In [2]:

```
import numpy as np
import pandas as pd
from collections import Counter

import seaborn as sns
sns.set(rc={'figure.figsize':(6,4)})
import matplotlib.pyplot as plt
%matplotlib inline

from tqdm import tqdm
import random
import pickle
import time

from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder

from sklearn.preprocessing import MinMaxScaler
from sklearn.preprocessing import StandardScaler
from sklearn.preprocessing import MaxAbsScaler
from sklearn.preprocessing import RobustScaler
from sklearn.preprocessing import QuantileTransformer
from sklearn.preprocessing import PowerTransformer
from sklearn.preprocessing import Normalizer

from sklearn.linear_model import LogisticRegression
from sklearn.neighbors import KNeighborsClassifier
from sklearn.naive_bayes import GaussianNB
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import RandomForestClassifier

from sklearn.model_selection import cross_val_score
from sklearn.metrics import accuracy_score
from sklearn.metrics import log_loss
from sklearn.metrics import cohen_kappa_score
from sklearn.metrics import confusion_matrix
from sklearn import metrics

# Root Mean Square Error
from sklearn.metrics import mean_squared_error
from math import sqrt

# for ignore warnings
import warnings
warnings.filterwarnings("ignore")
```

In [10]:
```
df = pd.read_csv('dataset\dataset_Software developement methodology recommender system.csv')
df.head()
```

Out[10]:

	Requirements are defined early in the SDLC	Requirements are easily defined and understandable	Requirements are changed frequently	Change Management_in the early process	Limited User Involvement	User Participation in All Phases	No Experience of Participating in Similar Projects	Little Experience of Similar Projects	Little Domain Knowledge_new to Technology	Little experience on Tools	Training Availability when needed	Improvement of an Old System	Stable Funding
0	No	No	No	No	No	No	No	No	No	No	No	No	No
1	No	No	No	No	No	No	No	No	No	No	No	No	No
2	No	No	No	No	No	No	No	No	No	No	No	No	No
3	No	No	No	No	No	No	No	No	No	No	No	No	No
4	No	No	No	No	No	No	No	No	No	No	No	No	No

In [12]:
```
df.columns[:-3]
```

Out[12]:
```
Index(['Requirements are defined early in the SDLC',
       'Requirements are easily defined and understandable',
       'Requirements are changed frequently',
       'Change Management_in the early process', 'Limited User Involvement',
       'User Participation in All Phases',
       'No Experience of Participating in Similar Projects',
       'Little Experience of Similar Projects',
       'Little Domain Knowledge_new to Technology',
       'Little experience on Tools', 'Training Availability when needed',
       'Improvement of an Old System', 'Stable Funding',
       'Highly Reliable Requirements', 'Tight Schedule', 'Reusable Components',
       'Scarce Resources'],
      dtype='object')
```

In [13]:
```
for indx, col_name in enumerate(df.columns[:-3]):
    df.rename(columns = {col_name:f'Q{indx+1}'}, inplace = True)
```

In [14]:
```
df.head()
```

Out[14]:

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Waterfall Model	Prototyping Model	Rapid Application Development
0	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
1	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	Yes	Yes	No	No
2	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	Yes	No	No	Yes	Yes
3	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
4	No	No	No	No	No	No	No	No	No	No	No	No	No	No	Yes	No	No	Yes	No	Yes

In [15]:
```
df = df.sample(frac=1).reset_index(drop=True)
df.head()
```

Out[15]:

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Waterfall Model	Prototyping Model	Rapid Application Development
0	Yes	No	No	Yes	No	Yes	Yes	Yes	No	Yes	No	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes
1	Yes	No	No	No	Yes	Yes	Yes	Yes	No	Yes	Yes	No	Yes	No	Yes	No	No	Yes	Yes	Yes
2	Yes	Yes	Yes	No	No	Yes	Yes	No	No	No	No	No	No	No	Yes	Yes	No	Yes	Yes	Yes
3	Yes	No	No	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
4	Yes	No	No	Yes	Yes	No	Yes	No	No	Yes	Yes	No	No	No	Yes	No	Yes	Yes	Yes	Yes

In [17]:
```
le = LabelEncoder()

for col_name in df.columns:
    df[col_name]= le.fit_transform(df[col_name])
#     no = 0 and yes = 1

df.head()
```

Out[17]:

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Waterfall Model	Prototyping Model	Rapid Application Development
0	1	0	0	1	0	1	1	1	0	1	0	1	1	0	0	1	1	1	1	1
1	1	0	0	0	1	1	1	1	0	1	1	0	1	0	1	0	0	1	1	1
2	1	1	1	0	0	1	1	0	0	0	0	0	0	0	1	1	0	1	1	1
3	1	0	0	1	1	1	0	0	1	1	1	1	1	0	1	1	1	1	1	1
4	1	0	0	1	1	0	1	0	0	1	1	0	0	0	1	0	1	1	1	1

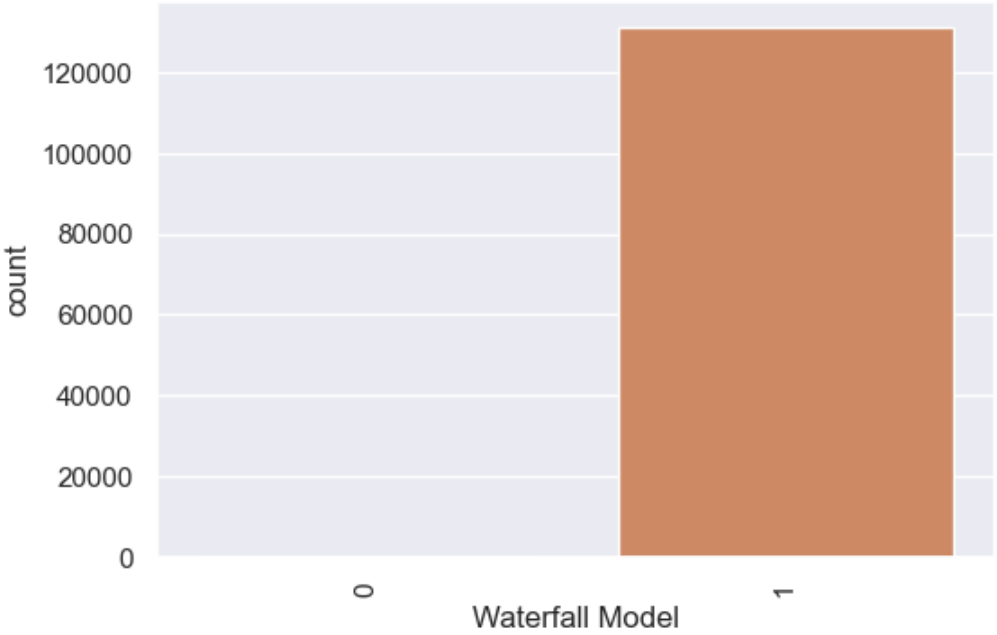
In [23]:
```
round(df.corr(method='pearson'),2)
```

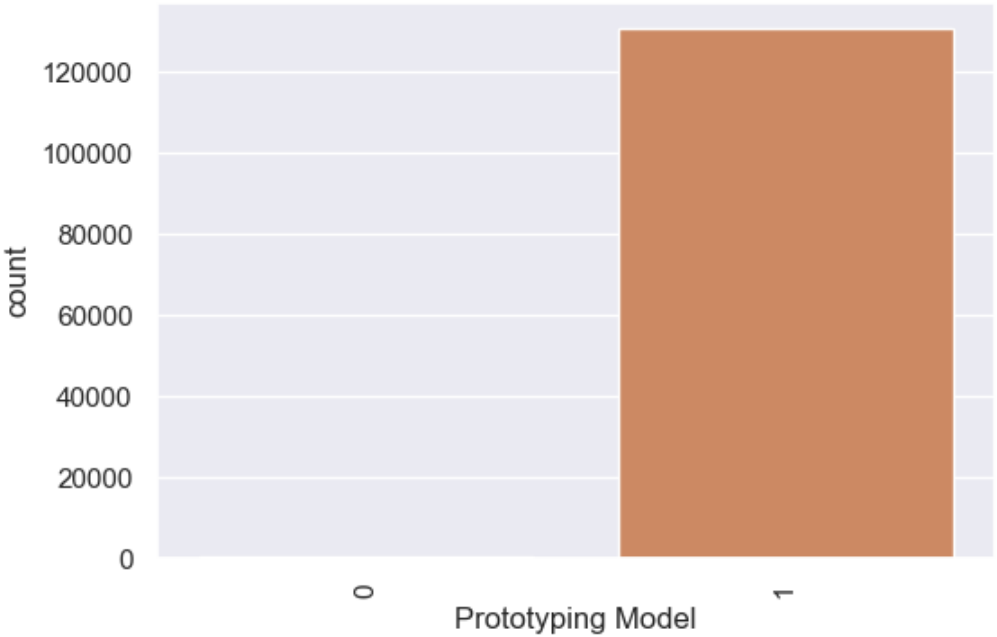
Out[23]:

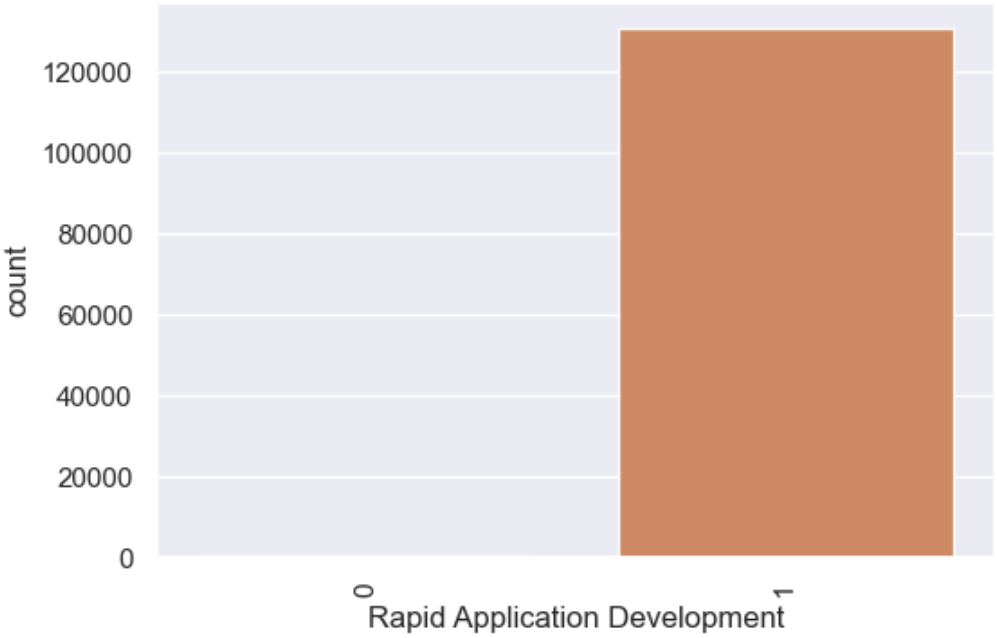
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Waterfall Model	Prototyping Model	Rapid Application Development
Q1	1.00	0.00	0.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00	0.00	-0.00	-0.00	0.00	-0.00	-0.00	-0.00	0.01	0.00	0.00
Q2	0.00	1.00	-0.00	-0.00	0.00	-0.00	0.00	0.00	0.00	0.00	0.00	-0.00	0.00	-0.00	-0.00	-0.00	-0.00	0.01	0.00	0.00
Q3	0.00	-0.00	1.00	0.00	-0.00	-0.00	-0.00	0.00	-0.00	0.00	-0.00	-0.00	-0.00	0.00	0.00	-0.00	-0.00	-0.00	0.06	0.06
Q4	0.00	-0.00	0.00	1.00	-0.00	-0.00	0.00	-0.00	0.00	-0.00	0.00	0.00	0.00	-0.00	-0.00	0.00	0.00	0.01	0.06	0.06
Q5	0.00	0.00	-0.00	-0.00	1.00	0.00	-0.00	0.00	-0.00	-0.00	-0.00	0.00	0.00	-0.00	0.00	-0.00	-0.00	0.01	0.00	0.00
Q6	0.00	-0.00	-0.00	-0.00	0.00	1.00	-0.00	-0.00	0.00	0.00	0.00	0.00	0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.06	0.06
Q7	-0.00	0.00	-0.00	0.00	-0.00	-0.00	1.00	-0.00	0.00	0.00	0.00	-0.00	-0.00	0.00	0.00	0.00	0.00	0.01	-0.00	0.06
Q8	0.00	0.00	0.00	-0.00	0.00	-0.00	-0.00	1.00	-0.00	-0.00	0.00	0.00	0.00	-0.00	-0.00	-0.00	0.00	0.01	0.06	-0.00
Q9	0.00	0.00	-0.00	0.00	-0.00	0.00	0.00	-0.00	1.00	0.00	-0.00	-0.00	0.00	0.00	0.00	0.00	-0.00	0.01	0.06	0.06
Q10	0.00	0.00	0.00	-0.00	-0.00	0.00	0.00	-0.00	0.00	1.00	0.00	-0.00	0.00	0.00	-0.00	0.00	0.00	0.01	0.06	0.00
Q11	0.00	0.00	-0.00	0.00	-0.00	0.00	0.00	0.00	-0.00	0.00	1.00	-0.00	-0.00	0.00	0.00	0.00	-0.00	-0.00	0.06	0.06
Q12	-0.00	-0.00	-0.00	0.00	0.00	0.00	-0.00	0.00	-0.00	-0.00	-0.00	1.00	0.00	0.00	-0.00	-0.00	-0.00	0.01	-0.00	-0.00
Q13	-0.00	0.00	-0.00	0.00	0.00	0.00	-0.00	0.00	0.00	0.00	-0.00	0.00	1.00	0.00	-0.00	-0.00	0.00	0.01	0.00	0.00
Q14	0.00	-0.00	0.00	-0.00	-0.00	-0.00	0.00	-0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	-0.00	0.01	0.00	0.00
Q15	-0.00	-0.00	0.00	-0.00	0.00	-0.00	0.00	-0.00	0.00	-0.00	0.00	-0.00	-0.00	0.00	1.00	0.00	0.00	0.01	-0.00	0.06
Q16	-0.00	-0.00	-0.00	0.00	-0.00	-0.00	0.00	-0.00	0.00	0.00	0.00	-0.00	-0.00	0.00	0.00	1.00	-0.00	0.00	0.06	0.06
Q17	-0.00	-0.00	-0.00	0.00	-0.00	-0.00	0.00	0.00	-0.00	0.00	-0.00	-0.00	0.00	-0.00	0.00	-0.00	1.00	0.01	0.00	0.00
Waterfall Model	0.01	0.01	-0.00	0.01	0.01	-0.00	0.01	0.01	0.01	0.01	-0.00	0.01	0.01	0.01	0.01	0.00	0.01	1.00	0.01	0.01
Prototyping Model	0.00	0.00	0.06	0.06	0.00	0.06	-0.00	0.06	0.06	0.06	0.06	-0.00	0.00	0.00	-0.00	0.06	0.00	0.01	1.00	0.25
Rapid Application Development	0.00	0.00	0.06	0.06	0.00	0.06	0.06	-0.00	0.06	0.00	0.06	-0.00	0.00	0.00	0.06	0.06	0.00	0.01	0.25	1.00

