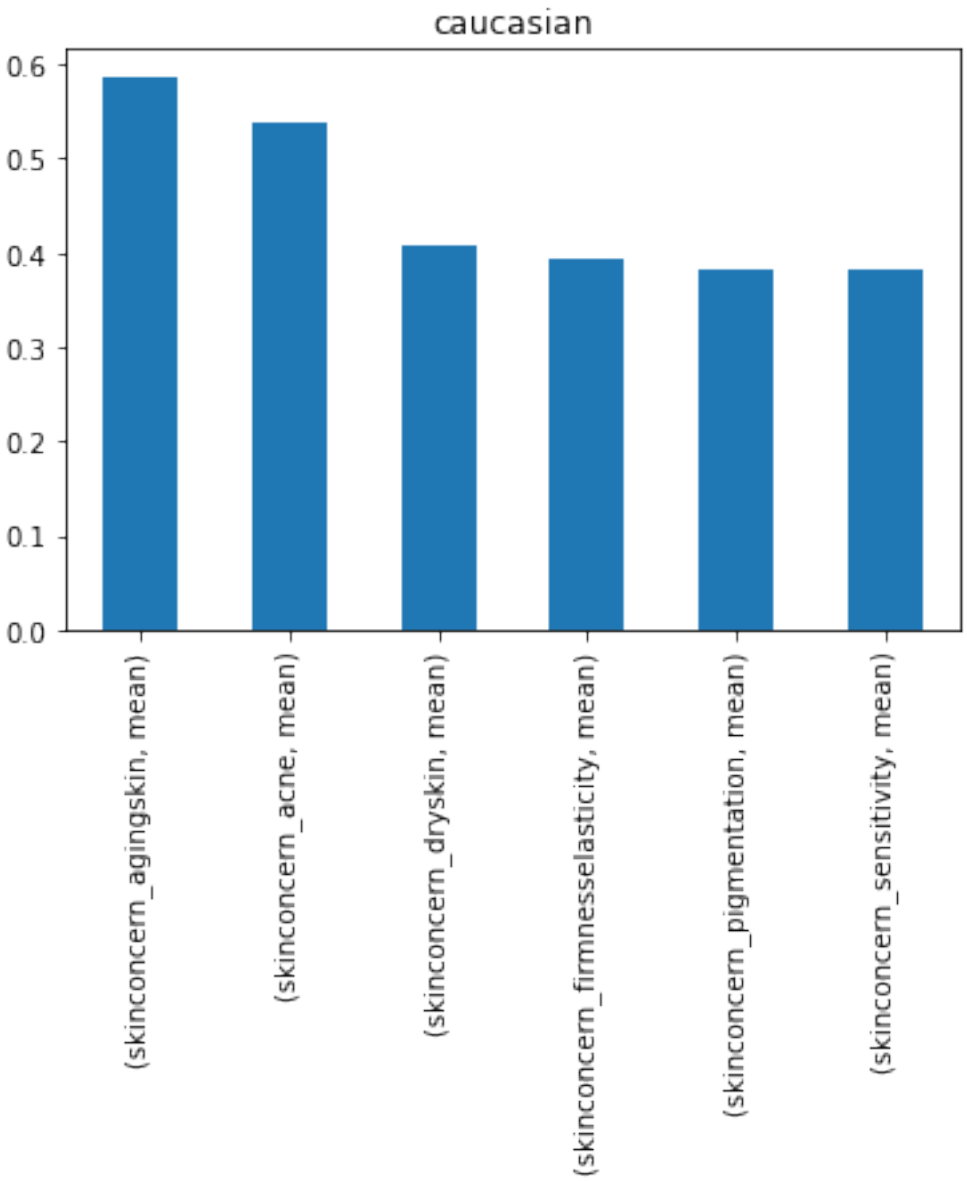
	skinconcern_sensitivity	
	sum	mean
ethnicity		
africanamerican	321.0	0.375878
asian	321.0	0.323263
caucasian	2290.0	0.380905
hispanic	444.0	0.334086
nativeamerican	78.0	0.433333
pacific	74.0	0.413408

