

Coffee Survey Project

Below are a few ideas for extensions.

1. Which sweeteners should your coffee shop stockpile to meet customer preferences?
2. Which brewing methods should your coffee shop offer to optimize customer satisfaction?

Overview of the data frame

```
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('coffee-survey-results.csv')
df
```

		SubmissionID	What is your age range?	How many cups of coffee do you consume per day?	Do you add any dairy or dairy alternative to your coffee?	What kind of dairy? (Whole milk)	What kind of dairy? (Skim milk)	What kind of dairy? (Half and half)	What kind of dairy? (Coffee creamer)	What kind of dairy? (Flavored creamer)	What kind of dairy? (Oat milk)	What kind of sugar or sweetener? (Raw Sugar)	How do you brew your coffee? (Pour over)	How do you brew your coffee? (French press)	How do you brew your coffee? (Espresso)	How do you brew your coffee? (Coffee brewing machine)
0	VZbv6N	35-44	< 1	1.0	1.0	0.0	1.0	0.0	0.0	0.0	...	1.0	0.0	1.0	0.0	0.0
1	jBGZBQ	25-34	< 1	0.0	NaN	NaN	NaN	NaN	NaN	NaN	...	1.0	1.0	1.0	0.0	0.0
2	7d9809	18-24	1	1.0	1.0	0.0	1.0	0.0	0.0	0.0	...	0.0	0.0	1.0	0.0	1.0
3	bl62ee	45-54	1	1.0	0.0	0.0	1.0	0.0	0.0	0.0	...	1.0	0.0	0.0	0.0	0.0
4	BGbJ9Q	25-34	2	1.0	1.0	0.0	1.0	0.0	0.0	1.0	...	NaN	0.0	0.0	0.0	1.0
...
1165	ZdR7Dy	18-24	2	1.0	0.0	0.0	0.0	0.0	0.0	1.0	...	NaN	1.0	1.0	1.0	0.0
1166	LdMr0l	35-44	2	1.0	1.0	0.0	0.0	0.0	0.0	0.0	...	NaN	0.0	0.0	1.0	1.0
1167	ylqaYX	18-24	< 1	0.0	NaN	NaN	NaN	NaN	NaN	NaN	...	0.0	0.0	1.0	0.0	0.0
1168	KMBEyK	25-34	3	1.0	1.0	0.0	0.0	0.0	0.0	0.0	...	NaN	1.0	0.0	0.0	1.0
1169	g5ggRM	18-24	1	1.0	0.0	0.0	0.0	1.0	1.0	1.0	...	NaN	0.0	0.0	1.0	0.0

1170 rows × 30 columns

To display the questions from the survey, we used **df.columns**.

```
#Display Columns for better visualization
df.columns
```

```
Index(['SubmissionID', 'What is your age range?',
      'How many cups of coffee do you consume per day?',
      'Do you add any dairy or dairy alternative to your coffee?',
      'What kind of dairy? (Whole milk)', 'What kind of dairy? (Skim milk)',
      'What kind of dairy? (Half and half)',
      'What kind of dairy? (Coffee creamer)',
      'What kind of dairy? (Flavored creamer)',
      'What kind of dairy? (Oat milk)', 'What kind of dairy? (Almond milk)',
      'What kind of dairy? (Soy milk)',
      'Do you add any sugar or sweetener to your coffee?',
      'What kind of sugar or sweetener? (Granulated Sugar)',
      'What kind of sugar or sweetener? (Artificial Sweetener)',
      'What kind of sugar or sweetener? (Honey)',
      'What kind of sugar or sweetener? (Maple Syrup)',
      'What kind of sugar or sweetener? (Stevia)',
      'What kind of sugar or sweetener? (Agave Nectar)',
      'What kind of sugar or sweetener? (Brown Sugar)',
      'What kind of sugar or sweetener? (Raw Sugar)',
      'How do you brew your coffee? (Pour over)',
      'How do you brew your coffee? (French press)',
      'How do you brew your coffee? (Espresso)',
      'How do you brew your coffee? (Coffee brewing machine)',
      'How do you brew your coffee? (Pod/capsule machine)',
      'How do you brew your coffee? (Instant coffee)',
      'How do you brew your coffee? (Bean-to-cup machine)',
      'How do you brew your coffee? (Cold brew)',
      'How do you brew your coffee? (Coffee extract)'],
      dtype='object')
```

To answer the first question, we have to create a data frame which consists only of the columns we actually need (Sweeteners).

```
#Select needed columns
needed_columns = ['What kind of sugar or sweetener? (Granulated Sugar)',
                  'What kind of sugar or sweetener? (Artificial Sweetener)',
                  'What kind of sugar or sweetener? (Honey)',
                  'What kind of sugar or sweetener? (Maple Syrup)',
                  'What kind of sugar or sweetener? (Stevia)',
                  'What kind of sugar or sweetener? (Agave Nectar)',
                  'What kind of sugar or sweetener? (Brown Sugar)',
                  'What kind of sugar or sweetener? (Raw Sugar)']

df = df[needed_columns]
df
```

