Weeks 5&6 Exercises

Author: Jordan,Andrew

Date: 07/14/2022

In [276…

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from datetime import datetime
%pwd
```

Out[276]:

```
'C:\\Users\\Andrew\\Documents\\Grad School\\DSC 540 - Data Preparation\\Assignments'
```

In [277…

```python
#met = pd.read_csv("data/Weeks 5 & 6 Data/MetObjects.csv")
candy17 = pd.read_csv("data/Weeks 5 & 6 Data/candyhierarchy2017.csv", encoding = "ISO-8859-1", header=0)
candy15 = pd.read_excel("data/Weeks 5 & 6 Data/CANDY-HIERARCHY-2015-SURVEY-Responses.xlsx")
candy15.head()
```

Out[277]:

	Timestamp	How old are you?	Are you going actually going trick or treating yourself?	[Butterfinger]	[100 Grand Bar]	[Anonymous brown globs that come in black and orange wrappers]	[Any full-sized candy bar]	[Black Jacks]	[Bonkers]	[Bottle Caps]	...	[Necco Wafers]	Which day do you prefer, Friday or Sunday?	Ple estim degree separat you h from follow fo [Br L
0	2015-10-23 08:46:20.451	35.0	No	JOY	NaN	DESPAIR	JOY	NaN	NaN	NaN	...	NaN	NaN	N
1	2015-10-23 08:46:51.583	41.0	No	JOY	JOY	DESPAIR	JOY	DESPAIR	DESPAIR	JOY	...	DESPAIR	NaN	N
2	2015-10-23 08:47:34.285	33.0	No	DESPAIR	DESPAIR	DESPAIR	JOY	DESPAIR	DESPAIR	DESPAIR	...	DESPAIR	NaN	N
3	2015-10-23 08:47:58.964	31.0	No	JOY	JOY	DESPAIR	JOY	DESPAIR	DESPAIR	JOY	...	DESPAIR	NaN	N
4	2015-10-23 08:48:11.719	30.0	No	NaN	JOY	DESPAIR	JOY	NaN	NaN	NaN	...	NaN	NaN	N

5 rows × 124 columns

In [278…
```python
#Remove blank row
candy17 = candy17.drop(index=candy17.index[0])
candy17.head(2)
```

Out[278]:

	Internal ID	Q1: GOING OUT?	Q2: GENDER	Q3: AGE	Q4: COUNTRY	Q5: STATE, PROVINCE, COUNTY, ETC	Q6 | 100 Grand Bar	Q6 | Anonymous brown globs that come in black and orange wrappers\t(a.k.a. Mary Janes)	Q6 | Any full-sized candy bar	Q6 | Black Jacks	...	Q8: DESPAIR OTHER	Q9: OTHER COMMENTS	Q10: DRESS
1	90272821	No	Male	44	USA	NM	MEH	DESPAIR	JOY	MEH	...	NaN	Bottom line is Twix is really the only candy w...	White and gold
2	90272829	NaN	Male	49	USA	Virginia	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN

2 rows × 120 columns

In [279...

```
#Check dataframe for null values
find_nulls = candy17.isnull().sum()
for key,value in find_nulls.iteritems():
    print(key,",",value)
```

```
Internal ID , 0
Q1: GOING OUT? , 109
Q2: GENDER , 40
Q3: AGE , 83
Q4: COUNTRY , 63
Q5: STATE, PROVINCE, COUNTY, ETC , 99
Q6 | 100 Grand Bar , 746
Q6 | Anonymous brown globs that come in black and orange wrappers        (a.k.a. Mary Janes) , 733
Q6 | Any full-sized candy bar , 671
Q6 | Black Jacks , 957
Q6 | Bonkers (the candy) , 991
Q6 | Bonkers (the board game) , 1005
Q6 | Bottle Caps , 764
Q6 | Box'o'Raisins , 687
Q6 | Broken glow stick , 705
Q6 | Butterfinger , 681
Q6 | Cadbury Creme Eggs , 682
Q6 | Candy Corn , 677
Q6 | Candy that is clearly just the stuff given out for free at restaurants , 690
Q6 | Caramellos , 751
Q6 | Cash, or other forms of legal tender , 679
Q6 | Chardonnay , 742
Q6 | Chick-o-Sticks (we donÕt know what that is) , 945
Q6 | Chiclets , 710
Q6 | Coffee Crisp , 852
Q6 | Creepy Religious comics/Chick Tracts , 703
Q6 | Dental paraphenalia , 691
Q6 | Dots , 728
Q6 | Dove Bars , 701
Q6 | Fuzzy Peaches , 822
Q6 | Generic Brand Acetaminophen , 730
Q6 | Glow sticks , 696
Q6 | Goo Goo Clusters , 878
Q6 | Good N' Plenty , 733
Q6 | Gum from baseball cards , 715
Q6 | Gummy Bears straight up , 696
Q6 | Hard Candy , 694
Q6 | Healthy Fruit , 693
Q6 | Heath Bar , 711
Q6 | Hershey's Dark Chocolate , 672
Q6 | HersheyÕs Milk Chocolate , 671
Q6 | Hershey's Kisses , 677
Q6 | Hugs (actual physical hugs) , 712
Q6 | Jolly Rancher (bad flavor) , 693
```