In [26]:

```
# Count Plot: Displays the count of occurrences for each category as bars.
for col in df.columns[-3:]:
    plt.xticks(rotation=90)
    sns.countplot(data=df, x=col)
    plt.show()
```

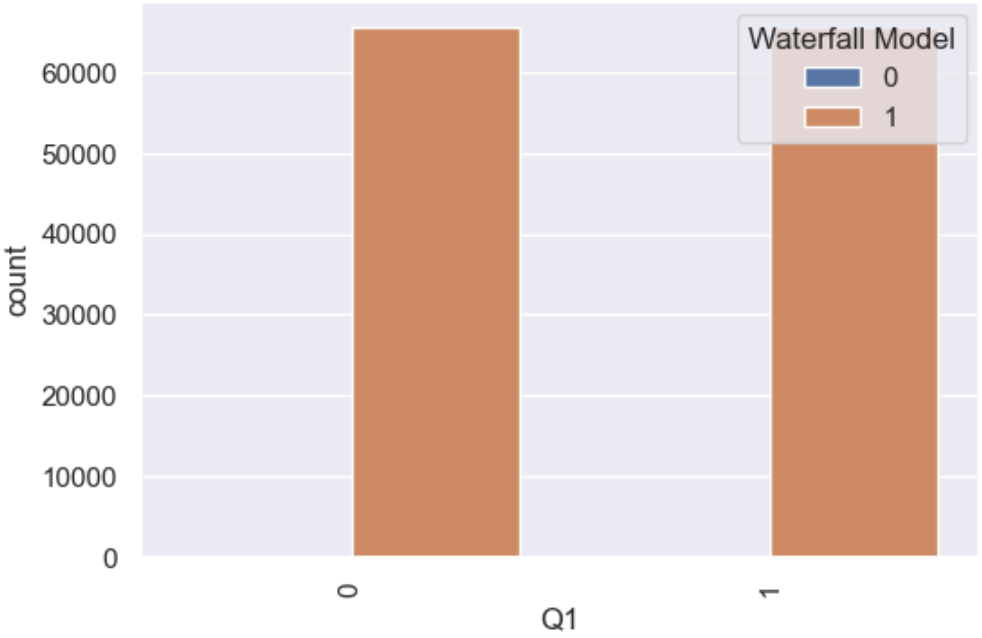


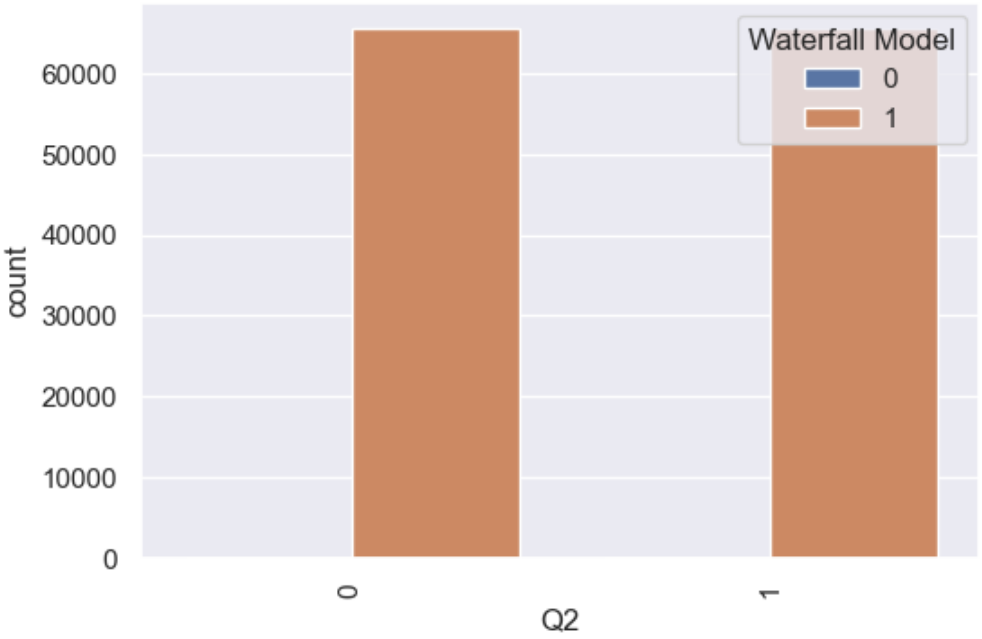

120000
100000
80000
60000
40000
20000
0
count
0
1
Prototyping Model

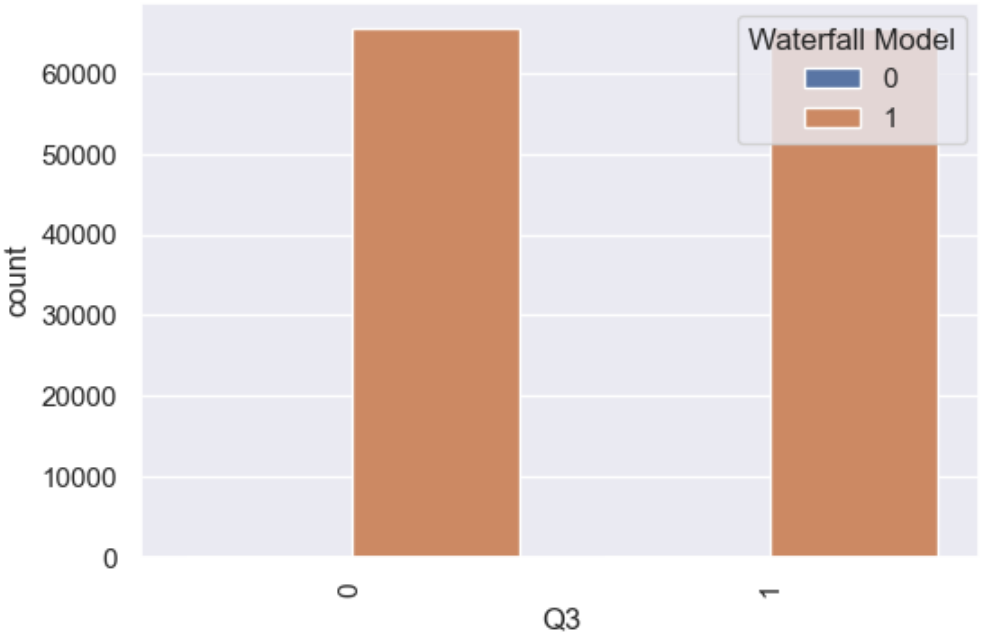

120000
100000
80000
60000
40000
20000
0
count
0
1
Rapid Application Development

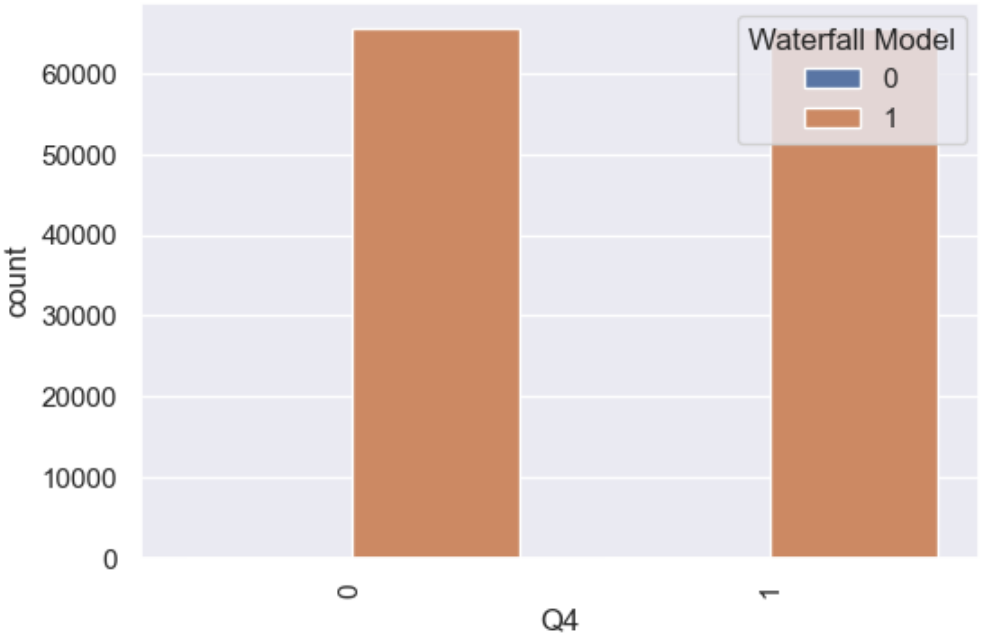
In [28]:

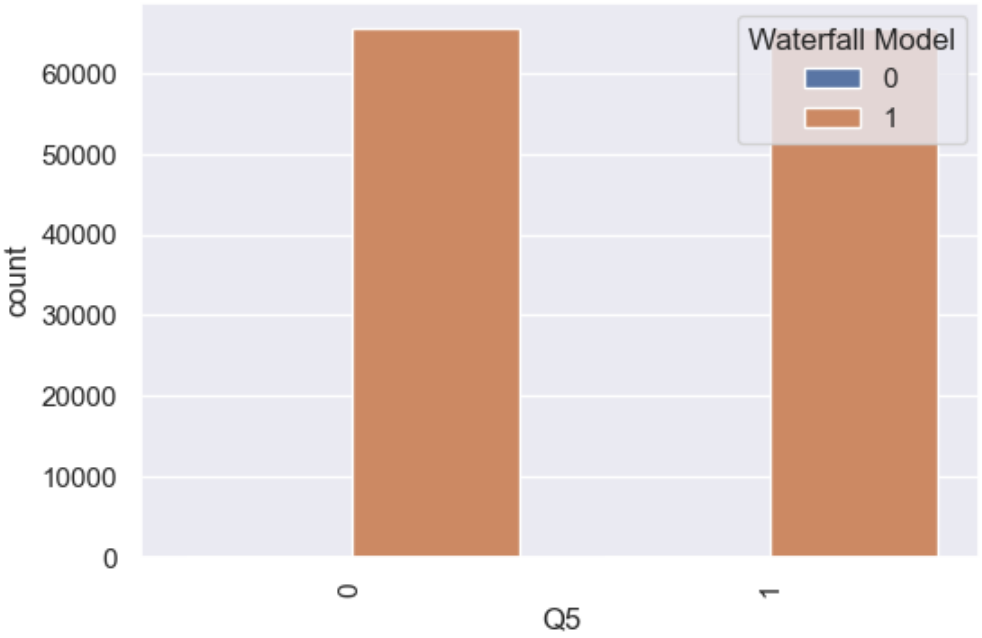
```
# Count Plot: Displays the count of occurrences for each category as bars.
for col in df.columns[:-3]:
    plt.xticks(rotation=90)
    sns.countplot(data=df, x=col, hue="Waterfall Model")
    plt.show()
```