We also rename the columns for a better visualization. We also use `df.dropna()` to remove empty rows, because in the survey, if you answer that you did not put any sweetener you will not have these questions.

```
#Rename the columns for better visualization
rename_group = {'What kind of sugar or sweetener? (Granulated Sugar)': 'Granulated Sugar',
                'What kind of sugar or sweetener? (Artificial Sweetener)': 'Artificial Sweetener',
                'What kind of sugar or sweetener? (Honey)': 'Honey',
                'What kind of sugar or sweetener? (Maple Syrup)': 'Maple Syrup',
                'What kind of sugar or sweetener? (Stevia)': 'Stevia',
                'What kind of sugar or sweetener? (Agave Nectar)': 'Agave Nectar',
                'What kind of sugar or sweetener? (Brown Sugar)': 'Brown Sugar',
                'What kind of sugar or sweetener? (Raw Sugar)': 'Raw Sugar'}

df = df.rename(columns=rename_group)
#Remove null values
df.dropna()
```

After preparing the data, we are now ready for some calculations. We basically use `df.sum()` to get the total number of people used each sweetener. We also sorted the data for better visualization.

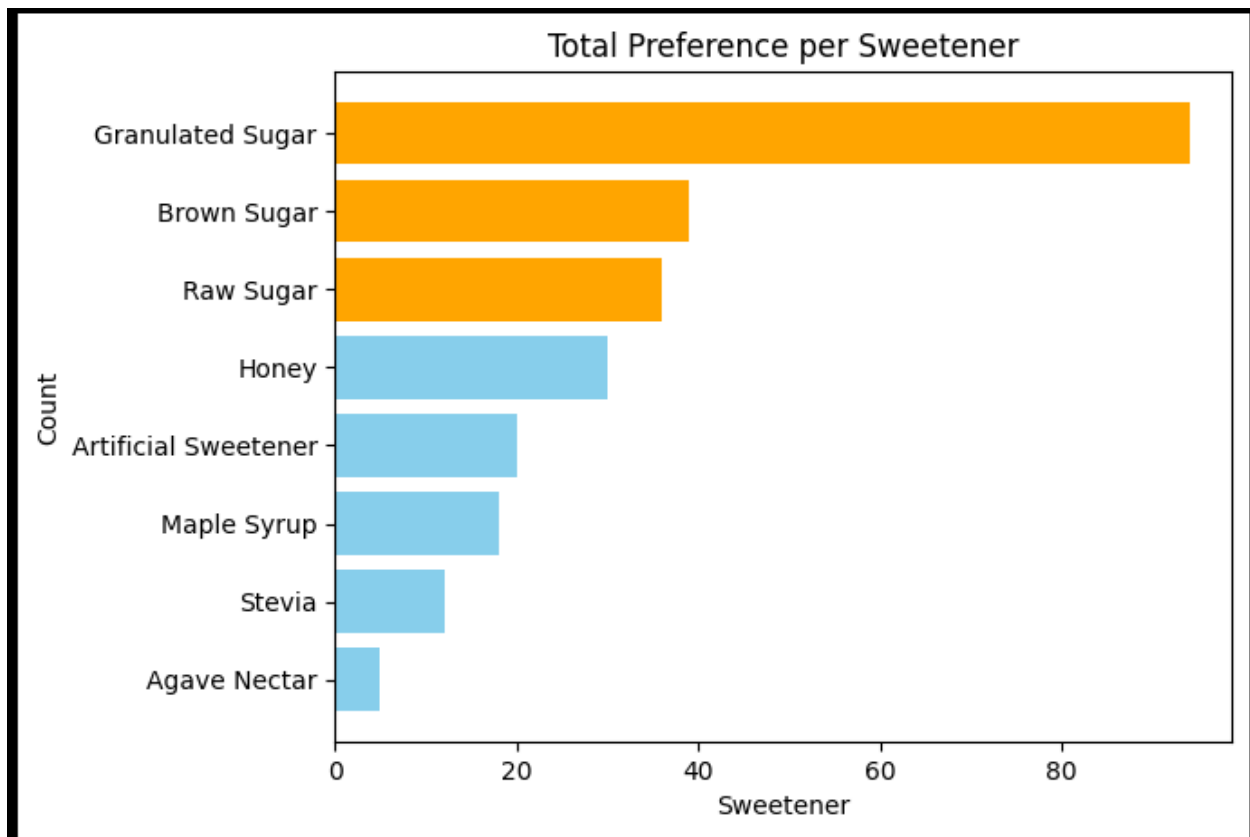
```
sweeteners_count = df.sum().reset_index().rename(columns={'count': 'count', 'index': 'sweetener'})
sweeteners_count.sort_values(by='count', ascending=True, inplace=True)
```

Plot code:

```
# Get the top 3 sweeteners
top3 = sweeteners_count.tail(3).index

# Create a color list: highlight top 3, others default
colors = ['orange' if i in top3 else 'skyblue' for i in sweeteners_count.index]

plt.barh(sweeteners_count['sweetener'], sweeteners_count['count'], color=colors)
plt.xlabel('Sweetener')
plt.ylabel('Count')
plt.title('Total Preference per Sweetener')
```



Inference:

1. Which sweeteners should your coffee shop stockpile to meet customer preferences?

We would recommend that the coffee shop should stockpile the top 3: Granulated Sugar, Brown and Raw Sugar. According to the graph, they are the most used sweetener by the customers.