```
Q6 | Jolly Ranchers (good flavor) , 694
Q6 | JoyJoy (Mit Iodine!) , 1025
Q6 | Junior Mints , 697
Q6 | Senior Mints , 941
Q6 | Kale smoothie , 743
Q6 | Kinder Happy Hippo , 944
Q6 | Kit Kat , 673
Q6 | LaffyTaffy , 735
Q6 | LemonHeads , 729
Q6 | Licorice (not black) , 685
Q6 | Licorice (yes black) , 684
Q6 | Lindt Truffle , 717
Q6 | Lollipops , 690
Q6 | Mars , 724
Q6 | Maynards , 1023
Q6 | Mike and Ike , 728
Q6 | Milk Duds , 692
Q6 | Milky Way , 687
Q6 | Regular M&Ms , 674
Q6 | Peanut M&MÕs , 670
Q6 | Blue M&M's , 726
Q6 | Red M&M's , 728
Q6 | Green Party M&M's , 763
Q6 | Independent M&M's , 812
Q6 | Abstained from M&M'ing. , 941
Q6 | Minibags of chips , 723
Q6 | Mint Kisses , 775
Q6 | Mint Juleps , 810
Q6 | Mr. Goodbar , 739
Q6 | Necco Wafers , 743
Q6 | Nerds , 722
Q6 | Nestle Crunch , 697
Q6 | Now'n'Laters , 816
Q6 | Peeps , 709
Q6 | Pencils , 708
Q6 | Pixy Stix , 721
Q6 | Real Housewives of Orange County Season 9 Blue-Ray , 752
Q6 | ReeseÕs Peanut Butter Cups , 678
Q6 | Reese's Pieces , 690
Q6 | Reggie Jackson Bar , 1013
Q6 | Rolos , 713
Q6 | Sandwich-sized bags filled with BooBerry Crunch , 775
Q6 | Skittles , 705
Q6 | Smarties (American) , 724
```

```
Q6 | Smarties (Commonwealth) , 900
Q6 | Snickers , 689
Q6 | Sourpatch Kids (i.e. abominations of nature) , 737
Q6 | Spotted Dick , 881
Q6 | Starburst , 692
Q6 | Sweet Tarts , 707
Q6 | Swedish Fish , 714
Q6 | Sweetums (a friend to diabetes) , 1001
Q6 | Take 5 , 916
Q6 | Tic Tacs , 713
Q6 | Those odd marshmallow circus peanut things , 735
Q6 | Three Musketeers , 707
Q6 | Tolberone something or other , 705
Q6 | Trail Mix , 707
Q6 | Twix , 689
Q6 | Vials of pure high fructose corn syrup, for main-lining into your vein , 791
Q6 | Vicodin , 788
Q6 | Whatchamacallit Bars , 822
Q6 | White Bread , 756
Q6 | Whole Wheat anything , 746
Q6 | York Peppermint Patties , 704
Q7: JOY OTHER , 1548
Q8: DESPAIR OTHER , 1742
Q9: OTHER COMMENTS , 2073
Q10: DRESS , 743
Unnamed: 113 , 2450
Q11: DAY , 724
Q12: MEDIA [Daily Dish] , 2374
Q12: MEDIA [Science] , 1097
Q12: MEDIA [ESPN] , 2360
Q12: MEDIA [Yahoo] , 2392
Click Coordinates (x, y) , 854
```

In [280… `candy17.shape`

Out[280]: `(2459, 120)`

1. Filter out Missing Data

```
In [281...   #Create copy of dataframe for modifications
             candy17_copy = candy17.copy()
             #Remove data with null values
             candy17_copy = candy17_copy.dropna()
             candy17_copy.isnull().sum()
```

```
Out[281]:   Internal ID                 0.0
            Q1: GOING OUT?              0.0
            Q2: GENDER                  0.0
            Q3: AGE                     0.0
            Q4: COUNTRY                 0.0
                                        ...
            Q12: MEDIA [Daily Dish]     0.0
            Q12: MEDIA [Science]        0.0
            Q12: MEDIA [ESPN]           0.0
            Q12: MEDIA [Yahoo]          0.0
            Click Coordinates (x, y)    0.0
            Length: 120, dtype: float64
```

```
In [282...   candy17_copy.shape
```

```
Out[282]:   (0, 120)
```

As can be seen above, every row had at least one missing value resulting in the removal of all data. This is not ideal and would not be the preferred method when working with this data.

1.5 Fill in missing data

```
In [283...   #Replace null values
             candy17 = candy17.fillna("No Response")
             candy17.head()
```

	Internal ID	Q1: GOING OUT?	Q2: GENDER	Q3: AGE	Q4: COUNTRY	Q5: STATE, PROVINCE, COUNTY, ETC	Q6 | 100 Grand Bar	Q6 | Anonymous brown globs that come in black and orange wrappers\t(a.k.a. Mary Janes)	Q6 | Any full-sized candy bar	Q6 | Black Jacks	...	Q8: DESPAIR OTHER	Q9: OT COMMI
1	90272821	No	Male	44	USA	NM	MEH	DESPAIR	JOY	MEH	...	No Response	Botton is Tu reall only c
2	90272829	No Response	Male	49	USA	Virginia	No Response	No Response	No Response	No Response	...	No Response	Resp
3	90272840	No	Male	40	us	or	MEH	DESPAIR	JOY	MEH	...	No Response	Raisin go to
4	90272841	No	Male	23	usa	exton pa	JOY	DESPAIR	JOY	DESPAIR	...	No Response	Resp
5	90272852	No	Male	No Response	No Response	No Response	JOY	DESPAIR	JOY	No Response	...	No Response	Resp

5 rows × 120 columns

2. Replace Values

```python
#Replace united states country value
candy17['Q4: COUNTRY'] = candy17['Q4: COUNTRY'].replace({'United States' : 'USA', 'Usa':'USA','US':'USA',
                                                          'usa':'USA','us':'USA','Murica':'USA'})
candy17.head(10)
```

`Out[284]:`

	Internal ID	Q1: GOING OUT?	Q2: GENDER	Q3: AGE	Q4: COUNTRY	Q5: STATE, PROVINCE, COUNTY, ETC	Q6 | 100 Grand Bar	Q6 | Anonymous brown globs that come in black and orange wrappers\t(a.k.a. Mary Janes)	Q6 | Any full-sized candy bar	Q6 | Black Jacks	...	Q8: DESPAIR OTHER	Q9: COMM
1	90272821	No	Male	44	USA	NM	MEH	DESPAIR	JOY	MEH	...	No Response	Botto is rea only
2	90272829	No Response	Male	49	USA	Virginia	No Response	No Response	No Response	No Response	...	No Response	Res
3	90272840	No	Male	40	USA	or	MEH	DESPAIR	JOY	MEH	...	No Response	Raisi go
4	90272841	No	Male	23	USA	exton pa	JOY	DESPAIR	JOY	DESPAIR	...	No Response	Res
5	90272852	No	Male	No Response	No Response	No Response	JOY	DESPAIR	JOY	No Response	...	No Response	Res
6	90272853	No	Male	53	USA	Colorado	No Response	No Response	No Response	No Response	...	No Response	Res
7	90272854	No	Male	33	canada	ontario	JOY	DESPAIR	JOY	DESPAIR	...	when the little kids get the big chocolate bar...	first to gi cans o
8	90272858	No	Male	40	Canada	Ontario	JOY	DESPAIR	JOY	MEH	...	No Response	Res
9	90272859	No	Female	53	Us	Wa	MEH	DESPAIR	JOY	MEH	...	No Response	Res
10	90272861	Yes	Male	43	No Response	No Response	No Response	No Response	No Response	No Response	...	No Response	Res

10 rows × 120 columns

3. Create hierarchical index