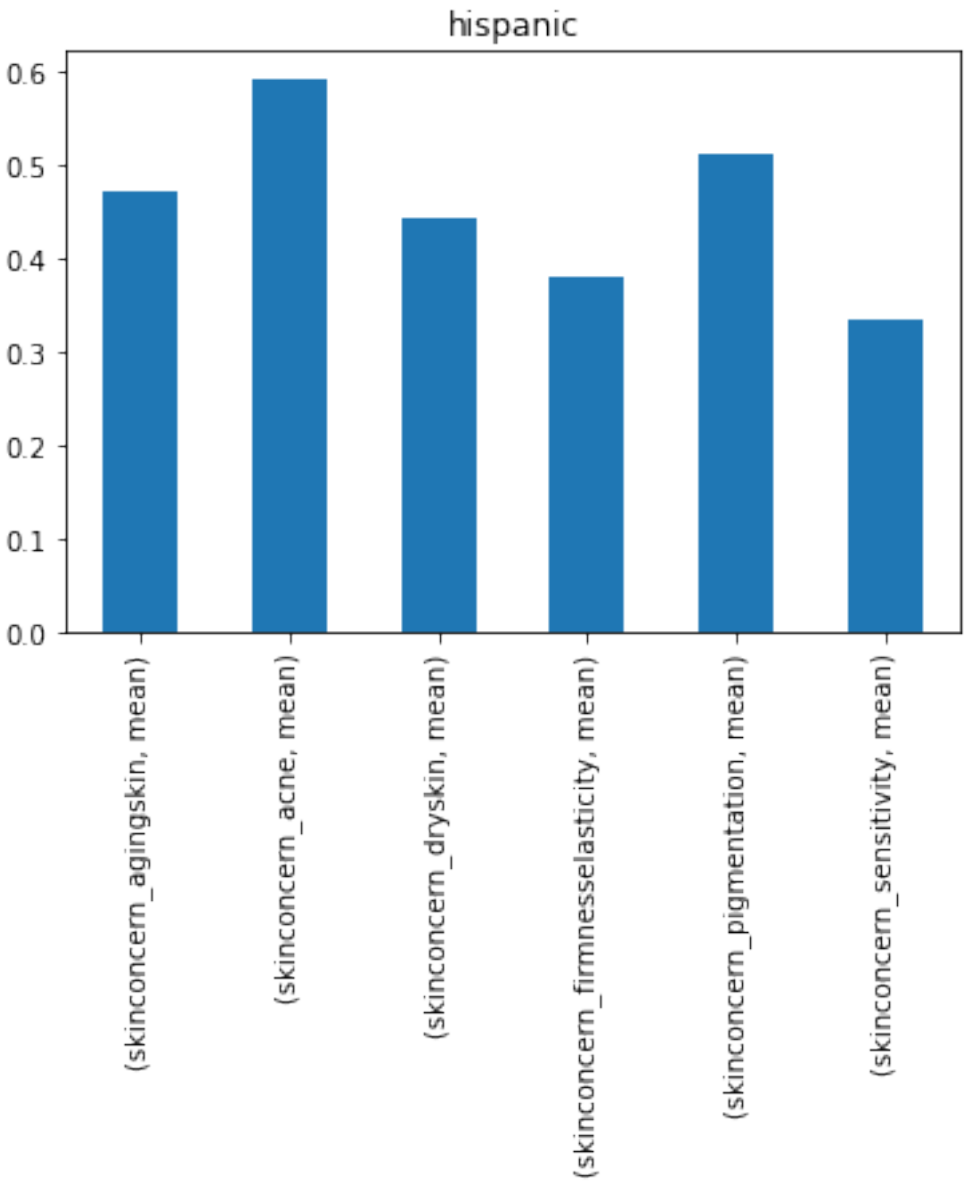
[214]:
```
aggDataEth = cdata.groupby('ethnicity')
aggDataEth = aggDataEth.agg(['mean'])
```

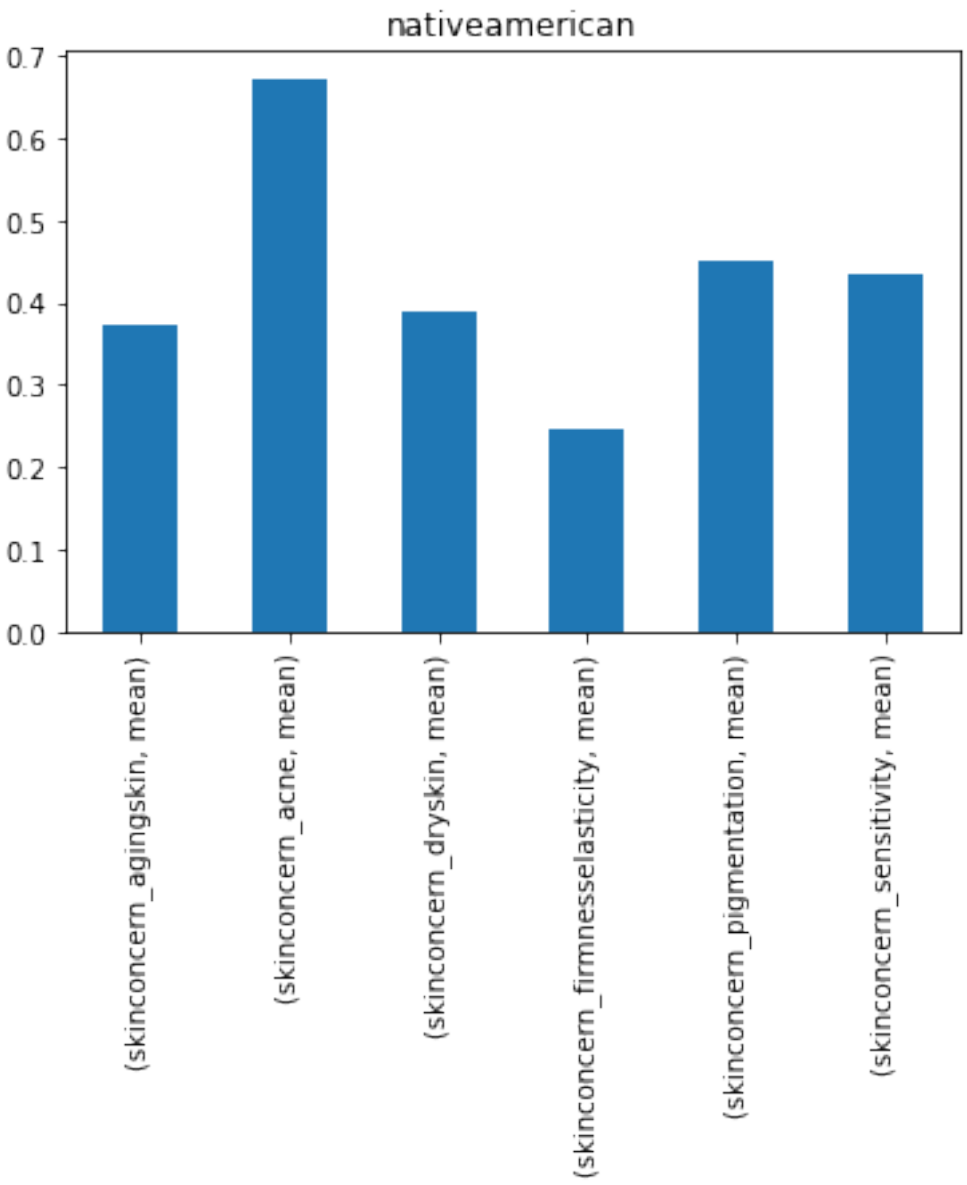
[206]:
```
eths =␣
 ↪['africanamerican','asian','caucasian','hispanic','nativeamerican','pacific']
for ii in range(0,6):
    aggDataEth.iloc[ii,11:17].plot(subplots = True,kind = 'bar')
    plt.title(eths[ii])
    plt.show()
```

