Homesite_QuoteConversion_Final

August 26, 2022

1. BUSINESS PROBLEM

DESCRIPTION

Before asking someone on a date or skydiving, it's important to know your likelihood of success. The same goes for quoting home insurance prices to a potential customer. Homesite, a leading provider of homeowners insurance, does not currently have a dynamic conversion rate model that can give them confidence a quoted price will lead to a purchase.

Using an anonymized database of information on customer and sales activity, including property and coverage information, Homesite is challenging you to predict which customers will purchase a given quote. Accurately predicting conversion would help Homesite better understand the impact of proposed pricing changes and maintain an ideal portfolio of customer segments.

PROBLEM STATEMENT

1. Every organization that we come across in our day-to-day life works on a limited resources and each of the processes that are carried out in an organization consumes resources and for the long-term survival organization it is important that these processes that are carried out in an organization are carried out in the most efficient way possible.
2. Homesite is looking out for optimizing one such process of quoting home insurance prices to potential customers where they want to build a model which can achieve the best possible conversion rate for them.
3. For this task they have provided us with a dataset which represents the activity of large number of customers who are interested in buying policies from there website. The provided features for each of these activities are anonymized and provide a rich representation of perspective customer and policy.
4. So, our task as a ML Engineer is that given a set of features we have to predict whether the quote given by Homesite to a customer will end up in a successful conversion or not.

BUSINESS OBJECTIVES & CONSTRAINTS

1. If the company is working with less resources then the cost of misclassification can be high because if the model goes on to predict a quote as a successful conversion when it is not then a lot of company resources will be wasted working on something that will not reap any benifits for the company.
2. What is the probability of successful conversion is required so that any threshold of choice can be taken based on the requirement. Meaning if the company wants more and more conversions and does not care much about unsuccessful converions being marked as successful one then they can go on to reduce the threshold for predicting a quote as successful but if working on limited resources then they cannot afford a missclassification in that case increase the

threshold to include only those quotes as successful for which the model is very sure that it will end up being a successful conversion.
3. No strict latency requirement is there for the given problem.
4. High interepretibilty of the model is desired as it would help the management understand what all factors have influenced the model to decide a successful or unsuccessful conversion.

2. MAPPING BUSINESS PROBLEM TO A ML PROBLEM

DATA OVERVIEW

1. Data for solving the problem is there in a 207 MB file train.csv which has 260753 quotes with each datapoint having 299 columns out of which majority of features are anonymized.
2. Features include specific coverage information, sales information, personal information, property information, and geographic information.

TYPE OF ML PROBLEM

Its a BINARY CLASSIFICATION problem where the task is to predict QuoteConversion_Flag for each QuoteNumber in the test set.

PERFORMANCE METRICS

1. LOG LOSS - Because it is a binary classification task and we are working with probability scores and this is a useful metric in such scenarios.
2. BINARY CONFUSION MATRIX - Gives us an insight into how our model is performing with various classes.
3. ROC Curve - Provides us with a value to compare performance amongst various models and helps in making a choice of right threshold by providing us with a best tradeoff possible between FPR & TPR
4. F1_Score - Since the nature of the dataset is such that there exists a natural imbalance in the dataset and we are more concerned about how well our model performs on successful conversions so in such scenarios F1_Score becomes a important metric.(Similar scenarios exists in Fraud Detection , Cancer Detection etc.)

TRAIN & CV SET CONSTRUCTION

Since the features being used are broadly of a NON-TEMPORAL nature. So, for such problem statements RANDOM SPLITTING is the stratergy we should opt for.

[1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import re
import math
import seaborn as sns
import scikitplot as skplt
import pickle
%matplotlib inline
from sklearn.preprocessing import OneHotEncoder
from sklearn.model_selection import train_test_split
from sklearn.feature_selection import VarianceThreshold
```

3. DATA ACQUISITION

[2]:
```python
data = pd.read_csv('homesite-quote-conversion/train.csv')
data.head()
```

[2]:
```
   QuoteNumber Original_Quote_Date  QuoteConversion_Flag Field6  Field7  \
0            1          2013-08-16                     0      B      23
1            2          2014-04-22                     0      F       7
2            4          2014-08-25                     0      F       7
3            6          2013-04-15                     0      J      10
4            8          2014-01-25                     0      E      23

   Field8  Field9 Field10  Field11 Field12  CoverageField1A  CoverageField1B  \
0  0.9403  0.0006     965   1.0200       N               17               23
1  1.0006  0.0040     548   1.2433       N                6                8
2  1.0006  0.0040     548   1.2433       N                7               12
3  0.9769  0.0004   1,165   1.2665       N                3                2
4  0.9472  0.0006   1,487   1.3045       N                8               13

   CoverageField2A  CoverageField2B  CoverageField3A  CoverageField3B  \
0               17               23               15               22
1                6                8                5                7
2                7               12                6               10
3                3                2                2                2
4                8               13                7               11

   CoverageField4A  CoverageField4B  CoverageField5A  CoverageField5B  \
0               16               22               13               22
1                5                8               13               22
2                7               11               25               25
3                3                2               13               22
4                7               13               13               22

   CoverageField6A  CoverageField6B CoverageField8 CoverageField9  \
0               13               23              T              D
1               13               23              T              E
2               13               23              T              J
3               13               23              Y              F
4               13               23              T              F

   CoverageField11A  CoverageField11B  SalesField1A  SalesField1B  \
0                 2                 1             7            18
1                 5                 9             5            14
2                 4                 6             3            10
3                15                23             8            19
4                 4                 6             3             6
```

	SalesField2A	SalesField2B	SalesField3	SalesField4	SalesField5	\
0	3	8	0	5	5	
1	6	18	1	5	5	
2	4	11	1	5	5	
3	14	24	0	5	5	
4	3	6	1	5	5	

	SalesField6	SalesField7	SalesField8	SalesField9	SalesField10	\
0	24	V	48649	0	0	
1	11	P	26778	0	0	
2	11	K	8751	0	0	
3	23	V	43854	0	0	
4	7	R	12505	1	0	

	SalesField11	SalesField12	SalesField13	SalesField14	SalesField15	\
0	0	0	0	0	0	
1	1	1	0	0	0	
2	2	2	0	0	0	
3	0	0	0	0	0	
4	0	0	0	0	0	

	PersonalField1	PersonalField2	PersonalField4A	PersonalField4B	\
0	0	0	-1	-1	
1	1	1	14	19	
2	1	1	16	21	
3	1	1	2	2	
4	1	1	20	24	

	PersonalField5	PersonalField6	PersonalField7	PersonalField8	\
0	7	0	N	1	
1	7	0	N	1	
2	7	0	N	1	
3	6	1	N	1	
4	7	0	N	1	

	PersonalField9	PersonalField10A	PersonalField10B	PersonalField11	\
0	2	5	7	0	
1	2	24	25	0	
2	2	7	16	0	
3	3	-1	-1	0	
4	2	5	8	0	

	PersonalField12	PersonalField13	PersonalField14	PersonalField15	\
0	1	2	2	24	
1	1	2	5	19	
2	1	2	3	7	
3	5	2	3	13	

```
4               1               2               2               24

  PersonalField16 PersonalField17 PersonalField18 PersonalField19  \
0              ZA              ZE              XR              XD
1              XB              YJ              YE              XT
2              ZH              XS              YP              XC
3              XO              XE              YI              XX
4              ZA              ZE              XR              XD

   PersonalField22  PersonalField23  PersonalField24  PersonalField25  \
0                1                0                0                0
1                1                0                0                0
2                1                0                0                0
3                1                0                0                0
4                1                0                0                0

   PersonalField26  PersonalField27  PersonalField28  PersonalField29  \
0                0                1                2                2
1                0                0                1                1
2                0                0                1                1
3                0                0                1                1
4                0                1                2                2

   PersonalField30  PersonalField31  PersonalField32  PersonalField33  \
0                0                1                1                1
1                0                0                0                0
2                0                0                0                0
3                0                0                0                0
4                0                1                1                1

   PersonalField34  PersonalField35  PersonalField36  PersonalField37  \
0                1                0                0                0
1                1                0                0                0
2                1                0                0                0
3                1                0                0                0
4                1                0                0                0

   PersonalField38  PersonalField39  PersonalField40  PersonalField41  \
0                0                0                0                0
1                0                0                0                0
2                0                0                0                0
3                0                0                0                0
4                0                0                0                0

   PersonalField42  PersonalField43  PersonalField44  PersonalField45  \
0                0                1                0                1
1                0                1                0                0
```

```
2              0              1              0              0
3              0              1              0              0
4              0              1              0              1

   PersonalField46  PersonalField47  PersonalField48  PersonalField49  \
0                1                1                2                0
1                0                0                1                0
2                0                0                1                0
3                0                0                1                0
4                1                1                2                0

   PersonalField50  PersonalField51  PersonalField52  PersonalField53  \
0                0                0                0                1
1                0                0                0                1
2                0                0                0                1
3                0                0                0                1
4                0                0                0                1

   PersonalField54  PersonalField55  PersonalField56  PersonalField57  \
0                0                0                0                0
1                0                0                0                0
2                0                0                0                0
3                0                0                0                0
4                0                0                0                0

   PersonalField58  PersonalField59  PersonalField60  PersonalField61  \
0                1                0                0                0
1                1                0                0                0
2                1                0                0                0
3                1                0                0                0
4                1                0                0                0

   PersonalField62  PersonalField63  PersonalField64  PersonalField65  \
0                0                1                0                0
1                0                1                0                0
2                0                1                0                0
3                0                1                0                0
4                0                1                0                0

   PersonalField66  PersonalField67  PersonalField68  PersonalField69  \
0                0                0                1                0
1                0                0                1                0
2                0                0                1                0
3                0                0                1                0
4                0                0                1                0

   PersonalField70  PersonalField71  PersonalField72  PersonalField73  \
```

```
0               0                0                0                1
1               0                0                0                1
2               0                0                0                1
3               0                0                0                1
4               0                0                0                1

   PersonalField74  PersonalField75  PersonalField76  PersonalField77  \
0                0                1                1                1
1                0                0                0                0
2                0                0                0                0
3                0                0                0                0
4                0                1                1                1

   PersonalField78  PersonalField79  PersonalField80  PersonalField81  \
0                2                0                0                0
1                1                0                0                0
2                1                0                0                0
3                1                0                0                0
4                2                0                0                0

   PersonalField82  PersonalField83  PersonalField84  PropertyField1A  \
0                0                1              2.0                5
1                0                1              NaN               17
2                0                1              NaN               12
3                0                1              NaN                1
4                0                1              2.0               10

   PropertyField1B  PropertyField2A  PropertyField2B PropertyField3  \
0                7               -1               19              N
1               21               -1               22              N
2               16               -1               13              Y
3                1               -1                2              N
4               13               -1               19              Y

  PropertyField4 PropertyField5  PropertyField6 PropertyField7  \
0              N              Y               0              O
1              N              Y               0              N
2              Y              Y               0              R
3              N              Y               0              R
4              Y              Y               0              D

   PropertyField8  PropertyField9  PropertyField10  PropertyField11A  \
0               0               0                1                -1
1               0               0                1                -1
2               1               0                1                -1
3               1               0                1                -1
4               0               0                1                -1
```

```
  PropertyField11B  PropertyField12  PropertyField13 PropertyField14  \
0               21                4                2               C
1               23                3                0               B
2               21                4                2               C
3               21                4                2               C
4               21                1                2               A

  PropertyField15  PropertyField16A  PropertyField16B  PropertyField17  \
0               4                14                22                0
1               4                 7                15                0
2               4                 9                17                1
3               1                 3                 6                0
4               4                 5                11                0

  PropertyField18  PropertyField19  PropertyField20  PropertyField21A  \
0               3                0                0                17
1               1                0                0                 6
2               3                0                0                 7
3               2                0                0                 3
4               0                0                0                 8

  PropertyField21B  PropertyField22  PropertyField23  PropertyField24A  \
0               23                2                4                15
1                8                2                1                 6
2               12                2                1                15
3                2                2                1                 7
4               13                2                1                13

  PropertyField24B  PropertyField25  PropertyField26A  PropertyField26B  \
0               21              3.0                 5                 5
1                8              2.0                 4                 3
2               21              1.5                14                20
3                9              1.0                13                18
4               19              3.0                 4                 4

  PropertyField27 PropertyField28  PropertyField29 PropertyField30  \
0               6               B              0.0               N
1               3               B              NaN               N
2               3               B              NaN               N
3              10               B              NaN               N
4               4               B              0.0               N

 PropertyField31 PropertyField32 PropertyField33 PropertyField34  \
0               N               Y               G               Y
1               O               N               H               Y
2               K               Y               H               Y
```

```
3              O              Y              G              N
4              O              N              H              N

   PropertyField35 PropertyField36 PropertyField37 PropertyField38  \
0                2               N               N               N
1                2               N               N               N
2                2               N               N               N
3                2               N               Y               N
4                0               N               N               N

   PropertyField39A  PropertyField39B  GeographicField1A  GeographicField1B  \
0                 7                 5                  3                  6
1                11                12                  8                 17
2                23                24                  2                  4
3                24                25                  2                  3
4                18                22                  2                  2

   GeographicField2A  GeographicField2B  GeographicField3A  GeographicField3B  \
0                 15                 14                 12                 13
1                 19                 20                 12                 14
2                 20                 22                 18                 20
3                 10                  8                  8                  9
4                 17                 18                 13                 15

   GeographicField4A  GeographicField4B  GeographicField5A  GeographicField5B  \
0                 20                 22                 -1                 13
1                 16                 18                 -1                 19
2                 16                 18                 -1                 21
3                 21                 23                 -1                 23
4                  9                  7                 -1                 13

   GeographicField6A  GeographicField6B  GeographicField7A  GeographicField7B  \
0                  2                  5                  1                  1
1                  4                 13                  9                 13
2                  4                 12                 10                 14
3                 13                 21                 22                 22
4                  9                 18                 11                 16

   GeographicField8A  GeographicField8B  GeographicField9A  GeographicField9B  \
0                  1                  1                  2                  9
1                  5                 13                  5                 12
2                  5                 13                  6                 14
3                 14                 21                 18                 21
4                  9                 17                 12                 17

   GeographicField10A  GeographicField10B  GeographicField11A  \
0                  -1                  25                   1
```

```
1                  -1                  25                   4
2                  -1                  25                   4
3                  -1                  25                  15
4                  -1                  25                   9

   GeographicField11B  GeographicField12A  GeographicField12B  \
0                   1                   1                   1
1                  15                   8                  13
2                  11                   9                  15
3                  22                  21                  22
4                  17                  12                  16

   GeographicField13A  GeographicField13B  GeographicField14A  \
0                   1                   1                  -1
1                   5                  14                  -1
2                   5                  12                  -1
3                  15                  22                  -1
4                   9                  17                  -1

   GeographicField14B  GeographicField15A  GeographicField15B  \
0                   7                   1                   2
1                  13                   9                  11
2                  12                  10                  16
3                  15                  24                  23
4                  22                   9                   8

   GeographicField16A  GeographicField16B  GeographicField17A  \
0                   1                   2                  13
1                   5                  12                   2
2                   6                  14                   2
3                   9                  18                   2
4                  10                  20                   2

   GeographicField17B  GeographicField18A  GeographicField18B  \
0                  22                  -1                  12
1                  16                  -1                  22
2                  13                  -1                  17
3                   3                  -1                   6
4                   6                  -1                   4

   GeographicField19A  GeographicField19B  GeographicField20A  \
0                   3                   6                   4
1                  23                  24                  20
2                  21                  21                  20
3                  10                  13                   2
4                  17                  16                   5
```

```
   GeographicField20B  GeographicField21A  GeographicField21B  \
0                  12                  -1                  10
1                  22                  -1                  22
2                  21                  -1                  19
3                   3                  -1                   8
4                  13                  -1                  22

   GeographicField22A  GeographicField22B  GeographicField23A  \
0                  -1                  15                  -1
1                  -1                  22                  -1
2                  -1                  19                  -1
3                  -1                  15                  -1
4                  25                  25                  -1

   GeographicField23B  GeographicField24A  GeographicField24B  \
0                   5                  14                   9
1                  13                  10                   4
2                  10                  18                  17
3                  24                  19                  20
4                  15                  17                  16

   GeographicField25A  GeographicField25B  GeographicField26A  \
0                  16                  16                  15
1                  10                   5                   5
2                  12                   8                   7
3                  13                  10                   7
4                  17                  17                  10

   GeographicField26B  GeographicField27A  GeographicField27B  \
0                  18                  17                  16
1                   5                  13                   6
2                   8                  15                  11
3                  10                  16                  14
4                  13                  21                  20

   GeographicField28A  GeographicField28B  GeographicField29A  \
0                  11                  18                  24
1                   7                  10                  21
2                   6                   9                   8
3                   3                   3                  11
4                   9                  14                   5

   GeographicField29B  GeographicField30A  GeographicField30B  \
0                  25                   5                   4
1                  24                  10                  19
2                  11                   7                  12
3                  16                   2                   1
```

```
4                  4                 10                 18

   GeographicField31A  GeographicField31B  GeographicField32A  \
0                   7                  10                  13
1                   7                  10                   9
2                  13                  21                  11
3                  10                  15                   3
4                   8                  11                   6

   GeographicField32B  GeographicField33A  GeographicField33B  \
0                  20                  14                  21
1                  15                   4                   3
2                  18                  12                  19
3                   3                  14                  21
4                   8                  11                  17

   GeographicField34A  GeographicField34B  GeographicField35A  \
0                  12                  14                  14
1                  13                  17                   6
2                  15                  20                  10
3                   1                   1                   9
4                  16                  21                   7

   GeographicField35B  GeographicField36A  GeographicField36B  \
0                  22                   3                   2
1                   9                  17                  23
2                  17                   9                  13
3                  16                   3                   2
4                  11                   8                  11

   GeographicField37A  GeographicField37B  GeographicField38A  \
0                  15                  24                   2
1                  11                  21                  13
2                   4                  10                  14
3                   4                   8                   4
4                   3                   6                  12

   GeographicField38B  GeographicField39A  GeographicField39B  \
0                   2                  15                  23
1                  21                   9                  17
2                  22                  17                  24
3                   3                   9                  18
4                  19                   4                   6

   GeographicField40A  GeographicField40B  GeographicField41A  \
0                   1                   1                  15
1                  12                  21                  25
```

```
2                 10                 16                 24
3                  2                  1                 13
4                 10                 15                 18

   GeographicField41B  GeographicField42A  GeographicField42B  \
0                 13                  9                 10
1                 25                 15                 21
2                 22                 12                 15
3                  9                  7                  7
4                 17                 24                 25

   GeographicField43A  GeographicField43B  GeographicField44A  \
0                  2                  2                  8
1                 10                 13                 23
2                 12                 20                 21
3                 25                 25                  3
4                 10                 12                 24

   GeographicField44B  GeographicField45A  GeographicField45B  \
0                  4                 20                 22
1                 24                 11                 15
2                 22                 24                 25
3                  1                 14                 22
4                 25                  9                 11

   GeographicField46A  GeographicField46B  GeographicField47A  \
0                 10                  8                  6
1                 21                 24                  6
2                 20                 22                  7
3                  6                  2                  7
4                 25                 25                  5

   GeographicField47B  GeographicField48A  GeographicField48B  \
0                  5                 15                 13
1                 11                 21                 21
2                 13                 23                 23
3                 14                 11                  8
4                  3                 22                 22

   GeographicField49A  GeographicField49B  GeographicField50A  \
0                 19                 18                 16
1                 18                 15                 20
2                 20                 19                 20
3                 19                 18                 18
4                 21                 21                 17

   GeographicField50B  GeographicField51A  GeographicField51B  \
```

0	14	21	23
1	20	13	12
2	20	18	20
3	16	13	12
4	15	25	25

	GeographicField52A	GeographicField52B	GeographicField53A	\
0	21	23	16	
1	12	12	15	
2	19	21	20	
3	13	12	17	
4	25	25	17	

	GeographicField53B	GeographicField54A	GeographicField54B	\
0	11	22	24	
1	9	13	11	
2	19	11	8	
3	13	5	2	
4	13	13	11	

	GeographicField55A	GeographicField55B	GeographicField56A	\
0	7	14	-1	
1	11	20	-1	
2	3	3	-1	
3	3	4	-1	
4	3	4	-1	

	GeographicField56B	GeographicField57A	GeographicField57B	\
0	17	15	17	
1	9	18	21	
2	5	21	24	
3	7	14	14	
4	7	11	9	

	GeographicField58A	GeographicField58B	GeographicField59A	\
0	14	18	9	
1	8	7	10	
2	12	15	15	
3	14	18	6	
4	10	10	18	

	GeographicField59B	GeographicField60A	GeographicField60B	\
0	9	-1	8	
1	10	-1	11	
2	18	-1	21	
3	5	-1	10	
4	22	-1	10	

```
   GeographicField61A  GeographicField61B  GeographicField62A  \
0                  -1                  18                  -1
1                  -1                  17                  -1
2                  -1                  11                  -1
3                  -1                   9                  -1
4                  -1                  11                  -1

   GeographicField62B GeographicField63 GeographicField64
0                  10                 N                CA
1                  20                 N                NJ
2                   8                 N                NJ
3                  21                 N                TX
4                  12                 N                IL
```

4. EXPLORATORY DATA ANALYSIS

[3]:
```
def extract_feature_dataset(feature,data):
    """
    This functions takes feature name as input and a dataset containing all
    the sub features realted to that particular feature
    """
    features = []
    for column in data.columns:
        if re.search('^'+feature+'\w+$',column) != None:
            features.append(column)
    return pd.DataFrame(data.loc[:,features + ['QuoteConversion_Flag']])
```

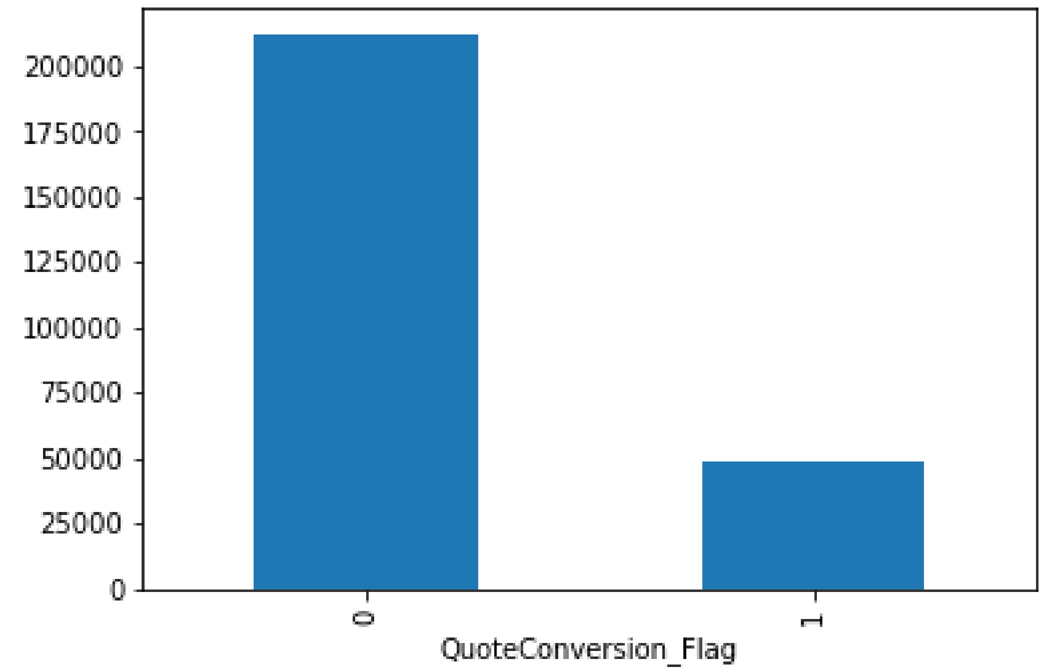
[4]:
```
def get_numerical_features(dataset):
    """Returns a list of all the numerical features present in the dataset"""
    numerical_features = []
    for feature in dataset.columns:
        if dataset[feature].dtypes != 'O':
            numerical_features.append(feature)
    return numerical_features
```

[5]:
```
def get_categorical_features(dataset):
    """Returns a list of all the categorical features present in the dataset"""
    categorical_features = []
    for feature in dataset.columns:
        if dataset[feature].dtypes == 'O':
            categorical_features.append(feature)
    return categorical_features
```

DISTIBUTION OF DATASET AMONG OUTPUT CLASSES (SUCCESSFUL AND NOT SUCCESSFUL QUOTE CONVERSIONS)

[6]:
```
data.groupby('QuoteConversion_Flag')['QuoteConversion_Flag'].count().plot.bar()
```

[6]: <AxesSubplot:xlabel='QuoteConversion_Flag'>



[7]:
```
print('Total number of quotes for which conversion was unsuccessful are', data.
 ↪groupby('QuoteConversion_Flag')['QuoteConversion_Flag'].count()[0]/
 ↪data['QuoteConversion_Flag'].count() * 100)
print('Total number of quotes for which conversion was successful are', data.
 ↪groupby('QuoteConversion_Flag')['QuoteConversion_Flag'].count()[1]/
 ↪data['QuoteConversion_Flag'].count() * 100)
```

```
Total number of quotes for which conversion was unsuccessful are
81.24892139304247
Total number of quotes for which conversion was successful are 18.75107860695754
```

OBSERVATION: The given dataset is a IMBALANCED DATASET.

MISSING VALUES

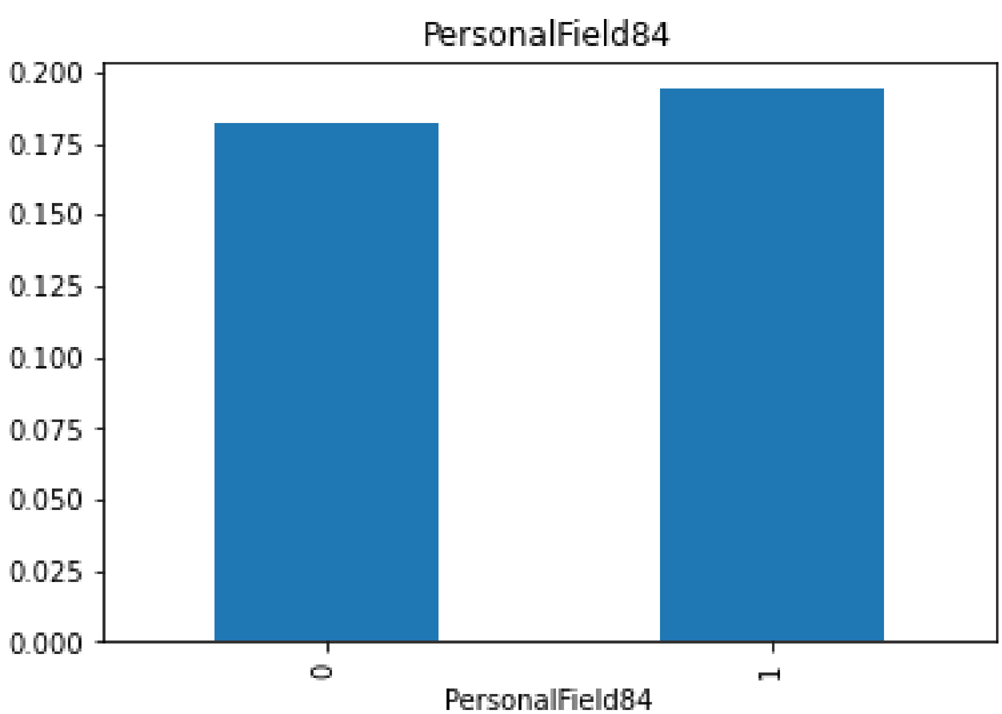
[8]:
```
def identify_features_na(data):
    """This function takes a dataset as input and return a list of
    columns for which contain a null value"""
    features_with_na = []
    for column in data.columns:
        if data[column].isnull().sum() > 1:
            features_with_na.append(column)
    return features_with_na
```

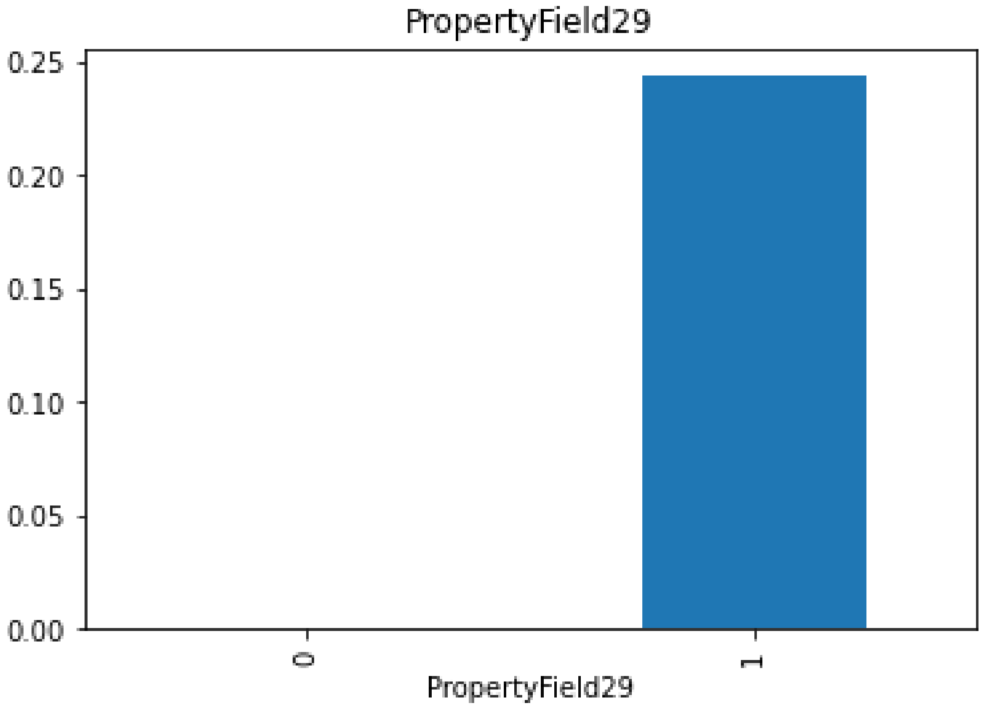
[9]:
```
for feature in identify_features_na(data):
    print(feature, np.round(data[feature].isnull().mean(),5)* 100, '% missing␣
↪values')
```

```
PersonalField7 0.043 % missing values
PersonalField84 47.634 % missing values
PropertyField3 0.031 % missing values
PropertyField4 0.024 % missing values
PropertyField29 76.964 % missing values
PropertyField32 0.027 % missing values
PropertyField34 0.027 % missing values
PropertyField36 0.043 % missing values
PropertyField38 0.468 % missing values
```

Since only two features PersonalField84 & PropertyField29 have significant amount of null values will explore if the presence of null value in both these features is providing us with any meaningful information. For the rest of the features since the number of values that are null are extremely low in number so would avoid making any conclusions out of it.

[10]:
```
for feature in ['PersonalField84', 'PropertyField29']:
    d = data.copy()
    d[feature] = np.where(data[feature].isnull(),1,0)
    d.groupby(feature)['QuoteConversion_Flag'].mean().plot.bar()
    plt.title(feature)
    plt.show()
```





OBSERVATIONS:

1. If PropertyField29 is NOT NULL then the chance of successful conversion is extremely low. so we can say that presence of a NULL value for this feature provides some meaningful information which is helping decide successful and not successful conversions.
2. Whereas PersonalField84 has equal proportion (Approximately 17.5% of transactions as successful and rest as unsucessful) of successful and not successful conversions in both the cases where its value is NULL and where it is NOT NULL.So, presence of NULL value in this case does not seem to provide any additional information

HANDLING MISSING VALUES FOR CATEGORICAL FEATURES

[11]:
```
features_categorical = get_categorical_features(data.loc[:
 ↪,identify_features_na(data)])
features_categorical
```

[11]:
```
['PersonalField7',
 'PropertyField3',
 'PropertyField4',
 'PropertyField32',
 'PropertyField34',
 'PropertyField36',
 'PropertyField38']
```

[12]:
```
def replace_categorical_feature_na(data, categorical_feature_nan):
    dataset = data.copy()
    dataset[categorical_feature_nan] = dataset[categorical_feature_nan].
 ↪fillna('Missing')
    return dataset
```

[13]:
```
data = replace_categorical_feature_na(data, features_categorical)
data[features_categorical].isnull().sum()
```

[13]:
```
PersonalField7     0
PropertyField3     0
PropertyField4     0
PropertyField32    0
PropertyField34    0
PropertyField36    0
PropertyField38    0
dtype: int64
```

HANDLING MISSING VALUES FOR NUMERICAL FEATURES

[14]:
```
features_numerical = get_numerical_features(data.loc[:
 ↪,identify_features_na(data)])
features_numerical
```

[14]:
```
['PersonalField84', 'PropertyField29']
```

[15]:
```
data[['PersonalField84_nan', 'PropertyField29_nan']] = np.
 ↪where(data[features_numerical].isnull(),1,0)
```

[16]:
```
data[features_numerical] = data[features_numerical].fillna(100)
```

ANALYSING FEILD, COVERAGE, SALES, PERSONAL, PROPERTY & GEOGRAPHIC FEATURES

FIELD FEATURES

[17]:
```
dataset = extract_feature_dataset('Field',data)
```

[18]:
```
for feature in dataset.columns:
    if feature != 'QuoteConversion_Flag':
        print('{} has {} unique values'.format(feature,len(dataset[feature].
 ↪unique())))
```

```
Field6 has 8 unique values
Field7 has 28 unique values
Field8 has 38 unique values
Field9 has 5 unique values
Field10 has 8 unique values
```

```
Field11 has 11 unique values
Field12 has 2 unique values
```

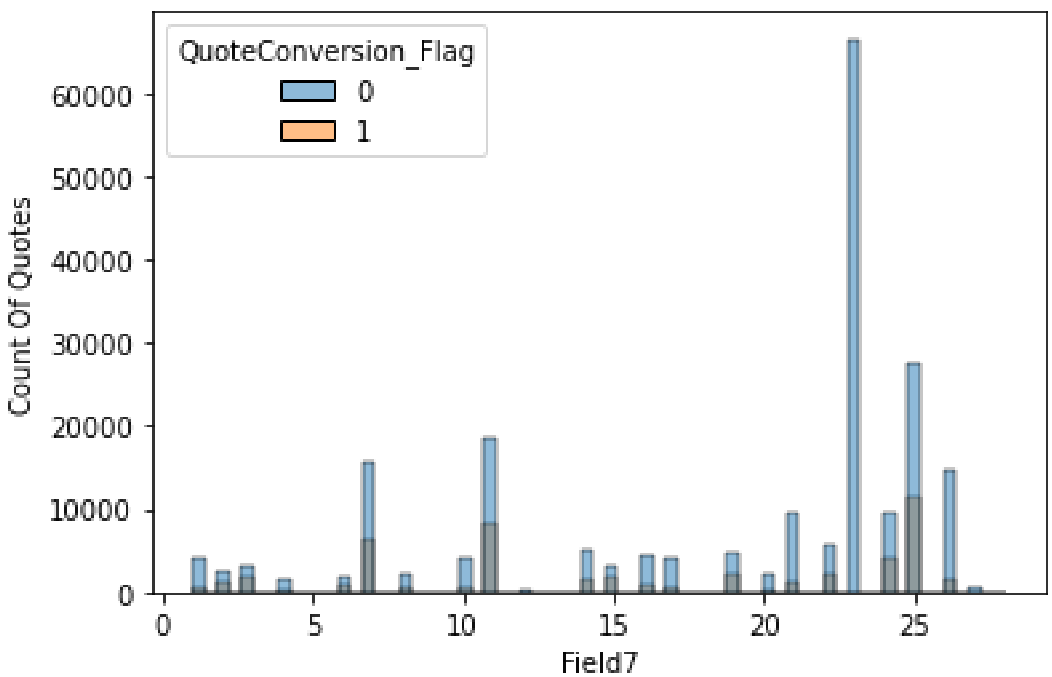
[19]:
```
numerical_features = get_numerical_features(dataset)
categorical_features = get_categorical_features(dataset)
```

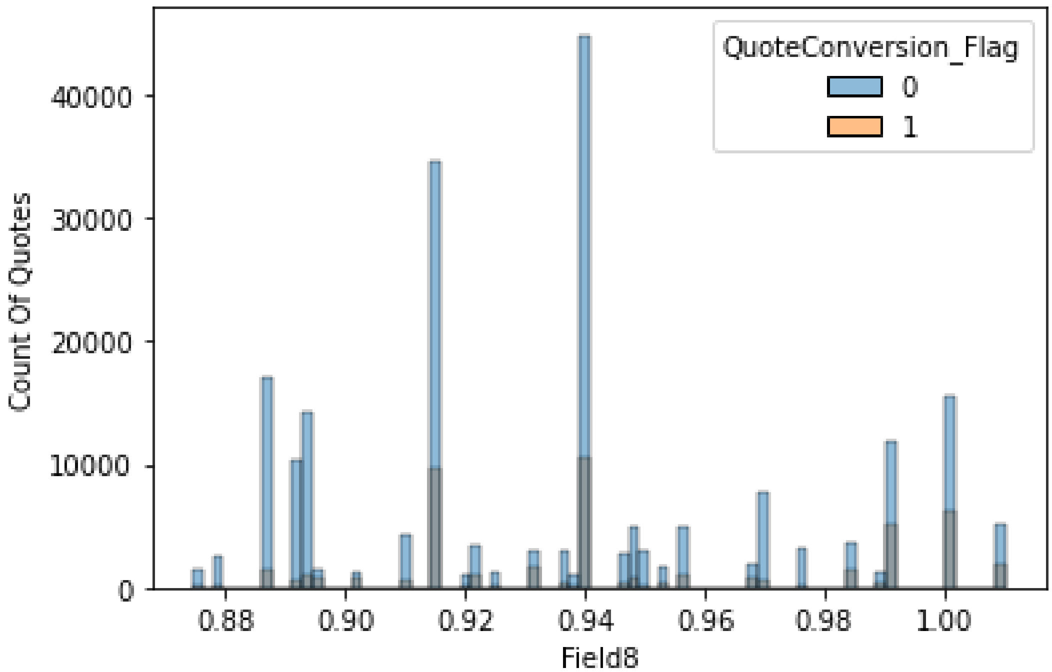
[20]:
```
print('Numerical Features : ', numerical_features)
print('Categorical Features : ', categorical_features)
```

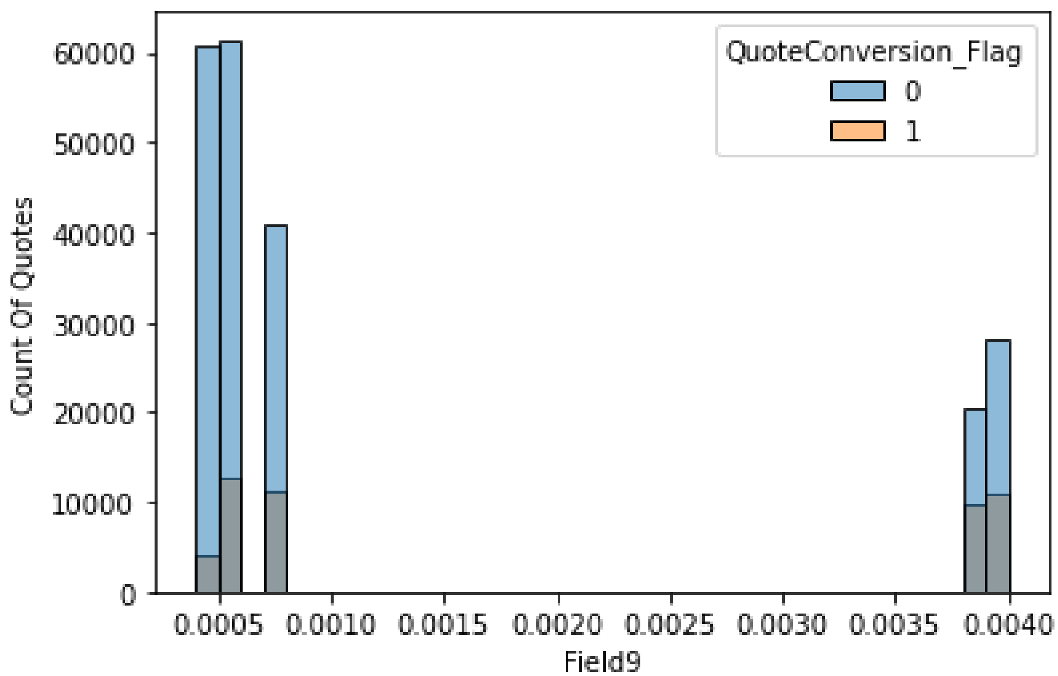
```
Numerical Features :  ['Field7', 'Field8', 'Field9', 'Field11',
'QuoteConversion_Flag']
Categorical Features :  ['Field6', 'Field10', 'Field12']
```

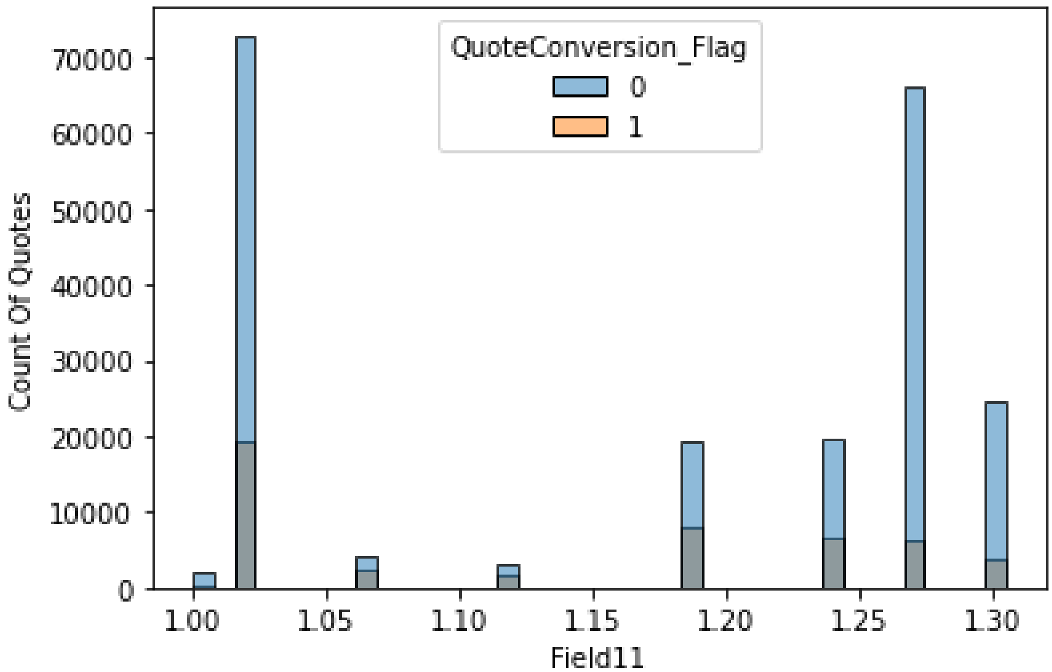
DISCRETE NUMERICAL FEATURES

[21]:
```
for feature in get_numerical_features(dataset):
    if feature != 'QuoteConversion_Flag' and len(dataset[feature].unique()) <␣
 ↪40:
        sns.histplot(data = dataset, x = feature, hue = 'QuoteConversion_Flag')
        plt.xlabel(feature)
        plt.ylabel('Count Of Quotes')
        plt.show()
```



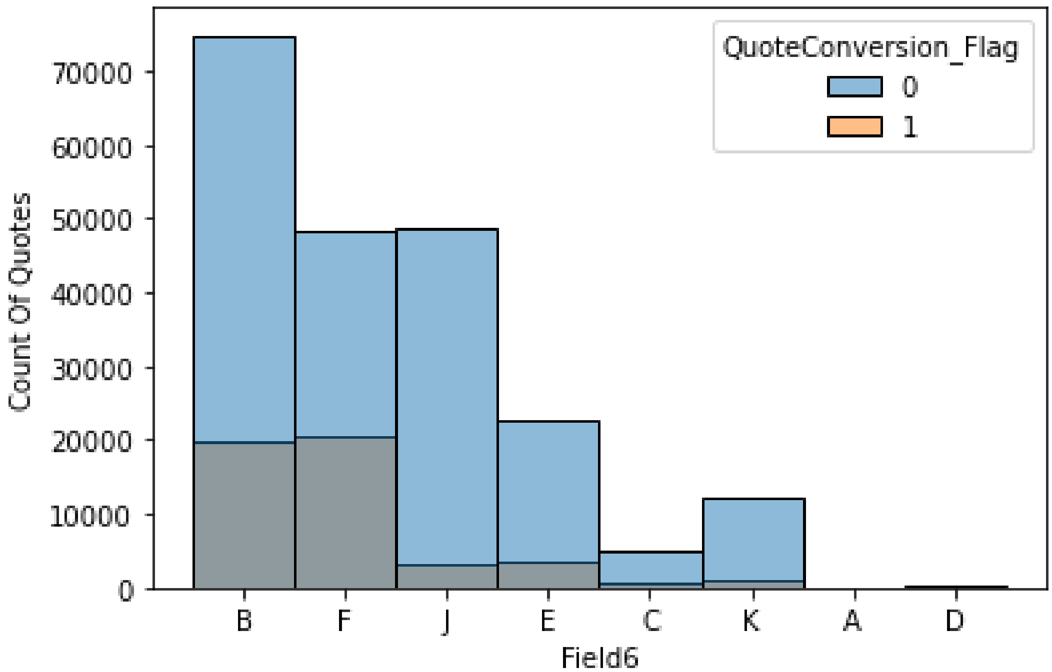

QuoteConversion_Flag
0
1
Count Of Quotes
40000
30000
20000
10000
0
0.88
0.90
0.92
0.94
0.96
0.98
1.00
Field8

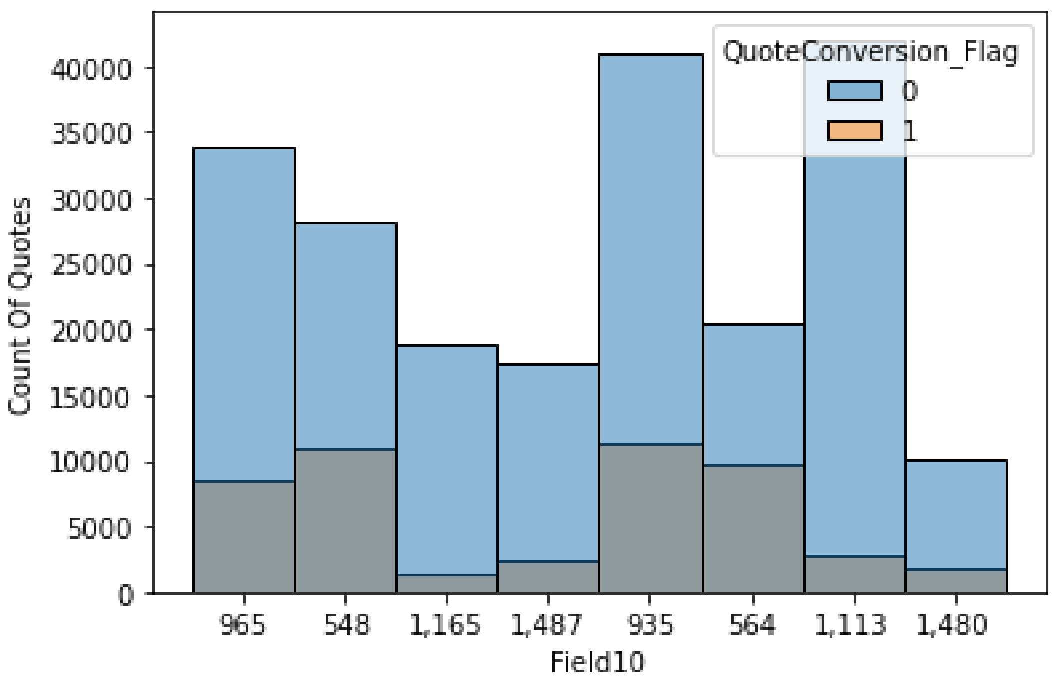

QuoteConversion_Flag
0
1
Count Of Quotes
60000
50000
40000
30000
20000
10000
0
0.0005
0.0010
0.0015
0.0020
0.0025
0.0030
0.0035
0.0040
Field9

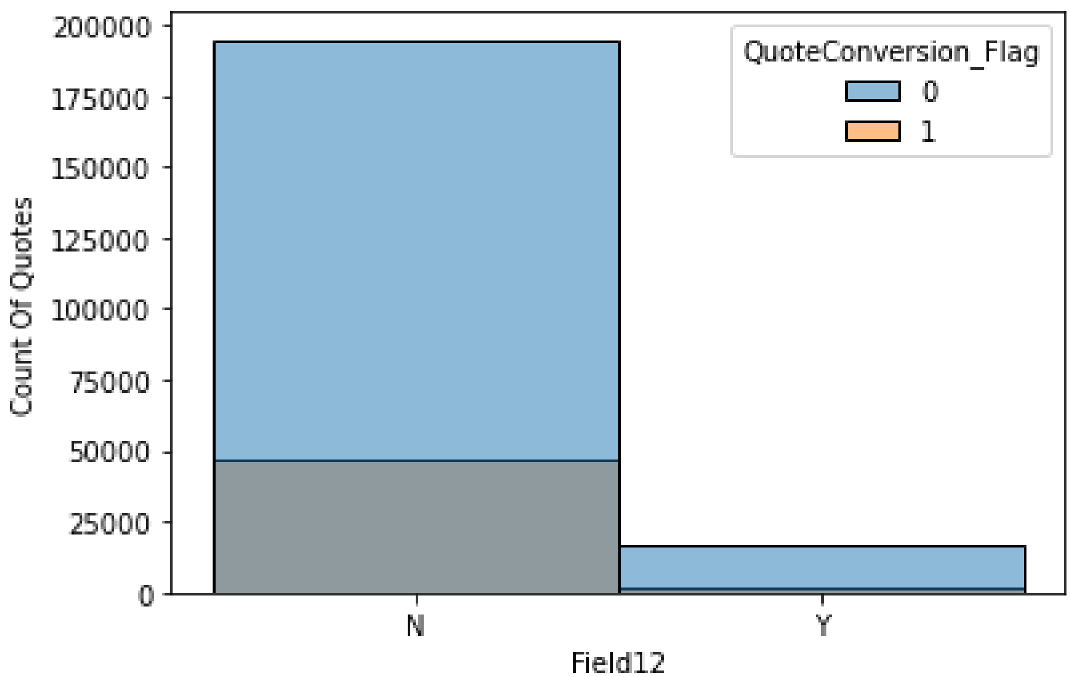


CATEGORICAL FEATURES

[22]:
```
for feature in get_categorical_features(dataset):
        sns.histplot(data = dataset, x = feature, hue = 'QuoteConversion_Flag')
        plt.xlabel(feature)
        plt.ylabel('Count Of Quotes')
        plt.show()
```


QuoteConversion_Flag
0
1
Count Of Quotes
70000
60000
50000
40000
30000
20000
10000
0
B
F
J
E
C
K
A
D
Field6


QuoteConversion_Flag
0
1
Count Of Quotes
40000
35000
30000
25000
20000
15000
10000
5000
0
965
548
1,165
1,487
935
564
1,113
1,480
Field10



OBSERVATIONS:

1. DISCRETE NUMERICAL FEATURES

   A. The dicrete feature distributions does not seem to be following any standard distibutions. But what can be obsereved is that for certian discrete values proportion of successful conversion out of total quotes made having that discrete value is high and in some cases there seems to be no or negligible conversion ratio which can be a useful information. Like for example we can say that the chance of a quote getting converted into a successful one is high when we have a higher value of Field9 in our quote.

2. CATEGORICAL NUMERICAL FEATURES

   A. Field6 taking a value B or F dractically increases its chance of being a successful conversion.

   B. Filed10 there seems to be a realtionship b/w magnitude of the value that feature takes and its chance of being a successful conversion. The smaller the value the higher the chance.

   C. Field12 assuming a value Y drastically reduces the chance of quote being a successful conversion.

COVERAGE FIELD

[23]:
```
dataset = extract_feature_dataset('CoverageField',data)
```

[24]:
```
for feature in dataset.columns:
    if feature != 'QuoteConversion_Flag':
        print('{} has {} unique values'.format(feature,len(dataset[feature].
 ↪unique())))
```

