

```
In [1]: import pandas as pd
import numpy as np
import seaborn as sns

In [2]: from openpyxl import *

In [3]: d_ProductLaunch = pd.read_csv('C:/Users/Lenovo/Documents/Project/Food & Beverage/Product Launch.csv')

In [4]: d_FlavourClass = pd.read_csv('C:/Users/Lenovo/Documents/Project/Food & Beverage/Flavour Classification.csv')

In [5]: d_PosCatMap = pd.read_csv('C:/Users/Lenovo/Documents/Project/Food & Beverage/Positioning category mapping.csv')
```

Dropping Rows with Null values

```
In [6]: df = d_ProductLaunch.dropna()

In [7]: df.isnull().sum()

Out[7]: Product id      0
Flavor              0
Market Subcategory  0
Launch Date         0
Country             0
Region              0
Positioning         0
dtype: int64
```

Only North America

```
In [8]: # Dataframe for North America
dfNA = df.loc[df['Region'] == 'North America']
```

Flavor Split

```
In [9]: # Splitting the flavours
flav_split = dfNA['Flavor'].str.split('; ', expand=True)

In [10]: # Renaming the column for flavours
flav_split = flav_split.rename(columns={0:'Flavour1', 1:'Flavour2', 2:'Flavour3'})

In [11]: # Converting flavour data to lower case
flav_split['Flavour1'] = flav_split['Flavour1'].str.lower()
flav_split['Flavour2'] = flav_split['Flavour2'].str.lower()
flav_split['Flavour3'] = flav_split['Flavour3'].str.lower()

In [12]: # Importing product id in flavor split for joining it again with main table
flav_split['Pid'] = dfNA['Product id']

In [13]: # Converting flavour data to lower case for right table
d_FlavourClass['Flavor'] = d_FlavourClass['Flavor'].str.lower()
```

Merging flavor 1 category

```
In [14]: # Merging for Flavour 1

Flav1 = pd.merge(flav_split, d_FlavourClass[['Flavor_Group', 'Flavor']], left_on = 'Flavour1', right_on = 'Flavor',
how = 'left')

Flav1

Out[14]:
```

		Flavour1	Flavour2	Flavour3	Pid	Flavor_Group	Flavor
	0	fruit, not specified	None	None	7	Fruit	fruit, not specified
	1	cherry, not specified	vanilla, not specified	None	9	Fruit	cherry, not specified
	2	orange, not specified	None	None	10	Fruit	orange, not specified
	3	superfruit, pomegranate	berry, blackberry	None	11	Fruit	superfruit, pomegranate
	4	superfruit, acai	herbs, not specified	None	15	Fruit	superfruit, acai

	19186	berry, wildberry	None	None	114685	Fruit	berry, wildberry
	19187	raspberry, not specified	None	None	114686	Fruit	raspberry, not specified
	19188	tea, white	None	None	114687	Tea	tea, white
	19189	tea, green	citrus, not specified	None	114688	Tea	tea, green
	19190	cherry, not specified	None	None	114689	Fruit	cherry, not specified

19191 rows × 6 columns

```
In [15]: # Rearranging the index after 1st merge
new_col = ['Pid', 'Flavour1', 'Flavor_Group', 'Flavour2', 'Flavour3']
Flav1 = Flav1[new_col]

In [16]: # Renaming the Flavour Group
Flav1 = Flav1.rename(columns={'Flavor_Group':'Flavor_Group1'})
```

Merging flavor 2 category

```
In [17]: # Merging for Flavour 2

Flav2 = pd.merge(Flav1, d_FlavourClass[['Flavor_Group', 'Flavor']], left_on = 'Flavour2', right_on = 'Flavor',
how = 'left')

In [18]: Flav2 = Flav2.rename(columns={'Flavor_Group':'Flavor_Group2'})
```

Merging flavor 3 category

```
In [19]: # Merging for flavour 3
Flav3 = pd.merge(Flav2, d_FlavourClass[['Flavor_Group', 'Flavor']], left_on = 'Flavour3', right_on = 'Flavor',
how = 'left')

In [20]: # Renaming th column
Flav3 = Flav3.rename(columns={'Flavor_Group':'Flavor_Group3'})

In [21]: # Rearranging the index for final split
new_col = ['Pid', 'Flavour1', 'Flavor_Group1', 'Flavour2', 'Flavor_Group2', 'Flavour3', 'Flavor_Group3']
Flavour_Final = Flav3[new_col]

Flavour_Final
```

```
Out[21]:
```

	Pid	Flavour1	Flavor_Group1	Flavour2	Flavor_Group2	Flavour3	Flavor_Group3	
	0	7	fruit, not specified	Fruit	None	NaN	None	NaN
	1	9	cherry, not specified	Fruit	vanilla, not specified	Vanilla	None	NaN
	2	10	orange, not specified	Fruit	None	NaN	None	NaN
	3	11	superfruit, pomegranate	Fruit	berry, blackberry	Fruit	None	NaN
	4	15	superfruit, acai	Fruit	herbs, not specified	Herbs	None	NaN