```
In [285…    #Set hierarchical index
            candy17_hier = candy17.set_index(['Internal ID','Q2: GENDER'])
            candy17_hier
```

`Out[285]:`

Internal ID	Q2: GENDER	Q1: GOING OUT?	Q3: AGE	Q4: COUNTRY	Q5: STATE, PROVINCE, COUNTY, ETC	Q6 | 100 Grand Bar	Q6 | Anonymous brown globs that come in black and orange wrappers\t(a.k.a. Mary Janes)	Q6 | Any full-sized candy bar	Q6 | Black Jacks	Q6 | Bonkers (the candy)	Q6 | Bonkers (the board game)	...
90272821	Male	No	44	USA	NM	MEH	DESPAIR	JOY	MEH	DESPAIR	DESPAIR	...
90272829	Male	No Response	49	USA	Virginia	No Response	No Response	No Response	No Response	No Response	No Response	...
90272840	Male	No	40	USA	or	MEH	DESPAIR	JOY	MEH	MEH	DESPAIR	...
90272841	Male	No	23	USA	exton pa	JOY	DESPAIR	JOY	DESPAIR	MEH	DESPAIR	...
90272852	Male	No	No Response	No Response	No Response	JOY	DESPAIR	JOY	No Response	No Response	No Response	...
...	...	...	...	...	...	...	...	...	...	...	...	...
90314359	Male	No	24	USA	MD	JOY	DESPAIR	MEH	DESPAIR	DESPAIR	MEH	...
90314580	Female	No	33	USA	New York	MEH	DESPAIR	JOY	No Response	No Response	No Response	...
90314634	Female	No	26	USA	Tennessee	MEH	DESPAIR	JOY	DESPAIR	MEH	JOY	...
90314658	Male	No	58	USA	North Carolina	No Response	No Response	No Response	No Response	No Response	No Response	...
90314802	Female	No	66	USA	Pennsylvania	DESPAIR	DESPAIR	JOY	DESPAIR	DESPAIR	DESPAIR	...

Internal ID	Q2: GENDER	Q1: GOING OUT?	Q3: AGE	Q4: COUNTRY	Q5: STATE, PROVINCE, COUNTY, ETC	Q6 | 100 Grand Bar	Q6 | Anonymous brown globs that come in black and orange wrappers\t(a.k.a. Mary Janes)	Q6 | Any full-sized candy bar	Q6 | Black Jacks	Q6 | Bonkers (the candy)	Q6 | Bonkers (the board game)	...

2459 rows × 118 columns

4. Reshape

```
In [286…  #Pivot columns into rows with stack
          stacked = candy17_hier.stack()
          stacked
```

```
Out[286]:  Internal ID  Q2: GENDER
           90272821     Male         Q1: GOING OUT?                          No
                                     Q3: AGE                                 44
                                     Q4: COUNTRY                            USA
                                     Q5: STATE, PROVINCE, COUNTY, ETC        NM
                                     Q6 | 100 Grand Bar                     MEH
                                                                          ...
           90314802     Female       Q12: MEDIA [Daily Dish]               1.0
                                     Q12: MEDIA [Science]          No Response
                                     Q12: MEDIA [ESPN]             No Response
                                     Q12: MEDIA [Yahoo]            No Response
                                     Click Coordinates (x, y)         (19, 26)
           Length: 290162, dtype: object
```

