Week 7 & 8

In [1]:
```
#Import required libraries
import pandas as pd
import numpy as np
import warnings
warnings.filterwarnings('ignore', category=UserWarning)
```

In [2]:
```
# Loading candy survey data for 2015 and 2016 in separate dataframes
candy_2015_df = pd.read_excel("CANDY-HIERARCHY-2015-SURVEY-Responses.xlsx", engine = 'op
print(candy_2015_df.shape)
candy_2015_df.head(5)
```

```
(5630, 124)
```

Out[2]:

	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]	...	V
0	2015-10-23 08:46:20.451	35	No	JOY	NaN	DESPAIR	JOY	NaN	NaN	NaN	...	
1	2015-10-23 08:46:51.583	41	No	JOY	JOY	DESPAIR	JOY	DESPAIR	DESPAIR	JOY	...	D
2	2015-10-23 08:47:34.285	33	No	DESPAIR	DESPAIR	DESPAIR	JOY	DESPAIR	DESPAIR	DESPAIR	...	D
3	2015-10-23 08:47:58.964	31	No	JOY	JOY	DESPAIR	JOY	DESPAIR	DESPAIR	JOY	...	D
4	2015-10-23 08:48:11.719	30	No	NaN	JOY	DESPAIR	JOY	NaN	NaN	NaN	...	

5 rows × 124 columns

In [3]:
```
candy_2016_df = pd.read_excel("BOING-BOING-CANDY-HIERARCHY-2016-SURVEY-Responses.xlsx")
print(candy_2016_df.shape)
candy_2016_df.head(5)
```

```
(1259, 123)
```

Out[3]:

	Timestamp	Are you going actually going trick or treating yourself?	Your gender:	How old are you?	Which country do you live in?	Which state, province, county do you live in?	[100 Grand Bar]	[Anonymous brown globs that come in black and orange wrappers]	[Any full-sized candy bar]	[Black Jacks]	...	Pleas estimat th degree(s c separatio you hav from th followin celebritie [J Rowling

0	2016-10-24 05:09:23.033	No	Male	22	Canada	Ontario	JOY	DESPAIR	JOY	MEH	...	3 or highe
1	2016-10-24 05:09:54.798	No	Male	45	usa	il	MEH	MEH	JOY	JOY	...	3 or highe
2	2016-10-24 05:13:06.734	No	Female	48	US	Colorado	JOY	DESPAIR	JOY	MEH	...	3 or highe
3	2016-10-24 05:14:17.192	No	Male	57	usa	il	JOY	MEH	JOY	MEH	...	3 or highe
4	2016-10-24 05:14:24.625	Yes	Male	42	USA	South Dakota	MEH	DESPAIR	JOY	DESPAIR	...	3 or highe

5 rows × 123 columns

Chapter 7

Filter out missing data

Dropping null rows from both datasets

In [4]:
```
print('Shape before dropping NA :', candy_2015_df.shape)
candy_2015_df.dropna(how='all', inplace=True)
print('Shape after dropping NA :',candy_2015_df.shape)
```

```
Shape before dropping NA : (5630, 124)
Shape after dropping NA : (5630, 124)
```

In [5]:
```
print('Shape before dropping NA :', candy_2016_df.shape)
candy_2016_df.dropna(how='all', inplace=True)
print('Shape after dropping NA :',candy_2016_df.shape)
```

```
Shape before dropping NA : (1259, 123)
Shape after dropping NA : (1259, 123)
```

Validating columns with all Null values

In [6]:
```
[i for i in candy_2015_df.columns if candy_2015_df[i].isnull().sum() >= len(candy_2015_d
```

Out[6]:
```
['Fill in the blank: "Taylor Swift is a force for _____"',
 'Please estimate the degrees of separation you have from the following folks [Bruce Le
e]',
 'Please estimate the degrees of separation you have from the following folks [JK Rowlin
g]',
 'Please estimate the degrees of separation you have from the following folks [Malala Yo
usafzai]',
 'Please estimate the degrees of separation you have from the following folks [Thom York
e]',
 'Please estimate the degrees of separation you have from the following folks [JJ Abram
s]',
 'Please estimate the degrees of separation you have from the following folks [Hillary C
linton]',
 'Please estimate the degrees of separation you have from the following folks [Donald Tr
ump]',
 'Please estimate the degrees of separation you have from the following folks [Beyoncé K
nowles]']
```

In [7]:
```
[i for i in candy_2016_df.columns if candy_2016_df[i].isnull().sum() >= len(candy_2016_d
```

Out[7]:
```
[' [York Peppermint Patties] Ignore']
```

Dropping null columns from both dataframes

In [8]:
```
#Dropping null columns from 2015 dataframe
print('Shape before dropping NA columns :', candy_2015_df.shape)
candy_2015_df.dropna(axis = 1, how='all', inplace=True)
print('Shape after dropping NA columns:',candy_2015_df.shape)
```

```
Shape before dropping NA columns : (5630, 124)
Shape after dropping NA columns: (5630, 115)
```

In [9]:
```
#Dropping null columns from 2016 dataframe
print('Shape before dropping NA columns :', candy_2016_df.shape)
candy_2016_df.dropna(axis = 1, how='all', inplace=True)
print('Shape after dropping NA columns:',candy_2016_df.shape)
```

```
Shape before dropping NA columns : (1259, 123)
Shape after dropping NA columns: (1259, 122)
```

One null column - "[York Peppermint Patties] Ignore" dropped in candy_2016_df and 9 columns dropped from candy_2015_df

Fill in missing data

Update missing values in numeric columns with the mean

In [10]:
```
#Identify numeric columns in dataframe candy_2015_df
num_columns = candy_2015_df.select_dtypes(include=np.number).columns.tolist()
num_columns
```

Out[10]:
```
[]
```

In [11]:
```
#Identify numeric columns in dataframe candy_2016_df
num_columns = candy_2016_df.select_dtypes(include=np.number).columns.tolist()
num_columns
```

Out[11]:
```
[]
```

There are no numeric columns with missing values

In [12]:
```
#Update the non-numeric columns (object) with the most frequent values - Using mode
```