```
CoverageField1A has 26 unique values
CoverageField1B has 26 unique values
CoverageField2A has 25 unique values
CoverageField2B has 25 unique values
CoverageField3A has 25 unique values
CoverageField3B has 25 unique values
CoverageField4A has 25 unique values
CoverageField4B has 25 unique values
CoverageField5A has 3 unique values
CoverageField5B has 4 unique values
CoverageField6A has 3 unique values
CoverageField6B has 4 unique values
CoverageField8 has 7 unique values
CoverageField9 has 12 unique values
CoverageField11A has 26 unique values
CoverageField11B has 26 unique values
```

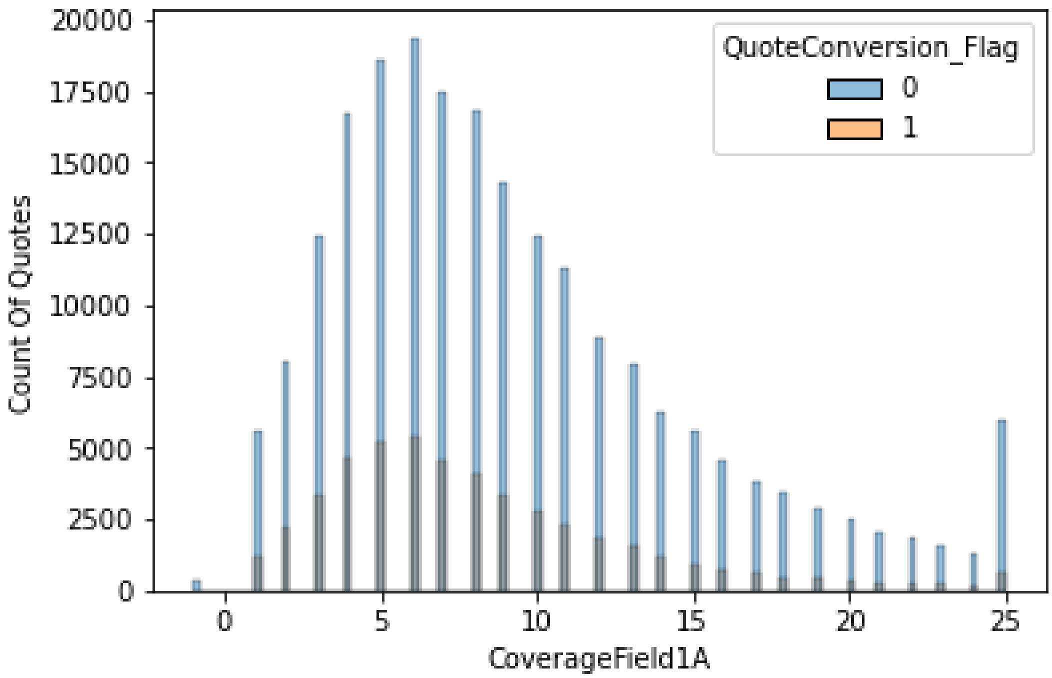
[25]:
```
numerical_features = get_numerical_features(dataset)
categorical_features = get_categorical_features(dataset)
```

[26]:
```
print('Numerical Features : ', numerical_features)
print('Categorical Features : ', categorical_features)
```

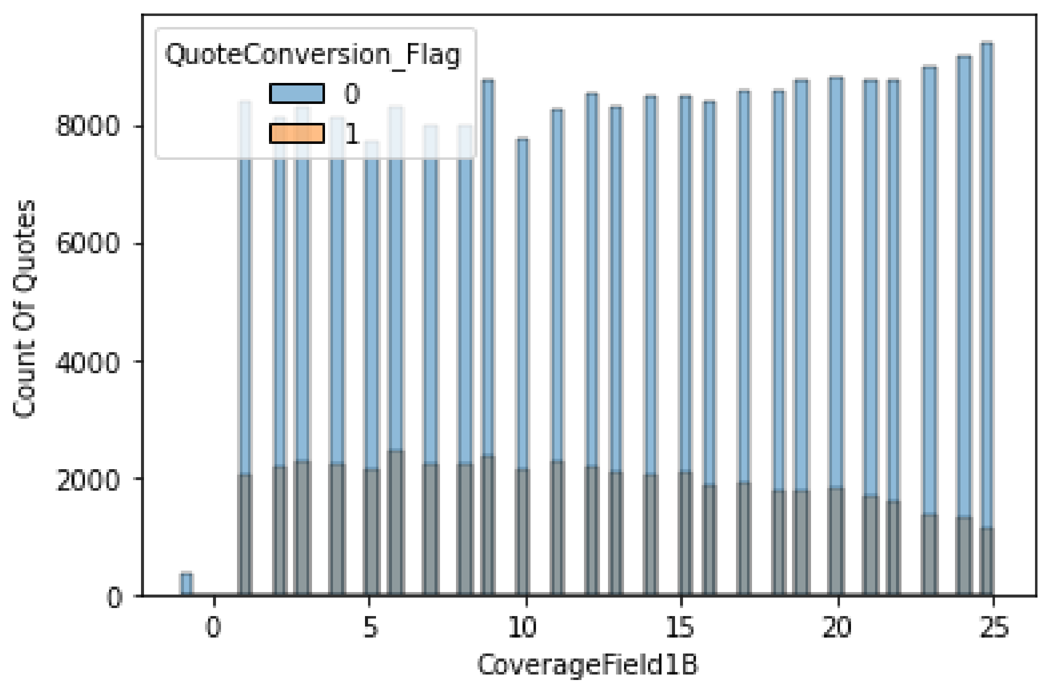
```
Numerical Features :  ['CoverageField1A', 'CoverageField1B', 'CoverageField2A',
'CoverageField2B', 'CoverageField3A', 'CoverageField3B', 'CoverageField4A',
'CoverageField4B', 'CoverageField5A', 'CoverageField5B', 'CoverageField6A',
'CoverageField6B', 'CoverageField11A', 'CoverageField11B',
'QuoteConversion_Flag']
Categorical Features :  ['CoverageField8', 'CoverageField9']
```

DISCRETE NUMERICAL FEATURES

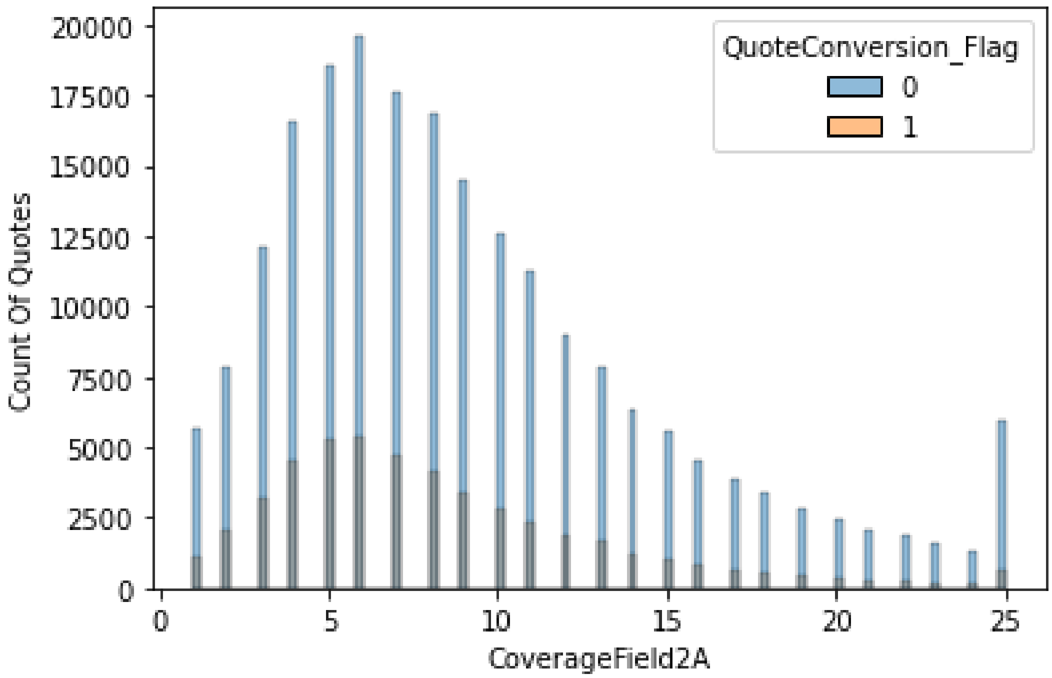
[27]:
```
for feature in get_numerical_features(dataset):
    if feature != 'QuoteConversion_Flag' and len(dataset[feature].unique()) <␣
↪30:
        sns.histplot(data = dataset, x = feature, hue = 'QuoteConversion_Flag')
        plt.xlabel(feature)
        plt.ylabel('Count Of Quotes')
        plt.show()
```



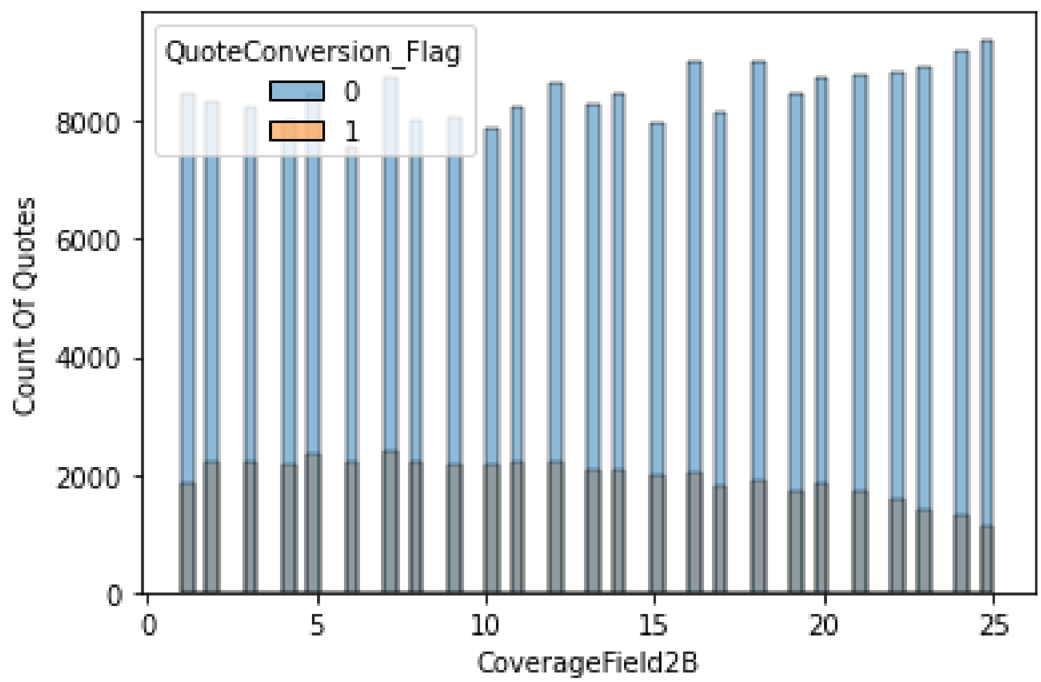
QuoteConversion_Flag
0
1
Count Of Quotes
CoverageField1A



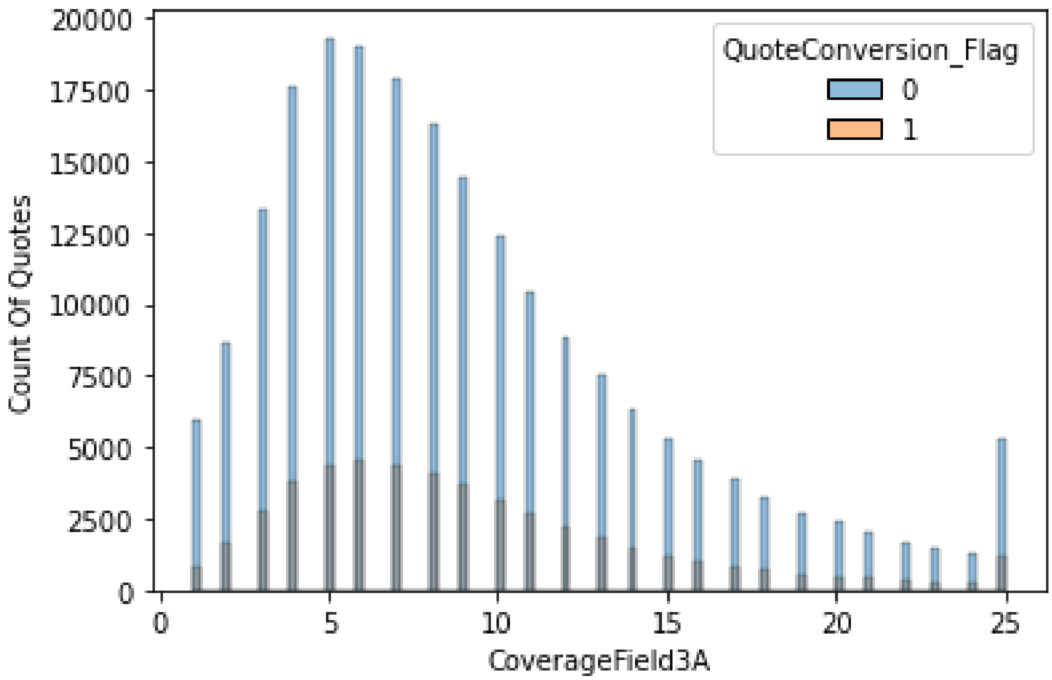
QuoteConversion_Flag
0
1
Count Of Quotes
CoverageField1B

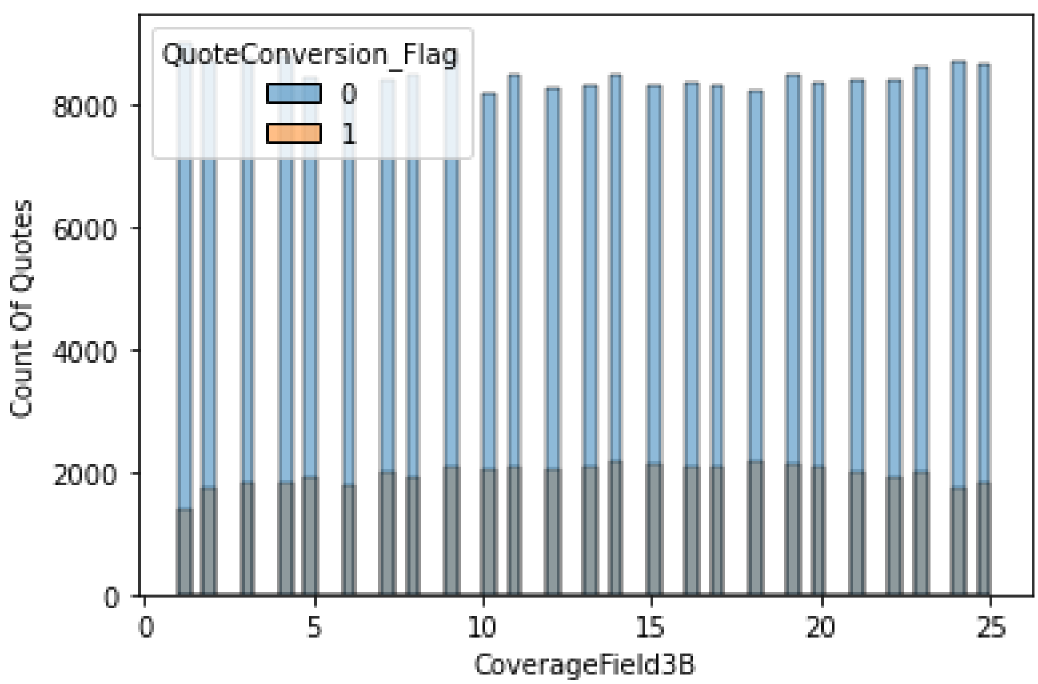


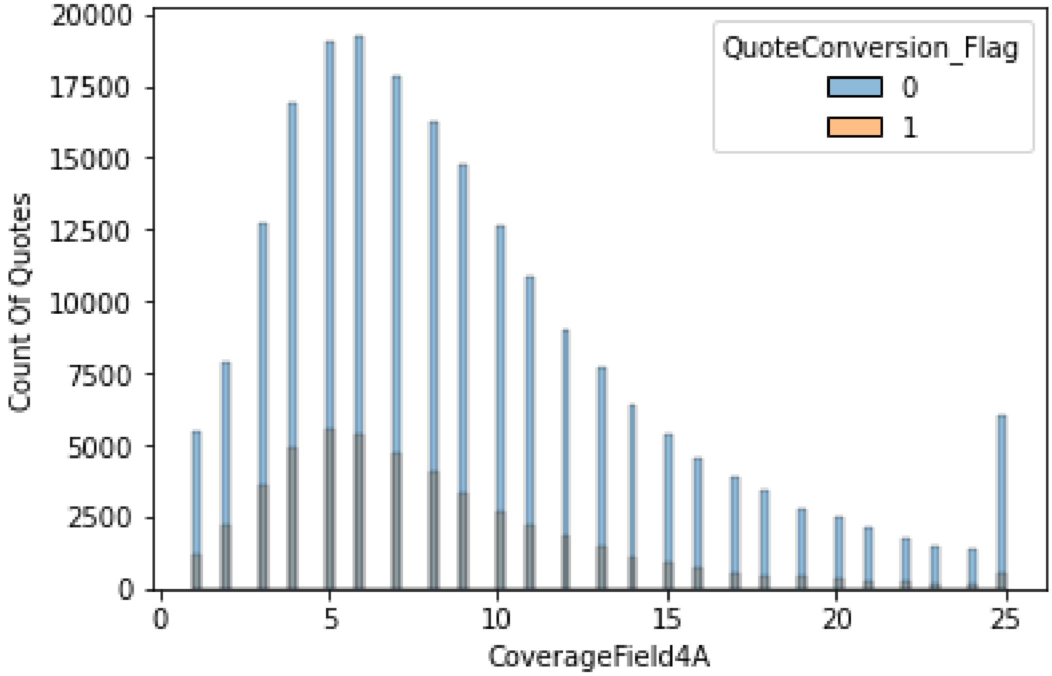
QuoteConversion_Flag
0
1
Count Of Quotes
CoverageField2A

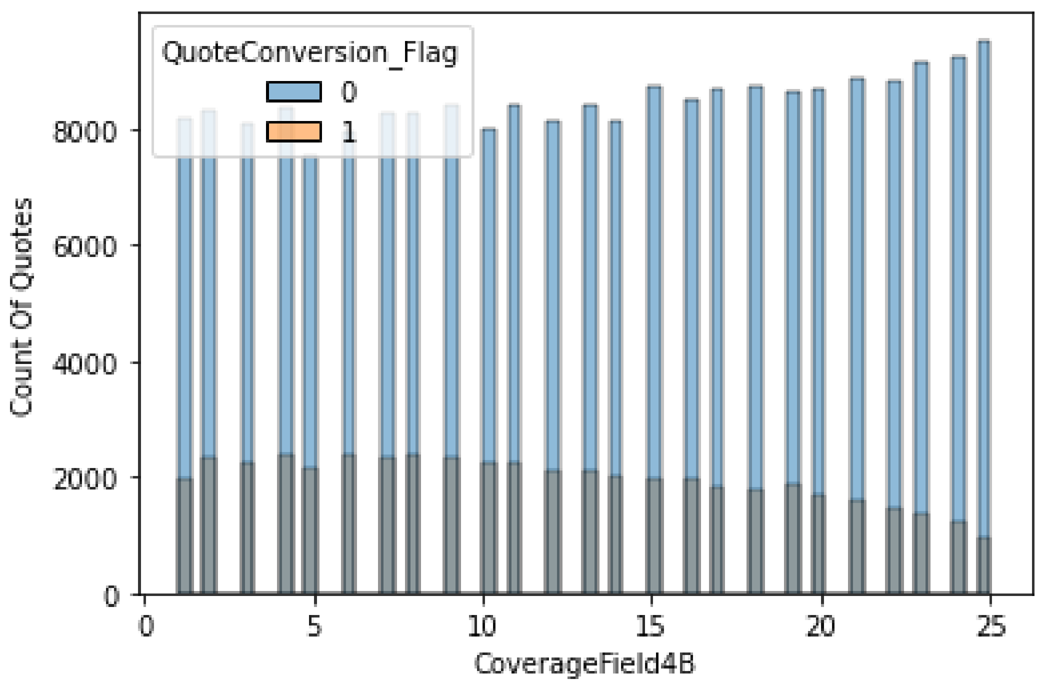


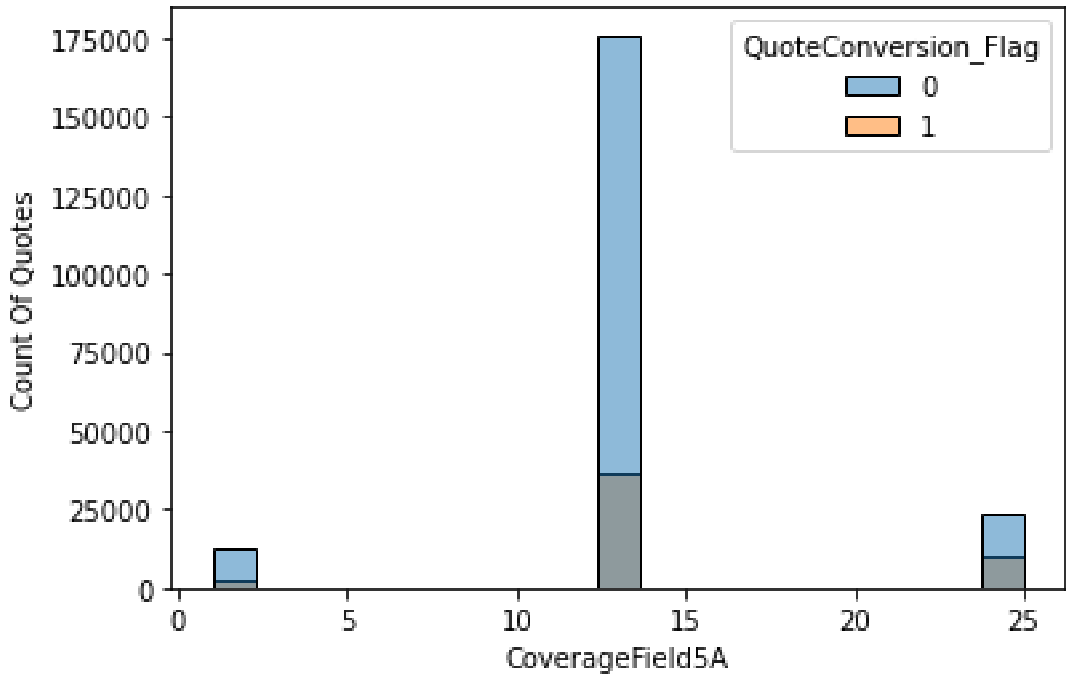
QuoteConversion_Flag
0
1
Count Of Quotes
CoverageField2B


QuoteConversion_Flag
0
1
Count Of Quotes
CoverageField3A
0
2500
5000
7500
10000
12500
15000
17500
20000
0
5
10
15
20
25

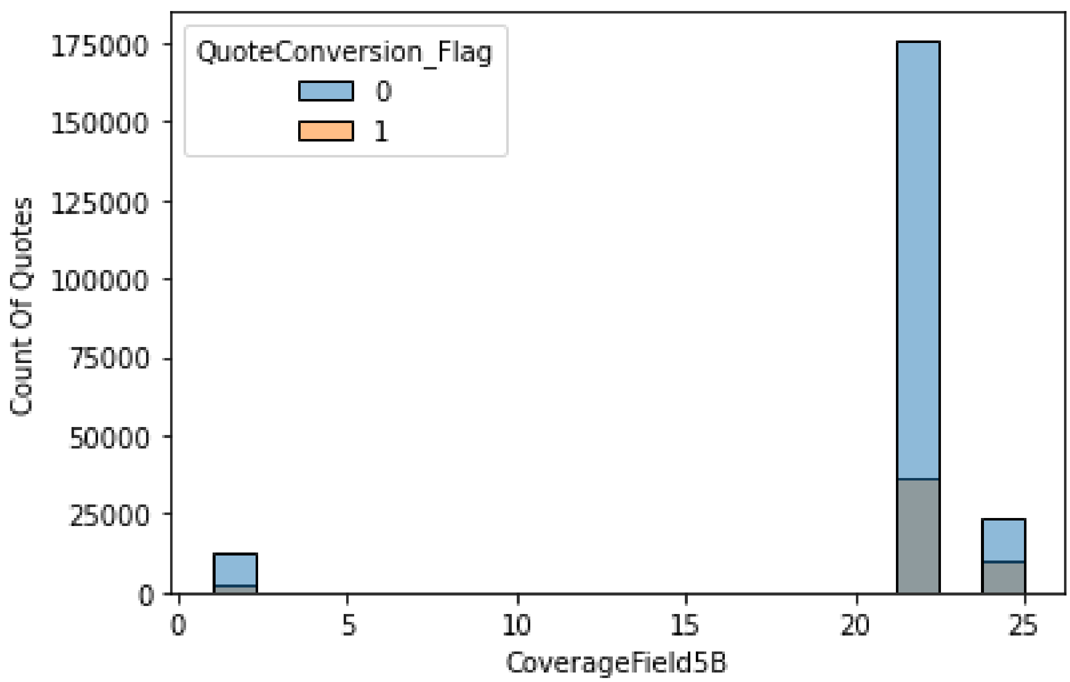

QuoteConversion_Flag
0
1
Count Of Quotes
CoverageField3B
0
2000
4000
6000
8000
0
5
10
15
20
25


QuoteConversion_Flag
0
1
Count Of Quotes
CoverageField4A
20000
17500
15000
12500
10000
7500
5000
2500
0
0
5
10
15
20
25

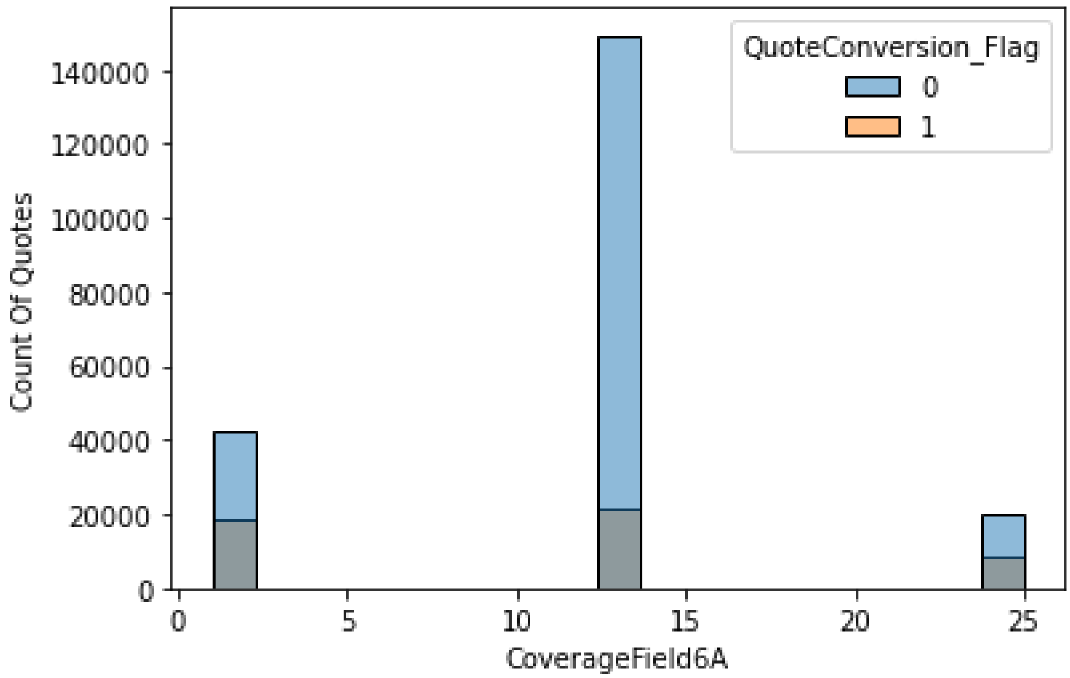

QuoteConversion_Flag
0
1
Count Of Quotes
CoverageField4B
8000
6000
4000
2000
0
0
5
10
15
20
25

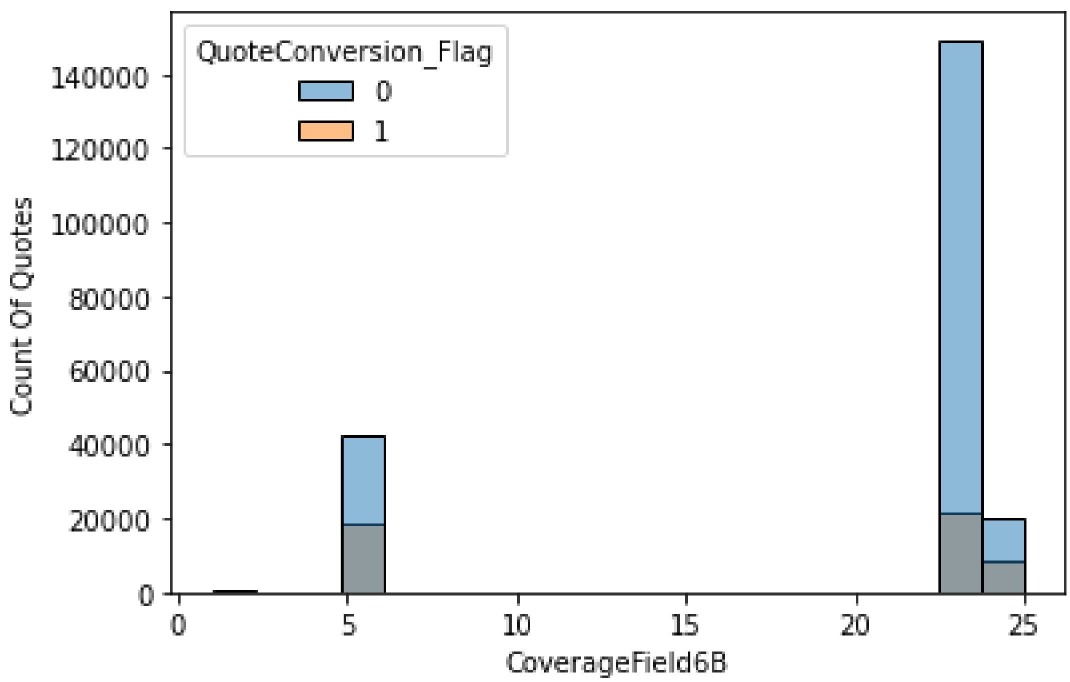


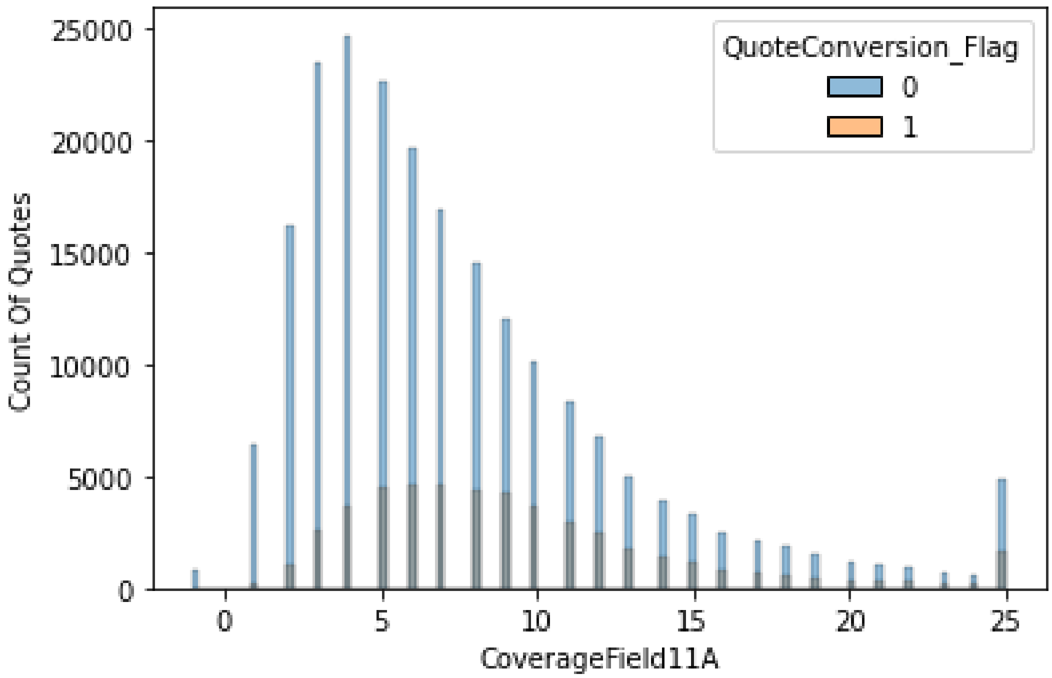
QuoteConversion_Flag
0
1
Count Of Quotes
CoverageField5A

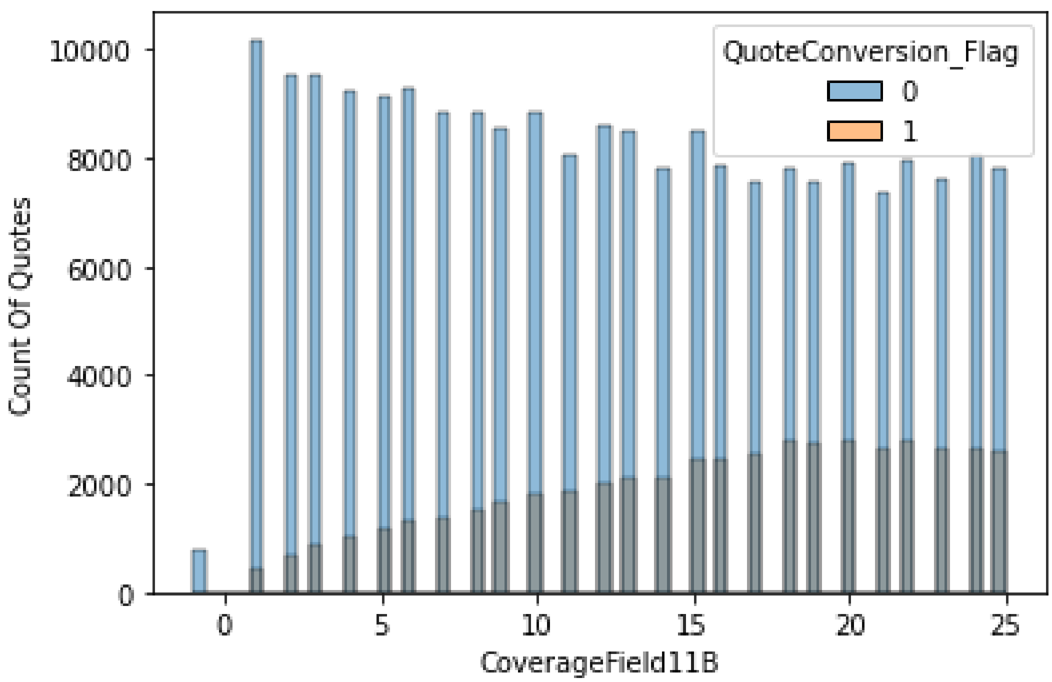


QuoteConversion_Flag
0
1
Count Of Quotes
CoverageField5B


QuoteConversion_Flag
0
1
Count Of Quotes
0
20000
40000
60000
80000
100000
120000
140000
0
5
10
15
20
25
CoverageField6A

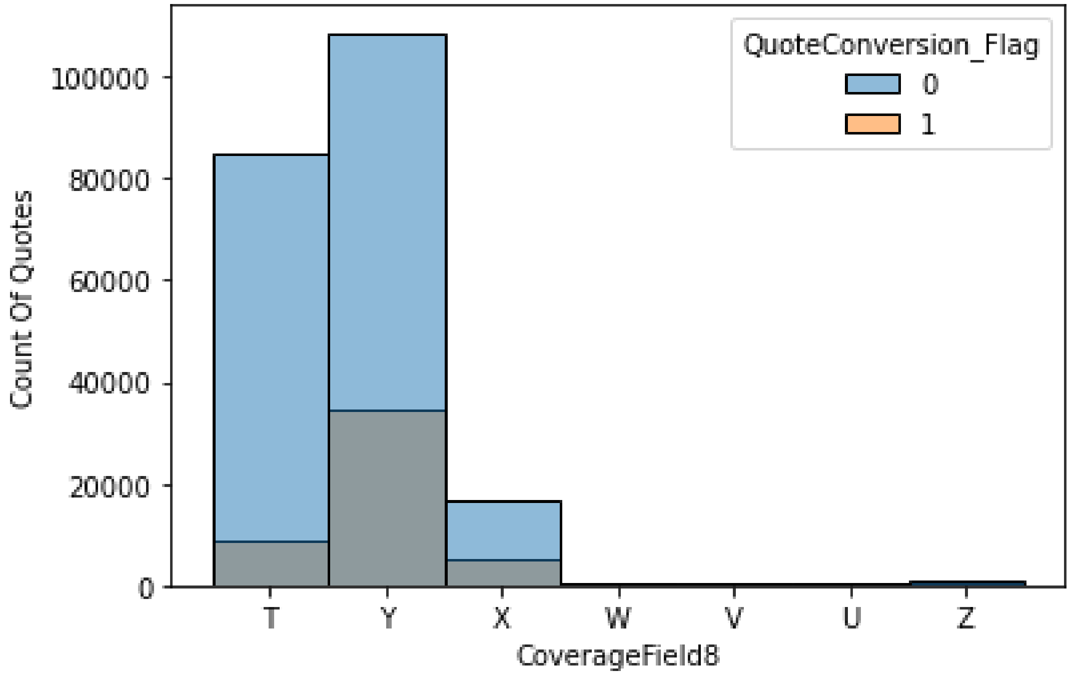

QuoteConversion_Flag
0
1
Count Of Quotes
0
20000
40000
60000
80000
100000
120000
140000
0
5
10
15
20
25
CoverageField6B

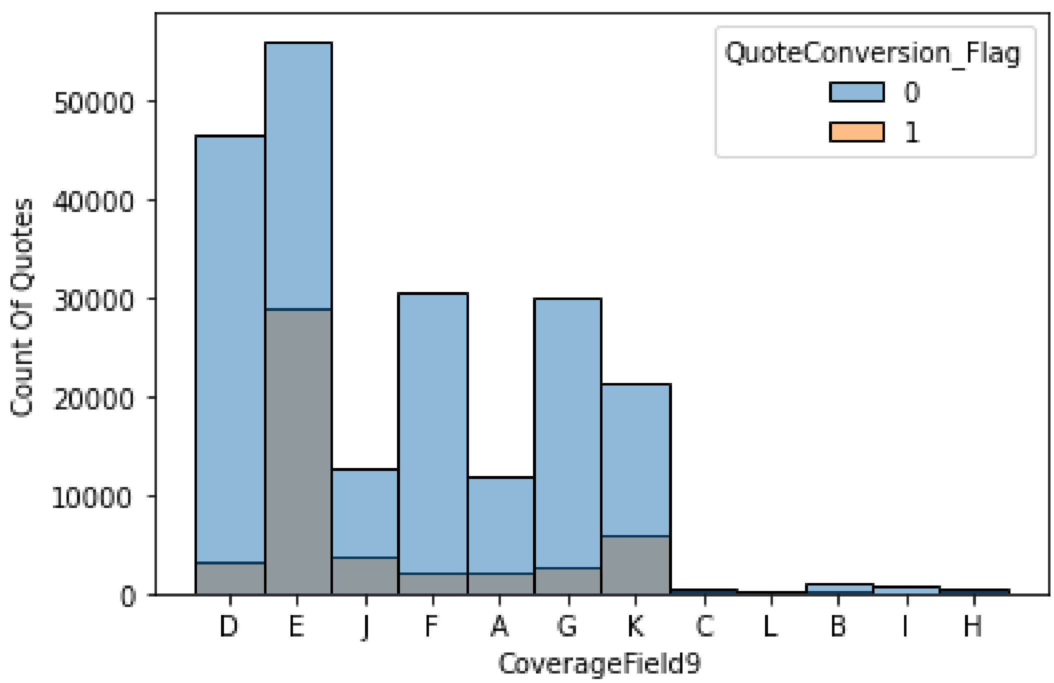




CATEGORICAL FEATURES

[28]:
```
for feature in get_categorical_features(dataset):
        sns.histplot(data = dataset, x = feature, hue = 'QuoteConversion_Flag')
        plt.xlabel(feature)
        plt.ylabel('Count Of Quotes')
        plt.show()
```





OBSERVATIONS:

1. DISCRETE NUMERICAL FEATURES

   A. CoverageField 1A,2A,3A,4A,11A seems to be having a log normal like distribution with the distribution being right skewed.

   B. Chance of a Quote being converted successfully reduces with increase in value of CoverageField 1B,2B,4B.

   C. Chance of a Quote being converted successfully increases with increase in value of CoverageField 11B.

2. CATEGORICAL NUMERICAL FEATURES

   A. Quote having a value of T, X and Y for CoverageField8 has more chance of being a successful conversion.For any other values the chances of it being successful conversion is extremely low.

   B. CoverageField9 seems to have a similar behaviour where the chances of successful conversion is high for only certain values and low for the rest.

SALES FEATURES

[29]:
```
dataset = extract_feature_dataset('SalesField',data)
```

[30]:
```
for feature in dataset.columns:
    if feature != 'QuoteConversion_Flag':
        print('{} has {} unique values'.format(feature,len(dataset[feature].
 ↪unique())))
```

```
SalesField1A has 25 unique values
SalesField1B has 25 unique values
SalesField2A has 26 unique values
SalesField2B has 26 unique values
SalesField3 has 2 unique values
SalesField4 has 5 unique values
SalesField5 has 5 unique values
SalesField6 has 24 unique values
SalesField7 has 7 unique values
SalesField8 has 61530 unique values
SalesField9 has 2 unique values
SalesField10 has 19 unique values
SalesField11 has 20 unique values
SalesField12 has 22 unique values
SalesField13 has 8 unique values
SalesField14 has 12 unique values
SalesField15 has 12 unique values
```

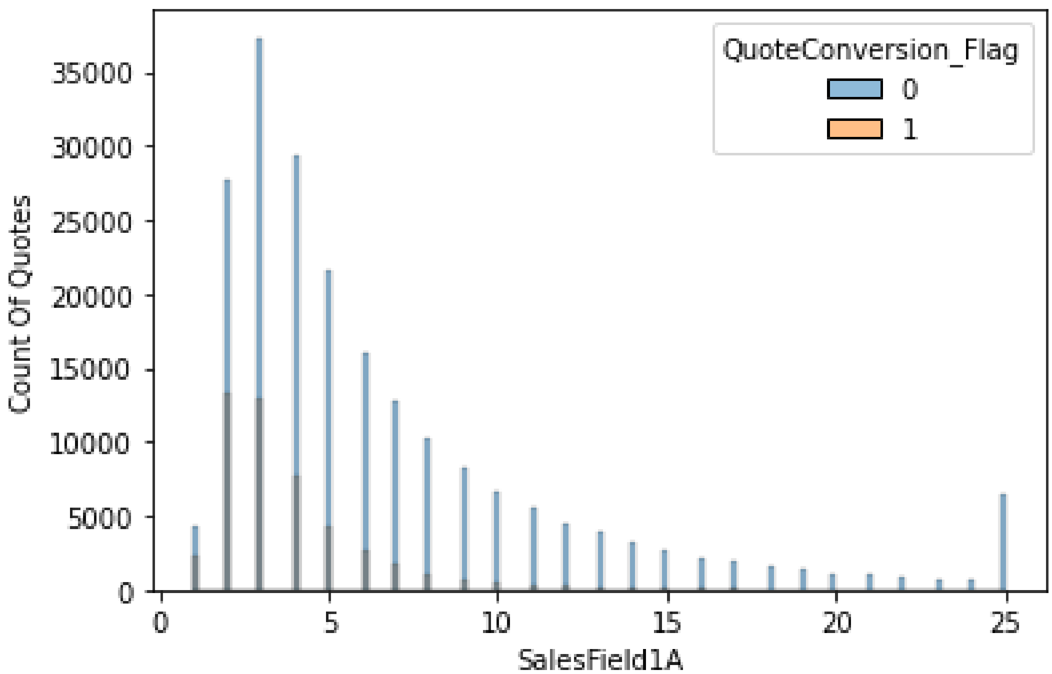
[31]:
```
numerical_features = get_numerical_features(dataset)
categorical_features = get_categorical_features(dataset)
```

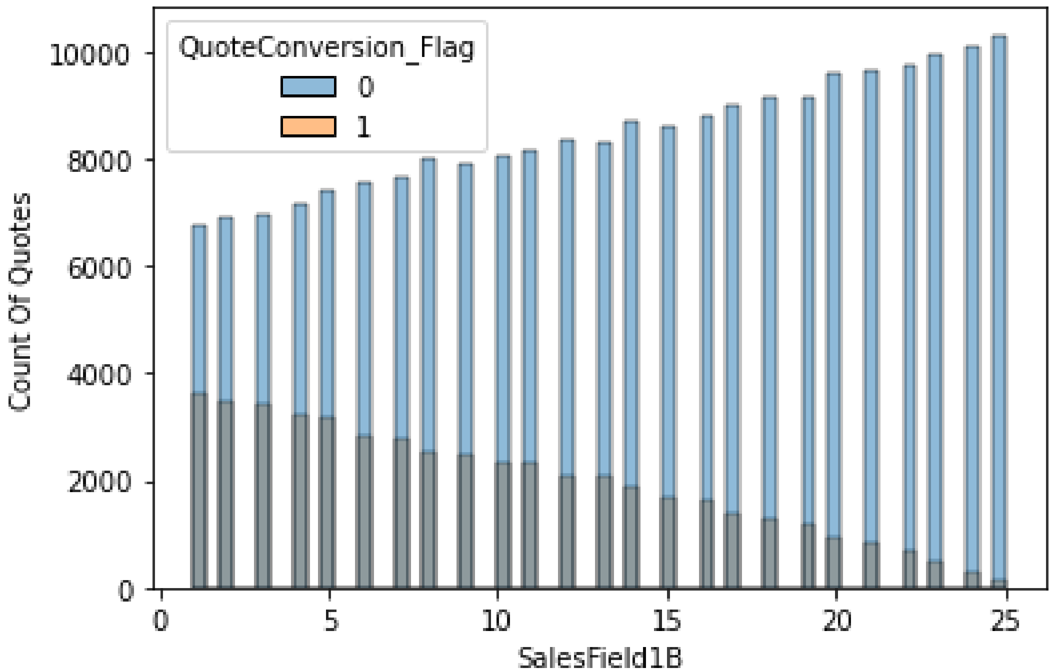
[32]:
```
print('Numerical Features : ', numerical_features)
print('Categorical Features : ', categorical_features)
```

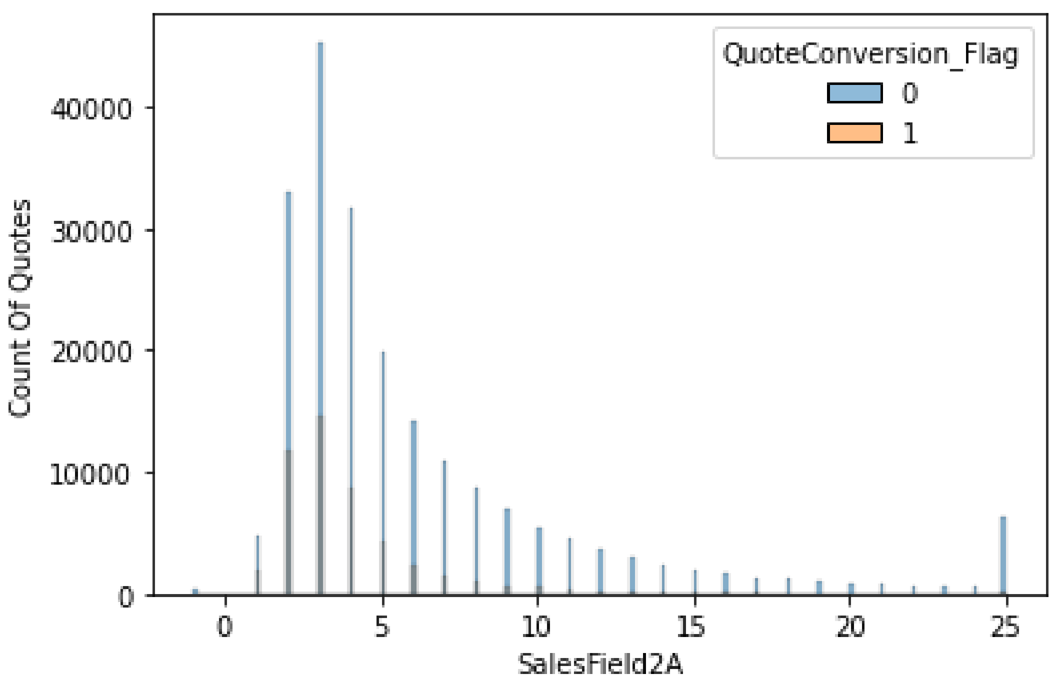
```
Numerical Features :  ['SalesField1A', 'SalesField1B', 'SalesField2A',
'SalesField2B', 'SalesField3', 'SalesField4', 'SalesField5', 'SalesField6',
'SalesField8', 'SalesField9', 'SalesField10', 'SalesField11', 'SalesField12',
'SalesField13', 'SalesField14', 'SalesField15', 'QuoteConversion_Flag']
Categorical Features :  ['SalesField7']
```

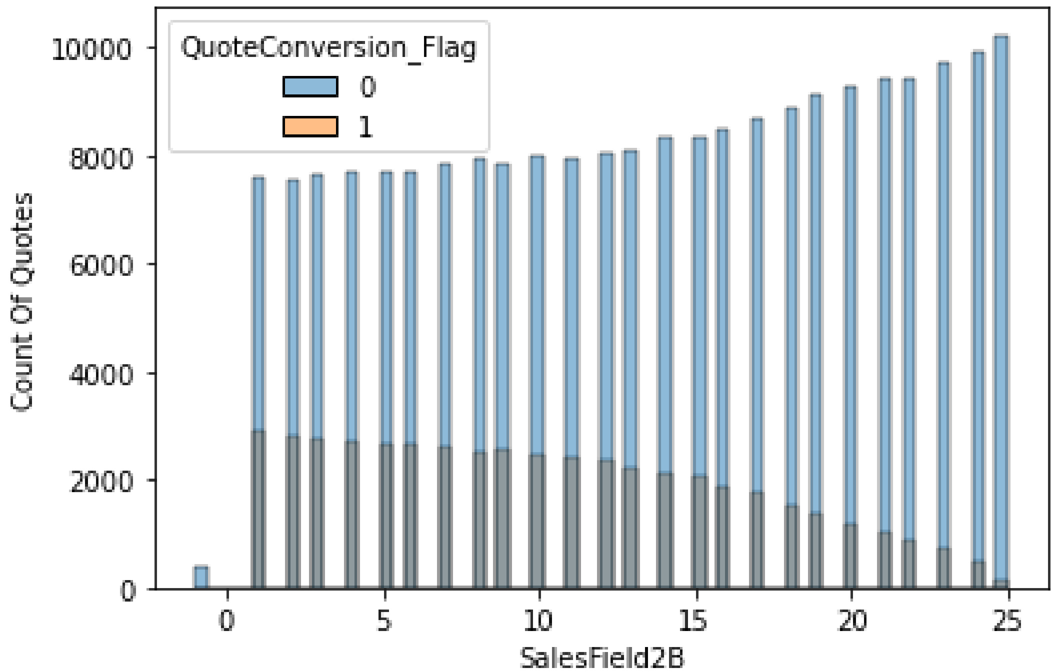
DISCRETE NUMERICAL FEATURES

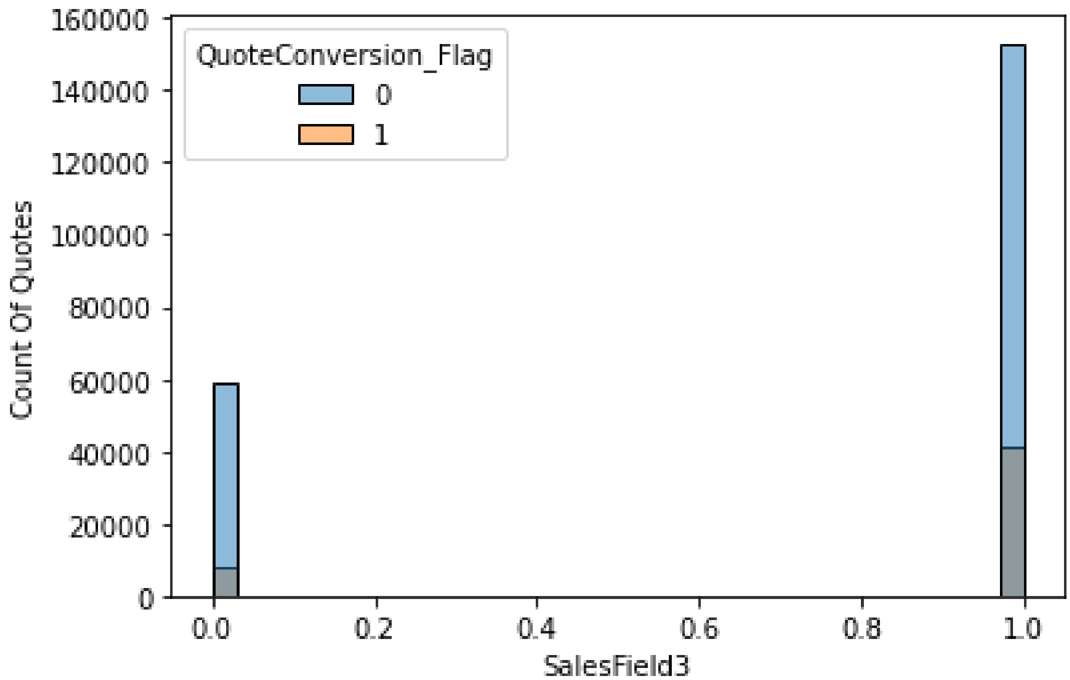
[33]:
```
for feature in get_numerical_features(dataset):
    if feature != 'QuoteConversion_Flag' and len(dataset[feature].unique()) <␣
↪30:
        sns.histplot(data = dataset, x = feature, hue = 'QuoteConversion_Flag')
        plt.xlabel(feature)
        plt.ylabel('Count Of Quotes')
        plt.show()
```

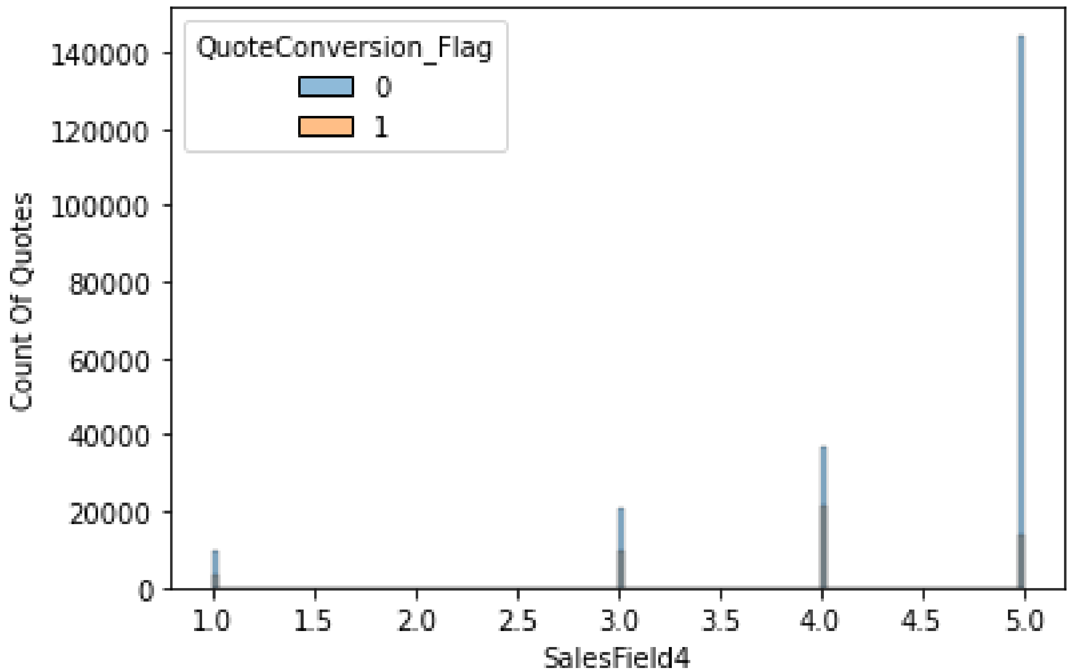


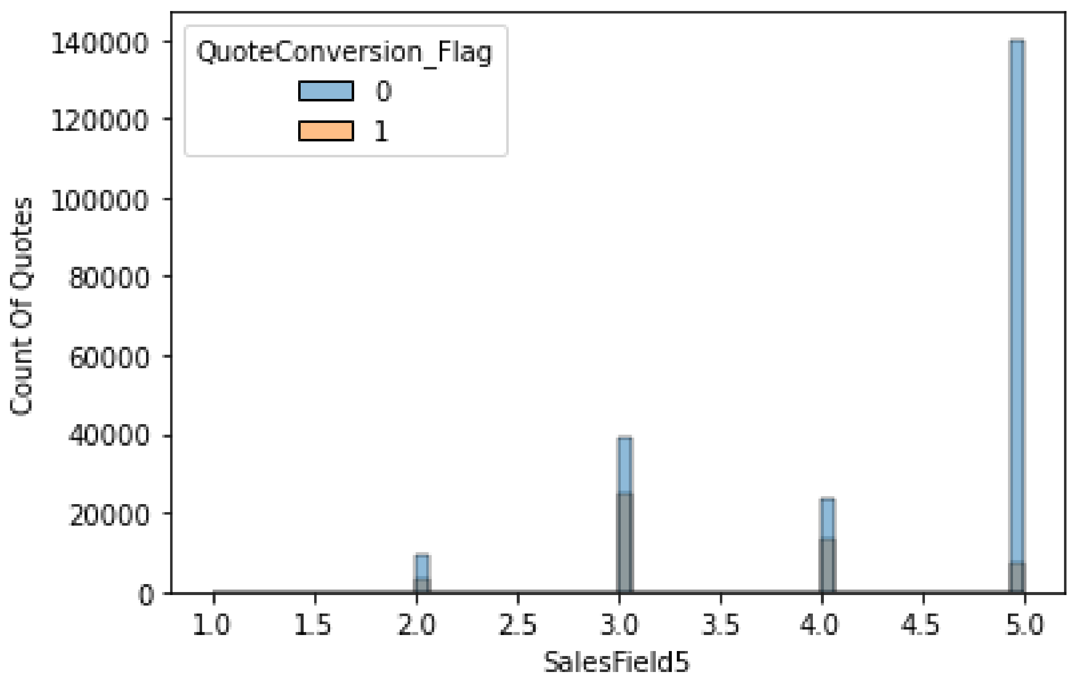

QuoteConversion_Flag
0
1
Count Of Quotes
0
2000
4000
6000
8000
10000
0
5
10
15
20
25
SalesField1B

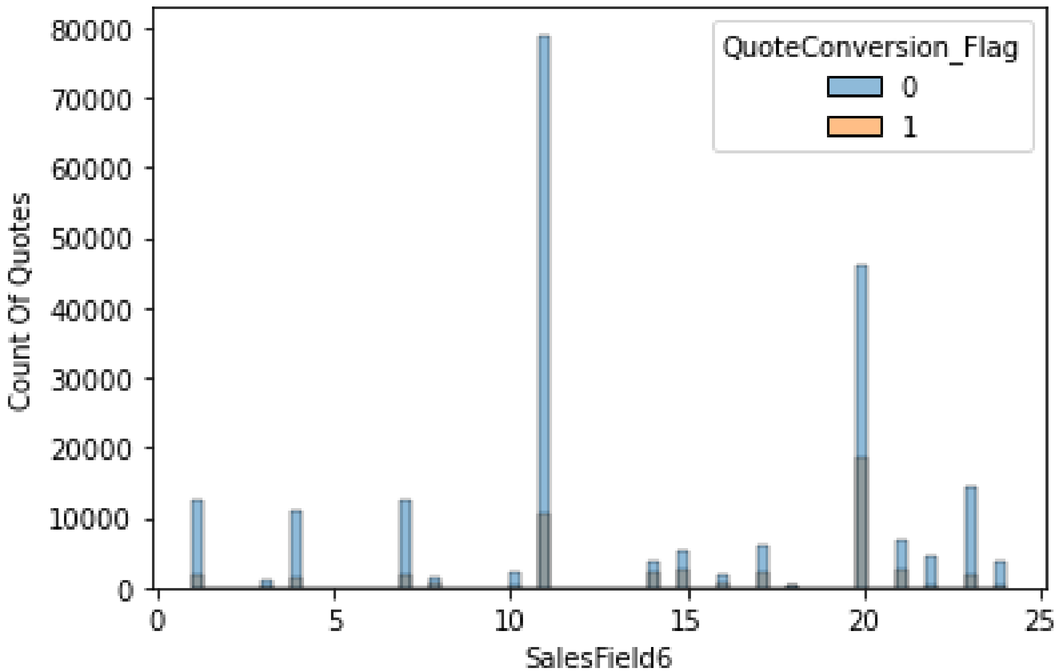

QuoteConversion_Flag
0
1
Count Of Quotes
0
10000
20000
30000
40000
0
5
10
15
20
25
SalesField2A

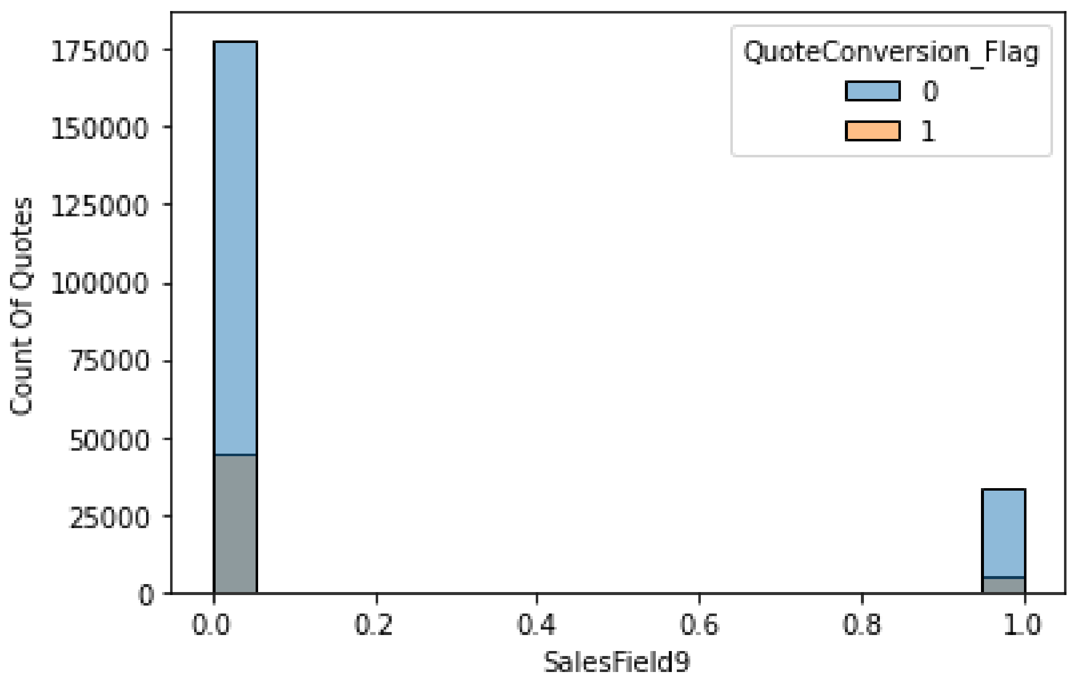

QuoteConversion_Flag
0
1
Count Of Quotes
10000
8000
6000
4000
2000
0
0
5
10
15
20
25
SalesField2B

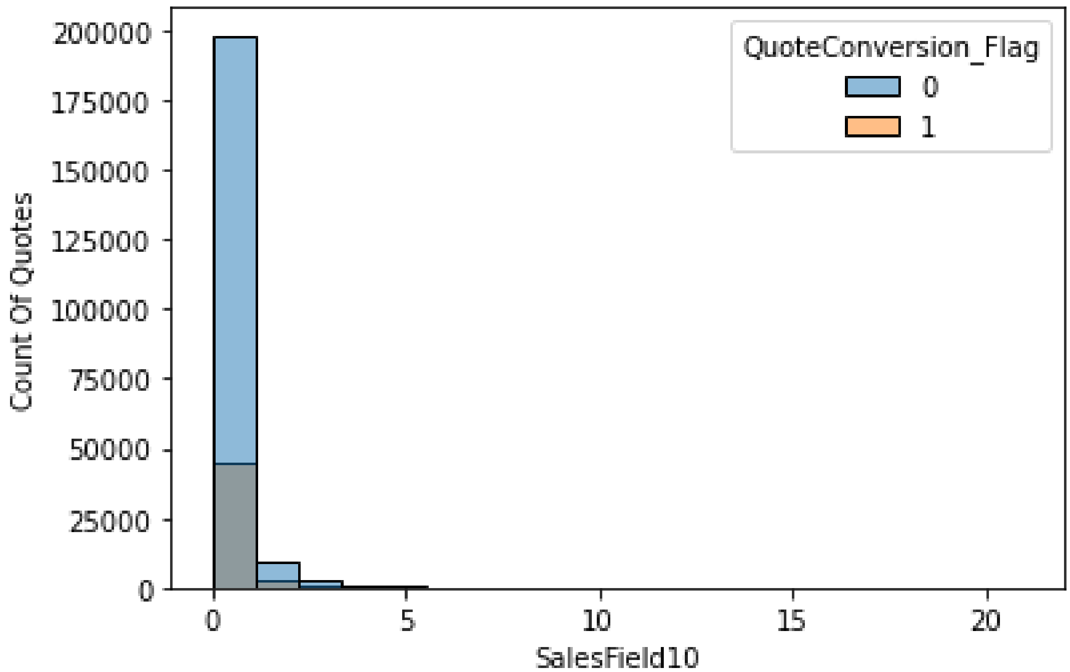

QuoteConversion_Flag
0
1
Count Of Quotes
160000
140000
120000
100000
80000
60000
40000
20000
0
0.0
0.2
0.4
0.6
0.8
1.0
SalesField3

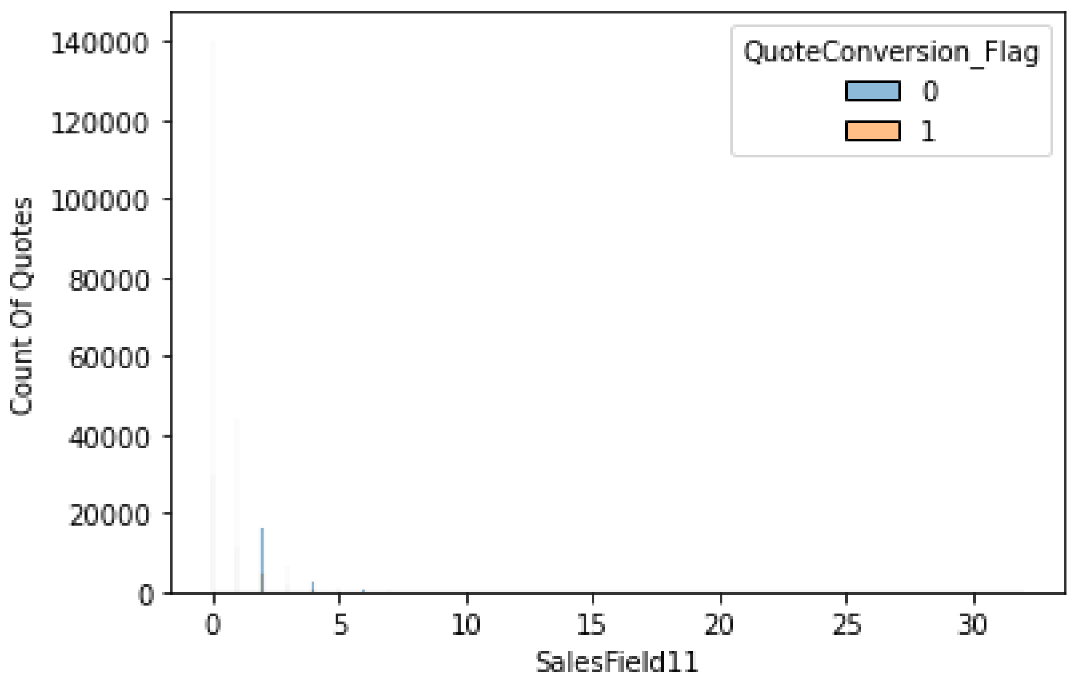

QuoteConversion_Flag
0
1
140000
120000
100000
80000
60000
40000
20000
0
Count Of Quotes
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
SalesField4

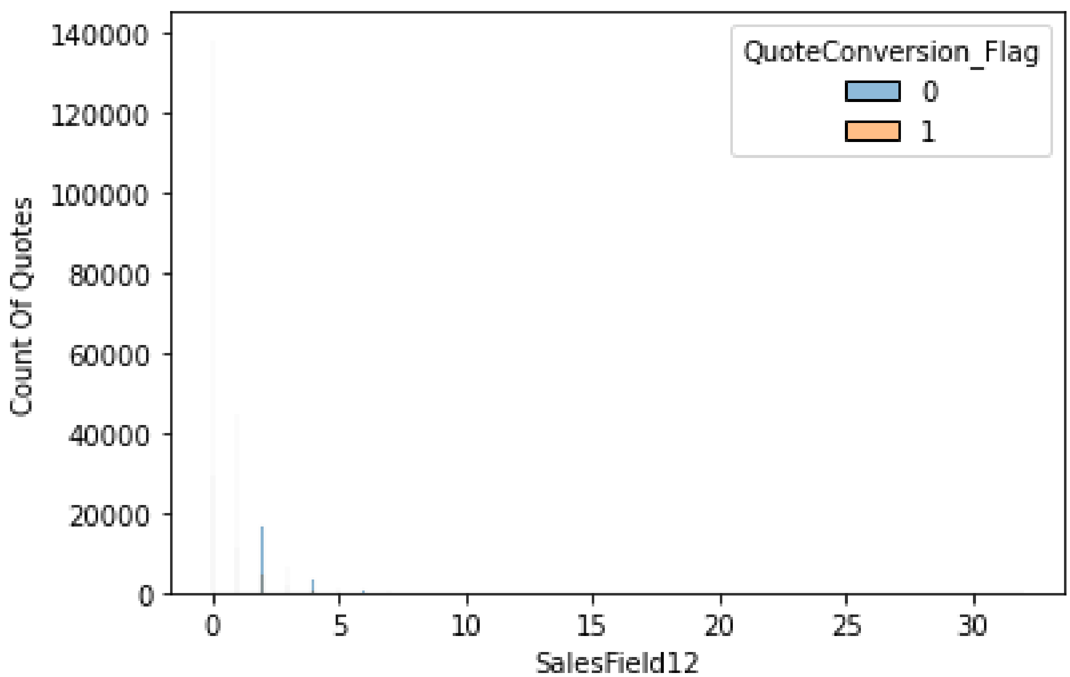

QuoteConversion_Flag
0
1
140000
120000
100000
80000
60000
40000
20000
0
Count Of Quotes
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
SalesField5

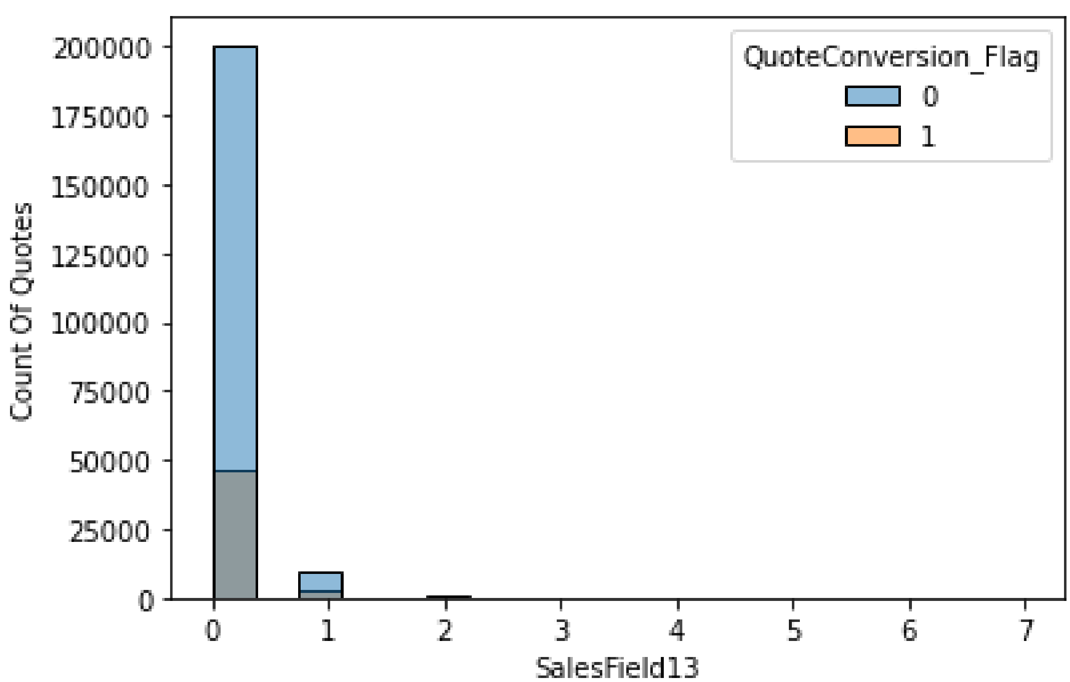

QuoteConversion_Flag
0
1
Count Of Quotes
80000
70000
60000
50000
40000
30000
20000
10000
0
0
5
10
15
20
25
SalesField6

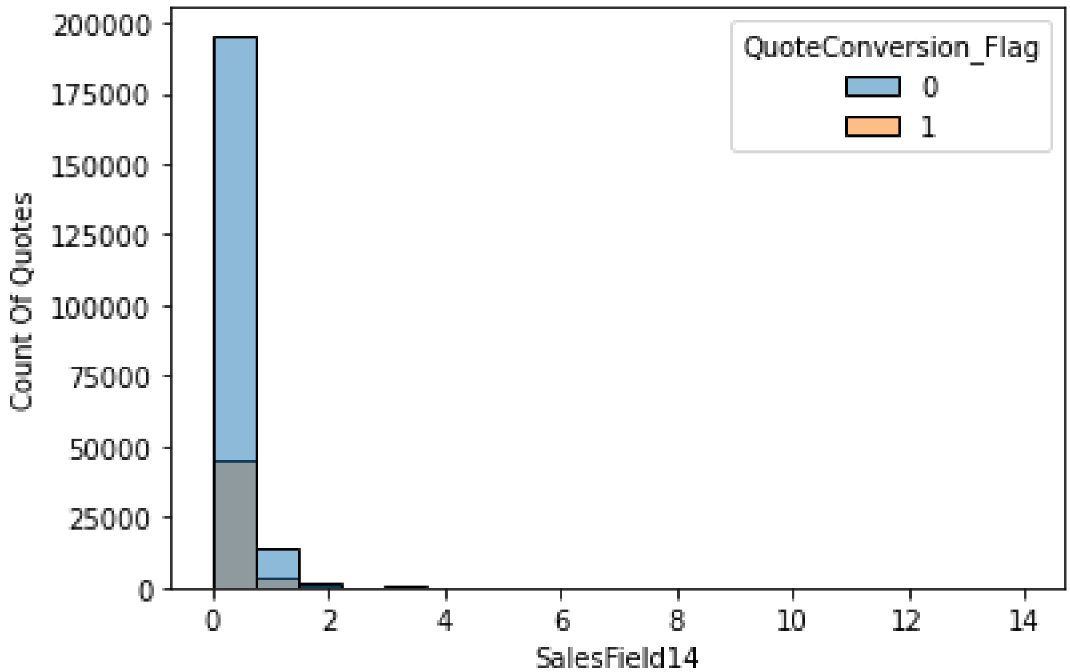

QuoteConversion_Flag
0
1
Count Of Quotes
175000
150000
125000
100000
75000
50000
25000
0
0.0
0.2
0.4
0.6
0.8
1.0
SalesField9

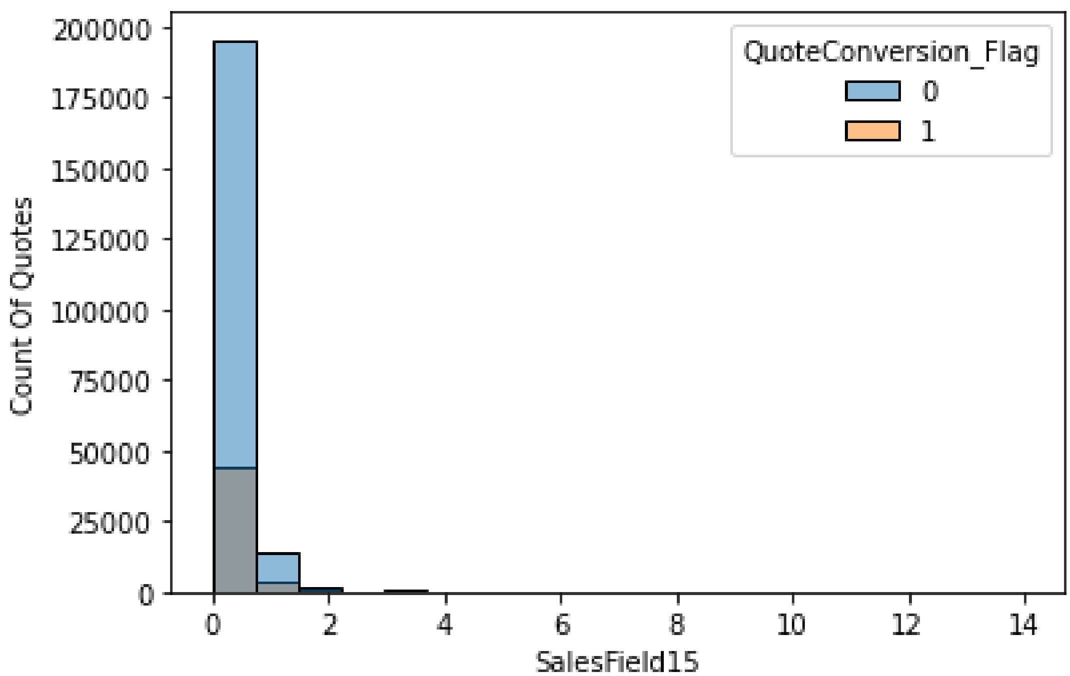

200000
175000
150000
125000
100000
75000
50000
25000
0
Count Of Quotes
0
5
10
15
20
SalesField10
QuoteConversion_Flag
0
1


140000
120000
100000
80000
60000
40000
20000
0
Count Of Quotes
0
5
10
15
20
25
30
SalesField11
QuoteConversion_Flag
0
1


QuoteConversion_Flag
0
1
Count Of Quotes
140000
120000
100000
80000
60000
40000
20000
0
0
5
10
15
20
25
30
SalesField12

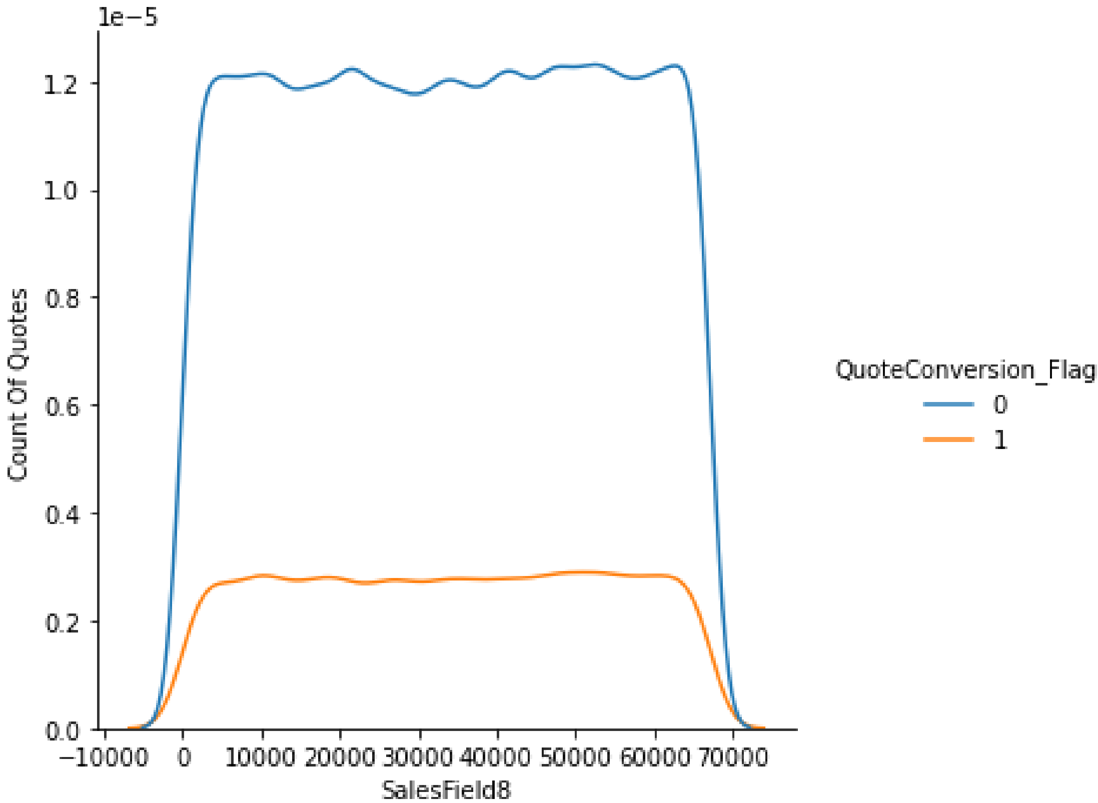

QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
0
1
2
3
4
5
6
7
SalesField13





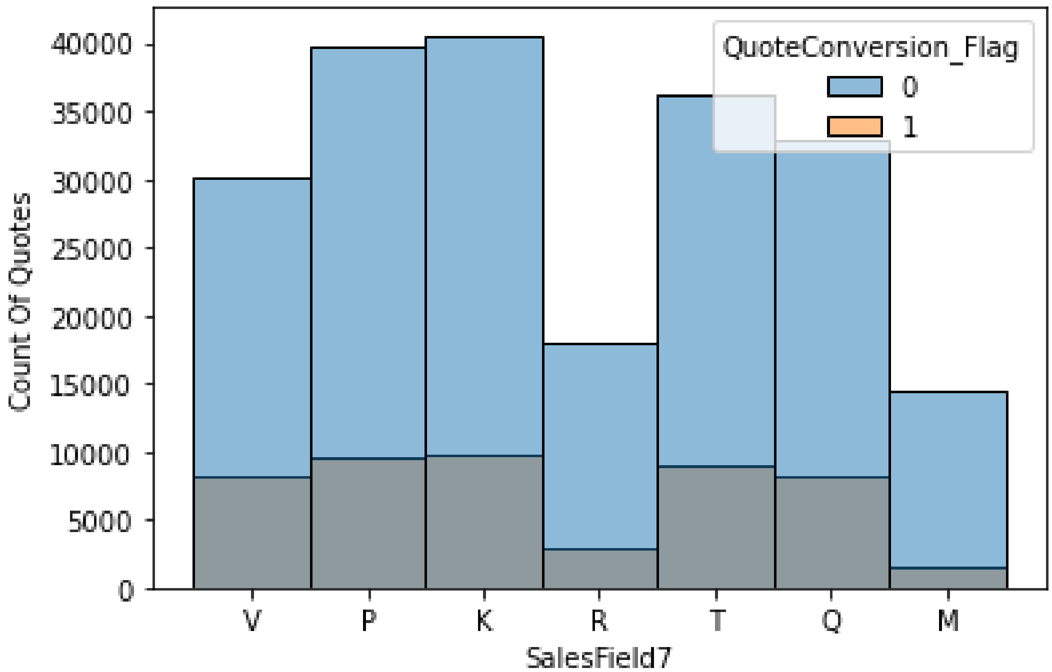
CONTINOUS NUMERICAL FEATURES

[34]:
```
sns.displot(data=dataset, x="SalesField8", hue="QuoteConversion_Flag", kind =␣
↪'kde')
plt.xlabel('SalesField8')
plt.ylabel('Count Of Quotes')
plt.show()
```



CATEGORICAL FEATURES

[35]:
```
for feature in get_categorical_features(dataset):
        sns.histplot(data = dataset, x = feature, hue = 'QuoteConversion_Flag')
        plt.xlabel(feature)
        plt.ylabel('Count Of Quotes')
        plt.show()
```



OBSERVATIONS:

1. DISCRETE NUMERICAL FEATURES

   A. SalesField 1A,2A seems to be having a log normal like distribution with the distribution being right skewed.

   B. SalesField 1B,2B seems to be having a inverse relationship b/w the magnitude and the chance of a quote being a sucessful conversion.

   C. SalesFeild3 having a value of 1 increases the chance of quote being successfully converted.

   D. Mid Range values for SalesField4 and SalesField5 increases the chance of quote being successfully converted.

   E. SalesField9 having a value of 0 increases the chance of quote being successfully converted.

   F. SalesField10-15 have a extremely skewed distribution with chances of conversion drastically dropping with increase in their respective values.

2. CONTINOUS NUMERICAL FEATURES

   A. Count of quotes being successfully converted is uniformly distributed accross all the values of SalesField8.

PERSONAL FEATURES

[36]:
```
dataset = extract_feature_dataset('PersonalField',data)
```

[37]:
```
for feature in dataset.columns:
    if feature != 'QuoteConversion_Flag':
        print('{} has {} unique values'.format(feature,len(dataset[feature].
↪unique())))
```

