By:Brian Kim

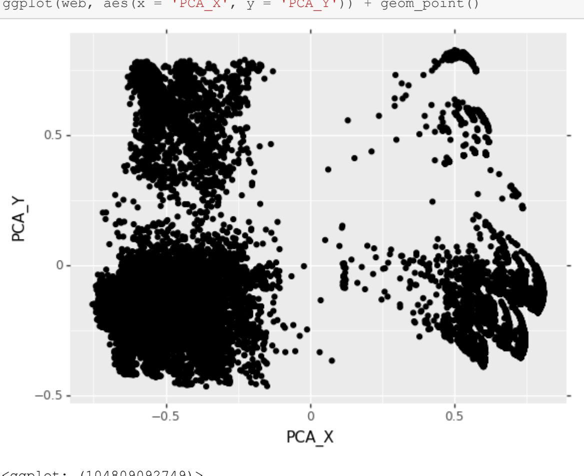
In [1]:
```
import numpy as np
import pandas as pd
from plotnine import *
from sklearn.cluster import KMeans

import warnings
warnings.filterwarnings('ignore')
```

In [2]:
```
web = pd.read_csv("Subscriber Information (Clean) Version 5.csv")
```

K-Means Clustering

Web

In [3]:
```
ggplot(web, aes(x = 'PCA_X', y = 'PCA_Y')) + geom_point()
```



Out[3]: <ggplot: (104809092749)>

In [4]:
```
web.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25291 entries, 0 to 25290
Data columns (total 71 columns):
Purchase_Amounts                          25291 non-null float64
Duration                                  25291 non-null int64
Send_Count                                25291 non-null int64
Open_Count                                25291 non-null int64
Click_Count                               25291 non-null int64
Unique_Open_Count                         25291 non-null int64
Unique_Click_Count                        25291 non-null int64
Start                                     25291 non-null int64
Other                                     25291 non-null int64
Completed                                 25291 non-null int64
NULL                                      25291 non-null int64
Onboarding                                25291 non-null int64
App_Launch_Times                          25291 non-null int64
Total_Time_Launched                       25291 non-null int64
Language_ALL                              25291 non-null float64
Language_ENG                              25291 non-null float64
Language_ESP                              25291 non-null float64
Language_FRA                              25291 non-null float64
Language_ITA                              25291 non-null float64
Language_Other                            25291 non-null float64
Subscription_Type_Lifetime                25291 non-null float64
Subscription_Type_Limited                 25291 non-null float64
Subscription_Event_Type_INITIAL_PURCHASE  25291 non-null float64
Subscription_Event_Type_RENEWAL           25291 non-null float64
Purchase_Store_App                        25291 non-null float64
Purchase_Store_Web                        25291 non-null float64
Demo_User_No                              25291 non-null float64
Demo_User_Yes                             25291 non-null float64
Free_Trial_User_No                        25291 non-null float64
Free_Trial_User_Yes                       25291 non-null float64
Auto_Renew_Off                            25291 non-null float64
Auto_Renew_On                             25291 non-null float64
Country_Europe                            25291 non-null float64
Country_Other                             25291 non-null float64
Country_US/Canada                         25291 non-null float64
User_Type_Consumer                        25291 non-null float64
User_Type_Other                           25291 non-null float64
Lead_Platform_App                         25291 non-null float64
Lead_Platform_Unknown                     25291 non-null float64
Lead_Platform_Web                         25291 non-null float64
Email_Subscriber_No                       25291 non-null float64
Email_Subscriber_Yes                      25291 non-null float64
Push_Notifications_No                     25291 non-null float64
Push_Notifications_Yes                    25291 non-null float64
PCA_X                                     25291 non-null float64
PCA_Y                                     25291 non-null float64
PCA1                                      25291 non-null float64
PCA2                                      25291 non-null float64
PCA3                                      25291 non-null float64
PCA4                                      25291 non-null float64
PCA5                                      25291 non-null float64
PCA6                                      25291 non-null float64
PCA7                                      25291 non-null float64
PCA8                                      25291 non-null float64
PCA9                                      25291 non-null float64
PCA10                                     25291 non-null float64
PCA11                                     25291 non-null float64
PCA12                                     25291 non-null float64
PCA13                                     25291 non-null float64
Language                                  25291 non-null object
Subscription_Type                         25291 non-null object
Subscription_Event_Type                   25291 non-null object
Purchase_Store                            25291 non-null object
Demo_User                                 25291 non-null object
Free_Trial_User                           25291 non-null object
Auto_Renew                                25291 non-null object
Country                                   25291 non-null object
User_Type                                 25291 non-null object
Lead_Platform                             25291 non-null object
Email_Subscriber                          25291 non-null object
Push_Notifications                        25291 non-null object
dtypes: float64(46), int64(13), object(12)
memory usage: 13.7+ MB
```

K-Means off all Variables

In [5]:
```
x = web.iloc[:,0:44]
x
```

Out[5]:

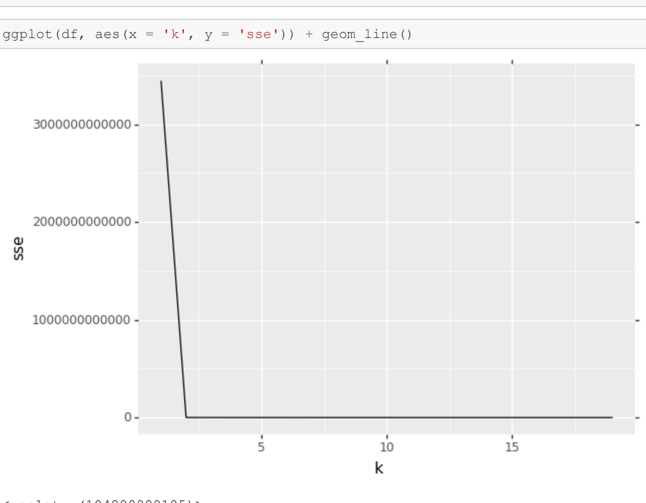
	Purchase_Amounts	Duration	Send_Count	Open_Count	Click_Count	Unique_Open_Count	Unique_Click_Count	Start	Other	Con
0	39.00	92	4	3	0	1	0	12	25	
1	0.00	365	1	0	0	0	0	0	2	
2	38.34	92	162	1	0	1	0	3	21	
3	79.00	113	2	0	0	0	0	7	9	
4	38.40	92	25	17	4	7	2	13	21	
5	43.16	97	52	11	0	5	0	15	1	
6	35.97	96	94	0	0	0	0	6	7	
7	13.43	427	70	70	0	2	0	42	67	
8	143.76	730	174	37	6	28	2	2	3	
9	199.00	28771	8	6	3	2	1	2	8	
10	98.58	564	2	1	0	1	0	54	70	
11	13.69	366	0	0	0	0	0	37	66	
12	39.50	91	8	2	0	1	0	5	22	
13	0.00	29136	2	0	0	0	0	11	13	
14	0.00	365	3	2	0	1	0	21	24	
15	65.94	286	0	0	0	0	0	33	69	
16	22.76	365	0	0	0	0	0	45	43	
17	9.59	731	0	0	0	0	0	6	8	
18	199.00	28775	23	7	5	3	2	3	3	
19	35.97	95	0	0	0	0	0	6	15	
20	0.00	366	3	2	2	1	1	11	6	
21	35.97	274	3	1	0	1	0	0	0	
22	38.89	94	19	11	4	7	2	4	7	
23	39.00	95	79	56	1	38	1	18	39	
24	0.00	182	17	0	0	0	0	0	0	
25	50.63	183	6	6	0	2	0	15	19	
26	199.00	28789	6	1	0	1	0	0	3	
27	40.18	181	0	0	0	0	0	2	3	
28	0.00	29083	0	0	0	0	0	4	8	
29	65.94	184	38	3	0	3	0	51	59	
...	...	...	...	...	...	...	...	...	...	
25261	0.00	29238	1	0	0	0	0	0	0	
25262	35.97	91	8	4	2	3	1	0	0	
25263	216.66	28767	60	12	6	4	1	0	0	
25264	0.00	181	38	16	0	16	0	0	0	
25265	79.00	275	6	5	0	1	0	0	0	
25266	210.94	28781	2	0	0	0	0	0	0	
25267	35.97	274	9	8	0	2	0	0	0	
25268	16.88	92	19	0	0	0	0	0	0	
25269	0.00	174	0	0	0	0	0	0	0	
25270	79.00	273	0	0	0	0	0	0	0	
25271	0.00	417	2	2	0	1	0	0	0	
25272	174.90	365	20	14	5	4	1	0	0	
25273	54.45	92	0	0	0	0	0	0	1	
25274	33.58	184	8	8	0	1	0	0	0	
25275	0.00	182	0	0	0	0	0	0	0	
25276	19.63	31	0	0	0	0	0	0	0	
25277	79.00	274	9	4	0	1	0	0	0	
25278	0.00	731	0	0	0	0	0	0	0	
25279	0.00	365	12	0	0	0	0	0	0	
25280	39.16	92	7	0	0	0	0	0	0	
25281	35.97	366	5	4	0	2	0	0	0	
25282	39.00	91	5	4	0	1	0	0	0	
25283	212.13	28791	6	0	0	0	0	0	0	
25284	35.97	92	16	10	0	1	0	0	0	
25285	35.97	183	4	0	0	0	0	0	0	
25286	19.63	31	40	0	0	0	0	0	0	
25287	212.13	28781	6	2	0	2	0	0	0	
25288	0.00	92	0	0	0	0	0	0	0	
25289	48.36	299	1	0	0	0	0	0	0	
25290	12.40	93	0	0	0	0	0	0	1	

25291 rows × 44 columns

In [6]:
```
sse = {}
for k in range(1,20):
    km = KMeans(n_clusters = k, max_iter = 100)
    km.fit(x)
    sse[k] = km.inertia_
```

In [7]:
```
df = pd.DataFrame({"k":list(sse.keys()), "sse":list(sse.values())})
```

In [8]:
```
ggplot(df, aes(x = 'k', y = 'sse')) + geom_line()
```



Out[8]: <ggplot: (104809909105)>

In [9]:
```
from sklearn.cluster import KMeans
km = KMeans(n_clusters = 4, random_state = 20)
km.fit(x)
```

Out[9]:
```
KMeans(algorithm='auto', copy_x=True, init='k-means++', max_iter=300,
       n_clusters=4, n_init=10, n_jobs=None, precompute_distances='auto',
       random_state=20, tol=0.0001, verbose=0)
```