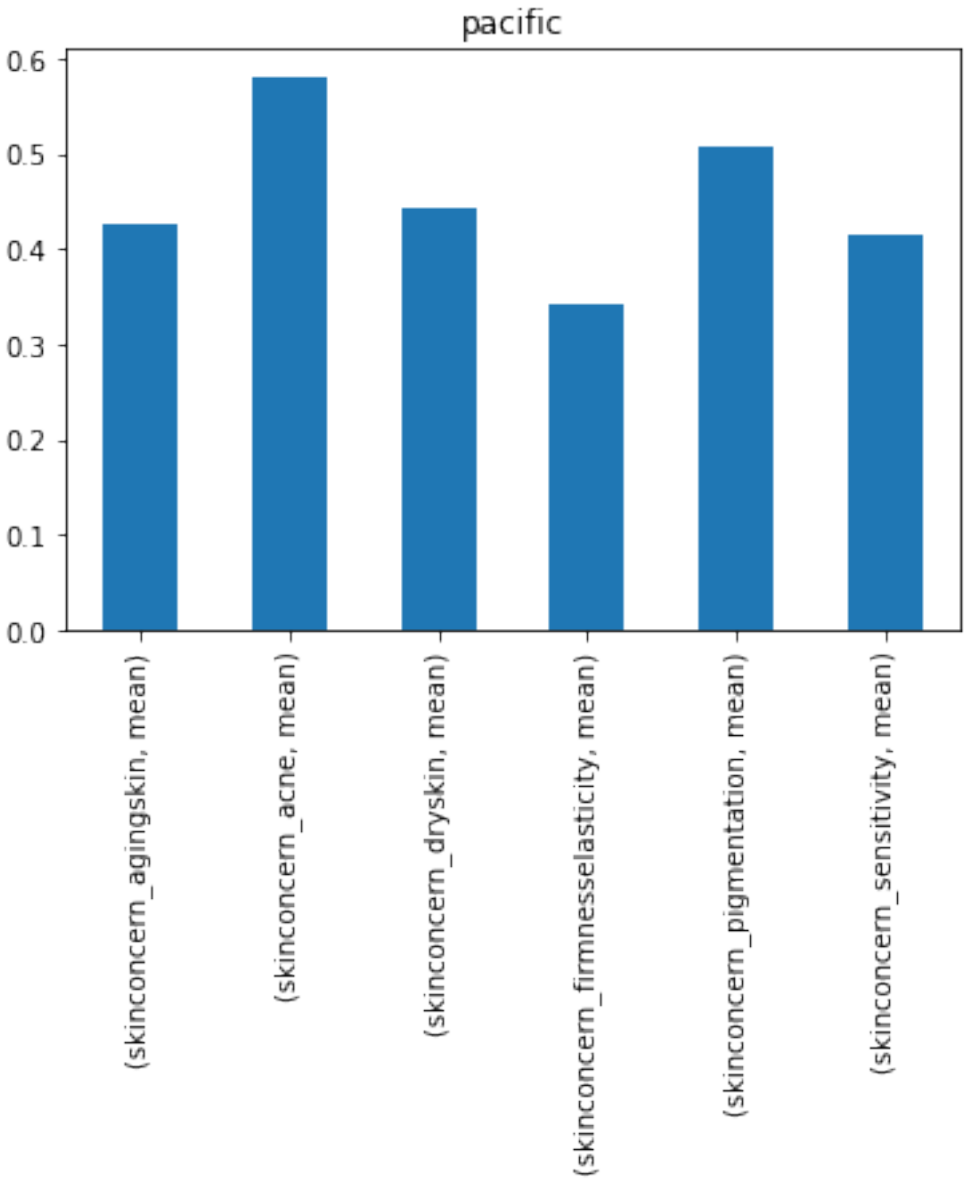



asian
0.6
0.5
0.4
0.3
0.2
0.1
0.0
(skinconcern_agingskin, mean)
(skinconcern_acne, mean)
(skinconcern_dryskin, mean)
(skinconcern_firmnesselasticity, mean)
(skinconcern_pigmentation, mean)
(skinconcern_sensitivity, mean)


caucasian
0.6
0.5
0.4
0.3
0.2
0.1
0.0
(skinconcern_agingskin, mean)
(skinconcern_acne, mean)
(skinconcern_dryskin, mean)
(skinconcern_firmnesselasticity, mean)
(skinconcern_pigmentation, mean)
(skinconcern_sensitivity, mean)


hispanic
0.6
0.5
0.4
0.3
0.2
0.1
0.0
(skinconcern_agingskin, mean)
(skinconcern_acne, mean)
(skinconcern_dryskin, mean)
(skinconcern_firmnesselasticity, mean)
(skinconcern_pigmentation, mean)
(skinconcern_sensitivity, mean)


nativeamerican
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
(skinconcern_agingskin, mean)
(skinconcern_acne, mean)
(skinconcern_dryskin, mean)
(skinconcern_firmnesselasticity, mean)
(skinconcern_pigmentation, mean)
(skinconcern_sensitivity, mean)



Africanamerican: Acne and pigmentaiton are the biggest concerns. African Americans care about Pigmentation the most out of all the ethnicites in this dataset. I can't say exactly why without some major assumptions since I'm white, but there is certainly something interesting going on here.