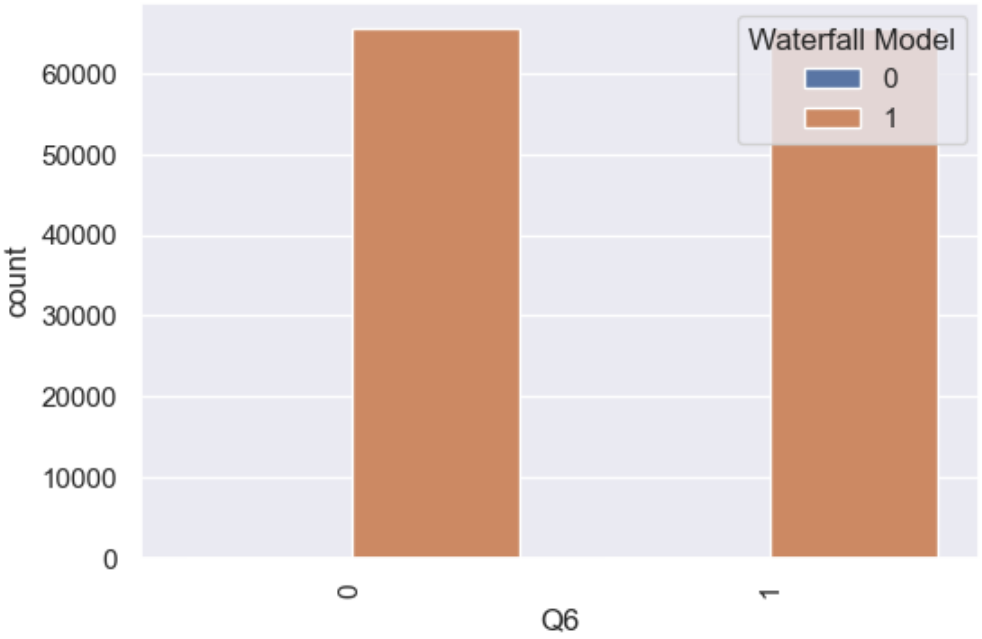


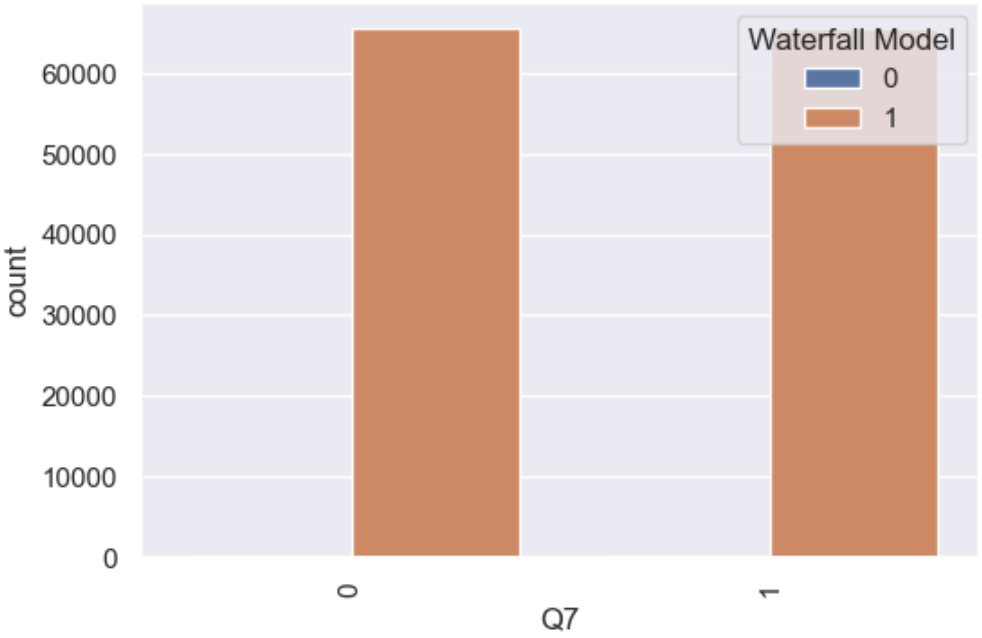

Waterfall Model
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q2

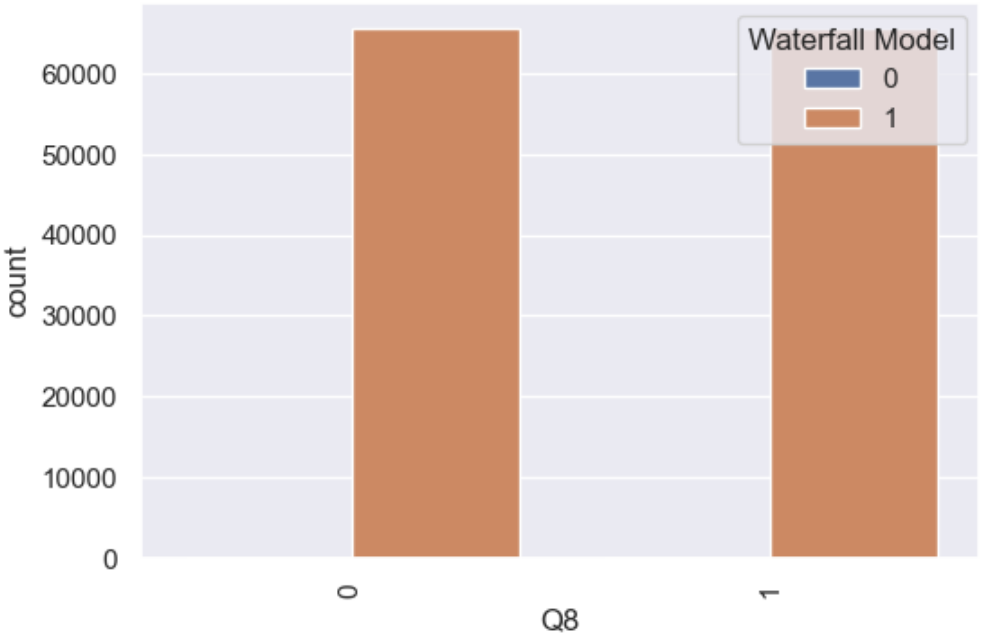

Waterfall Model
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q3

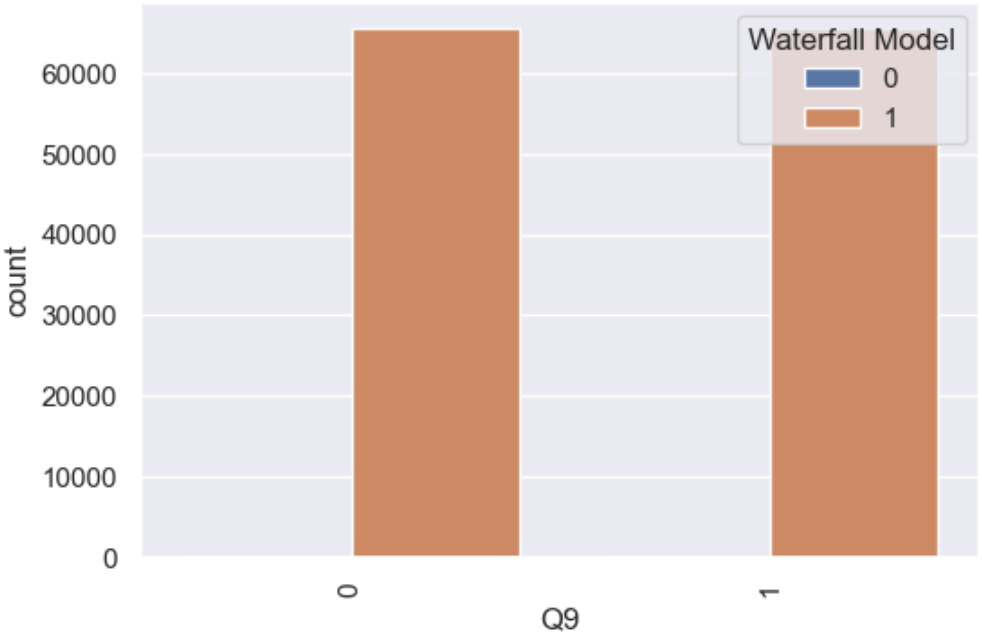

Waterfall Model
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q4

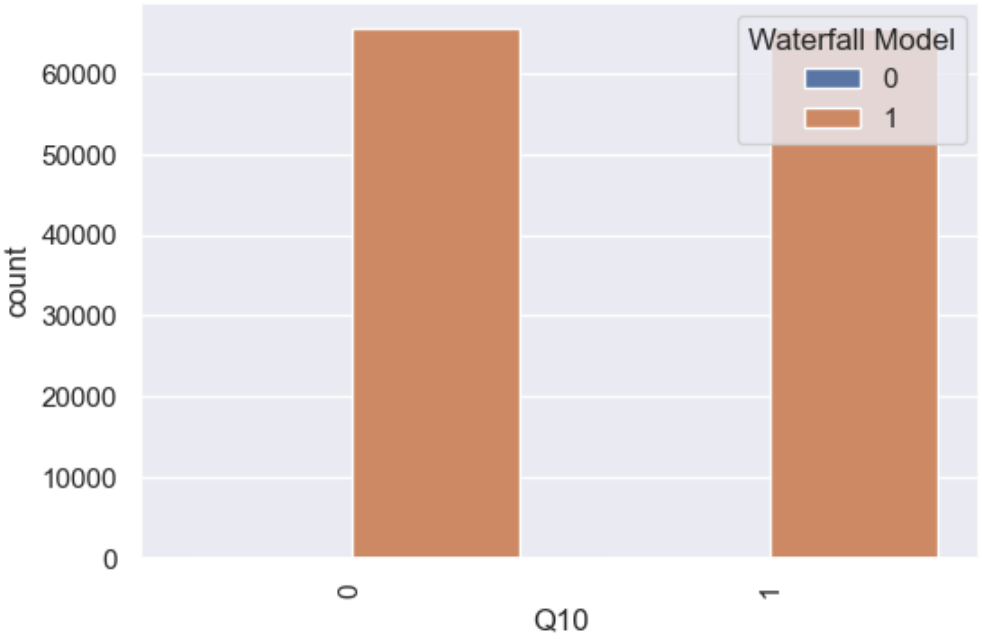

Waterfall Model
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q5

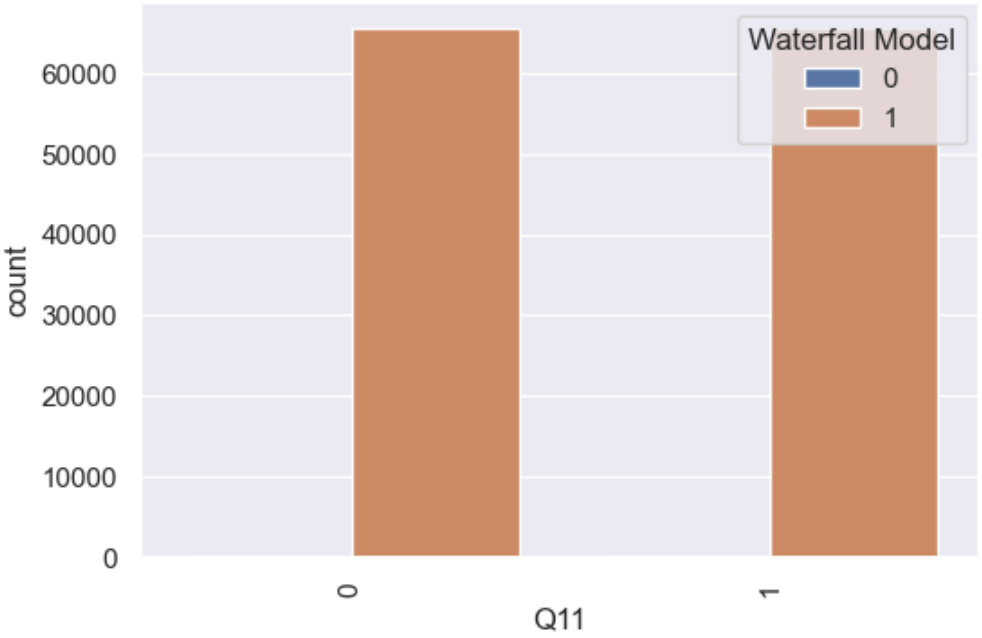

Waterfall Model
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q6

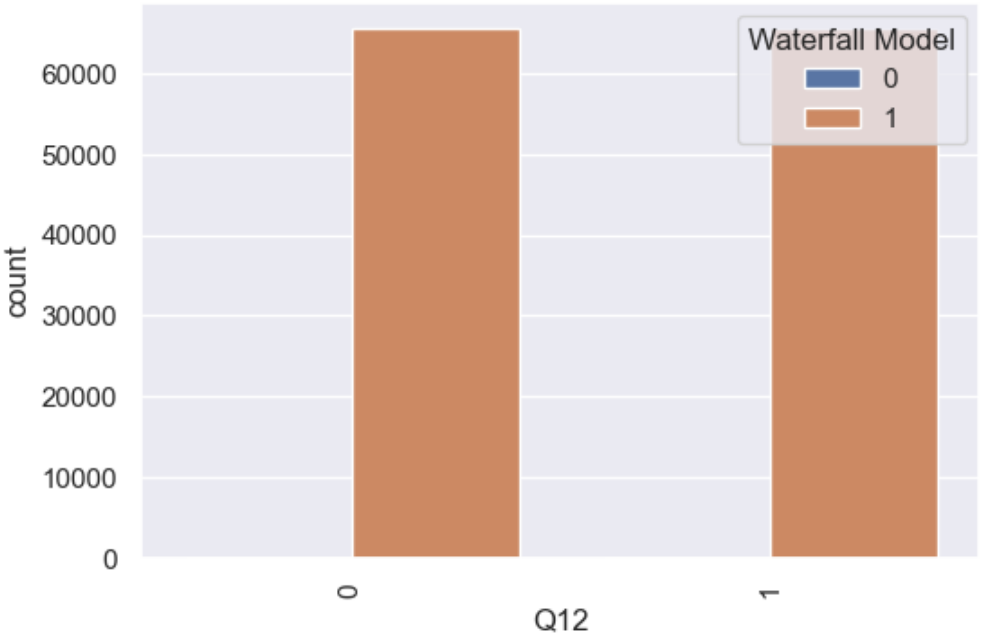

Waterfall Model
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q7

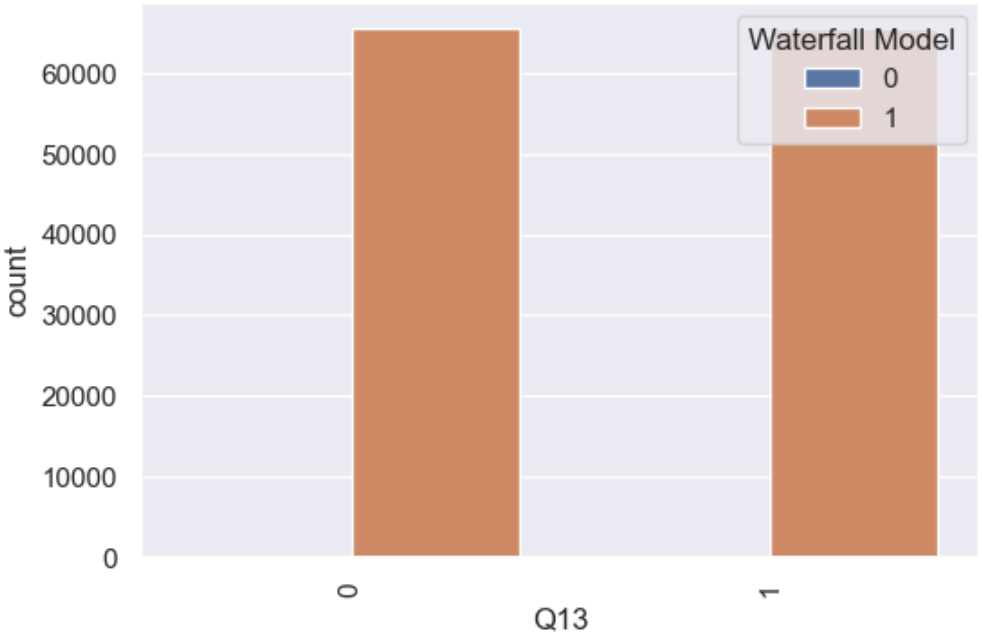

Waterfall Model
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q8

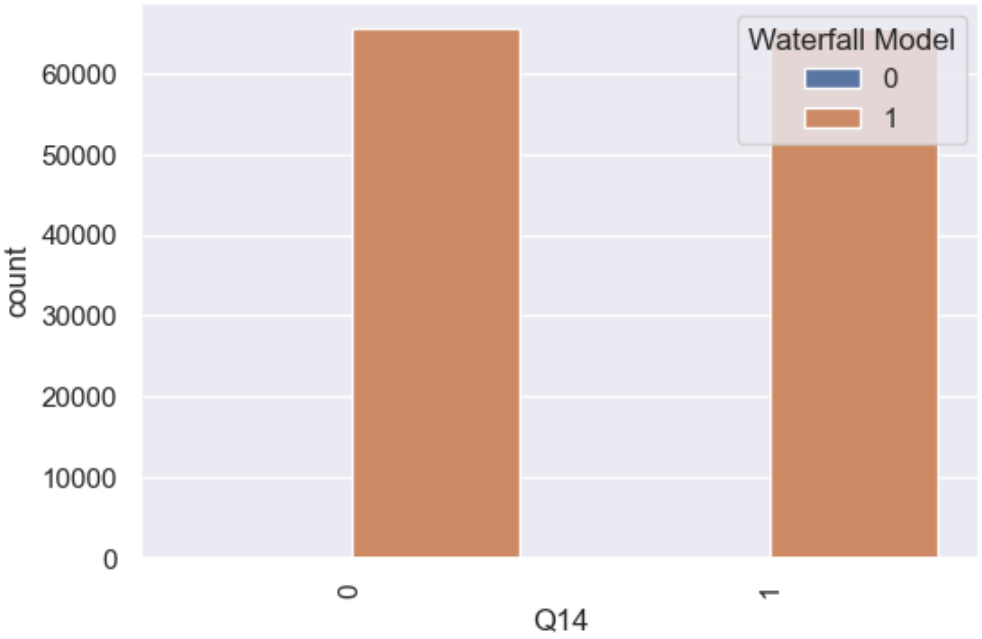

Waterfall Model
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q9

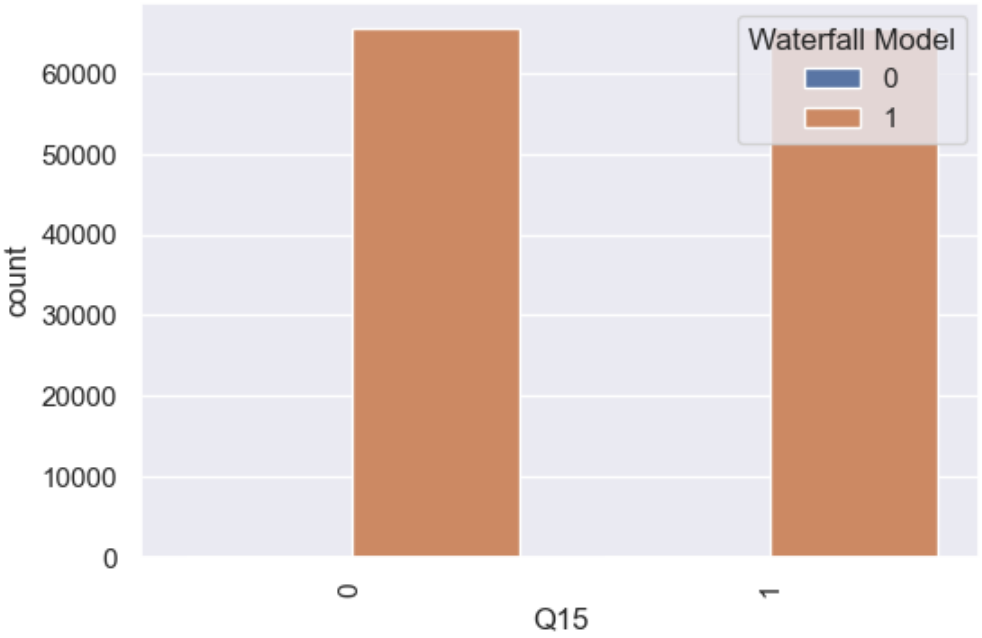

Waterfall Model
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q10