```
In [287…  #Return data to original state
          stacked.unstack()
```

Out[287]:

Internal ID	Q2: GENDER	Q1: GOING OUT?	Q3: AGE	Q4: COUNTRY	Q5: STATE, PROVINCE, COUNTY, ETC	Q6 | 100 Grand Bar	Q6 | Anonymous brown globs that come in black and orange wrappers\t(a.k.a. Mary Janes)	Q6 | Any full-sized candy bar	Q6 | Black Jacks	Q6 | Bonkers (the candy)	Q6 | Bonkers (the board game)	...
90272821	Male	No	44	USA	NM	MEH	DESPAIR	JOY	MEH	DESPAIR	DESPAIR	...
90272829	Male	No Response	49	USA	Virginia	No Response	No Response	No Response	No Response	No Response	No Response	...
90272840	Male	No	40	USA	or	MEH	DESPAIR	JOY	MEH	MEH	DESPAIR	...
90272841	Male	No	23	USA	exton pa	JOY	DESPAIR	JOY	DESPAIR	MEH	DESPAIR	...
90272852	Male	No	No Response	No Response	No Response	JOY	DESPAIR	JOY	No Response	No Response	No Response	...
...	...	...	...	...	...	...	...	...	...	...	...	...
90314359	Male	No	24	USA	MD	JOY	DESPAIR	MEH	DESPAIR	DESPAIR	MEH	...
90314580	Female	No	33	USA	New York	MEH	DESPAIR	JOY	No Response	No Response	No Response	...
90314634	Female	No	26	USA	Tennessee	MEH	DESPAIR	JOY	DESPAIR	MEH	JOY	...
90314658	Male	No	58	USA	North Carolina	No Response	No Response	No Response	No Response	No Response	No Response	...
90314802	Female	No	66	USA	Pennsylvania	DESPAIR	DESPAIR	JOY	DESPAIR	DESPAIR	DESPAIR	...

	Q1: GOING OUT?	Q3: AGE	Q4: COUNTRY	Q5: STATE, PROVINCE, COUNTY, ETC	Q6 | 100 Grand Bar	Q6 | Anonymous brown globs that come in black and orange wrappers\t(a.k.a. Mary Janes)	Q6 | Any full-sized candy bar	Q6 | Black Jacks	Q6 | Bonkers (the candy)	Q6 | Bonkers (the board game)	...
Internal ID	Q2: GENDER											

2459 rows × 118 columns

5. Grouping with Index Levels

```
In [288…   #Group hierarchical indexed df by gender responses
           candy17_hier.groupby(level='Q2: GENDER').count()
```

Q2: GENDER	Q1: GOING OUT?	Q3: AGE	Q4: COUNTRY	Q5: STATE, PROVINCE, COUNTY, ETC	Q6 | 100 Grand Bar	Q6 | Anonymous brown globs that come in black and orange wrappers\t(a.k.a. Mary Janes)	Q6 | Any full-sized candy bar	Q6 | Black Jacks	Q6 | Bonkers (the candy)	Q6 | Bonkers (the board game)	...	Q8: DESPAIR OTHER	Q9: OTHER COMMENTS	Q9 DF
Female	839	839	839	839	839	839	839	839	839	839	...	839	839	
I'd rather not say	83	83	83	83	83	83	83	83	83	83	...	83	83	
Male	1467	1467	1467	1467	1467	1467	1467	1467	1467	1467	...	1467	1467	1
No Response	40	40	40	40	40	40	40	40	40	40	...	40	40	
Other	30	30	30	30	30	30	30	30	30	30	...	30	30	

5 rows × 118 columns

6. Split/Apply/Combine

```python
#Replace non-numeric values with Nan
candy17['Q3: AGE'] = candy17['Q3: AGE'].apply(lambda x: pd.to_numeric(x, errors='coerce') )
#Replace null values with zero
candy17['Q3: AGE'] = candy17['Q3: AGE'].fillna(0)
#Check for remaining null values
candy17['Q3: AGE'].isnull().sum()
```

0

```python
#Create function to find 10 oldest respondents
def you_are_old (candy17, n=10, column='Q3: AGE'):
    return candy17.sort_values(by=column)[-n:]

