

```
In [1]: import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

```
In [2]: df = pd.read_csv('KickstarterData.csv')
df.head()
```

Out[2]:

	Donate ID	Donate Date	Gender	Deposit Amount	Preferred Color of Device	Ice Cream Products Consumed Per Week	Favorite Flavor Of Ice Cream	Donated To Kick Starter Before	Household Income	Do you own a Keurig	
0	1	01-07-19	female	100.0	white	4.0	vanilla	yes	<50K	yes	
1	2	01-07-19	female	100.0	blue	5.0	vanilla	yes	<50K	yes	
2	3	01-07-19	female	100.0	silver	6.0	vanilla	yes	<50K	yes	
3	4	01-07-19	female	100.0	silver	5.0	vanilla	yes	<50K	yes	
4	5	01-07-19	male	100.0	white	1.0	vanilla	yes	<50K	yes	

```
In [3]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 11 columns):
Donate ID                10000 non-null int64
Donate Date              10000 non-null object
Gender                   10000 non-null object
Deposit Amount           9960 non-null float64
Preferred Color of Device 10000 non-null object
Ice Cream Products Consumed Per Week 9896 non-null float64
Favorite Flavor Of Ice Cream 10000 non-null object
Donated To Kick Starter Before 10000 non-null object
Household Income         5199 non-null object
Do you own a Keurig      10000 non-null object
How many desserts do you eat a week 10000 non-null int64
dtypes: float64(2), int64(2), object(7)
memory usage: 859.5+ KB
```

```
In [4]: df.drop('Donate ID', axis = 1, inplace = True)
```

```
In [5]: df.head()
```

Out[5]:

	Donate Date	Gender	Deposit Amount	Preferred Color of Device	Ice Cream Products Consumed Per Week	Favorite Flavor Of Ice Cream	Donated To Kick Starter Before	Household Income	Do you own a Keurig	How many desserts do you eat a week
0	01-07-19	female	100.0	white	4.0	vanilla	yes	<50K	yes	7
1	01-07-19	female	100.0	blue	5.0	vanilla	yes	<50K	yes	5
2	01-07-19	female	100.0	silver	6.0	vanilla	yes	<50K	yes	1
3	01-07-19	female	100.0	silver	5.0	vanilla	yes	<50K	yes	4
4	01-07-19	male	100.0	white	1.0	vanilla	yes	<50K	yes	7

In [6]: `df.isnull().sum()`

```
Out[6]: Donate Date      0
Gender      0
Deposit Amount      40
Preferred Color of Device      0
Ice Cream Products Consumed Per Week      104
Favorite Flavor Of Ice Cream      0
Donated To Kick Starter Before      0
Household Income      4801
Do you own a Keurig      0
How many desserts do you eat a week      0
dtype: int64
```

In [7]: `df.drop(['Household Income'], axis = 1)`

Out[7]:

	Donate Date	Gender	Deposit Amount	Preferred Color of Device	Ice Cream Products Consumed Per Week	Favorite Flavor Of Ice Cream	Donated To Kick Starter Before	Do you own a Keurig	How many desserts do you eat a week
0	01-07-19	female	100.0	white	4.0	vanilla	yes	yes	7
1	01-07-19	female	100.0	blue	5.0	vanilla	yes	yes	5
2	01-07-19	female	100.0	silver	6.0	vanilla	yes	yes	1
3	01-07-19	female	100.0	silver	5.0	vanilla	yes	yes	4
4	01-07-19	male	100.0	white	1.0	vanilla	yes	yes	7
5	01-07-19	male	100.0	white	0.0	vanilla	yes	yes	9

	Donate Date	Gender	Deposit Amount	Preferred Color of Device	Ice Cream Products Consumed Per Week	Favorite Flavor Of Ice Cream	Donated To Kick Starter Before	Do you own a Keurig	How many desserts do you eat a week
6	01-07-19	female	100.0	blue	2.0	vanilla	yes	yes	5
7	01-07-19	female	100.0	black	5.0	vanilla	yes	yes	9
8	01-07-19	male	100.0	blue	2.0	vanilla	yes	yes	7
9	01-07-19	male	100.0	no preference	1.0	vanilla	yes	yes	5
10	01-07-19	male	100.0	blue	10.0	vanilla	yes	yes	10
11	01-07-19	female	100.0	no preference	8.0	vanilla	yes	yes	2
12	01-07-19	male	100.0	silver	9.0	vanilla	yes	yes	2
13	01-07-19	male	100.0	black	4.0	vanilla	yes	yes	6
14	01-07-19	male	100.0	no preference	5.0	vanilla	yes	yes	6
15	01-07-19	male	100.0	white	0.0	vanilla	yes	yes	3
16	01-07-19	male	100.0	black	5.0	vanilla	yes	yes	5
17	01-07-19	female	100.0	no preference	7.0	vanilla	yes	yes	4
18	01-07-19	male	100.0	no preference	8.0	vanilla	yes	yes	1
19	01-07-19	male	100.0	red	0.0	vanilla	yes	yes	5
20	01-07-19	male	100.0	red	7.0	vanilla	yes	yes	10
21	01-07-19	male	100.0	black	4.0	vanilla	yes	yes	2
22	01-07-19	female	100.0	blue	8.0	vanilla	yes	yes	9
23	01-07-19	female	100.0	blue	7.0	vanilla	yes	yes	6
24	01-07-19	male	100.0	white	6.0	swirl	yes	yes	3
25	01-07-19	male	100.0	no preference	6.0	vanilla	yes	yes	2

	Donate Date	Gender	Deposit Amount	Preferred Color of Device	Ice Cream Products Consumed Per Week	Favorite Flavor Of Ice Cream	Donated To Kick Starter Before	Do you own a Keurig	How many desserts do you eat a week
26	01-07-19	male	100.0	no preference	10.0	vanilla	yes	yes	5
27	01-07-19	female	100.0	no preference	2.0	no preference	yes	yes	2
28	01-07-19	male	100.0	white	5.0	vanilla	yes	yes	5
29	01-07-19	male	100.0	black	7.0	no preference	yes	yes	1
...
9970	09-07-19	female	100.0	silver	6.0	no preference	yes	no	6
9971	09-07-19	female	100.0	white	5.0	no preference	yes	no	7
9972	09-07-19	male	100.0	red	1.0	swirl	no	no	7
9973	09-07-19	male	100.0	red	2.0	swirl	yes	no	8
9974	09-07-19	male	100.0	no preference	1.0	specialty	yes	no	5
9975	09-07-19	male	100.0	white	4.0	vanilla	no	yes	7
9976	09-07-19	male	100.0	black	0.0	no preference	no	yes	6
9977	09-07-19	male	100.0	blue	5.0	vanilla	yes	yes	8
9978	09-07-19	female	100.0	red	0.0	swirl	no	yes	7
9979	09-07-19	male	100.0	red	9.0	vanilla	no	yes	5
9980	09-07-19	male	100.0	black	6.0	vanilla	yes	no	8
9981	09-07-19	male	100.0	black	6.0	chocolate	no	no	5
9982	09-07-19	male	NaN	blue	3.0	swirl	yes	yes	6
9983	09-07-19	female	NaN	silver	3.0	vanilla	yes	no	6
9984	09-07-19	male	NaN	black	4.0	chocolate	yes	yes	6

	Donate Date	Gender	Deposit Amount	Preferred Color of Device	Ice Cream Products Consumed Per Week	Favorite Flavor Of Ice Cream	Donated To Kick Starter Before	Do you own a Keurig	How many desserts do you eat a week
9985	09-07-19	female	100.0	blue	10.0	swirl	yes	no	8
9986	09-07-19	male	100.0	white	5.0	chocolate	yes	yes	8
9987	09-07-19	male	100.0	black	1.0	vanilla	yes	yes	8
9988	09-07-19	female	100.0	no preference	7.0	vanilla	no	yes	8
9989	09-07-19	female	100.0	red	9.0	swirl	yes	no	6
9990	09-07-19	male	100.0	silver	4.0	vanilla	no	yes	4
9991	09-07-19	male	100.0	silver	3.0	chocolate	no	no	5
9992	09-07-19	male	100.0	white	0.0	specialty	yes	yes	4
9993	09-07-19	male	100.0	black	8.0	vanilla	no	no	6
9994	09-07-19	male	100.0	red	4.0	chocolate	no	yes	6
9995	09-07-19	male	100.0	black	4.0	chocolate	no	no	5
9996	09-07-19	male	100.0	silver	6.0	no preference	yes	yes	6
9997	09-07-19	male	NaN	blue	3.0	chocolate	yes	yes	5
9998	09-07-19	male	NaN	blue	6.0	vanilla	no	no	5
9999	09-07-19	male	NaN	black	7.0	chocolate	yes	yes	4

10000 rows × 9 columns

```
In [8]: meanOfIceCreamProds = round(df['Ice Cream Products Consumed Per Week'].mean(), 2)
meanOfIceCreamProds
```

Out[8]: 4.97

```
In [9]: df['Ice Cream Products Consumed Per Week'].fillna(meanOfIceCreamProds, inplace = True)
```

```
In [10]: df['Ice Cream Products Consumed Per Week'].isnull().sum()
```

Out[10]: 0

```
In [11]: modeOfDepositAmount = df['Deposit Amount'].mode()[0]
modeOfDepositAmount
```

```
Out[11]: 100.0
```

```
In [12]: df['Deposit Amount'].fillna(modeOfDepositAmount, inplace = True)
df['Deposit Amount'].isnull().sum()
```

```
Out[12]: 0
```

```
In [13]: list(set(df.dtypes.tolist()))
df_num = df.select_dtypes(include = ['float64', 'int64'])
df_num.head()
```

```
Out[13]:
```

	Deposit Amount	Ice Cream Products Consumed Per Week	How many desserts do you eat a week
0	100.0	4.0	7
1	100.0	5.0	5
2	100.0	6.0	1
3	100.0	5.0	4
4	100.0	1.0	7

```
In [14]: df_num.hist(figsize=(20, 20), bins=100, xlabelsize=12, ylabelsize=12)
```

C:\Users\hp\anaconda3\lib\site-packages\pandas\plotting_matplotlib\tools.py:307: MatplotlibDeprecationWarning:

The rowNum attribute was deprecated in Matplotlib 3.2 and will be removed two minor releases later. Use ax.get_subplotspec().rowspan.start instead.

layout[ax.rowNum, ax.colNum] = ax.get_visible()

C:\Users\hp\anaconda3\lib\site-packages\pandas\plotting_matplotlib\tools.py:307: MatplotlibDeprecationWarning:

The colNum attribute was deprecated in Matplotlib 3.2 and will be removed two minor releases later. Use ax.get_subplotspec().colspan.start instead.

layout[ax.rowNum, ax.colNum] = ax.get_visible()

C:\Users\hp\anaconda3\lib\site-packages\pandas\plotting_matplotlib\tools.py:313: MatplotlibDeprecationWarning:

The rowNum attribute was deprecated in Matplotlib 3.2 and will be removed two minor releases later. Use ax.get_subplotspec().rowspan.start instead.

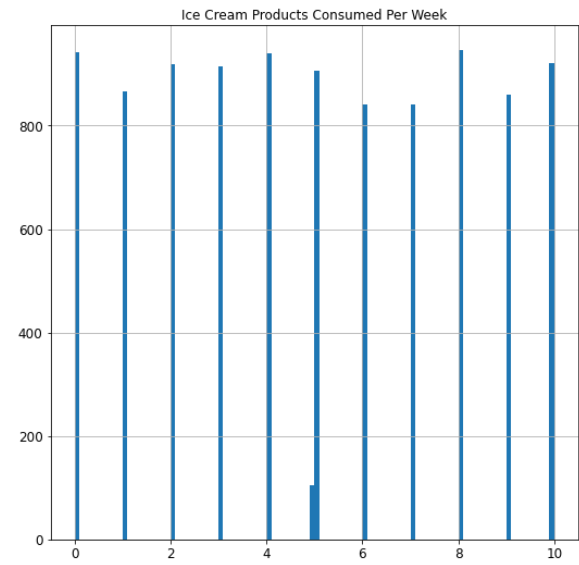
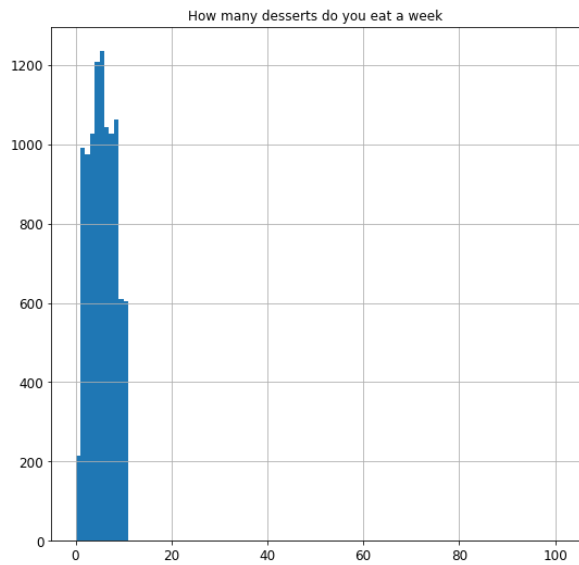
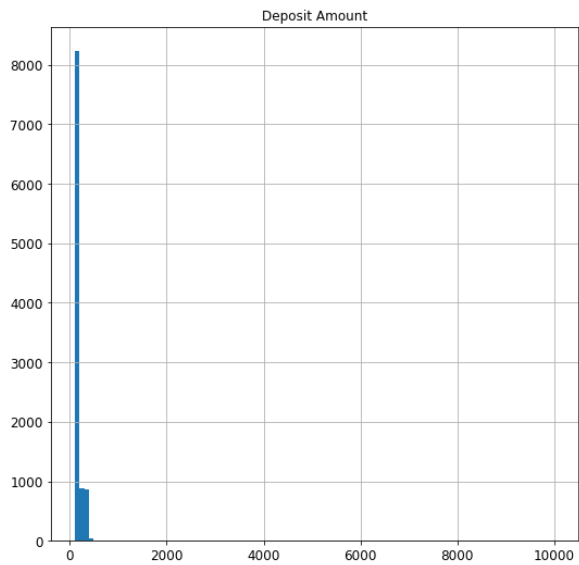
if not layout[ax.rowNum + 1, ax.colNum]:

C:\Users\hp\anaconda3\lib\site-packages\pandas\plotting_matplotlib\tools.py:313: MatplotlibDeprecationWarning:

The colNum attribute was deprecated in Matplotlib 3.2 and will be removed two minor releases later. Use ax.get_subplotspec().colspan.start instead.

if not layout[ax.rowNum + 1, ax.colNum]:

