Case Study - 2

In [42]:
```
# Importing the required libraries

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

In [43]:
```
Total_Revenue=pd.read_csv("casestudy.csv")
Total_Revenue.head()
```

Out[43]:

	Unnamed: 0	customer_email	net_revenue	year
0	0	nhknapwsbx@gmail.com	249.92	2015
1	1	joiuzbvcpn@gmail.com	87.61	2015
2	2	ukkjctepxt@gmail.com	168.38	2015
3	3	gykatilzrt@gmail.com	62.40	2015
4	4	mmsgsrtxah@gmail.com	43.08	2015

In [44]:
```
Total_Revenue.tail()
```

Out[44]:

	Unnamed: 0	customer_email	net_revenue	year
685922	685922	qzqttwiftu@gmail.com	184.58	2017
685923	685923	pjodiifjop@gmail.com	133.03	2017
685924	685924	appaplmgko@gmail.com	200.98	2017
685925	685925	wvkpmwsgck@gmail.com	235.35	2017
685926	685926	aregboumbw@gmail.com	208.43	2017

1. Total revenue for the current year

In [45]:
```
# Displaying the total revenue for each year i.e years - 2015, 2016 and 2017

Total_Revenue['year'].value_counts()
```

Out[45]:
```
2017    249987
2015    231294
2016    204646
Name: year, dtype: int64
```

In [46]:
```
print("Total Revenue for each Year:")
Total_Revenue.groupby('year').agg({'net_revenue':'sum'})
```

```
Total Revenue for each Year:
```

Out[46]:

year	net_revenue
2015	29036749.19
2016	25730943.59
2017	31417495.03

2. New Customer Revenue e.g., new customers not present in previous year only

In [47]:
```
Total_Revenue_17 = Total_Revenue[Total_Revenue['year']==2017]
Total_Revenue_16 = Total_Revenue[Total_Revenue['year']==2016]
Total_Revenue_15 = Total_Revenue[Total_Revenue['year']==2015]
```

In [48]:
```
# new customer revnue in 2017

Total_Revenue_17.loc[~Total_Revenue_17['customer_email'].isin(Total_Revenue_16['customer_email']),'net_rever
```

Out[48]:
```
28776235.039999995
```

In [49]:
```
# new customer revnue in 2016

Total_Revenue_16.loc[~Total_Revenue_16['customer_email'].isin(Total_Revenue_15['customer_email']),'net_rever
```

Out[49]: 18245491.01

3. Existing Customer Growth. To calculate this, use the Revenue of existing customers for current year –(minus) Revenue of existing customers from the previous year

In [50]:
```
# Existing customer growth in 2017
Total_Revenue_17.loc[Total_Revenue_17['customer_email'].isin(Total_Revenue_16['customer_email']),'net_revenu
```

Out[50]: 20611.340000000317

In [51]:
```
# existing customer growth in 2016
Total_Revenue_16.loc[Total_Revenue_16['customer_email'].isin(Total_Revenue_15['customer_email']),'net_revenu
```

Out[51]: 20335.459999999963

4. Revenue lost from attrition

In [52]:
```
# Revenue lost from attrition for year 2017

Total_Revenue_16.loc[~Total_Revenue_16['customer_email'].isin(Total_Revenue_17['customer_email']),'net_rever
```

Out[52]: 23110294.939999998

In [53]:
```
# Revenue lost from attrition for year 2016

Total_Revenue_15.loc[~Total_Revenue_15['customer_email'].isin(Total_Revenue_16['customer_email']),'net_rever
```

Out[53]: 21571632.070000004

5. Existing Customer Revenue Current Year

In [54]:
```
# existing customer revnue for current year 2017

Total_Revenue_17.loc[Total_Revenue_17['customer_email'].isin(Total_Revenue_16['customer_email']),'net_revenu
```

Out[54]: 2641259.99

In [55]:
```
# existing customer revnue for current year 2016

Total_Revenue_16.loc[Total_Revenue_16['customer_email'].isin(Total_Revenue_15['customer_email']),'net_revenu
```

Out[55]: 7485452.58

6. Existing Customer Revenue Prior Year

In [56]:
```
# existing customer revenue prior year 2016

Total_Revenue_16.loc[Total_Revenue_16['customer_email'].isin(Total_Revenue_17['customer_email']),'net_revenu
```

Out[56]: 2620648.65

In [57]:
```
# existing customer revenue prior year 2015

Total_Revenue_15.loc[Total_Revenue_15['customer_email'].isin(Total_Revenue_16['customer_email']),'net_revenu
```

Out[57]: 7465117.12

7. Total Customers Current Year

In [58]:
```
# Total customers for current year 2017

Total_Revenue_17.loc[Total_Revenue_17['customer_email'].isin(Total_Revenue_16['customer_email']),'customer_e
```

Out[58]: 20959

In [59]:
```
# Total customers for current year 2016

Total_Revenue_16.loc[Total_Revenue_16['customer_email'].isin(Total_Revenue_15['customer_email']),'customer_e
```

Out[59]: 59584

In [60]:
```
# Total customers for year 2017

Total_Revenue.loc[Total_Revenue['year']==2017,'customer_email'].nunique()
```

Out[60]: 249987

In [61]:
```
# Total customers for year 2016

Total_Revenue.loc[Total_Revenue['year']==2016,'customer_email'].nunique()
```

Out[61]: 204646

In [62]:
```
# Total customers for year 2015

Total_Revenue.loc[Total_Revenue['year']==2015,'customer_email'].nunique()
```

Out[62]: 231294

8. Total Customers Previous Year

In [63]:
```
# Total Customers for previous year 2016

Total_Revenue_16.loc[Total_Revenue_16['customer_email'].isin(Total_Revenue_17['customer_email']),'customer_e
```

Out[63]: 20959

In [64]:
```
# Total Customers for previous year 2015

Total_Revenue_15.loc[Total_Revenue_15['customer_email'].isin(Total_Revenue_16['customer_email']),'customer_e
```

Out[64]: 59584

9. New customers

In [65]:
```
# new customers in year 2017

Total_Revenue_17.loc[~Total_Revenue_17['customer_email'].isin(Total_Revenue_16['customer_email']),'customer_
```

Out[65]: 229028

In [66]:
```
# new customers in year 2016

Total_Revenue_16.loc[~Total_Revenue_16['customer_email'].isin(Total_Revenue_15['customer_email']),'customer_
```

Out[66]: 145062

10. Lost Customers

In [67]:
```
# lost customers in year 2017

Total_Revenue_16.loc[~Total_Revenue_16['customer_email'].isin(Total_Revenue_17['customer_email']),'customer_
```

Out[67]: 183687

In [68]:
```
# lost customers in year 2016

Total_Revenue_15.loc[~Total_Revenue_15['customer_email'].isin(Total_Revenue_16['customer_email']),'customer_
```

Out[68]: 171710

In [69]:
```
# lost customers in year 2017

Total_Revenue_16.loc[~Total_Revenue_16['customer_email'].isin(Total_Revenue_17['customer_email']),'customer_
```

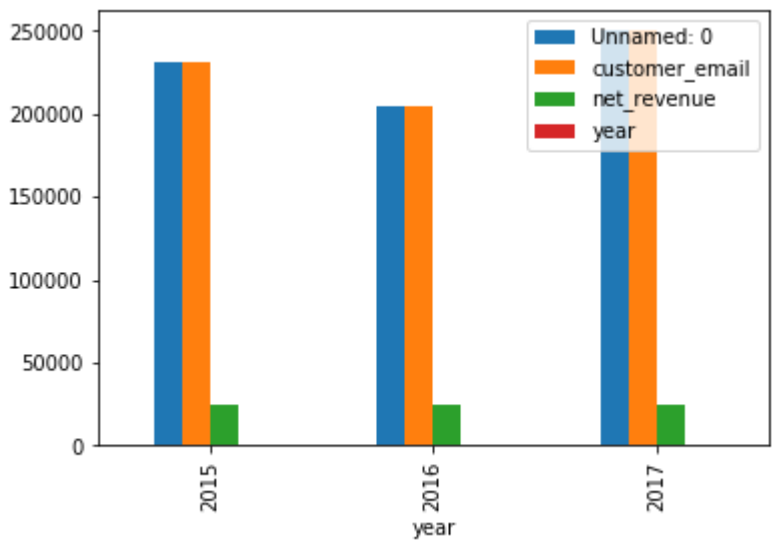
Out[69]:
```
array([' mwrossuukz@gmail.com', 'gkwsoupawk@gmail.com',
       'vlyigtgfzs@gmail.com', ..., ' rdotspqdxi@gmail.com',
       'pidugzoeej@gmail.com', ' kxqglfdktu@gmail.com'], dtype=object)
```

In [70]:
```
# lost customers in year 2016

Total_Revenue_15.loc[~Total_Revenue_15['customer_email'].isin(Total_Revenue_16['customer_email']),'customer_
```

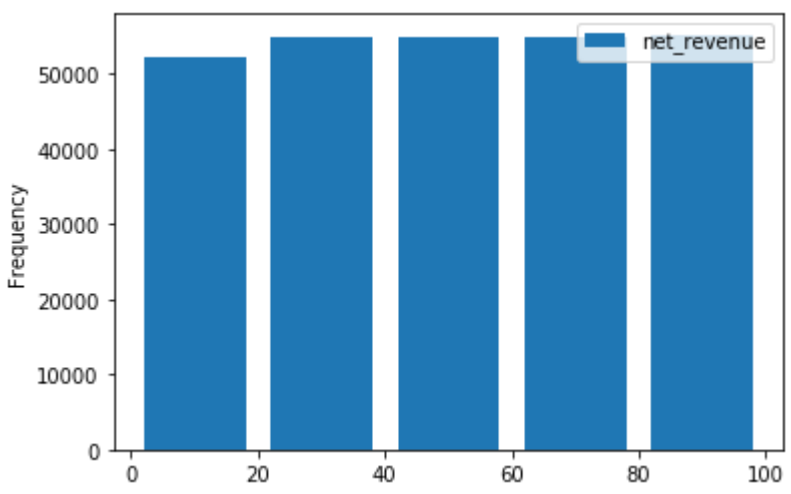
Out[70]:
```
array([' nhknapwsbx@gmail.com', 'joiuzbvcpn@gmail.com',
       'ukkjctepxt@gmail.com', ..., ' vanasezjpw@gmail.com',
       'dnpremlztb@gmail.com', 'qsgswrpycl@gmail.com'], dtype=object)
```

Additionally, generate a few unique plots highlighting some information from the dataset.

In [71]:
```
# The below plot shows the trends of each column in the dataset for each year

import matplotlib.pyplot as plt
import pandas as pd

Total_Revenue.groupby('year').nunique().plot(kind='bar')
plt.show()
```



In [72]:
```
# The below plot shows the frequency of net_revenue. As shown below, there is uniform distribution of net_re

Total_Revenue[['net_revenue']].plot(kind='hist',bins=[0,20,40,60,80,100],rwidth=0.8)
plt.show()
```



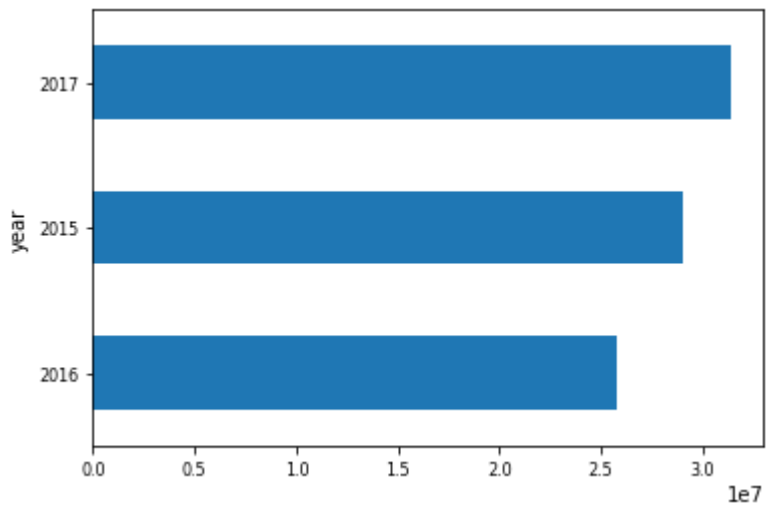
In [73]:
```
YearbyRevenue = Total_Revenue.groupby("year")["net_revenue"].sum().sort_values()
```

In [74]:
```
YearbyRevenue
```

Out[74]:
```
year
2016    25730943.59
2015    29036749.19
2017    31417495.03
Name: net_revenue, dtype: float64
```

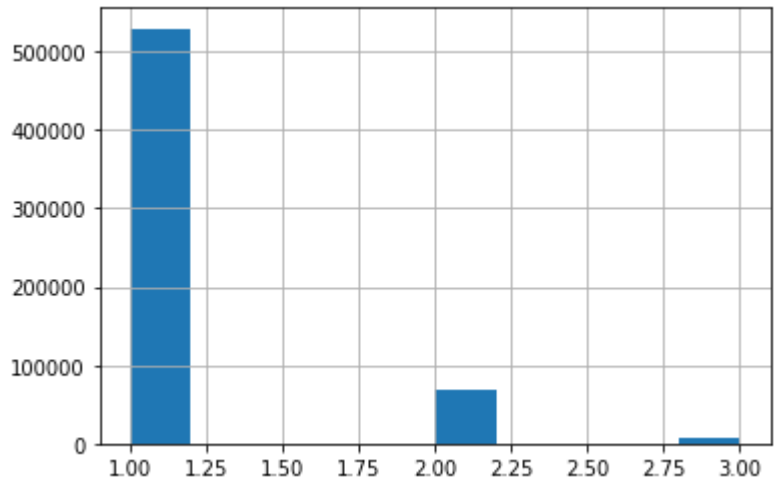
In [75]:
```
# The below plot displays the total net_revenue for each year