candy17.groupby('Q2: GENDER').apply(you_are_old)
```

Out[290]:

Q2: GENDER		Internal ID	Q1: GOING OUT?	Q2: GENDER	Q3: AGE	Q4: COUNTRY	Q5: STATE, PROVINCE, COUNTY, ETC	Q6 | 100 Grand Bar	Q6 | Anonymous brown globs that come in black and orange wrappers\t(a.k.a. Mary Janes)	Q6 | Any full-sized candy bar	Q6 | Black Jacks	...	
Female	438	90274015	Yes	Female	68.0	USA	California	DESPAIR	DESPAIR	MEH	DESPAIR	...	F
	504	90274333	No Response	Female	70.0	USA	Ohio	No Response	No Response	No Response	No Response	...	F
	304	90273584	Yes	Female	70.0	USA	ca	No Response	No Response	No Response	No Response	...	F
	1109	90277292	No	Female	70.0	USA	Ny	No Response	No Response	No Response	No Response	...	F
	645	90275325	Yes	Female	71.0	USA	ohio	JOY	JOY	JOY	MEH	...	F
	445	90274035	No	Female	71.0	USA	Arizona	MEH	DESPAIR	JOY	DESPAIR	...	Hc
	1557	90280519	No	Female	71.0	USA	California, Ventura County	No Response	No Response	No Response	No Response	...	F
	2207	90288438	No	Female	72.0	USA	Massachusetts	No Response	No Response	No Response	No Response	...	F
	2148	90287209	No	Female	73.0	USA	VA	No Response	No Response	No Response	No Response	...	F
	2300	90292039	No	Female	73.0	USA	Northampton County, PA	No Response	JOY	JOY	No Response	...	Tr hi

Q2: GENDER		Internal ID	Q1: GOING OUT?	Q2: GENDER	Q3: AGE	Q4: COUNTRY	Q5: STATE, PROVINCE, COUNTY, ETC	Q6 | 100 Grand Bar	Q6 | Anonymous brown globs that come in black and orange wrappers\t(a.k.a. Mary Janes)	Q6 | Any full-sized candy bar	Q6 | Black Jacks	...	
I'd rather not say	980												Tu S
		90276854	No	I'd rather not say	60.0	USA	California	JOY	DESPAIR	JOY	DESPAIR	...	
	711	90275719	No	I'd rather not say	67.0	usas	oakland,ca	No Response	No Response	No Response	No Response	...	F
	454	90274065	Yes	I'd rather not say	71.0	USA	CA	JOY	DESPAIR	DESPAIR	MEH	...	Jaw
	728	90275789	No	I'd rather not say	88.0	USA	Maryland	JOY	DESPAIR	JOY	MEH	...	F
	451	90274056	No	I'd rather not say	90.0	Earth	Sub Earth	MEH	DESPAIR	JOY	MEH	...	3D t
	575	90274994	No	I'd rather not say	99.0	USA	NY	MEH	DESPAIR	JOY	DESPAIR	...	F
	1495	90279712	No	I'd rather not say	100.0	USA	New york	MEH	DESPAIR	JOY	MEH	...	F
	1441	90279273	No	I'd rather not say	100.0	No Response	No Response	No Response	No Response	No Response	No Response	...	F
	884	90276469	No	I'd rather not say	102.0	A	A	MEH	DESPAIR	JOY	DESPAIR	...	

Q2: GENDER		Internal ID	Q1: GOING OUT?	Q2: GENDER	Q3: AGE	Q4: COUNTRY	Q5: STATE, PROVINCE, COUNTY, ETC	Q6 | 100 Grand Bar	Q6 | Anonymous brown globs that come in black and orange wrappers\t(a.k.a. Mary Janes)	Q6 | Any full-sized candy bar	Q6 | Black Jacks	...	
	1502	90279807	Yes	I'd rather not say	1000.0	Atlantis	Emerald City, Petrolia	MEH	JOY	JOY	MEH	...	
Male	**2312**	90292940	No	Male	73.0	USA	Pa. Northampton county	No Response	No Response	No Response	No Response	...	
	2390	90302627	No	Male	73.0	USA	pennsylvania, northampton couny, easton	JOY	JOY	No Response	MEH	...	wit
	781	90276043	No	Male	74.0	USA	Oakland County, Michigan	DESPAIR	DESPAIR	MEH	MEH	...	cc
	2010	90284966	No	Male	75.0	U.S.	Va	DESPAIR	DESPAIR	MEH	DESPAIR	...	
	1223	90277746	No	Male	75.0	USA	CA	MEH	MEH	JOY	JOY	...	
	987	90276872	No	Male	75.0	USA	Cook	MEH	DESPAIR	MEH	MEH	...	
	2329	90293566	No	Male	76.0	USA	California	No Response	JOY	MEH	JOY	...	
	1488	90279632	No	Male	76.0	USA	mo	MEH	No Response	JOY	MEH	...	