In [13]:
```
#Get the number of empty rows in each non-numeric column of both dataframes
```

In [14]:
```
candy_2015_df.select_dtypes([object]).isnull().sum()
```

Out[14]:
```
How old are you?
                                199
Are you going actually going trick or treating yourself?
                                  0
 [Butterfinger]
                                383
 [100 Grand Bar]
                                656
 [Anonymous brown globs that come in black and orange wrappers]
                                298

                               ...
Please estimate the degree(s) of separation you have from the following celebrities [Kev
in Bacon]                       555
Please estimate the degree(s) of separation you have from the following celebrities [Fra
ncis Bacon (1561 - 1626)]       565
 [Sea-salt flavored stuff, probably chocolate, since this is the "it" flavor of the yea
r]                               580
 [Necco Wafers]
                                683
Which day do you prefer, Friday or Sunday?
                               1037
Length: 114, dtype: int64
```

In [15]:
```
candy_2016_df.select_dtypes([object]).isnull().sum()
```

Out[15]:
```
Are you going actually going trick or treating yourself?
                                                     0
Your gender:
                                                     9
How old are you?
                                                    33
Which country do you live in?
                                                    21
Which state, province, county do you live in?
                                                    79

                                                    ..
Please estimate the degree(s) of separation you have from the following celebrities [Kev
in Bacon]                                           80
Please estimate the degree(s) of separation you have from the following celebrities [Fra
ncis Bacon (1561 - 1626)]                           85
Which day do you prefer, Friday or Sunday?
                                                    24
Do you eat apples the correct way, East to West (side to side) or do you eat them like a
freak of nature, South to North (bottom to top)?    45
When you see the above image of the 4 different websites, which one would you most likel
y check out (please be honest).                     37
Length: 121, dtype: int64
```

In [16]:
```
# Get the mode of the cateorical columns and the number of null values in column (2015 d
[print(col,' - ', candy_2015_df[col].mode(), ' : ', candy_2015_df[col].isnull().sum())
 for col in candy_2015_df.columns if (isinstance(candy_2015_df[col], object) and candy_2
```

```
How old are you?  -  0    35
Name: How old are you?, dtype: object  :  199
 [Butterfinger]  -  0    JOY
Name:  [Butterfinger], dtype: object  :  383
 [100 Grand Bar]  -  0    JOY
```

Name: [100 Grand Bar], dtype: object : 656
[Anonymous brown globs that come in black and orange wrappers] - 0 DESPAIR
Name: [Anonymous brown globs that come in black and orange wrappers], dtype: object : 298
[Any full-sized candy bar] - 0 JOY
Name: [Any full-sized candy bar], dtype: object : 277
[Black Jacks] - 0 DESPAIR
Name: [Black Jacks], dtype: object : 991
[Bonkers] - 0 DESPAIR
Name: [Bonkers], dtype: object : 1182
[Bottle Caps] - 0 DESPAIR
Name: [Bottle Caps], dtype: object : 596
[Box'o' Raisins] - 0 DESPAIR
Name: [Box'o' Raisins], dtype: object : 404
[Brach products (not including candy corn)] - 0 DESPAIR
Name: [Brach products (not including candy corn)], dtype: object : 698
[Bubble Gum] - 0 DESPAIR
Name: [Bubble Gum], dtype: object : 459
[Cadbury Creme Eggs] - 0 JOY
Name: [Cadbury Creme Eggs], dtype: object : 378
[Candy Corn] - 0 DESPAIR
Name: [Candy Corn], dtype: object : 394
[Vials of pure high fructose corn syrup, for main-lining into your vein] - 0 DESPAIR
Name: [Vials of pure high fructose corn syrup, for main-lining into your vein], dtype: object : 561
[Candy that is clearly just the stuff given out for free at restaurants] - 0 DESPAIR
Name: [Candy that is clearly just the stuff given out for free at restaurants], dtype: object : 366
[Cash, or other forms of legal tender] - 0 JOY
Name: [Cash, or other forms of legal tender], dtype: object : 303
[Chiclets] - 0 DESPAIR
Name: [Chiclets], dtype: object : 508
[Caramellos] - 0 JOY
Name: [Caramellos], dtype: object : 597
[Snickers] - 0 JOY
Name: [Snickers], dtype: object : 290
[Dark Chocolate Hershey] - 0 JOY
Name: [Dark Chocolate Hershey], dtype: object : 370
[Dental paraphenalia] - 0 DESPAIR
Name: [Dental paraphenalia], dtype: object : 336
[Dots] - 0 DESPAIR
Name: [Dots], dtype: object : 493
[Fuzzy Peaches] - 0 DESPAIR
Name: [Fuzzy Peaches], dtype: object : 655
[Generic Brand Acetaminophen] - 0 DESPAIR
Name: [Generic Brand Acetaminophen], dtype: object : 532
[Glow sticks] - 0 JOY
Name: [Glow sticks], dtype: object : 460
[Broken glow stick] - 0 DESPAIR
Name: [Broken glow stick], dtype: object : 350
[Goo Goo Clusters] - 0 DESPAIR
Name: [Goo Goo Clusters], dtype: object : 900
[Good N' Plenty] - 0 DESPAIR
Name: [Good N' Plenty], dtype: object : 583
[Gum from baseball cards] - 0 DESPAIR
Name: [Gum from baseball cards], dtype: object : 444
[Gummy Bears straight up] - 0 JOY
Name: [Gummy Bears straight up], dtype: object : 430
[Creepy Religious comics/Chick Tracts] - 0 DESPAIR
Name: [Creepy Religious comics/Chick Tracts], dtype: object : 335
[Healthy Fruit] - 0 DESPAIR
Name: [Healthy Fruit], dtype: object : 500
[Heath Bar] - 0 JOY
Name: [Heath Bar], dtype: object : 478