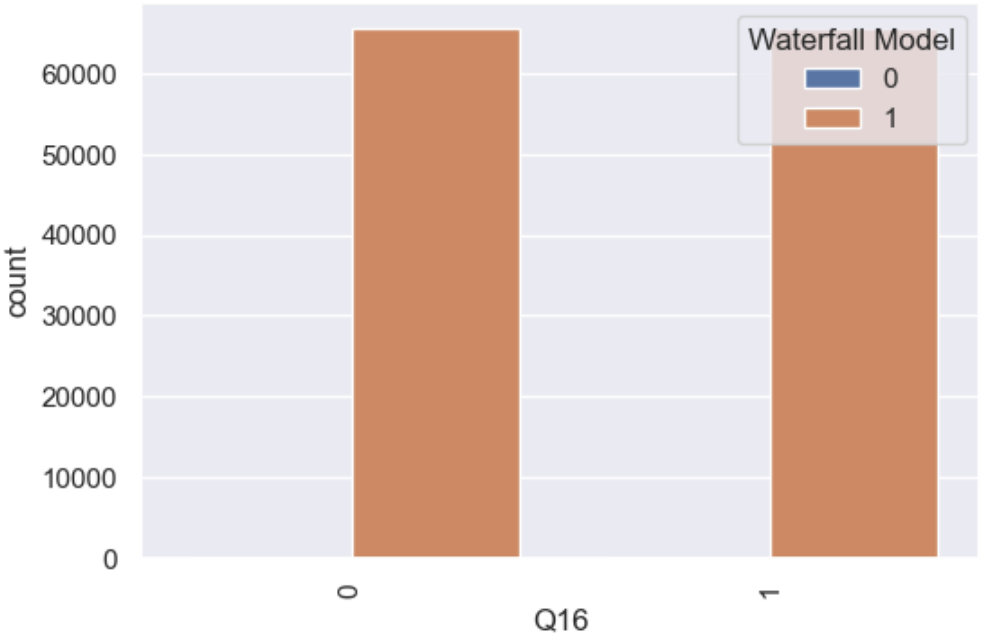

Waterfall Model
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q11

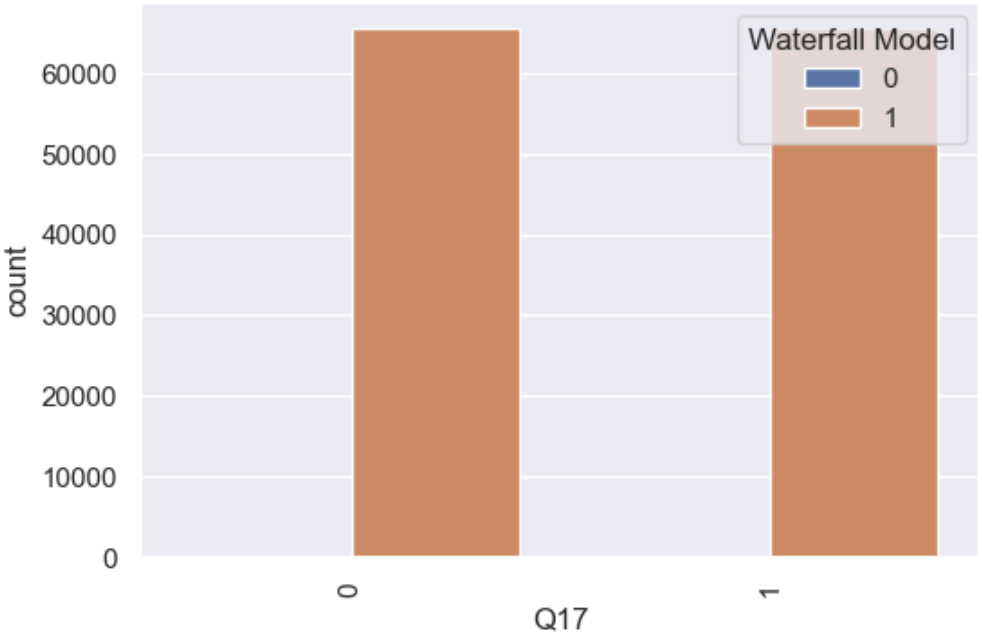

Waterfall Model
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q12


Waterfall Model
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q13


Waterfall Model
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q14

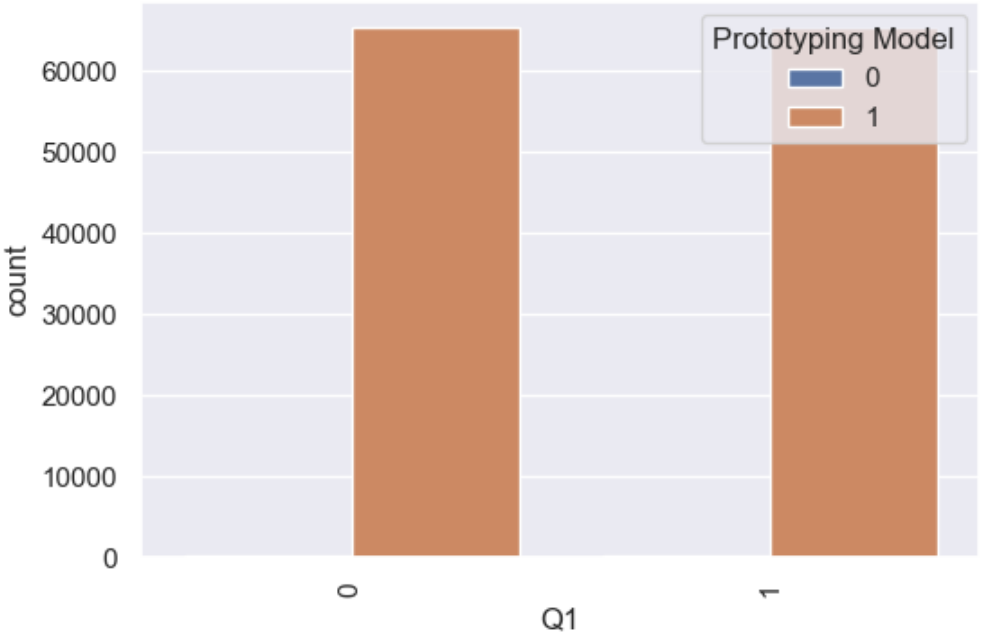

Waterfall Model
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q15

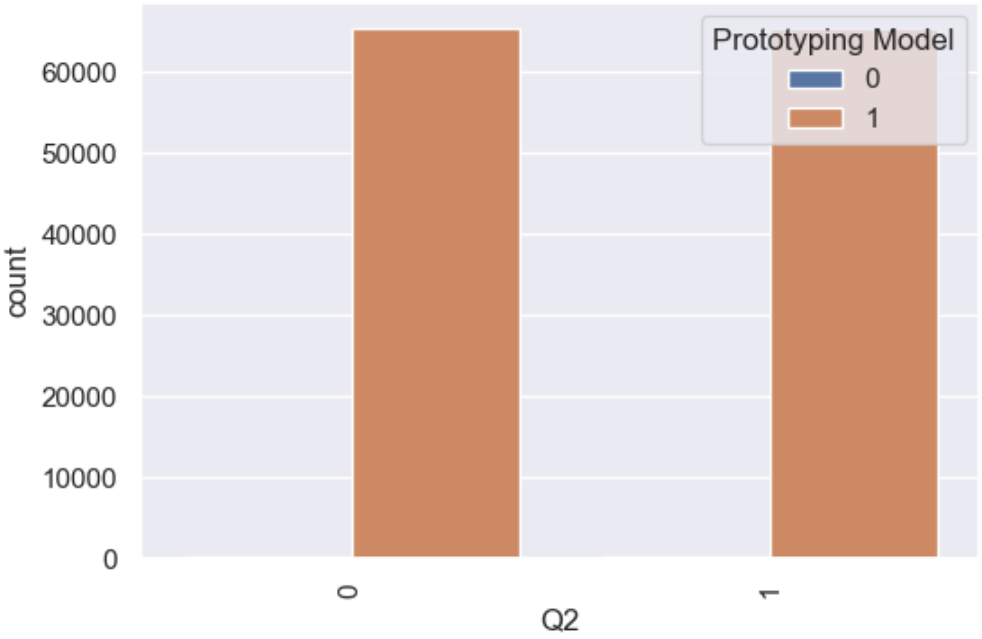

Waterfall Model
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q16

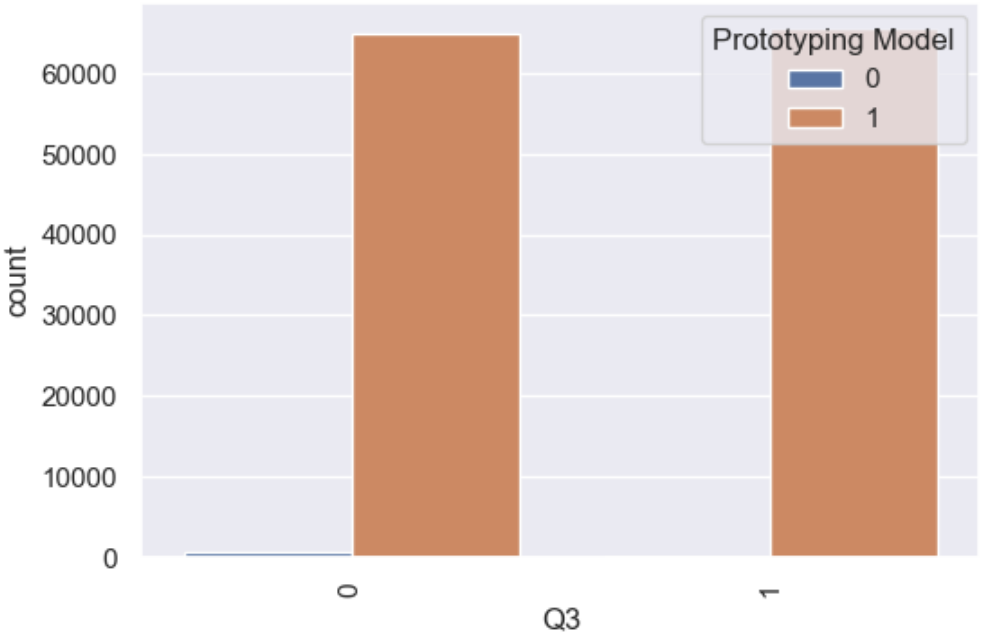

Waterfall Model
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q17

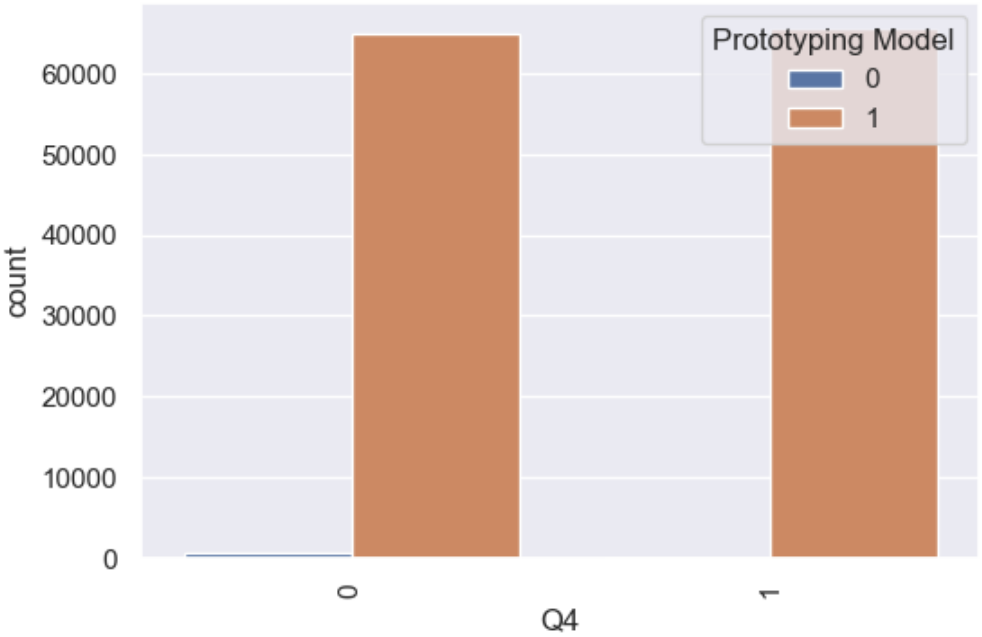
In [29]:

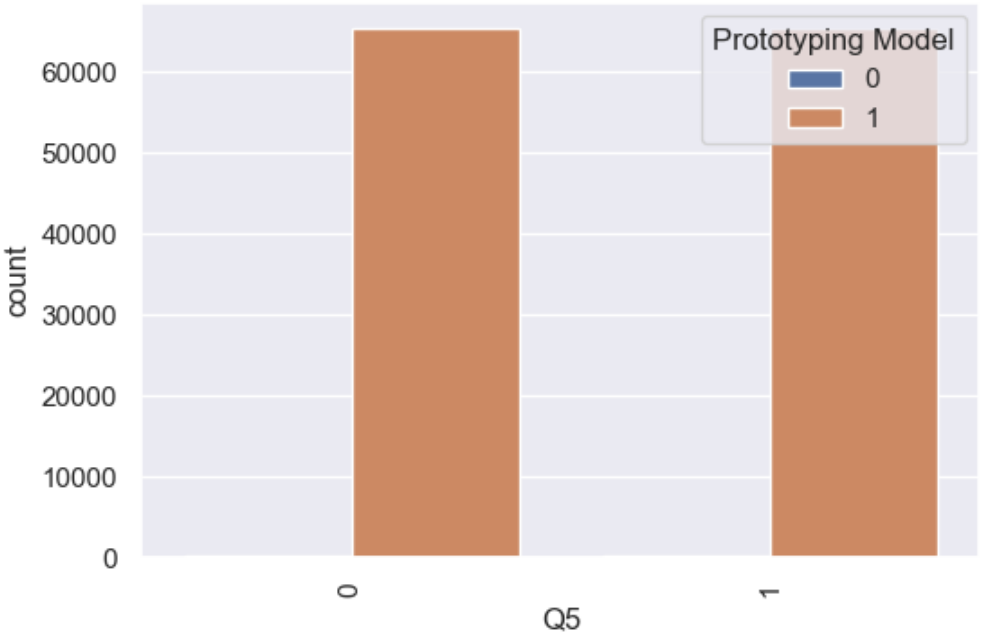
```
# Count Plot: Displays the count of occurrences for each category as bars.
for col in df.columns[:-3]:
    plt.xticks(rotation=90)
    sns.countplot(data=df, x=col, hue="Prototyping Model")
    plt.show()
```