	1830	90282301	No	Male	77.0	US of A	North Carolina	No Response	DESPAIR	JOY	DESPAIR	...	

		Internal ID	Q1: GOING OUT?	Q2: GENDER	Q3: AGE	Q4: COUNTRY	Q5: STATE, PROVINCE, COUNTY, ETC	Q6 | 100 Grand Bar	Q6 | Anonymous brown globs that come in black and orange wrappers\t(a.k.a. Mary Janes)	Q6 | Any full-sized candy bar	Q6 | Black Jacks	...	
Q2: GENDER													
	1411	90279092	Yes	Male	90.0	America	Arkansas	JOY	MEH	JOY	MEH	...	
No Response	**838**	90276266	No Response	No Response	0.0	No Response	No Response	No Response	No Response	No Response	No Response	...	
	845	90276295	No	No Response	0.0	No Response	No Response	No Response	No Response	No Response	No Response	...	
	708	90275703	Yes	No Response	0.0	No Response	No Response	No Response	No Response	No Response	No Response	...	
	540	90274509	No	No Response	30.0	USA	ma	No Response	No Response	No Response	No Response	...	
	280	90273520	No	No Response	31.0	USA	No Response	No Response	No Response	No Response	No Response	...	
	487	90274247	No	No Response	37.0	USA	Ohio	No Response	DESPAIR	JOY	DESPAIR	...	
	2327	90293388	No	No Response	45.0	America	Alabama	MEH	DESPAIR	MEH	MEH	...	
	1877	90282868	No	No Response	49.0	Canada	Nova Scotia	No Response	DESPAIR	JOY	No Response	...	
	532	90274468	No Response	No Response	65.0	USA	OR	No Response	No Response	No Response	No Response	...	
	1784	90282015	No	No Response	70.0	USA	North carolina	MEH	DESPAIR	JOY	MEH	...	

	Internal ID	Q1: GOING OUT?	Q2: GENDER	Q3: AGE	Q4: COUNTRY	Q5: STATE, PROVINCE, COUNTY, ETC	Q6 | 100 Grand Bar	Q6 | Anonymous brown globs that come in black and orange wrappers\t(a.k.a. Mary Janes)	Q6 | Any full-sized candy bar	Q6 | Black Jacks	...	
Q2: GENDER												
Other **353**	90273704	No	Other	38.0	USA	Illinois	JOY	DESPAIR	JOY	DESPAIR	...	
1553	90280480	No	Other	39.0	USA	Connecticut	MEH	DESPAIR	MEH	DESPAIR	...	
879	90276452	Yes	Other	40.0	USA	Massachusetts	JOY	DESPAIR	JOY	No Response	...	
560	90274604	No	Other	40.0	USA	Maine	MEH	JOY	JOY	DESPAIR	...	
1122	90277339	No	Other	42.0	USA	Brooklyn, kings county ny	No Response	JOY	JOY	No Response	...	
1989	90284627	No	Other	43.0	united states	Wisconsin	MEH	JOY	JOY	MEH	...	
548	90274538	No	Other	45.0	USA	Maryland	JOY	DESPAIR	JOY	DESPAIR	...	
430	90273988	No	Other	50.0	Canada	Ontario	MEH	JOY	JOY	DESPAIR	...	

	Internal ID	Q1: GOING OUT?	Q2: GENDER	Q3: AGE	Q4: COUNTRY	Q5: STATE, PROVINCE, COUNTY, ETC	Q6 | 100 Grand Bar	Q6 | Anonymous brown globs that come in black and orange wrappers\t(a.k.a. Mary Janes)	Q6 | Any full-sized candy bar	Q6 | Black Jacks	...	
Q2: GENDER												
1173	90277544	No	Other	50.0	USA	California	No Response	No Response	No Response	No Response	...	
506	90274337	No	Other	312.0	Canada	BC	MEH	MEH	JOY	DESPAIR	...	pil rub o pur

50 rows × 120 columns

6.5 Remove obnoxious age reponses I just noticed (Outliers)

In [291...
```python
#Really? You're 1000 years old? You're a highlander?
candy17['Q3: AGE'].nlargest(n=10)
```

Out[291]:
```
1502    1000.0
506      312.0
884      102.0
1441     100.0
1495     100.0
575       99.0
451       90.0
1411      90.0
728       88.0
1830      77.0
Name: Q3: AGE, dtype: float64
```

In [292...
```python
#Remove highlanders' responses
candy17 = candy17[candy17['Q3: AGE']<120]
candy17['Q3: AGE'].nlargest(n=10)
```