[Hershey’s Kissables] - 0 JOY
Name: [Hershey’s Kissables], dtype: object : 628
[Hershey’s Milk Chocolate] - 0 JOY
Name: [Hershey’s Milk Chocolate], dtype: object : 479
[Hugs (actual physical hugs)] - 0 DESPAIR
Name: [Hugs (actual physical hugs)], dtype: object : 503
[Jolly Rancher (bad flavor)] - 0 DESPAIR
Name: [Jolly Rancher (bad flavor)], dtype: object : 434
[Jolly Ranchers (good flavor)] - 0 JOY
Name: [Jolly Ranchers (good flavor)], dtype: object : 403
[Kale smoothie] - 0 DESPAIR
Name: [Kale smoothie], dtype: object : 427
[Kinder Happy Hippo] - 0 DESPAIR
Name: [Kinder Happy Hippo], dtype: object : 917
[Kit Kat] - 0 JOY
Name: [Kit Kat], dtype: object : 290
[Hard Candy] - 0 DESPAIR
Name: [Hard Candy], dtype: object : 646
[Lapel Pins] - 0 DESPAIR
Name: [Lapel Pins], dtype: object : 625
[LemonHeads] - 0 JOY
Name: [LemonHeads], dtype: object : 540
[Licorice] - 0 DESPAIR
Name: [Licorice], dtype: object : 400
[Licorice (not black)] - 0 DESPAIR
Name: [Licorice (not black)], dtype: object : 509
[Lindt Truffle] - 0 JOY
Name: [Lindt Truffle], dtype: object : 435
[Lollipops] - 0 DESPAIR
Name: [Lollipops], dtype: object : 616
[Mars] - 0 JOY
Name: [Mars], dtype: object : 614
[Mary Janes] - 0 DESPAIR
Name: [Mary Janes], dtype: object : 952
[Maynards] - 0 DESPAIR
Name: [Maynards], dtype: object : 1214
[Milk Duds] - 0 JOY
Name: [Milk Duds], dtype: object : 504
[LaffyTaffy] - 0 DESPAIR
Name: [LaffyTaffy], dtype: object : 608
[Minibags of chips] - 0 DESPAIR
Name: [Minibags of chips], dtype: object : 647
[JoyJoy (Mit Iodine)] - 0 DESPAIR
Name: [JoyJoy (Mit Iodine)], dtype: object : 1158
[Reggie Jackson Bar] - 0 DESPAIR
Name: [Reggie Jackson Bar], dtype: object : 1210
[Pixy Stix] - 0 DESPAIR
Name: [Pixy Stix], dtype: object : 579
[Nerds] - 0 JOY
Name: [Nerds], dtype: object : 433
[Nestle Crunch] - 0 JOY
Name: [Nestle Crunch], dtype: object : 448
[Now'n'Laters] - 0 DESPAIR
Name: [Now'n'Laters], dtype: object : 767
[Pencils] - 0 DESPAIR
Name: [Pencils], dtype: object : 586
[Milky Way] - 0 JOY
Name: [Milky Way], dtype: object : 465
[Reese’s Peanut Butter Cups] - 0 JOY
Name: [Reese’s Peanut Butter Cups], dtype: object : 257
[Tolberone something or other] - 0 JOY
Name: [Tolberone something or other], dtype: object : 485
[Runts] - 0 DESPAIR
Name: [Runts], dtype: object : 683
[Junior Mints] - 0 JOY
Name: [Junior Mints], dtype: object : 442