	19207	114685	berry, wildberry	Fruit	None	NaN	None	NaN
	19208	114686	raspberry, not specified	Fruit	None	NaN	None	NaN
	19209	114687	tea, white	Tea	None	NaN	None	NaN
	19210	114688	tea, green	Tea	citrus, not specified	Fruit	None	NaN
	19211	114689	cherry, not specified	Fruit	None	NaN	None	NaN

19212 rows × 7 columns

Merging this flavor split into the main table based on Product id

```
In [22]: # Merging Flavor split with main table based on product id

Table = pd.merge(dfNA, Flavour_Final, left_on = 'Product id', right_on = 'Pid',
how = 'left')

In [23]: # Rearranging the columns after merge
new_col = ['Product id', 'Flavor', 'Flavour1', 'Flavor_Group1', 'Flavour2', 'Flavor_Group2', 'Flavour3',
'Flavor_Group3', 'Market Subcategory', 'Launch Date', 'Country', 'Region', 'Positioning']
Table = Table[new_col]

In [24]: # Final table is ready for analysis
Table.head()
```

```
Out[24]:
```

	Product id	Flavor	Flavour1	Flavor_Group1	Flavour2	Flavor_Group2	Flavour3	Flavor_Group3	Market Subcategory	Launch Date	Country	Region	Positioning	
	0	7	Fruit, not specified	fruit, not specified	Fruit	None	NaN	None	NaN	Drink Concentrates & Mixes	1/1/2001	NA1	North America	Dry, Low Fat, Low Sodium, Low Calorie, Low Carb
	1	9	Cherry, Not specified; Vanilla, Not specified	cherry, not specified	Fruit	vanilla, not specified	Vanilla	None	NaN	Bottled Water - Flavoured	1/1/2001	NA1	North America	Convenience - Packaging
	2	10	Orange, not specified	orange, not specified	Fruit	None	NaN	None	NaN	Juice & Juice Drinks	1/1/2001	NA1	North America	Low Calorie, Low Sugar
	3	11	Superfruit, Pomegranate; Berry, Blackberry	superfruit, pomegranate	Fruit	berry, blackberry	Fruit	None	NaN	Carbonates	1/1/2001	NA1	North America	Female, Antioxidant, Low Calorie, Low Carb, Su...
	4	15	Superfruit, acai; Herbs, not specified	superfruit, acai	Fruit	herbs, not specified	Herbs	None	NaN	Juice & Juice Drinks	1/1/2001	NA2	North America	Heart Health, Antioxidant, Anti-Aging/Aging-We...

Analysis

```
In [25]: # Filtering records before 2006
Table['Launch Date'] = pd.to_datetime(Table['Launch Date'])
Tab = Table[(~(Table['Launch Date'] < '1/1/2006'))]
```

Clubbing all flavors and their groups into one

```
In [26]: # Putting all 3 flavours into 1 column
T = pd.DataFrame()
Q = pd.DataFrame()
R = pd.DataFrame()
T['flavor'] = Tab['Flavour1']
Q['flavor'] = Tab['Flavour2']
R['flavor'] = Tab['Flavour3']

In [27]: # Concatenating T and Q
U1 = pd.concat([T, Q])

In [28]: U2 = pd.concat([U1, R])

In [29]: U2.shape

Out[29]: (33966, 1)

In [30]: # Merging the clubbed flavours with their group
F = pd.merge(U2, d_FlavourClass[['Flavor_Group', 'Flavor']], left_on = 'flavor', right_on = 'Flavor',
how = 'left')

In [31]: # Rearranging columns after merge
new_col = ['Flavor_Group', 'Flavor']
F = F[new_col]

In [32]: # creating a pivot table for Flavour
P1 = F.pivot_table(index = ['Flavor'],
values = ['Flavor'],
aggfunc = ['count'])

In [33]: P1

Out[33]:
```

	Flavor_Group	
	Flavor	
	acacia	2
	allspice	2
	aloe vera	103
	anise	1
	aniseed	1

	wine, white	1
	wintergreen	2
	worcestershire	1
	yam, not specified	1
	yogurt	11

412 rows × 1 columns

```
In [34]: # creating a pivot table for Flavour group
P2 = F.pivot_table(index = ['Flavor_Group'],
values = ['Flavor_Group'],
aggfunc = ['count'])

In [35]: P2
```

```
Out[35]:
```

	Flavor	
	Flavor_Group	
	Alcohol	254
	Brown Flavors	1711
	Cake, cookie & pie	17
	Dairy Flavors	109
	Ethnic	47
	Fantasy Flavors	18
	Flowers	143
	Fruit	11903
	Grains	31
	Herbs	258
	Mint & Menthol	190
	Nuts	80
	Oil & Vinegar	22
	Poultry, meat, fish	8
	Rice	1
	Sauce & Condiment	1
	Smoke & Roasted	16
	Spices & Seeds	557
	Tea	1936
	Unflavored	579
	Vanilla	233
	Vegetable	763

```
In [ ]:

In [ ]:
```