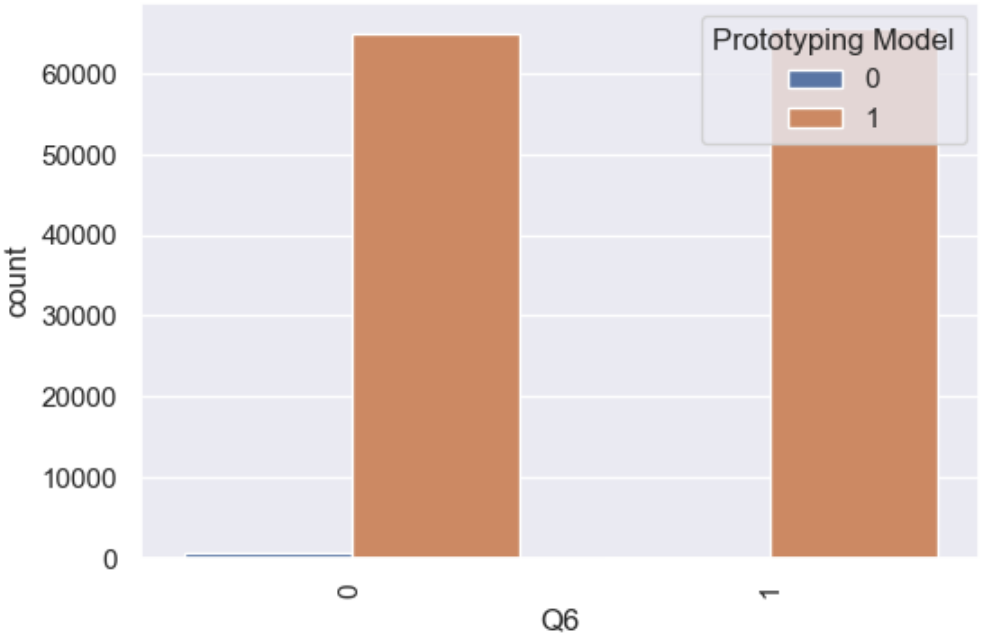


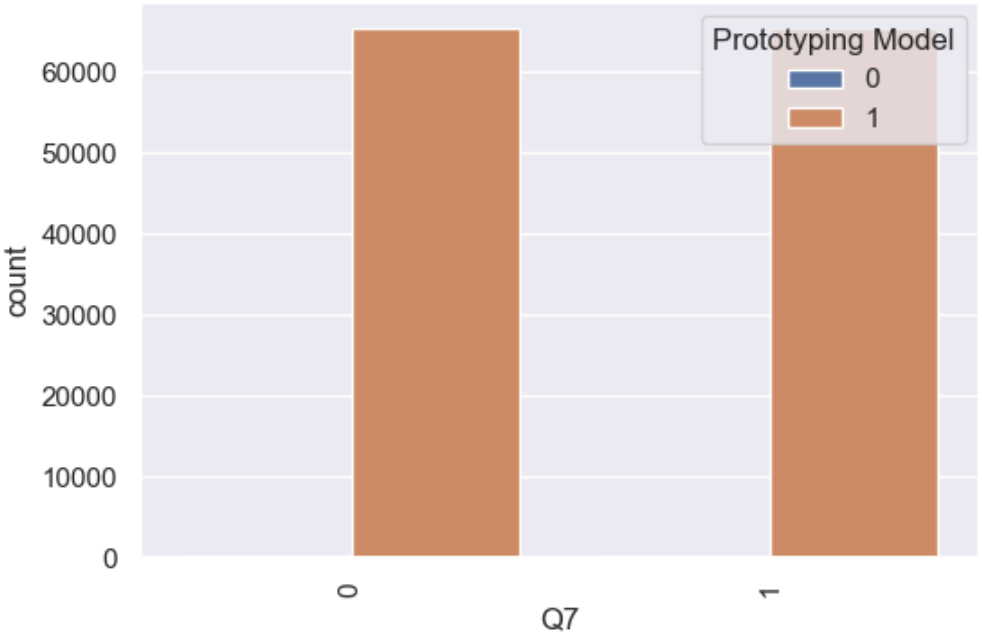

Prototyping Model
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q2

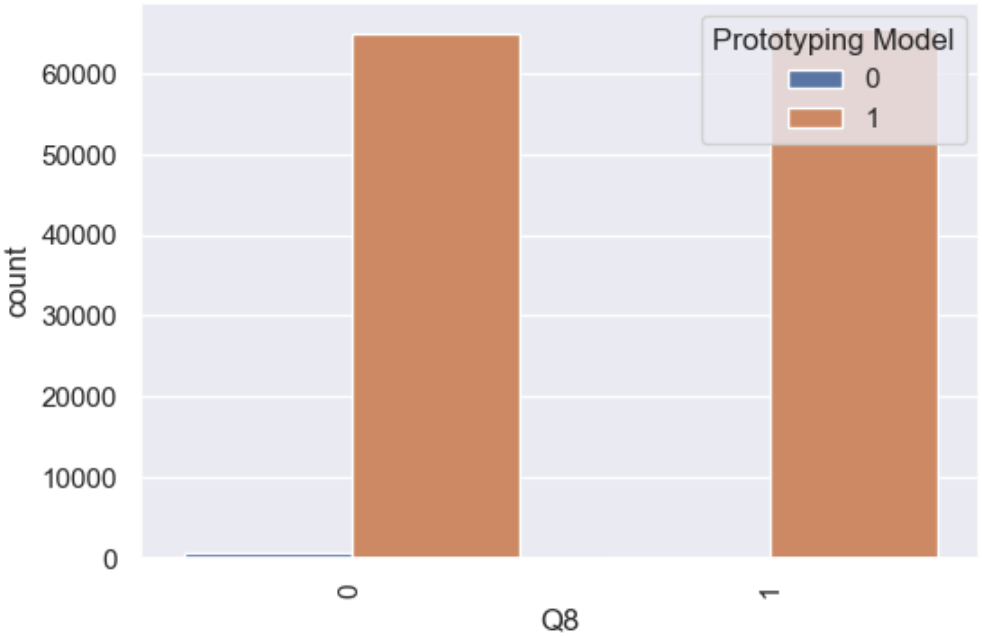

Prototyping Model
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q3

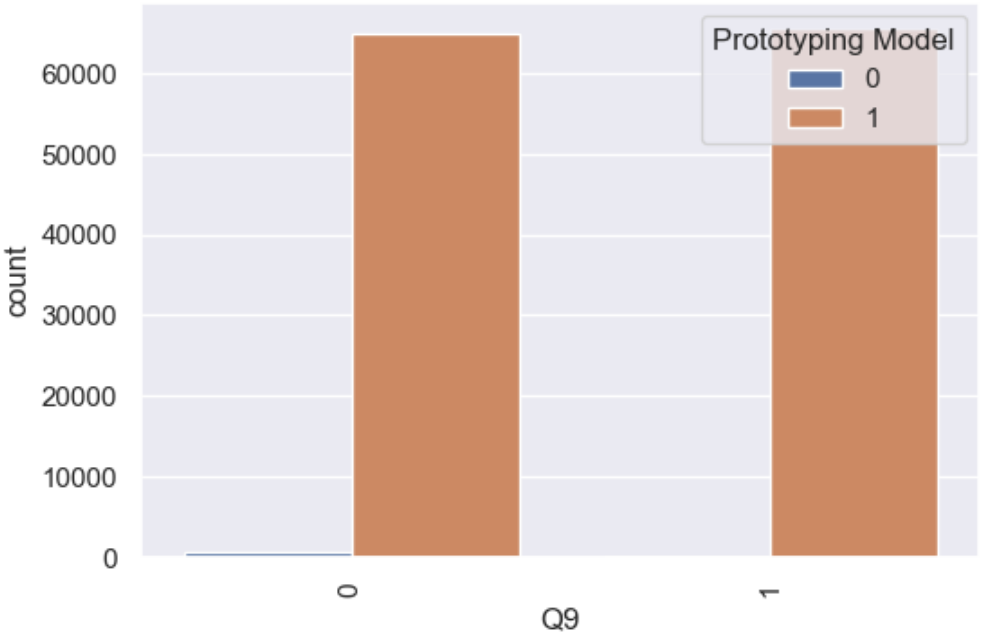

Prototyping Model
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q4

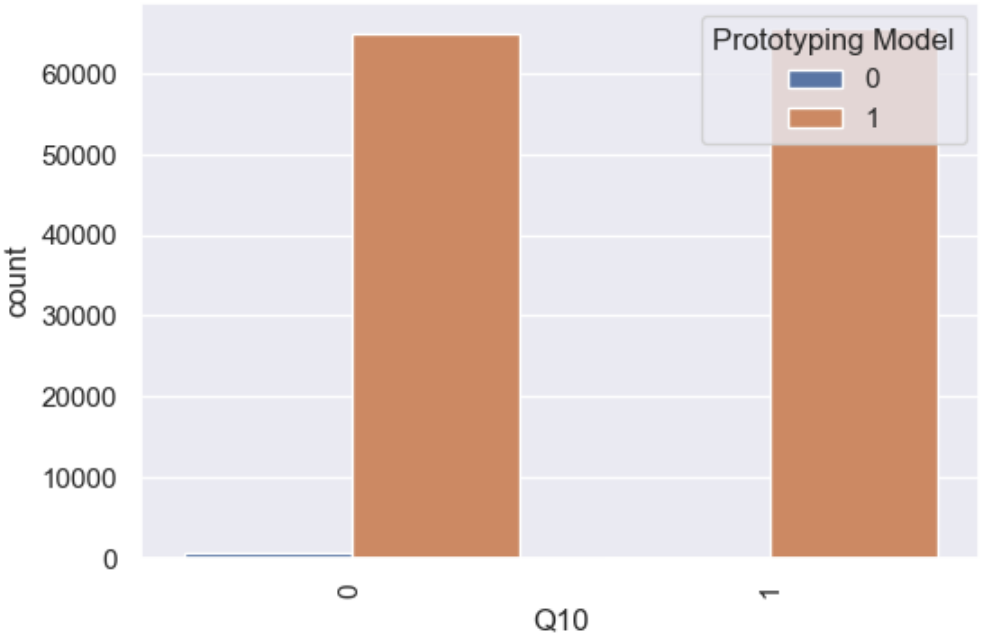

Prototyping Model
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q5

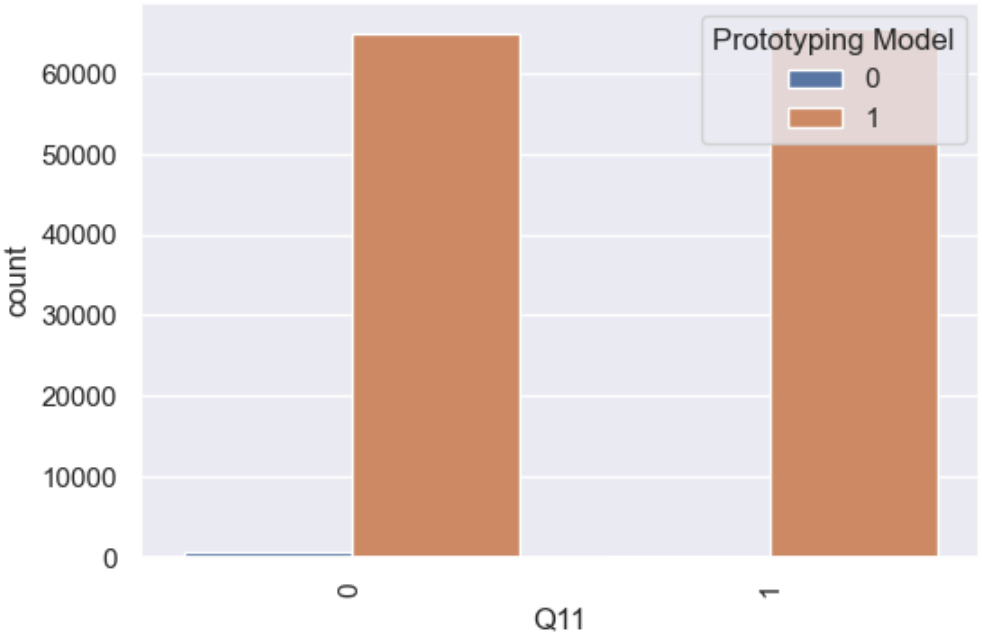

Prototyping Model
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q6

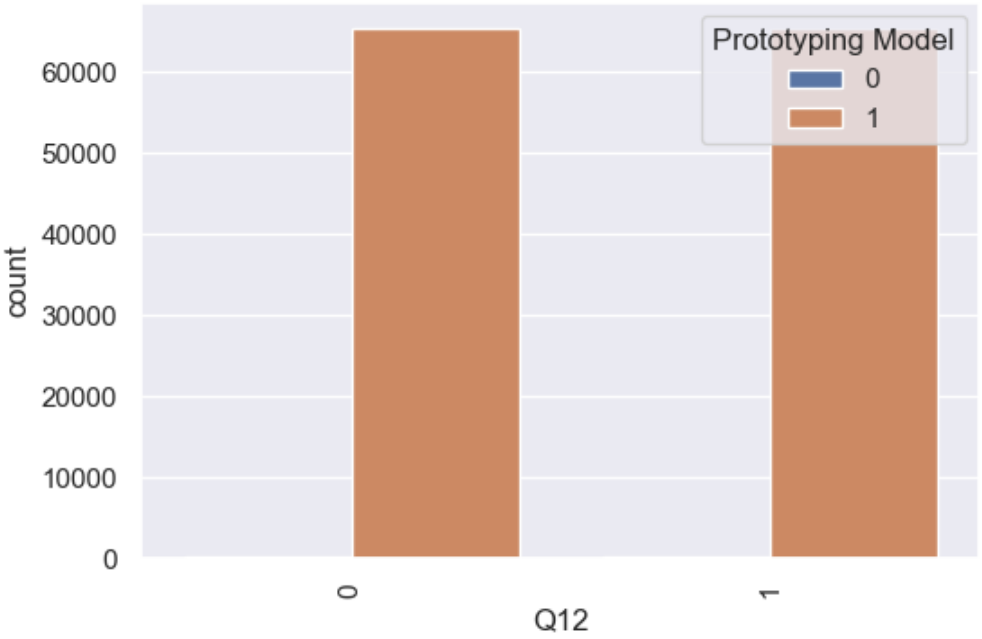

Prototyping Model
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q7

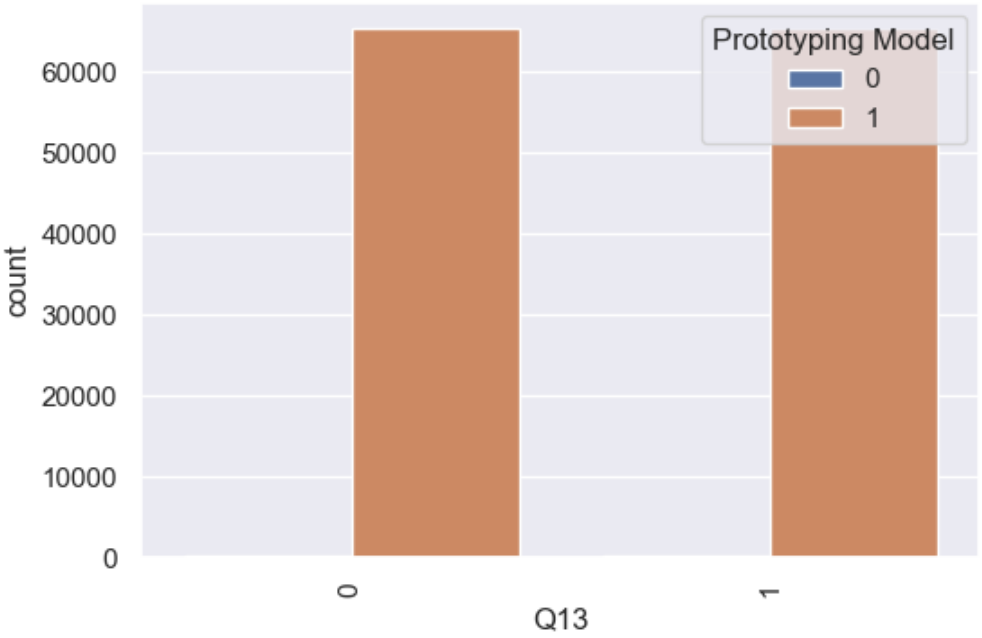

Prototyping Model
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q8

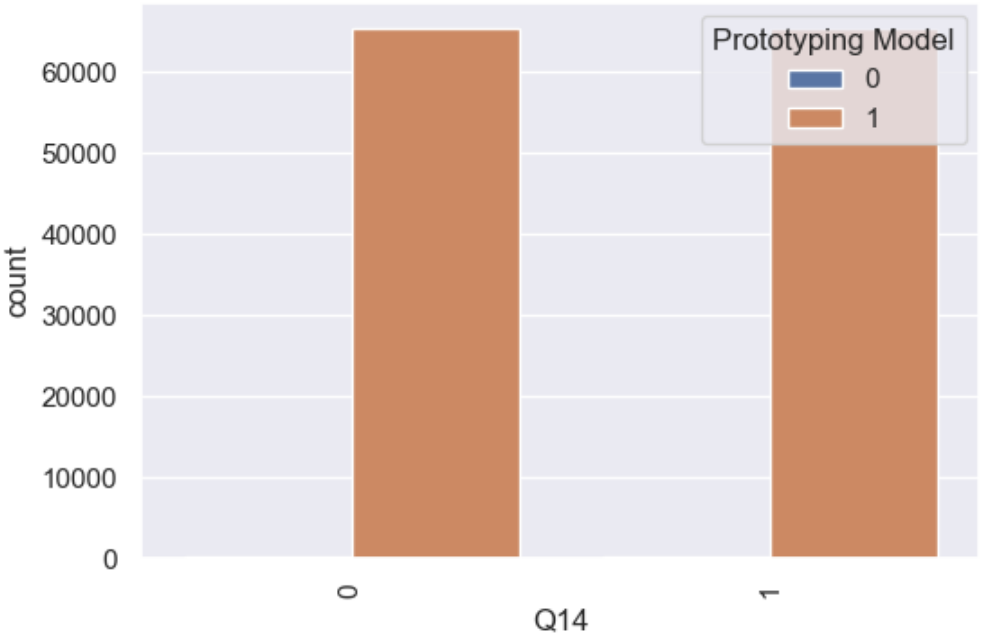