```
PersonalField1 has 2 unique values
PersonalField2 has 2 unique values
PersonalField4A has 26 unique values
PersonalField4B has 26 unique values
PersonalField5 has 9 unique values
PersonalField6 has 2 unique values
PersonalField7 has 3 unique values
PersonalField8 has 3 unique values
PersonalField9 has 3 unique values
PersonalField10A has 26 unique values
PersonalField10B has 26 unique values
PersonalField11 has 5 unique values
PersonalField12 has 5 unique values
PersonalField13 has 4 unique values
PersonalField14 has 30 unique values
PersonalField15 has 22 unique values
PersonalField16 has 50 unique values
PersonalField17 has 66 unique values
PersonalField18 has 61 unique values
PersonalField19 has 57 unique values
PersonalField22 has 7 unique values
PersonalField23 has 13 unique values
PersonalField24 has 14 unique values
PersonalField25 has 14 unique values
PersonalField26 has 14 unique values
PersonalField27 has 17 unique values
PersonalField28 has 7 unique values
PersonalField29 has 7 unique values
PersonalField30 has 12 unique values
PersonalField31 has 12 unique values
PersonalField32 has 12 unique values
PersonalField33 has 12 unique values
PersonalField34 has 7 unique values
PersonalField35 has 6 unique values
PersonalField36 has 6 unique values
PersonalField37 has 6 unique values
PersonalField38 has 6 unique values
PersonalField39 has 9 unique values
PersonalField40 has 10 unique values
PersonalField41 has 10 unique values
PersonalField42 has 10 unique values
PersonalField43 has 7 unique values
```

```
PersonalField44 has 12 unique values
PersonalField45 has 12 unique values
PersonalField46 has 12 unique values
PersonalField47 has 12 unique values
PersonalField48 has 7 unique values
PersonalField49 has 7 unique values
PersonalField50 has 7 unique values
PersonalField51 has 7 unique values
PersonalField52 has 7 unique values
PersonalField53 has 7 unique values
PersonalField54 has 11 unique values
PersonalField55 has 11 unique values
PersonalField56 has 12 unique values
PersonalField57 has 13 unique values
PersonalField58 has 7 unique values
PersonalField59 has 7 unique values
PersonalField60 has 7 unique values
PersonalField61 has 7 unique values
PersonalField62 has 7 unique values
PersonalField63 has 7 unique values
PersonalField64 has 3 unique values
PersonalField65 has 3 unique values
PersonalField66 has 3 unique values
PersonalField67 has 3 unique values
PersonalField68 has 4 unique values
PersonalField69 has 6 unique values
PersonalField70 has 7 unique values
PersonalField71 has 7 unique values
PersonalField72 has 8 unique values
PersonalField73 has 7 unique values
PersonalField74 has 8 unique values
PersonalField75 has 9 unique values
PersonalField76 has 10 unique values
PersonalField77 has 11 unique values
PersonalField78 has 7 unique values
PersonalField79 has 13 unique values
PersonalField80 has 14 unique values
PersonalField81 has 14 unique values
PersonalField82 has 14 unique values
PersonalField83 has 7 unique values
PersonalField84 has 8 unique values
PersonalField84_nan has 2 unique values
```

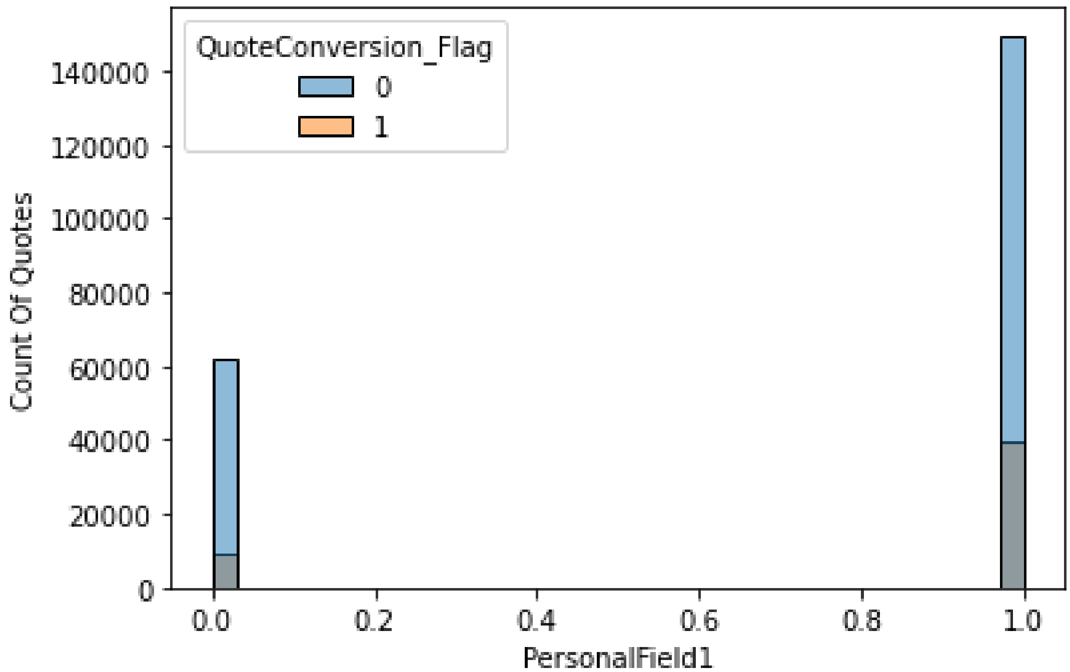
[38]:
```
numerical_features = get_numerical_features(dataset)
categorical_features = get_categorical_features(dataset)
```

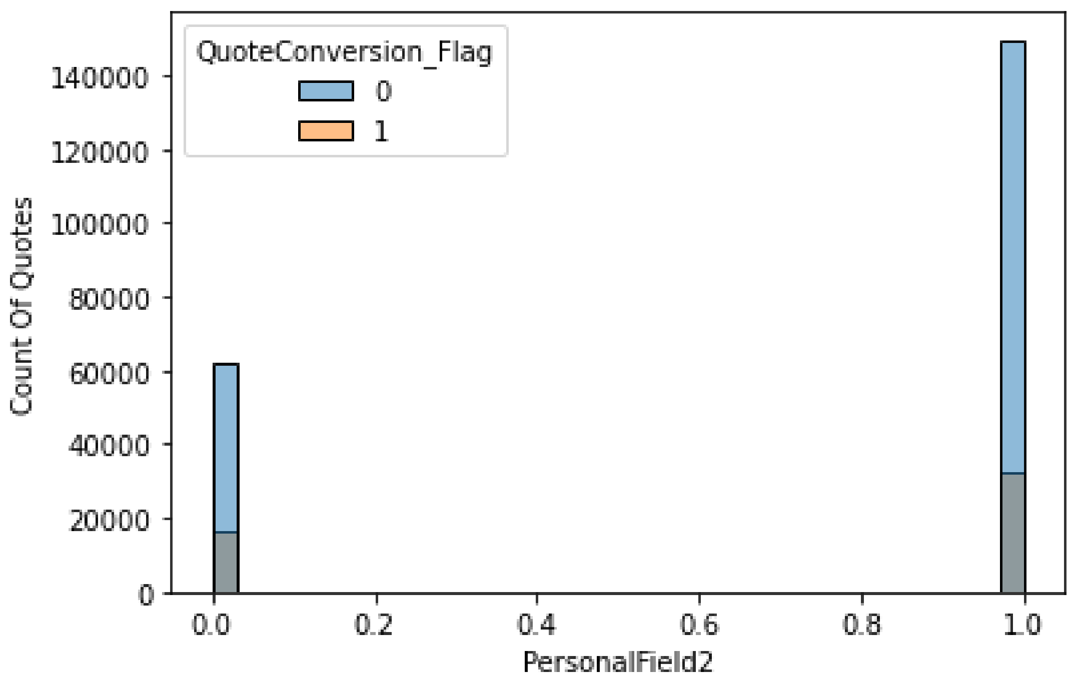
[39]:
```
print('Numerical Features : ', numerical_features)
print('Categorical Features : ', categorical_features)
```

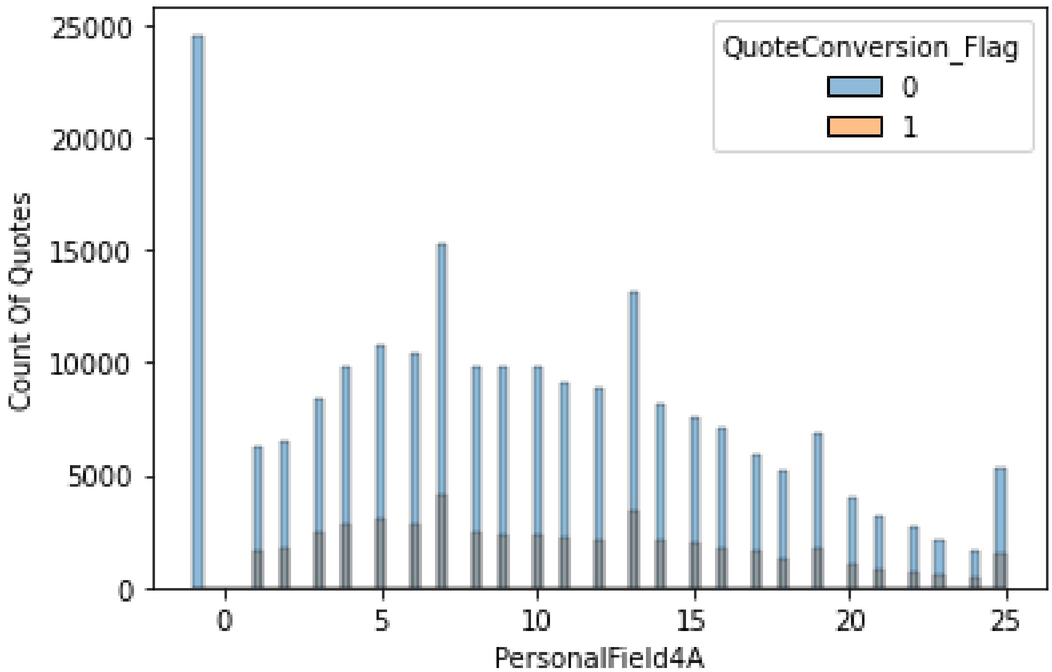
```
Numerical Features :  ['PersonalField1', 'PersonalField2', 'PersonalField4A',
'PersonalField4B', 'PersonalField5', 'PersonalField6', 'PersonalField8',
'PersonalField9', 'PersonalField10A', 'PersonalField10B', 'PersonalField11',
'PersonalField12', 'PersonalField13', 'PersonalField14', 'PersonalField15',
'PersonalField22', 'PersonalField23', 'PersonalField24', 'PersonalField25',
'PersonalField26', 'PersonalField27', 'PersonalField28', 'PersonalField29',
'PersonalField30', 'PersonalField31', 'PersonalField32', 'PersonalField33',
'PersonalField34', 'PersonalField35', 'PersonalField36', 'PersonalField37',
'PersonalField38', 'PersonalField39', 'PersonalField40', 'PersonalField41',
'PersonalField42', 'PersonalField43', 'PersonalField44', 'PersonalField45',
'PersonalField46', 'PersonalField47', 'PersonalField48', 'PersonalField49',
'PersonalField50', 'PersonalField51', 'PersonalField52', 'PersonalField53',
'PersonalField54', 'PersonalField55', 'PersonalField56', 'PersonalField57',
'PersonalField58', 'PersonalField59', 'PersonalField60', 'PersonalField61',
'PersonalField62', 'PersonalField63', 'PersonalField64', 'PersonalField65',
'PersonalField66', 'PersonalField67', 'PersonalField68', 'PersonalField69',
'PersonalField70', 'PersonalField71', 'PersonalField72', 'PersonalField73',
'PersonalField74', 'PersonalField75', 'PersonalField76', 'PersonalField77',
'PersonalField78', 'PersonalField79', 'PersonalField80', 'PersonalField81',
'PersonalField82', 'PersonalField83', 'PersonalField84', 'PersonalField84_nan',
'QuoteConversion_Flag']
Categorical Features :  ['PersonalField7', 'PersonalField16', 'PersonalField17',
'PersonalField18', 'PersonalField19']
```

DISCRETE NUMERICAL FEATURES

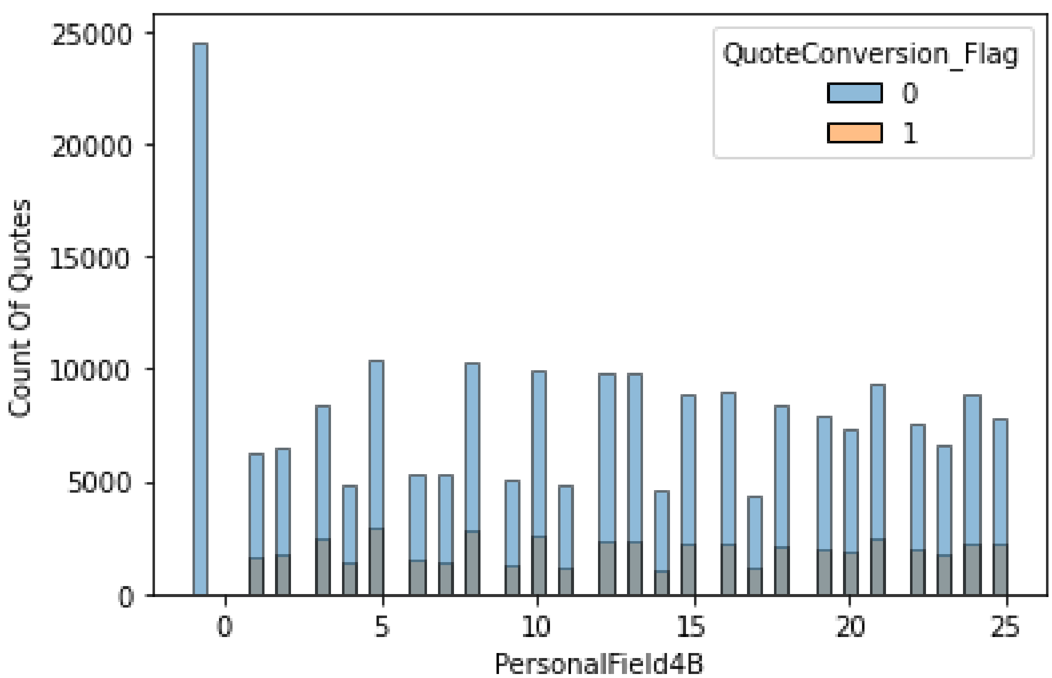
[40]:
```
for feature in get_numerical_features(dataset):
    if feature != 'QuoteConversion_Flag' and len(dataset[feature].unique()) <␣
↪30:
        sns.histplot(data = dataset, x = feature, hue = 'QuoteConversion_Flag')
        plt.xlabel(feature)
        plt.ylabel('Count Of Quotes')
        plt.show()
```


QuoteConversion_Flag
0
1
140000
120000
100000
80000
60000
40000
20000
0
Count Of Quotes
0.0
0.2
0.4
0.6
0.8
1.0
PersonalField1

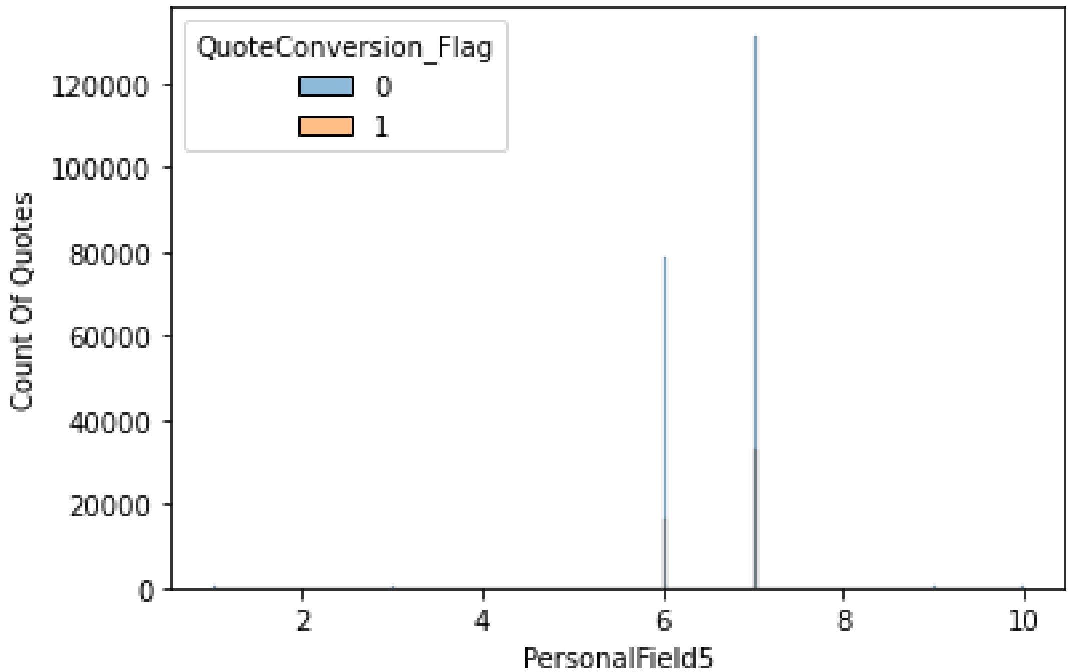

QuoteConversion_Flag
0
1
140000
120000
100000
80000
60000
40000
20000
0
Count Of Quotes
0.0
0.2
0.4
0.6
0.8
1.0
PersonalField2

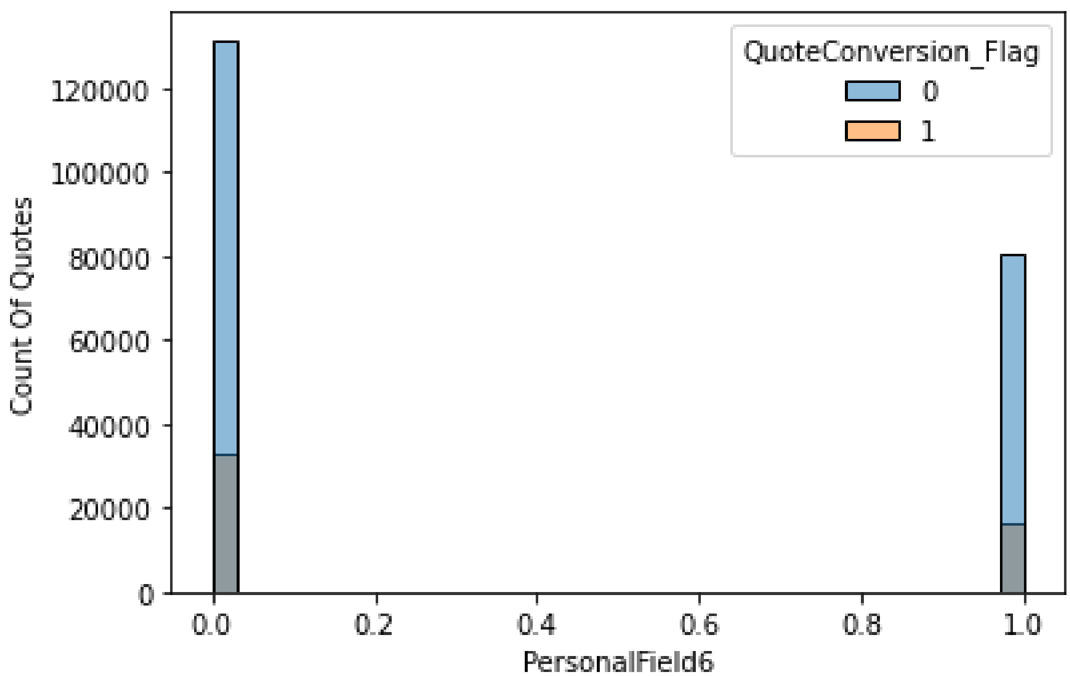


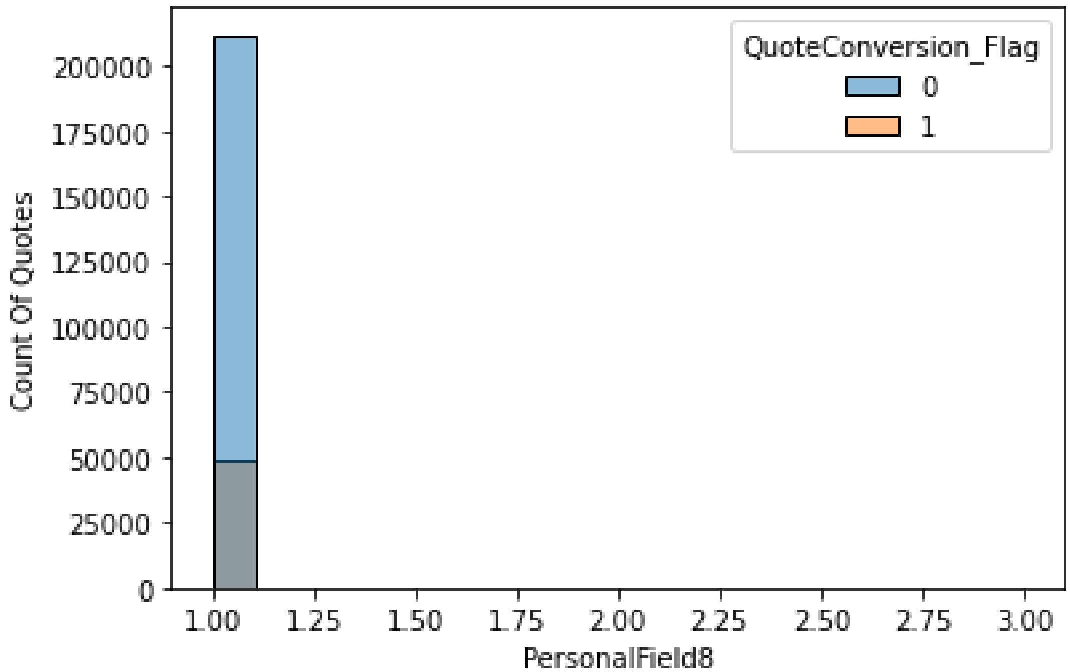
QuoteConversion_Flag
0
1
Count Of Quotes
PersonalField4A

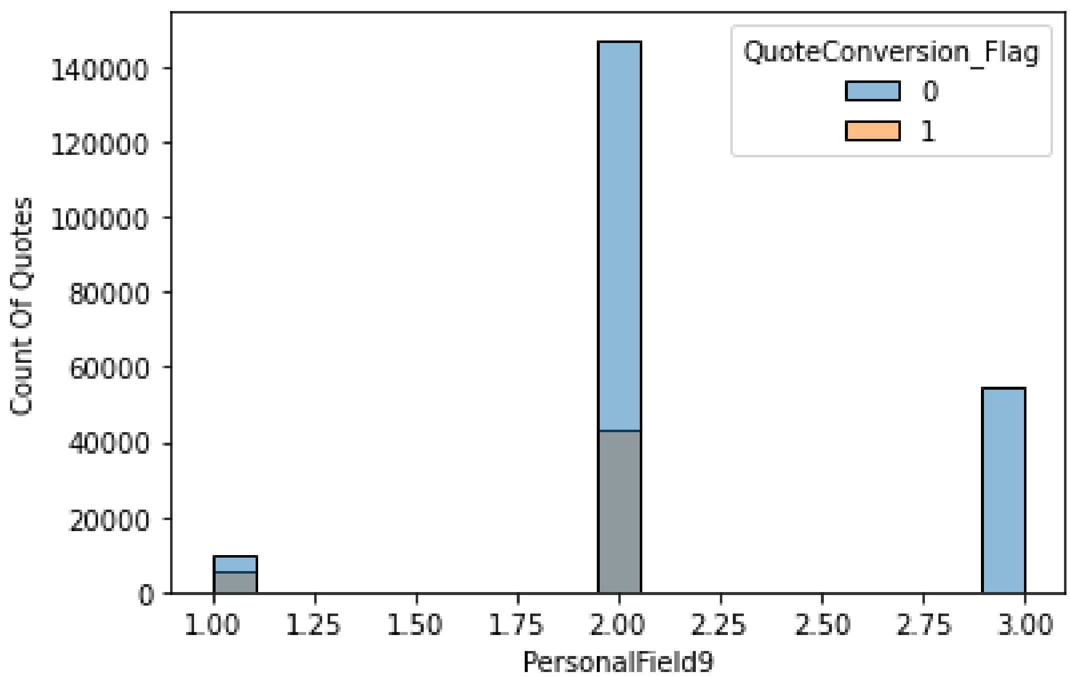


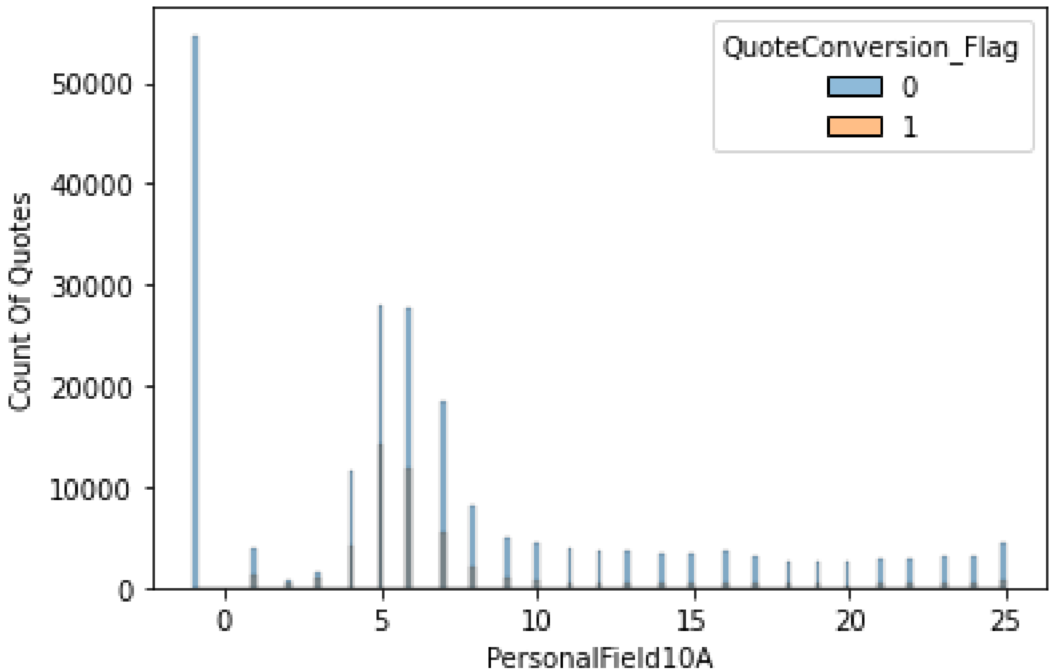
QuoteConversion_Flag
0
1
Count Of Quotes
PersonalField4B

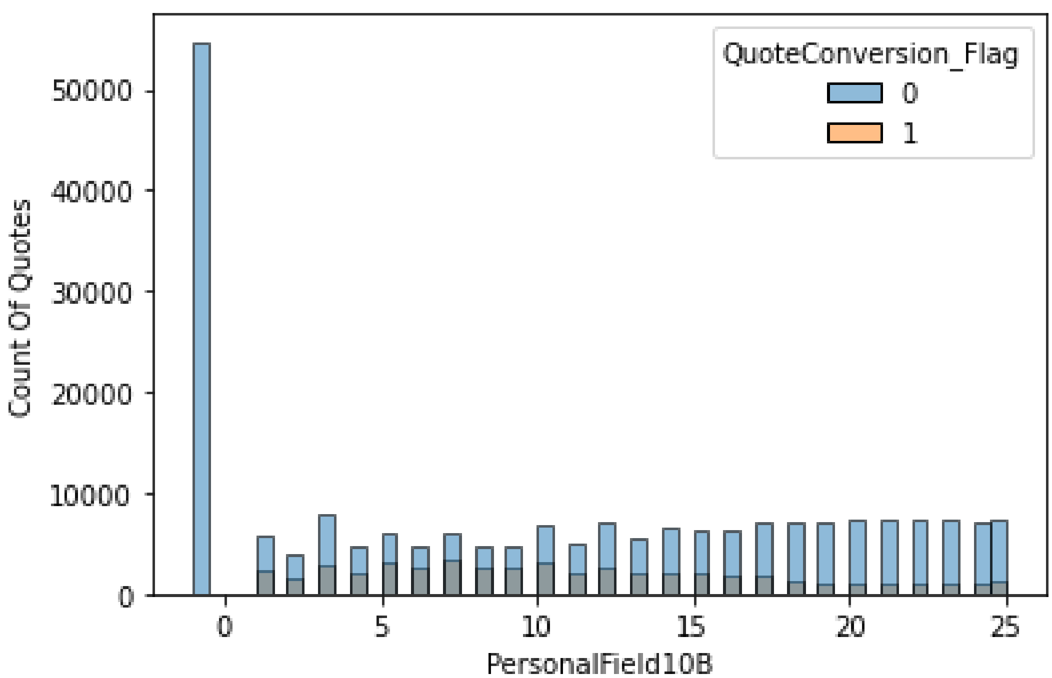

QuoteConversion_Flag
0
1
Count Of Quotes
120000
100000
80000
60000
40000
20000
0
2
4
6
8
10
PersonalField5

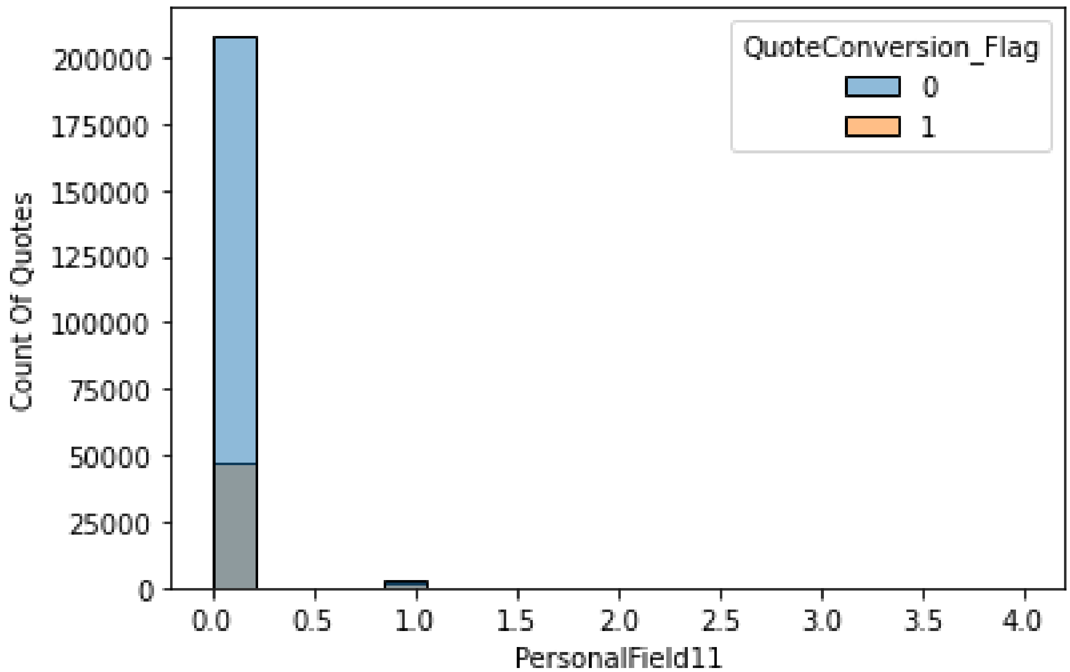

QuoteConversion_Flag
0
1
Count Of Quotes
120000
100000
80000
60000
40000
20000
0
0.0
0.2
0.4
0.6
0.8
1.0
PersonalField6

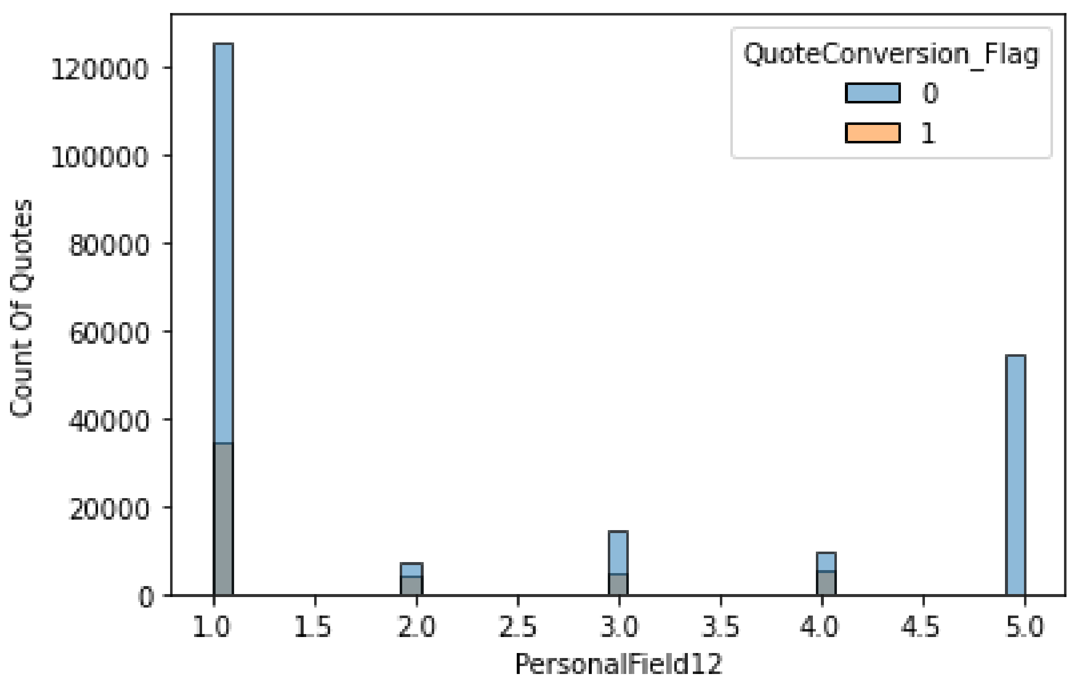

QuoteConversion_Flag
0
1
200000
175000
150000
125000
100000
75000
50000
25000
0
Count Of Quotes
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
PersonalField8

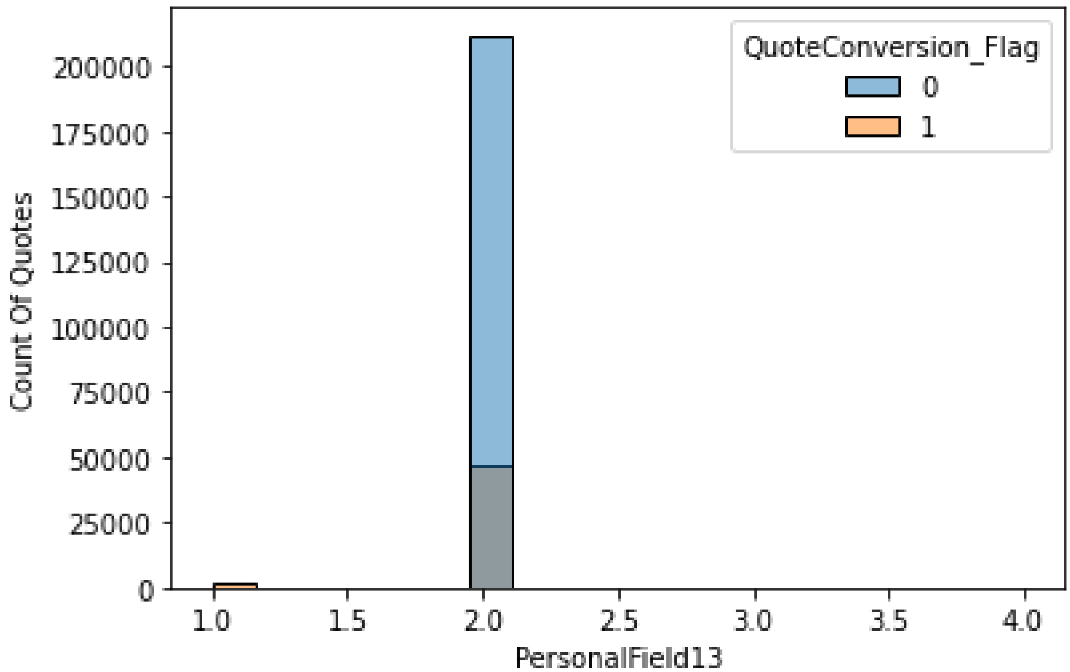

QuoteConversion_Flag
0
1
140000
120000
100000
80000
60000
40000
20000
0
Count Of Quotes
1.00
1.25
1.50