For the #2 question, we will do the same method but with different columns. We will use the columns about Brewing method.

```
import pandas as pd
import matplotlib.pyplot as plt

#Reload the csv file to start fresh for the second question
df2 = pd.read_csv('coffee-survey-results.csv')
df2
#New columns for the second question
needed_columns_2 = ['How do you brew your coffee? (Pour over)',
                    'How do you brew your coffee? (French press)',
                    'How do you brew your coffee? (Espresso)',
                    'How do you brew your coffee? (Coffee brewing machine)',
                    'How do you brew your coffee? (Pod/capsule machine)',
                    'How do you brew your coffee? (Instant coffee)',
                    'How do you brew your coffee? (Bean-to-cup machine)',
                    'How do you brew your coffee? (Cold brew)',
                    'How do you brew your coffee? (Coffee extract)']
df2 = df2[needed_columns_2]
df2

#Rename the columns for better visualization
rename_group_2 = {'How do you brew your coffee? (Pour over)' : 'Pour over',
                  'How do you brew your coffee? (French press)' : 'French press',
                  'How do you brew your coffee? (Espresso)' : 'Espresso',
                  'How do you brew your coffee? (Coffee brewing machine)' : 'Coffee brewing machine',
                  'How do you brew your coffee? (Pod/capsule machine)' : 'Pod/capsule machine',
                  'How do you brew your coffee? (Instant coffee)' : 'Instant coffee',
                  'How do you brew your coffee? (Bean-to-cup machine)' : 'Bean-to-cup machine',
                  'How do you brew your coffee? (Cold brew)' : 'Cold brew',
                  'How do you brew your coffee? (Coffee extract)' : 'Coffee extract'}
df2 = df2.rename(columns=rename_group_2)
df2.dropna()
```

✓ 0.0s

Python

```
#Sum the values of each brewing method
brewing_count = df2.sum().reset_index().rename(columns={0:'count', 'index':'brewing method'})
brewing_count.sort_values(by='count', ascending=False, inplace=True)
brewing_count
```

✓ 0.0s

Python

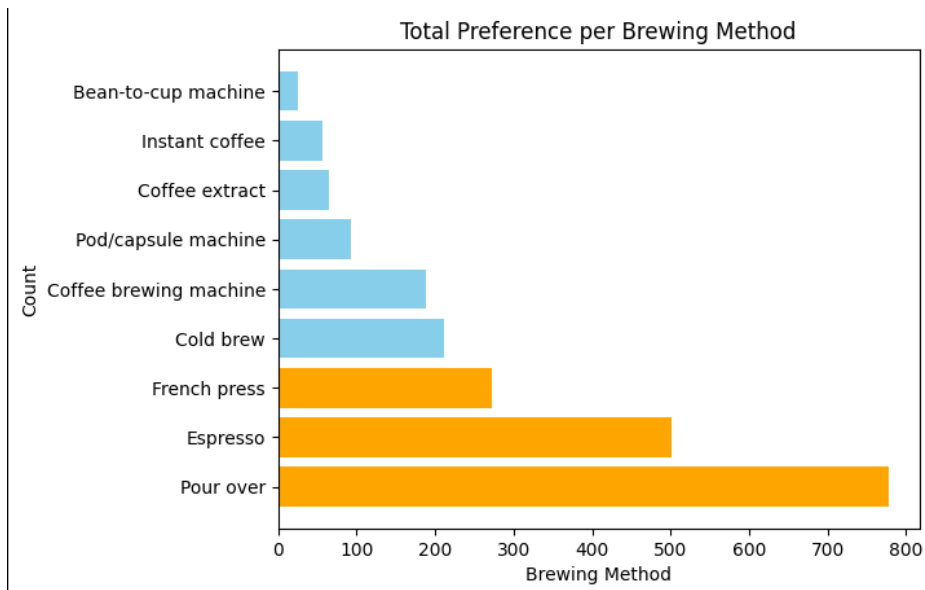
```
#Get the top 3 brewing methods
top3_brewing = brewing_count.head(3).index

#Create a color list: highlight top 3, others default
colors_brewing = ['orange' if i in top3_brewing else 'skyblue' for i in brewing_count.index]

plt.barh(brewing_count['brewing method'], brewing_count['count'], color=colors_brewing)
plt.xlabel('Brewing Method')
plt.ylabel('Count')
plt.title('Total Preference per Brewing Method')
```

✓ 0.1s

Python



Inference:

2. Which brewing methods should your coffee shop offer to optimize customer satisfaction?

The coffee shop should be advised to optimize three methods: Pour over, Espresso and French Press. These are the top 3 most preferred method by the customers. With Pour Over, greatly outranks the others.