Prototyping Model
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q9

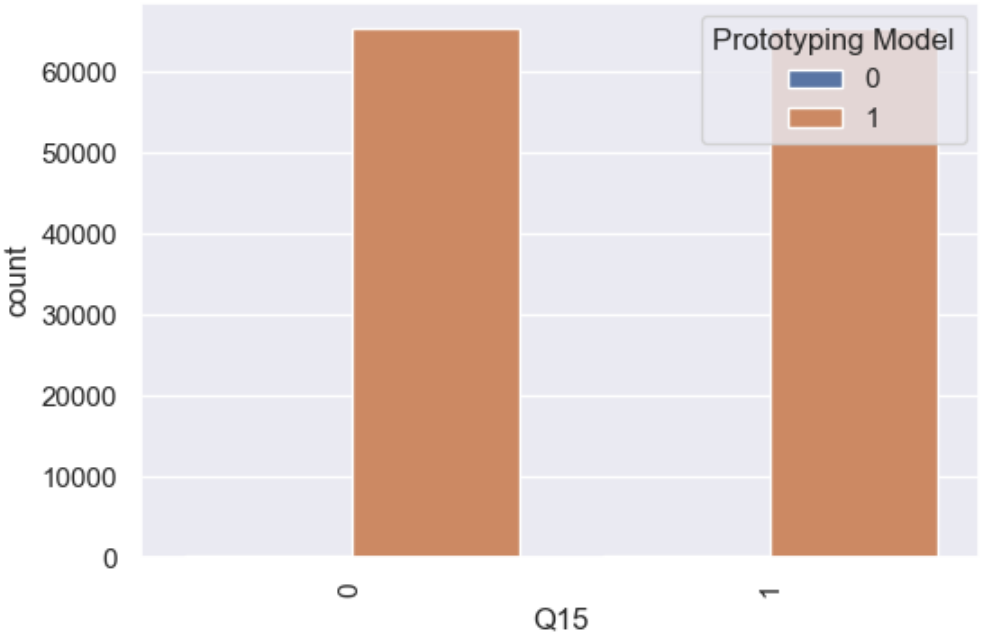

Prototyping Model
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q10

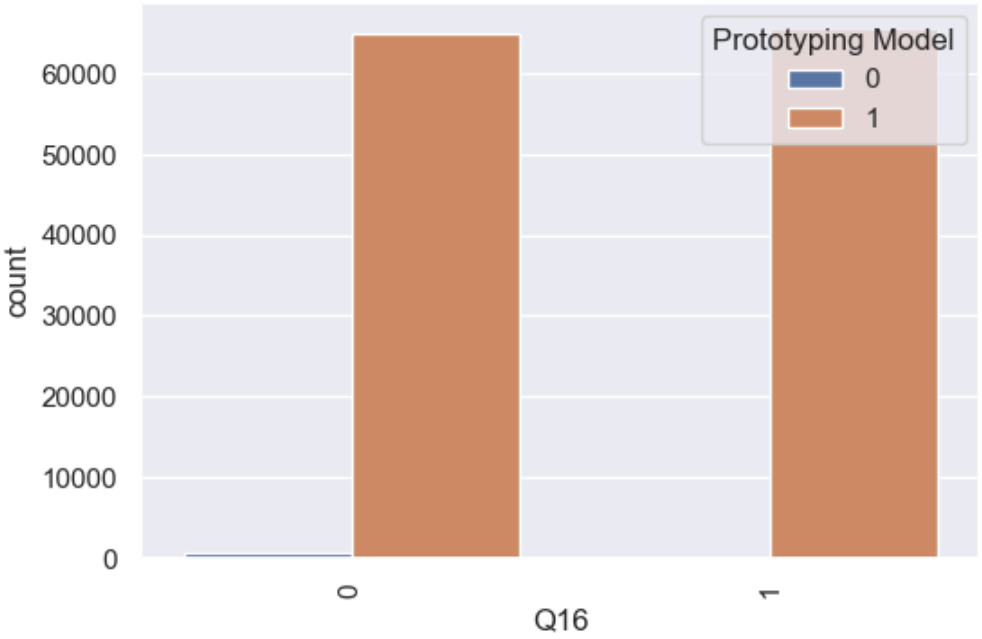

Prototyping Model
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q11

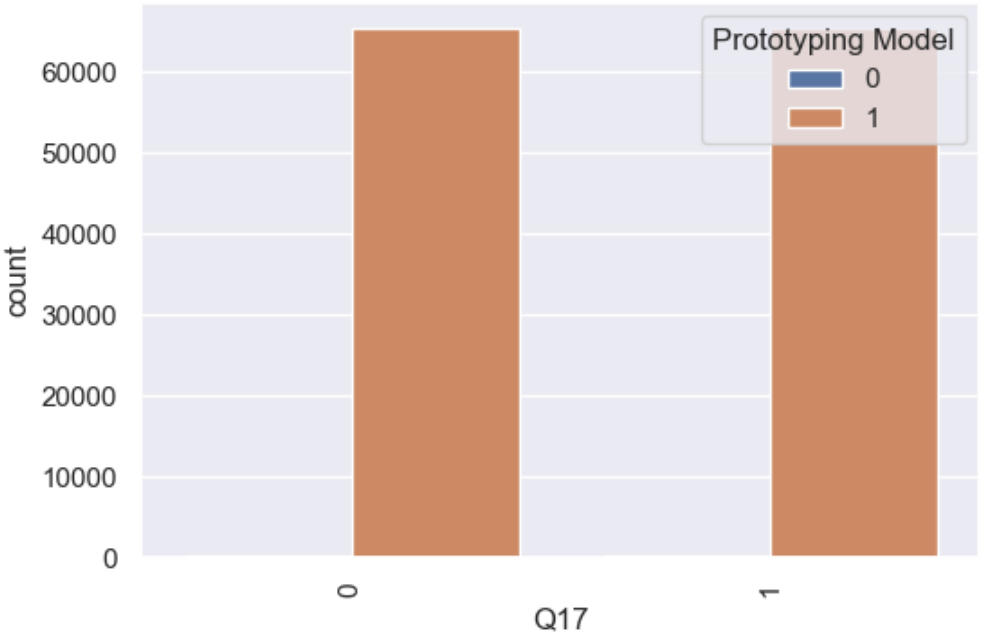

Prototyping Model
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q12


Prototyping Model
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q13


Prototyping Model
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q14

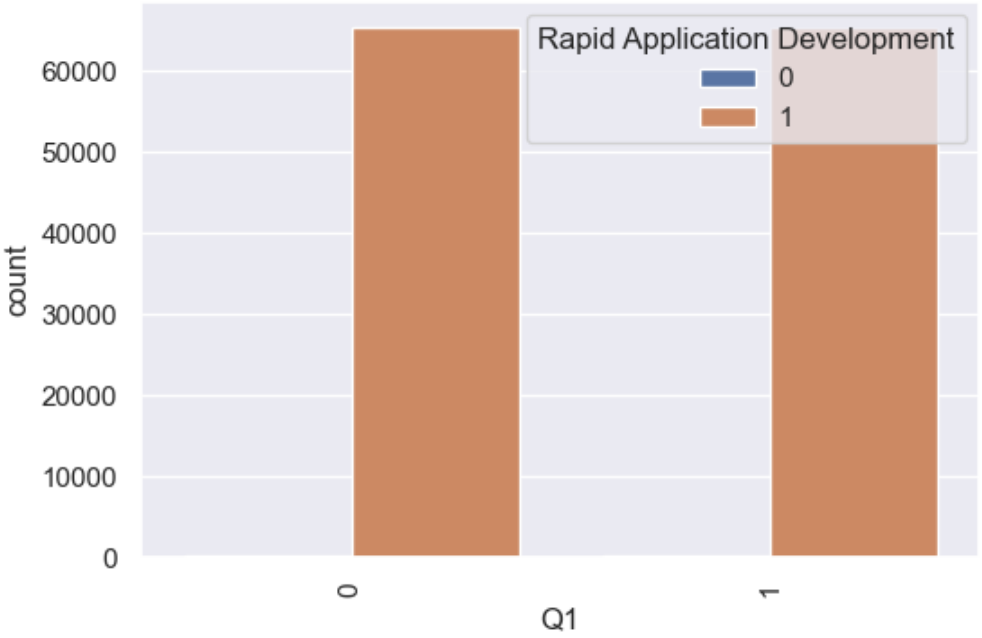

Prototyping Model
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q15

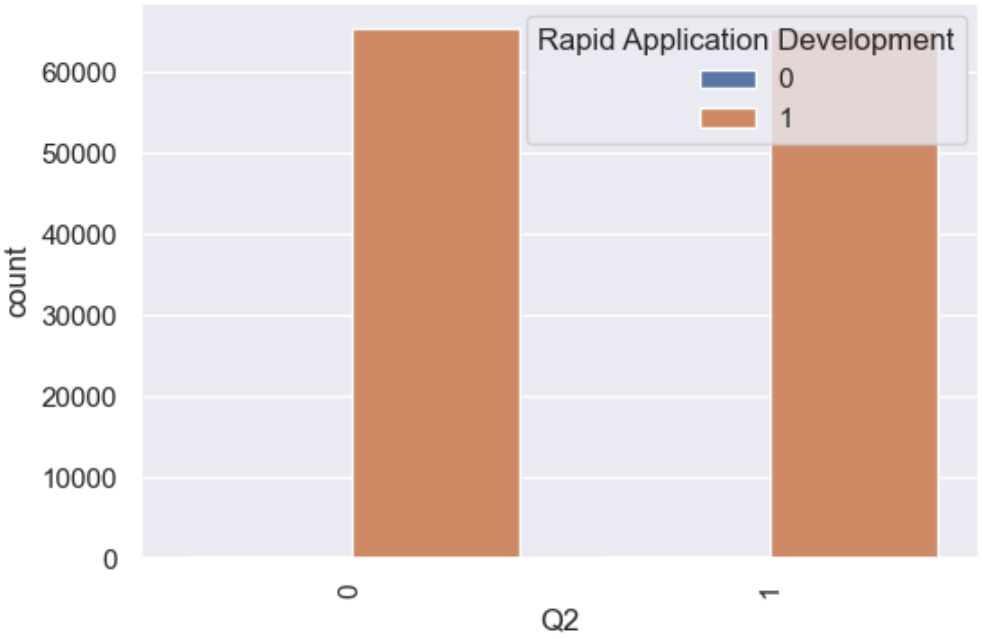

Prototyping Model
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q16

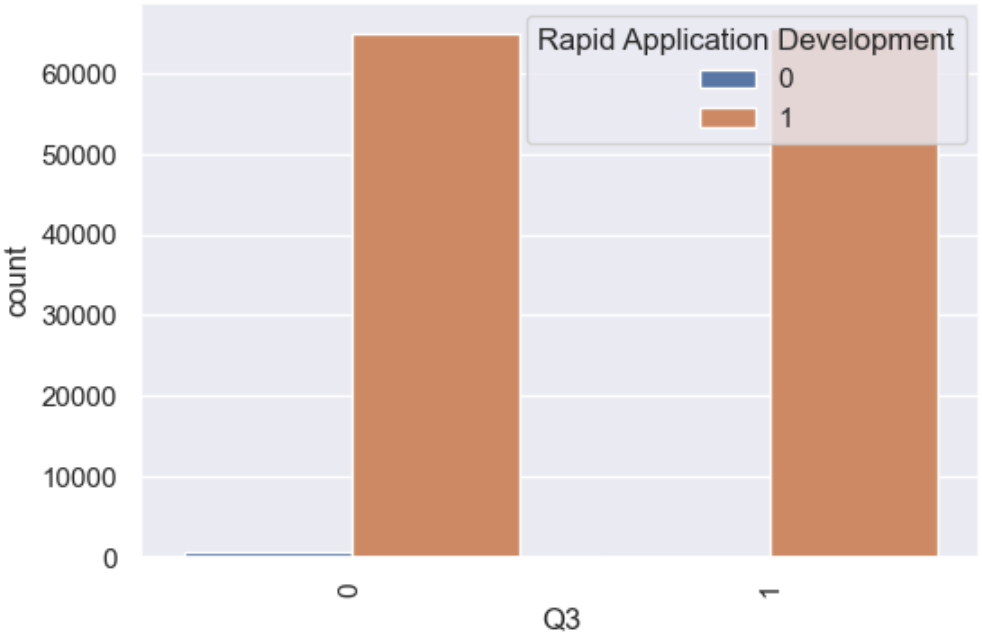

Prototyping Model