1.75
2.00
2.25
2.50
2.75
3.00
PersonalField9

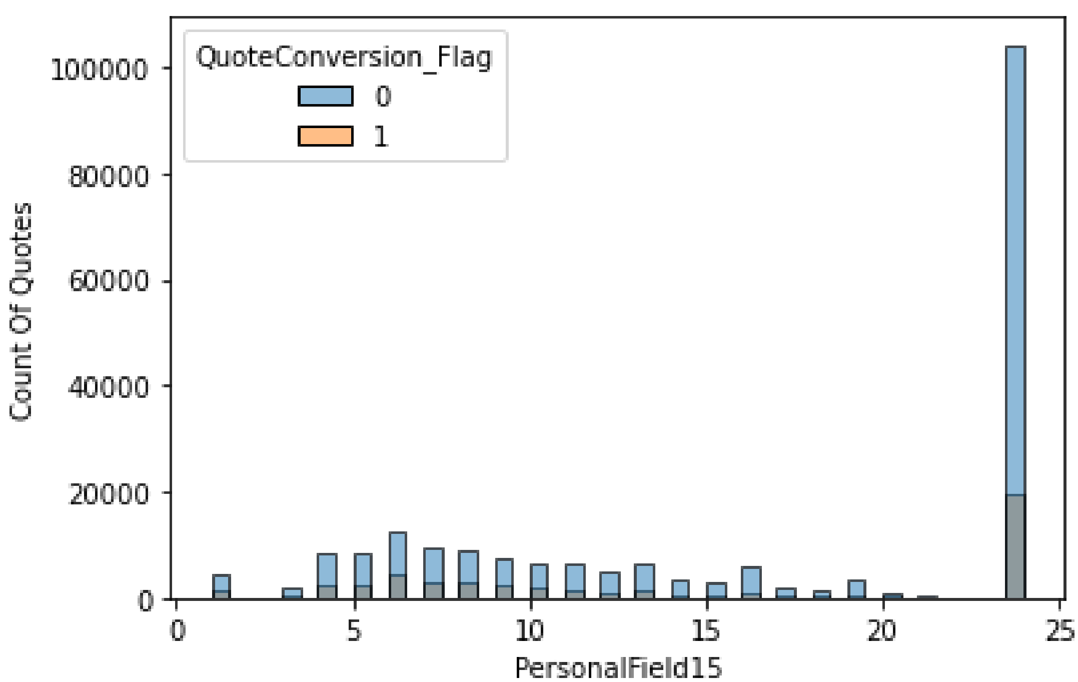

QuoteConversion_Flag
0
1
Count Of Quotes
PersonalField10A
0
10000
20000
30000
40000
50000
5
10
15
20
25

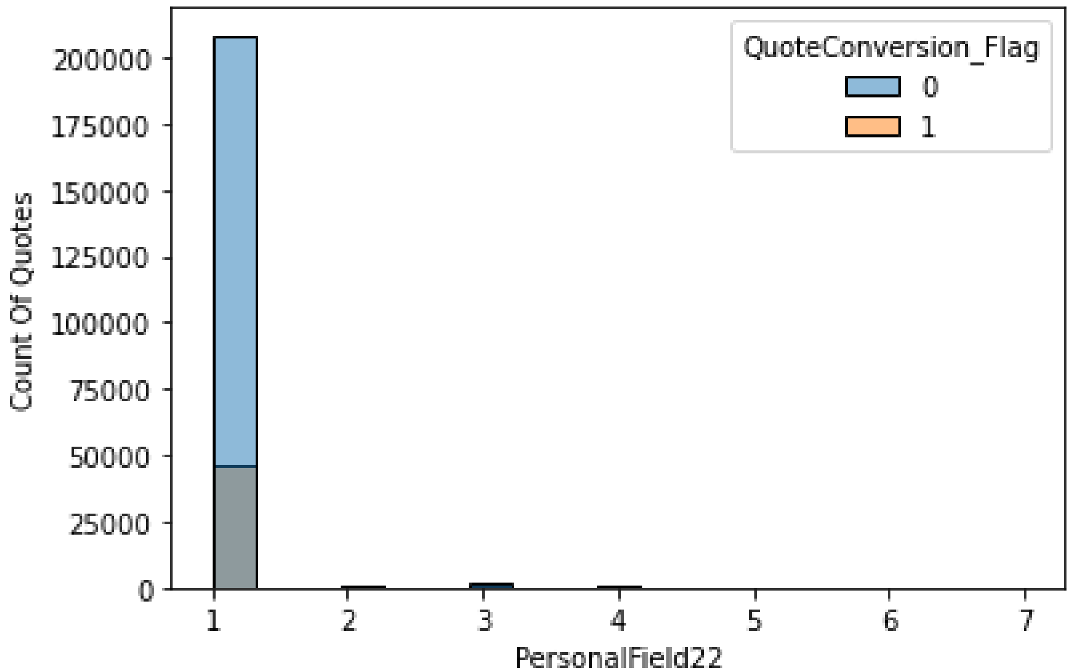

QuoteConversion_Flag
0
1
Count Of Quotes
PersonalField10B
0
10000
20000
30000
40000
50000
5
10
15
20
25

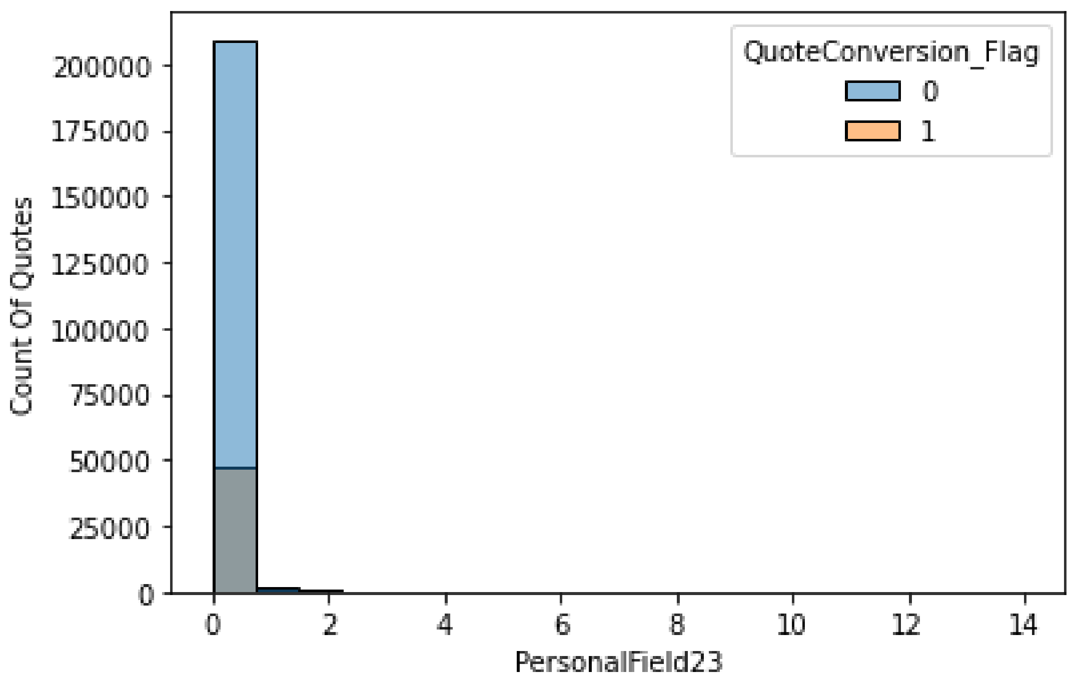

QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
PersonalField11

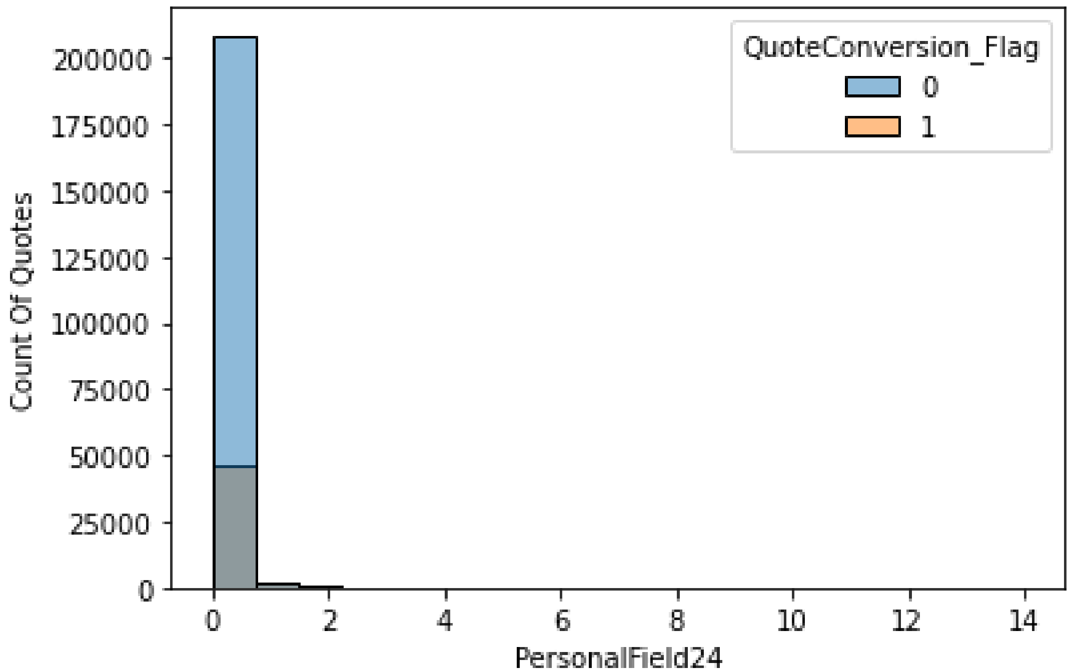

QuoteConversion_Flag
0
1
Count Of Quotes
120000
100000
80000
60000
40000
20000
0
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
PersonalField12

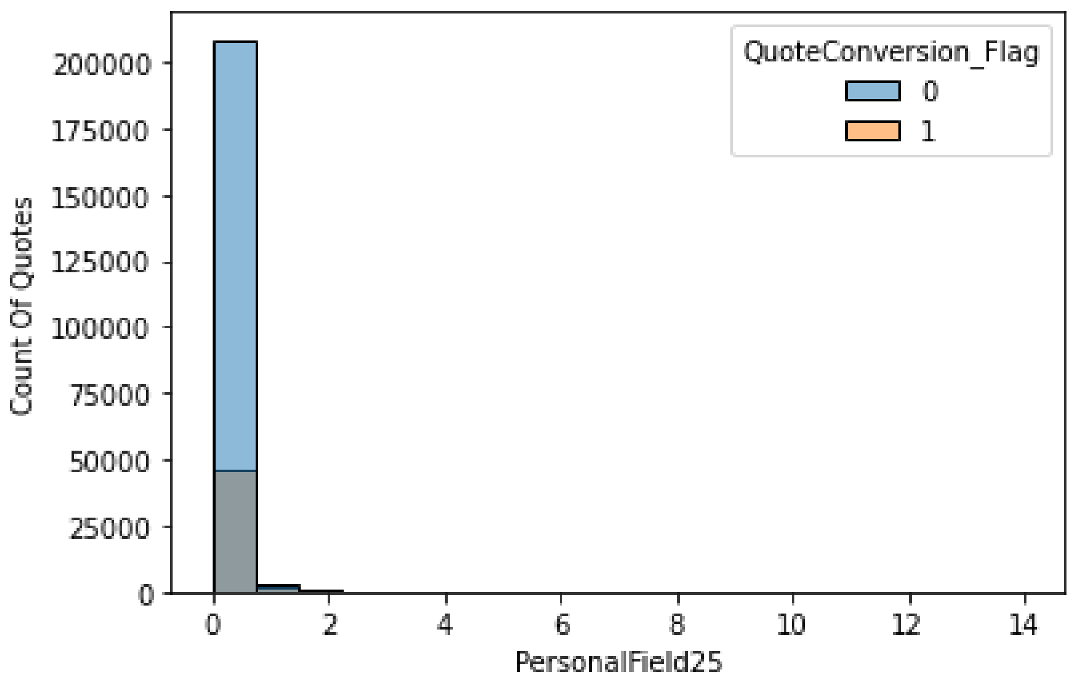

QuoteConversion_Flag
0
1
200000
175000
150000
125000
100000
75000
50000
25000
0
Count Of Quotes
1.0
1.5
2.0
2.5
3.0
3.5
4.0
PersonalField13

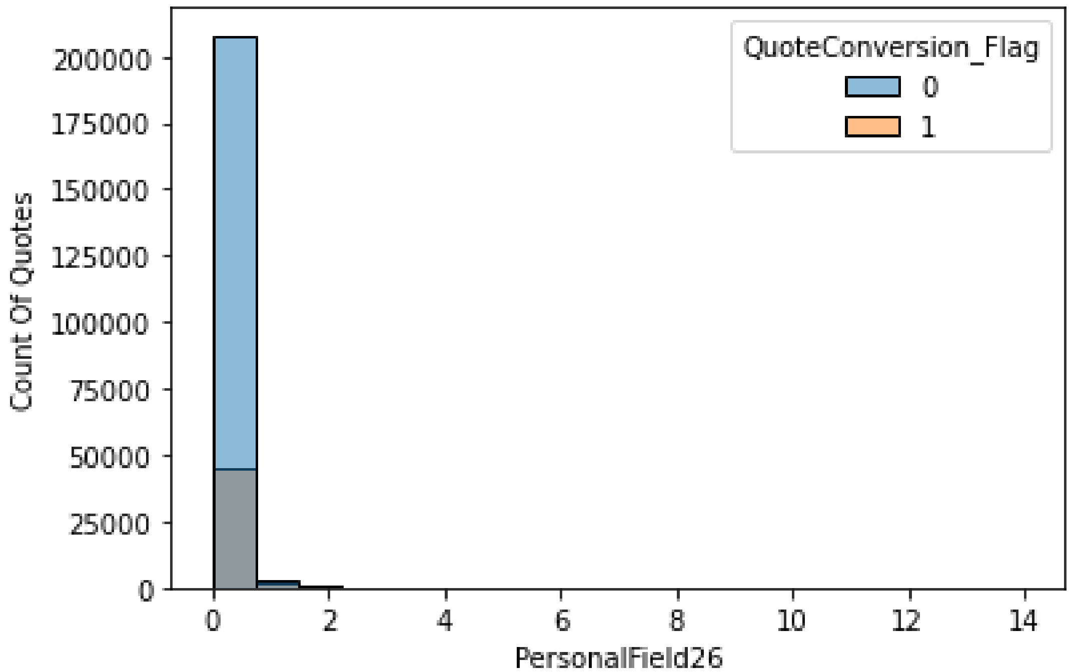

QuoteConversion_Flag
0
1
100000
80000
60000
40000
20000
0
Count Of Quotes
0
5
10
15
20
25
PersonalField15


QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
1
2
3
4
5
6
7
PersonalField22

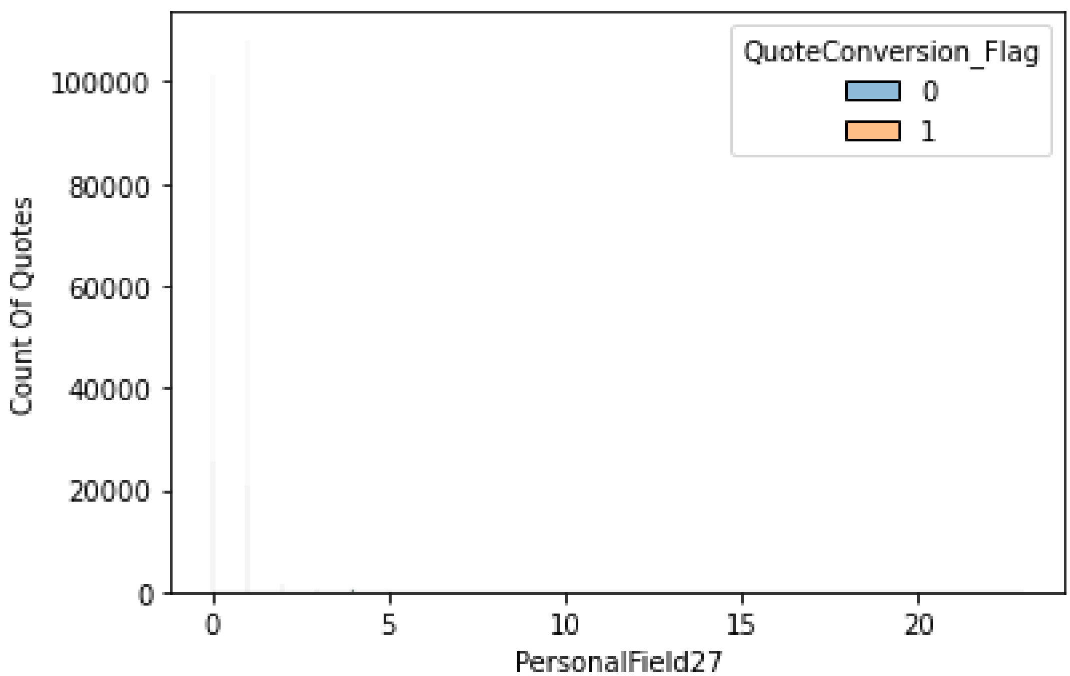

QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
0
2
4
6
8
10
12
14
PersonalField23


QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
0
2
4
6
8
10
12
14
PersonalField24

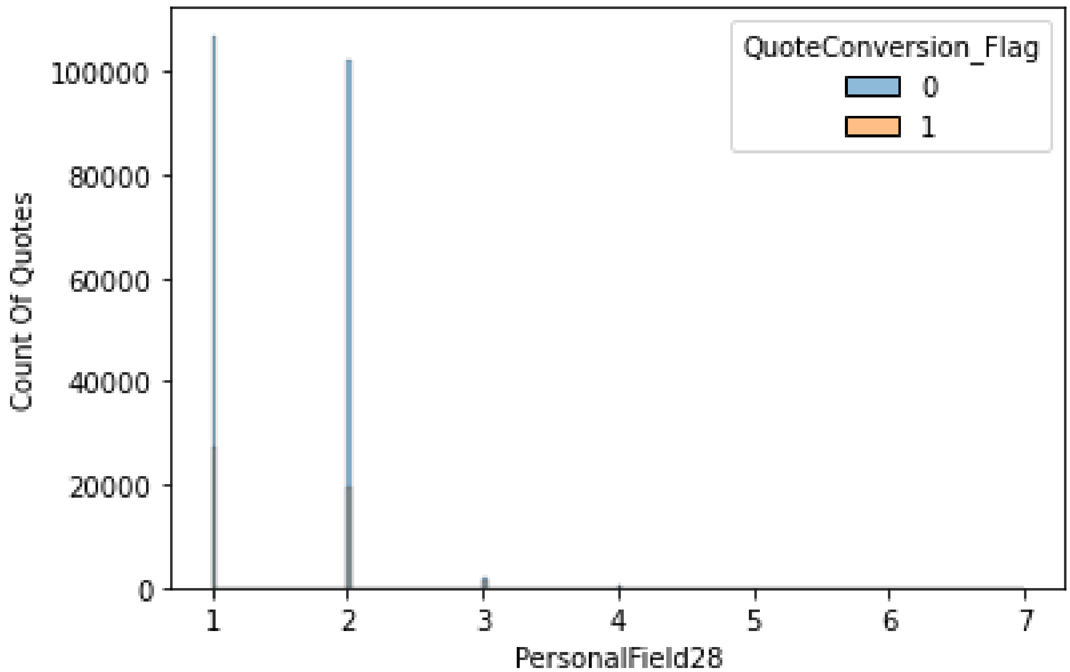

QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
0
2
4
6
8
10
12
14
PersonalField25

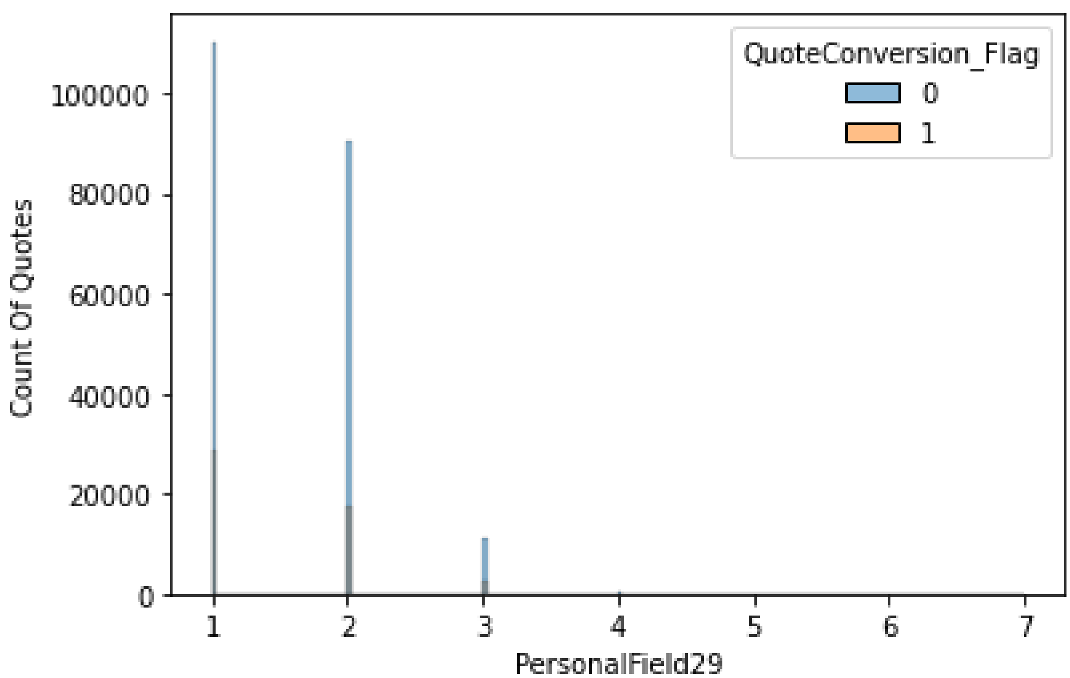


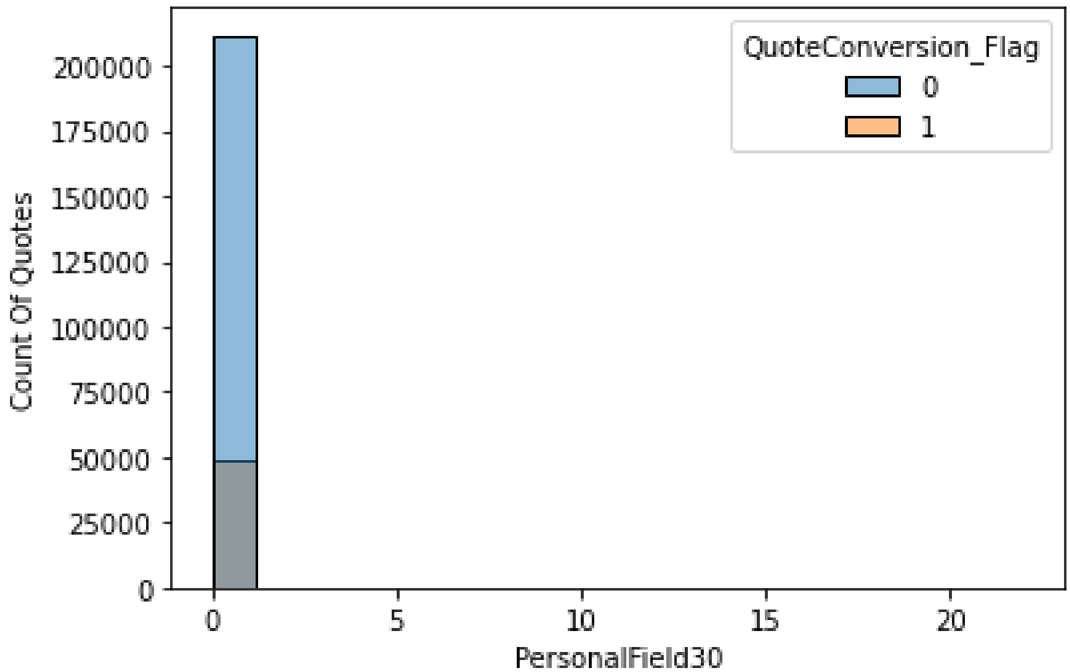
QuoteConversion_Flag
0
1
Count Of Quotes
PersonalField26



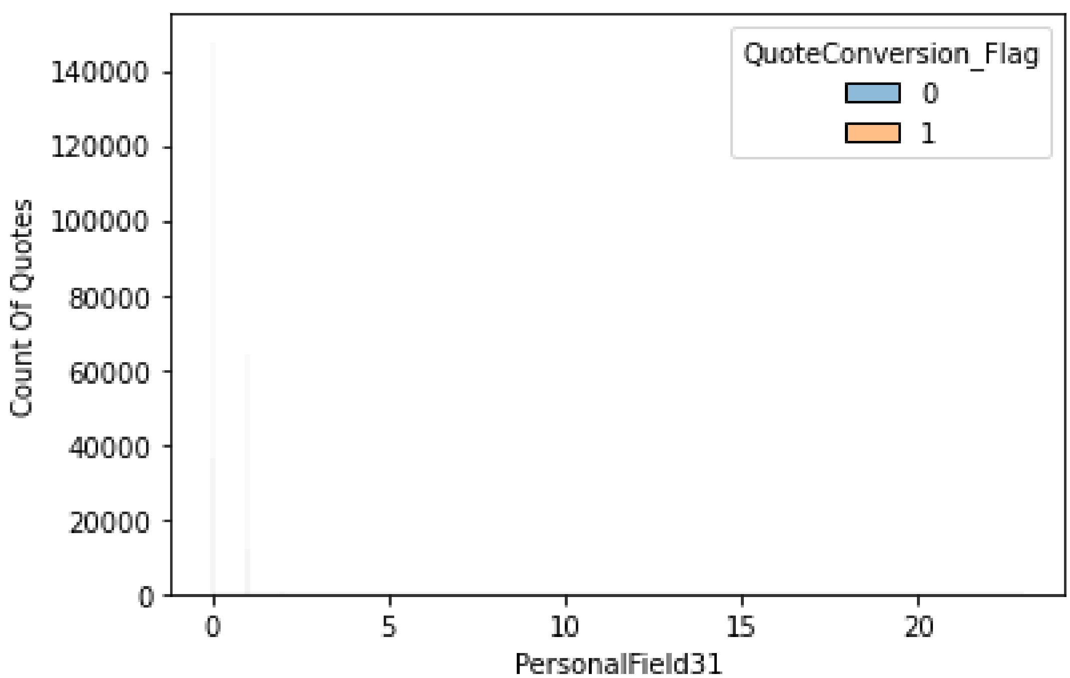
QuoteConversion_Flag
0
1
Count Of Quotes
PersonalField27


QuoteConversion_Flag
0
1
100000
80000
60000
40000
20000
0
Count Of Quotes
1
2
3
4
5
6
7
PersonalField28

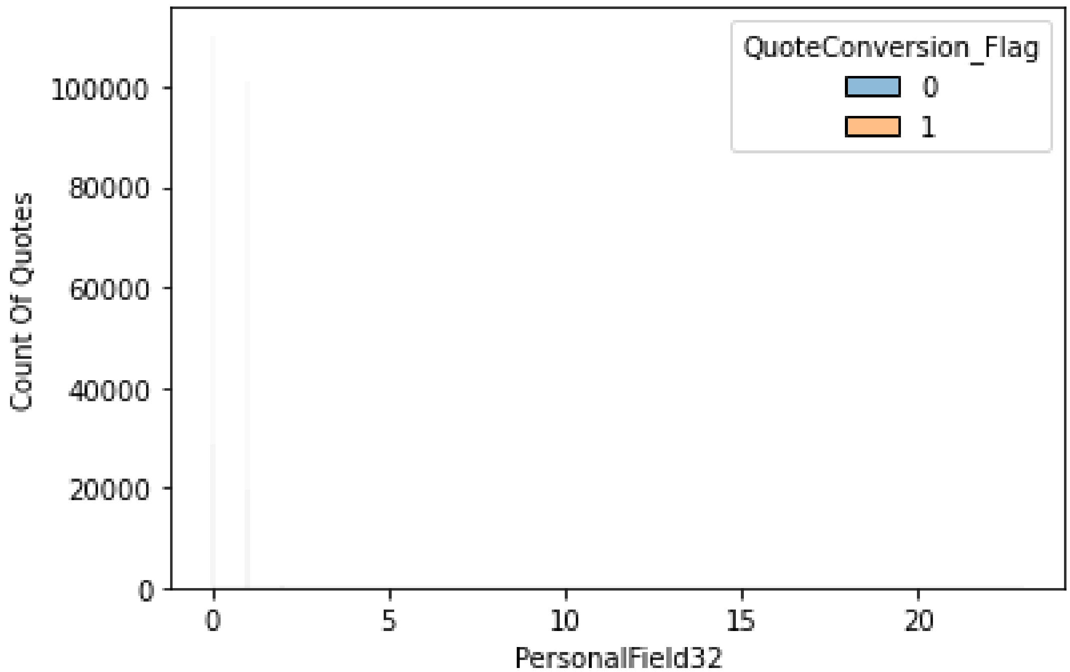

QuoteConversion_Flag
0
1
100000
80000
60000
40000
20000
0
Count Of Quotes
1
2
3
4
5
6
7
PersonalField29

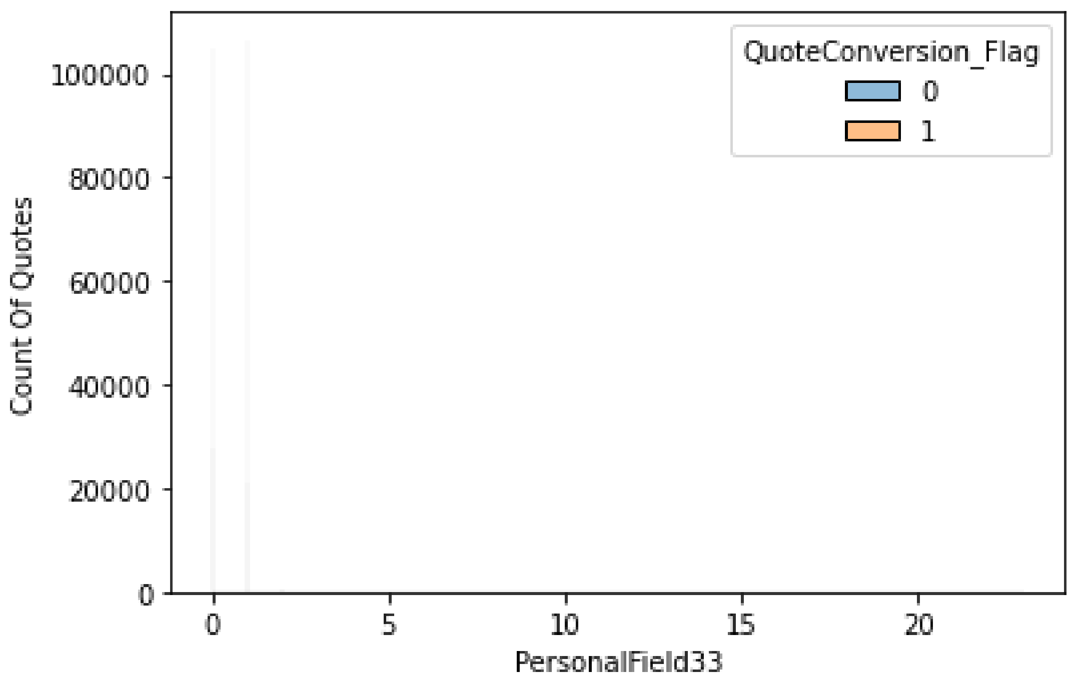


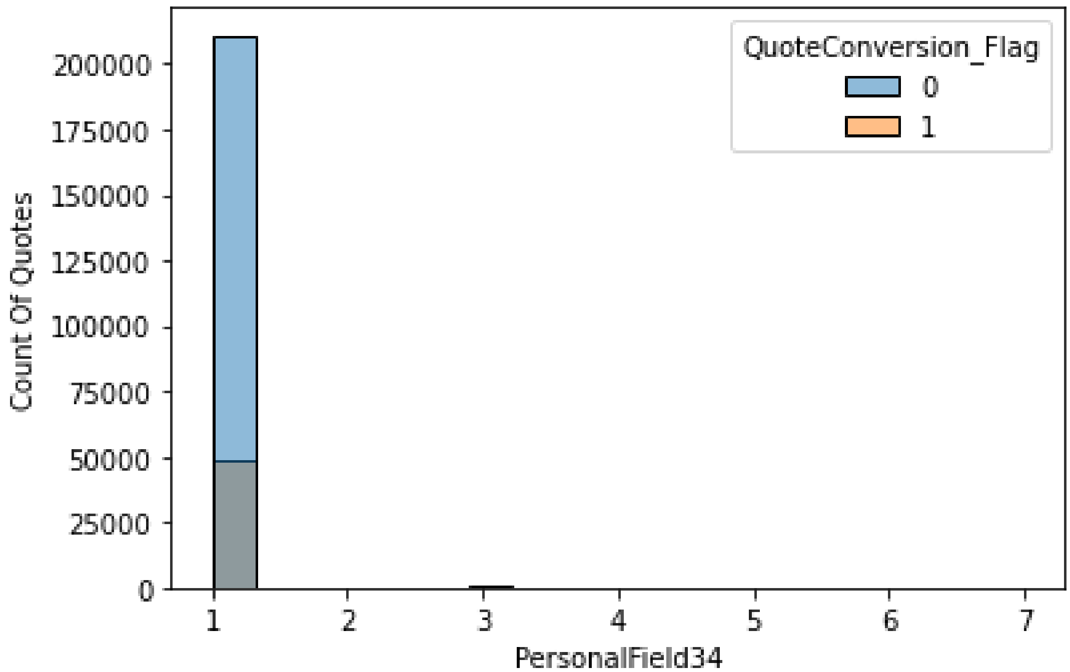
QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
0
5
10
15
20
PersonalField30

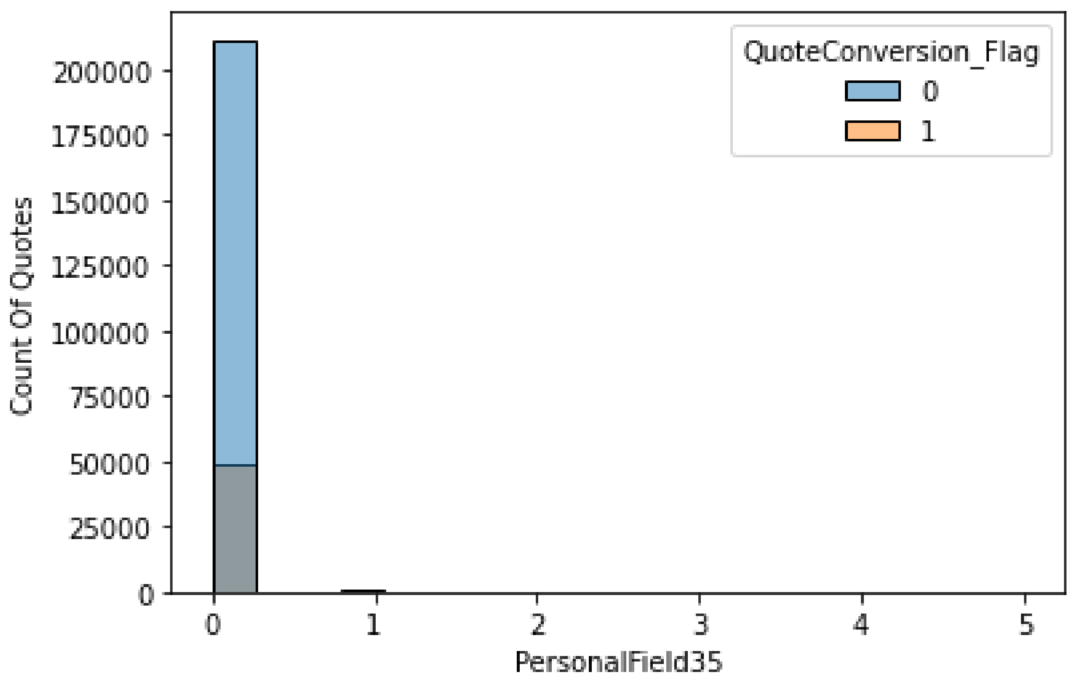


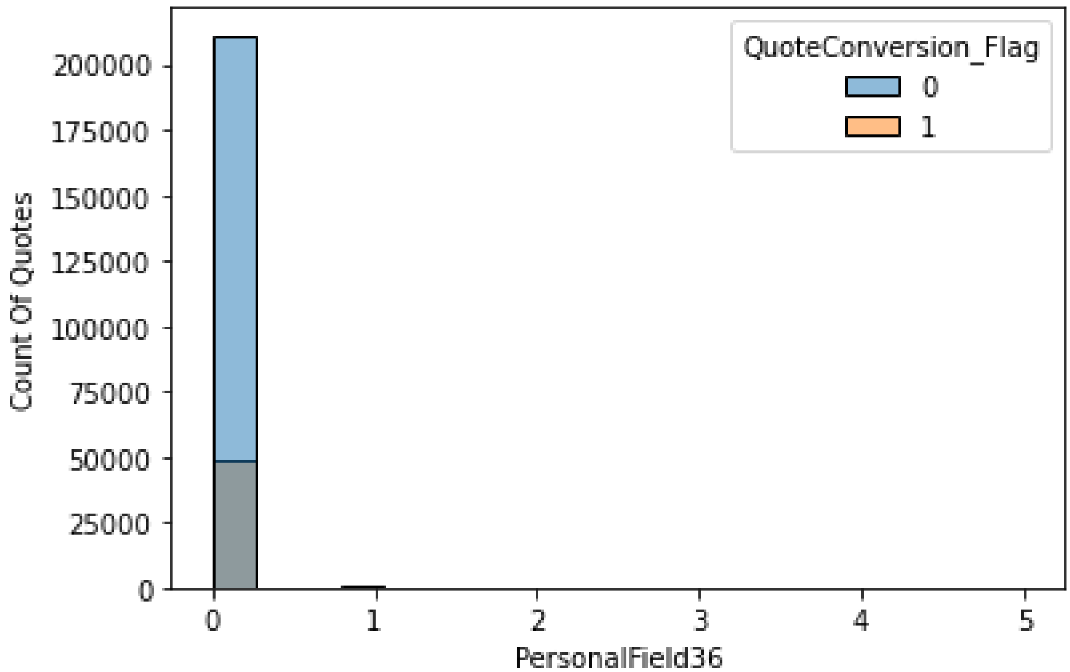
QuoteConversion_Flag
0
1
Count Of Quotes
140000
120000
100000
80000
60000
40000
20000
0
0
5
10
15
20
PersonalField31

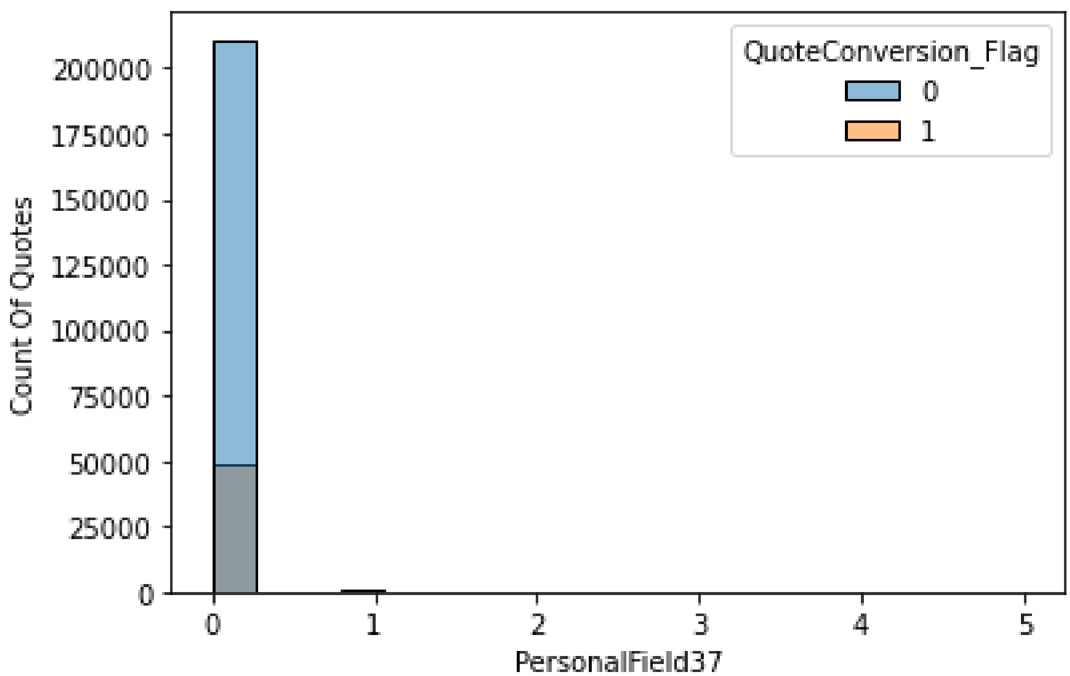

QuoteConversion_Flag
0
1
Count Of Quotes
0
20000
40000
60000
80000
100000
0
5
10
15
20
PersonalField32

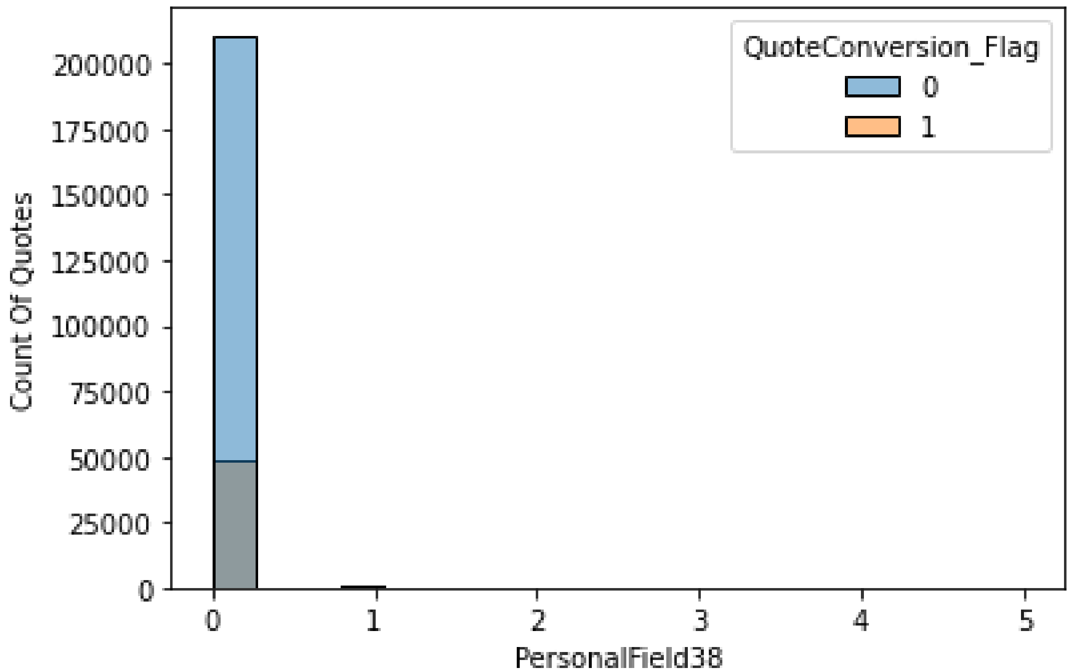

QuoteConversion_Flag
0
1
Count Of Quotes
0
20000
40000
60000
80000
100000
0
5
10
15
20
PersonalField33

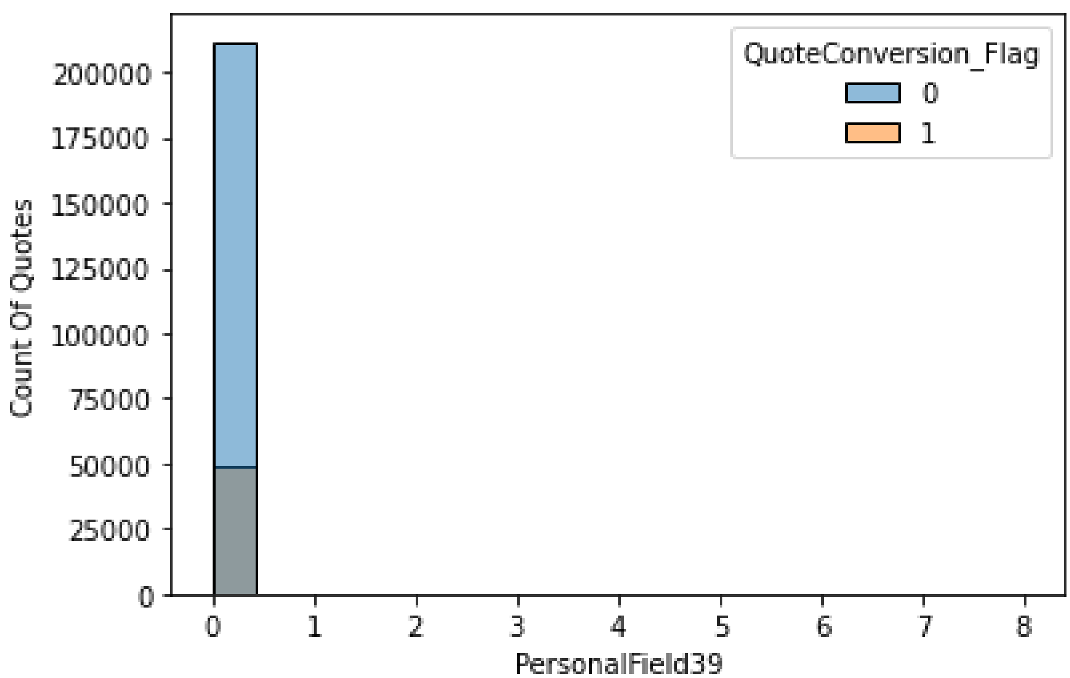

QuoteConversion_Flag
0
1
Count Of Quotes
0
25000
50000
75000
100000
125000
150000
175000
200000
1
2
3
4
5
6
7
PersonalField34

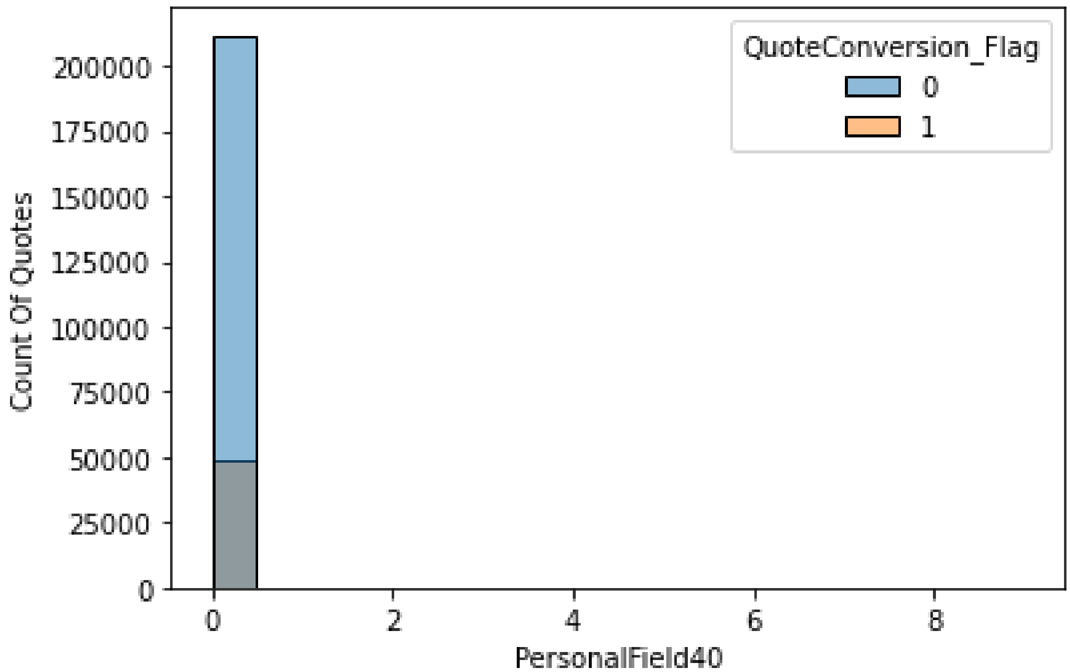

QuoteConversion_Flag
0
1
Count Of Quotes
0
25000
50000
75000
100000
125000
150000
175000
200000
0
1
2
3
4
5
PersonalField35

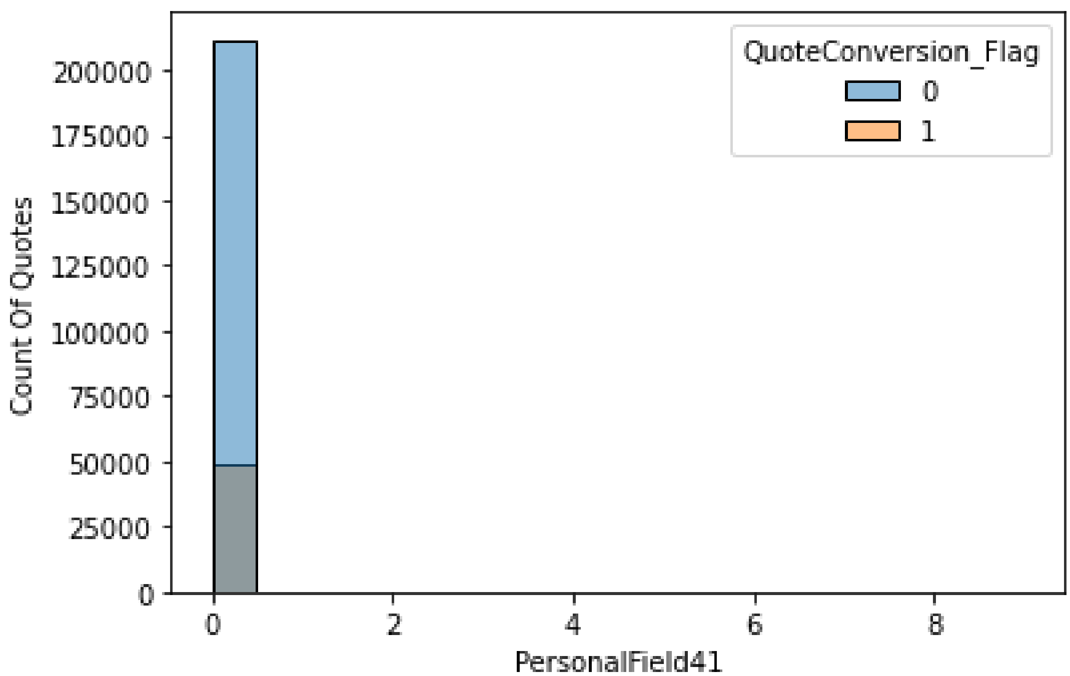

QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
0
1
2
3
4
5
PersonalField36

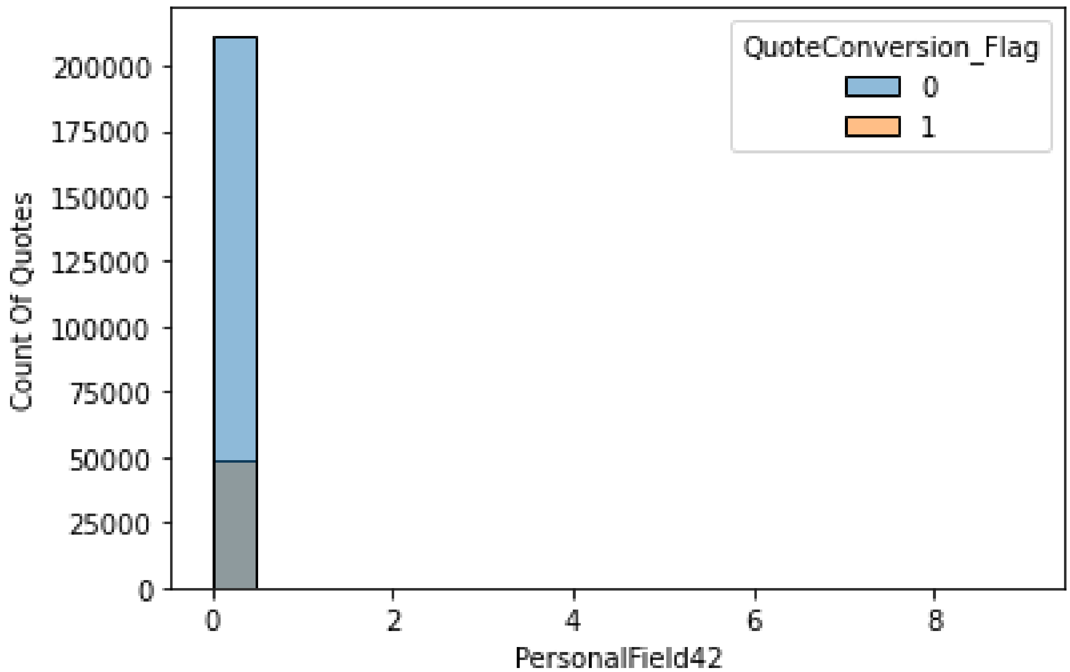

QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
0
1
2
3
4
5
PersonalField37

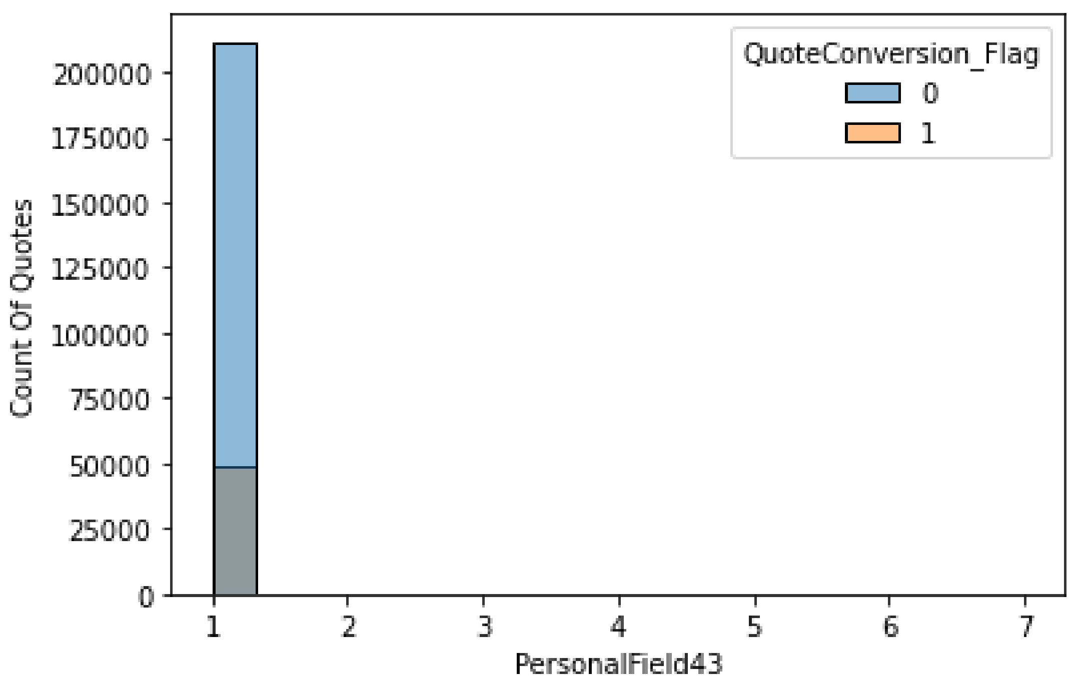

QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
0
1
2
3
4
5
PersonalField38

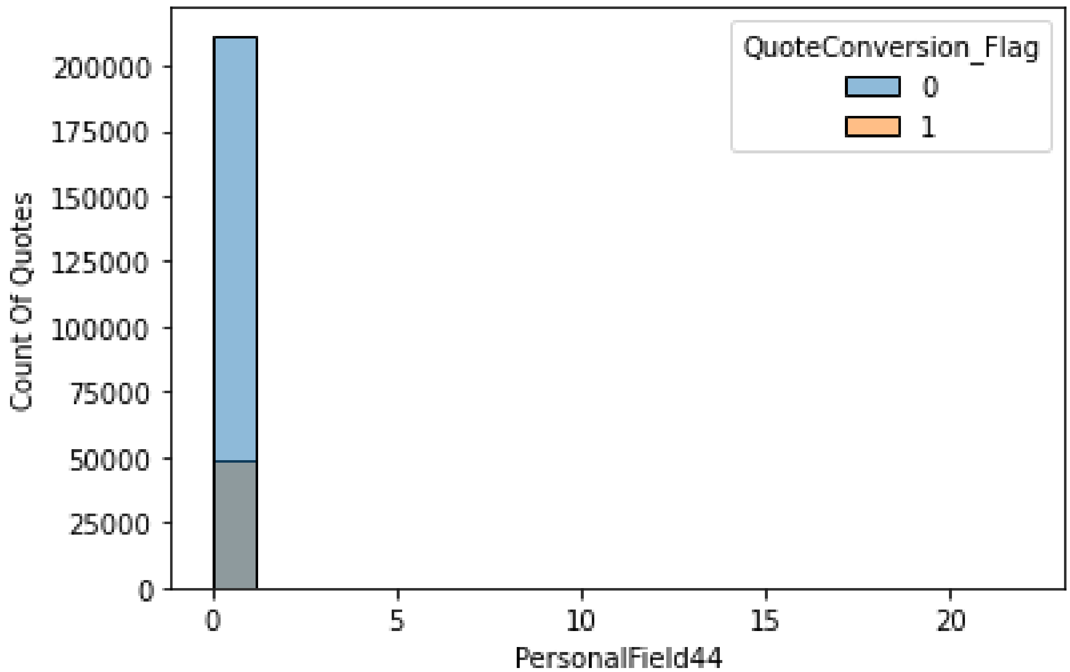

QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
0
1
2
3
4
5
6
7
8
PersonalField39

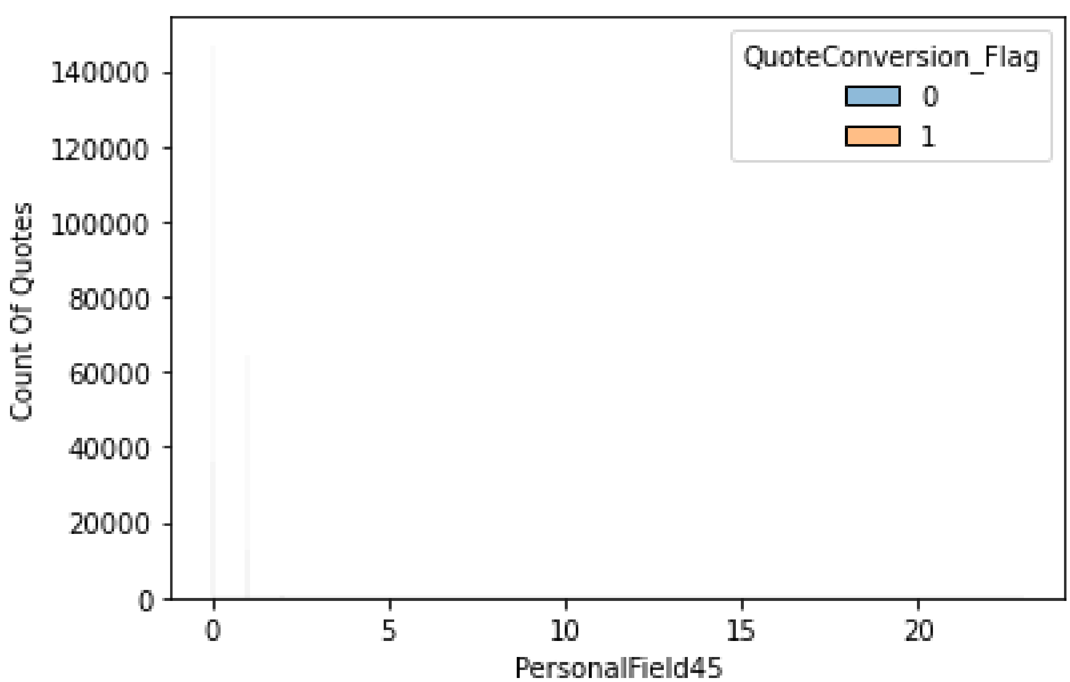

QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
0
2
4
6
8
PersonalField40

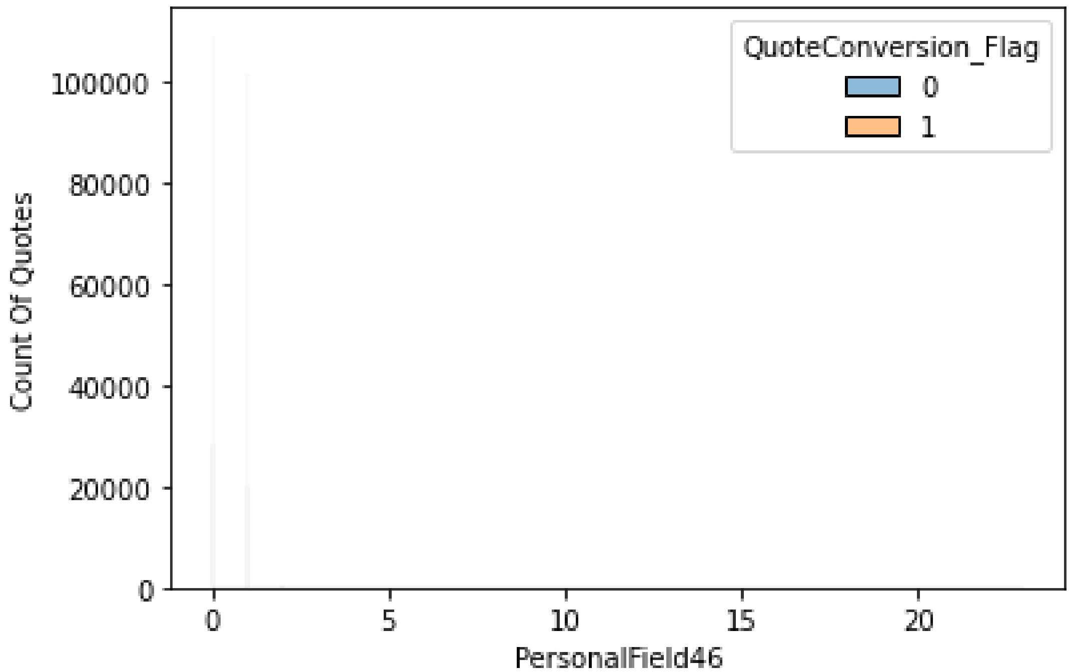

QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
0
2
4
6
8
PersonalField41

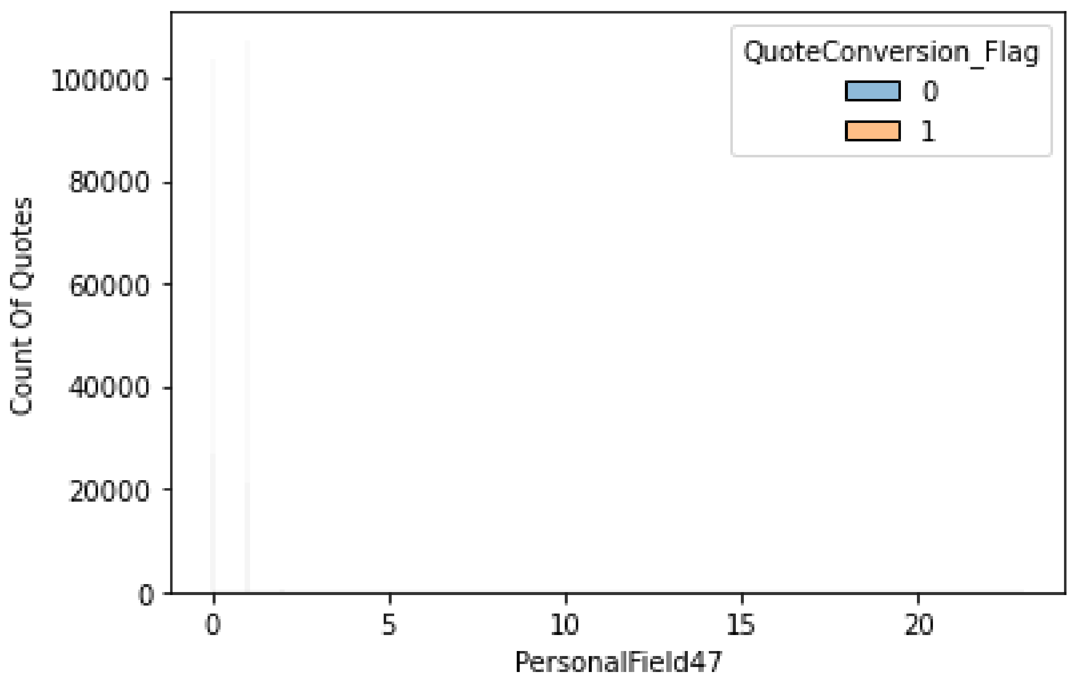

QuoteConversion_Flag
0
1
200000
175000
150000
125000
100000
75000
50000
25000
0
Count Of Quotes
0
2
4
6
8
PersonalField42

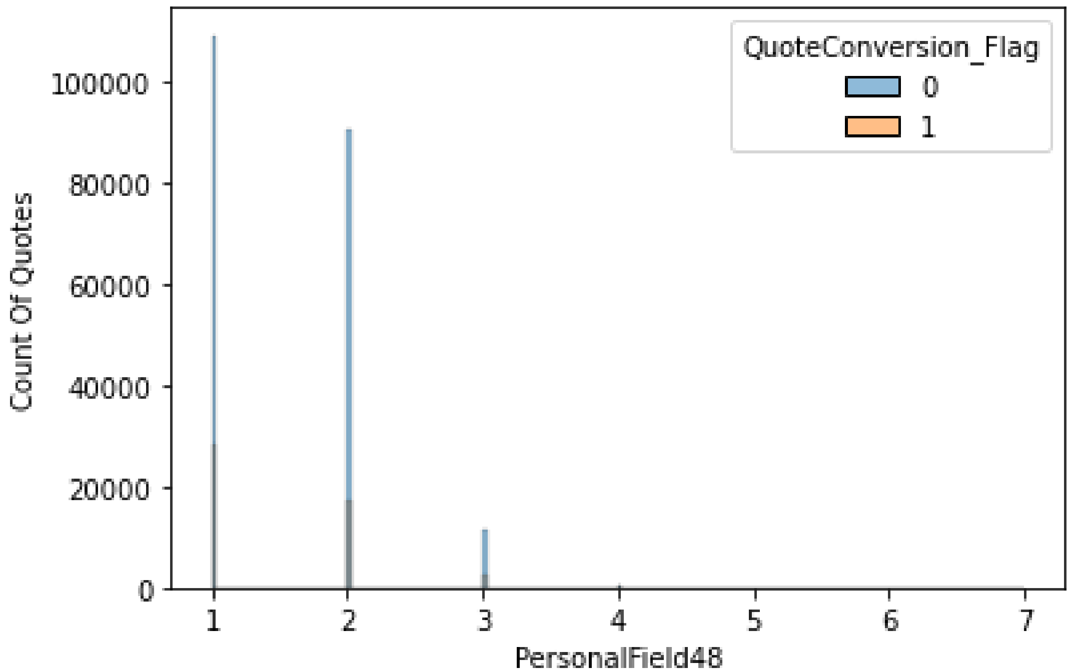

QuoteConversion_Flag
0
1
200000
175000
150000
125000
100000
75000
50000
25000
0
Count Of Quotes
1
2
3
4
5
6
7
PersonalField43

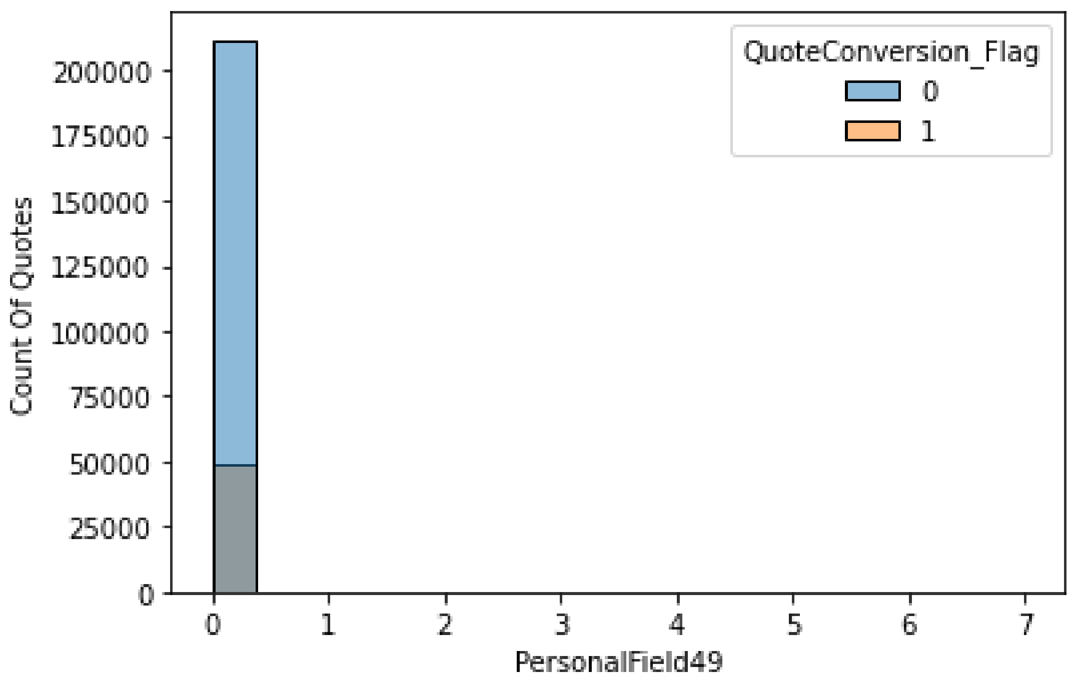

QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
0
5
10
15
20
PersonalField44

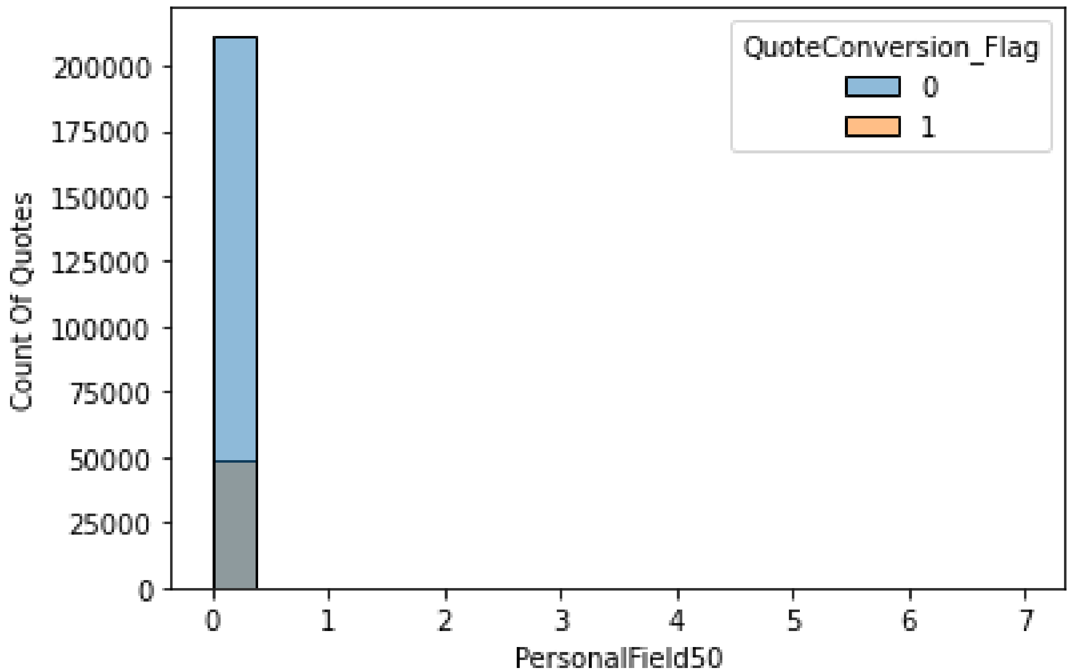

QuoteConversion_Flag
0
1
Count Of Quotes
140000
120000
100000
80000
60000
40000
20000
0
0
5
10
15
20
PersonalField45


QuoteConversion_Flag
0
1
Count Of Quotes
PersonalField46
0
20000
40000
60000
80000
100000
5
10
15
20

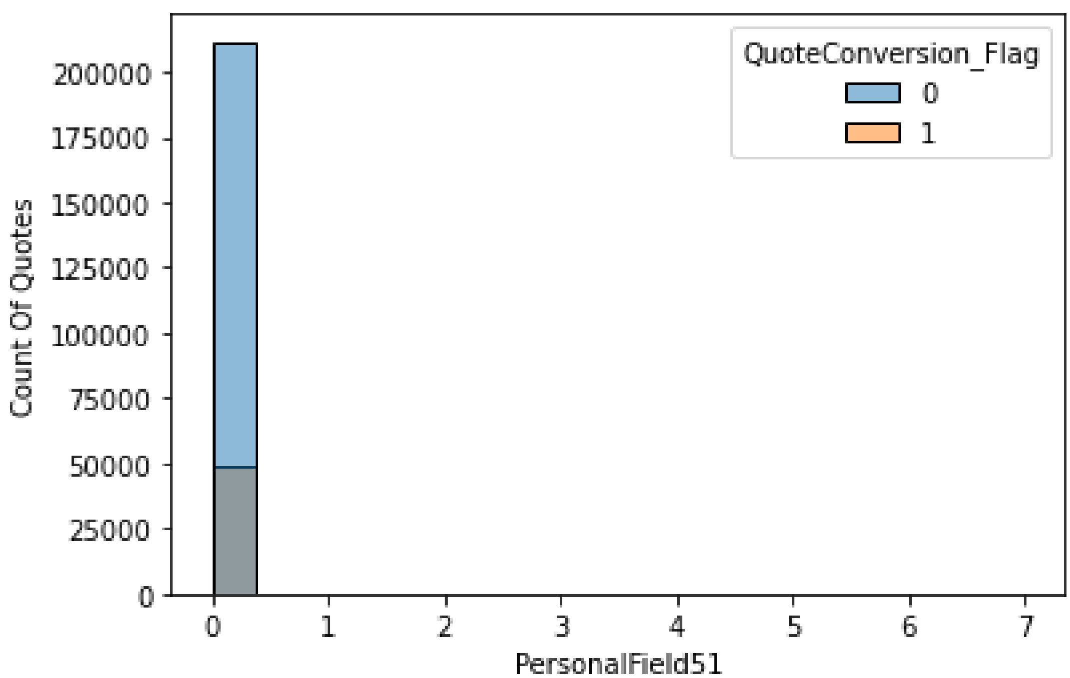

QuoteConversion_Flag
0
1
Count Of Quotes
PersonalField47
0
20000
40000
60000
80000
100000
5
10
15
20


QuoteConversion_Flag
0
1
Count Of Quotes
100000
80000
60000
40000
20000
0
1
2
3
4
5
6
7
PersonalField48

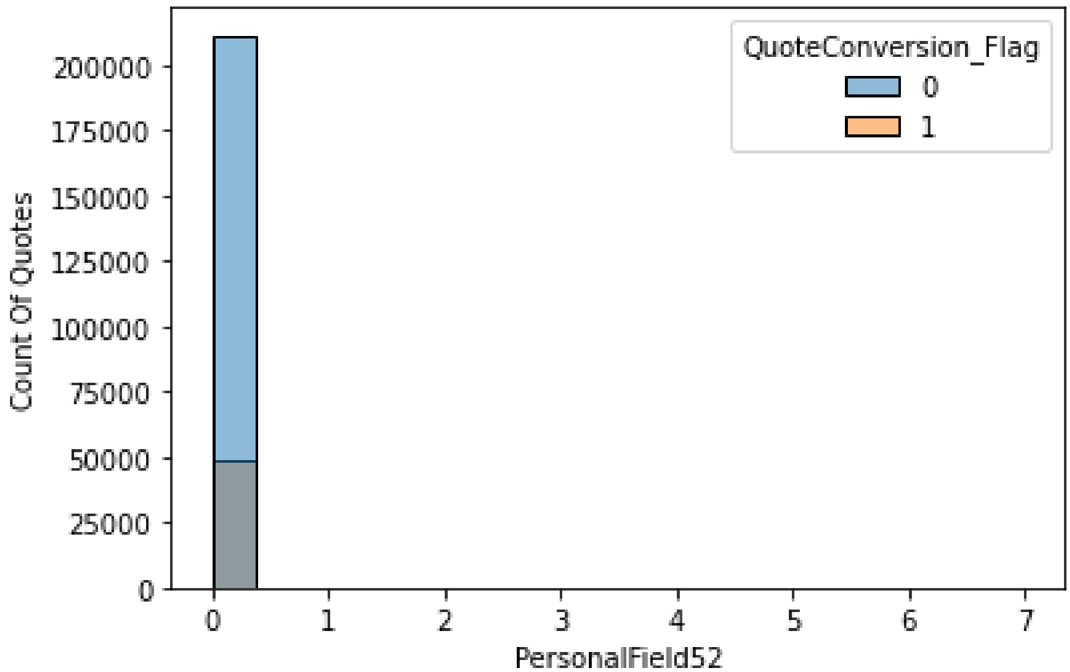

QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
0
1
2
3
4
5
6
7
PersonalField49

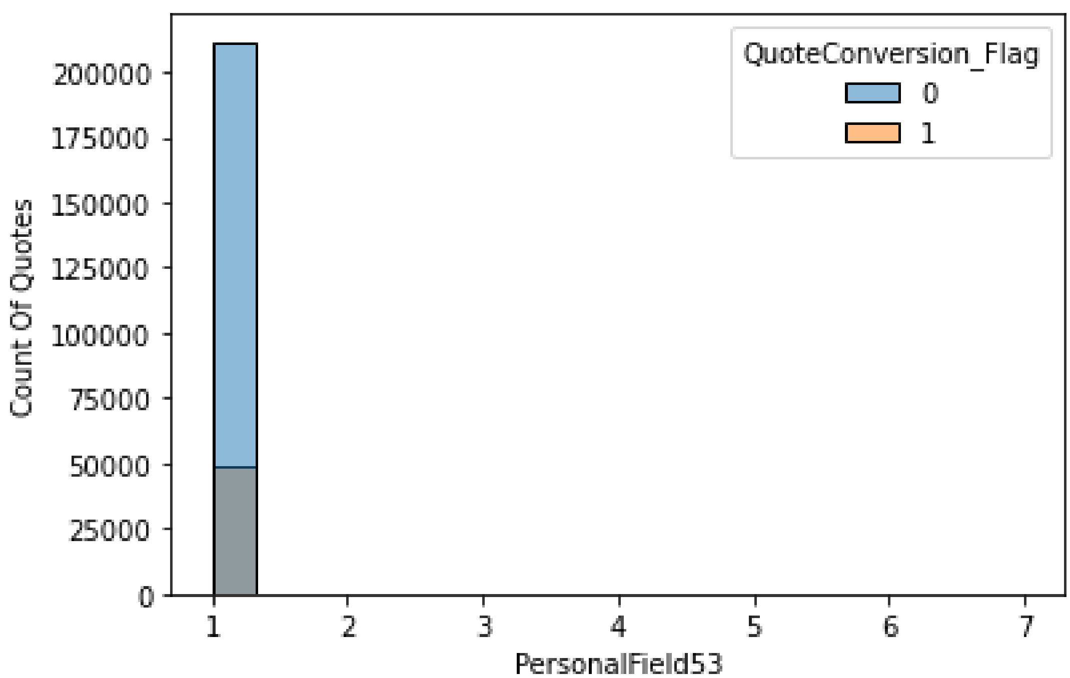


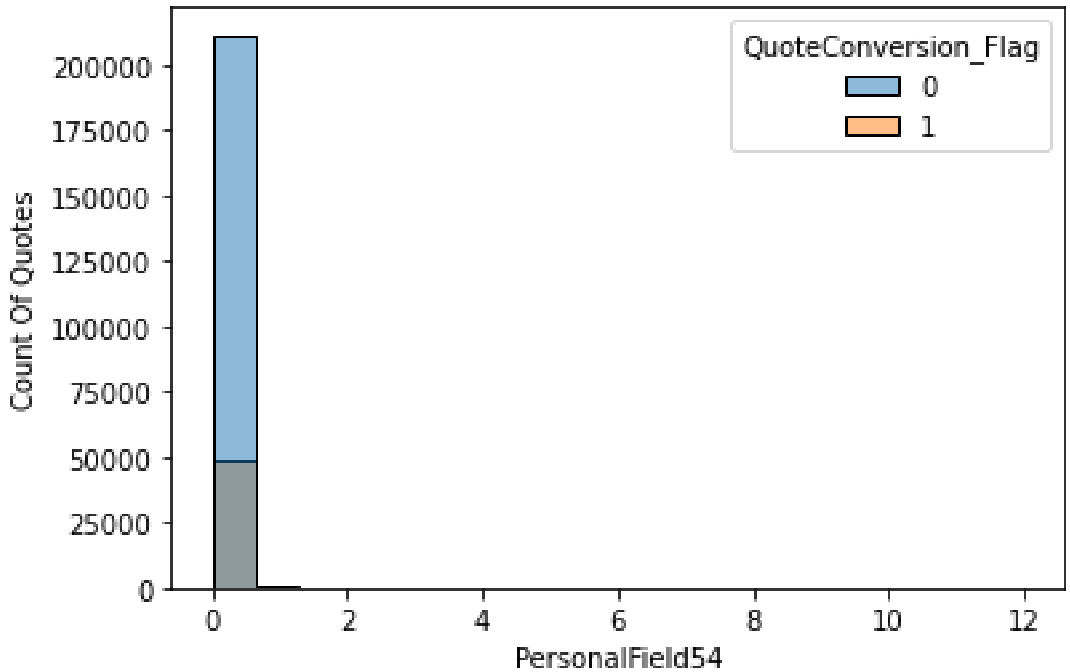
QuoteConversion_Flag
0
1
Count Of Quotes
0
25000
50000
75000
100000
125000
150000
175000
200000
0 1 2 3 4 5 6 7
PersonalField50

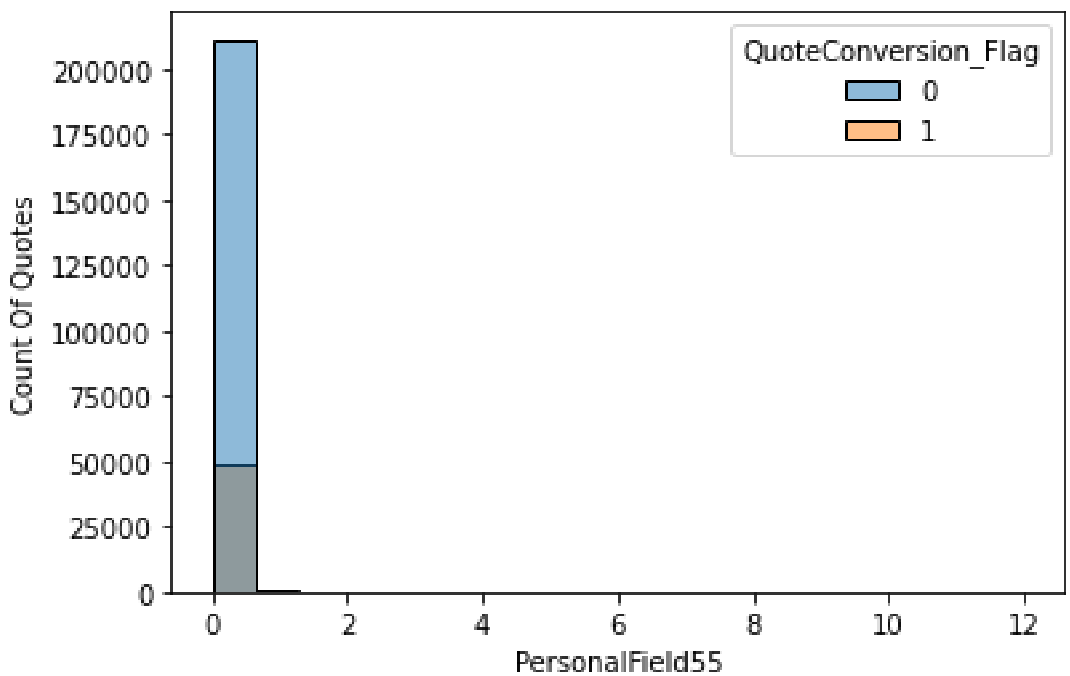


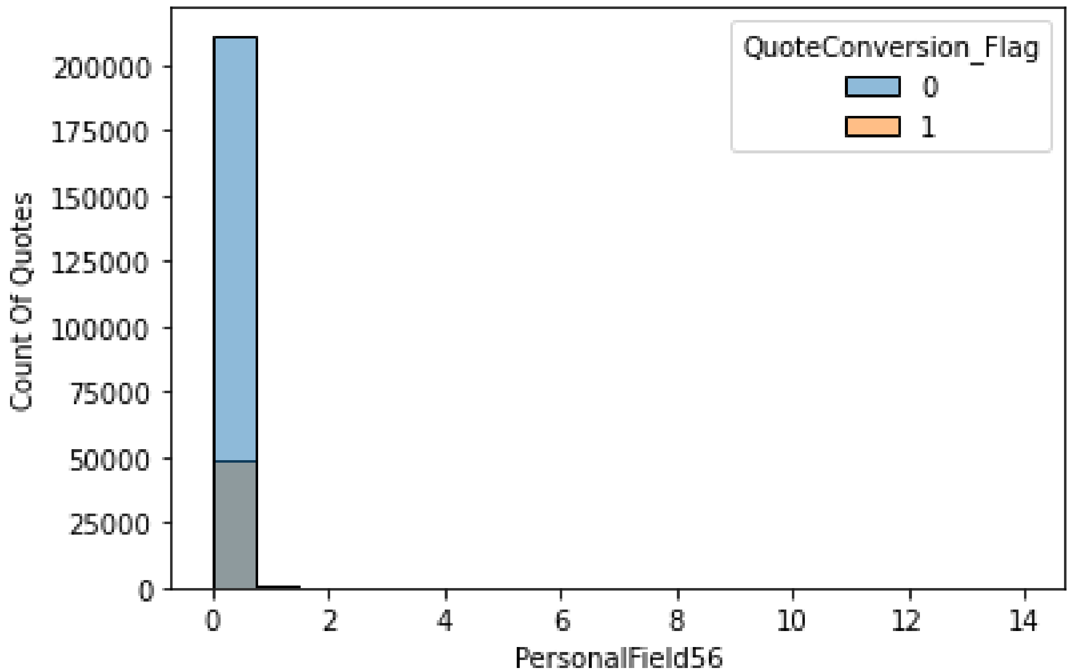
QuoteConversion_Flag
0
1
Count Of Quotes
0
25000
50000
75000
100000
125000
150000
175000
200000
0 1 2 3 4 5 6 7
PersonalField51