Asian: Pigmentation is the greatest concern here. I'd imagine that the reason why Asians are concerned about pigmentation is the same or similar to African Americans. Acne is the 2nd highest here.

Caucasian: Aging skin is the greatest concern followed by acne.

Hispanic: Acne and pigmentation here again.

Nativeamerican: Acne is by far the greatest concern here. However, the total data points for this

ethnicity is relatively low so there is a high vulneravility of noise

Pacific: Everything is about the same here with acne being the highest, again.

7 Part 6: Conclusions, and extra thoughts

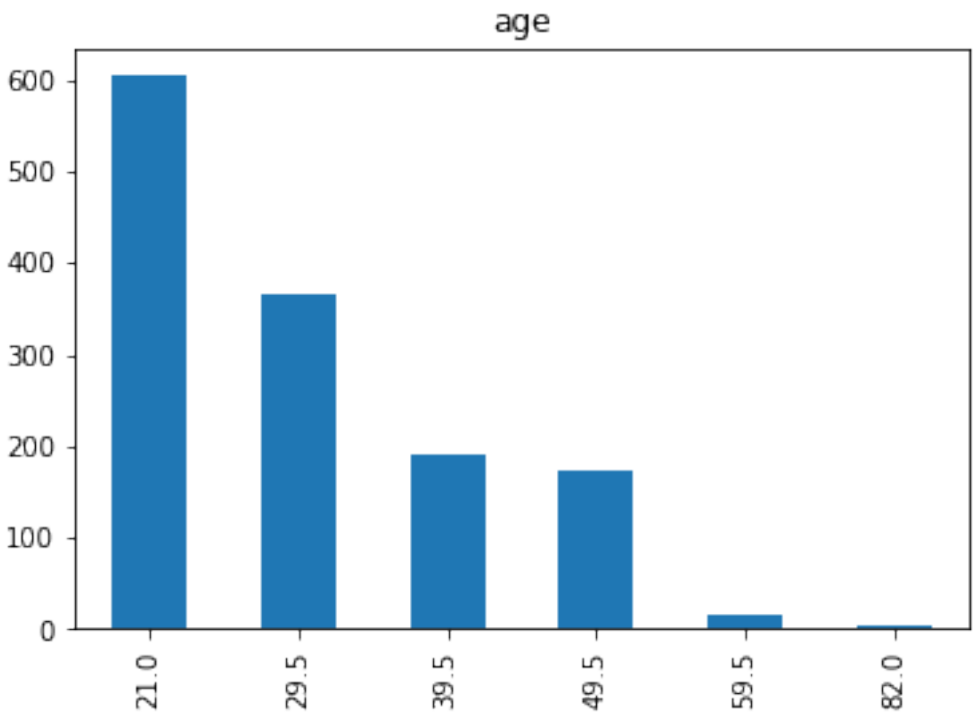
Overall the dataset returns what's to be expected. I think there could be some interesting analysis to do on subsets of the data, for example cut out the acne focused data points and see what's left. People's skin concerns not correlating to many of the goals was also interesting and a deep dive into what's really going on there could produce a psychological view on how we perceive our skin problems.

As a side note it could be fun to build an ML algorithm to predict skin concern / goal or product based on skin type, skin after wash, t zone oily, and clogged pores. Could even throw in age/gender/eth into the algo.

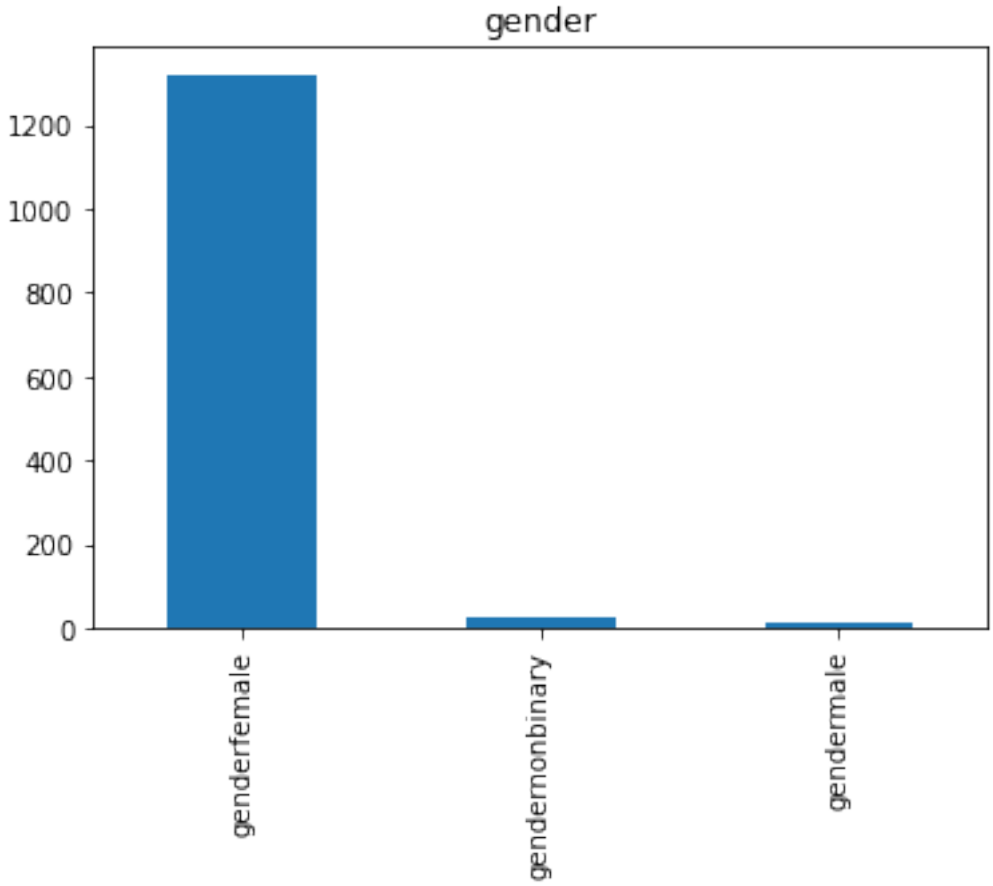
As a quick extra I'm curious about who are the vegans in the dataset so this is a quick look at the age, gender, ethnicity distribution of vegans