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q17

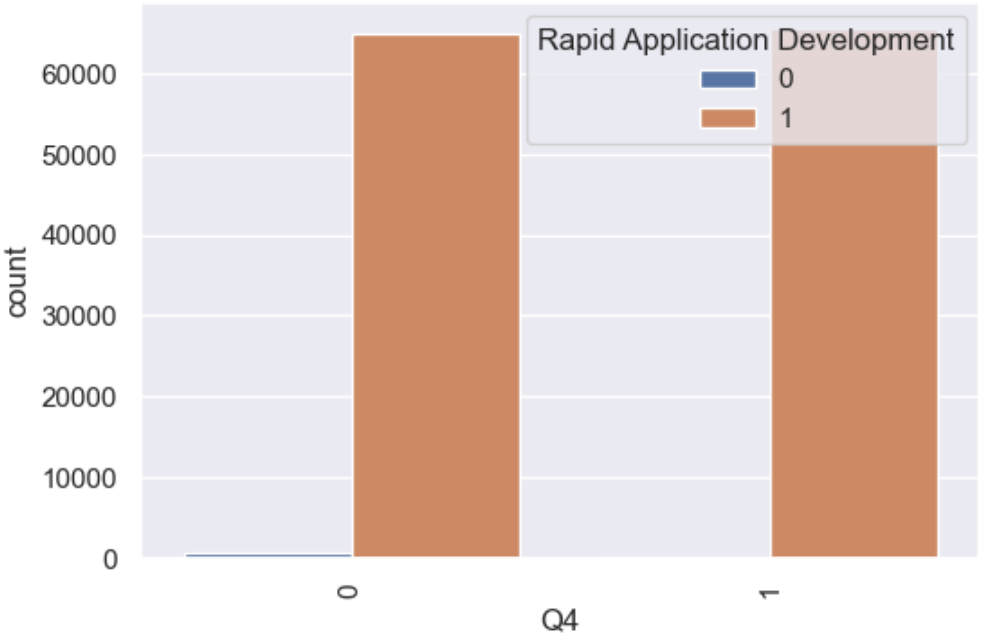
In [30]:

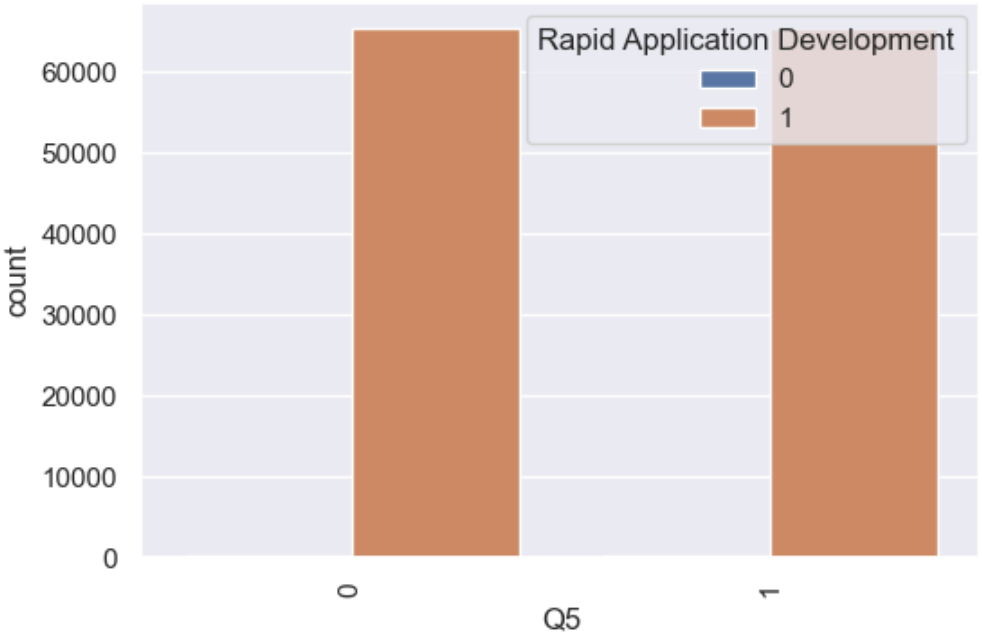
```
# Count Plot: Displays the count of occurrences for each category as bars.
for col in df.columns[:-3]:
    plt.xticks(rotation=90)
    sns.countplot(data=df, x=col, hue="Rapid Application Development")
    plt.show()
```

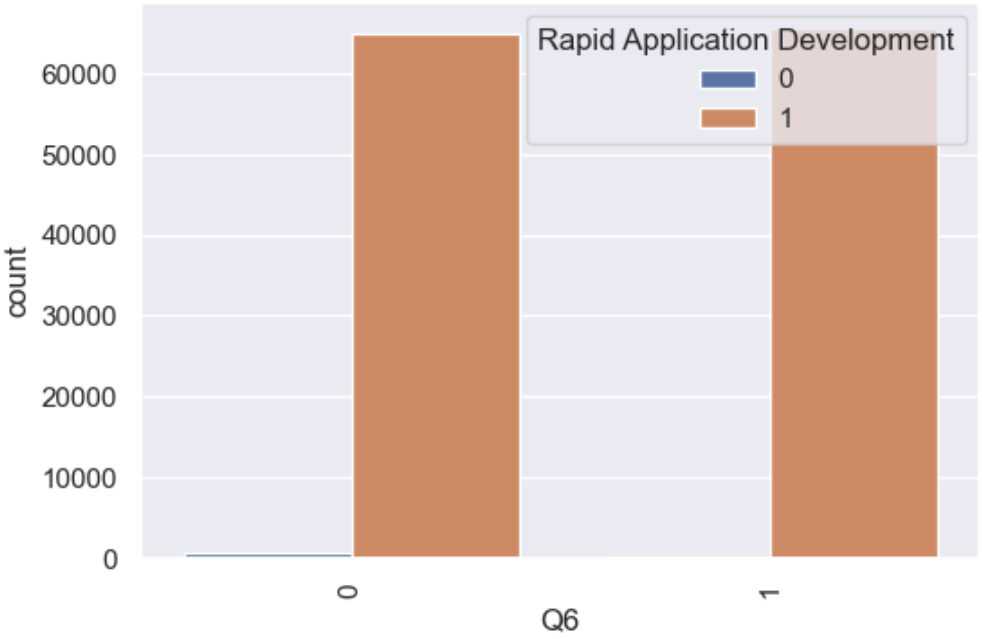


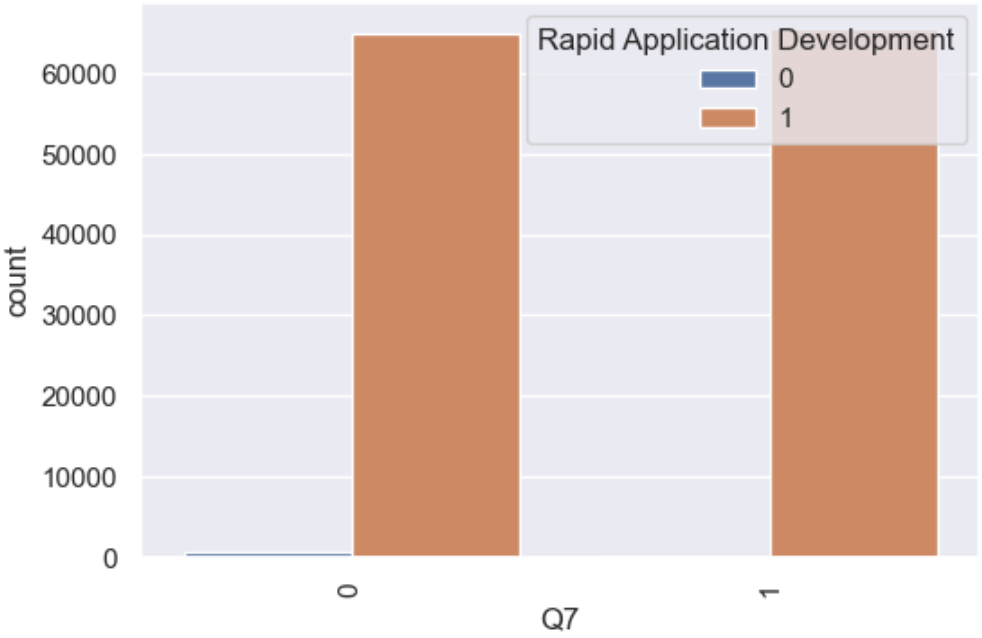

Rapid Application Development
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q2

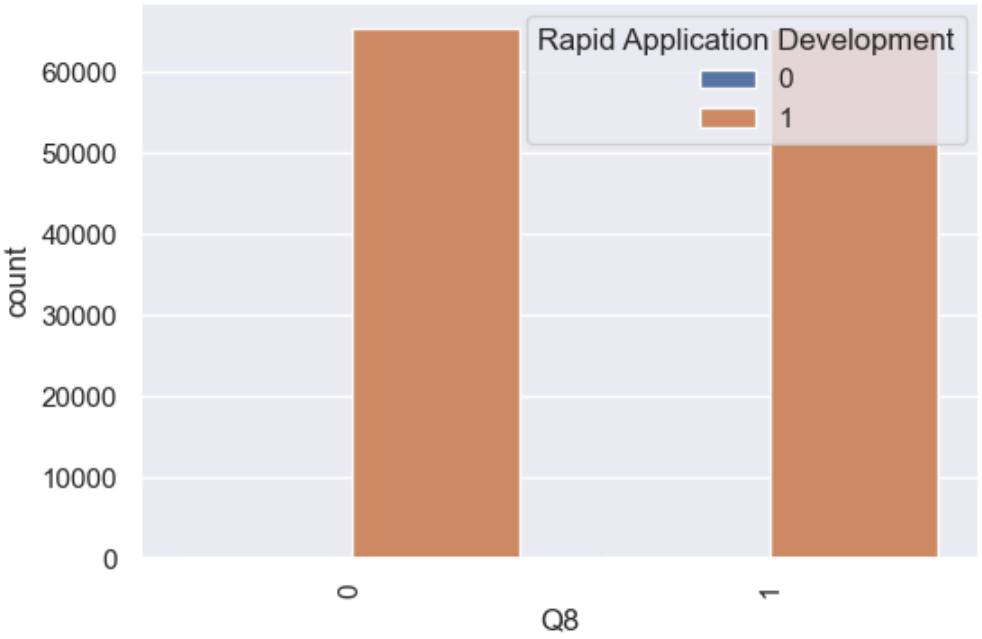

Rapid Application Development
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q3

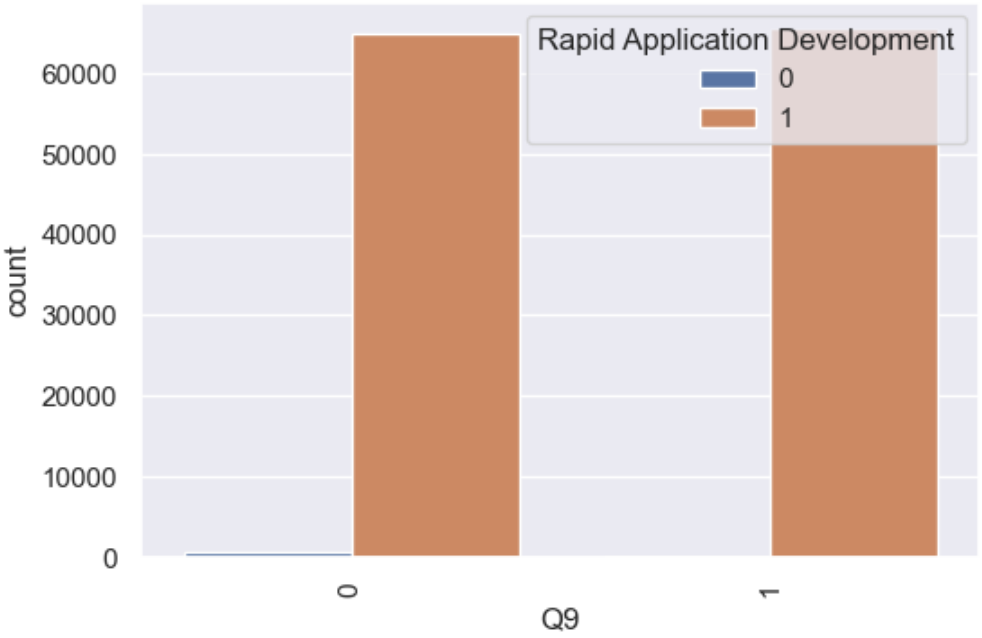

Rapid Application Development
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q4

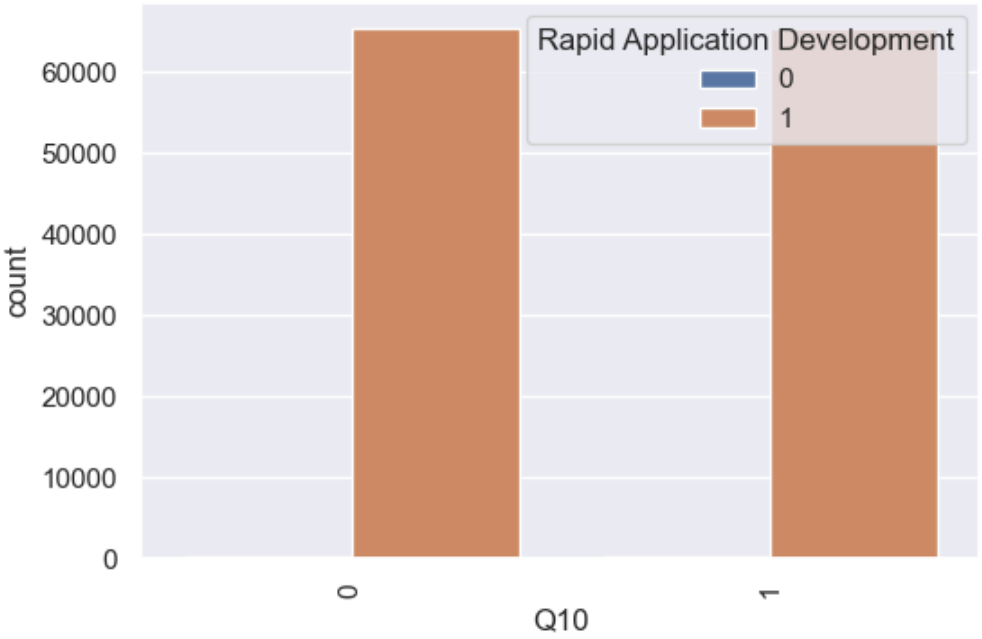

Rapid Application Development
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q5