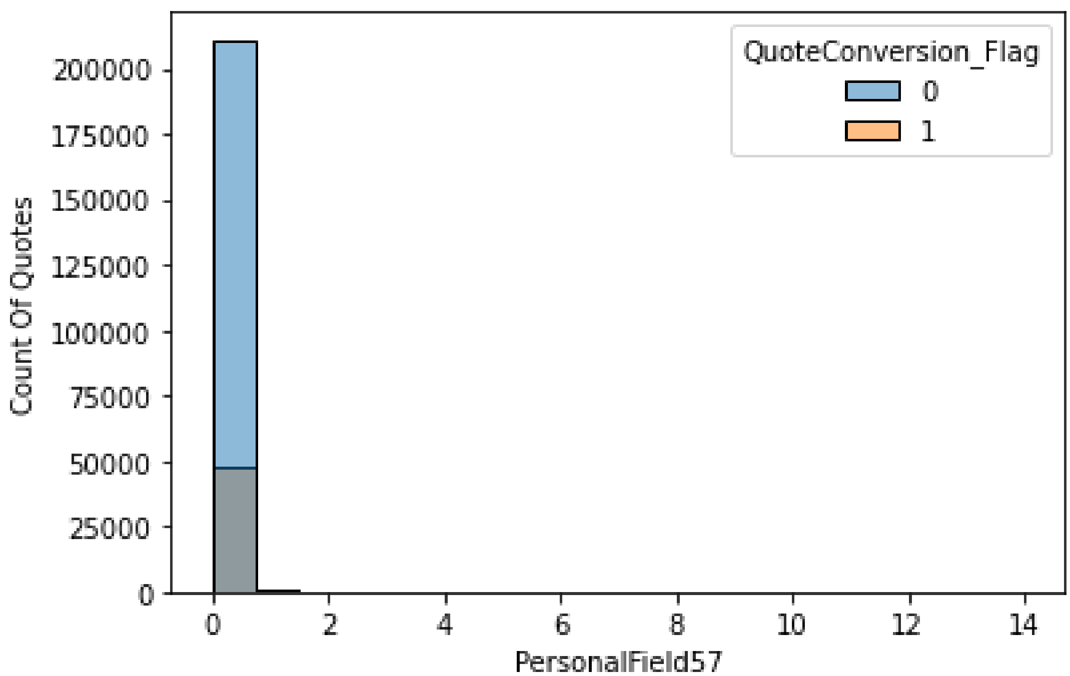

QuoteConversion_Flag
0
1
Count Of Quotes
PersonalField52

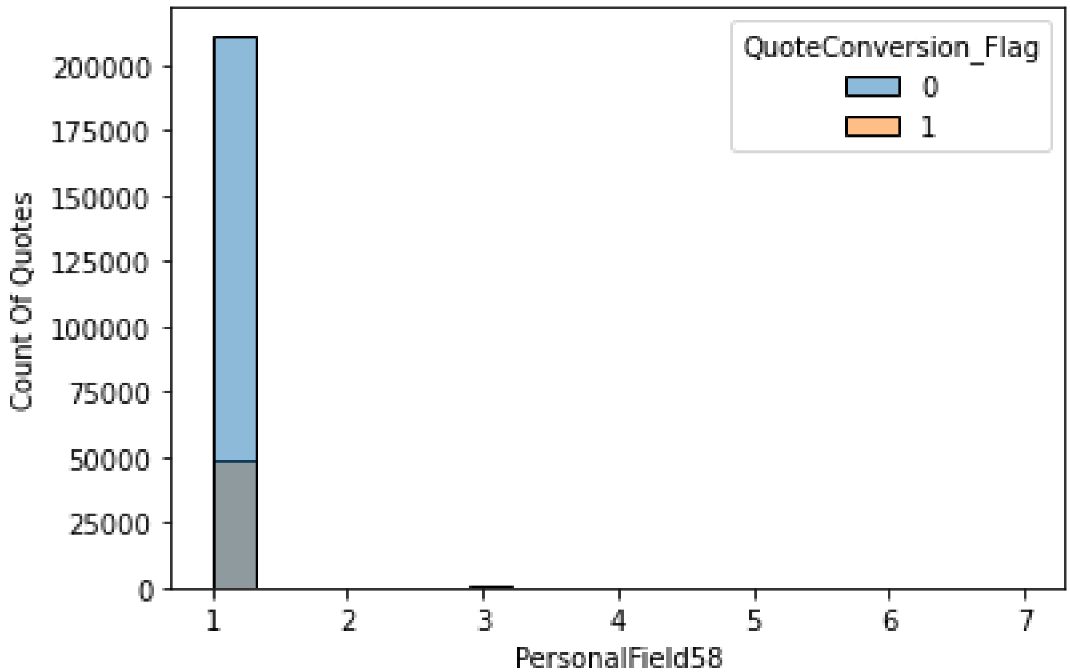

QuoteConversion_Flag
0
1
Count Of Quotes
PersonalField53

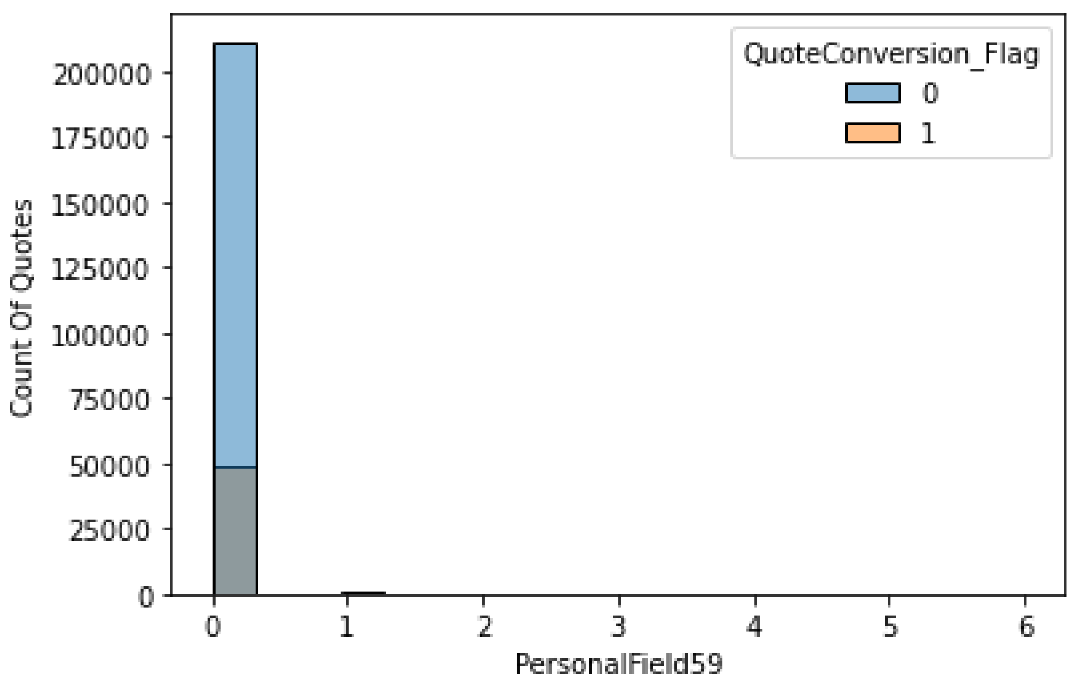

QuoteConversion_Flag
0
1
Count Of Quotes
PersonalField54
0
25000
50000
75000
100000
125000
150000
175000
200000
2
4
6
8
10
12

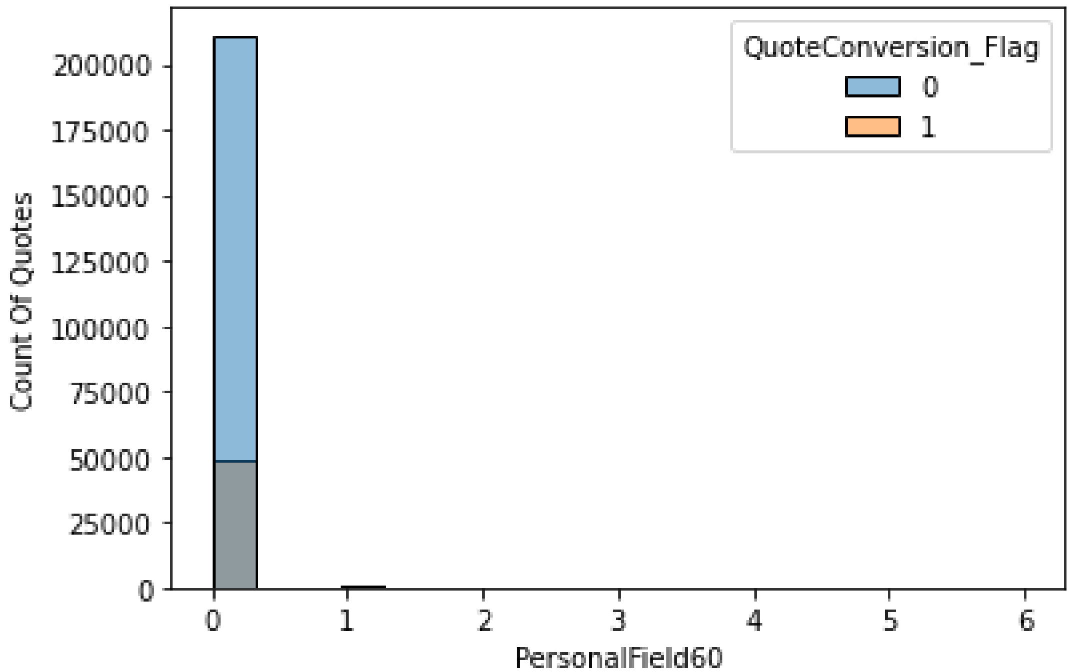

QuoteConversion_Flag
0
1
Count Of Quotes
PersonalField55
0
25000
50000
75000
100000
125000
150000
175000
200000
2
4
6
8
10
12

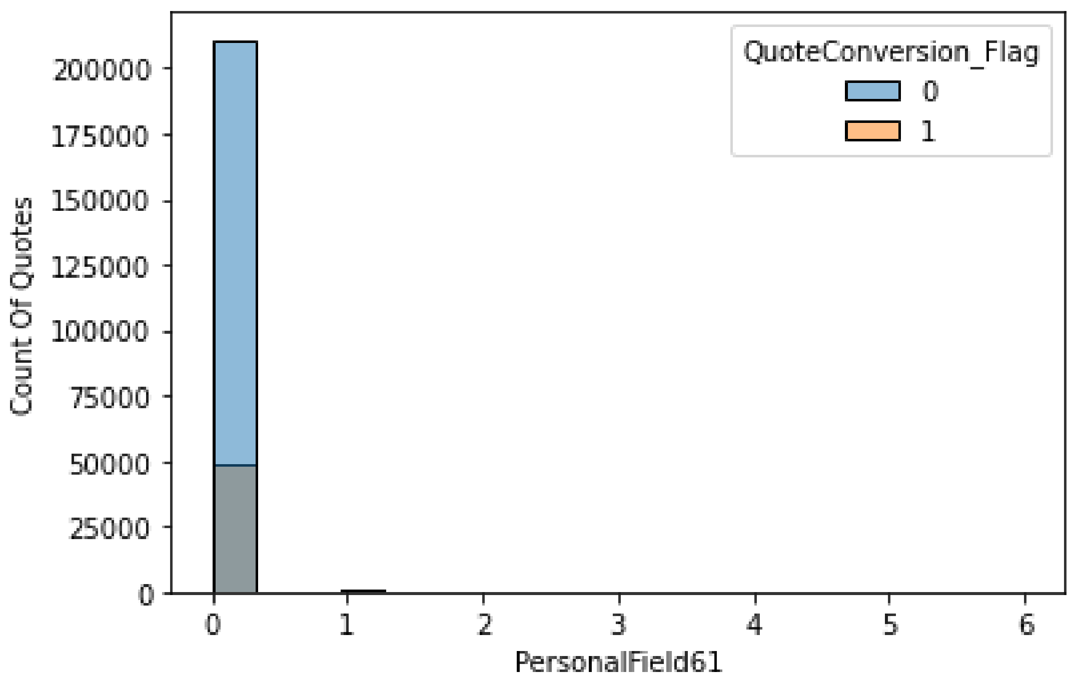

QuoteConversion_Flag
0
1
Count Of Quotes
PersonalField56

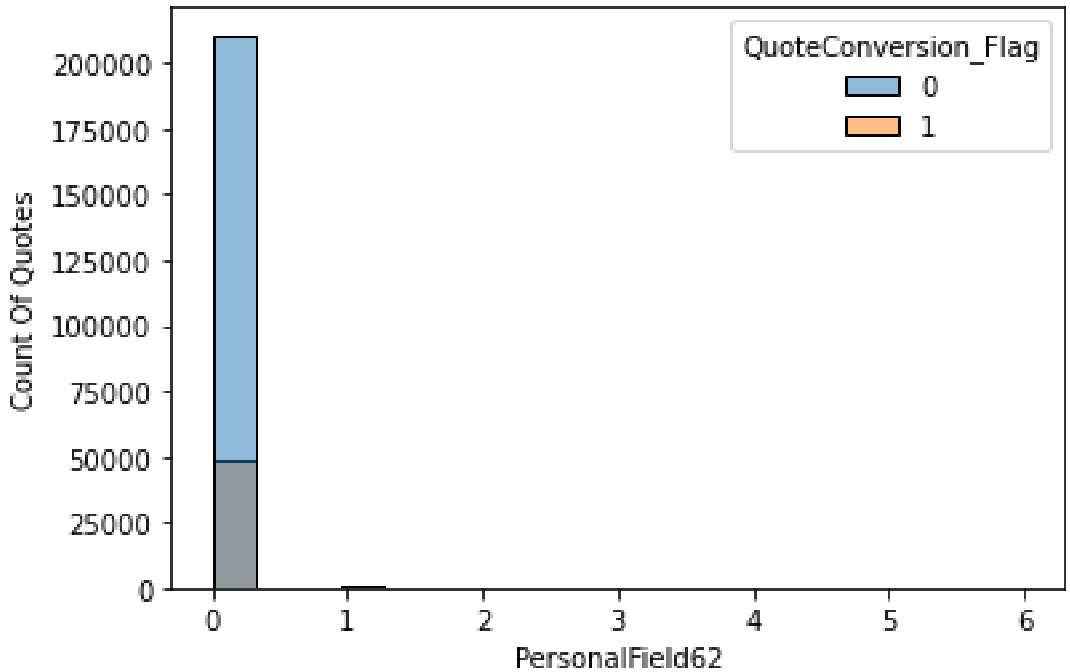

QuoteConversion_Flag
0
1
Count Of Quotes
PersonalField57

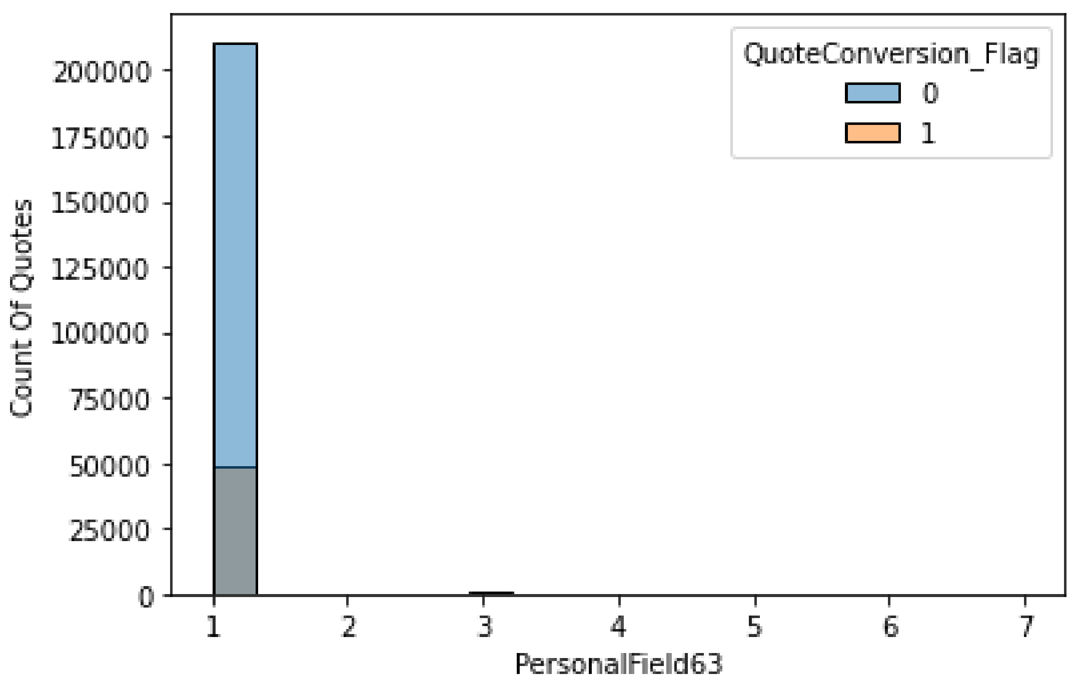

QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
1
2
3
4
5
6
7
PersonalField58

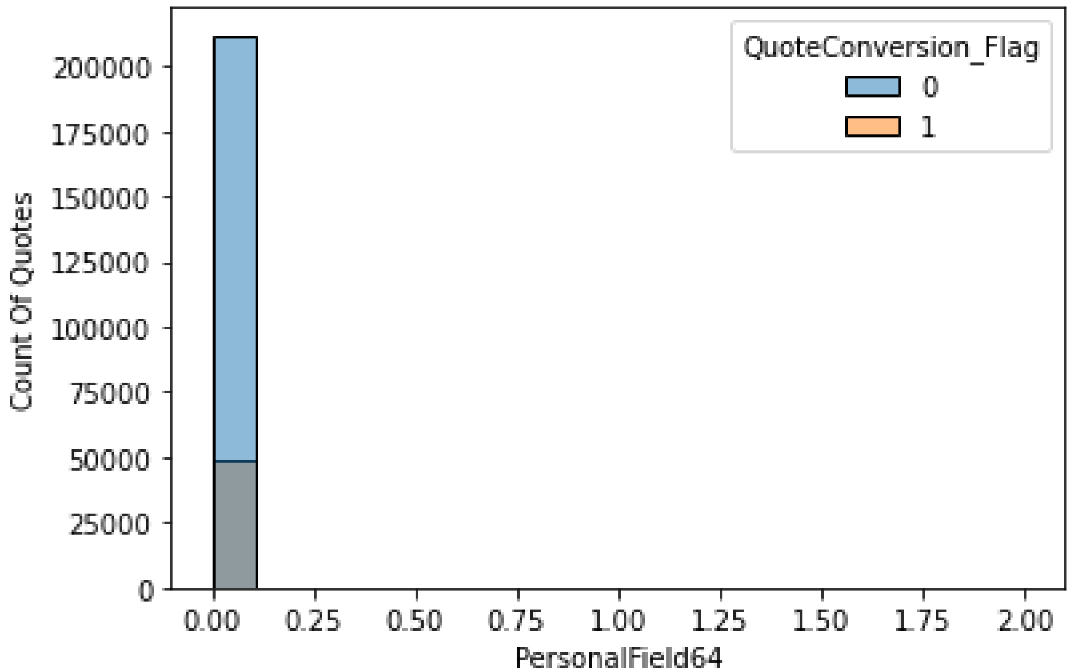

QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
0
1
2
3
4
5
6
PersonalField59

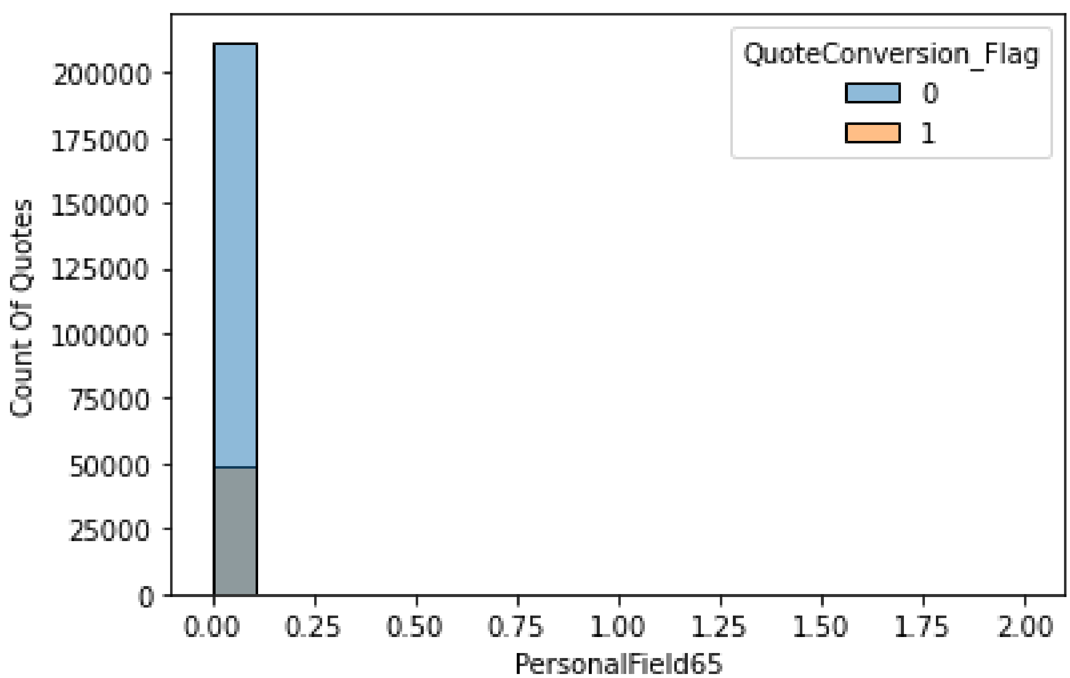

QuoteConversion_Flag
0
1
Count Of Quotes
0
25000
50000
75000
100000
125000
150000
175000
200000
0
1
2
3
4
5
6
PersonalField60

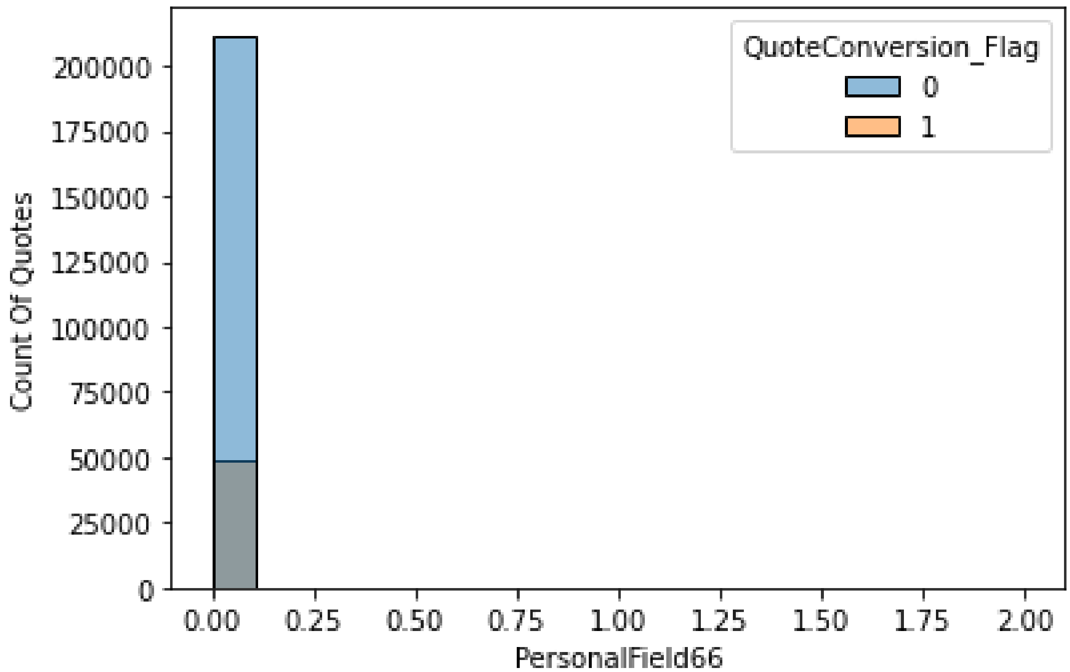

QuoteConversion_Flag
0
1
Count Of Quotes
0
25000
50000
75000
100000
125000
150000
175000
200000
0
1
2
3
4
5
6
PersonalField61

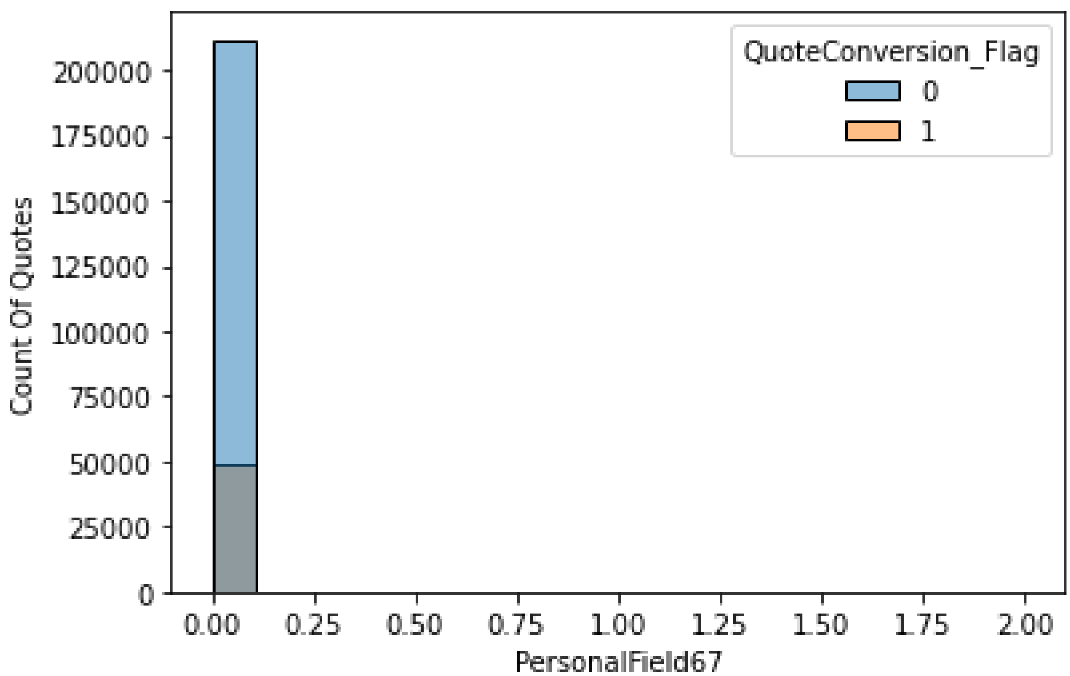

QuoteConversion_Flag
0
1
Count Of Quotes
PersonalField62

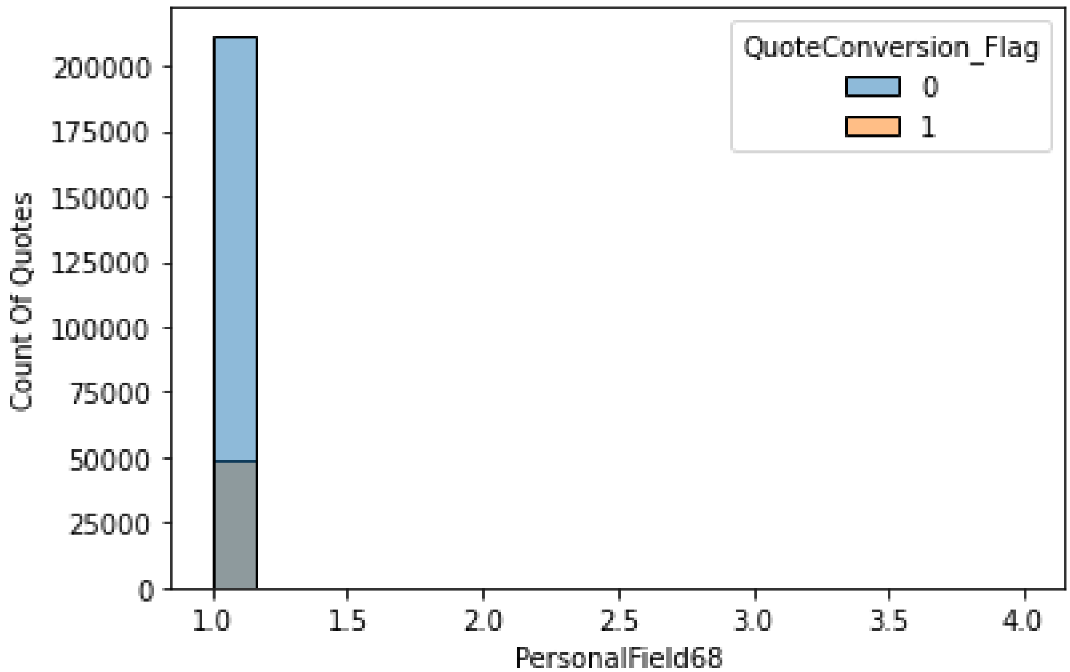

QuoteConversion_Flag
0
1
Count Of Quotes
PersonalField63

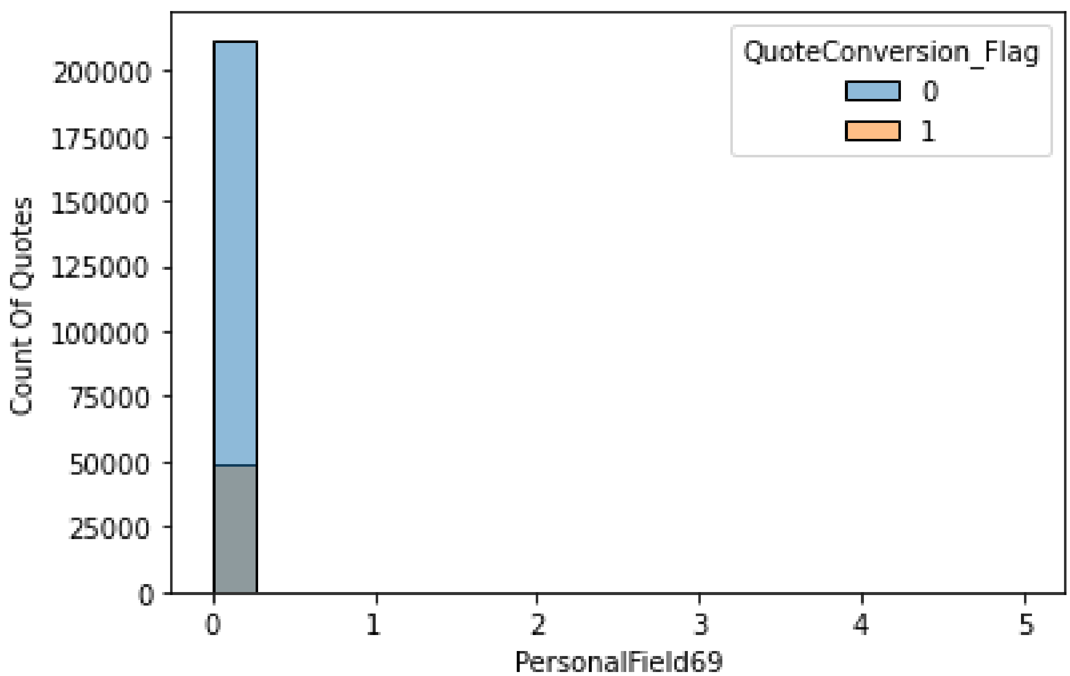

QuoteConversion_Flag
0
1
Count Of Quotes
PersonalField64
0
25000
50000
75000
100000
125000
150000
175000
200000
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00

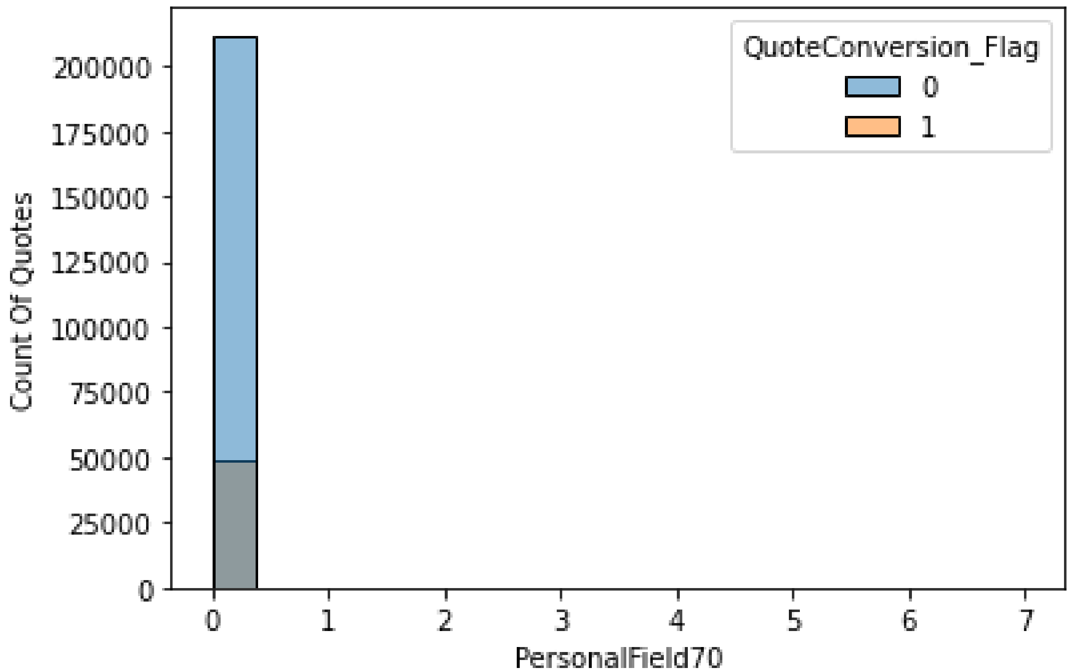

QuoteConversion_Flag
0
1
Count Of Quotes
PersonalField65
0
25000
50000
75000
100000
125000
150000
175000
200000
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00

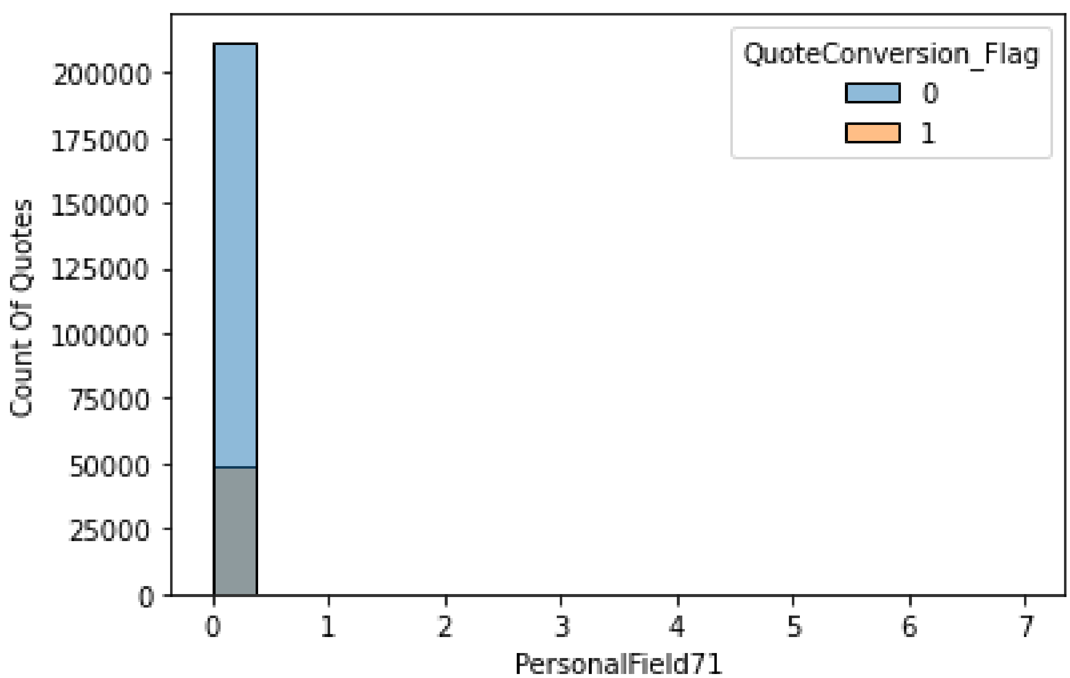

QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
PersonalField66

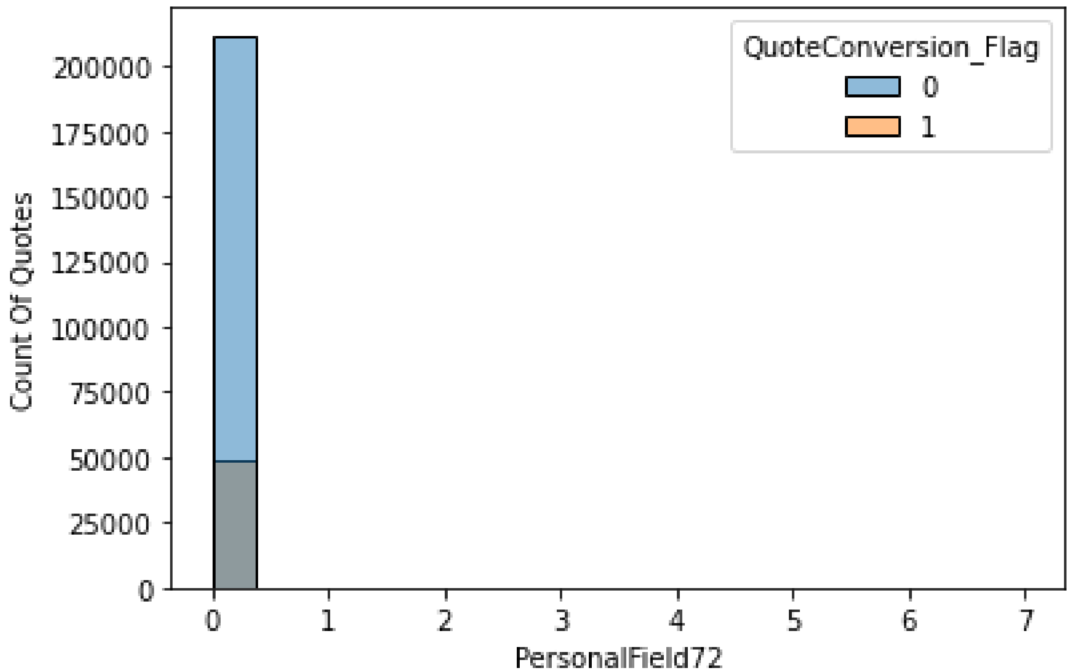

QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
PersonalField67

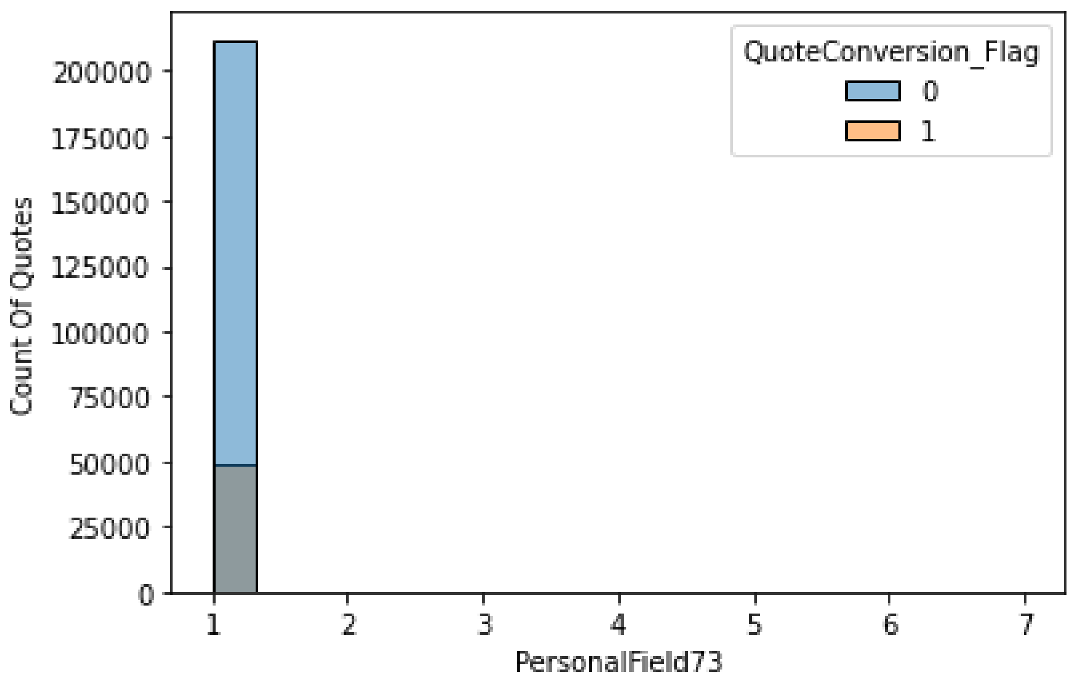

QuoteConversion_Flag
0
1
Count Of Quotes
PersonalField68
0
25000
50000
75000
100000
125000
150000
175000
200000
1.0
1.5
2.0
2.5
3.0
3.5
4.0

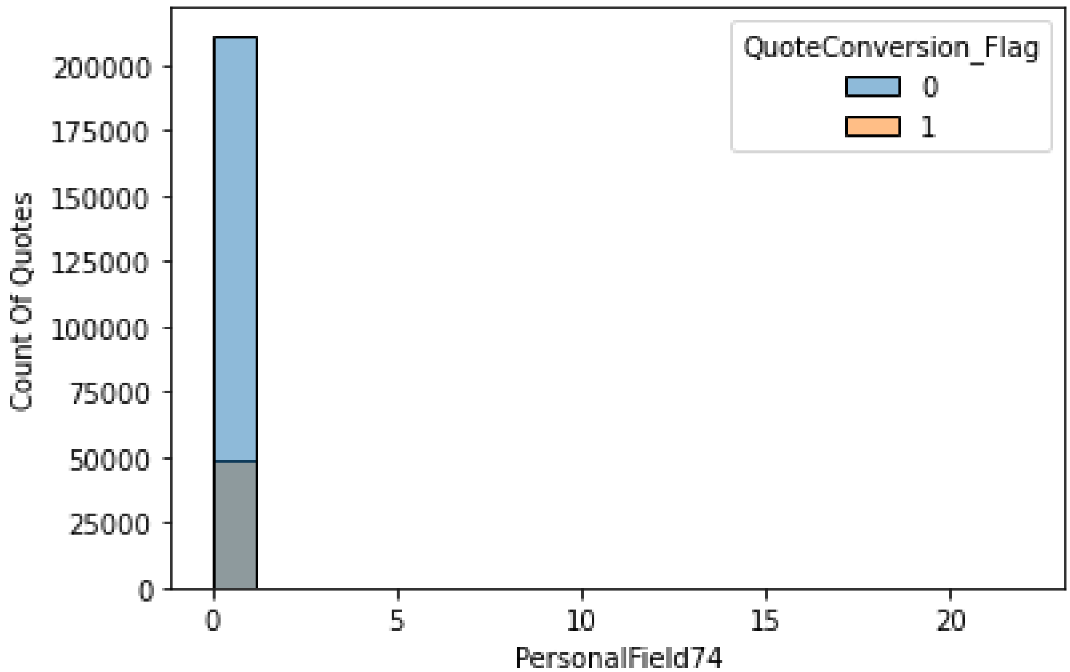

QuoteConversion_Flag
0
1
Count Of Quotes
PersonalField69
0
25000
50000
75000
100000
125000
150000
175000
200000
0
1
2
3
4
5

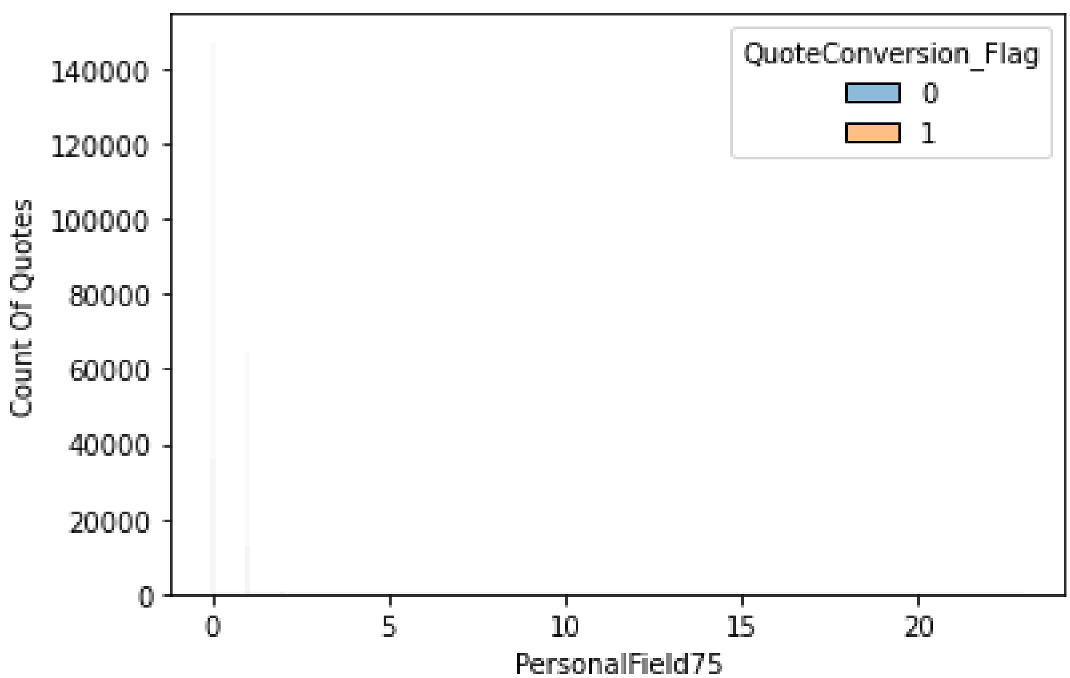

QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
0
1
2
3
4
5
6
7
PersonalField70

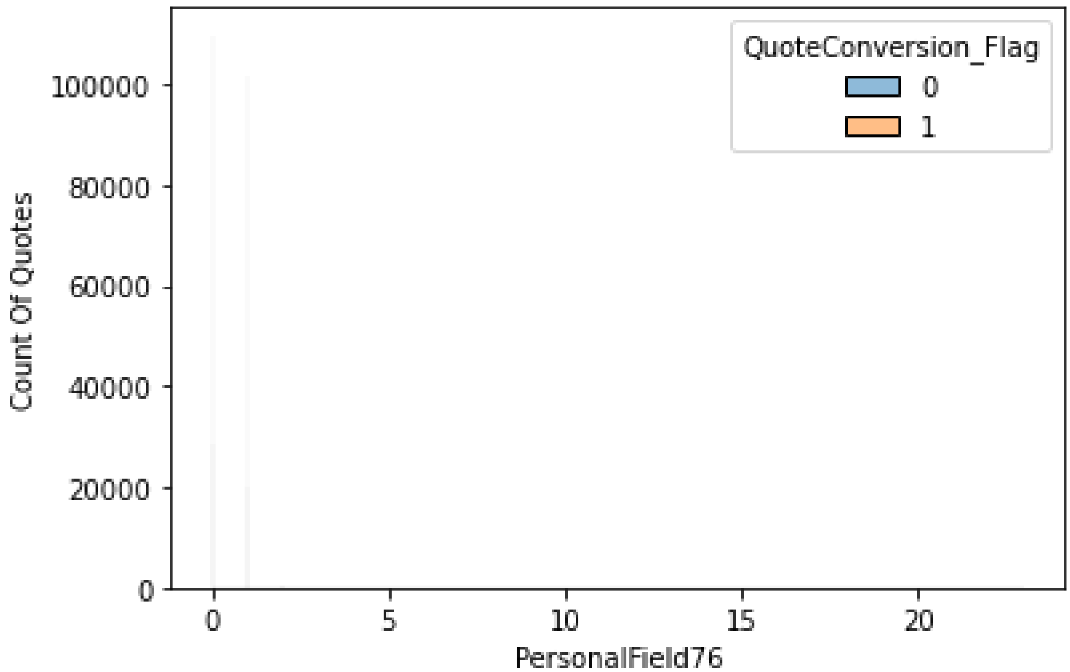

QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
0
1
2
3
4
5
6
7
PersonalField71

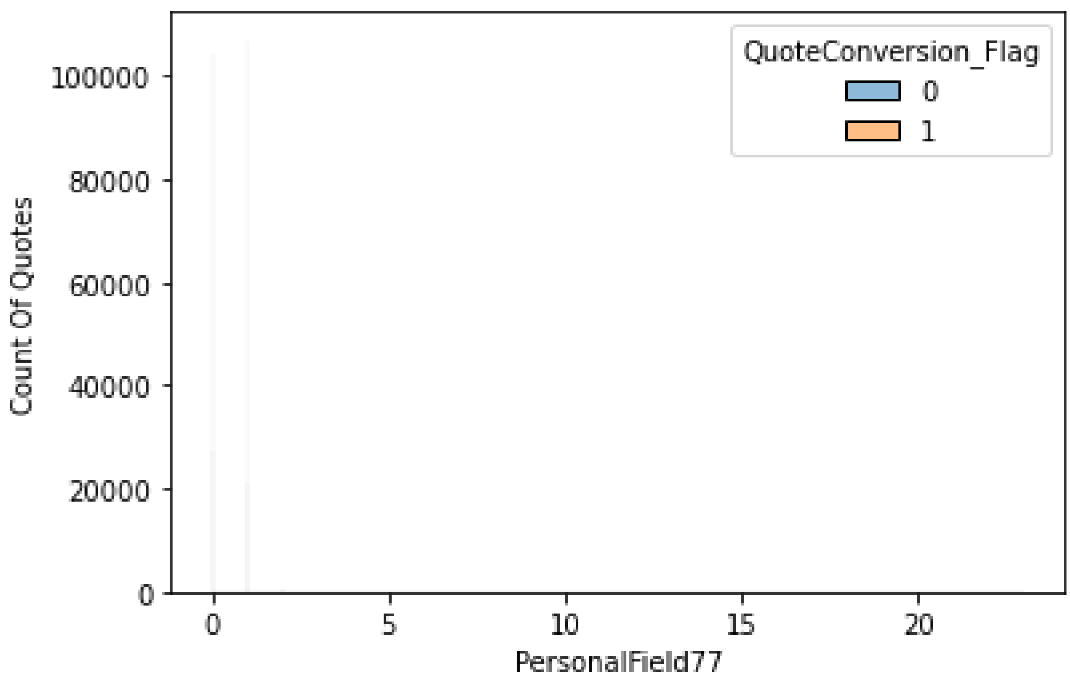

QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
0
1
2
3
4
5
6
7
PersonalField72

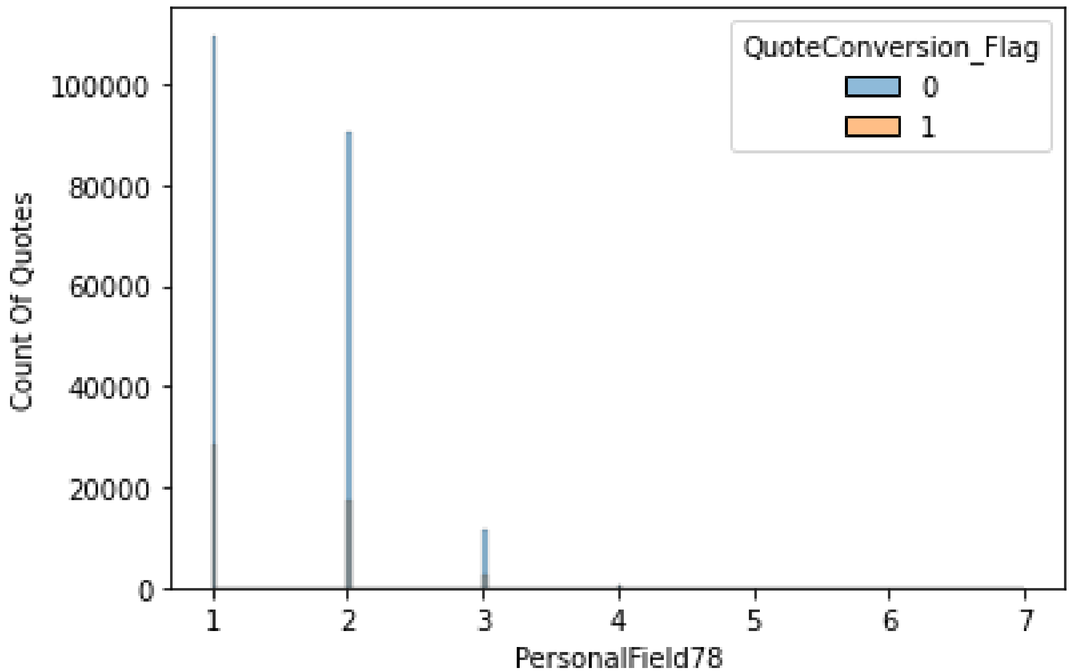

QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
1
2
3
4
5
6
7
PersonalField73

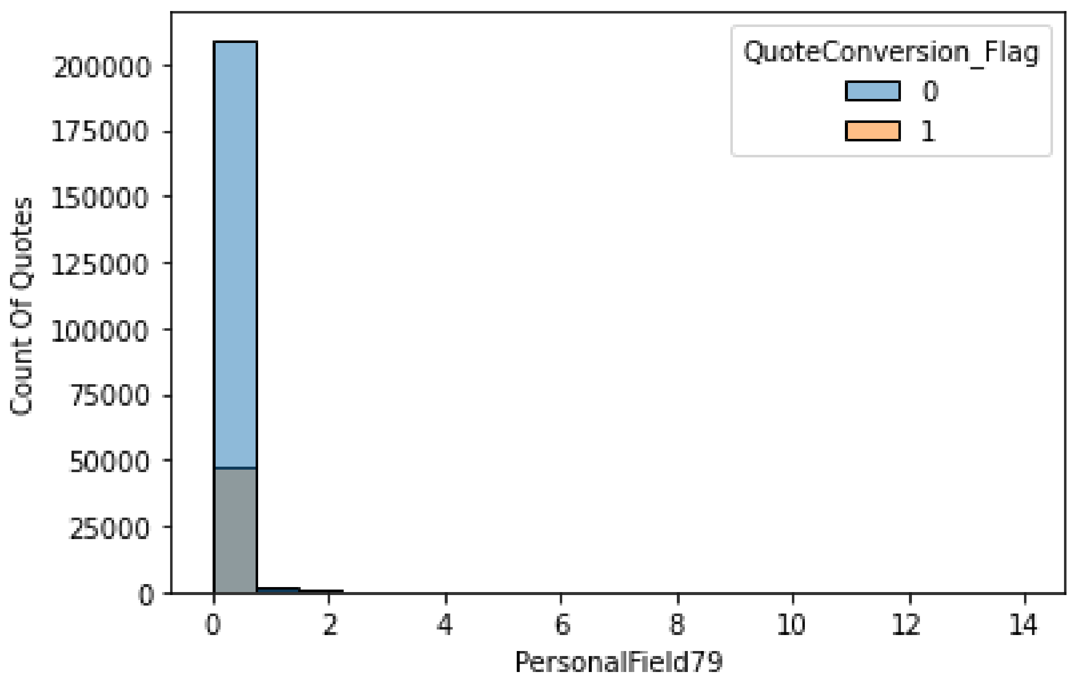

QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
0
5
10
15
20
PersonalField74

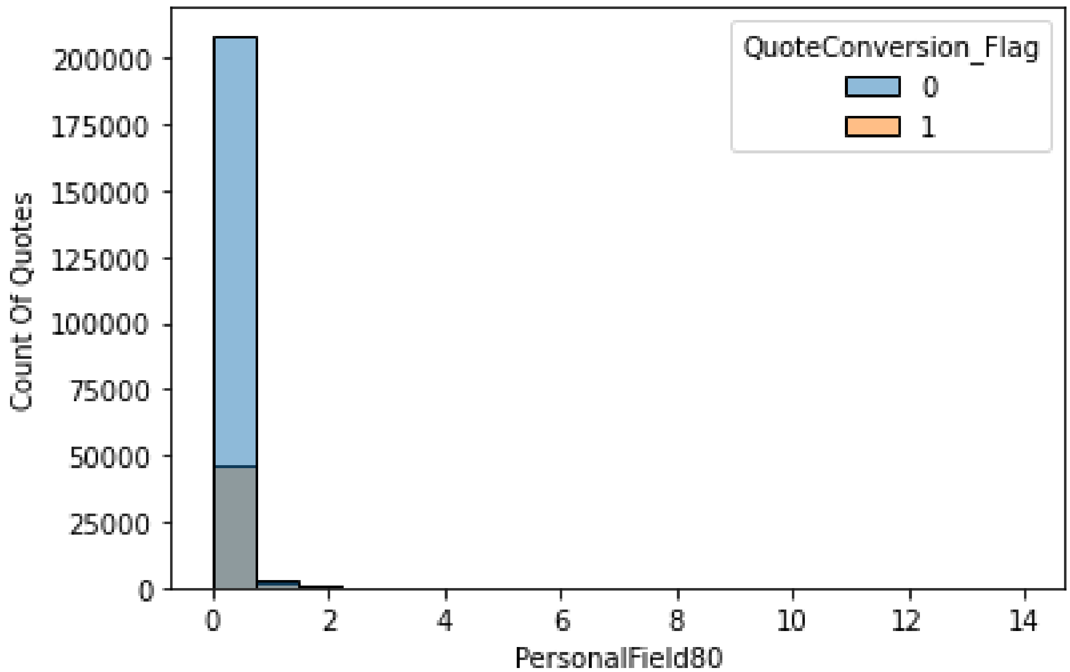

QuoteConversion_Flag
0
1
Count Of Quotes
140000
120000
100000
80000
60000
40000
20000
0
0
5
10
15
20
PersonalField75

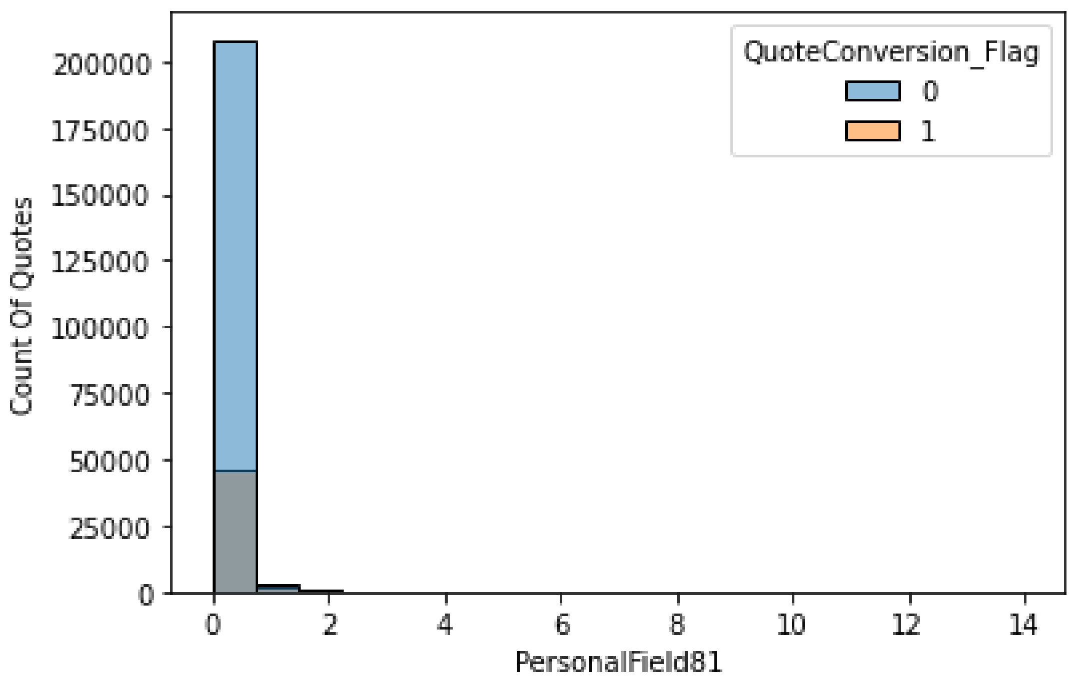

QuoteConversion_Flag
0
1
Count Of Quotes
0
20000
40000
60000
80000
100000
0
5
10
15
20
PersonalField76

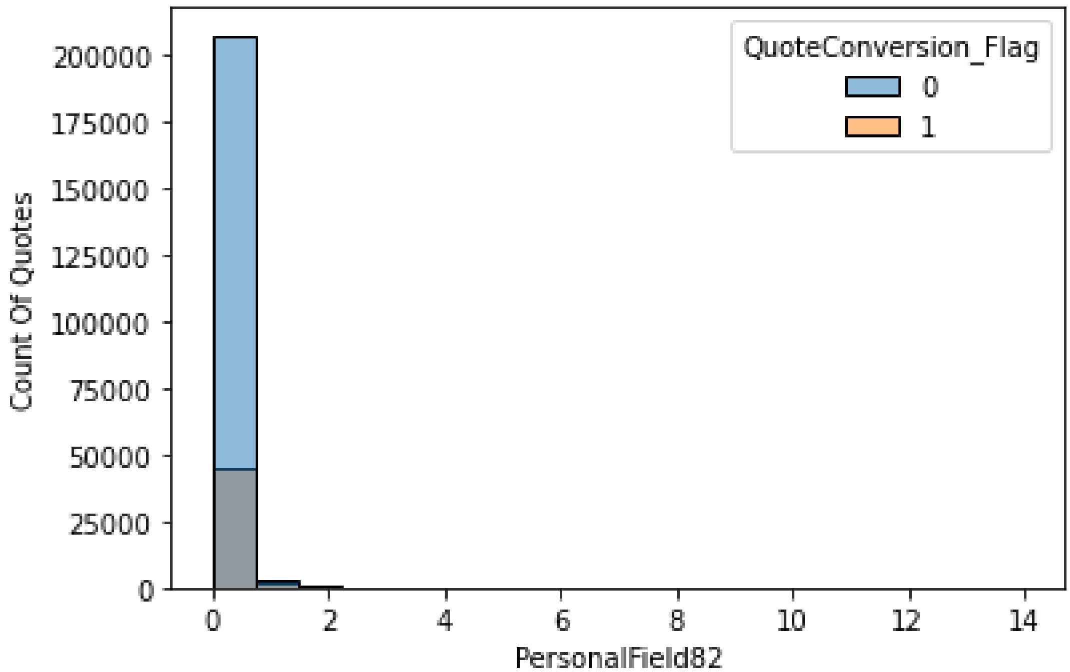

QuoteConversion_Flag
0
1
Count Of Quotes
0
20000
40000
60000
80000
100000
0
5
10
15
20
PersonalField77

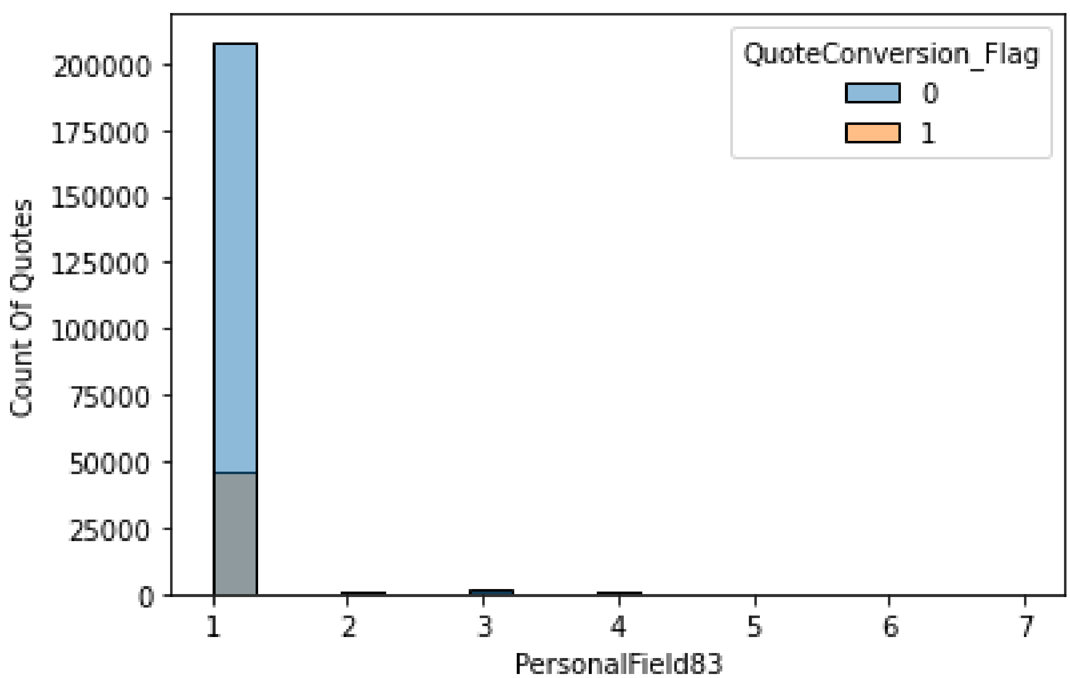

QuoteConversion_Flag
0
1
Count Of Quotes
PersonalField78
0
20000
40000
60000
80000
100000
1
2
3
4
5
6
7

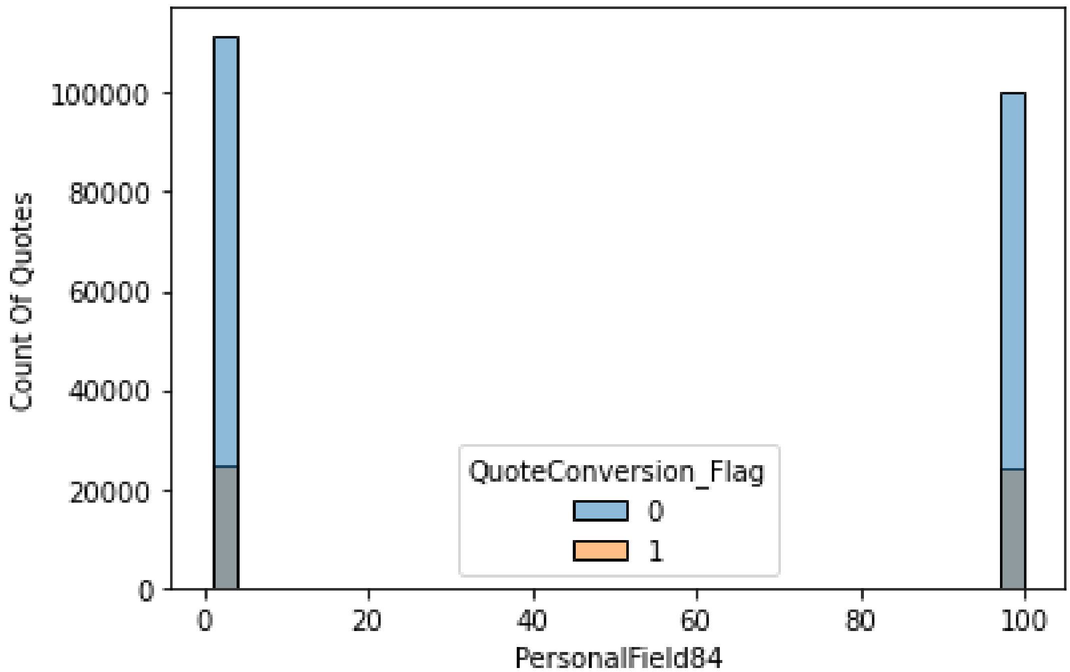

QuoteConversion_Flag
0
1
Count Of Quotes
PersonalField79
0
25000
50000
75000
100000
125000
150000
175000
200000
0
2
4
6
8
10
12
14

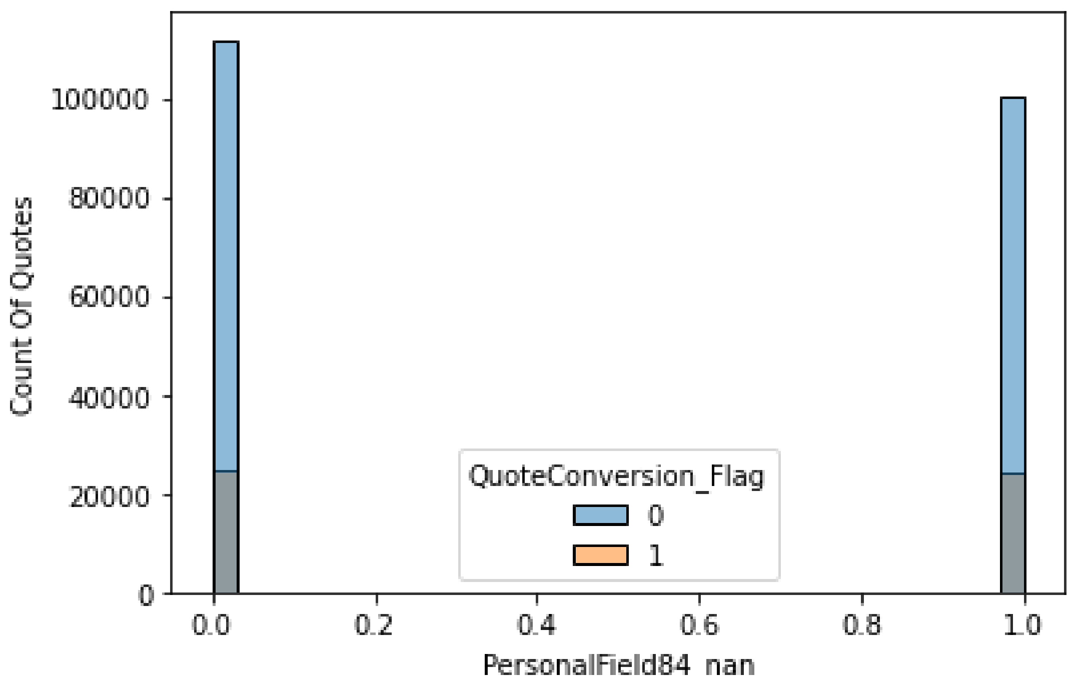

QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
0
2
4
6
8
10
12
14
PersonalField80


QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
0
2
4
6
8
10
12
14
PersonalField81


QuoteConversion_Flag
0
1
Count Of Quotes
PersonalField82

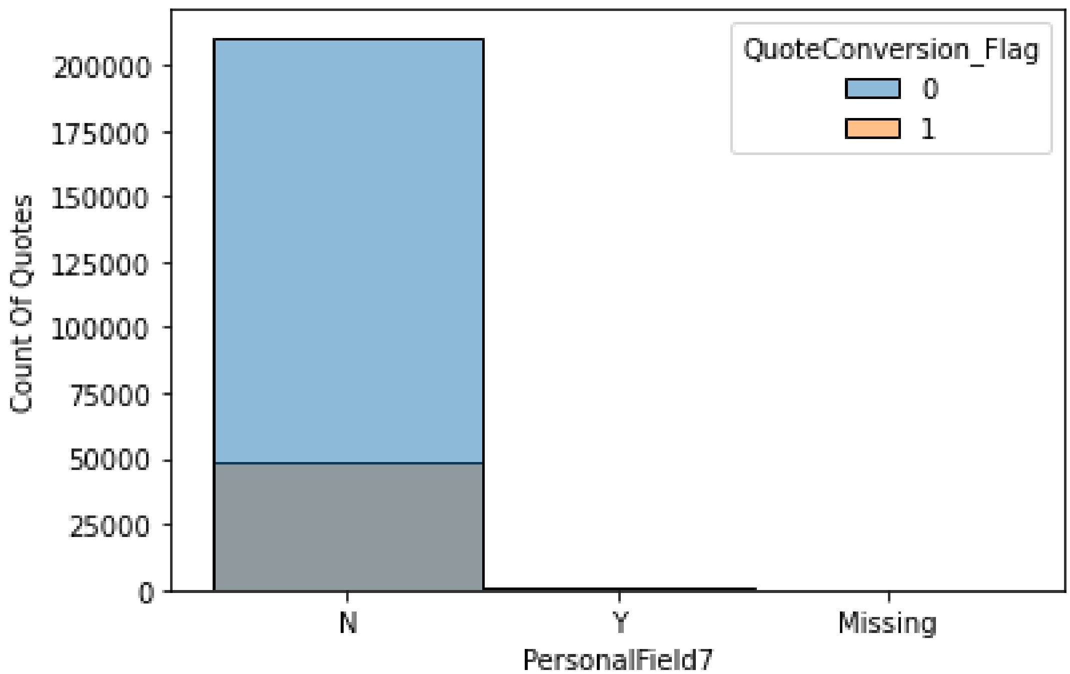

QuoteConversion_Flag
0
1
Count Of Quotes
PersonalField83

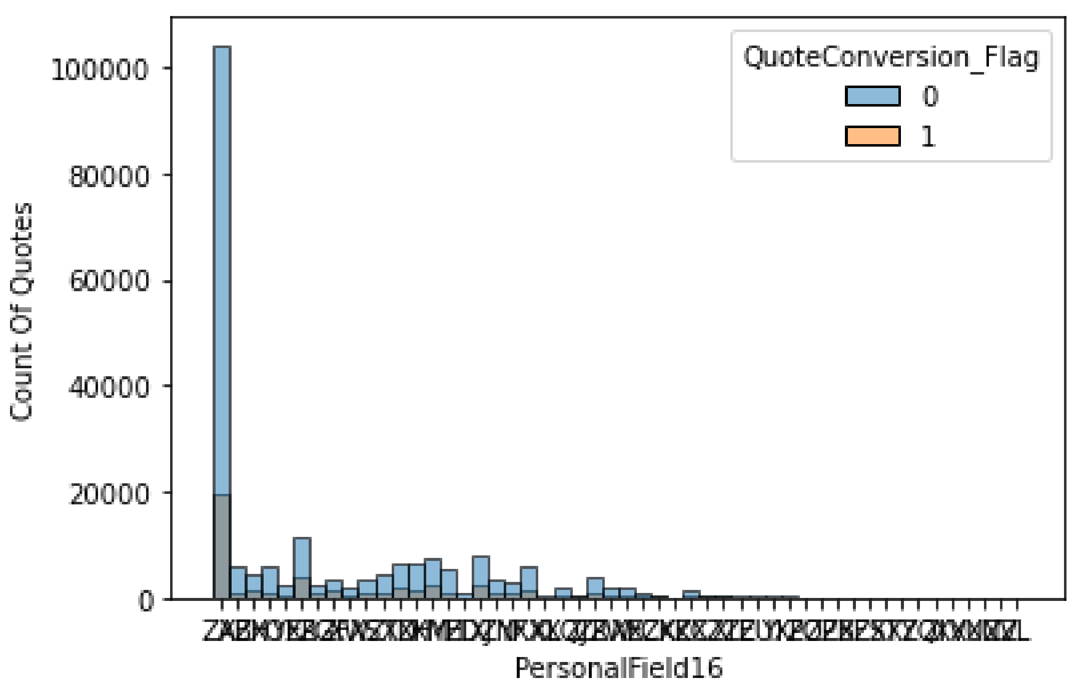


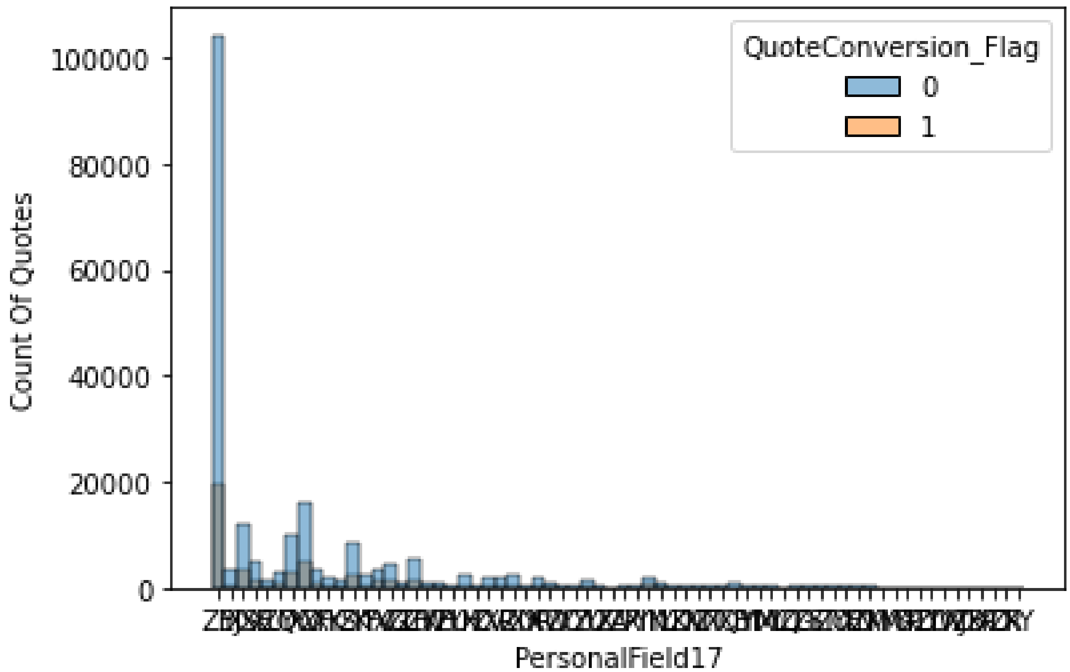


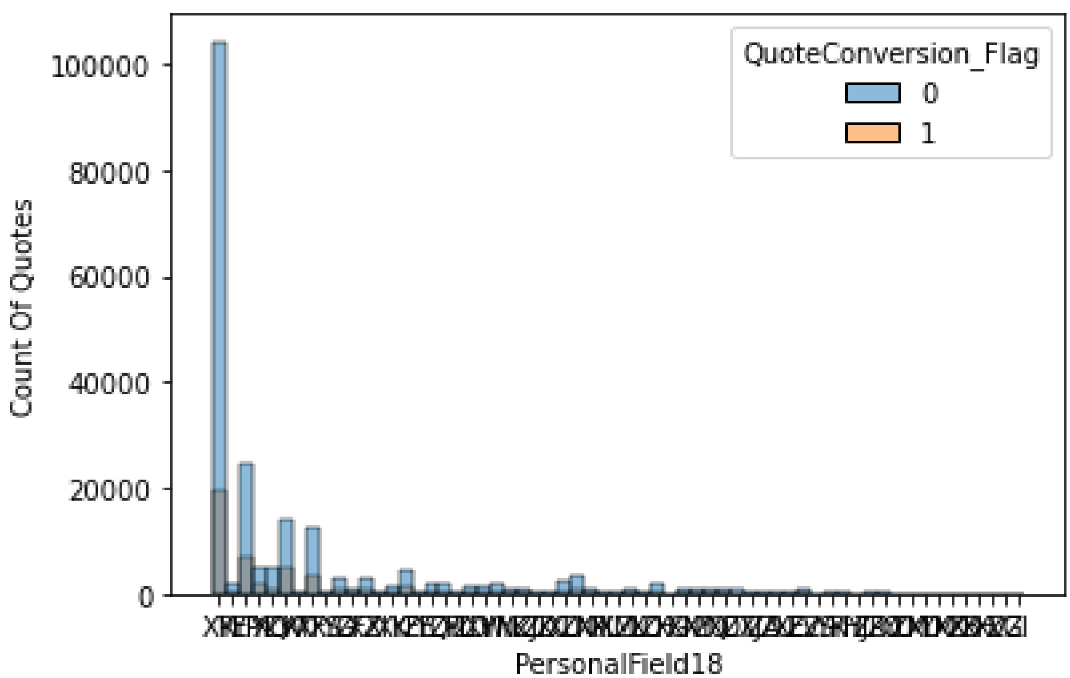
CATEGORICAL FEATURES

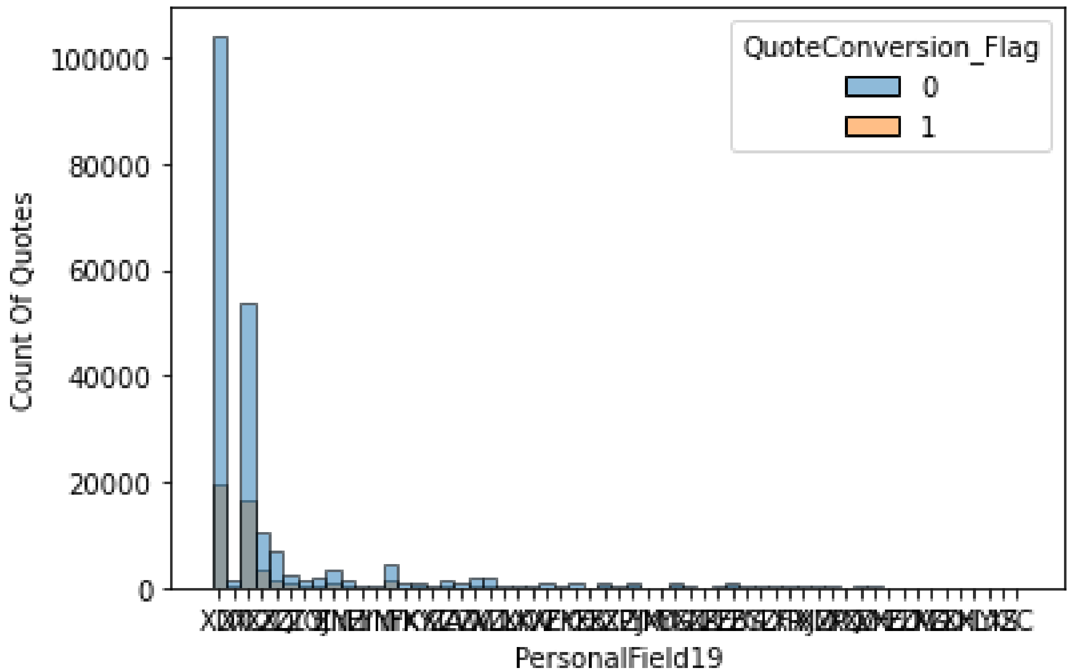
[41]:
```python
for feature in get_categorical_features(dataset):
        sns.histplot(data = dataset, x = feature, hue = 'QuoteConversion_Flag')
        plt.xlabel(feature)
        plt.ylabel('Count Of Quotes')
        plt.show()
```






QuoteConversion_Flag
0
1
Count Of Quotes
100000
80000
60000
40000
20000
0
PersonalField17


QuoteConversion_Flag
0
1
Count Of Quotes
100000
80000
60000
40000
20000
0
PersonalField18



PROPERTY FEATURES

[42]:
```
dataset = extract_feature_dataset('PropertyField',data)
```

[43]:
```
for feature in dataset.columns:
    if feature != 'QuoteConversion_Flag':
        print('{} has {} unique values'.format(feature,len(dataset[feature].
 ↪unique())))
```