```
Out[292]:  884      102.0
           1441     100.0
           1495     100.0
           575       99.0
           451       90.0
           1411      90.0
           728       88.0
           1830      77.0
           1488      76.0
           2329      76.0
           Name: Q3: AGE, dtype: float64
```

7. Convert between string and date time

```
In [293…  candy15.dtypes
```

```
Out[293]:  Timestamp                                                                      datetime64[ns]
           How old are you?                                                                       object
           Are you going actually going trick or treating yourself?                               object
            [Butterfinger]                                                                         object
            [100 Grand Bar]                                                                        object
                                                                                                      ...
           Please estimate the degrees of separation you have from the following folks [Thom Yorke]        float64
           Please estimate the degrees of separation you have from the following folks [JJ Abrams]          float64
           Please estimate the degrees of separation you have from the following folks [Hillary Clinton]    float64
           Please estimate the degrees of separation you have from the following folks [Donald Trump]       float64
           Please estimate the degrees of separation you have from the following folks [Beyoncé Knowles]    float64
           Length: 124, dtype: object
```

```
In [294…  #Convert Timestamp from datetime to string to so I can convert to datetime
          candy15['Timestamp'] = candy15['Timestamp'].astype(str)
          candy15.dtypes
```

```
Out[294]:  Timestamp                                                                              object
           How old are you?                                                                        object
           Are you going actually going trick or treating yourself?                                object
            [Butterfinger]                                                                         object
            [100 Grand Bar]                                                                        object
                                                                                                   ...
           Please estimate the degrees of separation you have from the following folks [Thom Yorke]        float64
           Please estimate the degrees of separation you have from the following folks [JJ Abrams]         float64
           Please estimate the degrees of separation you have from the following folks [Hillary Clinton]   float64
           Please estimate the degrees of separation you have from the following folks [Donald Trump]      float64
           Please estimate the degrees of separation you have from the following folks [Beyoncé Knowles]   float64
           Length: 124, dtype: object
```

```python
In [295…   #Convert Timestamp from string to datetime
           candy15['Timestamp'] = pd.to_datetime(candy15['Timestamp'])
           candy15.dtypes
```

```
Out[295]:  Timestamp                                                                              datetime64[ns]
           How old are you?                                                                        object
           Are you going actually going trick or treating yourself?                                object
            [Butterfinger]                                                                         object
            [100 Grand Bar]                                                                        object
                                                                                                   ...
           Please estimate the degrees of separation you have from the following folks [Thom Yorke]        float64
           Please estimate the degrees of separation you have from the following folks [JJ Abrams]         float64
           Please estimate the degrees of separation you have from the following folks [Hillary Clinton]   float64
           Please estimate the degrees of separation you have from the following folks [Donald Trump]      float64
           Please estimate the degrees of separation you have from the following folks [Beyoncé Knowles]   float64
           Length: 124, dtype: object
```

7.5 Split datetime into Date and Time columns for fun

```python
In [296…   candy15.insert(0,'Date',pd.to_datetime(candy15['Timestamp']).dt.date)
           candy15.insert(1,'Time',pd.to_datetime(candy15['Timestamp']).dt.time)
           candy15 = candy15.drop(columns=['Timestamp'])
```

```python
In [297…   candy15.head()
```

	Date	Time	How old are you?	Are you going actually going trick or treating yourself?	[Butterfinger]	[100 Grand Bar]	[Anonymous brown globs that come in black and orange wrappers]	[Any full-sized candy bar]	[Black Jacks]	[Bonkers]	...	[Necco Wafers]	Which day do you prefer, Friday or Sunday?
0	2015-10-23	08:46:20.451000	35.0	No	JOY	NaN	DESPAIR	JOY	NaN	NaN	...	NaN	NaN
1	2015-10-23	08:46:51.583000	41.0	No	JOY	JOY	DESPAIR	JOY	DESPAIR	DESPAIR	...	DESPAIR	NaN
2	2015-10-23	08:47:34.285000	33.0	No	DESPAIR	DESPAIR	DESPAIR	JOY	DESPAIR	DESPAIR	...	DESPAIR	NaN
3	2015-10-23	08:47:58.964000	31.0	No	JOY	JOY	DESPAIR	JOY	DESPAIR	DESPAIR	...	DESPAIR	NaN
4	2015-10-23	08:48:11.719000	30.0	No	NaN	JOY	DESPAIR	JOY	NaN	NaN	...	NaN	NaN

5 rows × 125 columns

8. Generate date range

```
#Create business week date range
candied_date_range = pd.date_range('2015-10-23', '2015-10-31',freq='B')
candied_date_range
```

Out[299]:
```
DatetimeIndex(['2015-10-23', '2015-10-26', '2015-10-27', '2015-10-28',
               '2015-10-29', '2015-10-30'],
              dtype='datetime64[ns]', freq='B')
```