```
Out[14]: array([[<AxesSubplot:title={'center': 'Deposit Amount'}>,
<AxesSubplot:title={'center': 'How many desserts do you eat a week'}>],
[<AxesSubplot:title={'center': 'Ice Cream Products Consumed Per Week'}>,
<AxesSubplot:>]], dtype=object)
```

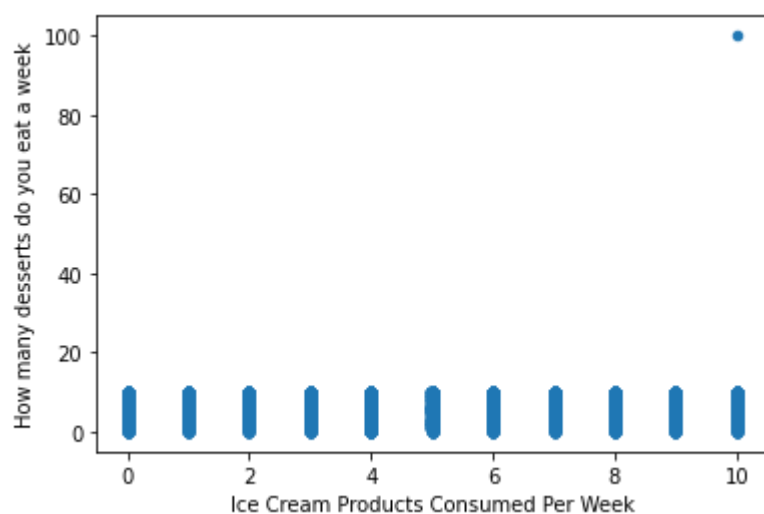


```
In [15]: df.describe()
```

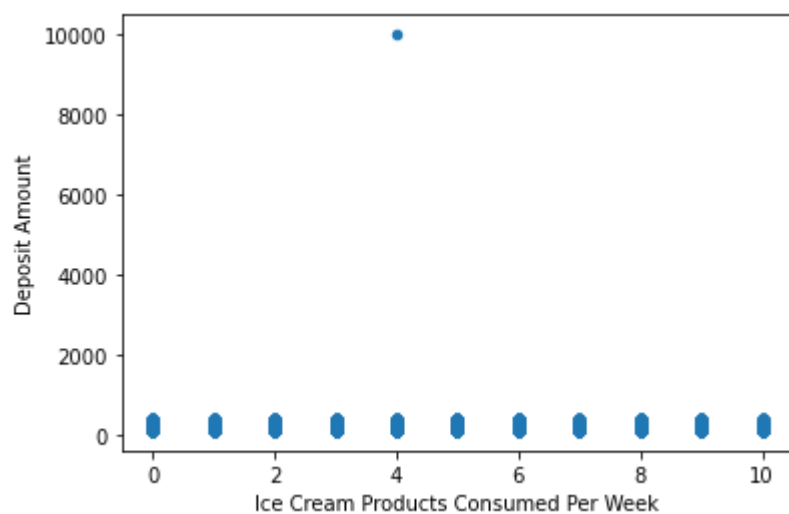
	Deposit Amount	Ice Cream Products Consumed Per Week	How many desserts do you eat a week
count	10000.00000	10000.00000	10000.00000
mean	140.90570	4.96888	5.060700
std	126.99502	3.158740	2.897008
min	100.00000	0.000000	0.000000
25%	100.00000	2.000000	3.000000
50%	100.00000	5.000000	5.000000
75%	119.00000	8.000000	7.000000
max	10000.00000	10.000000	100.000000

```
In [16]: df.plot.scatter(x="Ice Cream Products Consumed Per Week",y="How many desserts do you ea
```

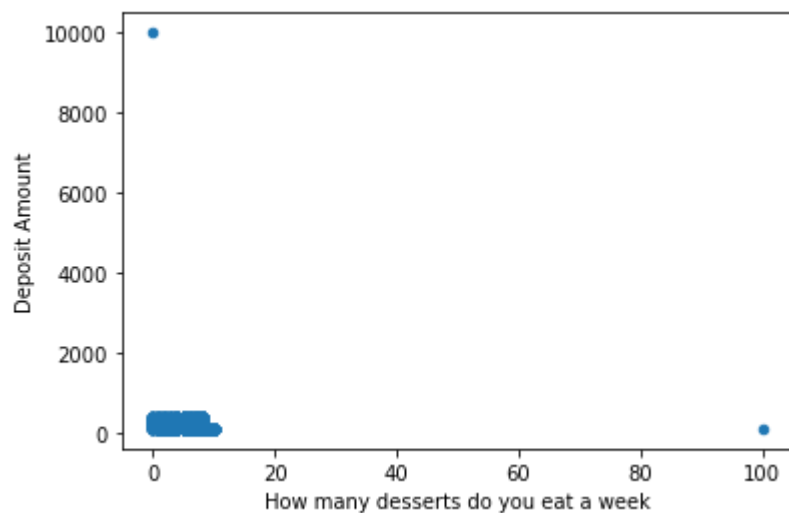
```
plt.show()
```



```
In [17]: df.plot.scatter(x="Ice Cream Products Consumed Per Week",y="Deposit Amount")  
plt.show()
```



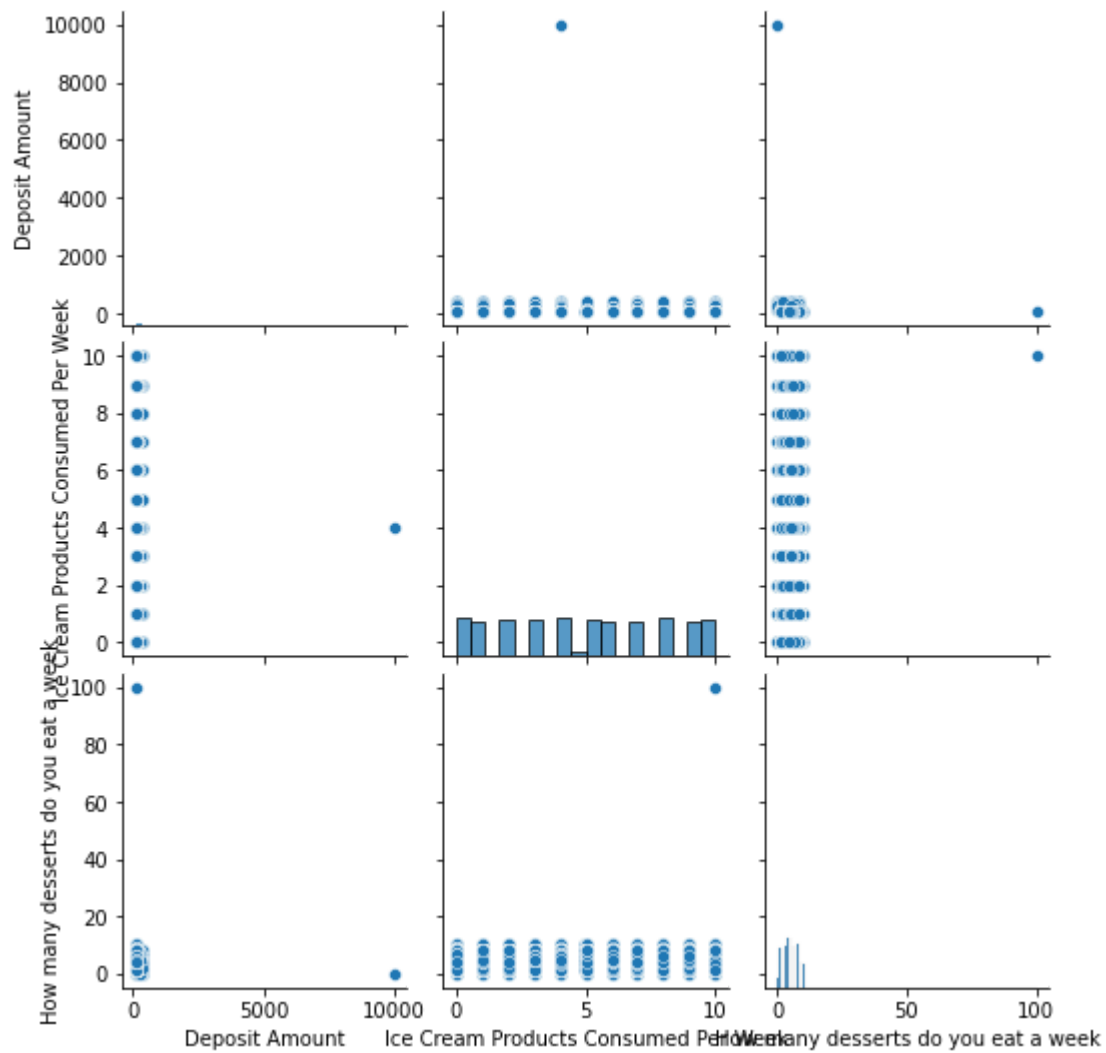
```
In [18]: df.plot.scatter(x="How many desserts do you eat a week",y="Deposit Amount")  
plt.show()
```



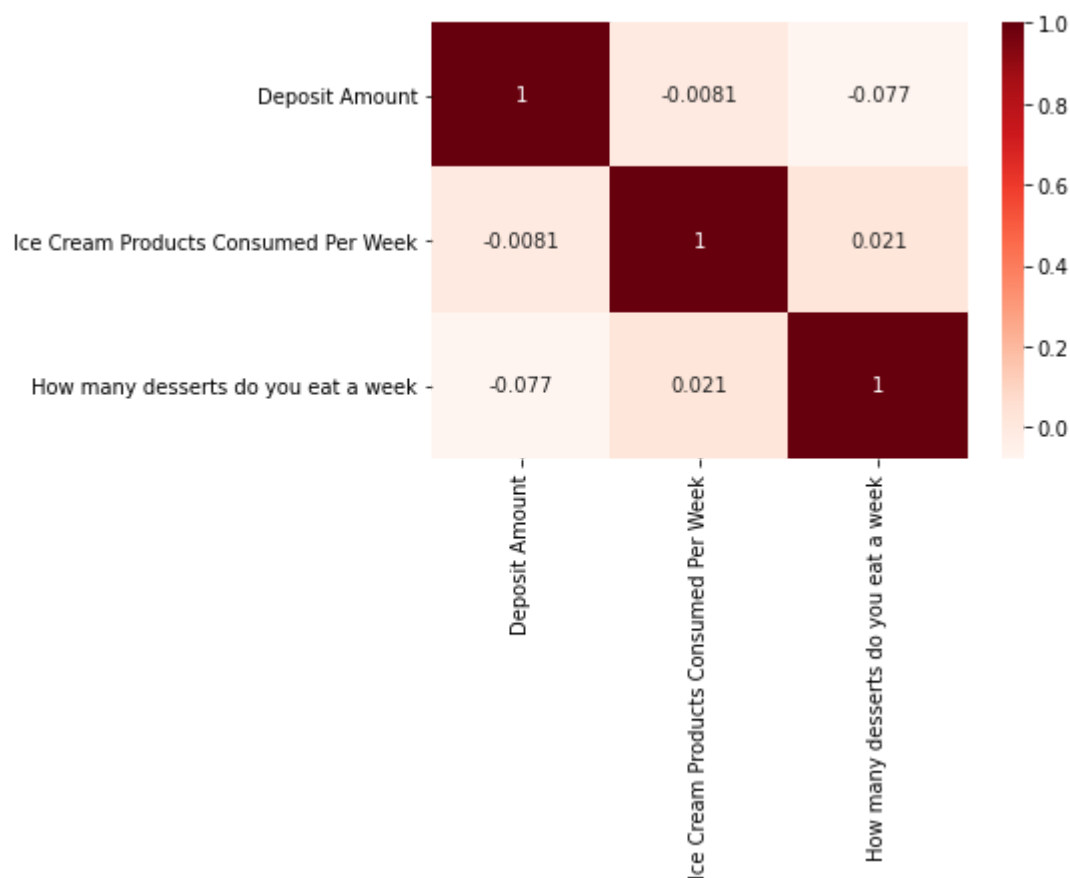
```
In [19]: sns.pairplot(data = df, vars=['Deposit Amount','Ice Cream Products Consumed Per Week'],'
```



```
plt.show()
```



```
In [20]: df[['Deposit Amount', 'Ice Cream Products Consumed Per Week', 'How many desserts do you eat a week']]
sns.heatmap(df[['Deposit Amount', 'Ice Cream Products Consumed Per Week', 'How many desserts do you eat a week']])
plt.show()
```



```
In [21]: df.groupby('Donated To Kick Starter Before')['Deposit Amount'].mean()
```

```
Out[21]: Donated To Kick Starter Before
no      111.647349
yes     154.954559
Name: Deposit Amount, dtype: float64
```

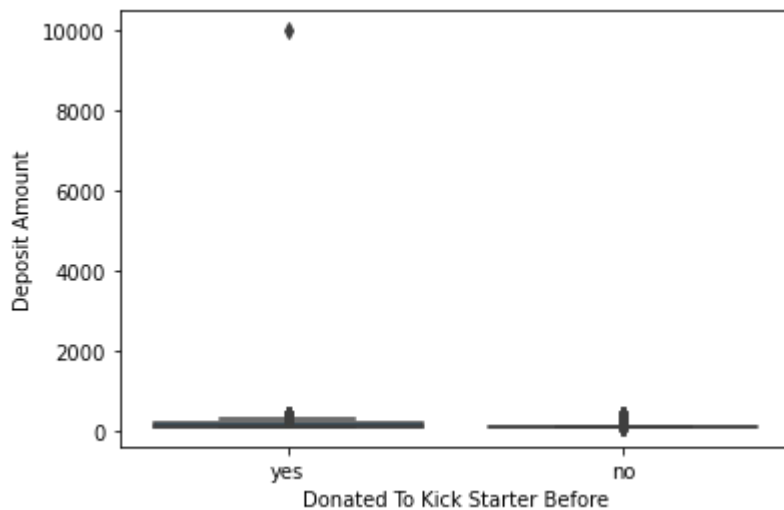
```
In [22]: df.groupby('Donated To Kick Starter Before')['Deposit Amount'].median()
```

```
Out[22]: Donated To Kick Starter Before
no       100.0
yes      100.0
Name: Deposit Amount, dtype: float64
```

```
In [23]: sns.boxplot(df['Donated To Kick Starter Before'], df['Deposit Amount'])
plt.show()
```

C:\Users\hp\anaconda3\lib\site-packages\seaborn_decorators.py:36: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

```
warnings.warn(
```



```
In [24]: df.groupby('Do you own a Keurig')['Deposit Amount'].mean()
```

```
Out[24]: Do you own a Keurig
no      137.581009
yes     141.667404
Name: Deposit Amount, dtype: float64
```

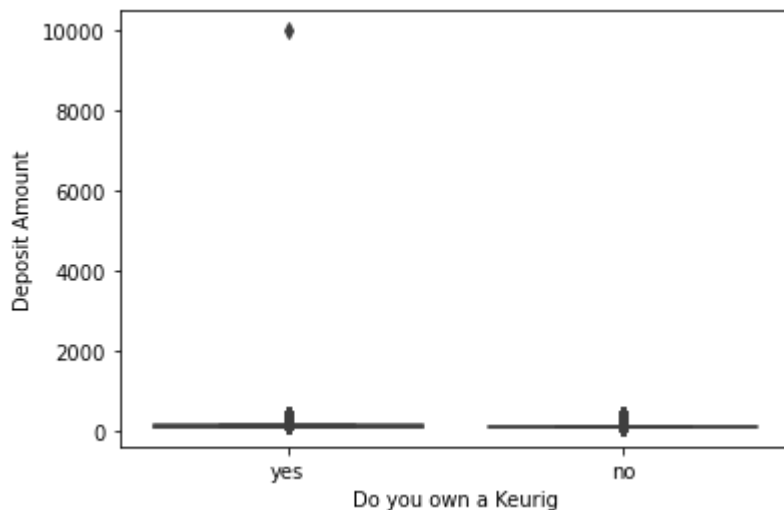
```
In [25]: df.groupby('Do you own a Keurig')['Deposit Amount'].median()
```

```
Out[25]: Do you own a Keurig
no      100.0
yes     100.0
Name: Deposit Amount, dtype: float64
```

```
In [26]: sns.boxplot(df['Do you own a Keurig'], df['Deposit Amount'])
plt.show()
```

C:\Users\hp\anaconda3\lib\site-packages\seaborn_decorators.py:36: FutureWarning: Pass the following variables as keyword args: x, y. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

warnings.warn(



```
In [27]: df['Donated_Before_Or_Not'] = np.where(df['Donated To Kick Starter Before']=='yes',1,0)
df['Donated_Before_Or_Not'].value_counts()
```

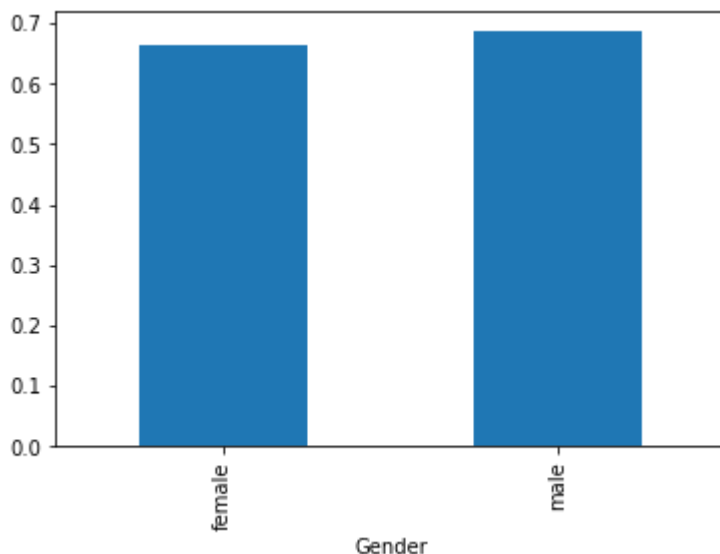
```
Out[27]: 1    6756
```

```
0    3244
Name: Donated_Before_Or_Not, dtype: int64
```

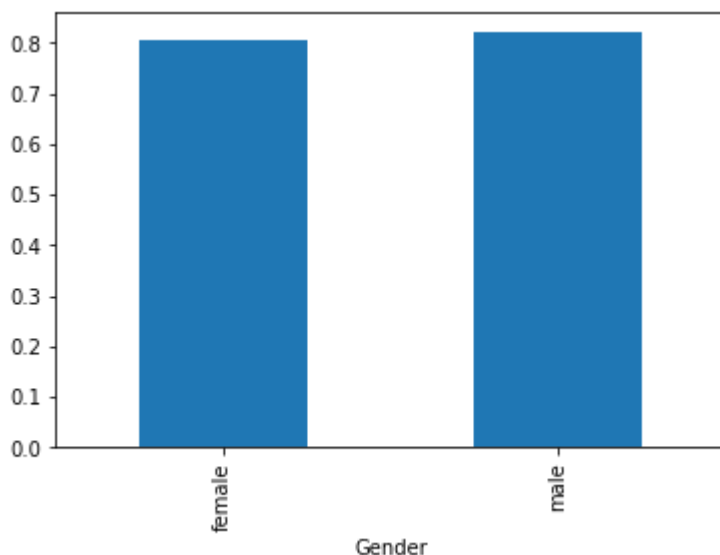
```
In [28]: df['Keurig_Or_Not'] = np.where(df['Do you own a Keurig']=='yes',1,0)
df['Keurig_Or_Not'].value_counts()
```

```
Out[28]: 1    8136
0    1864
Name: Keurig_Or_Not, dtype: int64
```

```
In [29]: df.groupby('Gender')['Donated_Before_Or_Not'].mean().plot.bar()
plt.show()
```



```
In [30]: df.groupby('Gender')['Keurig_Or_Not'].mean().plot.bar()
plt.show()
```



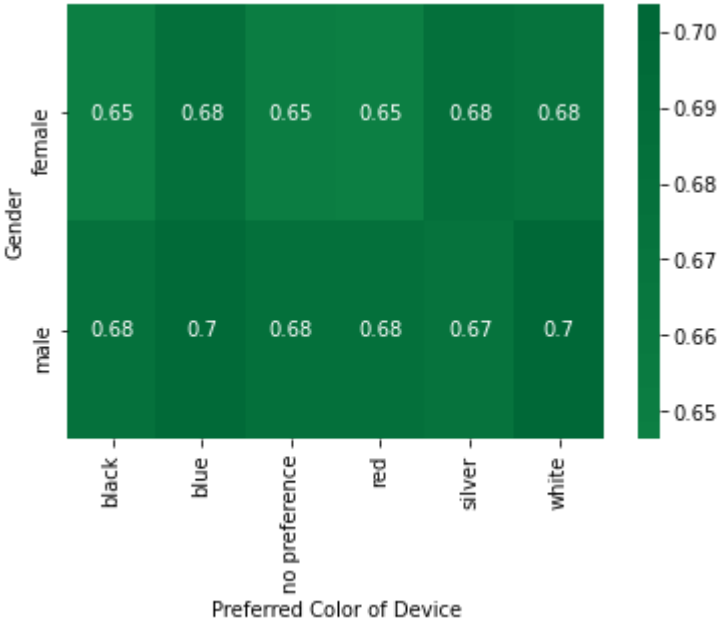
```
In [31]: result = pd.pivot_table(data=df, index='Gender', columns='Preferred Color of Device', va
print(result)

sns.heatmap(result, annot=True, cmap = 'RdYlGn', center=0.117)
plt.show()
```

```
Preferred Color of Device  black  blue  no preference  red \
Gender
```

female	0.646277	0.682741	0.649057	0.647482
male	0.679860	0.698782	0.681024	0.682243

Preferred Color of Device	silver	white
Gender		
female	0.681250	0.675033
male	0.672185	0.703704

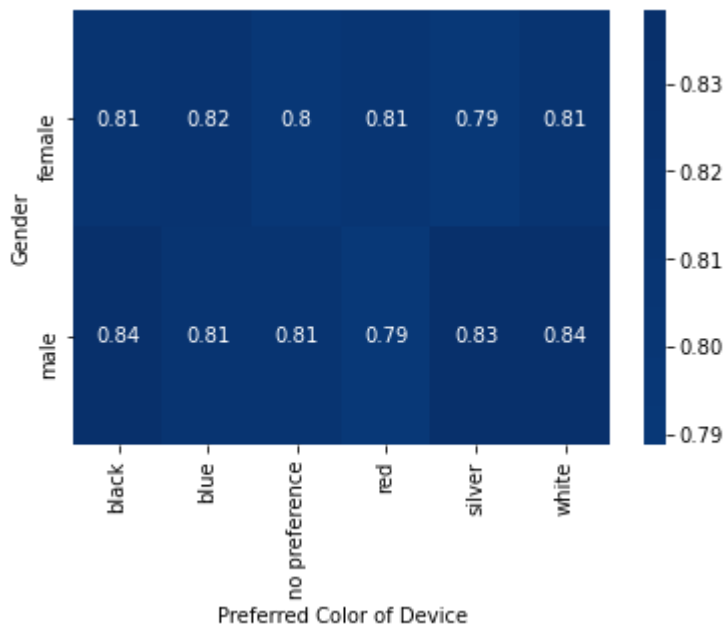


```
In [32]: result = pd.pivot_table(data=df, index='Gender', columns='Preferred Color of Device',va
print(result)

sns.heatmap(result, annot=True, cmap = 'Blues', center=0.117)
plt.show()
```

Preferred Color of Device	black	blue	no preference	red \
Gender				
female	0.81250	0.815990	0.797484	0.805755
male	0.83702	0.810631	0.811409	0.793224

Preferred Color of Device	silver	white
Gender		
female	0.788750	0.813738
male	0.833333	0.838384



```
In [33]: from autoviz.AutoViz_Class import AutoViz_Class
```

```
AV = AutoViz_Class()
```

Imported AutoViz_Class version: 0.0.81. Call using:

```
from autoviz.AutoViz_Class import AutoViz_Class
```

```
AV = AutoViz_Class()
```

```
AV.AutoViz(filename, sep=',', depVar='', dfte=None, header=0, verbose=0,
            lowess=False, chart_format='svg', max_rows_analyzed=150000, max
```

```
_cols_analyzed=30)
```

Note: verbose=0 or 1 generates charts and displays them in your local Jupyter notebook.
 verbose=2 saves plots in your local machine under AutoViz_Plots directory and does not display charts.

```
In [34]: filename = "KickstarterData.csv"
```

```
sep = ","
```

```
dft = AV.AutoViz(
```

```
    filename,
```

```
    sep=",",
```

```
    depVar="",
```

```
    dfte=None,
```

```
    header=0,
```

```
    verbose=0,
```

```
    lowess=False,
```

```
    chart_format="svg",
```

```
    max_rows_analyzed=15000,
```

```
    max_cols_analyzed=15,
```

```
)
```

Shape of your Data Set: (10000, 11)

C L A S S I F Y I N G V A R I A B L E S

Classifying variables in data set...

Number of Numeric Columns = 2

Number of Integer-Categorical Columns = 1

Number of String-Categorical Columns = 4

Number of Factor-Categorical Columns = 0

Number of String-Boolean Columns = 3

Number of Numeric-Boolean Columns = 0

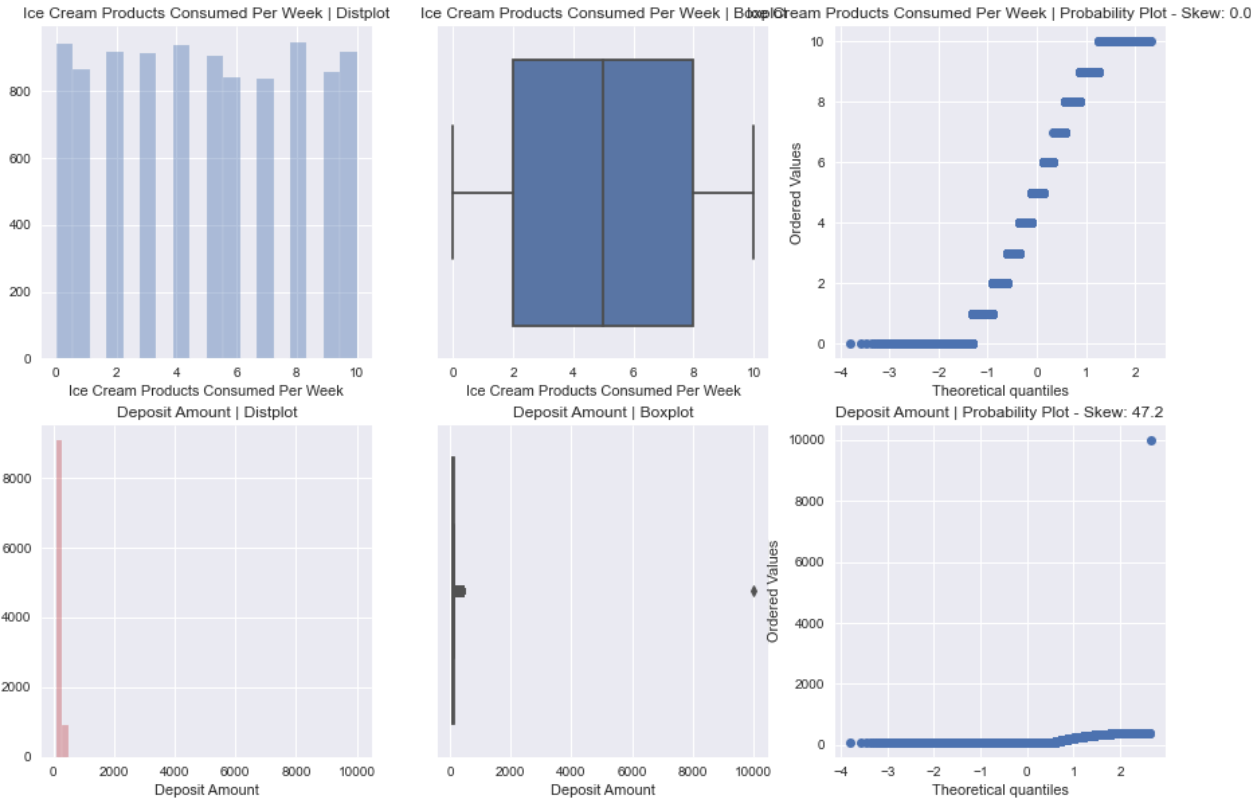
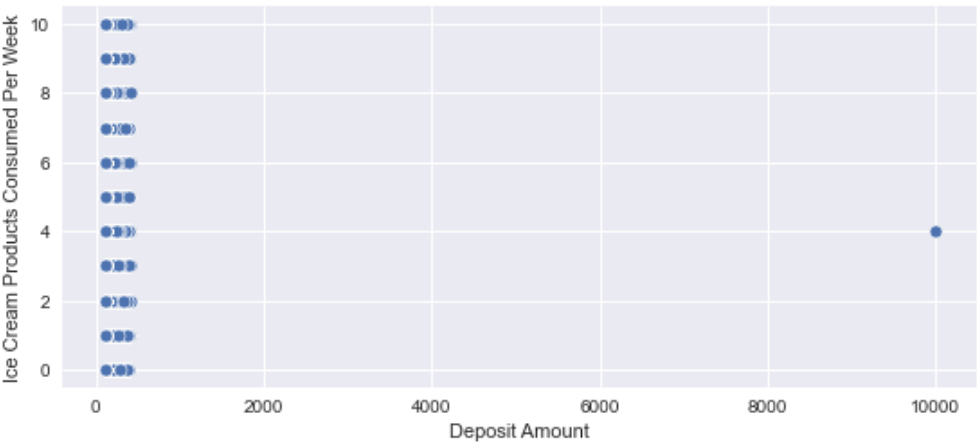
Number of Discrete String Columns = 0

Number of NLP String Columns = 0

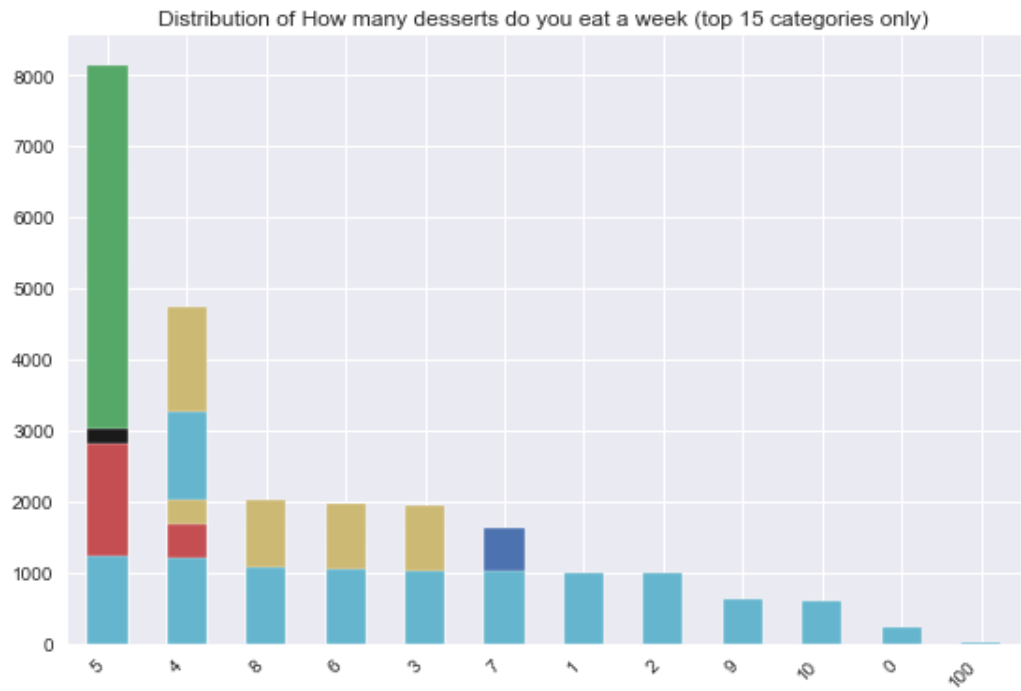
Number of Date Time Columns = 0

Number of ID Columns = 1
Number of Columns to Delete = 0
11 Predictors classified...
 This does not include the Target column(s)
 1 variables removed since they were ID or low-information variables
Number of All Scatter Plots = 3

Pair-wise Scatter Plot of all Continuous Variables



Histograms (KDE plots) of all Continuous Variables



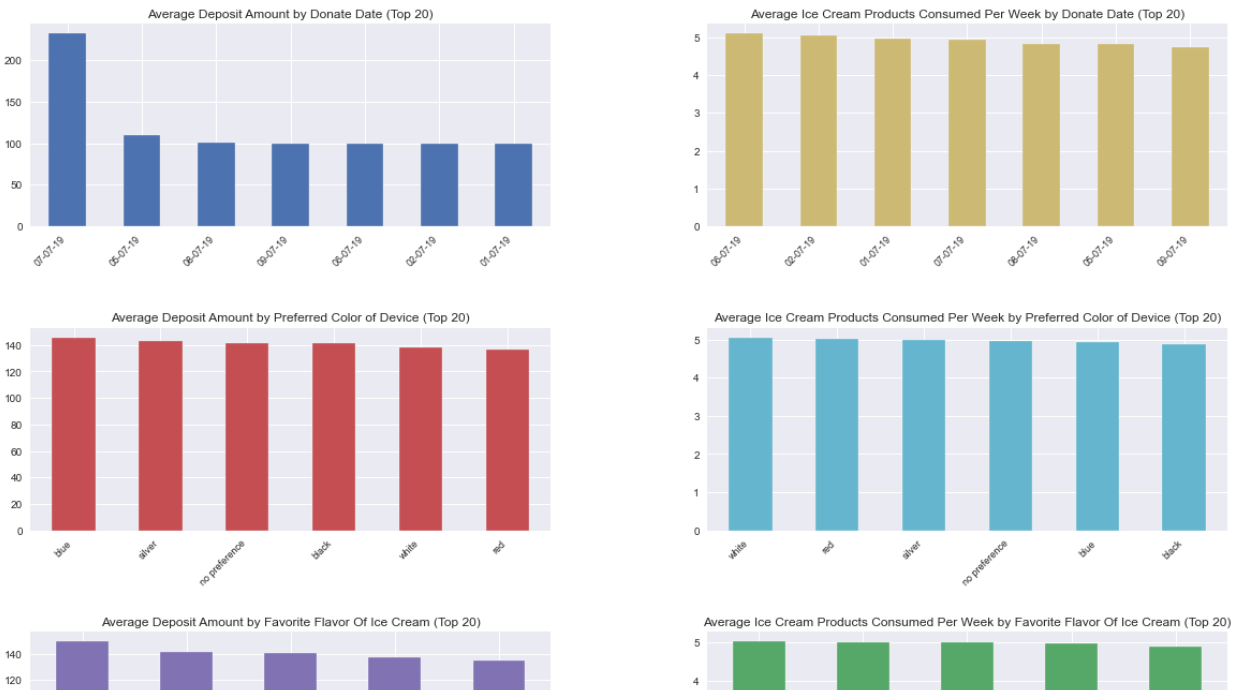
Violin Plot of all Continuous Variables

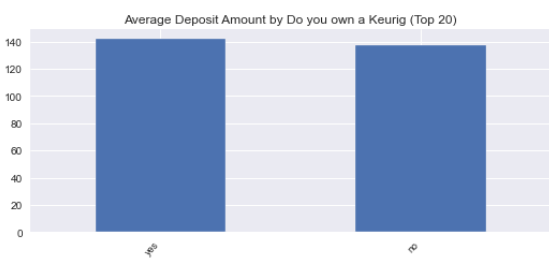
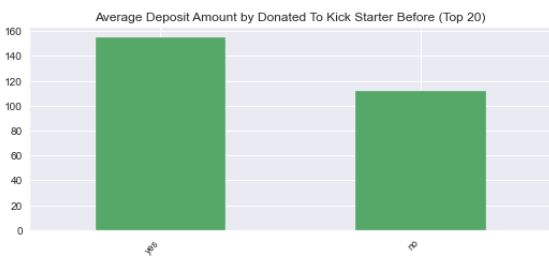
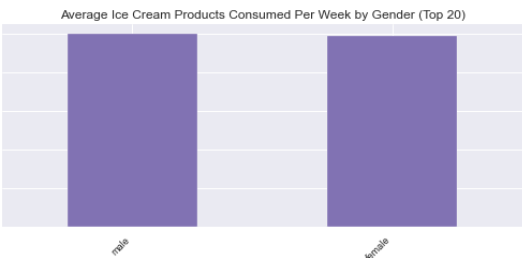
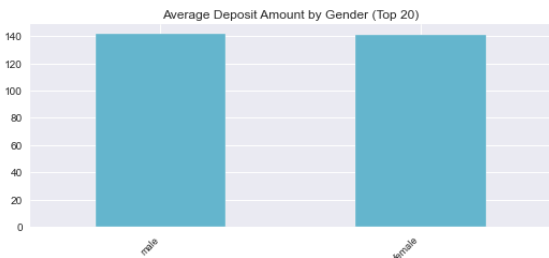
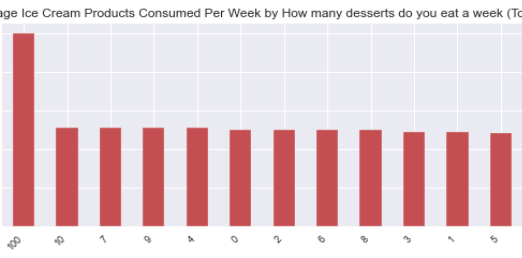
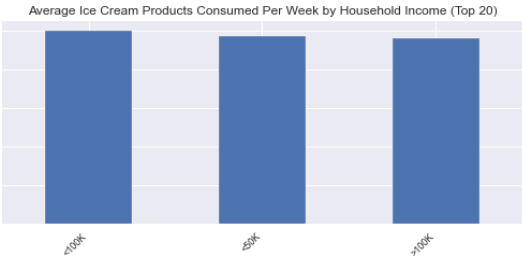
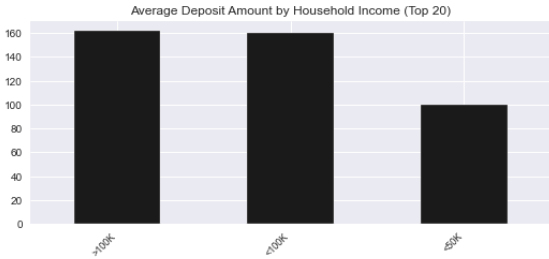
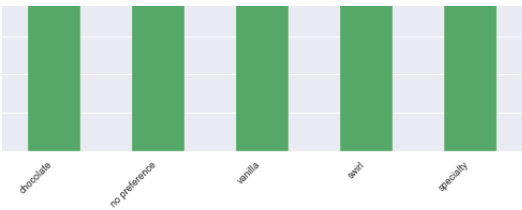
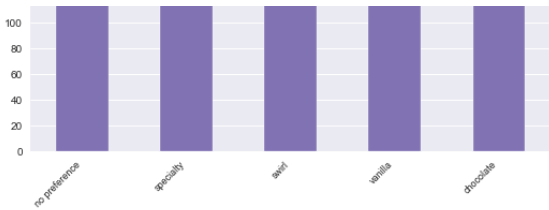


Heatmap of all Continuous Variables including target =



Bar plots for each Continuous by each Categorical variable





Time to run AutoViz (in seconds) = 6.840

VISUALIZATION Completed