```
PropertyField1A has 26 unique values
PropertyField1B has 26 unique values
PropertyField2A has 2 unique values
PropertyField2B has 21 unique values
PropertyField3 has 3 unique values
PropertyField4 has 3 unique values
PropertyField5 has 2 unique values
PropertyField6 has 1 unique values
PropertyField7 has 19 unique values
PropertyField8 has 2 unique values
PropertyField9 has 3 unique values
PropertyField10 has 5 unique values
PropertyField11A has 2 unique values
PropertyField11B has 5 unique values
PropertyField12 has 7 unique values
PropertyField13 has 4 unique values
```

```
PropertyField14 has 4 unique values
PropertyField15 has 15 unique values
PropertyField16A has 26 unique values
PropertyField16B has 26 unique values
PropertyField17 has 8 unique values
PropertyField18 has 10 unique values
PropertyField19 has 10 unique values
PropertyField20 has 3 unique values
PropertyField21A has 26 unique values
PropertyField21B has 26 unique values
PropertyField22 has 5 unique values
PropertyField23 has 13 unique values
PropertyField24A has 26 unique values
PropertyField24B has 26 unique values
PropertyField25 has 11 unique values
PropertyField26A has 26 unique values
PropertyField26B has 26 unique values
PropertyField27 has 17 unique values
PropertyField28 has 4 unique values
PropertyField29 has 3 unique values
PropertyField30 has 2 unique values
PropertyField31 has 4 unique values
PropertyField32 has 3 unique values
PropertyField33 has 4 unique values
PropertyField34 has 3 unique values
PropertyField35 has 3 unique values
PropertyField36 has 3 unique values
PropertyField37 has 2 unique values
PropertyField38 has 3 unique values
PropertyField39A has 26 unique values
PropertyField39B has 26 unique values
PropertyField29_nan has 2 unique values
```

[44]:
```
numerical_features = get_numerical_features(dataset)
categorical_features = get_categorical_features(dataset)
```

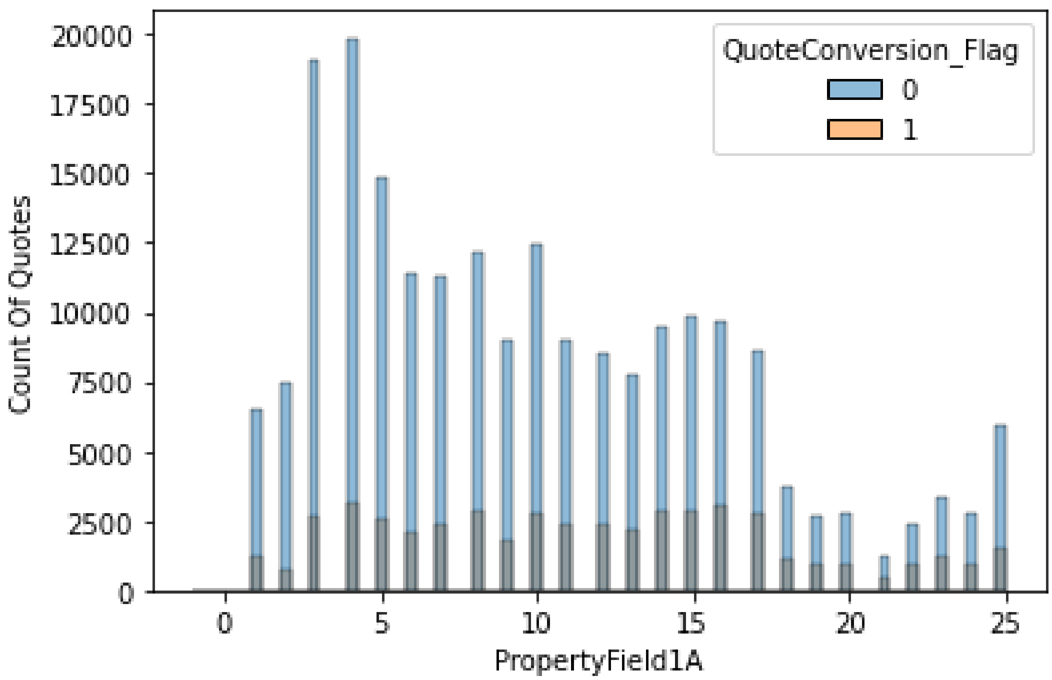
[45]:
```
print('Numerical Features : ', numerical_features)
print('Categorical Features : ', categorical_features)
```

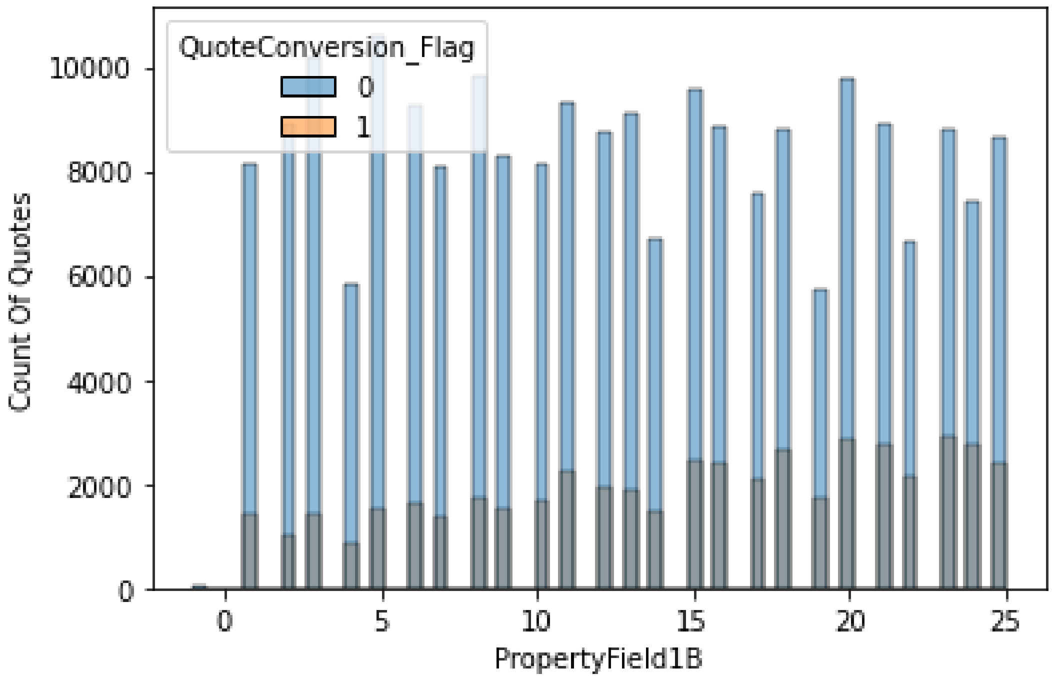
```
Numerical Features :  ['PropertyField1A', 'PropertyField1B', 'PropertyField2A',
'PropertyField2B', 'PropertyField6', 'PropertyField8', 'PropertyField9',
'PropertyField10', 'PropertyField11A', 'PropertyField11B', 'PropertyField12',
'PropertyField13', 'PropertyField15', 'PropertyField16A', 'PropertyField16B',
'PropertyField17', 'PropertyField18', 'PropertyField19', 'PropertyField20',
'PropertyField21A', 'PropertyField21B', 'PropertyField22', 'PropertyField23',
'PropertyField24A', 'PropertyField24B', 'PropertyField25', 'PropertyField26A',
'PropertyField26B', 'PropertyField27', 'PropertyField29', 'PropertyField35',
'PropertyField39A', 'PropertyField39B', 'PropertyField29_nan',
```

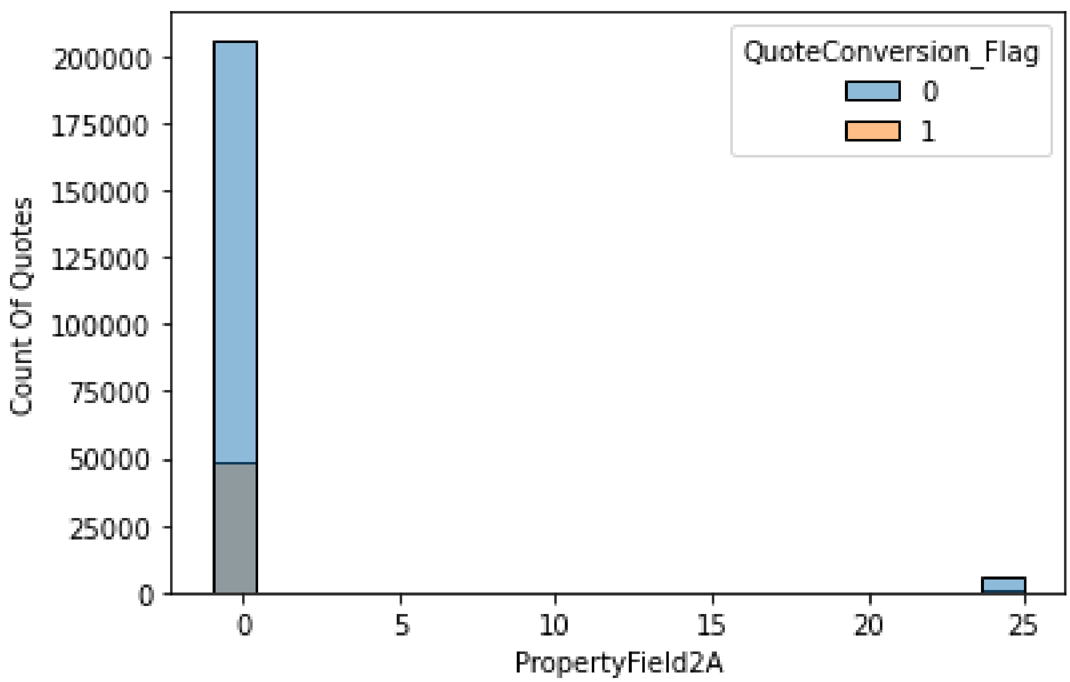
```
'QuoteConversion_Flag']
Categorical Features :  ['PropertyField3', 'PropertyField4', 'PropertyField5',
'PropertyField7', 'PropertyField14', 'PropertyField28', 'PropertyField30',
'PropertyField31', 'PropertyField32', 'PropertyField33', 'PropertyField34',
'PropertyField36', 'PropertyField37', 'PropertyField38']
```

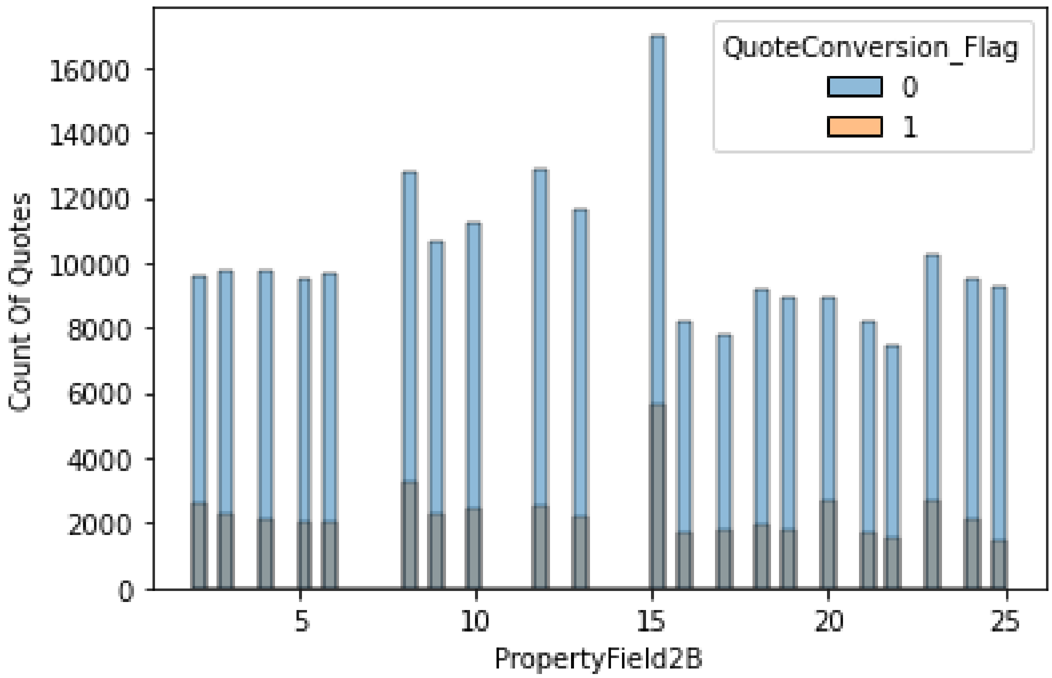
DISCRETE NUMERICAL FEATURES

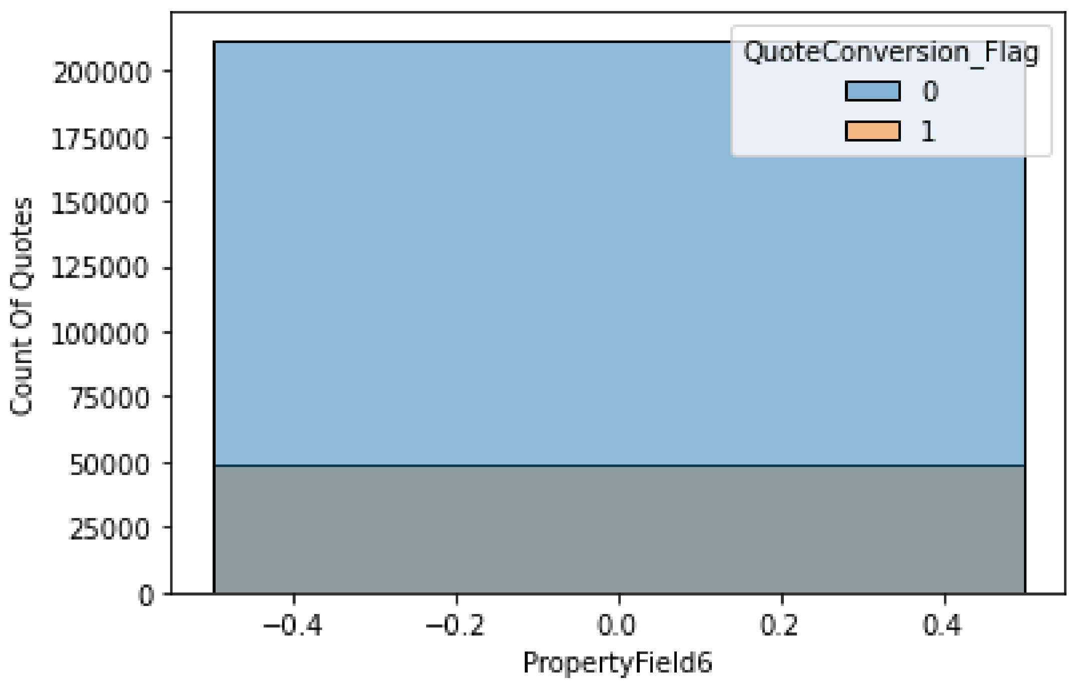
[46]:
```
for feature in get_numerical_features(dataset):
    if feature != 'QuoteConversion_Flag' and len(dataset[feature].unique()) <
 ↪30:
        sns.histplot(data = dataset, x = feature, hue = 'QuoteConversion_Flag')
        plt.xlabel(feature)
        plt.ylabel('Count Of Quotes')
        plt.show()
```

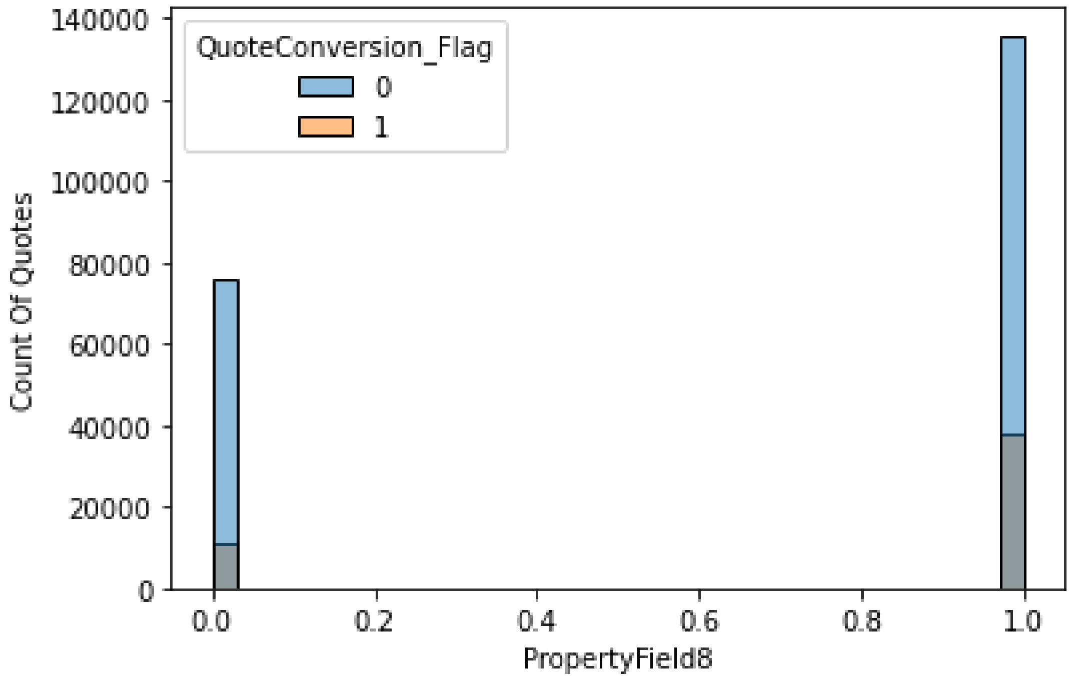


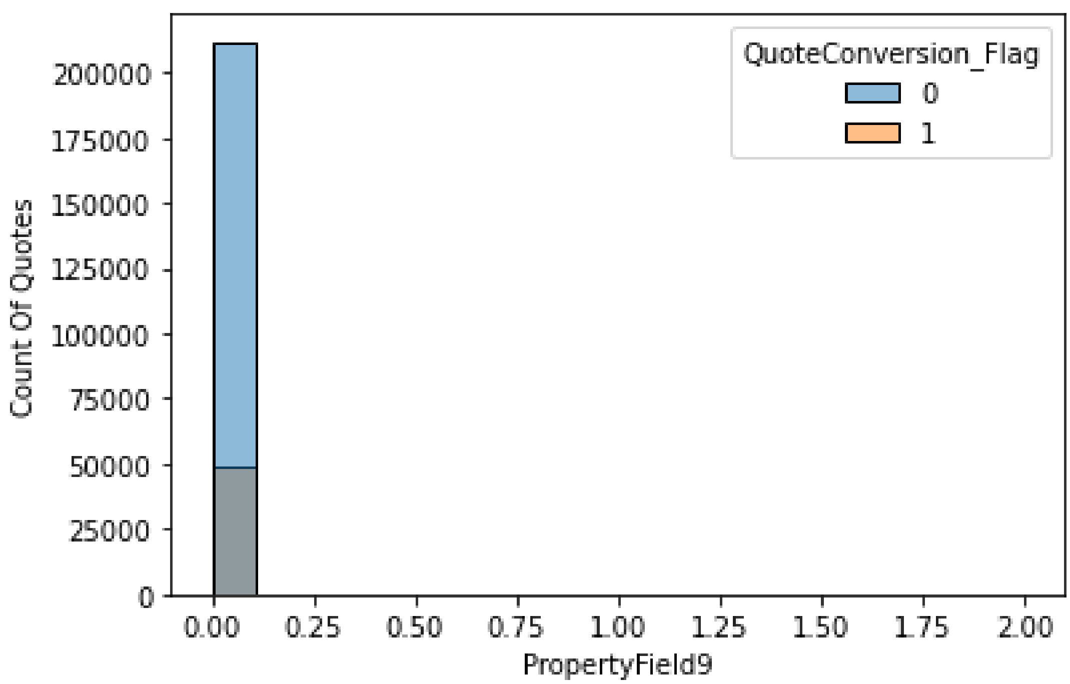

QuoteConversion_Flag
0
1
10000
8000
6000
4000
2000
0
Count Of Quotes
0
5
10
15
20
25
PropertyField1B

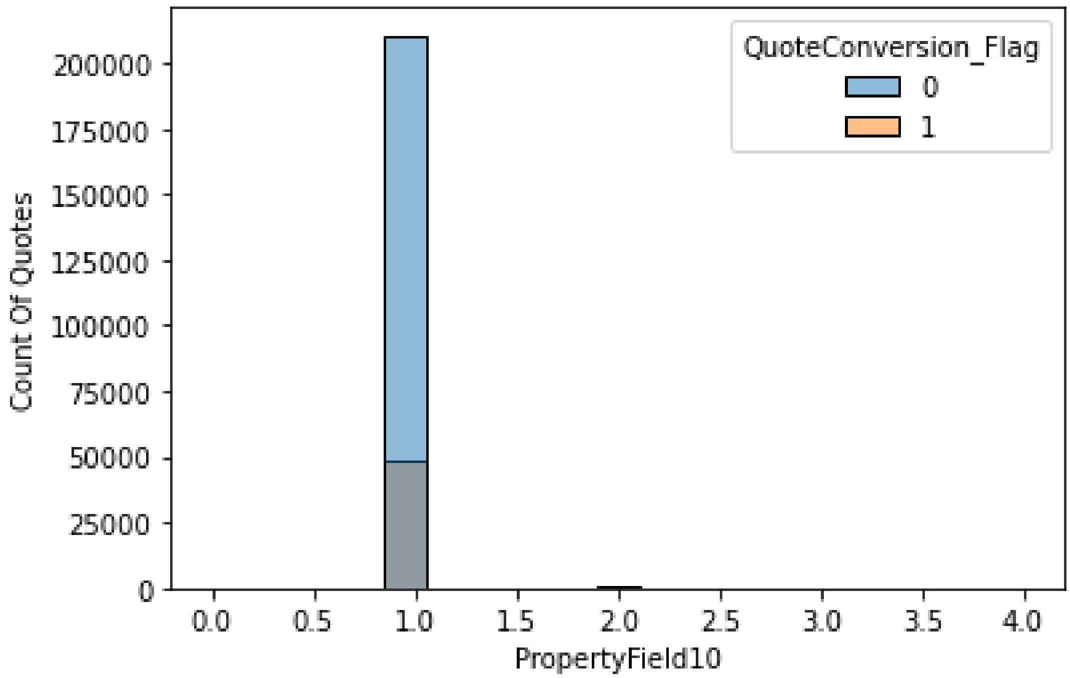

QuoteConversion_Flag
0
1
200000
175000
150000
125000
100000
75000
50000
25000
0
Count Of Quotes
0
5
10
15
20
25
PropertyField2A

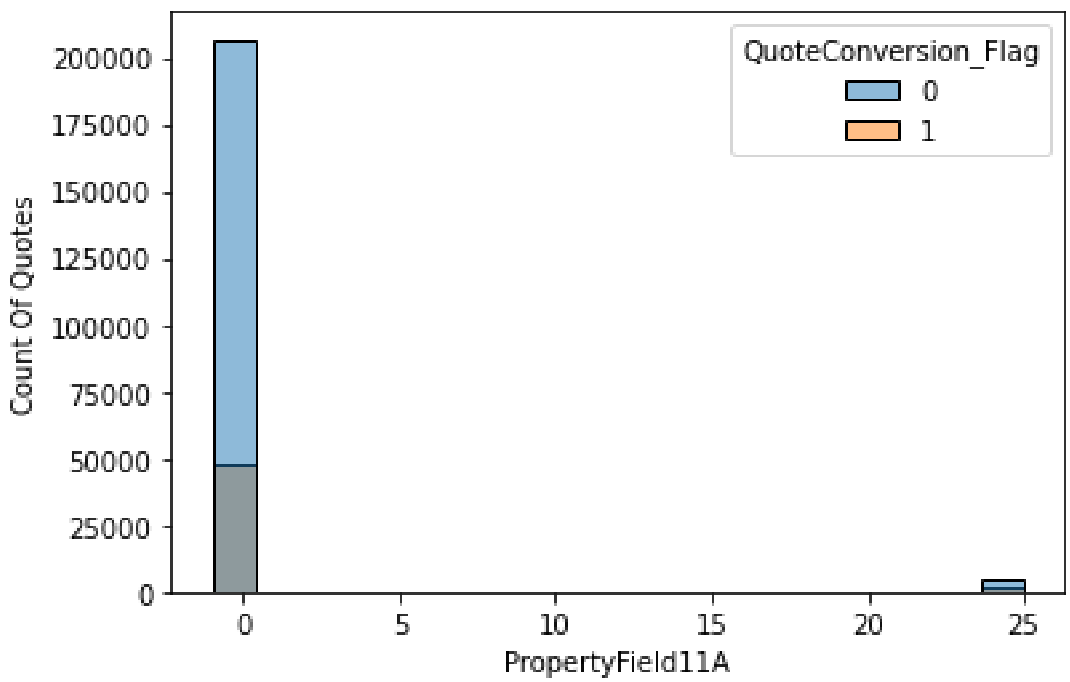

QuoteConversion_Flag
0
1
Count Of Quotes
0
2000
4000
6000
8000
10000
12000
14000
16000
5
10
15
20
25
PropertyField2B

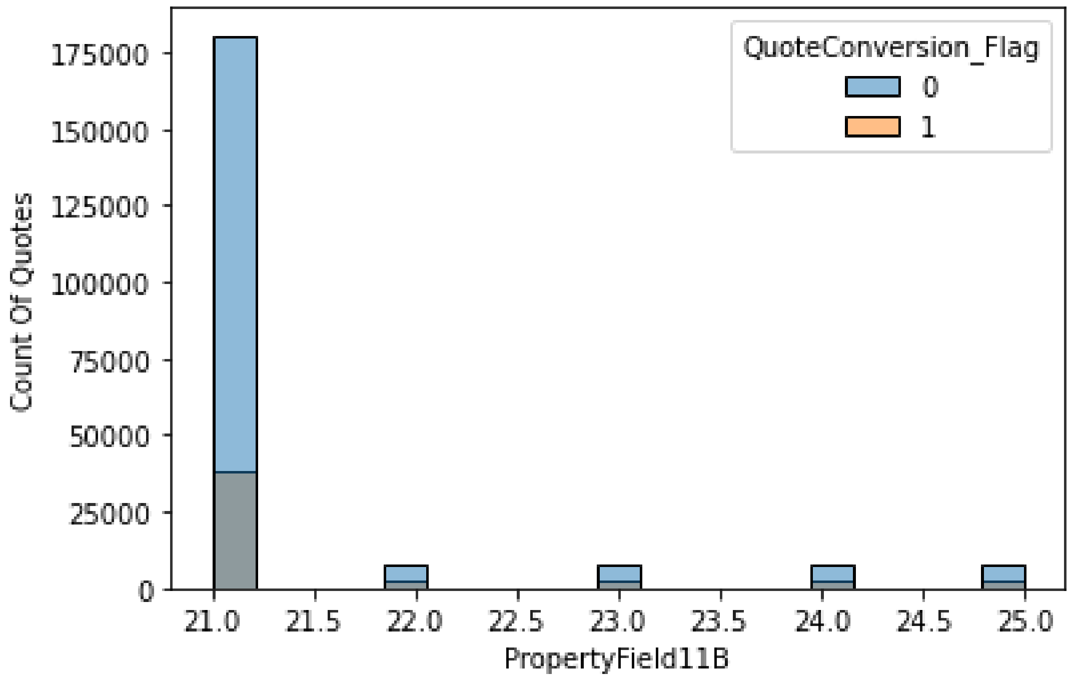

QuoteConversion_Flag
0
1
Count Of Quotes
0
25000
50000
75000
100000
125000
150000
175000
200000
−0.4
−0.2
0.0
0.2
0.4
PropertyField6


QuoteConversion_Flag
0
1
Count Of Quotes
PropertyField8
0
20000
40000
60000
80000
100000
120000
140000
0.0
0.2
0.4
0.6
0.8
1.0

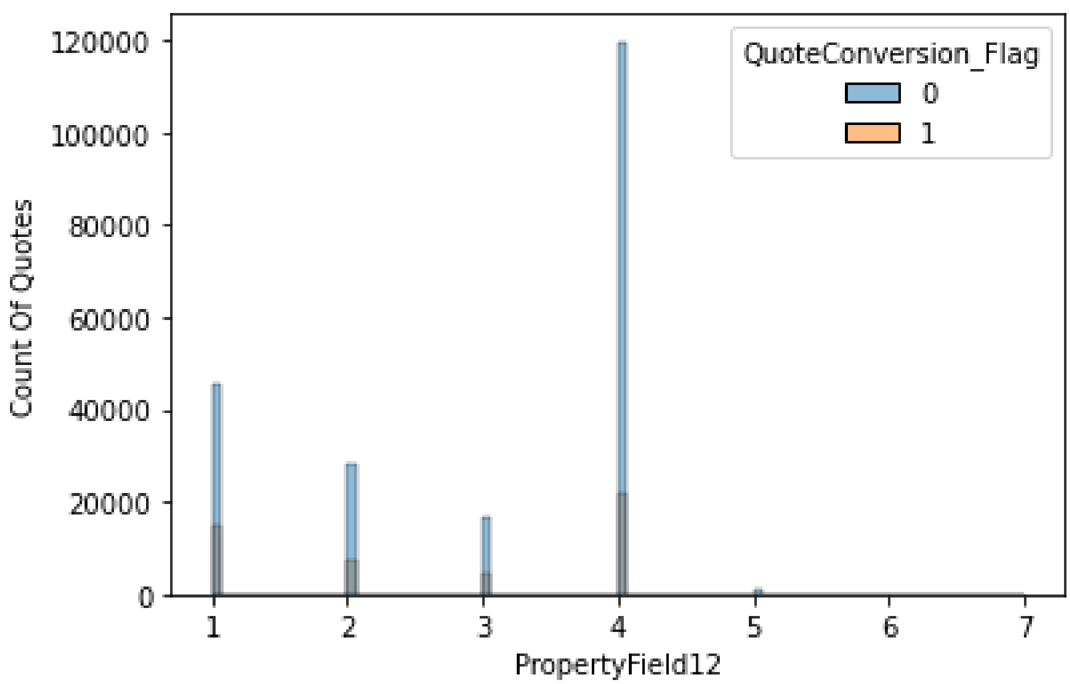

QuoteConversion_Flag
0
1
Count Of Quotes
PropertyField9
0
25000
50000
75000
100000
125000
150000
175000
200000
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00


QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
PropertyField10

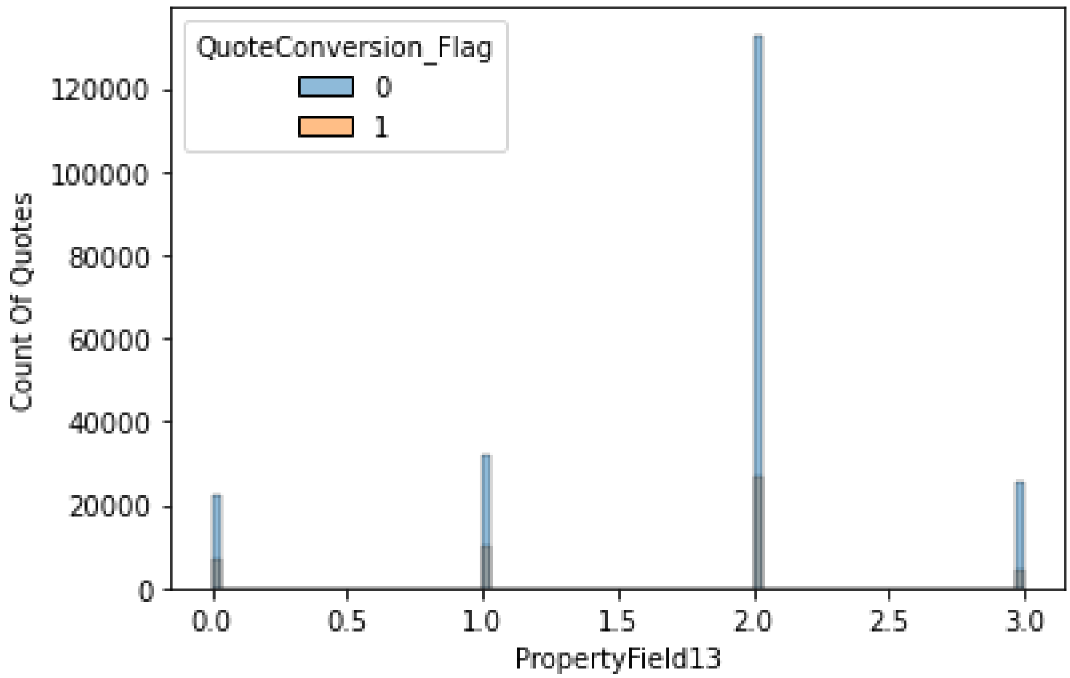

QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
0
5
10
15
20
25
PropertyField11A

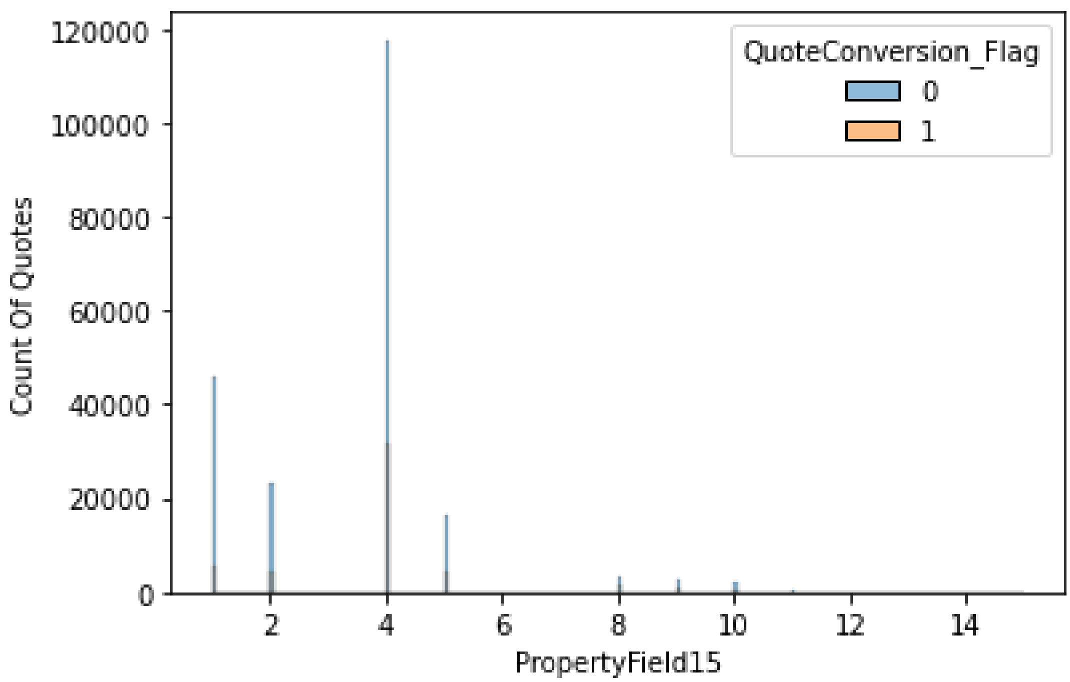


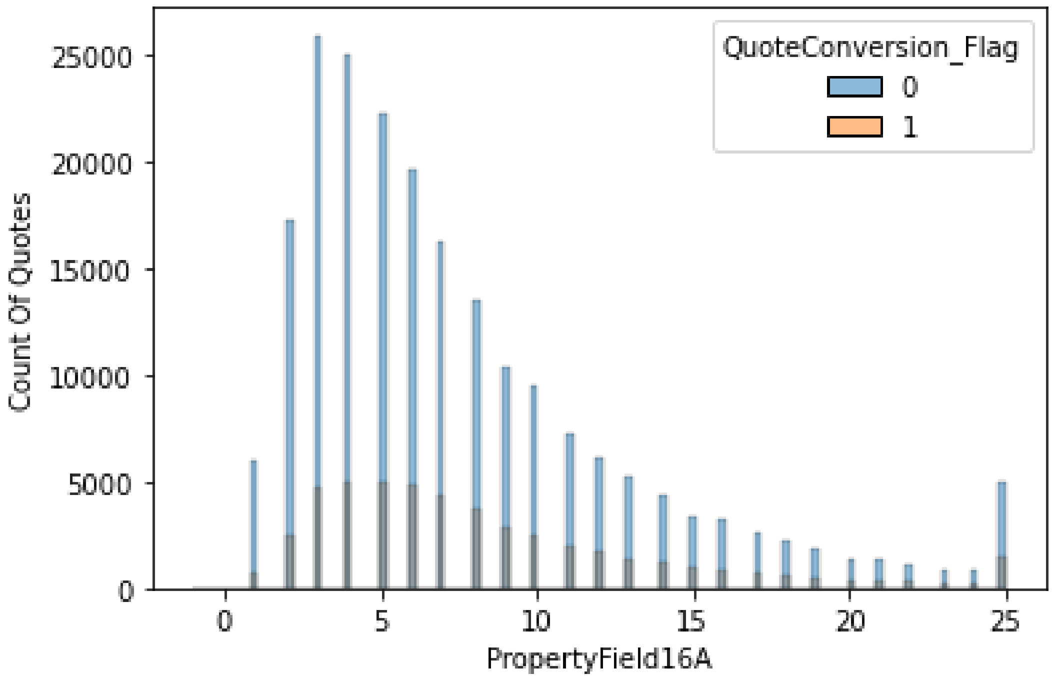
QuoteConversion_Flag
0
1
Count Of Quotes
175000
150000
125000
100000
75000
50000
25000
0
21.0
21.5
22.0
22.5
23.0
23.5
24.0
24.5
25.0
PropertyField11B

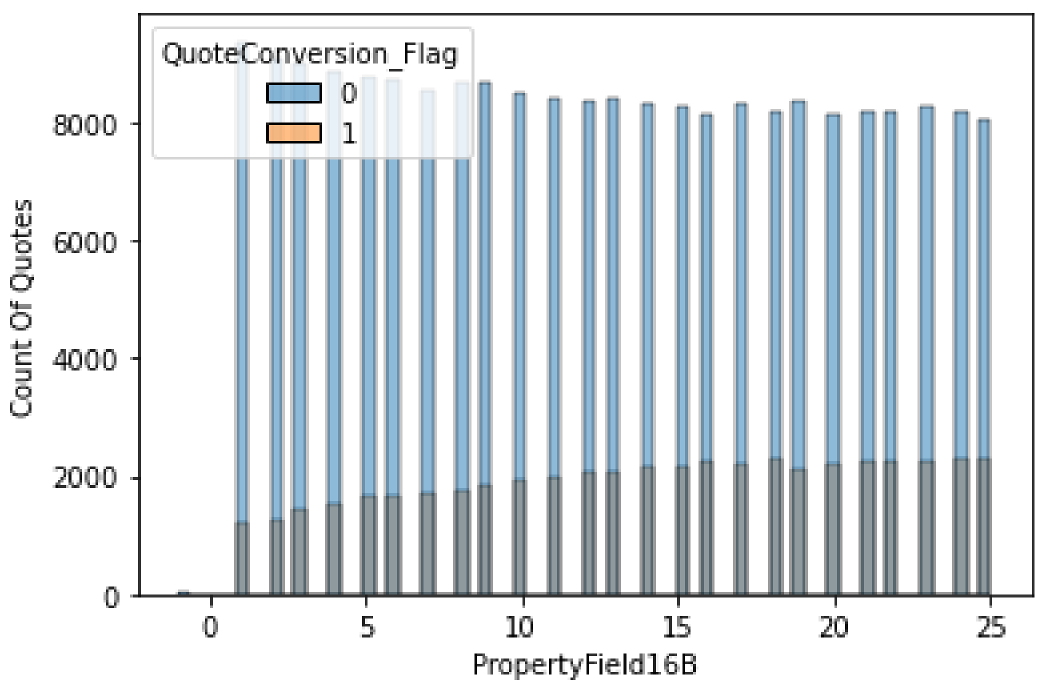


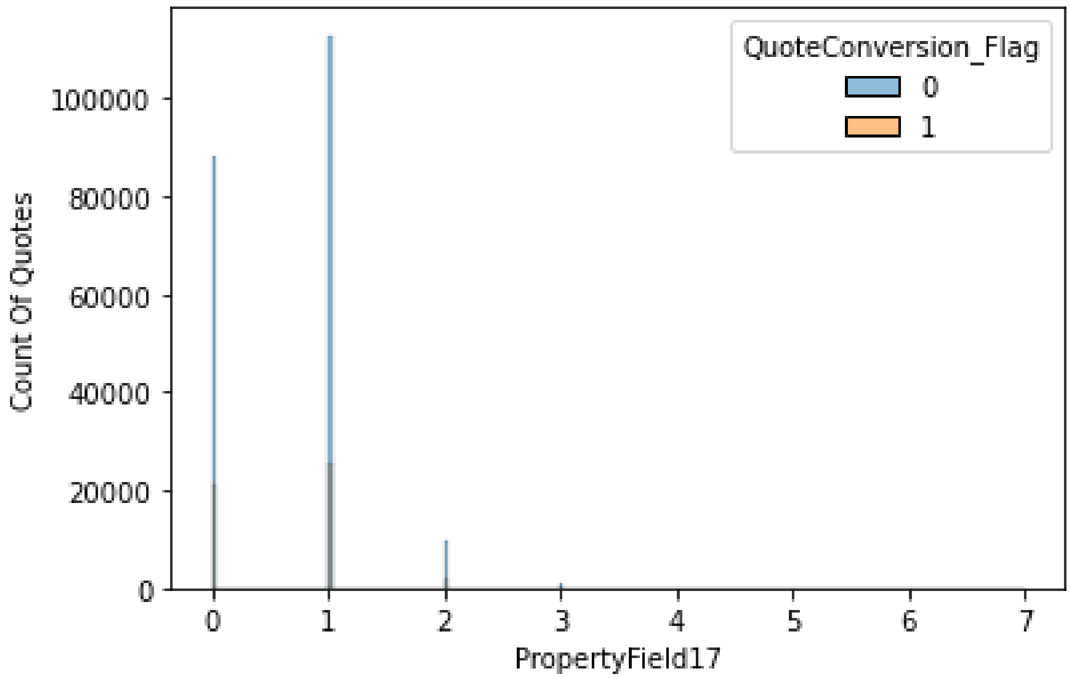
QuoteConversion_Flag
0
1
Count Of Quotes
120000
100000
80000
60000
40000
20000
0
1
2
3
4
5
6
7
PropertyField12

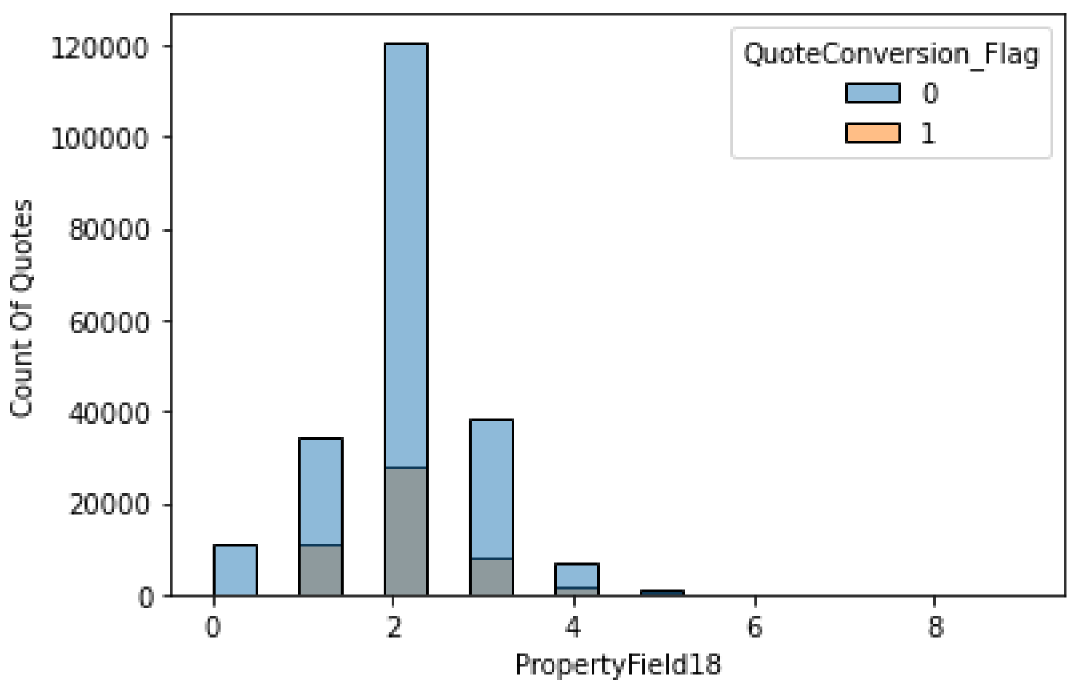

QuoteConversion_Flag
0
1
Count Of Quotes
PropertyField13
0
20000
40000
60000
80000
100000
120000
0.0
0.5
1.0
1.5
2.0
2.5
3.0

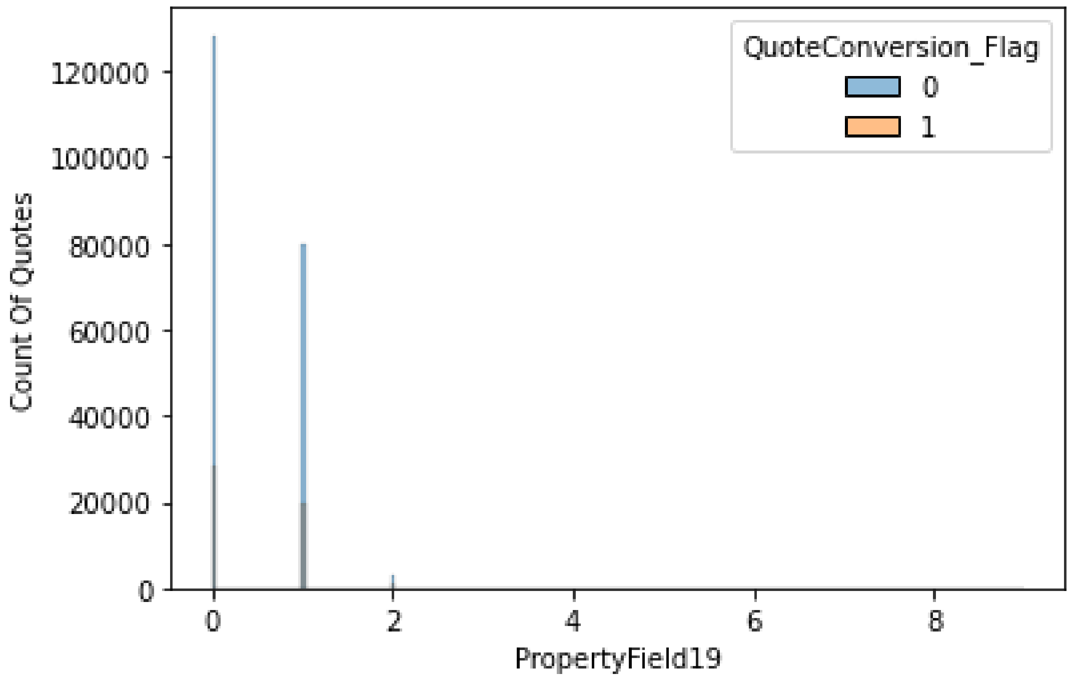

QuoteConversion_Flag
0
1
Count Of Quotes
PropertyField15
0
20000
40000
60000
80000
100000
120000
2
4
6
8
10
12
14