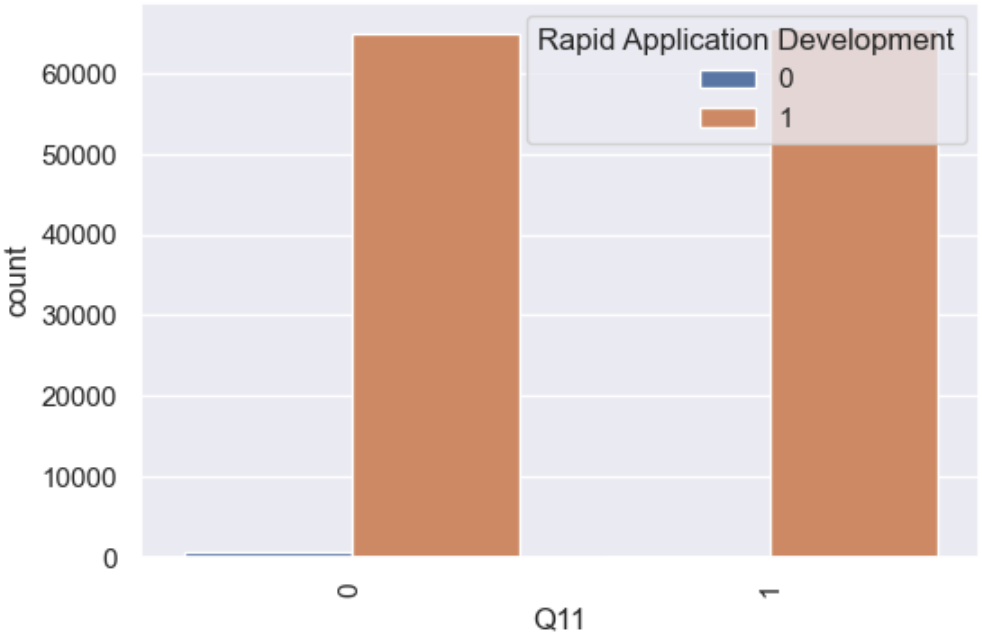

Rapid Application Development
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q6

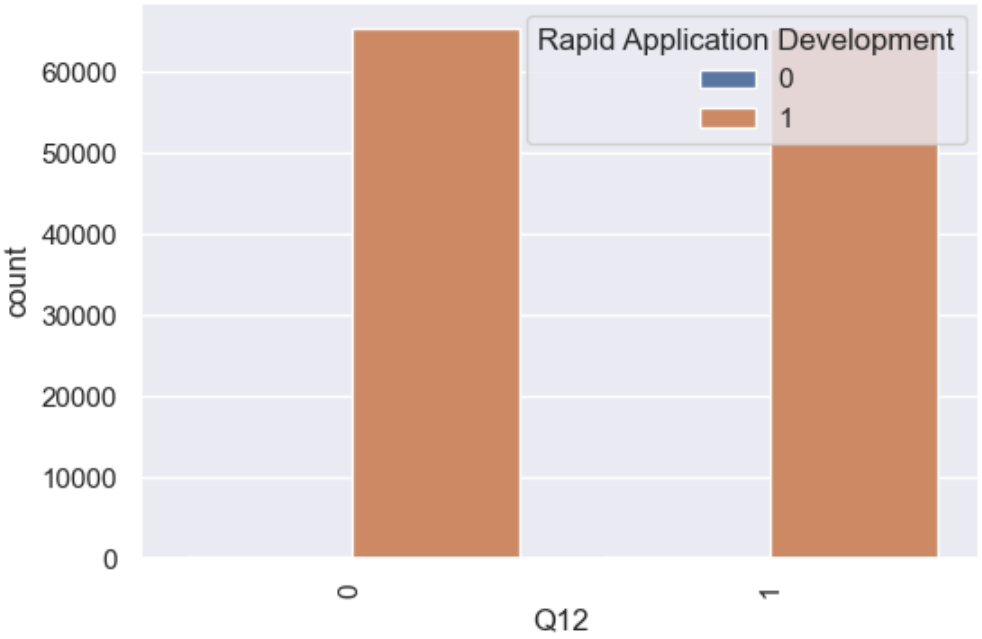

Rapid Application Development
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q7

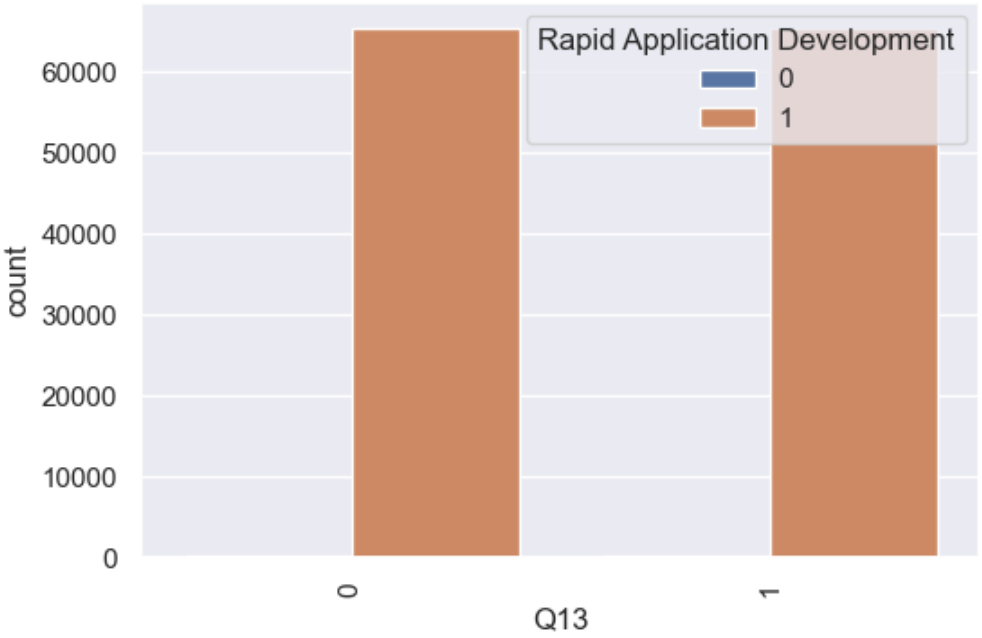

Rapid Application Development
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q8

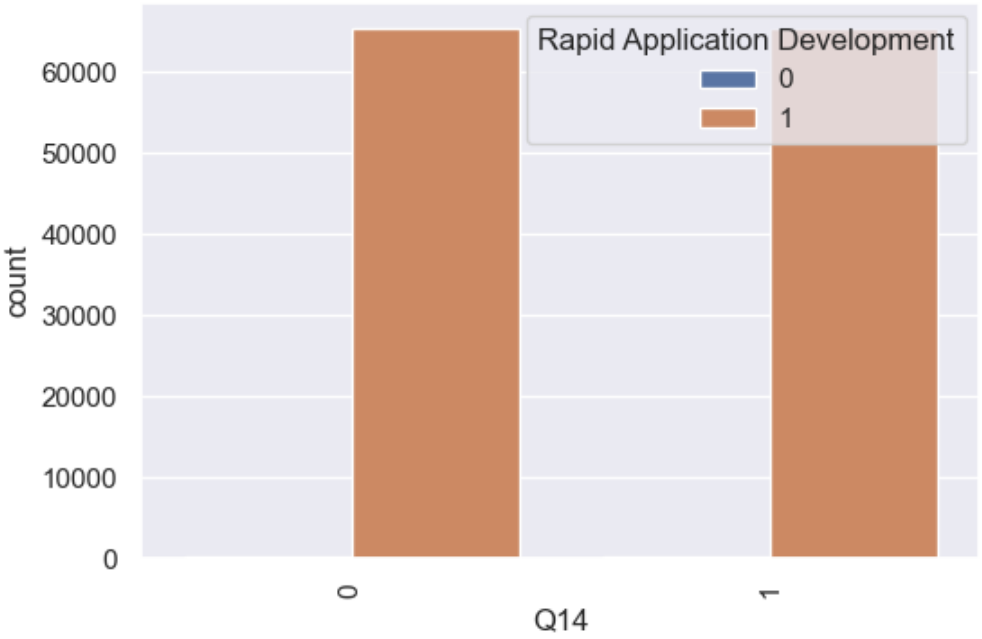

Rapid Application Development
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q9

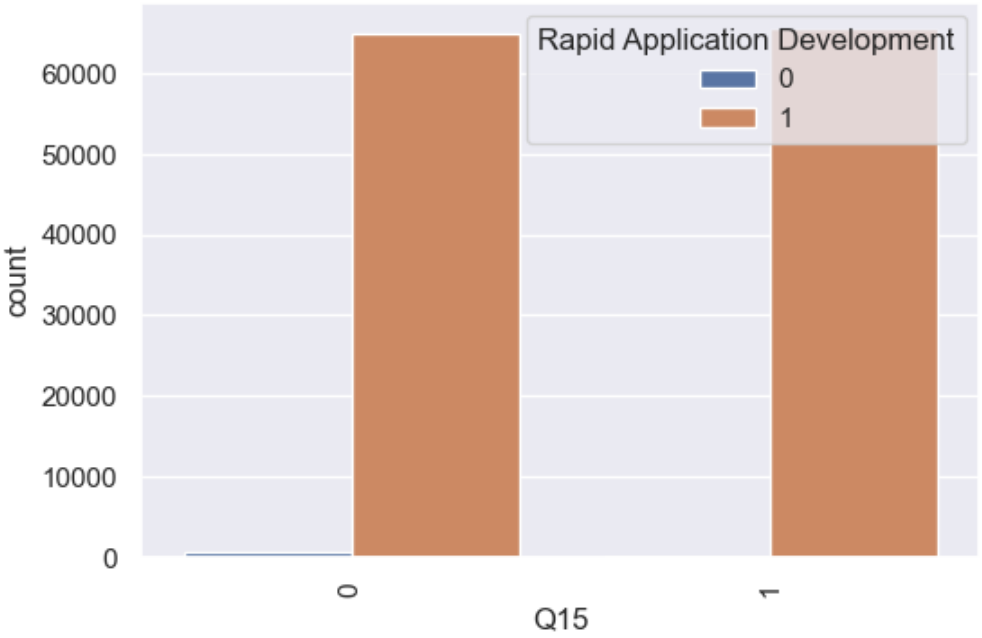

Rapid Application Development
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q10

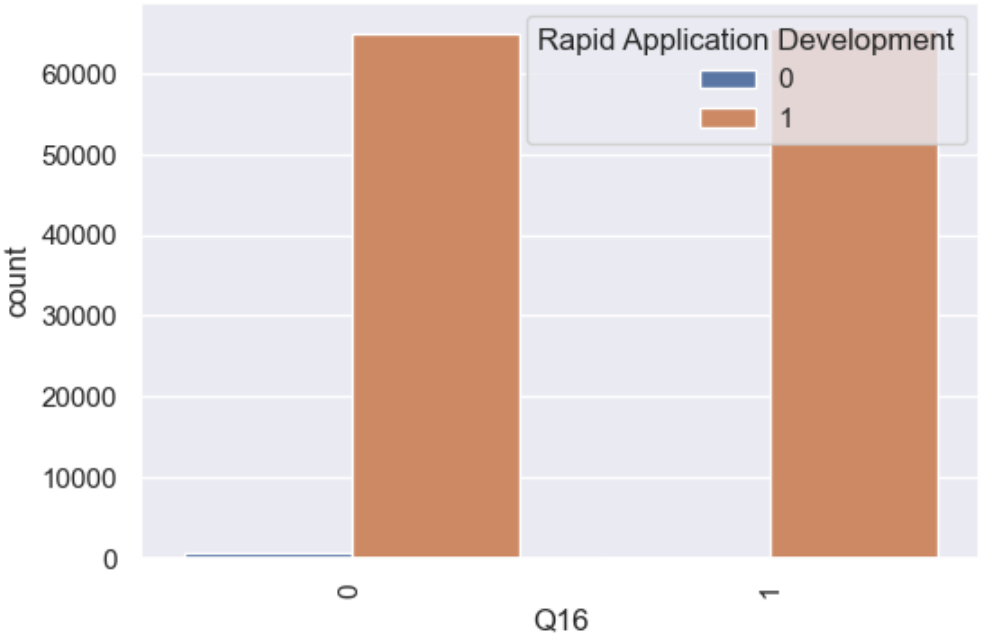

Rapid Application Development
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q11

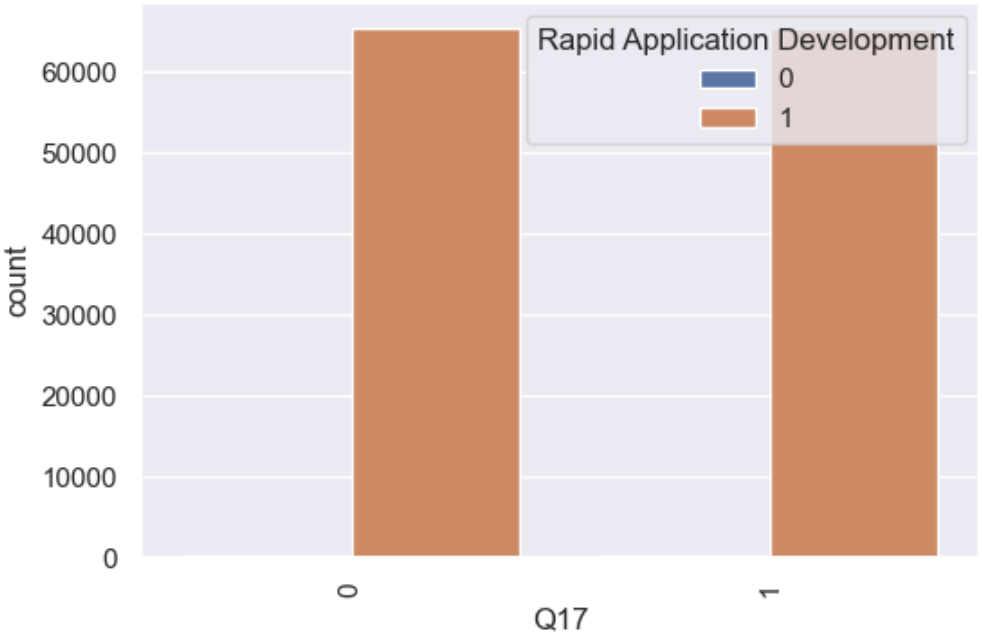

Rapid Application Development
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q12


Rapid Application Development
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q13


Rapid Application Development
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q14


Rapid Application Development
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q15


Rapid Application Development
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q16


Rapid Application Development
0
1
60000
50000
40000
30000
20000
10000
0
count
0
1
Q17

In [34]:
```
for col in df.columns[-3:]:
    print(col, ' :: \n', df[col].value_counts())
```

```
Waterfall Model  ::
 1    131056
0        16
Name: Waterfall Model, dtype: int64
Prototyping Model  ::
 1    130560
0       512
Name: Prototyping Model, dtype: int64
Rapid Application Development  ::
 1    130560
0       512
Name: Rapid Application Development, dtype: int64
```

In [ ]:
```

```