[217]:
```
vdata = cdata.loc[cdata['vegan']=='veganyes']
```

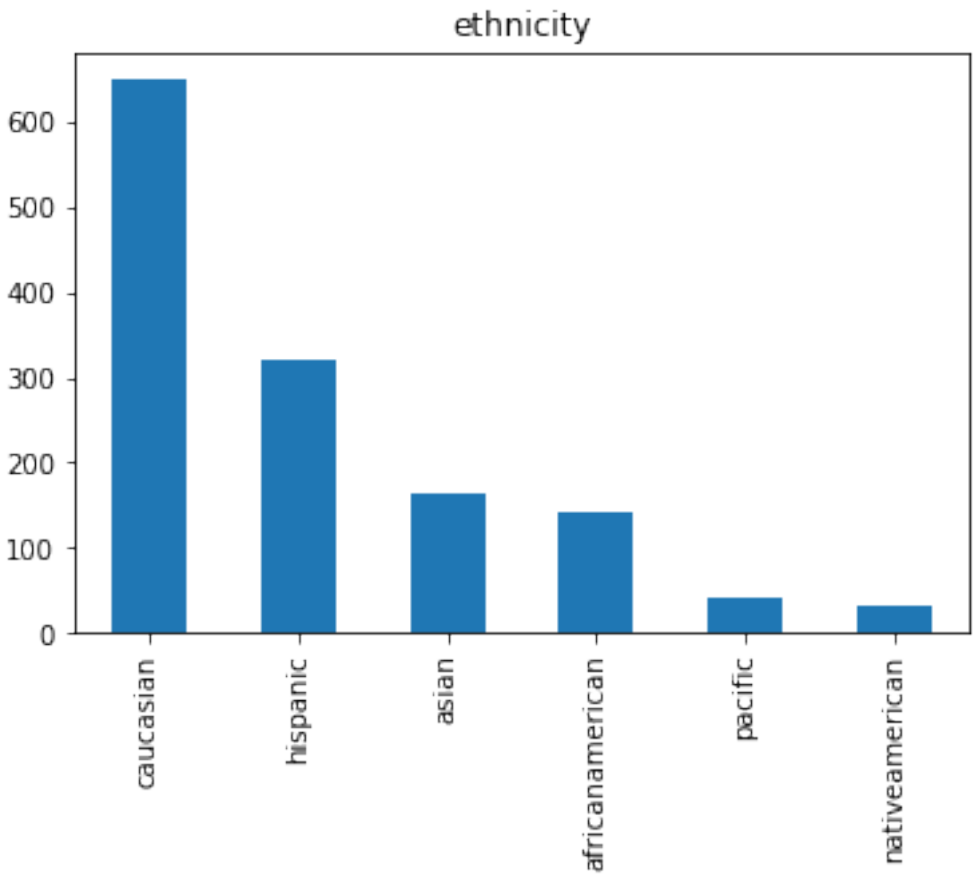
[224]:
```
vdata.age.value_counts().plot(subplots = True, kind = 'bar')
plt.show()
```



[225]:
```
vdata.gender.value_counts().plot(subplots = True, kind = 'bar')
plt.show()
```



[226]:
```
vdata.ethnicity.value_counts().plot(subplots = True, kind = 'bar')
plt.show()
```



All of these look almost idential to the full dataset except for ethnicity, which is kind of interesting.

[ ]: