In [1]:
```
# Import modules
import pandas as pd
```

In [2]:
```
# Load the data frame
df = pd.read_csv("casestudy.csv")
df = df.drop(df.columns[0], axis=1)
```

In [3]:
```
years = sorted(list(set(df["year"])))
```

In [4]:
```
total_revenue_list = list()
for y in years:
    df_y = df[df["year"] == y]
    total_revenue = sum(df_y["net_revenue"])
    total_revenue_list.append(total_revenue)
    print("Total Revenue in %d: %.2f" % (y, total_revenue))
```

```
Total Revenue in 2015: 29036749.19
Total Revenue in 2016: 25730943.59
Total Revenue in 2017: 31417495.03
```

In [5]:
```
new_revenue_list = list()
for y in years:
    if y == years[0]:
        continue
    df_y = df[df["year"] == y]
    df_previous_y = df[df["year"] == y-1]
    new_customs = set(df_y["customer_email"]) - set(df_previous_y["customer_email"])
    new_revenue = df_y[df_y["customer_email"].isin(new_customs)]
    print("New Custom Revenue in %d"  % (y))
    display(new_revenue)
```

New Custom Revenue in 2016

	customer_email	net_revenue	year
231294	mwrossuukz@gmail.com	197.26	2016
231295	gkwsoupawk@gmail.com	38.28	2016
231296	vlyigtgfzs@gmail.com	154.57	2016
231297	yfirychuhk@gmail.com	102.47	2016
231298	trnzgqinuu@gmail.com	32.03	2016
...	...	...	...
435933	sqodpufunf@gmail.com	29.62	2016
435934	dcakqgznnm@gmail.com	146.47	2016
435936	rdotspqdxi@gmail.com	169.79	2016
435937	fkweqlmmjw@gmail.com	79.71	2016
435938	pidugzoeej@gmail.com	116.80	2016

145062 rows × 3 columns

New Custom Revenue in 2017

	customer_email	net_revenue	year
435940	mwrossuukz@gmail.com	96.61	2017
435941	qvjfterwnk@gmail.com	176.00	2017
435942	vlyigtgfzs@gmail.com	73.47	2017
435943	yfirychuhk@gmail.com	138.17	2017
435944	fdkiioqtli@gmail.com	156.87	2017
...	...	...	...
685922	qzqttwiftu@gmail.com	184.58	2017
685923	pjodiifjop@gmail.com	133.03	2017
685924	appaplmgko@gmail.com	200.98	2017
685925	wvkpmwsgck@gmail.com	235.35	2017
685926	aregboumbw@gmail.com	208.43	2017

229028 rows × 3 columns

In [6]:
```
for y in years:
    if y == years[0]:
        continue
    df_y = df[df["year"] == y]
    df_previous_y = df[df["year"] == y-1]
    existing_customs = set(df_y["customer_email"]).intersection(set(df_previous_y["c
    current_year_df = df_y[df_y["customer_email"].isin(existing_customs)][["customer
    last_year_df = df_previous_y[df_previous_y["customer_email"].isin(existing_custo
    merged_table = pd.merge(current_year_df, last_year_df, on = "customer_email")
    merged_table['revenue_growth'] = merged_table["net_revenue_x"] - merged_table["n
    merged_table = merged_table[["customer_email", 'revenue_growth']]
    print("Revenue Growth in %d" % (y))
    display(merged_table)
```

Revenue Growth in 2016

	customer_email	revenue_growth
0	baiikostmd@gmail.com	-28.13
1	lfeafnigbu@gmail.com	-19.93
2	tqxsjlgjpi@gmail.com	-74.21
3	hxshgpdxtr@gmail.com	36.76
4	zvhsssvgor@gmail.com	22.29
...	...	...
59579	tpdtoiokyt@gmail.com	-77.28
59580	yeserbsmyf@gmail.com	37.12
59581	nyunstvnpc@gmail.com	-6.41
59582	mezrjfkced@gmail.com	-4.02
59583	kxqglfdktu@gmail.com	-36.01

59584 rows × 2 columns

Revenue Growth in 2017

	customer_email	revenue_growth
0	yixtbjnxce@gmail.com	-67.46
1	gwqjyelgct@gmail.com	149.48
2	hxshgpdxtr@gmail.com	-39.25
3	gipqsrvngd@gmail.com	16.84
4	cdvwonzpwu@gmail.com	-37.97
...	...	...
20954	flkeldljhv@gmail.com	-31.67
20955	ecrvkbfunu@gmail.com	-25.07
20956	tpdtoiokyt@gmail.com	237.03
20957	dcakqgznnm@gmail.com	-83.81
20958	fkweqlmmjw@gmail.com	-57.93

20959 rows × 2 columns

In [7]:
```
# https://smallbusiness.chron.com/calculate-revenue-attrition-79349.html
for i, y in enumerate(years):
    if i == 0:
        continue
```

In [8]:
```
# Existing Customer Revenue Current Year
# Existing Customer Revenue Prior Year
for y in years:
    if y == years[0]:
        continue
    df_y = df[df["year"] == y]
    df_previous_y = df[df["year"] == y-1]
    existing_customs = set(df_y["customer_email"]).intersection(set(df_previous_y["c
    current_year_df = df_y[df_y["customer_email"].isin(existing_customs)][["customer
    last_year_df = df_previous_y[df_previous_y["customer_email"].isin(existing_custo
    merged_table = pd.merge(current_year_df, last_year_df, on = "customer_email")
    merged_table.rename(columns = {'net_revenue_x' : 'current_year', 'net_revenue_y'
    print("Existing Customer Revenue in %d" % (y))
    display(merged_table)
```