In [10]:
```
centers = km.cluster_centers_
centers
```

Out[10]:
```
array([[ 4.24444229e+01,  3.62281633e+02,  2.56964898e+01,
         5.67330612e+00,  1.20685714e+00,  2.57289796e+00,
         2.39673469e-01,  2.95624490e+00,  4.11020408e+00,
         3.49844898e+00,  4.01306122e-01,  9.79591837e-04,
         5.75412245e+00,  1.67213061e+01,  6.17142857e-02,
         7.90204082e-02,  2.82938776e-01,  1.28653061e-01,
         6.53061224e-02,  3.82367347e-01,  1.17683641e-14,
         1.00000000e+00,  6.48816327e-01,  3.51183673e-01,
         8.81632653e-03,  9.91183673e-01,  8.24653061e-01,
         1.75346939e-01,  9.18204082e-01,  8.17959184e-02,
         5.22612245e-01,  4.77387755e-01,  1.15591837e-01,
         4.15183673e-01,  4.69224490e-01,  6.00653061e-01,
         3.99346939e-01,  1.25387755e-01,  3.79591837e-01,
         4.95020408e-01,  5.70448980e-01,  4.29551020e-01,
         3.79591837e-01,  6.20408163e-01],
       [ 1.88906369e+02,  2.88265564e+04,  3.92572816e+01,
         1.24626587e+01,  3.71340553e+00,  5.79387603e+00,
         8.51755041e-01,  3.05601195e+00,  3.45761763e+00,
         2.80937267e+00,  4.67699776e-01,  1.12023898e-03,
         5.02371173e+00,  1.48155340e+01,  8.31404033e-01,
         8.96191187e-03,  3.93950709e-02,  2.33383122e-02,
         9.14861837e-03,  8.77520538e-02,  1.00000000e+00,
         1.02362563e-13,  1.00000000e+00,  4.99600361e-15,
         3.73412995e-04,  9.99626587e-01,  7.94996266e-01,
         2.05003734e-01,  9.16168783e-01,  8.38312173e-02,
         9.99253174e-01,  7.46825990e-04,  5.95593727e-02,
         4.66952950e-01,  4.73487677e-01,  5.48730396e-01,
         4.51269604e-01,  1.71209858e-01,  4.40253921e-01,
         3.88536221e-01,  5.39395071e-01,  4.60604929e-01,
         4.40440627e-01,  5.59559373e-01],
       [ 3.16945211e+01,  1.28312834e+02,  2.44217916e+01,
         4.02291128e+00,  1.09336779e+00,  1.96425495e+00,
         1.98449612e-01,  2.51283376e+00,  3.22101637e+00,
         2.66511628e+00,  5.05943152e-01,  2.58397933e-04,
         4.33514212e+00,  1.32403101e+01,  1.72265289e-04,
         8.80275624e-02,  2.85271318e-01,  1.36175711e-01,
         7.23514212e-02,  4.18001723e-01, -2.42306175e-14,
         1.00000000e+00,  8.74677003e-01,  1.25322997e-01,
         6.63221361e-03,  9.93367786e-01,  8.28768303e-01,
         1.71231697e-01,  8.67441860e-01,  1.32558140e-01,
         6.93453919e-01,  3.06546081e-01,  1.27906977e-01,
         4.91559001e-01,  3.80534022e-01,  5.32558140e-01,
         4.67441860e-01,  1.14642550e-01,  4.59517657e-01,
         4.25839793e-01,  6.69078381e-01,  3.30921619e-01,
         4.59603790e-01,  5.40396210e-01],
       [ 6.08876727e+01,  7.08873182e+02,  8.05681818e+00,
         3.17045455e+00,  8.08636364e-01,  1.11727273e+00,
         1.41363636e-01,  4.35090909e+00,  4.60909091e+00,
         4.81954545e+00,  4.51363636e-01, -5.96311195e-18,
         7.82409091e+00,  2.20550000e+01,  8.95454545e-02,
         3.90909091e-02,  1.68636364e-01,  1.14090909e-01,
         5.59090909e-02,  5.32727273e-01,  4.21884749e-15,
         1.00000000e+00,  7.75454545e-01,  2.24545455e-01,
         1.81818182e-03,  9.98181818e-01,  8.35454545e-01,
         1.64545455e-01,  9.61363636e-01,  3.86363636e-02,
         2.56363636e-01,  7.43636364e-01,  2.25000000e-01,
         4.18181818e-01,  3.56818182e-01,  5.35000000e-01,
         4.65000000e-01,  1.22727273e-01,  4.00909091e-01,
         4.76363636e-01,  5.71818182e-01,  4.28181818e-01,
         4.00909091e-01,  5.99090909e-01]])
```

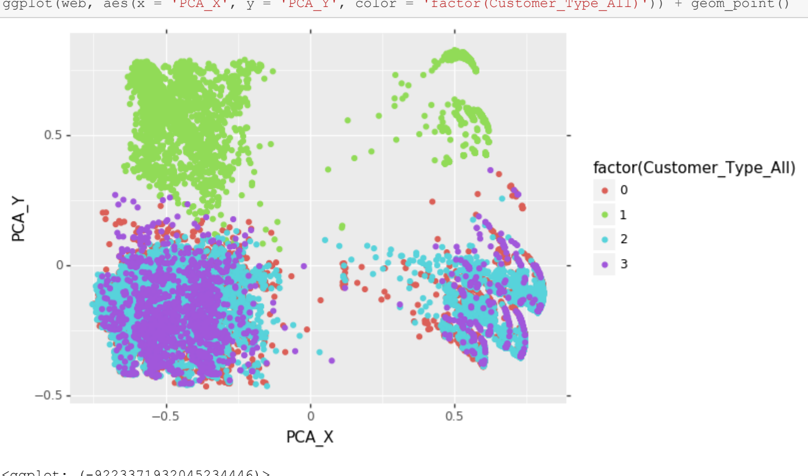
In [11]:
```
labels = km.labels_
```

In [12]:
```
labels
```

Out[12]: array([2, 0, 2, ..., 2, 0, 2])

In [13]:
```
web['Customer_Type_All'] = labels
```

In [14]:
```
ggplot(web, aes(x = 'PCA_X', y = 'PCA_Y', color = 'factor(Customer_Type_All)')) + geom_point()
```



Out[14]: <ggplot: (-9223371932045234446)>

K-means Off PCA

In [15]:
```
x = web.iloc[:,46:59]
x
```

Out[15]:

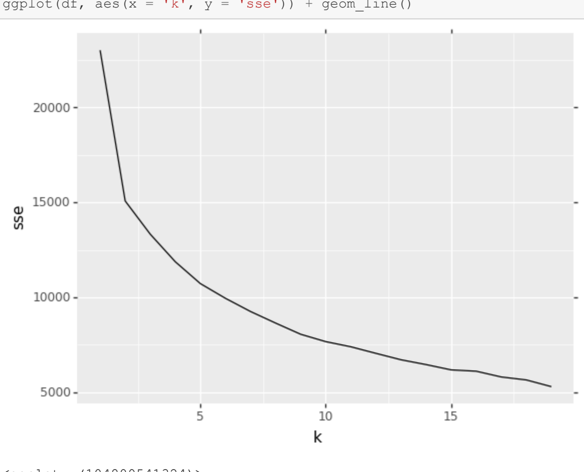
	PCA1	PCA2	PCA3	PCA4	PCA5	PCA6	PCA7	PCA8	PCA9	PCA10	PCA11	PCA12
0	-0.292194	-0.160964	-0.259184	0.051102	0.404735	0.162471	0.204449	-0.129026	-0.133562	0.239844	0.201701	0.446019
1	-0.581994	-0.139768	-0.213536	-0.072823	-0.229437	0.048782	0.170642	-0.148091	-0.162538	0.114197	0.174945	0.414098
2	-0.620542	-0.015141	0.045022	0.546917	-0.374148	-0.004509	0.103327	-0.181328	0.180809	0.050306	-0.016210	0.158139
3	-0.394892	-0.107622	0.409122	0.471256	-0.283308	0.417196	0.065522	-0.216234	-0.063017	-0.102674	-0.021113	0.097575
4	-0.604825	-0.137909	0.351037	0.505756	0.122261	-0.235885	0.035455	-0.117303	0.246212	0.078763	0.081814	0.318768
5	-0.406194	-0.263317	0.069380	-0.281943	0.595067	-0.132133	0.444822	-0.227996	0.134130	0.057659	-0.121977	-0.079076
6	-0.632558	-0.104463	-0.241098	-0.372699	-0.200424	0.039199	0.314547	-0.311037	0.284199	-0.268615	-0.098260	-0.031236
7	-0.318732	-0.113650	0.271210	-0.002947	0.318741	0.198069	0.312144	-0.128673	0.023695	0.444019	0.217103	0.514205
8	-0.557176	0.124488	0.174524	-0.253228	-0.124174	-0.295338	-0.221933	0.113846	0.406456	0.442413	-0.066711	-0.071950
9	-0.562990	0.620678	0.006194	-0.408794	0.143243	0.003673	0.213353	-0.160492	0.149628	-0.316220	-0.015088	-0.000159
10	-0.386065	-0.081742	0.156316	0.010644	0.339787	-0.006374	0.327659	-0.123174	-0.051739	0.317886	0.274534	0.587941
11	-0.360746	-0.113544	0.194980	0.217348	0.363409	0.003649	0.405244	-0.137255	-0.091587	0.312968	0.226179	0.505463
12	-0.554453	-0.013881	0.093560	0.541658	-0.077807	0.041825	0.200497	-0.004523	0.113836	-0.113151	0.545739	0.238467
13	-0.567497	0.362141	0.124027	0.487763	-0.237739	-0.023579	0.158606	-0.201522	-0.122383	-0.116378	0.035101	0.243303
14	-0.459220	-0.133786	-0.211971	-0.005122	0.083364	0.188527	0.416789	0.379960	-0.118115	0.113671	0.061498	0.479427
15	-0.300358	-0.009626	0.034777	0.062108	0.227538	0.127120	0.344823	-0.137166	-0.008779	0.251794	0.296350	0.613932
16	-0.347642	-0.085939	-0.063341	0.084637	0.419107	0.154759	0.383246	-0.138697	-0.138538	0.353463	0.225121	0.506566
17	-0.371320	-0.181434	-0.270214	0.018069	0.302051	0.188306	0.454039	0.479881	-0.326678	0.235092	-0.160385	0.084463
18	-0.547784	0.702008	0.235166	0.426912	-0.069474	-0.069197	-0.049622	-0.029303	0.035672	-0.125911	-0.031102	0.002813
19	-0.391643	-0.292090	0.067772	-0.272968	0.626424	-0.124848	0.479744	-0.235544	0.011345	-0.046162	-0.088146	-0.048242
20	-0.474910	-0.158003	-0.308416	-0.039721	0.022343	0.175655	0.332814	0.521864	-0.114163	0.073135	-0.074479	0.292423
21	0.468748	-0.259266	0.460226	-0.123761	-0.046671	0.201677	-0.033993	0.089507	-0.017083	0.074733	0.197971	-0.398897
22	-0.632253	-0.075714	-0.209887	-0.380011	-0.166915	0.046953	0.351703	-0.320274	0.331344	-0.190585	-0.116134	-0.039507
23	-0.411257	-0.025301	-0.065132	-0.109708	0.428809	0.156601	0.450488	-0.231203	0.384360	0.382385	-0.015494	0.161955
24	-0.610911	-0.166936	-0.339975	-0.112720	-0.170359	0.064775	-0.005479	0.605162	-0.035119	-0.171899	-0.103344	0.174710
25	-0.355159	-0.373939	0.409958	-0.131940	0.502452	0.278066	0.064770	-0.070773	-0.297325	0.345701	0.014634	0.139598
26	-0.515606	0.678386	0.210834	0.434911	-0.084725	-0.083457	-0.094820	-0.011901	-0.092068	-0.283108	-0.037205	-0.052801
27	-0.159409	-0.040975	0.111032	-0.029339	0.039752	-0.069221	0.146080	-0.035275	0.139207	-0.143919	0.144746	-0.048860
28	-0.326686	0.315683	0.098640	0.533994	0.182080	0.042628	0.268498	0.022300	-0.201793	-0.001625	0.298285	-0.160278
29	-0.425562	-0.066087	0.162608	-0.028175	0.121991	-0.026457	0.317841	-0.042052	0.045116	0.232203	0.534030	0.576543
...	...	...	...	...	...	...	...	...	...	...	...	...
25261	-0.439308	0.375044	-0.279884	-0.128728	0.103050	-0.055941	-0.424488	0.015309	-0.165693	-0.197950	0.000654	0.064109
25262	0.785737	-0.116550	-0.307227	0.130391	0.169334	0.044532	-0.146269	-0.086960	0.197285	0.038741	-0.035276	0.046278
25263	0.500069	0.826177	0.061576	-0.085213	0.005841	-0.024766	-0.031802	0.001751	0.059730	0.123093	-0.039125	-0.033085
25264	-0.275002	-0.340522	0.023198	-0.190629	0.148790	-0.388354	-0.379526	0.061940	0.181692	0.169076	-0.035654	-0.018940
25265	0.597590	-0.158057	0.218391	-0.089521	-0.345494	0.545797	-0.009793	-0.192782	-0.094449	0.047859	-0.083208	-0.006814
25266	0.548100	0.787303	0.046138	-0.075846	0.037217	-0.025909	-0.005379	-0.000414	-0.147063	-0.068303	0.007955	-0.002521
25267	0.478708	-0.263634	0.481883	-0.128469	-0.063151	0.185765	-0.044693	0.099000	0.058595	0.004622	0.459696	-0.073066
25268	-0.410429	-0.222455	-0.387532	-0.012076	0.361513	0.016363	-0.223065	0.025879	-0.184486	0.117283	-0.087152	-0.126891
25269	-0.428569	-0.136624	-0.303863	-0.089913	-0.175161	0.001717	-0.141539	0.194326	-0.016568	-0.171419	0.597806	-0.059439
25270	0.503246	-0.112327	0.206902	-0.038086	-0.093322	0.488354	-0.034585	0.067426	0.036664	-0.072432	0.489289	-0.062214
25271	0.729187	-0.351095	0.107498	-0.004469	0.207491	-0.363548	-0.139657	-0.032016	0.059669	-0.009023	-0.013340	0.048004
25272	0.628255	0.133478	0.328442	-0.145043	0.010960	-0.361495	0.027235	0.053691	0.109251	0.164287	-0.016233	-0.058406
25273	-0.322344	-0.242561	0.286743	0.613155	0.384291	-0.248470	0.095660	-0.060528	-0.198905	0.012170	-0.196041	-0.243371
25274	0.799984	-0.138556	-0.318437	0.135919	0.168873	0.044147	-0.150810	-0.084824	0.138635	-0.022699	-0.019036	0.056594
25275	0.788434	-0.181608	-0.337422	0.145448	0.165483	0.039881	-0.138242	-0.079510	0.084950	-0.074667	-0.002476	0.062613
25276	0.735731	-0.339900	0.117423	-0.008885	0.214679	-0.364088	-0.138399	-0.034717	0.037415	-0.029250	-0.010470	0.051728
25277	0.597278	-0.158181	0.217639	-0.089475	-0.347247	0.545823	-0.011307	-0.192610	-0.092963	0.048252	-0.083425	-0.007027
25278	0.729563	-0.351288	0.105824	-0.003916	0.209627	-0.362878	-0.137086	-0.032272	0.039694	-0.027910	-0.008947	0.050909
25279	0.787570	-0.177768	-0.340950	0.145691	0.158219	0.039144	-0.147382	-0.078586	0.109331	-0.056309	-0.007329	0.059590
25280	0.740388	-0.111370	-0.256200	0.038956	-0.390127	0.031545	0.179625	-0.199245	-0.091110	0.046386	-0.067359	0.017506
25281	0.565905	-0.337597	0.528359	-0.188414	-0.288752	0.186787	-0.015520	-0.144415	-0.112521	0.083749	-0.081913	-0.014052
25282	0.740657	-0.107390	-0.253452	0.037937	-0.391868	0.032725	0.180132	-0.200051	-0.070977	0.068530	-0.072627	0.014354
25283	0.546799	0.789443	0.046350	-0.076257	0.035353	-0.026073	-0.007417	-0.000175	-0.141268	-0.063683	0.006701	-0.003336
25284	0.680423	-0.278599	0.174282	-0.100953	-0.318054	-0.361011	0.151601	-0.143953	-0.060242	0.143411	-0.082940	0.000102
25285	0.567574	-0.342981	0.523015	-0.186327	-0.285158	0.185362	-0.013115	-0.144113	-0.138741	0.056316	-0.075224	-0.009809
25286	-0.419422	-0.214782	-0.388284	-0.013407	0.349196	0.015291	-0.235213	0.027728	-0.147909	0.145897	-0.094488	-0.131949
25287	0.545645	0.795293	0.048695	-0.077291	0.033628	-0.025572	-0.008930	0.000007	-0.122602	-0.044448	0.002204	-0.006425
25288	0.655628	-0.111970	-0.228576	0.074228	-0.100203	0.102681	0.228435	0.515715	0.079255	-0.107564	-0.149529	0.117096
25289	0.805763	-0.132038	-0.312521	0.133214	0.179185	0.043512	-0.143977	-0.087241	0.078191	-0.078526	-0.007400	0.065319
25290	-0.292134	-0.322424	0.028377	-0.064911	0.418593	0.574917	-0.144387	-0.027687	-0.363790	0.167062	-0.146544	-0.204512

25291 rows × 13 columns

In [16]:
```
sse = {}
for k in range(1,20):
    km = KMeans(n_clusters = k, max_iter = 100)
    km.fit(x)
    sse[k] = km.inertia_
```

In [17]:
```
df = pd.DataFrame({"k":list(sse.keys()), "sse":list(sse.values())})
```

In [18]:
```
ggplot(df, aes(x = 'k', y = 'sse')) + geom_line()
```



Out[18]: <ggplot: (104809541324)>

In [19]:
```
from sklearn.cluster import KMeans
km = KMeans(n_clusters = 4, random_state = 20)
km.fit(x)
```

Out[19]:
```
KMeans(algorithm='auto', copy_x=True, init='k-means++', max_iter=300,
       n_clusters=4, n_init=10, n_jobs=None, precompute_distances='auto',
       random_state=20, tol=0.0001, verbose=0)
```

In [20]:
```
centers = km.cluster_centers_
centers
```

Out[20]:
```
array([[ 5.46412166e-01,  7.46410136e-01,  2.79933009e-02,
        -6.47265309e-02,  3.57192383e-02, -2.01890980e-02,
        -1.15452318e-02, -4.68068259e-03, -1.16431532e-01,
        -4.55774711e-02, -3.57337904e-04,  3.31881369e-04,
        -3.69426459e-03],
       [ 6.70271964e-01, -1.82410917e-01,  3.05078662e-03,
         6.44100844e-03, -3.49690624e-02,  5.90262888e-03,
        -9.15155906e-03,  5.91598136e-03,  3.62340017e-02,
         3.92631471e-03,  5.28147873e-03,  1.53171709e-03,
         5.09297107e-03],
       [-4.86829650e-01, -1.77981508e-01, -2.09985574e-02,
         1.11297102e-02,  1.53693123e-02,  6.44982658e-03,
         2.16787053e-02, -1.35840945e-03, -1.42274484e-02,
        -3.56077875e-03,  3.44911728e-03, -2.40672357e-03,
        -1.22970509e-03],
       [-4.88710355e-01,  6.11129520e-01,  4.92925179e-02,
        -9.79703618e-03,  1.30948364e-02, -2.55350936e-02,
        -4.74328782e-02, -8.02524694e-03,  4.25895562e-02,
         3.83092005e-02, -2.79570267e-02,  4.54194939e-03,
        -6.94333441e-03]])
```

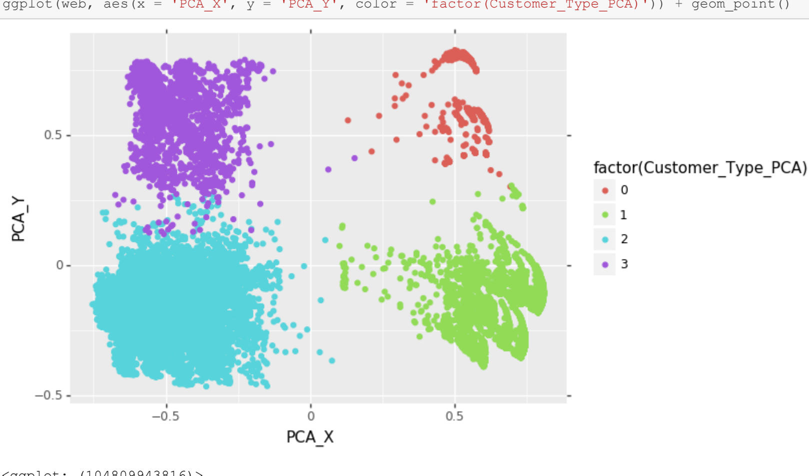
In [21]:
```
labels = km.labels_
```

In [22]:
```
labels
```

Out[22]: array([2, 2, 2, ..., 1, 1, 2])

In [23]:
```
web['Customer_Type_PCA'] = labels
```

In [24]:
```
ggplot(web, aes(x = 'PCA_X', y = 'PCA_Y', color = 'factor(Customer_Type_PCA)')) + geom_point()
```



Out[24]: <ggplot: (104809943816)>

In [25]:
```
web.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25291 entries, 0 to 25290
Data columns (total 73 columns):
Purchase_Amounts                          25291 non-null float64
Duration                                  25291 non-null int64
Send_Count                                25291 non-null int64
Open_Count                                25291 non-null int64
Click_Count                               25291 non-null int64
Unique_Open_Count                         25291 non-null int64
Unique_Click_Count                        25291 non-null int64
Start                                     25291 non-null int64
Other                                     25291 non-null int64
Completed                                 25291 non-null int64
NULL                                      25291 non-null int64
Onboarding                                25291 non-null int64
App_Launch_Times                          25291 non-null int64
Total_Time_Launched                       25291 non-null int64
Language_ALL                              25291 non-null float64
Language_ENG                              25291 non-null float64
Language_ESP                              25291 non-null float64
Language_FRA                              25291 non-null float64
Language_ITA                              25291 non-null float64
Language_Other                            25291 non-null float64
Subscription_Type_Lifetime                25291 non-null float64
Subscription_Type_Limited                 25291 non-null float64
Subscription_Event_Type_INITIAL_PURCHASE  25291 non-null float64
Subscription_Event_Type_RENEWAL           25291 non-null float64
Purchase_Store_App                        25291 non-null float64
Purchase_Store_Web                        25291 non-null float64
Demo_User_No                              25291 non-null float64
Demo_User_Yes                             25291 non-null float64
Free_Trial_User_No                        25291 non-null float64
Free_Trial_User_Yes                       25291 non-null float64
Auto_Renew_Off                            25291 non-null float64
Auto_Renew_On                             25291 non-null float64
Country_Europe                            25291 non-null float64
Country_Other                             25291 non-null float64
Country_US/Canada                         25291 non-null float64
User_Type_Consumer                        25291 non-null float64
User_Type_Other                           25291 non-null float64
Lead_Platform_App                         25291 non-null float64
Lead_Platform_Unknown                     25291 non-null float64
Lead_Platform_Web                         25291 non-null float64
Email_Subscriber_No                       25291 non-null float64
Email_Subscriber_Yes                      25291 non-null float64
Push_Notifications_No                     25291 non-null float64
Push_Notifications_Yes                    25291 non-null float64
PCA_X                                     25291 non-null float64
PCA_Y                                     25291 non-null float64
PCA1                                      25291 non-null float64
PCA2                                      25291 non-null float64
PCA3                                      25291 non-null float64
PCA4                                      25291 non-null float64
PCA5                                      25291 non-null float64
PCA6                                      25291 non-null float64
PCA7                                      25291 non-null float64
PCA8                                      25291 non-null float64
PCA9                                      25291 non-null float64
PCA10                                     25291 non-null float64
PCA11                                     25291 non-null float64
PCA12                                     25291 non-null float64
PCA13                                     25291 non-null float64
Language                                  25291 non-null object
Subscription_Type                         25291 non-null object
Subscription_Event_Type                   25291 non-null object
Purchase_Store                            25291 non-null object
Demo_User                                 25291 non-null object
Free_Trial_User                           25291 non-null object
Auto_Renew                                25291 non-null object
Country                                   25291 non-null object
User_Type                                 25291 non-null object
Lead_Platform                             25291 non-null object
Email_Subscriber                          25291 non-null object
Push_Notifications                        25291 non-null object
Customer_Type_All                         25291 non-null int32
Customer_Type_PCA                         25291 non-null int32
dtypes: float64(46), int32(2), int64(13), object(12)
memory usage: 13.9+ MB
```

In [26]:
```
web['Customer_Type_All'] = web['Customer_Type_All'].astype(str)
web['Customer_Type_PCA'] = web['Customer_Type_PCA'].astype(str)
```

In [27]:
```
clean = web
clean.drop(clean.iloc[:,14:59], inplace = True, axis = 1)
```

In [28]:
```
clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25291 entries, 0 to 25290
Data columns (total 28 columns):
Purchase_Amounts           25291 non-null float64
Duration                   25291 non-null int64
Send_Count                 25291 non-null int64
Open_Count                 25291 non-null int64
Click_Count                25291 non-null int64
Unique_Open_Count          25291 non-null int64
Unique_Click_Count         25291 non-null int64
Start                      25291 non-null int64
Other                      25291 non-null int64
Completed                  25291 non-null int64
NULL                       25291 non-null int64
Onboarding                 25291 non-null int64
App_Launch_Times           25291 non-null int64
Total_Time_Launched        25291 non-null int64
Language                   25291 non-null object
Subscription_Type          25291 non-null object
Subscription_Event_Type    25291 non-null object
Purchase_Store             25291 non-null object
Demo_User                  25291 non-null object
Free_Trial_User            25291 non-null object
Auto_Renew                 25291 non-null object
Country                    25291 non-null object
User_Type                  25291 non-null object
Lead_Platform              25291 non-null object
Email_Subscriber           25291 non-null object
Push_Notifications         25291 non-null object
Customer_Type_All          25291 non-null object
Customer_Type_PCA          25291 non-null object
dtypes: float64(1), int64(13), object(14)
memory usage: 5.4+ MB
```

In [29]:
```
clean.to_csv('Subscriber Information (Clean) Version 6.csv', index = False)
```

```
---------------------------------------------------------------------------
OSError                                   Traceback (most recent call last)
<ipython-input-29-f506dbe25b75> in <module>
----> 1 clean.to_csv('Subscriber Information (Clean) Version 6.csv', index = False)

~\Anaconda3\lib\site-packages\pandas\core\generic.py in to_csv(self, path_or_buf, sep, na_rep, float_format, columns, header, index, index_label, mode, encoding, compression, quoting, quotechar, line_terminator, chunksize, tupleize_cols, date_format, doublequote, escapechar, decimal)
   3018                                      doublequote=doublequote,
   3019                                      escapechar=escapechar, decimal=decimal)
-> 3020         formatter.save()
   3021
   3022         if path_or_buf is None:

~\Anaconda3\lib\site-packages\pandas\io\formats\csvs.py in save(self)
    170                 self.writer = UnicodeWriter(f, **writer_kwargs)
    171
--> 172             self._save()
    173
    174         finally:

~\Anaconda3\lib\site-packages\pandas\io\formats\csvs.py in _save(self)
    286                 break
    287
--> 288             self._save_chunk(start_i, end_i)
    289
    290     def _save_chunk(self, start_i, end_i):

~\Anaconda3\lib\site-packages\pandas\io\formats\csvs.py in _save_chunk(self, start_i, end_i)
    313
    314         libwriters.write_csv_rows(self.data, ix, self.nlevels,
--> 315                                   self.cols, self.writer)

pandas/_libs/writers.pyx in pandas._libs.writers.write_csv_rows()

OSError: [Errno 22] Invalid argument
```

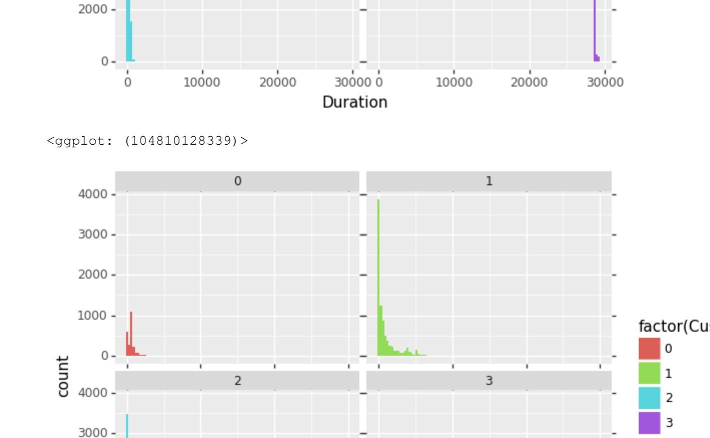
Graphing out each segment

Numerical

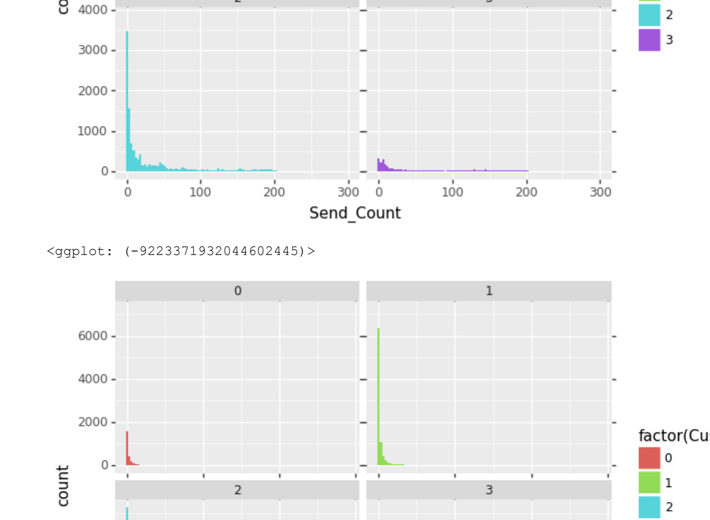
In [30]:
```
Numerical = ['Purchase_Amounts', 'Duration', 'Send_Count', 'Open_Count', 'Click_Count', 'Unique_Open_Count', 'Unique_Click_Count', 'Start', 'Other', 'Completed', 'NULL', 'Onboarding', 'App_Launch_Times', 'Total_Time_Launched']
for i in Numerical:
    print(ggplot(clean, aes(x = i, fill = 'factor(Customer_Type_PCA)')) + geom_histogram(bins=100) + facet_wrap('Customer_Type_PCA'))
```

```
<ggplot: (-9223371932045133859)>
```

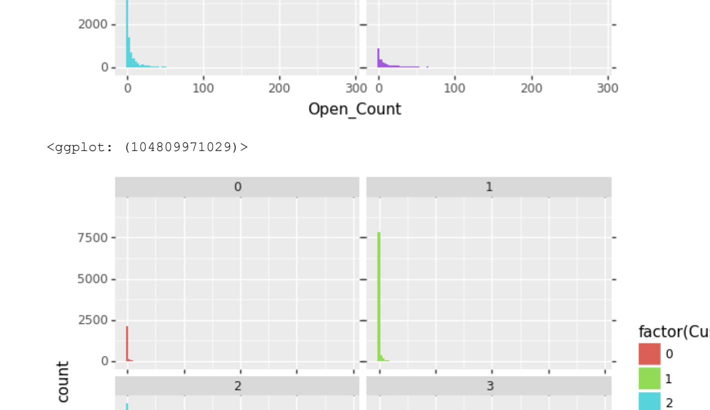
```
<ggplot: (104810128339)>
```



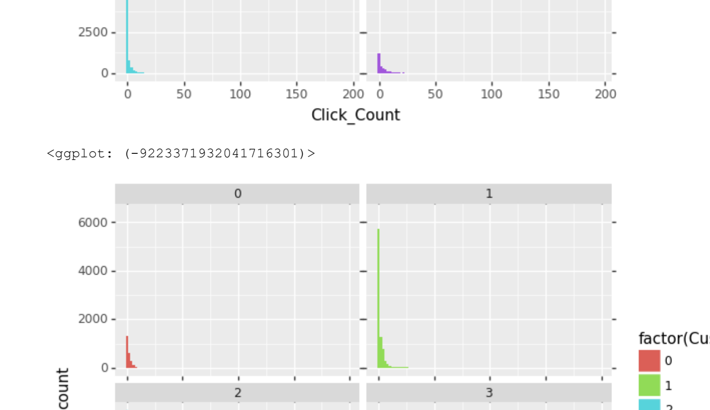
```
<ggplot: (-9223371932044602445)>
```



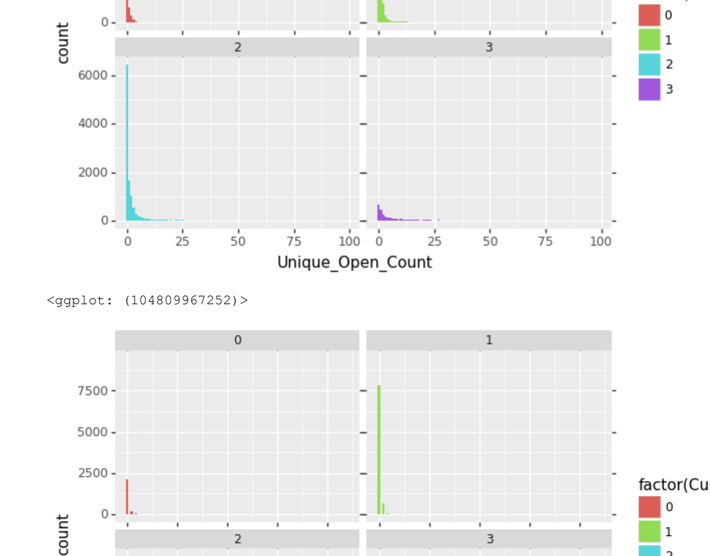
```
<ggplot: (104809971029)>
```



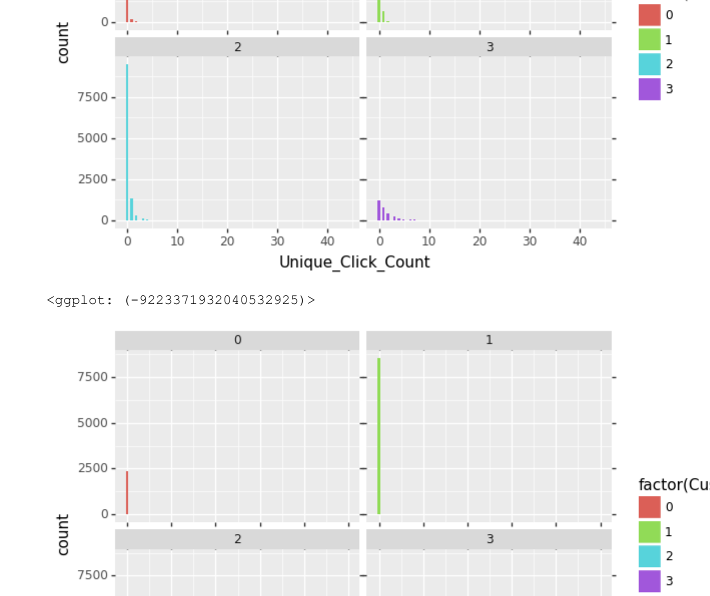
```
<ggplot: (-9223371932041716301)>
```



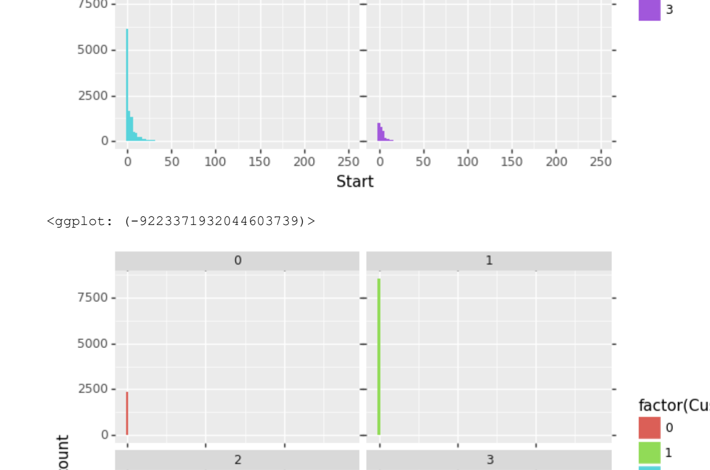
```
<ggplot: (104809967252)>
```



```
<ggplot: (-9223371932040532925)>
```

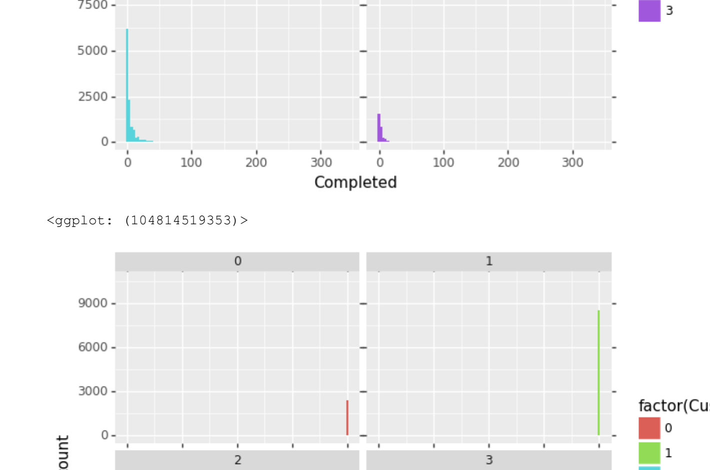


```
<ggplot: (-9223371932044603739)>
```

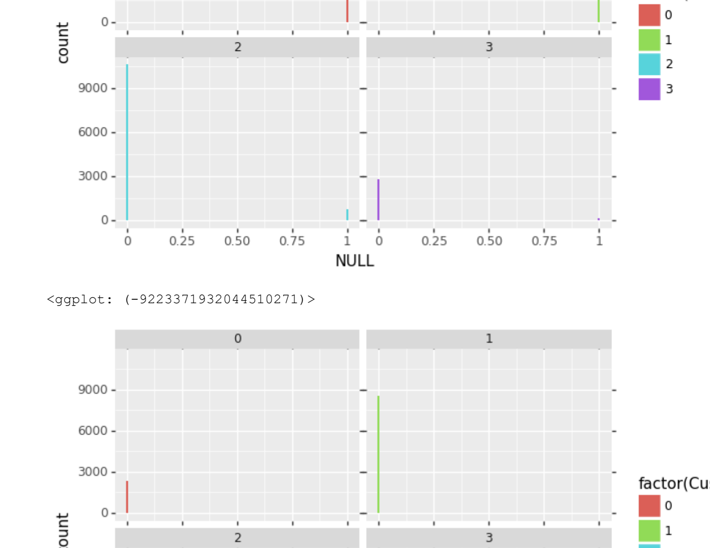


```
<ggplot: (-9223371932044829338)>
```

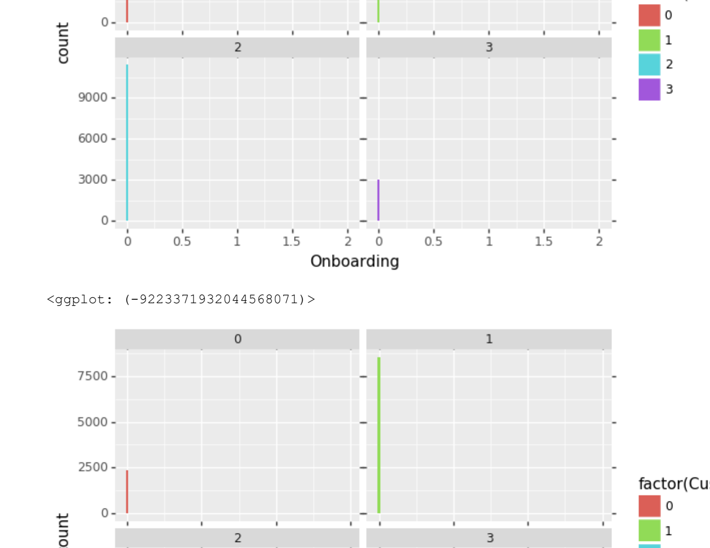
```
<ggplot: (104814519353)>
```



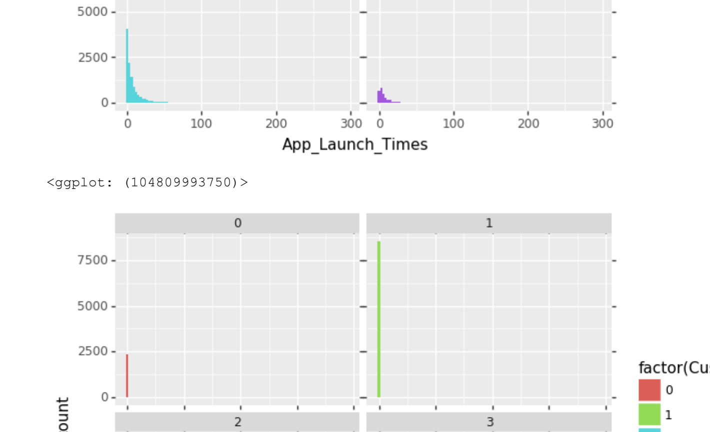
```
<ggplot: (-9223371932044510271)>
```



```
<ggplot: (-9223371932044568071)>
```

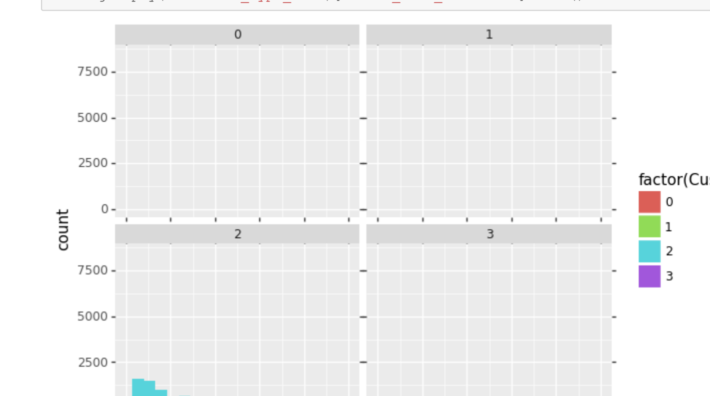


```
<ggplot: (104809993750)>
```



```
<ggplot: (-9223371932040596641)>
```

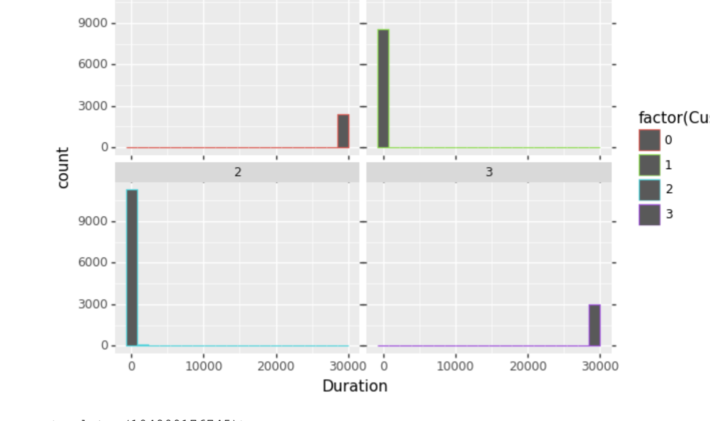
In [73]:
```
print(ggplot(clean,aes(x = 'Total_Time_Launched', fill = 'factor(Customer_Type_PCA)')) + geom_histogram(bins = 20) + facet_wrap('Customer_Type_PCA') + xlim(0,50))
clean.groupby('Customer_Type_PCA')['Total_Time_Launched'].mean()
```



```
<ggplot: (104814160504)>
```

Out[73]:
```
Customer_Type_PCA
0     1.000000
1     1.000000
2    26.850653
3    22.386288
Name: Total_Time_Launched, dtype: float64
```

In [63]:
```
print(ggplot(clean,aes(x = 'Duration', color = 'factor(Customer_Type_PCA)')) + geom_histogram(bins = 20) + facet_wrap('Customer_Type_PCA'))
clean.groupby('Customer_Type_PCA')['Duration'].mean()
```



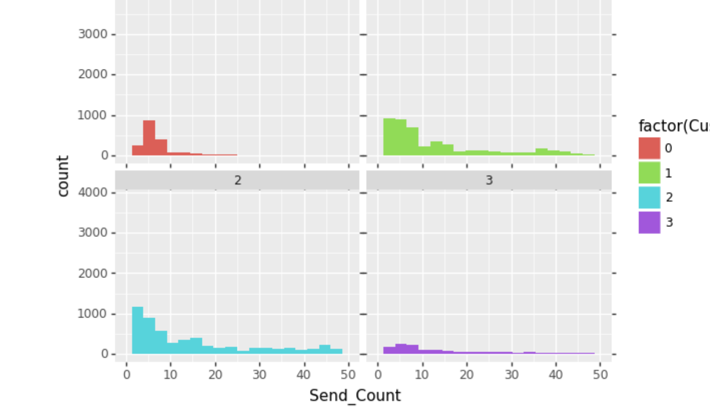
```
<ggplot: (104809176745)>
```

Out[63]:
```
Customer_Type_PCA
0    28779.715013
1      257.465457
2      359.312374
3    28596.883612
Name: Duration, dtype: float64
```

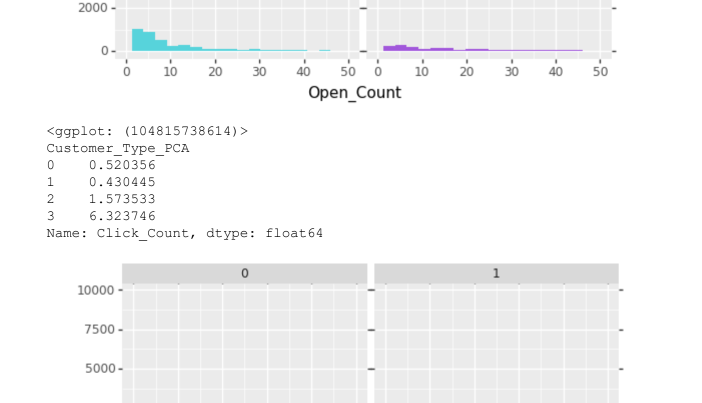
Email

In [70]:
```
email = ['Send_Count', 'Open_Count', 'Click_Count', 'Unique_Open_Count', 'Unique_Click_Count']
for i in email:
    print(clean.groupby('Customer_Type_PCA')[i].mean())
    print(ggplot(clean, aes(x = i, fill = 'factor(Customer_Type_PCA)')) + geom_histogram(bins = 20) + facet_wrap('Customer_Type_PCA') + xlim(0,50))
```

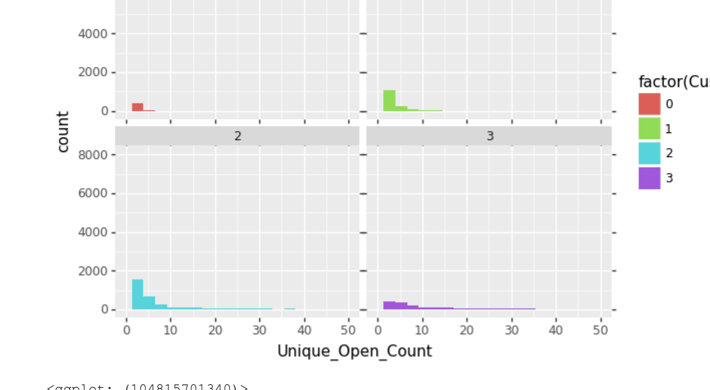
```
Customer_Type_PCA
0     5.782443
1     8.696019
2    33.431641
3    66.821070
Name: Send_Count, dtype: float64
```



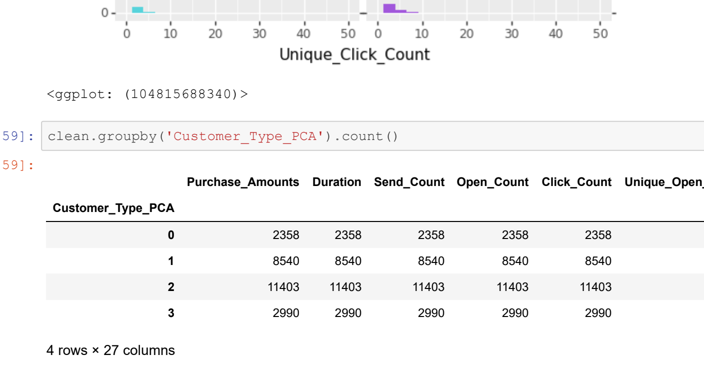
```
<ggplot: (104813103273)>
Customer_Type_PCA
0     2.336302
1     2.313700
2     5.915812
3    20.887625
Name: Open_Count, dtype: float64
```



```
<ggplot: (104815738614)>
Customer_Type_PCA
0    0.520356
1    0.430445
2    1.573533
3    6.323746
Name: Click_Count, dtype: float64
```



```
<ggplot: (104814635121)>
Customer_Type_PCA
0    0.821035
1    0.951171
2    2.818030
3    9.986957
Name: Unique_Open_Count, dtype: float64
```



```
<ggplot: (104815701340)>
Customer_Type_PCA
0    0.111535
1    0.089461
2    0.283785
3    1.465552
Name: Unique_Click_Count, dtype: float64
```



```
<ggplot: (104815688340)>
```

In [59]:
```
clean.groupby('Customer_Type_PCA').count()
```

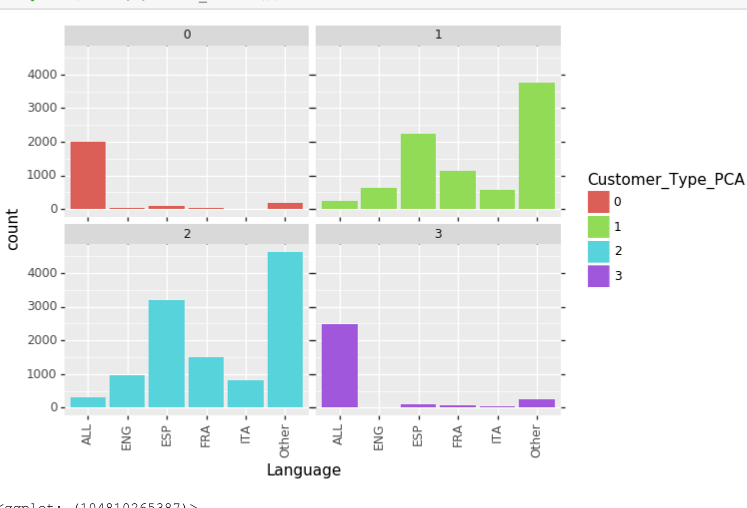
Out[59]:

	Purchase_Amounts	Duration	Send_Count	Open_Count	Click_Count	Unique_Open_Count	Unique_Click_Count	St
Customer_Type_PCA								
0	2358	2358	2358	2358	2358	2358	2358	23
1	8540	8540	8540	8540	8540	8540	8540	85
2	11403	11403	11403	11403	11403	11403	11403	114
3	2990	2990	2990	2990	2990	2990	2990	29

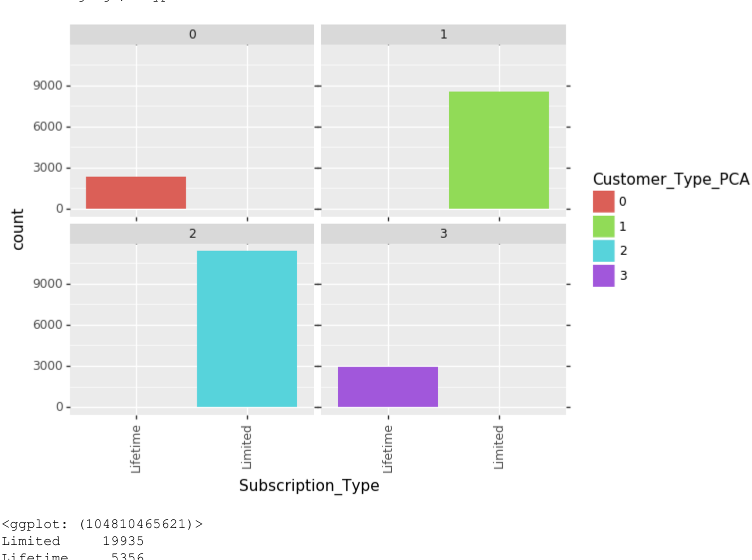
4 rows × 27 columns

Categorical

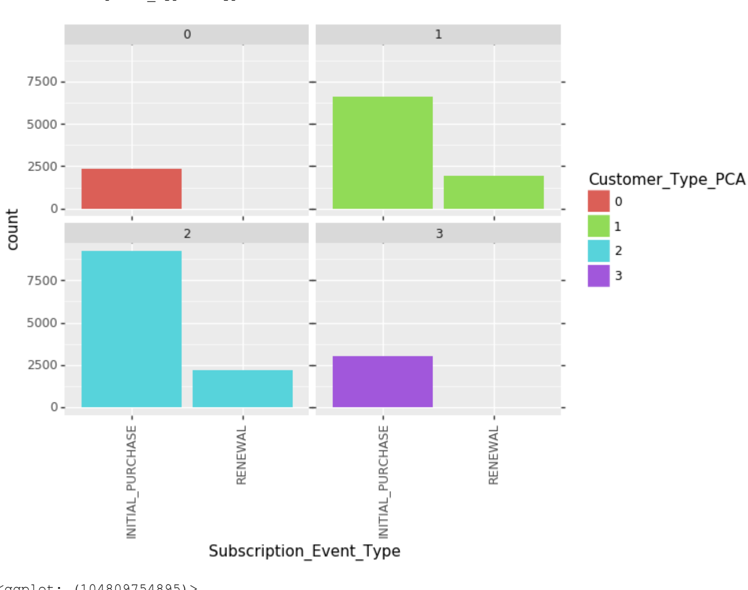
```
In [31]: categorical_columns = ['Language', 'Subscription_Type', 'Subscription_Event_Type', 'Purchase_Store', 'Demo_User', 'Free_Trial_User', 'Auto_Renew', 'Country', 'User_Type', 'Lead_Platform', 'Email_Subscriber', 'Push_Notifications']
for i in categorical_columns:
    print(ggplot(clean, aes(x = i, fill = 'Customer_Type_PCA')) + geom_bar() + theme(axis_text_x=element_text(angle = 90, hjust=2)) + facet_wrap('Customer_Type_PCA'))
    print(clean[i].value_counts())
```



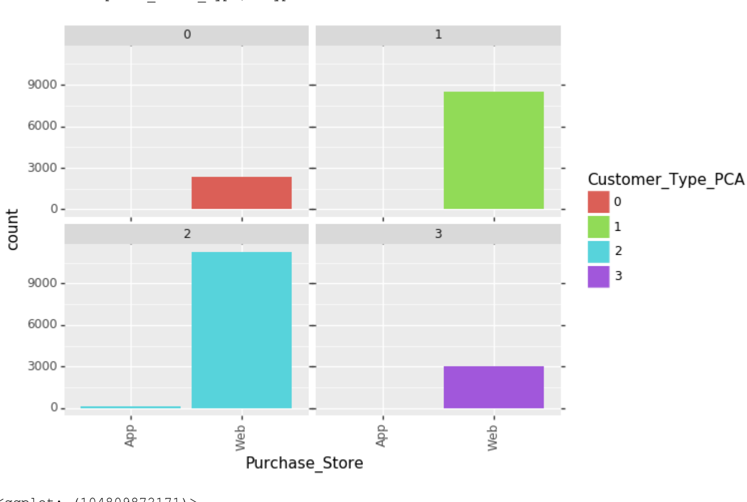

```
<ggplot: (104810265387)>
Other    8837
ESP      5627
ALL      5030
FRA      2745
ENG      1640
ITA      1412
Name: Language, dtype: int64
```



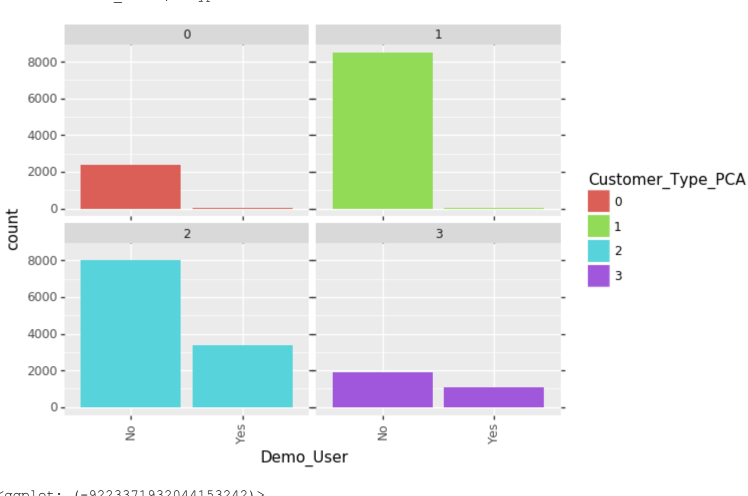

```
<ggplot: (104810465621)>
Limited     19935
Lifetime     5356
Name: Subscription_Type, dtype: int64
```



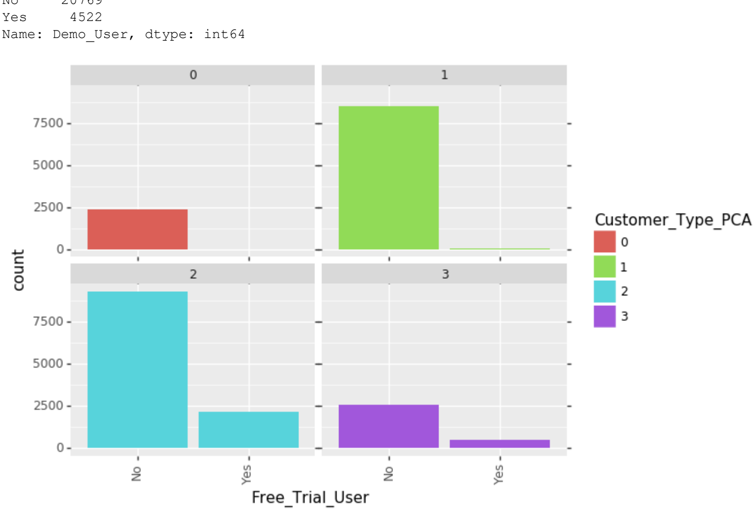

```
<ggplot: (104809754895)>
INITIAL_PURCHASE    21191
RENEWAL              4100
Name: Subscription_Event_Type, dtype: int64
```



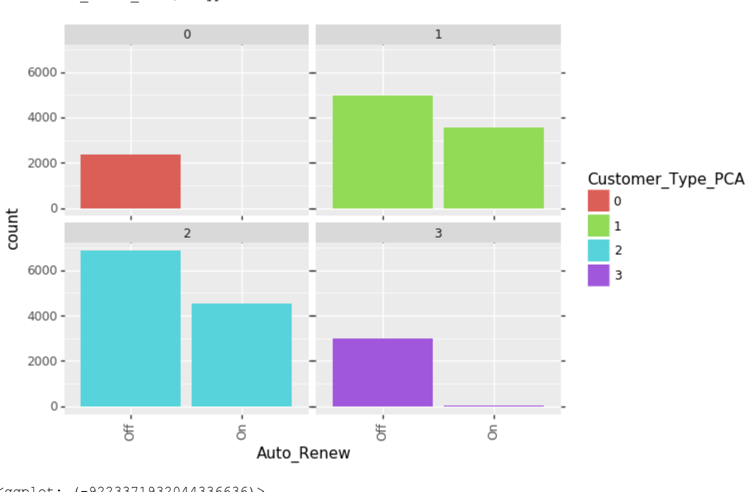

```
<ggplot: (104809872171)>
Web    25154
App      137
Name: Purchase_Store, dtype: int64
```



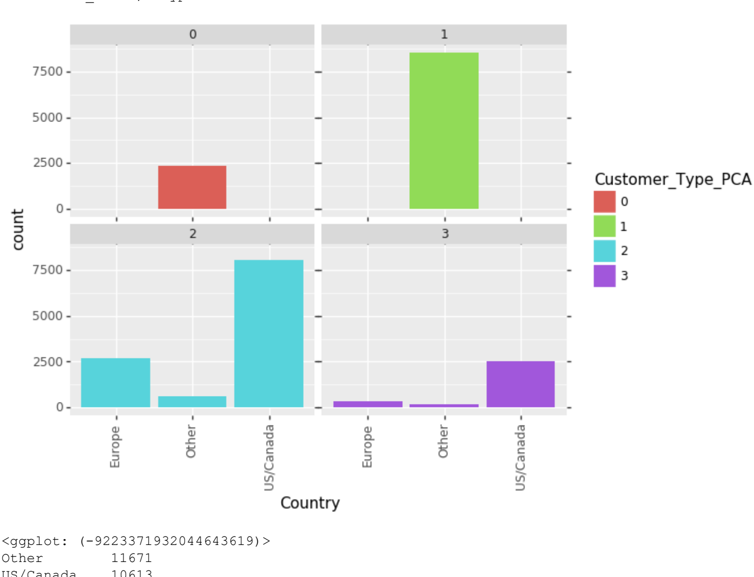

```
<ggplot: (-9223371932044153242)>
No     20769
Yes     4522
Name: Demo_User, dtype: int64
```



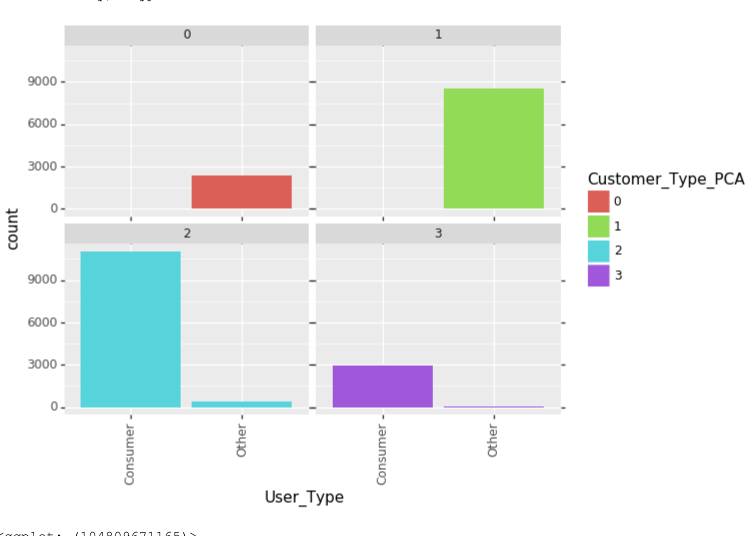

```
<ggplot: (-9223371932044802507)>
No     22717
Yes     2574
Name: Free_Trial_User, dtype: int64
```



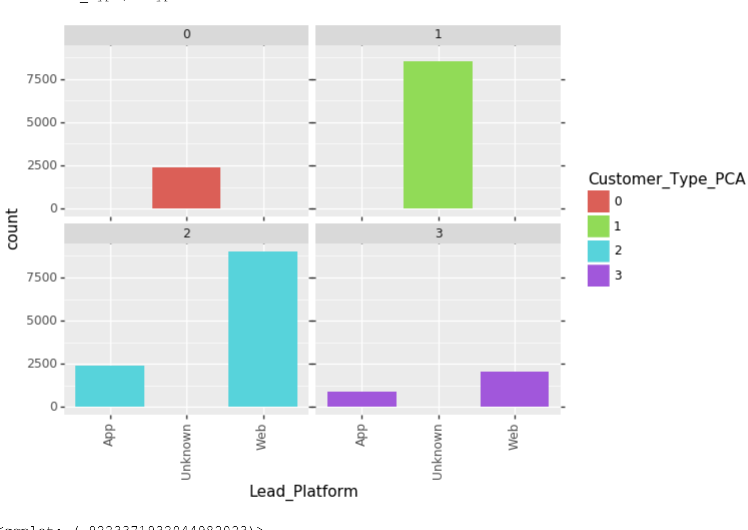

```
<ggplot: (-9223371932044336636)>
Off    17168
On      8123
Name: Auto_Renew, dtype: int64
```



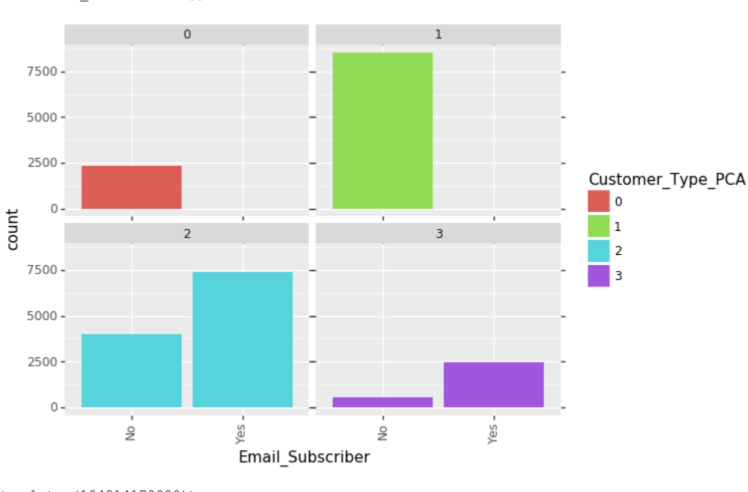

```
<ggplot: (-9223371932044643619)>
Other        11671
US/Canada    10613
Europe        3007
Name: Country, dtype: int64
```



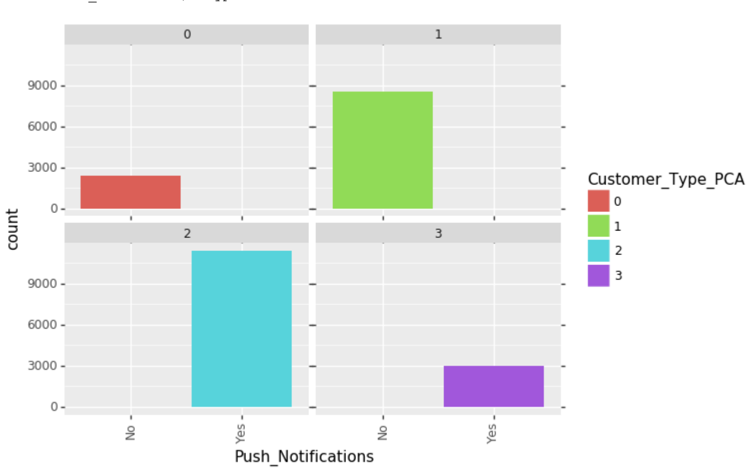

```
<ggplot: (104809671165)>
Consumer    13978
Other       11313
Name: User_Type, dtype: int64
```




```
<ggplot: (-9223371932044982023)>
Web        11105
Unknown    10900
App         3286
Name: Lead_Platform, dtype: int64
```




```
<ggplot: (104814179020)>
No     15409
Yes     9882
Name: Email_Subscriber, dtype: int64
```



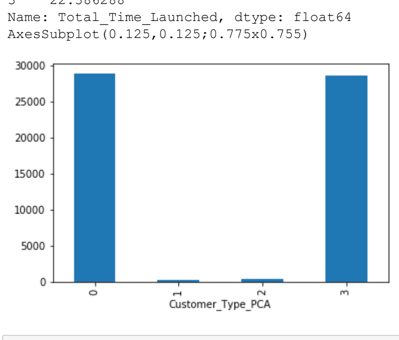

```
<ggplot: (-9223371932044662149)>
Yes    14389
No     10902
Name: Push_Notifications, dtype: int64
```

```
In [71]: print(ggplot(clean, aes(x = 'Customer_Type_PCA', fill = 'Customer_Type_PCA')) + geom_bar() + theme(axis_text_x=element_text(angle = 90, hjust=2)))
clean['Customer_Type_PCA'].value_counts()
```


```
<ggplot: (-9223371932037903644)>

Out[71]: 2    11403
1     8540
3     2990
0     2358
Name: Customer_Type_PCA, dtype: int64
```

```
In [33]: for i in Numerical:
    plot = clean.groupby('Customer_Type_PCA')[i].mean()
    print('---------------------'+i+'------------------------------')
    print(plot)
    print(clean.groupby('Customer_Type_PCA')[i].mean().plot(kind = 'bar'))
```

```
---------------------Purchase_Amounts------------------------------
Customer_Type_PCA
0    194.041968
1     40.825678
2     36.107986
3    185.865973
Name: Purchase_Amounts, dtype: float64
AxesSubplot(0.125,0.125;0.775x0.755)
---------------------Duration------------------------------
Customer_Type_PCA
0    28779.715013
1      257.465457
2      359.312374
3    28596.883612
Name: Duration, dtype: float64
AxesSubplot(0.125,0.125;0.775x0.755)
---------------------Send_Count------------------------------
Customer_Type_PCA
0     5.782443
1     8.696019
2    33.431641
3    66.821070
Name: Send_Count, dtype: float64
AxesSubplot(0.125,0.125;0.775x0.755)
---------------------Open_Count------------------------------
Customer_Type_PCA
0     2.336302
1     2.313700
2     5.915812
3    20.887625
Name: Open_Count, dtype: float64
AxesSubplot(0.125,0.125;0.775x0.755)
---------------------Click_Count------------------------------
Customer_Type_PCA
0    0.520356
1    0.430445
2    1.573533
3    6.323746
Name: Click_Count, dtype: float64
AxesSubplot(0.125,0.125;0.775x0.755)
---------------------Unique_Open_Count------------------------------
Customer_Type_PCA
0    0.821035
1    0.951171
2    2.818030
3    9.986957
Name: Unique_Open_Count, dtype: float64
AxesSubplot(0.125,0.125;0.775x0.755)
---------------------Unique_Click_Count------------------------------
Customer_Type_PCA
0    0.111535
1    0.089461
2    0.283785
3    1.465552
Name: Unique_Click_Count, dtype: float64
AxesSubplot(0.125,0.125;0.775x0.755)
---------------------Start------------------------------
Customer_Type_PCA
0    0.000000
1    0.000000
2    5.173288
3    4.759197
Name: Start, dtype: float64
AxesSubplot(0.125,0.125;0.775x0.755)
---------------------Other------------------------------
Customer_Type_PCA
0    0.000000
1    0.000000
2    6.570289
3    5.454515
Name: Other, dtype: float64
AxesSubplot(0.125,0.125;0.775x0.755)
---------------------Completed------------------------------
Customer_Type_PCA
0    0.000000
1    0.000000
2    5.764273
3    4.110368
Name: Completed, dtype: float64
AxesSubplot(0.125,0.125;0.775x0.755)
---------------------NULL------------------------------
Customer_Type_PCA
0    1.000000
1    1.000000
2    0.068754
3    0.049498
Name: NULL, dtype: float64
AxesSubplot(0.125,0.125;0.775x0.755)
---------------------Onboarding------------------------------
Customer_Type_PCA
0    0.000000
1    0.000000
2    0.001315
3    0.000000
Name: Onboarding, dtype: float64
AxesSubplot(0.125,0.125;0.775x0.755)
---------------------App_Launch_Times------------------------------
Customer_Type_PCA
0    0.000000
1    0.000000
2    9.272735
3    8.012709
Name: App_Launch_Times, dtype: float64
AxesSubplot(0.125,0.125;0.775x0.755)
---------------------Total_Time_Launched------------------------------
Customer_Type_PCA
0     1.000000
1     1.000000
2    26.850653
3    22.386288
Name: Total_Time_Launched, dtype: float64
AxesSubplot(0.125,0.125;0.775x0.755)
```




```
In [40]: plot = clean.groupby('Customer_Type_PCA')['Purchase_Amounts'].mean()
plot

Out[40]: Customer_Type_PCA
0    194.041968
1     40.825678
2     36.107986
3    185.865973
Name: Purchase_Amounts, dtype: float64
```

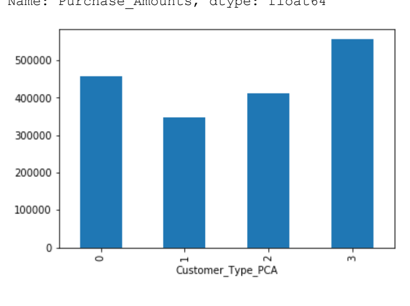
Graphing out each column

Numerical

- ID
- Purchase_Amounts
- Duration
- Send_Count
- Open_Count
- Click_Count
- Unique_Open_Count
- Unique_Click_Count
- Start
- Other
- Completed
- NULL
- Onboarding
- App_Launch_Times
- Total_Time_Launched

Purchase Amounts

```
In [46]: clean.groupby('Customer_Type_PCA')['Purchase_Amounts'].sum().plot(kind = 'bar')

sum = clean.groupby('Customer_Type_PCA')['Purchase_Amounts'].sum()
sum

Out[46]: Customer_Type_PCA
0    457550.96
1    348651.29
2    411739.36
3    555739.26
Name: Purchase_Amounts, dtype: float64
```



Purchase Amounts here we see that Segment 3 has the highest total revenue from all three segments which is important to note. Then we have segment 0, 2, then 1 trailing after. The lowest we see here is segment 1.

```
In [75]: clean.groupby('Customer_Type_PCA')['Purchase_Amounts'].mean().plot(kind = 'bar')
clean.groupby('Customer_Type_PCA')['Purchase_Amounts'].mean()

Out[75]: Customer_Type_PCA
0    194.041968
1     40.825678
2     36.107986
3    185.865973
Name: Purchase_Amounts, dtype: float64
```

We see here that there that the average of our segments are very different from each other. Segments 1 and 3 are high at around $190averagepercustomerwhiletheothertwosegmentsarearound$38 per customer.

```
In [43]: segments = clean.groupby('Customer_Type_PCA')

segments_Descriptions = segments.describe()
```

```
In [44]: for i in segments_Descriptions.columns:
    print(i)
    #print(ggplot(segments_Descriptions, aes(x = i)) + geom_bar())
```

```
('Purchase_Amounts', 'count')
('Purchase_Amounts', 'mean')
('Purchase_Amounts', 'std')
('Purchase_Amounts', 'min')
('Purchase_Amounts', '25%')
('Purchase_Amounts', '50%')
('Purchase_Amounts', '75%')
('Purchase_Amounts', 'max')
('Duration', 'count')
('Duration', 'mean')
('Duration', 'std')
('Duration', 'min')
('Duration', '25%')
('Duration', '50%')
('Duration', '75%')
('Duration', 'max')
('Send_Count', 'count')
('Send_Count', 'mean')
('Send_Count', 'std')
('Send_Count', 'min')
('Send_Count', '25%')
('Send_Count', '50%')
('Send_Count', '75%')
('Send_Count', 'max')
('Open_Count', 'count')
('Open_Count', 'mean')
('Open_Count', 'std')
('Open_Count', 'min')
('Open_Count', '25%')
('Open_Count', '50%')
('Open_Count', '75%')
('Open_Count', 'max')
('Click_Count', 'count')
('Click_Count', 'mean')
('Click_Count', 'std')
('Click_Count', 'min')
('Click_Count', '25%')
('Click_Count', '50%')
('Click_Count', '75%')
('Click_Count', 'max')
('Unique_Open_Count', 'count')
('Unique_Open_Count', 'mean')
('Unique_Open_Count', 'std')
('Unique_Open_Count', 'min')
('Unique_Open_Count', '25%')
('Unique_Open_Count', '50%')
('Unique_Open_Count', '75%')
('Unique_Open_Count', 'max')
('Unique_Click_Count', 'count')
('Unique_Click_Count', 'mean')
('Unique_Click_Count', 'std')
('Unique_Click_Count', 'min')
('Unique_Click_Count', '25%')
('Unique_Click_Count', '50%')
('Unique_Click_Count', '75%')
('Unique_Click_Count', 'max')
('Start', 'count')
('Start', 'mean')
('Start', 'std')
('Start', 'min')
('Start', '25%')
('Start', '50%')
('Start', '75%')
('Start', 'max')
('Other', 'count')
('Other', 'mean')
('Other', 'std')
('Other', 'min')
('Other', '25%')
('Other', '50%')
('Other', '75%')
('Other', 'max')
('Completed', 'count')
('Completed', 'mean')
('Completed', 'std')
('Completed', 'min')
('Completed', '25%')
('Completed', '50%')
('Completed', '75%')
('Completed', 'max')
('NULL', 'count')
('NULL', 'mean')
('NULL', 'std')
('NULL', 'min')
('NULL', '25%')
('NULL', '50%')
('NULL', '75%')
('NULL', 'max')
('Onboarding', 'count')
('Onboarding', 'mean')
('Onboarding', 'std')
('Onboarding', 'min')
('Onboarding', '25%')
('Onboarding', '50%')
('Onboarding', '75%')
('Onboarding', 'max')
('App_Launch_Times', 'count')
('App_Launch_Times', 'mean')
('App_Launch_Times', 'std')
('App_Launch_Times', 'min')
('App_Launch_Times', '25%')
('App_Launch_Times', '50%')
('App_Launch_Times', '75%')
('App_Launch_Times', 'max')
('Total_Time_Launched', 'count')
('Total_Time_Launched', 'mean')
('Total_Time_Launched', 'std')
('Total_Time_Launched', 'min')
('Total_Time_Launched', '25%')
('Total_Time_Launched', '50%')
('Total_Time_Launched', '75%')
('Total_Time_Launched', 'max')
```

```
In [39]: segments_Descriptions.to_csv('segments.csv', index = False)

---------------------------------------------------------------------------
OSError                                   Traceback (most recent call last)
<ipython-input-39-360dd67f3d0e> in <module>
----> 1 segments_Descriptions.to_csv('segments.csv', index = False)

~\Anaconda3\lib\site-packages\pandas\core\generic.py in to_csv(self, path_or_buf, sep, na_rep, float_format, columns, header, index, index_label, mode, encoding, compression, quoting, quotechar, line_terminator, chunksize, tupleize_cols, date_format, doublequote, escapechar, decimal)
   3018                                  doublequote=doublequote,
   3019                                  escapechar=escapechar, decimal=decimal)
-> 3020         formatter.save()
   3021
   3022         if path_or_buf is None:

~\Anaconda3\lib\site-packages\pandas\io\formats\csvs.py in save(self)
    185                 close = True
    186             if close:
--> 187                 f.close()
    188                 for _fh in handles:
    189                     _fh.close()

OSError: [Errno 22] Invalid argument
```

```
In [ ]:
```