[Senior Mints] - 0 DESPAIR
Name: [Senior Mints], dtype: object : 975
[Mint Kisses] - 0 DESPAIR
Name: [Mint Kisses], dtype: object : 837
[Mint Juleps] - 0 DESPAIR
Name: [Mint Juleps], dtype: object : 695
[Mint Leaves] - 0 DESPAIR
Name: [Mint Leaves], dtype: object : 839
[Peanut M&M's] - 0 JOY
Name: [Peanut M&M's], dtype: object : 329
[Regular M&Ms] - 0 JOY
Name: [Regular M&Ms], dtype: object : 398
[Mint M&Ms] - 0 DESPAIR
Name: [Mint M&Ms], dtype: object : 612
[Ribbon candy] - 0 DESPAIR
Name: [Ribbon candy], dtype: object : 776
[Rolos] - 0 JOY
Name: [Rolos], dtype: object : 471
[Skittles] - 0 JOY
Name: [Skittles], dtype: object : 454
[Smarties (American)] - 0 JOY
Name: [Smarties (American)], dtype: object : 472
[Smarties (Commonwealth)] - 0 DESPAIR
Name: [Smarties (Commonwealth)], dtype: object : 1001
[Chick-o-Sticks (we don't know what that is)] - 0 DESPAIR
Name: [Chick-o-Sticks (we don't know what that is)], dtype: object : 1085
[Spotted Dick] - 0 DESPAIR
Name: [Spotted Dick], dtype: object : 919
[Starburst] - 0 JOY
Name: [Starburst], dtype: object : 476
[Swedish Fish] - 0 JOY
Name: [Swedish Fish], dtype: object : 519
[Sweetums] - 0 DESPAIR
Name: [Sweetums], dtype: object : 1195
[Those odd marshmallow circus peanut things] - 0 DESPAIR
Name: [Those odd marshmallow circus peanut things], dtype: object : 520
[Three Musketeers] - 0 JOY
Name: [Three Musketeers], dtype: object : 499
[Peterson Brand Sidewalk Chalk] - 0 DESPAIR
Name: [Peterson Brand Sidewalk Chalk], dtype: object : 889
[Peanut Butter Bars] - 0 DESPAIR
Name: [Peanut Butter Bars], dtype: object : 714
[Peanut Butter Jars] - 0 DESPAIR
Name: [Peanut Butter Jars], dtype: object : 786
[Trail Mix] - 0 DESPAIR
Name: [Trail Mix], dtype: object : 687
[Twix] - 0 JOY
Name: [Twix], dtype: object : 369
[Vicodin] - 0 JOY
Name: [Vicodin], dtype: object : 638
[White Bread] - 0 DESPAIR
Name: [White Bread], dtype: object : 667
[Whole Wheat anything] - 0 DESPAIR
Name: [Whole Wheat anything], dtype: object : 683
[York Peppermint Patties] - 0 JOY
Name: [York Peppermint Patties], dtype: object : 464
Please leave any remarks or comments regarding your choices. - 0 Data is oppressio
n.
Name: Please leave any remarks or comments regarding your choices., dtype: object : 47
45
Please list any items not included above that give you JOY. - 0 Almond Joy
Name: Please list any items not included above that give you JOY., dtype: object : 319
6
Please list any items not included above that give you DESPAIR. - 0 Popcorn balls
Name: Please list any items not included above that give you DESPAIR., dtype: object :
3858

```
Guess the number of mints in my hand.  -  0    5
Name: Guess the number of mints in my hand., dtype: object  :  478
Betty or Veronica?  -  0    Veronica
Name: Betty or Veronica?, dtype: object  :  294
Check all that apply: "I cried tears of sadness at the end of  _____________"  -  0    Th
e last Pixar movie I saw
Name: Check all that apply: "I cried tears of sadness at the end of  _____________", dtyp
e: object  :  928
"That dress* that went viral early this year - when I first saw it, it was ________"  -
0    Blue and black
Name: "That dress* that went viral early this year - when I first saw it, it was _______
_", dtype: object  :  520
What is your favourite font?  -  0    Helvetica
Name: What is your favourite font?, dtype: object  :  1076
If you squint really hard, the words "Intelligent Design" would look like.  -  0    "Bul
lsh*t!"
Name: If you squint really hard, the words "Intelligent Design" would look like., dtype:
object  :  632
Fill in the blank: "Imitation is a form of _____________"  -  0    Lazily flattering some
one
Name: Fill in the blank: "Imitation is a form of _____________", dtype: object  :  409
Please estimate the degree(s) of separation you have from the following celebrities [JK
Rowling]  -  0    3 or higher
Name: Please estimate the degree(s) of separation you have from the following celebritie
s [JK Rowling], dtype: object  :  551
Please estimate the degree(s) of separation you have from the following celebrities [JJ
Abrams]  -  0    3 or higher
Name: Please estimate the degree(s) of separation you have from the following celebritie
s [JJ Abrams], dtype: object  :  557
Please estimate the degree(s) of separation you have from the following celebrities [Bey
oncé]  -  0    3 or higher
Name: Please estimate the degree(s) of separation you have from the following celebritie
s [Beyoncé], dtype: object  :  562
Please estimate the degree(s) of separation you have from the following celebrities [Bie
ber]  -  0    3 or higher
Name: Please estimate the degree(s) of separation you have from the following celebritie
s [Bieber], dtype: object  :  584
Please estimate the degree(s) of separation you have from the following celebrities [Kev
in Bacon]  -  0    3 or higher
Name: Please estimate the degree(s) of separation you have from the following celebritie
s [Kevin Bacon], dtype: object  :  555
Please estimate the degree(s) of separation you have from the following celebrities [Fra
ncis Bacon (1561 - 1626)]  -  0    3 or higher
Name: Please estimate the degree(s) of separation you have from the following celebritie
s [Francis Bacon (1561 - 1626)], dtype: object  :  565
 [Sea-salt flavored stuff, probably chocolate, since this is the "it" flavor of the yea
r]  -  0    JOY
Name:  [Sea-salt flavored stuff, probably chocolate, since this is the "it" flavor of th
e year], dtype: object  :  580
 [Necco Wafers]  -  0    DESPAIR
Name:  [Necco Wafers], dtype: object  :  683
Which day do you prefer, Friday or Sunday?  -  0    Friday
Name: Which day do you prefer, Friday or Sunday?, dtype: object  :  1037
```

Out[16]:

```
[None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
```

```
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
            None,
```

```
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None]
```

In [17]:
```
#Update the null values with the mode  (2015 df)
candy_2015_df.fillna(candy_2015_df.select_dtypes([object]).mode().iloc[0],inplace=True)
```

In [18]:
```
# Rerun to make sure all null values are updated with the mode  (2015 df)
[print(col,' - ', candy_2015_df[col].mode(), ' : ', candy_2015_df[col].isnull().sum())
 for col in candy_2015_df.columns if (isinstance(candy_2015_df[col], object) and candy_2
```

Out[18]:
```
[]
```

In [19]:
```
# Get the mode of the cateorical columns and the number of null values in column (2016 d
[print(col,' - ', candy_2016_df[col].mode(), ' : ', candy_2016_df[col].isnull().sum())
 for col in candy_2016_df.columns if (isinstance(candy_2016_df[col], object) and candy_2
```

```
Your gender:  -  0    Male
Name: Your gender:, dtype: object  :  9
How old are you?  -  0    35
Name: How old are you?, dtype: object  :  33
Which country do you live in?  -  0    USA
Name: Which country do you live in?, dtype: object  :  21
Which state, province, county do you live in?  -  0    California
Name: Which state, province, county do you live in?, dtype: object  :  79
 [100 Grand Bar]  -  0    MEH
Name:  [100 Grand Bar], dtype: object  :  78
 [Anonymous brown globs that come in black and orange wrappers]  -  0    DESPAIR
Name:  [Anonymous brown globs that come in black and orange wrappers], dtype: object  :
26
 [Any full-sized candy bar]  -  0    JOY
Name:  [Any full-sized candy bar], dtype: object  :  17
 [Black Jacks]  -  0    DESPAIR
```