Existing Customer Revenue in 2016

	customer_email	current_year	last_year
0	baiikostmd@gmail.com	142.57	170.70
1	lfeafnigbu@gmail.com	35.06	54.99
2	tqxsjlgjpi@gmail.com	33.50	107.71
3	hxshgpdxtr@gmail.com	217.35	180.59
4	zvhsssvgor@gmail.com	43.84	21.55
...	...	...	...
59579	tpdtoiokyt@gmail.com	5.60	82.88
59580	yeserbsmyf@gmail.com	83.72	46.60
59581	nyunstvnpc@gmail.com	18.56	24.97
59582	mezrjfkced@gmail.com	114.17	118.19
59583	kxqglfdktu@gmail.com	37.31	73.32

59584 rows × 3 columns

Existing Customer Revenue in 2017

	customer_email	current_year	last_year
0	yixtbjnxce@gmail.com	133.86	201.32
1	gwqjyelgct@gmail.com	164.10	14.62
2	hxshgpdxtr@gmail.com	178.10	217.35
3	gipqsrvngd@gmail.com	172.90	156.06
4	cdvwonzpwu@gmail.com	102.46	140.43
...	...	...	...
20954	flkeldljhv@gmail.com	180.01	211.68
20955	ecrvkbfunu@gmail.com	205.42	230.49
20956	tpdtoiokyt@gmail.com	242.63	5.60
20957	dcakqgznnm@gmail.com	62.66	146.47
20958	fkweqlmmjw@gmail.com	21.78	79.71

20959 rows × 3 columns

In [9]:
```
for y in years:
    if y == years[0]:
        continue
    df_y = df[df["year"] == y]
    df_previous_y = df[df["year"] == y-1]
    print("Total Customer Current in year %d: %d" %(y, len(df_y)))
    print("Total Customer Prior in year %d: %d" %(y, len(df_previous_y)))
```

```
Total Customer Current in year 2016: 204646
Total Customer Prior in year 2016: 231294
Total Customer Current in year 2017: 249987
Total Customer Prior in year 2017: 204646
```

In [10]:
```
# New Customers
# Lost Customers
```

```
for y in years:
    if y == years[0]:
        continue
    df_y = df[df["year"] == y]
    df_previous_y = df[df["year"] == y-1]
    new_customers = set(df_y["customer_email"]) - set(df_previous_y["customer_email"
    lost_customers = set(df_previous_y["customer_email"]) - set(df_y["customer_email
    new_df = df_y[df_y["customer_email"].isin(new_customers)][["customer_email","net
    lost_df = df_previous_y[df_previous_y["customer_email"].isin(lost_customers)][["
    print("New Customs in %d" % (y))
    display(new_df)
    print("Lost Customs in %d" % (y))
    display(lost_df)
```

New Customs in 2016

	customer_email	net_revenue
231294	mwrossuukz@gmail.com	197.26
231295	gkwsoupawk@gmail.com	38.28
231296	vlyigtgfzs@gmail.com	154.57
231297	yfirychuhk@gmail.com	102.47
231298	trnzgqinuu@gmail.com	32.03
...	...	...
435933	sqodpufunf@gmail.com	29.62
435934	dcakqgznnm@gmail.com	146.47
435936	rdotspqdxi@gmail.com	169.79
435937	fkweqlmmjw@gmail.com	79.71
435938	pidugzoeej@gmail.com	116.80

145062 rows × 2 columns

Lost Customs in 2016

	customer_email	net_revenue
0	nhknapwsbx@gmail.com	249.92
1	joiuzbvcpn@gmail.com	87.61
2	ukkjctepxt@gmail.com	168.38
3	gykatilzrt@gmail.com	62.40
4	mmsgsrtxah@gmail.com	43.08
...	...	...
231289	xtrpmgjbwp@gmail.com	216.89
231290	peeorxpsbr@gmail.com	39.16
231291	vanasezjpw@gmail.com	233.46
231292	dnpremlztb@gmail.com	136.27
231293	qsgswrpycl@gmail.com	60.24

171710 rows × 2 columns

New Customs in 2017

	customer_email	net_revenue
435940	mwrossuukz@gmail.com	96.61
435941	qvjfterwnk@gmail.com	176.00
435942	vlyigtgfzs@gmail.com	73.47
435943	yfirychuhk@gmail.com	138.17
435944	fdkiioqtli@gmail.com	156.87
...	...	...
685922	qzqttwiftu@gmail.com	184.58
685923	pjodiifjop@gmail.com	133.03
685924	appaplmgko@gmail.com	200.98
685925	wvkpmwsgck@gmail.com	235.35
685926	aregboumbw@gmail.com	208.43

229028 rows × 2 columns

Lost Customs in 2017

	customer_email	net_revenue
231294	mwrossuukz@gmail.com	197.26
231295	gkwsoupawk@gmail.com	38.28
231296	vlyigtgfzs@gmail.com	154.57
231297	yfirychuhk@gmail.com	102.47
231298	trnzgqinuu@gmail.com	32.03
...	...	...
435933	sqodpufunf@gmail.com	29.62
435935	mezrjfkced@gmail.com	114.17
435936	rdotspqdxi@gmail.com	169.79
435938	pidugzoeej@gmail.com	116.80
435939	kxqglfdktu@gmail.com	37.31

183687 rows × 2 columns

In [ ]:

In [ ]:

In [ ]: