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

In [2]:
```
df = pd.read_csv('/content/drive/MyDrive/2022F ㅂ□ㄱㅋ□ㄴㅌ□ㅅ□ㄷ□/Data/log_data.csv')
df
```

Out[2]:

	user_id	event	timestamp	mp_os	mp_app_version	date_cd
0	576409	StartLoanApply	2022-03-25 11:12:09	Android	3.8.2	2022-03-25
1	576409	ViewLoanApplyIntro	2022-03-25 11:12:09	Android	3.8.2	2022-03-25
2	72878	EndLoanApply	2022-03-25 11:14:44	Android	3.8.4	2022-03-25
3	645317	OpenApp	2022-03-25 11:15:09	iOS	3.6.1	2022-03-25
4	645317	UseLoanManage	2022-03-25 11:15:11	iOS	3.6.1	2022-03-25
...	...	...	...	...	...	...
17843988	242442	StartLoanApply	2022-05-27 16:08:21	Android	3.13.0	2022-05-27
17843989	242442	EndLoanApply	2022-05-27 16:08:21	Android	3.13.0	2022-05-27
17843990	593062	OpenApp	2022-05-27 16:19:10	Android	3.13.0	2022-05-27
17843991	593062	Login	2022-05-27 16:19:15	Android	3.13.0	2022-05-27
17843992	593062	UseLoanManage	2022-05-27 16:19:19	Android	3.13.0	2022-05-27

17843993 rows × 6 columns

In [3]:
```
df.columns
```

Out[3]:
```
Index(['user_id', 'event', 'timestamp', 'mp_os', 'mp_app_version', 'date_cd'], dtype='object')
```

In [4]:
```
df = df[['user_id', 'event']]
OHE = pd.get_dummies(df['event'])
```

In [5]:
```
OHE
```

Out[5]:

	CompleteIDCertification	EndLoanApply	GetCreditInfo	Login	OpenApp	SignUp	StartLoanApply	UseDSRCalc	UseLoanManage	Us
0	0	0	0	0	0	0	1	0	0	
1	0	0	0	0	0	0	0	0	0	
2	0	1	0	0	0	0	0	0	0	
3	0	0	0	0	1	0	0	0	0	
4	0	0	0	0	0	0	0	0	1	
...	...	...	...	...	...	...	...	...	...	
17843988	0	0	0	0	0	0	1	0	0	
17843989	0	1	0	0	0	0	0	0	0	
17843990	0	0	0	0	1	0	0	0	0	
17843991	0	0	0	1	0	0	0	0	0	
17843992	0	0	0	0	0	0	0	0	1	

17843993 rows × 11 columns

In [6]:
```
df = pd.concat([df, OHE], axis = 1)
```

In [7]:
```
df
```

Out[7]:

	user_id	event	CompleteIDCertification	EndLoanApply	GetCreditInfo	Login	OpenApp	SignUp	StartLoanApply	UseD
0	576409	StartLoanApply	0	0	0	0	0	0	1	
1	576409	ViewLoanApplyIntro	0	0	0	0	0	0	0	
2	72878	EndLoanApply	0	1	0	0	0	0	0	
3	645317	OpenApp	0	0	0	0	1	0	0	
4	645317	UseLoanManage	0	0	0	0	0	0	0	
...	...	...	...	...	...	...	...	...	...	
17843988	242442	StartLoanApply	0	0	0	0	0	0	1	
17843989	242442	EndLoanApply	0	1	0	0	0	0	0	
17843990	593062	OpenApp	0	0	0	0	1	0	0	
17843991	593062	Login	0	0	0	1	0	0	0	
17843992	593062	UseLoanManage	0	0	0	0	0	0	0	

17843993 rows × 13 columns

In [8]:
```
df.drop(['event'], axis = 1, inplace = True)
```

In [10]:
```
df = df.groupby('user_id').sum()
df
```

Out[10]:

user_id	CompleteIDCertification	EndLoanApply	GetCreditInfo	Login	OpenApp	SignUp	StartLoanApply	UseDSRCalc	UseLoanManage	Usel
1	5	5	5	5	5	5	5	5	5	
7	1	1	1	1	1	1	1	1	1	
9	3	3	3	3	3	3	3	3	3	
11	43	43	43	43	43	43	43	43	43	
12	120	120	120	120	120	120	120	120	120	
...	...	...	...	...	...	...	...	...	...	
879693	61	61	61	61	61	61	61	61	61	
879694	6	6	6	6	6	6	6	6	6	
879695	4	4	4	4	4	4	4	4	4	
879696	11	11	11	11	11	11	11	11	11	
879698	3	3	3	3	3	3	3	3	3	

584636 rows × 11 columns

In [ ]:
```
df.to_csv('/content/drive/MyDrive/2022F ㅂ ㄱ ㅋ ㄴ ㄷ ㅅ ㄷ /Data/countbylog.csv', index = False)
```