Name: [Black Jacks], dtype: object : 203
[Bonkers (the candy)] - 0 MEH
Name: [Bonkers (the candy)], dtype: object : 221
[Bonkers (the board game)] - 0 MEH
Name: [Bonkers (the board game)], dtype: object : 232
[Bottle Caps] - 0 MEH
Name: [Bottle Caps], dtype: object : 79
[Box'o'Raisins] - 0 DESPAIR
Name: [Box'o'Raisins], dtype: object : 29
[Broken glow stick] - 0 DESPAIR
Name: [Broken glow stick], dtype: object : 33
[Butterfinger] - 0 JOY
Name: [Butterfinger], dtype: object : 23
[Cadbury Creme Eggs] - 0 JOY
Name: [Cadbury Creme Eggs], dtype: object : 21
[Candy Corn] - 0 DESPAIR
Name: [Candy Corn], dtype: object : 17
[Candy that is clearly just the stuff given out for free at restaurants] - 0 DESPA
IR
Name: [Candy that is clearly just the stuff given out for free at restaurants], dtype:
object : 28
[Caramellos] - 0 JOY
Name: [Caramellos], dtype: object : 79
[Cash, or other forms of legal tender] - 0 JOY
Name: [Cash, or other forms of legal tender], dtype: object : 22
[Chardonnay] - 0 JOY
Name: [Chardonnay], dtype: object : 57
[Chick-o-Sticks (we don’t know what that is)] - 0 DESPAIR
Name: [Chick-o-Sticks (we don’t know what that is)], dtype: object : 211
[Chiclets] - 0 DESPAIR
Name: [Chiclets], dtype: object : 53
[Coffee Crisp] - 0 MEH
Name: [Coffee Crisp], dtype: object : 146
[Creepy Religious comics/Chick Tracts] - 0 DESPAIR
Name: [Creepy Religious comics/Chick Tracts], dtype: object : 46
[Dental paraphenalia] - 0 DESPAIR
Name: [Dental paraphenalia], dtype: object : 32
[Dots] - 0 DESPAIR
Name: [Dots], dtype: object : 62
[Dove Bars] - 0 JOY
Name: [Dove Bars], dtype: object : 45
[Fuzzy Peaches] - 0 DESPAIR
Name: [Fuzzy Peaches], dtype: object : 122
[Generic Brand Acetaminophen] - 0 DESPAIR
Name: [Generic Brand Acetaminophen], dtype: object : 68
[Glow sticks] - 0 JOY
Name: [Glow sticks], dtype: object : 36
[Goo Goo Clusters] - 0 MEH
Name: [Goo Goo Clusters], dtype: object : 176
[Good N' Plenty] - 0 DESPAIR
Name: [Good N' Plenty], dtype: object : 77
[Gum from baseball cards] - 0 DESPAIR
Name: [Gum from baseball cards], dtype: object : 46
[Gummy Bears straight up] - 0 JOY
Name: [Gummy Bears straight up], dtype: object : 35
[Hard Candy] - 0 MEH
Name: [Hard Candy], dtype: object : 31
[Healthy Fruit] - 0 DESPAIR
Name: [Healthy Fruit], dtype: object : 34
[Heath Bar] - 0 JOY
Name: [Heath Bar], dtype: object : 54
[Hershey's Dark Chocolate] - 0 JOY
Name: [Hershey's Dark Chocolate], dtype: object : 24
[Hershey’s Milk Chocolate] - 0 JOY
Name: [Hershey’s Milk Chocolate], dtype: object : 18
[Hershey's Kisses] - 0 JOY

```
Name:  [Hershey's Kisses], dtype: object  :  20
 [Hugs (actual physical hugs)]  -  0    DESPAIR
Name:  [Hugs (actual physical hugs)], dtype: object  :  37
 [Jolly Rancher (bad flavor)]  -  0    DESPAIR
Name:  [Jolly Rancher (bad flavor)], dtype: object  :  33
 [Jolly Ranchers (good flavor)]  -  0    JOY
Name:  [Jolly Ranchers (good flavor)], dtype: object  :  34
 [JoyJoy (Mit Iodine!)]  -  0    DESPAIR
Name:  [JoyJoy (Mit Iodine!)], dtype: object  :  243
 [Junior Mints]  -  0    JOY
Name:  [Junior Mints], dtype: object  :  41
 [Senior Mints]  -  0    DESPAIR
Name:  [Senior Mints], dtype: object  :  212
 [Kale smoothie]  -  0    DESPAIR
Name:  [Kale smoothie], dtype: object  :  76
 [Kinder Happy Hippo]  -  0    DESPAIR
Name:  [Kinder Happy Hippo], dtype: object  :  215
 [Kit Kat]  -  0    JOY
Name:  [Kit Kat], dtype: object  :  22
 [LaffyTaffy]  -  0    MEH
Name:  [LaffyTaffy], dtype: object  :  59
 [LemonHeads]  -  0    MEH
Name:  [LemonHeads], dtype: object  :  80
 [Licorice (not black)]  -  0    MEH
Name:  [Licorice (not black)], dtype: object  :  31
 [Licorice (yes black)]  -  0    DESPAIR
Name:  [Licorice (yes black)], dtype: object  :  28
 [Lindt Truffle]  -  0    JOY
Name:  [Lindt Truffle], dtype: object  :  51
 [Lollipops]  -  0    MEH
Name:  [Lollipops], dtype: object  :  37
 [Mars]  -  0    JOY
Name:  [Mars], dtype: object  :  50
 [Mary Janes]  -  0    MEH
Name:  [Mary Janes], dtype: object  :  194
 [Maynards]  -  0    MEH
Name:  [Maynards], dtype: object  :  262
 [Mike and Ike]  -  0    MEH
Name:  [Mike and Ike], dtype: object  :  65
 [Milk Duds]  -  0    JOY
Name:  [Milk Duds], dtype: object  :  50
 [Milky Way]  -  0    JOY
Name:  [Milky Way], dtype: object  :  32
 [Regular M&Ms]  -  0    JOY
Name:  [Regular M&Ms], dtype: object  :  27
 [Peanut M&M’s]  -  0    JOY
Name:  [Peanut M&M’s], dtype: object  :  25
 [Blue M&M's]  -  0    JOY
Name:  [Blue M&M's], dtype: object  :  65
 [Red M&M's]  -  0    JOY
Name:  [Red M&M's], dtype: object  :  63
 [Third Party M&M's]  -  0    DESPAIR
Name:  [Third Party M&M's], dtype: object  :  72
 [Minibags of chips]  -  0    MEH
Name:  [Minibags of chips], dtype: object  :  43
 [Mint Kisses]  -  0    MEH
Name:  [Mint Kisses], dtype: object  :  72
 [Mint Juleps]  -  0    JOY
Name:  [Mint Juleps], dtype: object  :  134
 [Mr. Goodbar]  -  0    JOY
Name:  [Mr. Goodbar], dtype: object  :  82
 [Necco Wafers]  -  0    DESPAIR
Name:  [Necco Wafers], dtype: object  :  79
 [Nerds]  -  0    JOY
Name:  [Nerds], dtype: object  :  48
 [Nestle Crunch]  -  0    JOY
```

```
Name:  [Nestle Crunch], dtype: object  :  41
 [Now'n'Laters]  -  0    MEH
Name:  [Now'n'Laters], dtype: object  :  129
 [Peeps]  -  0    DESPAIR
Name:  [Peeps], dtype: object  :  59
 [Pencils]  -  0    DESPAIR
Name:  [Pencils], dtype: object  :  52
 [Person of Interest Season 3 DVD Box Set (not including Disc 4 with hilarious outtake
s)]  -  0    DESPAIR
Name:  [Person of Interest Season 3 DVD Box Set (not including Disc 4 with hilarious out
takes)], dtype: object  :  123
 [Pixy Stix]  -  0    MEH
Name:  [Pixy Stix], dtype: object  :  56
 [Reese’s Peanut Butter Cups]  -  0    JOY
Name:  [Reese’s Peanut Butter Cups], dtype: object  :  27
 [Reese's Pieces]  -  0    JOY
Name:  [Reese's Pieces], dtype: object  :  29
 [Reggie Jackson Bar]  -  0    MEH
Name:  [Reggie Jackson Bar], dtype: object  :  246
 [Rolos]  -  0    JOY
Name:  [Rolos], dtype: object  :  47
 [Skittles]  -  0    JOY
Name:  [Skittles], dtype: object  :  33
 [Smarties (American)]  -  0    JOY
Name:  [Smarties (American)], dtype: object  :  47
 [Smarties (Commonwealth)]  -  0    MEH
Name:  [Smarties (Commonwealth)], dtype: object  :  163
 [Snickers]  -  0    JOY
Name:  [Snickers], dtype: object  :  36
 [Sourpatch Kids (i.e. abominations of nature)]  -  0    JOY
Name:  [Sourpatch Kids (i.e. abominations of nature)], dtype: object  :  62
 [Spotted Dick]  -  0    DESPAIR
Name:  [Spotted Dick], dtype: object  :  177
 [Starburst]  -  0    JOY
Name:  [Starburst], dtype: object  :  39
 [Sweet Tarts]  -  0    JOY
Name:  [Sweet Tarts], dtype: object  :  41
 [Swedish Fish]  -  0    JOY
Name:  [Swedish Fish], dtype: object  :  56
 [Sweetums (a friend to diabetes)]  -  0    DESPAIR
Name:  [Sweetums (a friend to diabetes)], dtype: object  :  261
 [Tic Tacs]  -  0    DESPAIR
Name:  [Tic Tacs], dtype: object  :  42
 [Those odd marshmallow circus peanut things]  -  0    DESPAIR
Name:  [Those odd marshmallow circus peanut things], dtype: object  :  70
 [Three Musketeers]  -  0    JOY
Name:  [Three Musketeers], dtype: object  :  45
 [Tolberone something or other]  -  0    JOY
Name:  [Tolberone something or other], dtype: object  :  37
 [Trail Mix]  -  0    DESPAIR
Name:  [Trail Mix], dtype: object  :  48
 [Twix]  -  0    JOY
Name:  [Twix], dtype: object  :  35
 [Vials of pure high fructose corn syrup, for main-lining into your vein]  -  0    DESPA
IR
Name:  [Vials of pure high fructose corn syrup, for main-lining into your vein], dtype:
object  :  115
 [Vicodin]  -  0    JOY
Name:  [Vicodin], dtype: object  :  94
 [Whatchamacallit Bars]  -  0    JOY
Name:  [Whatchamacallit Bars], dtype: object  :  131
 [White Bread]  -  0    DESPAIR
Name:  [White Bread], dtype: object  :  78
 [Whole Wheat anything]  -  0    DESPAIR
Name:  [Whole Wheat anything], dtype: object  :  76
 [York Peppermint Patties]  -  0    JOY
```

```
Name:  [York Peppermint Patties], dtype: object  :  52
Please list any items not included above that give you JOY.  -  0    Almond Joy
Name: Please list any items not included above that give you JOY., dtype: object  :  578
Please list any items not included above that give you DESPAIR.  -  0    Red Hots
Name: Please list any items not included above that give you DESPAIR., dtype: object  :
701
Please leave any witty, snarky or thoughtful remarks or comments regarding your choices.
  -  0    No
Name: Please leave any witty, snarky or thoughtful remarks or comments regarding your ch
oices., dtype: object  :  917
Guess the number of mints in my hand.  -  0    3
Name: Guess the number of mints in my hand., dtype: object  :  68
Betty or Veronica?  -  0    Betty
Name: Betty or Veronica?, dtype: object  :  39
"That dress* that went viral a few years back - when I first saw it, it was ________"  -
  0    White and gold
Name: "That dress* that went viral a few years back - when I first saw it, it was ______
__", dtype: object  :  46
What is your favourite font?  -  0    Helvetica
Name: What is your favourite font?, dtype: object  :  169
Please estimate the degree(s) of separation you have from the following celebrities [JK
Rowling]  -  0    3 or higher
Name: Please estimate the degree(s) of separation you have from the following celebritie
s [JK Rowling], dtype: object  :  82
Please estimate the degree(s) of separation you have from the following celebrities [JJ
Abrams]  -  0    3 or higher
Name: Please estimate the degree(s) of separation you have from the following celebritie
s [JJ Abrams], dtype: object  :  83
Please estimate the degree(s) of separation you have from the following celebrities [Bey
oncé]  -  0    3 or higher
Name: Please estimate the degree(s) of separation you have from the following celebritie
s [Beyoncé], dtype: object  :  84
Please estimate the degree(s) of separation you have from the following celebrities [Bie
ber]  -  0    3 or higher
Name: Please estimate the degree(s) of separation you have from the following celebritie
s [Bieber], dtype: object  :  84
Please estimate the degree(s) of separation you have from the following celebrities [Kev
in Bacon]  -  0    3 or higher
Name: Please estimate the degree(s) of separation you have from the following celebritie
s [Kevin Bacon], dtype: object  :  80
Please estimate the degree(s) of separation you have from the following celebrities [Fra
ncis Bacon (1561 - 1626)]  -  0    3 or higher
Name: Please estimate the degree(s) of separation you have from the following celebritie
s [Francis Bacon (1561 - 1626)], dtype: object  :  85
Which day do you prefer, Friday or Sunday?  -  0    Friday
Name: Which day do you prefer, Friday or Sunday?, dtype: object  :  24
Do you eat apples the correct way, East to West (side to side) or do you eat them like a
freak of nature, South to North (bottom to top)?  -  0    East to West
Name: Do you eat apples the correct way, East to West (side to side) or do you eat them
like a freak of nature, South to North (bottom to top)?, dtype: object  :  45
When you see the above image of the 4 different websites, which one would you most likel
y check out (please be honest).  -  0    Science: Latest News and Headlines
Name: When you see the above image of the 4 different websites, which one would you most
likely check out (please be honest)., dtype: object  :  37
```

Out[19]:
```
[None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
```

```
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
    None,
```

```
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None,
 None]
```

In [20]:
```
#Update the null values with the mode (2016 df)
candy_2016_df.fillna(candy_2016_df.select_dtypes([object]).mode().iloc[0],inplace=True)
```

In [21]:
```
# Rerun to make sure all null values are updated with the mode (2016 df)
[print(col,' - ', candy_2016_df[col].mode(), ' : ', candy_2016_df[col].isnull().sum())
 for col in candy_2016_df.columns if (isinstance(candy_2016_df[col], object) and candy_2
```

Out[21]:
```
[]
```

Remove duplicates

In [22]:
```
print('No. of rows before dropping dups: ', len(candy_2015_df))
candy_2015_df.drop_duplicates(keep='last',inplace = True)
print('No. of rows after dropping dups: ', len(candy_2015_df))
```

```
No. of rows before dropping dups:  5630
No. of rows after dropping dups:  5630
```

In [23]:
```
print('No. of rows before dropping dups: ', len(candy_2016_df))
candy_2016_df.drop_duplicates(keep='last',inplace = True)
```

```
print('No. of rows after dropping dups: ', len(candy_2016_df))
```

```
No. of rows before dropping dups:  1259
No. of rows after dropping dups:  1259
```

There are no duplicates in both dataframes.

Chapter 8

Combine and Merge Datasets (you will have to either create a new dataset from your existing data or create a relationship between the data I have provided)

In [24]:
```
#Merging the 2015 & 2016 candy dataframes
merged_df = pd.merge(candy_2015_df,candy_2016_df,on="How old are you?",how="left",suffix
print(merged_df.shape)
merged_df.head(5)
```

```
(185307, 236)
```

Out[24]:

	Timestamp_2015	How old are you?	Are you going actually going trick or treating yourself? _2015	[Butterfinger]_2015	[100 Grand Bar]_2015	[Anonymous brown globs that come in black and orange wrappers]_2015	[Any full-sized candy bar]_2015	[Black Jacks]_2015	[Bc
0	2015-10-23 08:46:20.451	35	No	JOY	JOY	DESPAIR	JOY	DESPAIR	D
1	2015-10-23 08:46:20.451	35	No	JOY	JOY	DESPAIR	JOY	DESPAIR	D
2	2015-10-23 08:46:20.451	35	No	JOY	JOY	DESPAIR	JOY	DESPAIR	D
3	2015-10-23 08:46:20.451	35	No	JOY	JOY	DESPAIR	JOY	DESPAIR	D
4	2015-10-23	35	No	JOY	JOY	DESPAIR	JOY	DESPAIR	D

08:46:20.451

5 rows × 236 columns

Pivot the data

In [25]:
```
#Using pivot function on a column in the merged dataframe
merged_df.pivot_table(values=[' [Butterfinger]_2015',' [Butterfinger]_2016'], index='How
```

Out[25]:

How old are you?	[Butterfinger]_2015	[Butterfinger]_2016
0	JOY	None
0.62	JOY	None
5	JOY	None
6	JOY	None
7	JOY	JOY
...	...	...
too old	DESPAIR	None
too old for this	JOY	None
very	DESPAIR	None
x	JOY	None
7 1 +	JOY	None

146 rows × 2 columns

Chapter 10

Grouping with Index Levels

In [26]:
```
# set index for grouping
index_df = merged_df.set_index(['Timestamp_2015','How old are you?'], inplace=True)
```

In [27]:
```
#Geoup by level (set index)
index_df = merged_df.groupby(level=['How old are you?','Timestamp_2015']).count()
index_df
```

Out[27]:

	Are you going actually going trick or treating yourself? _2015	[Butterfinger]_2015	[100 Grand Bar]_2015	[Anonymous brown globs that come in black and orange wrappers]_2015	[Any full-sized candy bar]_2015	[Black Jacks]_2015	[Bonk

How old are you?	Timestamp_2015						
0	2015-10-30 05:41:53.620	1	1	1	1	1	1
0.62	2015-10-28 19:49:06.308	1	1	1	1	1	1
5	2015-10-24 19:28:30.211	1	1	1	1	1	1
6	2015-10-23 10:44:29.949	1	1	1	1	1	1
	2015-10-23 10:44:47.028	1	1	1	1	1	1
...	...	...	...	...	...	...	...
very	2015-10-24 12:35:56.211	1	1	1	1	1	1
	2015-10-25 07:45:08.986	1	1	1	1	1	1
	2015-10-29 08:51:45.438	1	1	1	1	1	1
x	2015-10-28 19:43:31.164	1	1	1	1	1	1
7 1 +	2015-10-24 23:08:46.593	1	1	1	1	1	1

5630 rows × 234 columns

Split/Apply/Combine

In [28]:
```
#Use value_counts to get the count by unique values
merged_df[' [Butterfinger]_2015'].value_counts().sort_values(ascending=False)
```

Out[28]:
```
JOY        149889
DESPAIR     35418
Name:  [Butterfinger]_2015, dtype: int64
```

In [29]:
```
#Split.Apply.combine dataframe using groupby function
merged_df.groupby(' [Butterfinger]_2015')[' [Butterfinger]_2015'].count()
```

Out[29]:
```
 [Butterfinger]_2015
DESPAIR     35418
```

```
JOY         149889
Name:  [Butterfinger]_2015, dtype: int64
```

The above are 2 ways of getting the unique value and counts of a dataframe column

Cross Tabs

In [30]:
```
# Compute frequencies of a column using crosstab
pd.crosstab(merged_df['Are you going actually going trick or treating yourself?_2015'],
```

Out[30]:

Are you going actually going trick or treating yourself?_2015	[Butterfinger]_2015	DESPAIR	JOY	All
	No	32018	137168	169186
	Yes	3400	12721	16121
	All	35418	149889	185307

Chapter 11

Convert between string and date time

In [31]:
```
#data type of Timestamp_2015 column in the merged dataframe
merged_df = merged_df.reset_index()
merged_df['Timestamp_2015']
```

Out[31]:
```
0         2015-10-23 08:46:20.451
1         2015-10-23 08:46:20.451
2         2015-10-23 08:46:20.451
3         2015-10-23 08:46:20.451
4         2015-10-23 08:46:20.451
                    ...
185302    2015-10-31 06:41:31.904
185303    2015-10-31 06:41:31.904
185304    2015-10-31 06:41:31.904
185305    2015-10-31 06:41:31.904
185306    2015-10-31 06:41:31.904
Name: Timestamp_2015, Length: 185307, dtype: datetime64[ns]
```

In [32]:
```
#Converting the DateTime column to string
merged_df['Timestamp'] = merged_df['Timestamp_2015'].dt.strftime('%Y-%m-%d %H:%M')
```

In [33]:
```
#Validate the datatype after converting to string
merged_df['Timestamp']
```

Out[33]:
```
0         2015-10-23 08:46
1         2015-10-23 08:46
2         2015-10-23 08:46
3         2015-10-23 08:46
4         2015-10-23 08:46
                 ...
185302    2015-10-31 06:41
185303    2015-10-31 06:41
185304    2015-10-31 06:41
185305    2015-10-31 06:41
185306    2015-10-31 06:41
Name: Timestamp, Length: 185307, dtype: object
```

In [34]:
```
merged_df.head(5)
```

Out[34]:

	Timestamp_2015	How old are you?	Are you going actually going trick or treating yourself? _2015	[Butterfinger]_2015	[100 Grand Bar]_2015	[Anonymous brown globs that come in black and orange wrappers]_2015	[Any full-sized candy bar]_2015	[Black Jacks]_2015	[Bc
0	2015-10-23 08:46:20.451	35	No	JOY	JOY	DESPAIR	JOY	DESPAIR	D
1	2015-10-23 08:46:20.451	35	No	JOY	JOY	DESPAIR	JOY	DESPAIR	D
2	2015-10-23 08:46:20.451	35	No	JOY	JOY	DESPAIR	JOY	DESPAIR	D
3	2015-10-23 08:46:20.451	35	No	JOY	JOY	DESPAIR	JOY	DESPAIR	D
4	2015-10-23 08:46:20.451	35	No	JOY	JOY	DESPAIR	JOY	DESPAIR	D

5 rows × 237 columns

Convert timestamps to periods and back

In [35]:
```
#Converting the timestamp to period (by quarter) and assigning to a new column in the da
merged_df["Period"]=pd.DatetimeIndex(merged_df["Timestamp"]).to_period('Q')
```

In [36]:
```
merged_df.head(5)
```

Out[36]:

	Timestamp_2015	How old are you?	Are you going actually going trick or	[Butterfinger]_2015	[100 Grand Bar]_2015	[Anonymous brown globs that come in black and	[Any full-sized candy bar]_2015	[Black Jacks]_2015	[Bc

			treating yourself? _2015			orange wrappers]_2015			
0	2015-10-23 08:46:20.451	35	No	JOY	JOY	DESPAIR	JOY	DESPAIR	D
1	2015-10-23 08:46:20.451	35	No	JOY	JOY	DESPAIR	JOY	DESPAIR	D
2	2015-10-23 08:46:20.451	35	No	JOY	JOY	DESPAIR	JOY	DESPAIR	D
3	2015-10-23 08:46:20.451	35	No	JOY	JOY	DESPAIR	JOY	DESPAIR	D
4	2015-10-23 08:46:20.451	35	No	JOY	JOY	DESPAIR	JOY	DESPAIR	D

5 rows × 238 columns

In [37]:
```
#To convert back to timestamps, use to_timestamp
merged_df['Period'] = merged_df.Period.dt.to_timestamp(how='end')

merged_df.head(5)
```

Out[37]:

	Timestamp_2015	How old are you?	Are you going actually going trick or treating yourself? _2015	[Butterfinger]_2015	[100 Grand Bar]_2015	[Anonymous brown globs that come in black and orange wrappers]_2015	[Any full-sized candy bar]_2015	[Black Jacks]_2015	[Bc

0	2015-10-23 08:46:20.451	35	No	JOY	JOY	DESPAIR	JOY	DESPAIR	D
1	2015-10-23 08:46:20.451	35	No	JOY	JOY	DESPAIR	JOY	DESPAIR	D
2	2015-10-23 08:46:20.451	35	No	JOY	JOY	DESPAIR	JOY	DESPAIR	D
3	2015-10-23 08:46:20.451	35	No	JOY	JOY	DESPAIR	JOY	DESPAIR	D
4	2015-10-23 08:46:20.451	35	No	JOY	JOY	DESPAIR	JOY	DESPAIR	D

5 rows × 238 columns