YearbyRevenue.plot(kind="barh", fontsize=8)
```

Out[75]: <matplotlib.axes._subplots.AxesSubplot at 0x1f9ccb5b108>



In [78]:
```
# The below plot indicates the number of times each customer_email has been repeated.

Total_Revenue['customer_email'].value_counts().hist()
```

Out[78]: <matplotlib.axes._subplots.AxesSubplot at 0x1f9cf7ccb88>



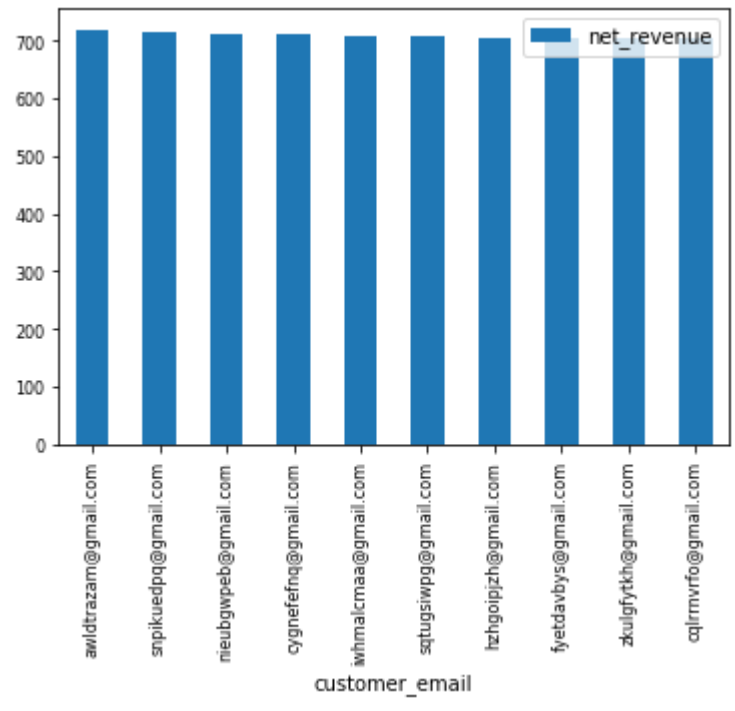
In [85]:
```
# The below plot shows the top 10 customers with highest net_revenue (in descending order)

Top_Customers=Total_Revenue.groupby('customer_email').agg({'net_revenue':'sum'}).sort_values('net_revenue',
Top_Customers
```

Out[85]:

customer_email	net_revenue
awldtrazam@gmail.com	719.58
snpikuedpq@gmail.com	713.76
nieubgwpeb@gmail.com	710.74
cygnefefnq@gmail.com	710.29
iwhmalcmaa@gmail.com	707.97
sqtugsiwpg@gmail.com	706.46
hzhgoipjzh@gmail.com	705.95
fyetdavbys@gmail.com	704.69
zkulgfytkh@gmail.com	703.99
cqlrrnvrfo@gmail.com	703.18

In [84]:
```
Top_Customers.plot(kind="bar", fontsize=8)
```

Out[84]: <matplotlib.axes._subplots.AxesSubplot at 0x1f9d53117c8>



Some of the interesting observations from above plots and graphs are :

1. The year 2017 has the maximum net_revenue.

2. Maximum number of customer emails are unique.

3. There is uniform distribution in the frequency of net_revenue.

4. Along with net_revenue, highest number of customer emails have been registered for the year 2017

In [ ]:
```

```