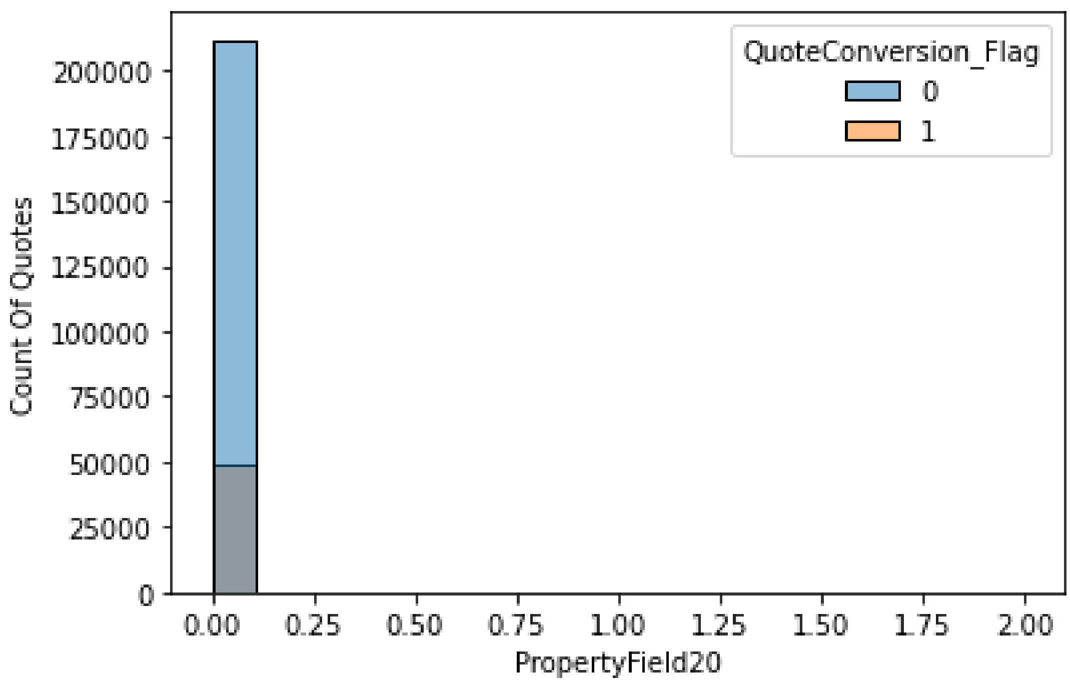

QuoteConversion_Flag
0
1
Count Of Quotes
25000
20000
15000
10000
5000
0
0
5
10
15
20
25
PropertyField16A

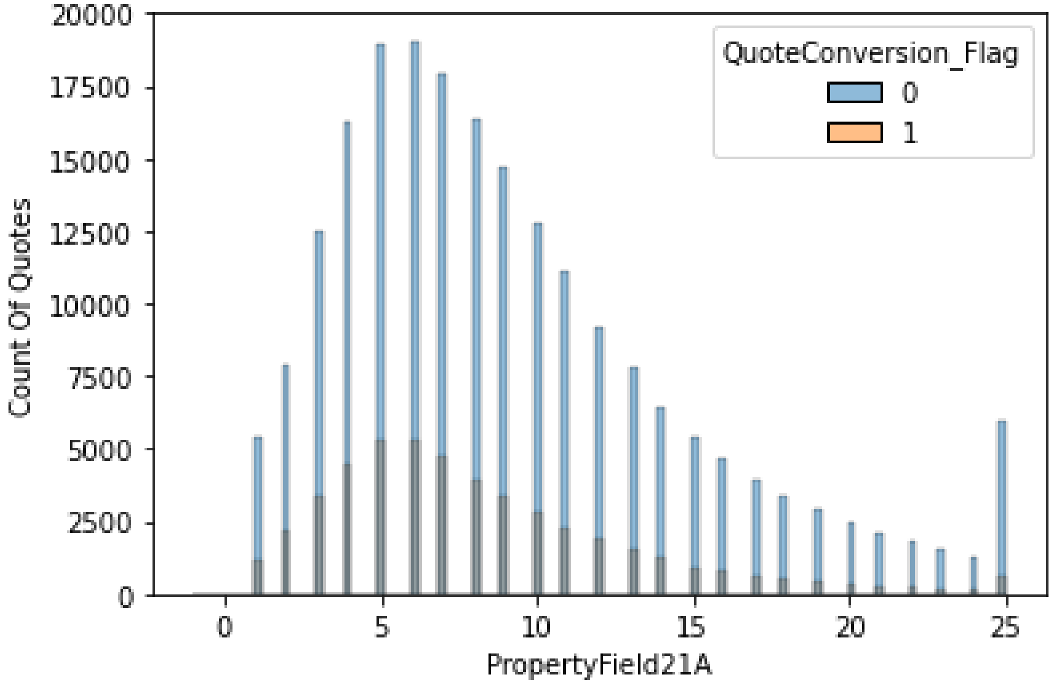

QuoteConversion_Flag
0
1
Count Of Quotes
8000
6000
4000
2000
0
0
5
10
15
20
25
PropertyField16B

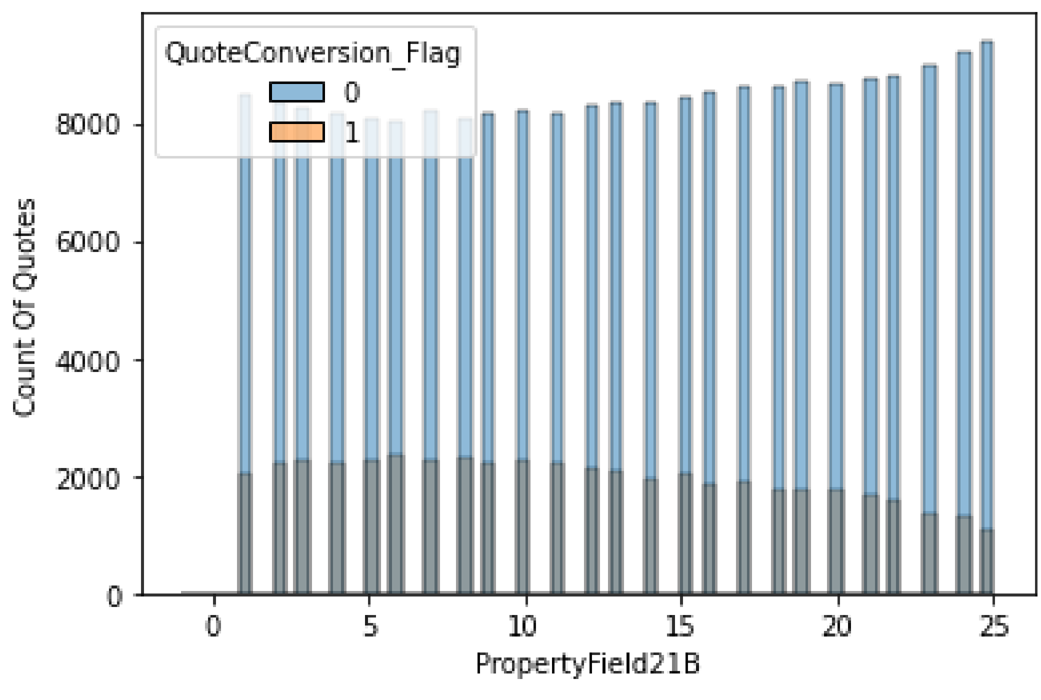

QuoteConversion_Flag
0
1
Count Of Quotes
100000
80000
60000
40000
20000
0
0
1
2
3
4
5
6
7
PropertyField17

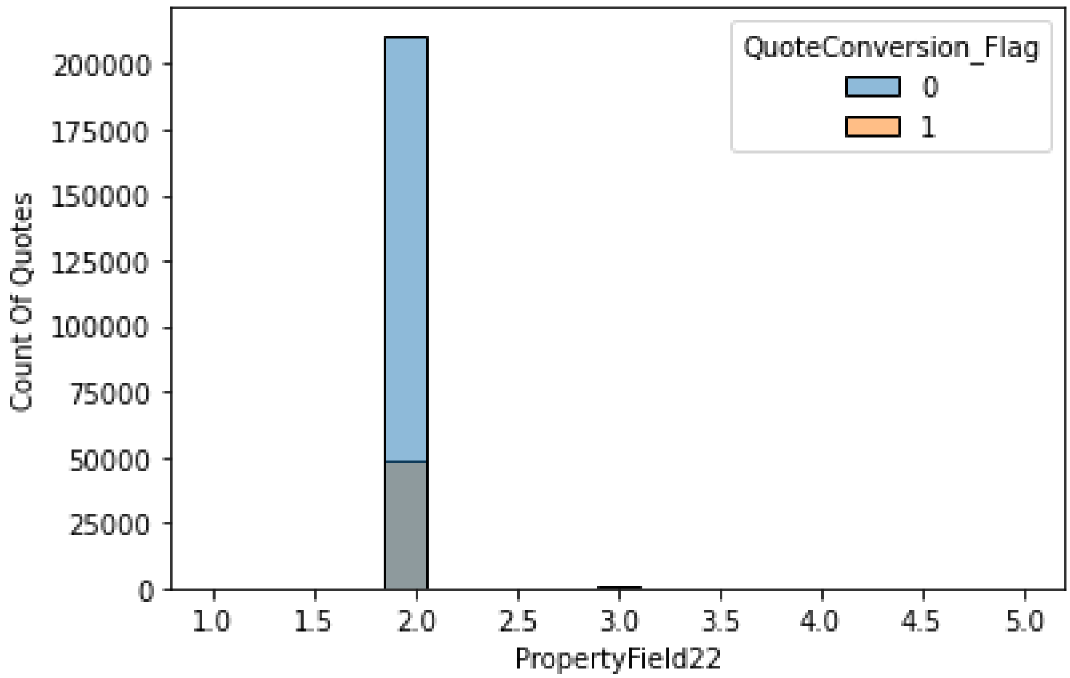

QuoteConversion_Flag
0
1
Count Of Quotes
120000
100000
80000
60000
40000
20000
0
0
2
4
6
8
PropertyField18

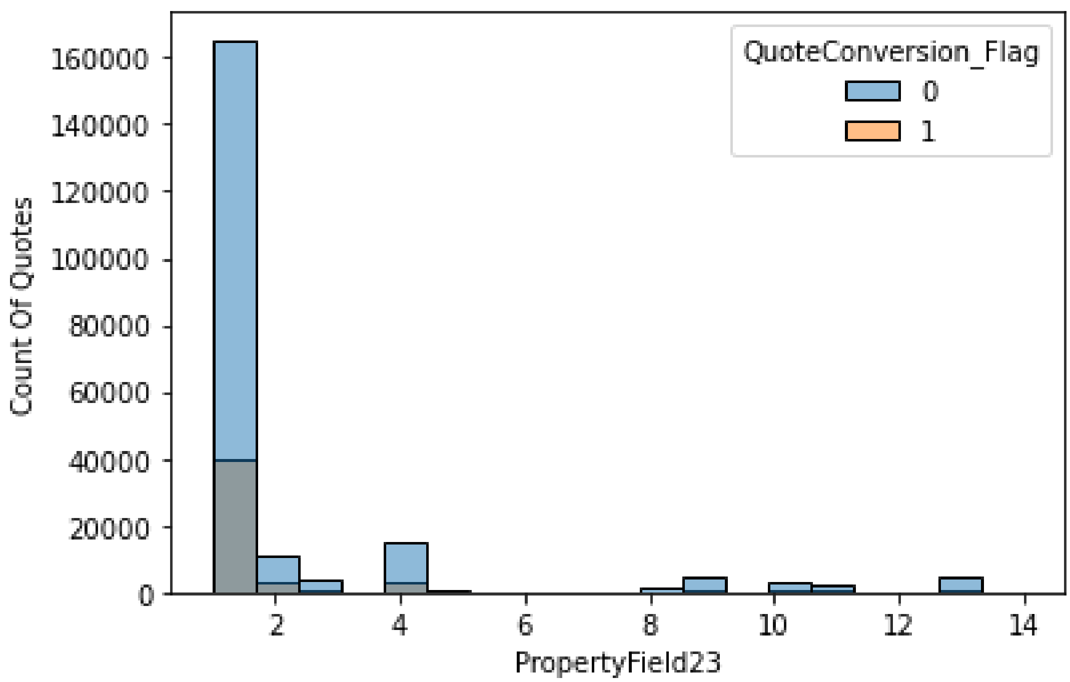

QuoteConversion_Flag
0
1
Count Of Quotes
120000
100000
80000
60000
40000
20000
0
0
2
4
6
8
PropertyField19

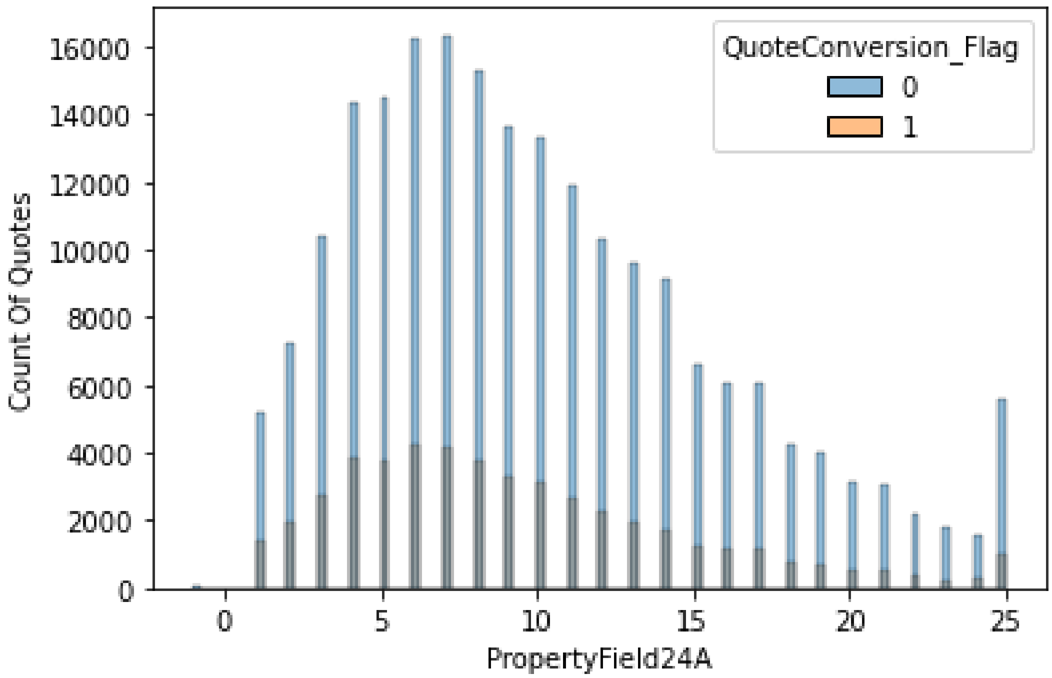

QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
PropertyField20

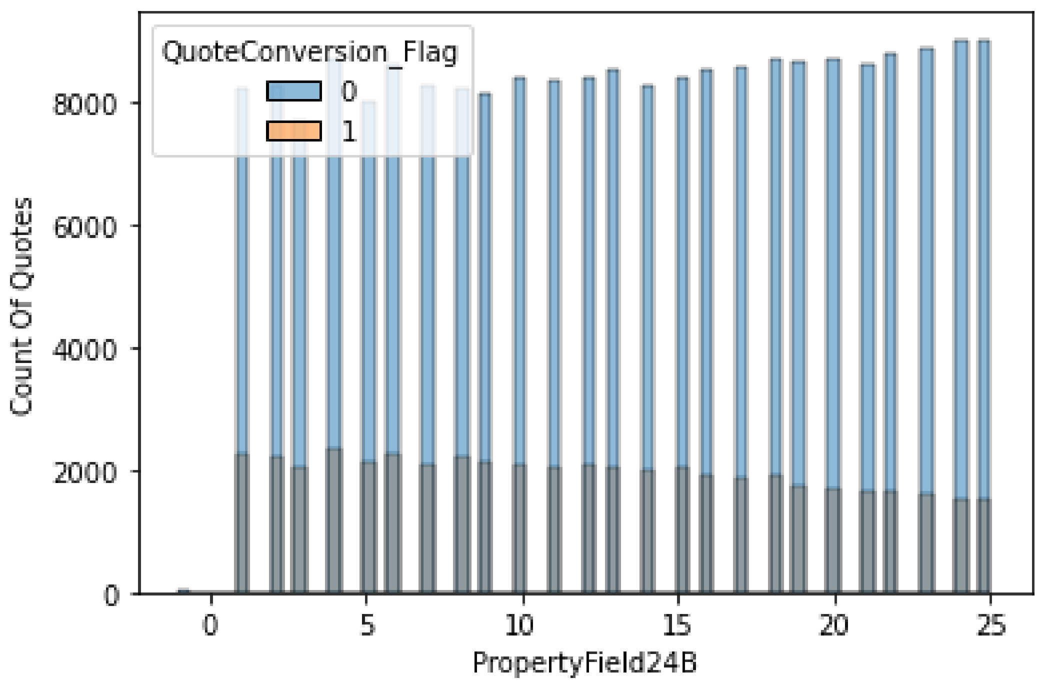

QuoteConversion_Flag
0
1
Count Of Quotes
PropertyField21A
20000
17500
15000
12500
10000
7500
5000
2500
0
0
5
10
15
20
25

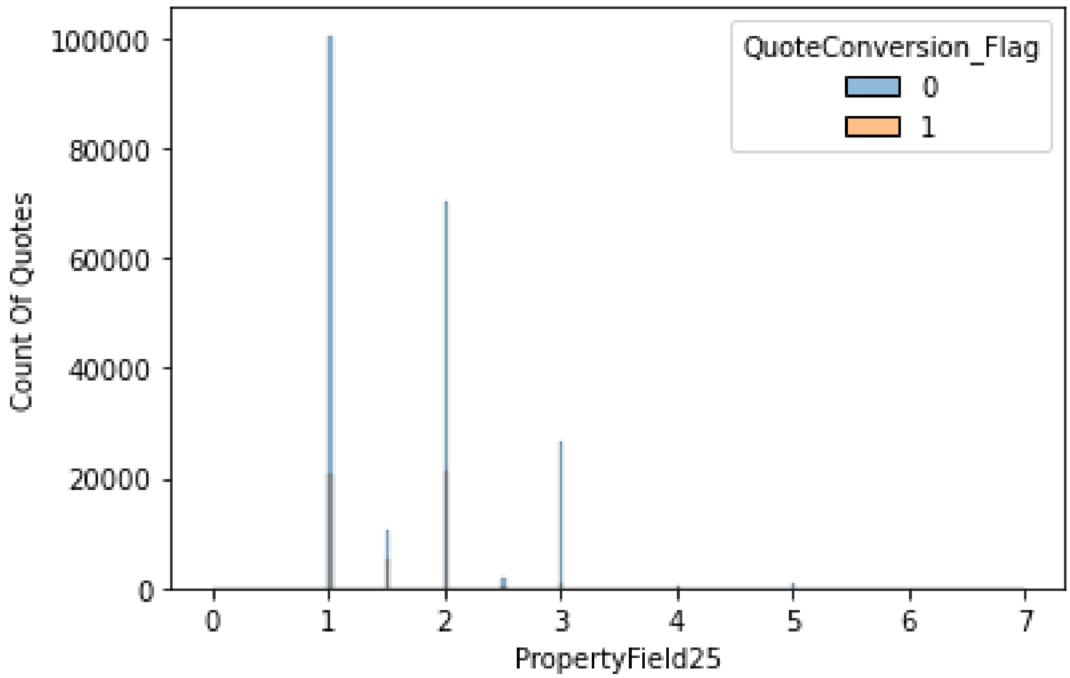

QuoteConversion_Flag
0
1
Count Of Quotes
PropertyField21B
8000
6000
4000
2000
0
0
5
10
15
20
25

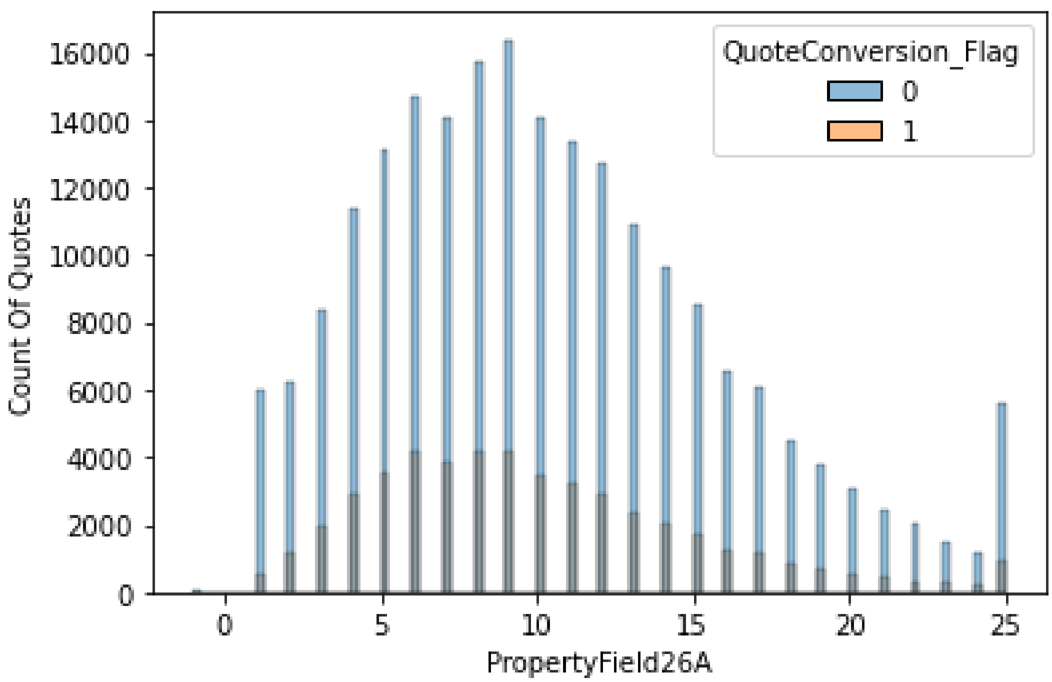

QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
PropertyField22

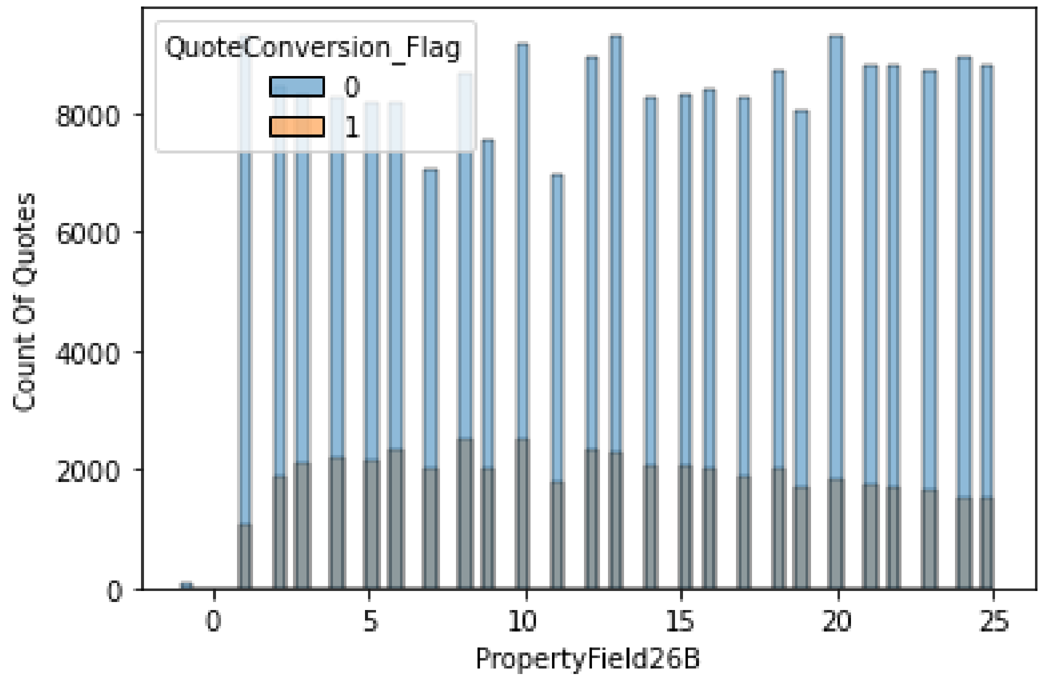

QuoteConversion_Flag
0
1
Count Of Quotes
160000
140000
120000
100000
80000
60000
40000
20000
0
2
4
6
8
10
12
14
PropertyField23

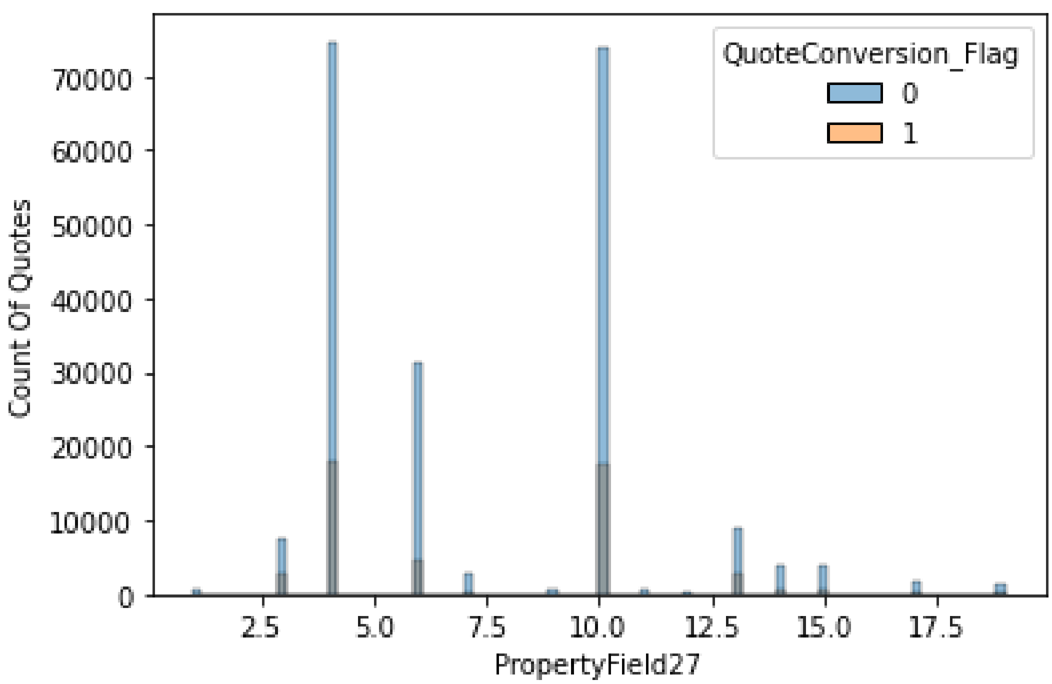

QuoteConversion_Flag
0
1
Count Of Quotes
PropertyField24A

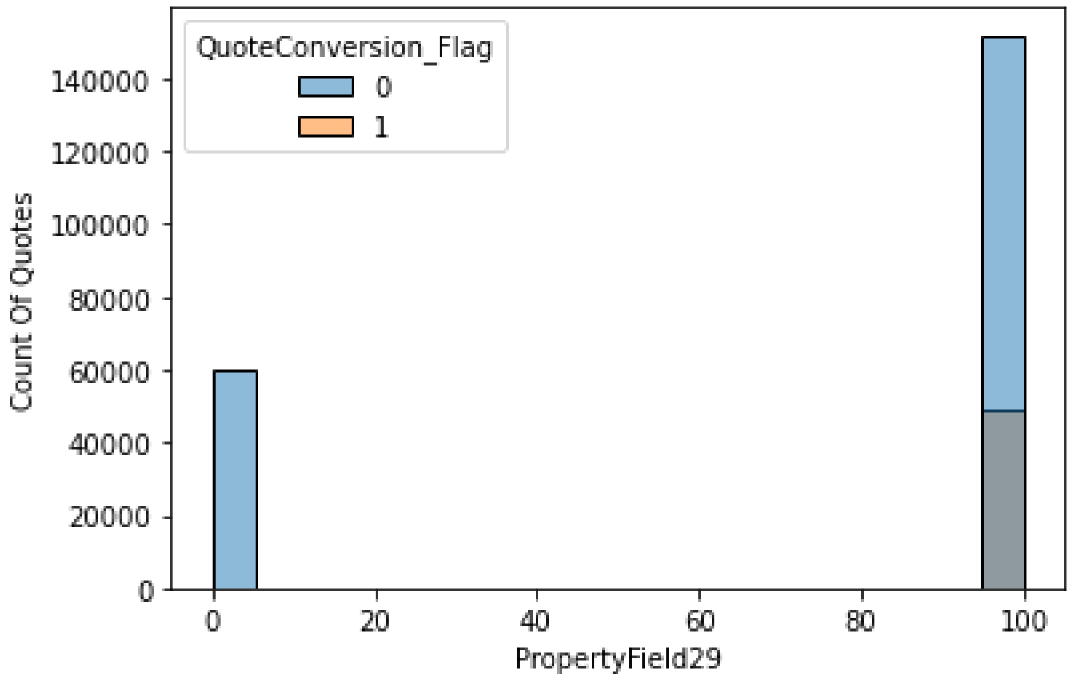

QuoteConversion_Flag
0
1
Count Of Quotes
PropertyField24B

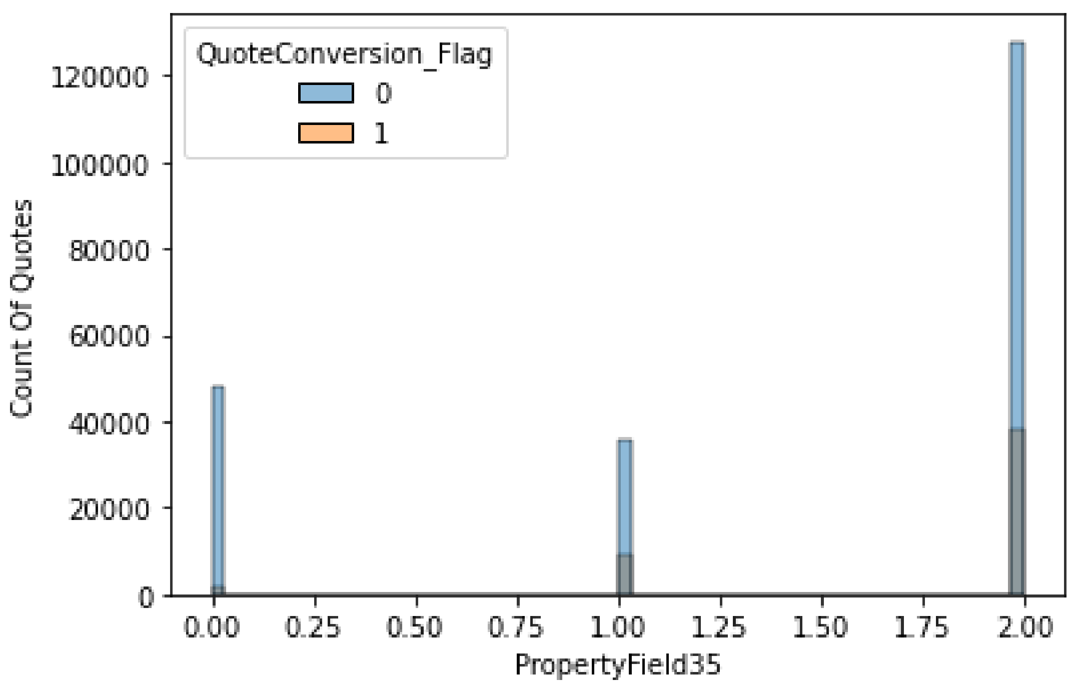

QuoteConversion_Flag
0
1
Count Of Quotes
PropertyField25

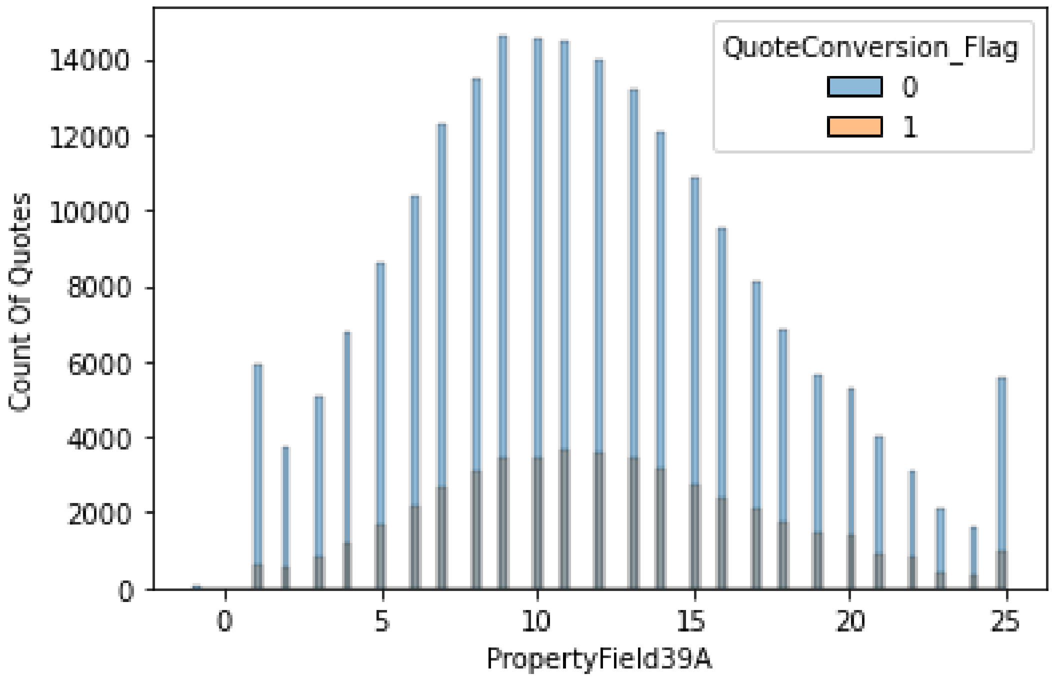

QuoteConversion_Flag
0
1
Count Of Quotes
PropertyField26A

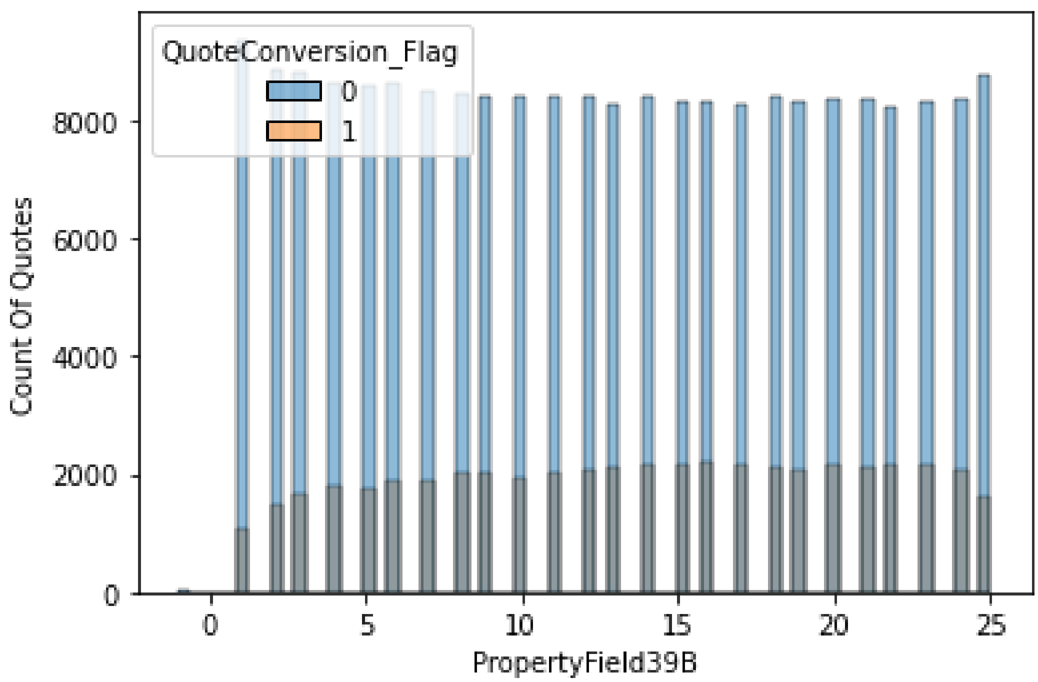

QuoteConversion_Flag
0
1
Count Of Quotes
PropertyField26B

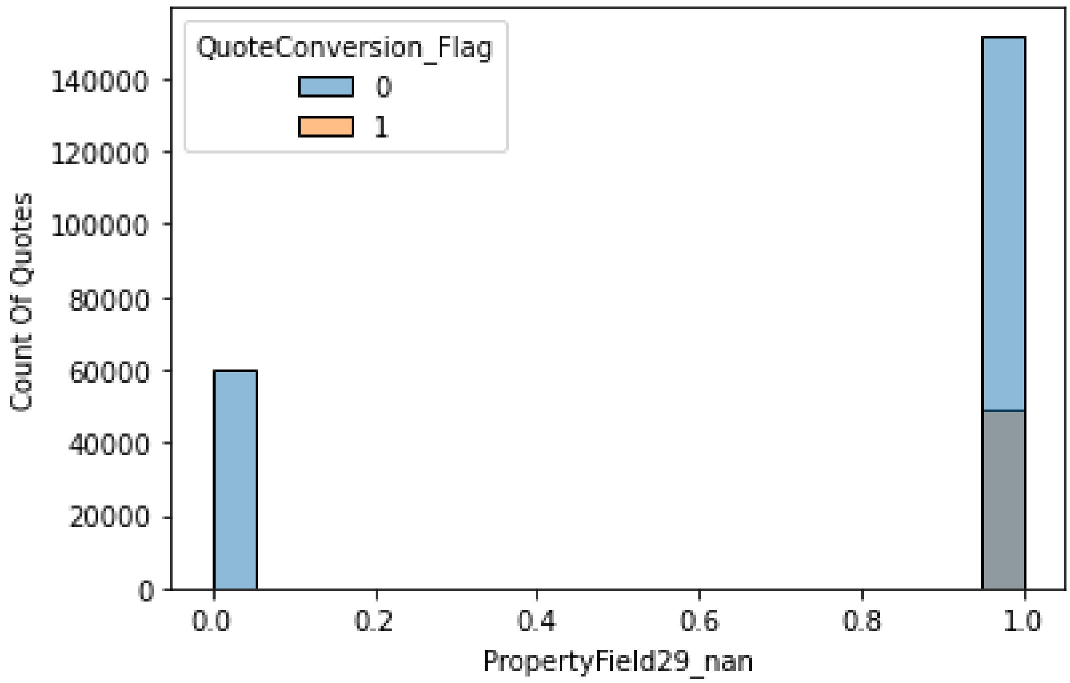

QuoteConversion_Flag
0
1
Count Of Quotes
PropertyField27


QuoteConversion_Flag
0
1
Count Of Quotes
0
20000
40000
60000
80000
100000
120000
140000
0
20
40
60
80
100
PropertyField29


QuoteConversion_Flag
0
1
Count Of Quotes
0
20000
40000
60000
80000
100000
120000
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
PropertyField35

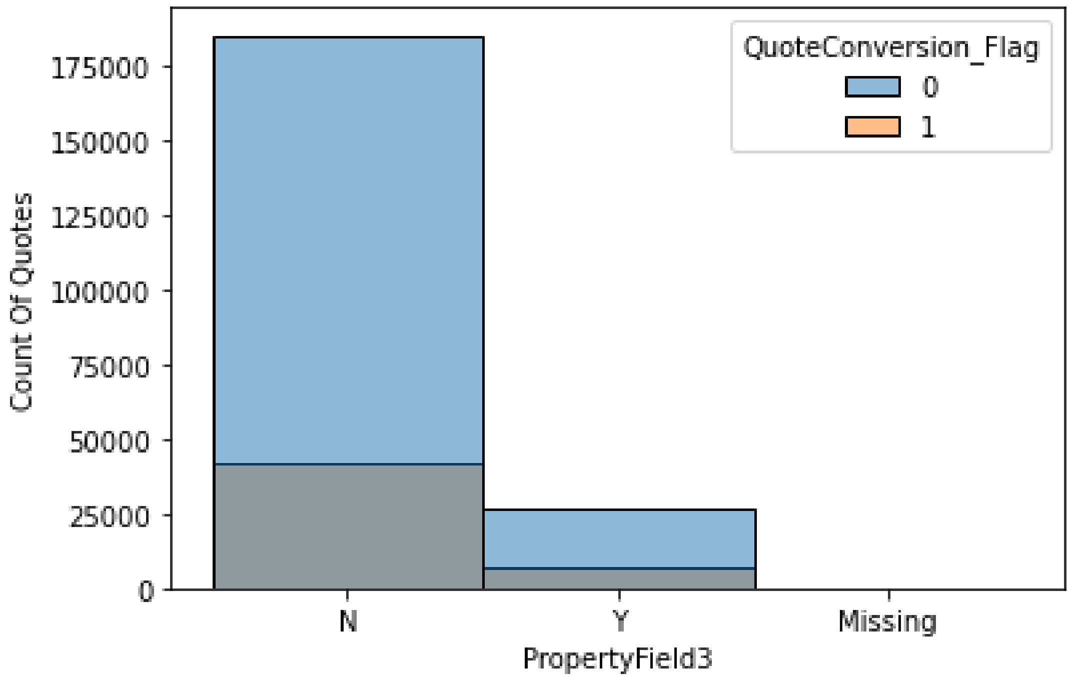

QuoteConversion_Flag
0
1
Count Of Quotes
0
2000
4000
6000
8000
10000
12000
14000
0
5
10
15
20
25
PropertyField39A

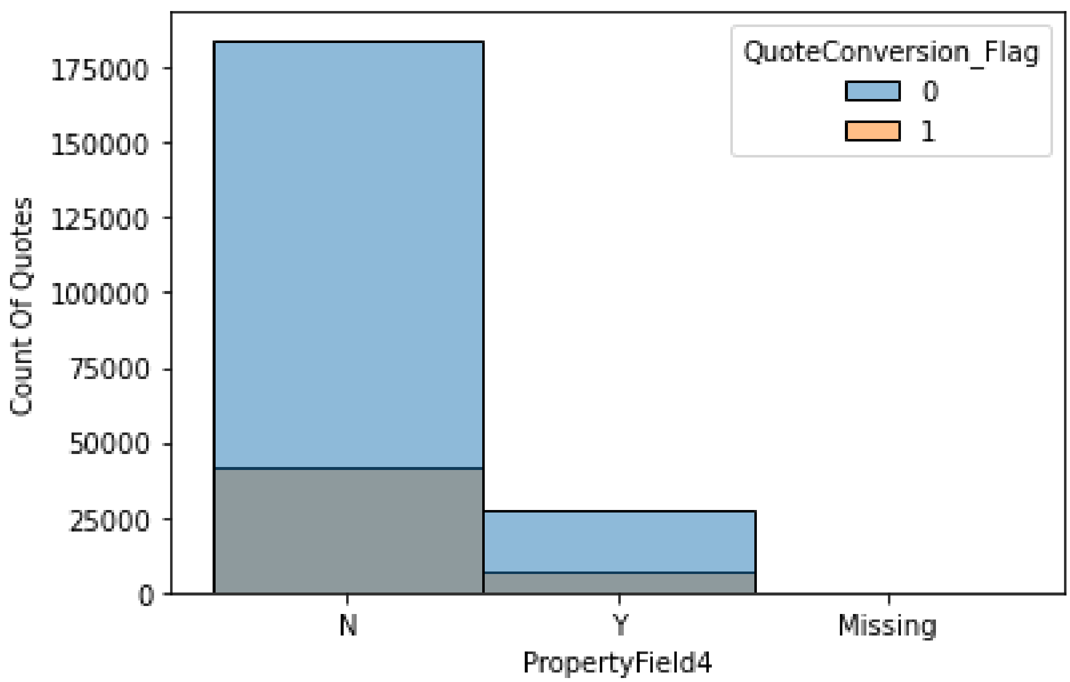

QuoteConversion_Flag
0
1
Count Of Quotes
0
2000
4000
6000
8000
0
5
10
15
20
25
PropertyField39B

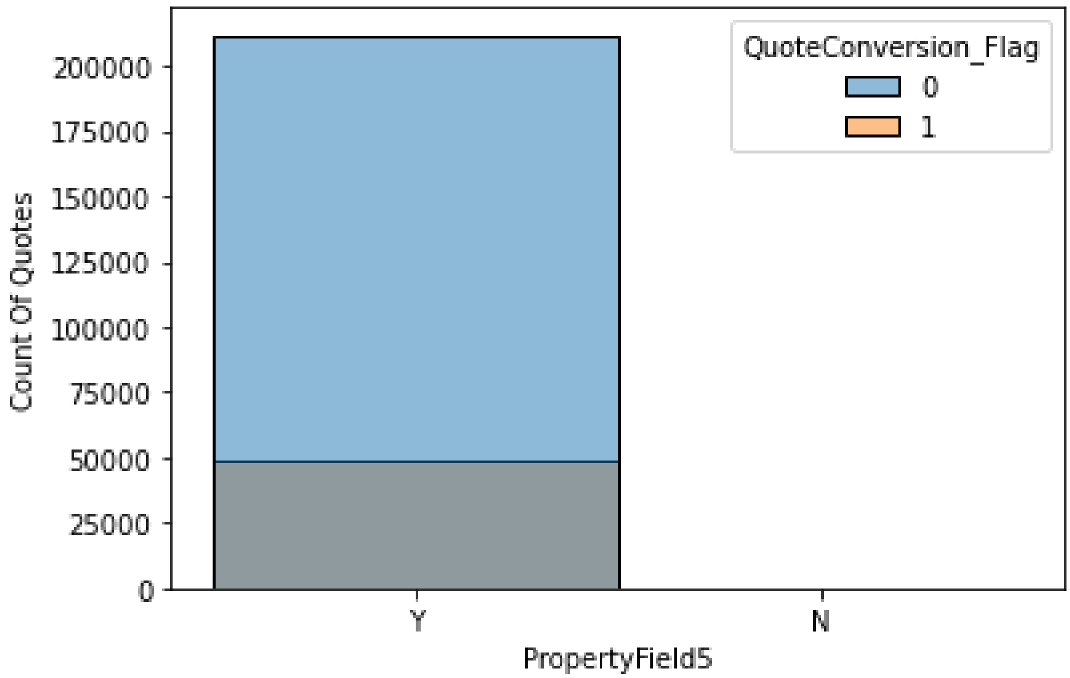


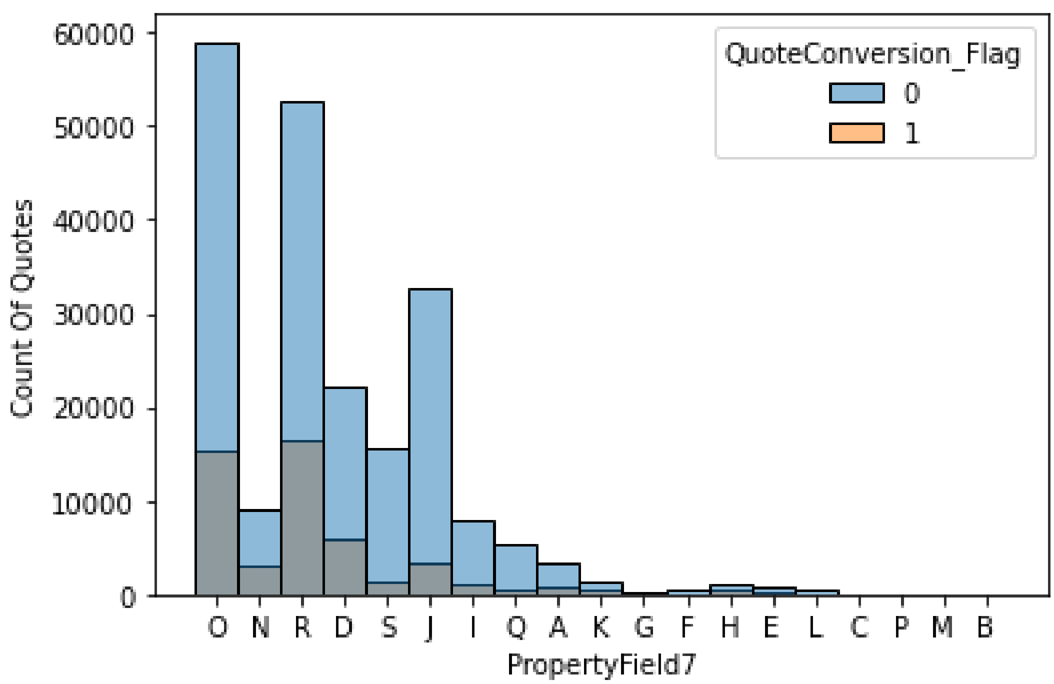
CATEGORICAL FEATURES

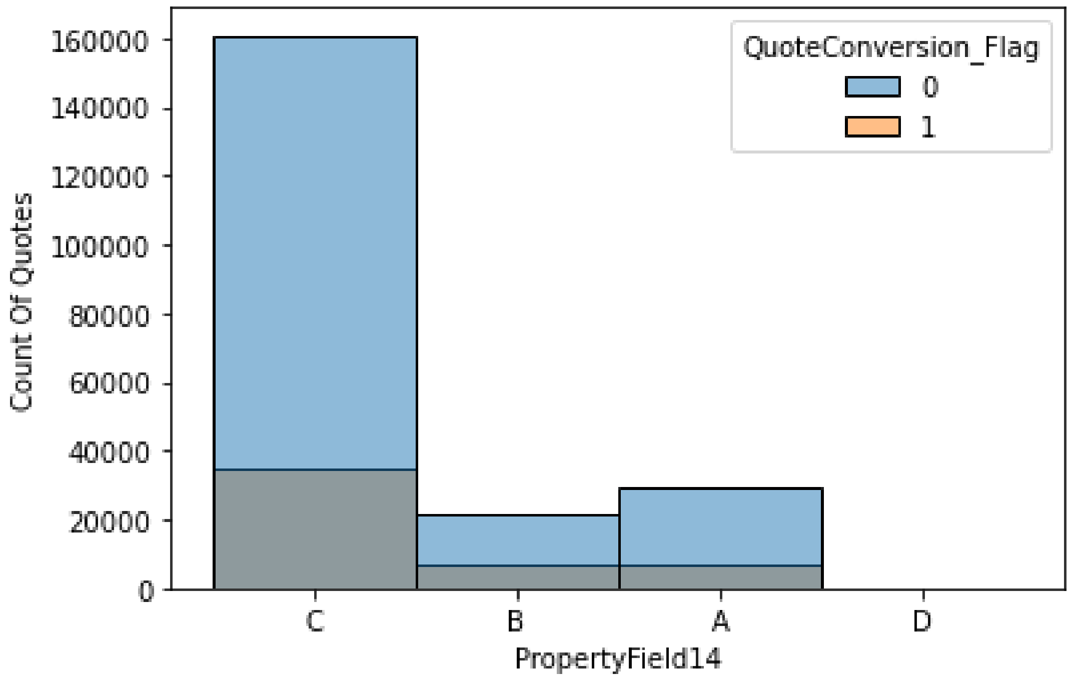
[47]:
```
for feature in get_categorical_features(dataset):
        sns.histplot(data = dataset, x = feature, hue = 'QuoteConversion_Flag')
        plt.xlabel(feature)
        plt.ylabel('Count Of Quotes')
        plt.show()
```

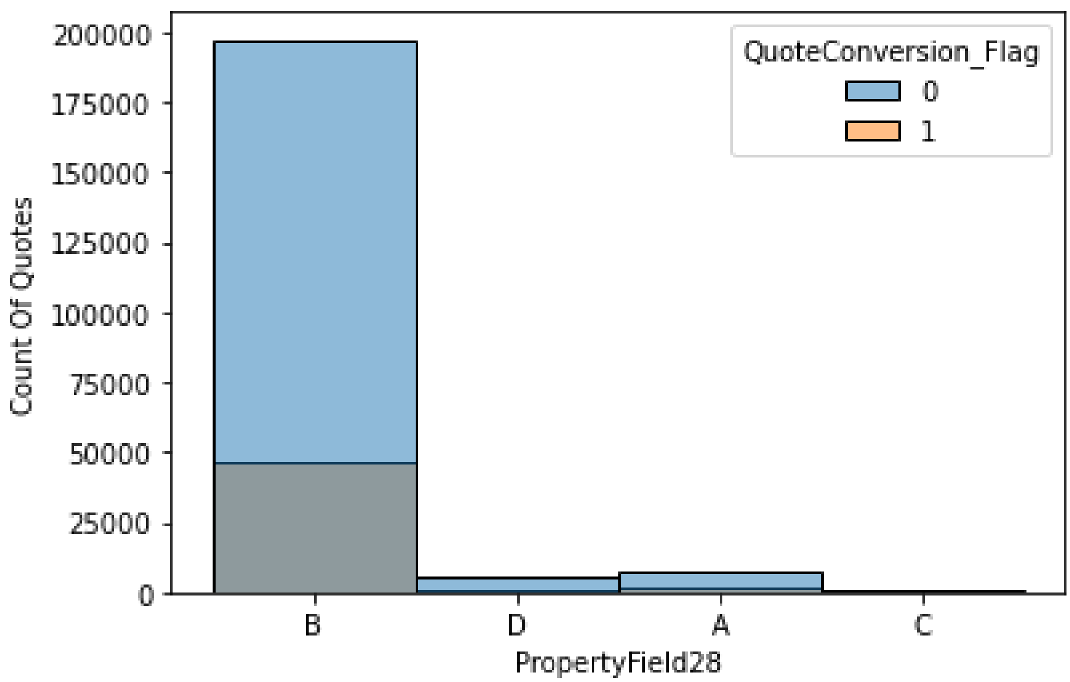

QuoteConversion_Flag
0
1
Count Of Quotes
175000
150000
125000
100000
75000
50000
25000
0
N
Y
Missing
PropertyField3

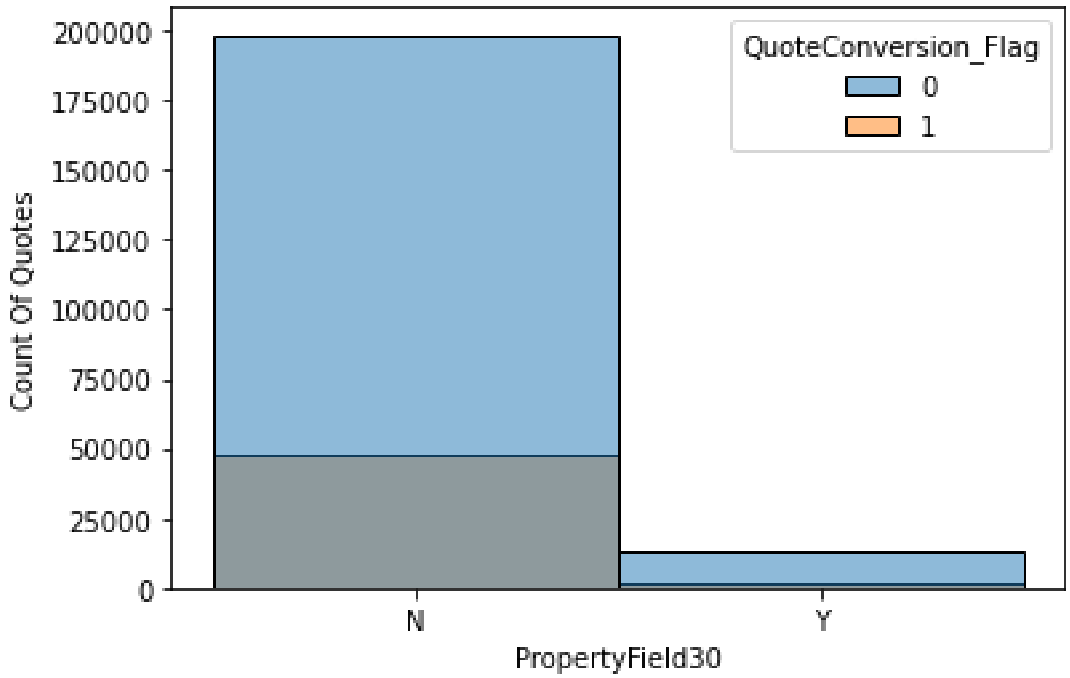

QuoteConversion_Flag
0
1
Count Of Quotes
175000
150000
125000
100000
75000
50000
25000
0
N
Y
Missing
PropertyField4


QuoteConversion_Flag
0
1
200000
175000
150000
125000
100000
75000
50000
25000
0
Count Of Quotes
Y
N
PropertyField5

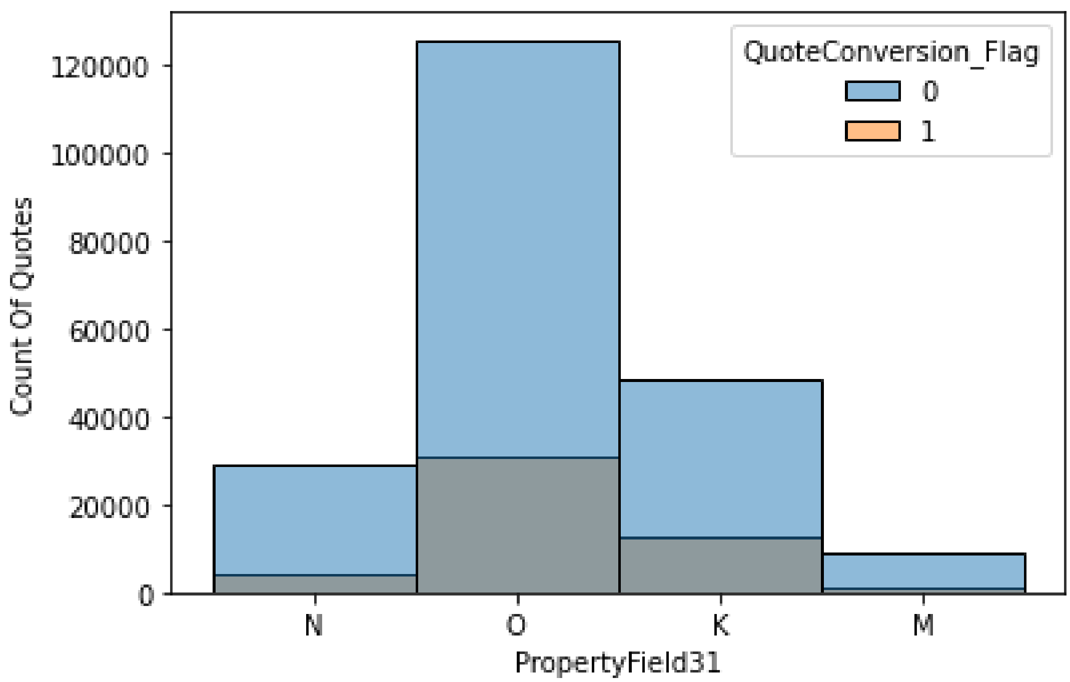

QuoteConversion_Flag
0
1
60000
50000
40000
30000
20000
10000
0
Count Of Quotes
O N R D S J I Q A K G F H E L C P M B
PropertyField7


QuoteConversion_Flag
0
1
Count Of Quotes
160000
140000
120000
100000
80000
60000
40000
20000
0
C
B
A
D
PropertyField14

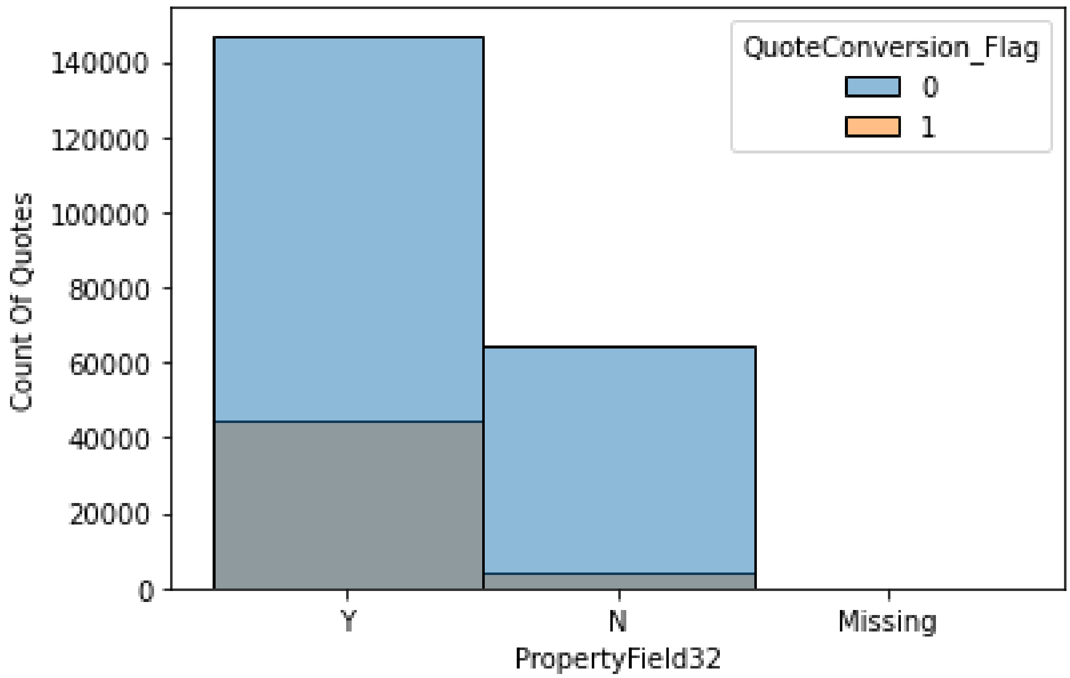

QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
B
D
A
C
PropertyField28

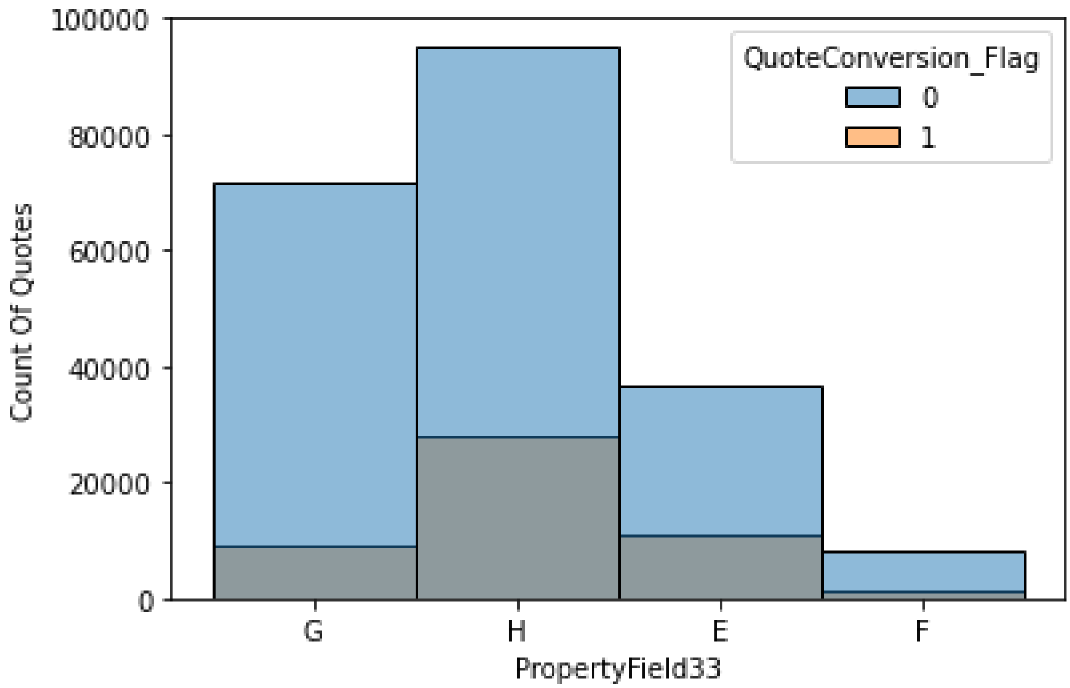


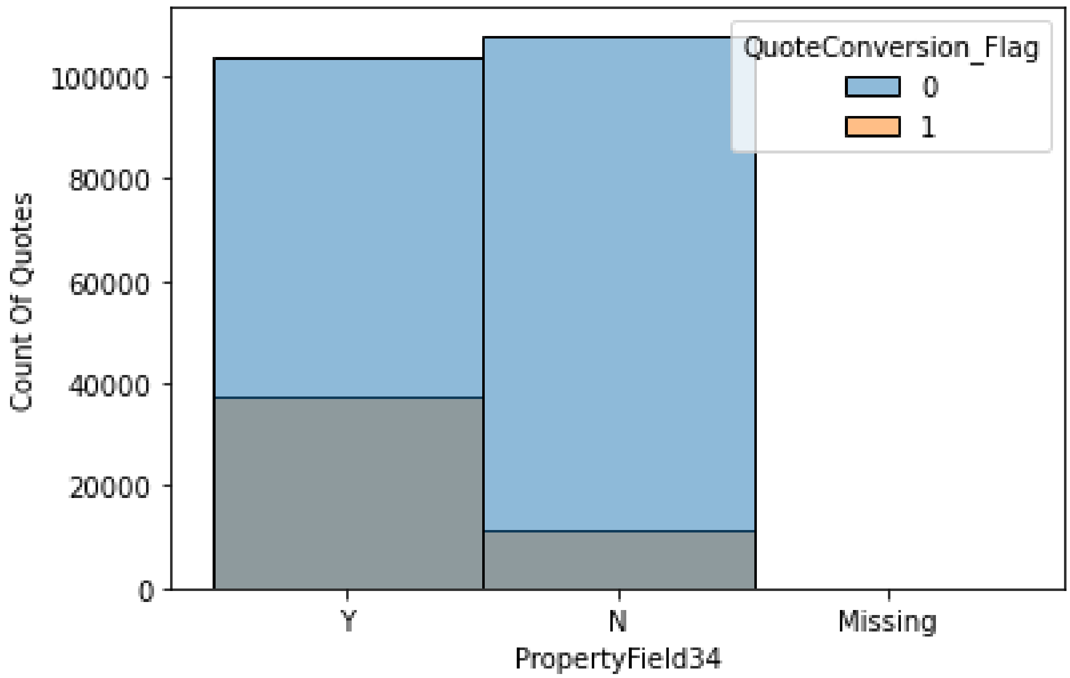
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
QuoteConversion_Flag
0
1
N
Y
PropertyField30



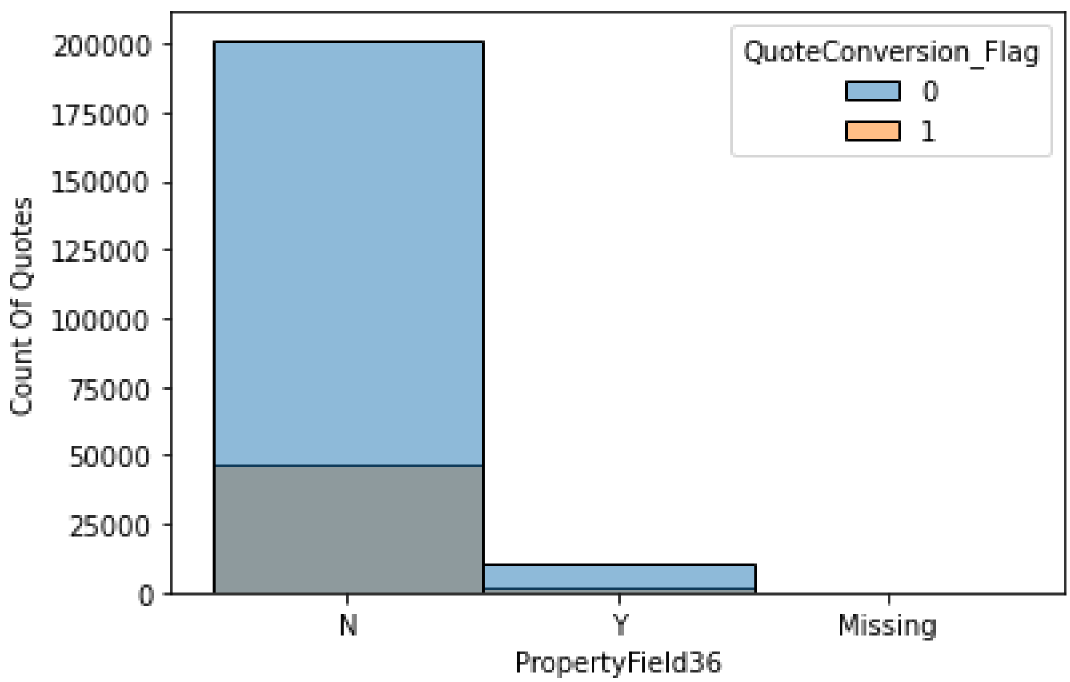
Count Of Quotes
120000
100000
80000
60000
40000
20000
0
QuoteConversion_Flag
0
1
N
O
K
M
PropertyField31


QuoteConversion_Flag
0
1
140000
120000
100000
80000
60000
40000
20000
0
Count Of Quotes
Y
N
Missing
PropertyField32

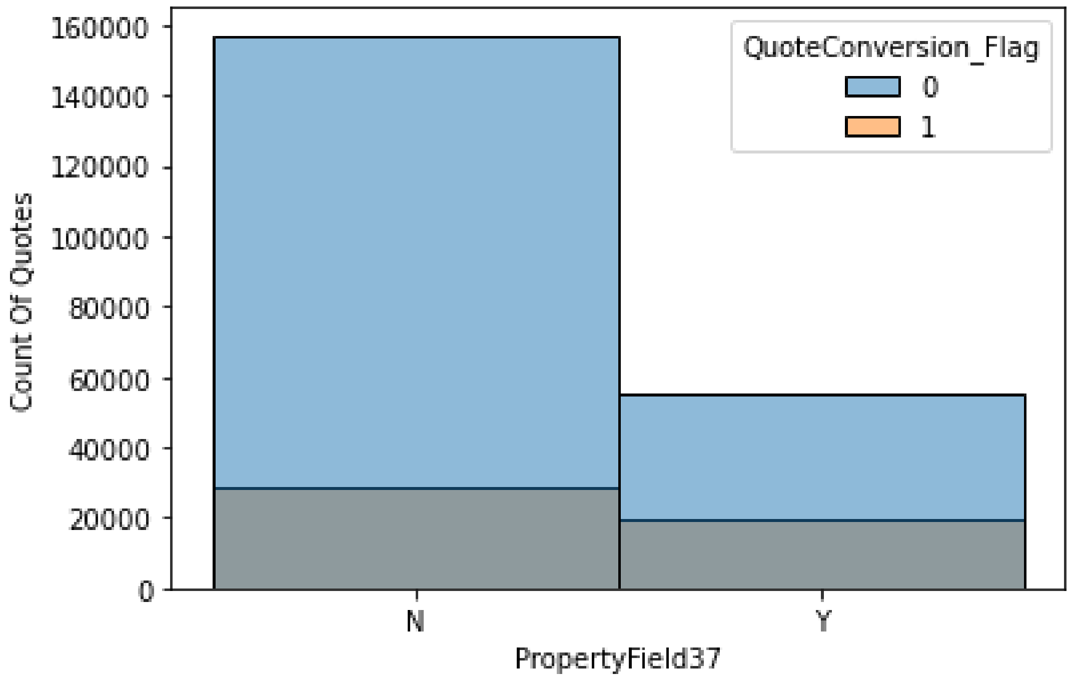

QuoteConversion_Flag
0
1
100000
80000
60000
40000
20000
0
Count Of Quotes
G
H
E
F
PropertyField33

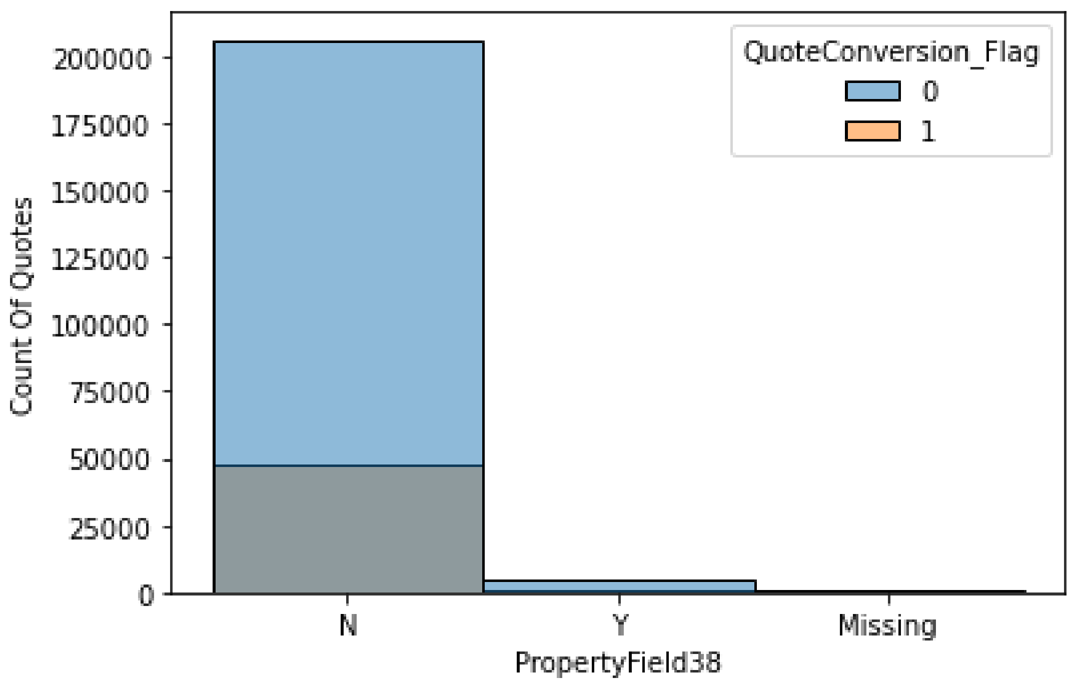


QuoteConversion_Flag
0
1
Count Of Quotes
100000
80000
60000
40000
20000
0
Y
N
Missing
PropertyField34



QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
N
Y
Missing
PropertyField36





GEOGRAPHIC FEATURES

[48]:
```
dataset = extract_feature_dataset('GeographicField',data)
```

```
[49]: for feature in dataset.columns:
          if feature != 'QuoteConversion_Flag':
              print('{} has {} unique values'.format(feature,len(dataset[feature].
       ↪unique())))
```

```
GeographicField1A has 26 unique values
GeographicField1B has 26 unique values
GeographicField2A has 26 unique values
GeographicField2B has 26 unique values
GeographicField3A has 25 unique values
GeographicField3B has 25 unique values
GeographicField4A has 25 unique values
GeographicField4B has 25 unique values
GeographicField5A has 2 unique values
GeographicField5B has 14 unique values
GeographicField6A has 26 unique values
GeographicField6B has 26 unique values
GeographicField7A has 25 unique values
GeographicField7B has 24 unique values
GeographicField8A has 26 unique values
GeographicField8B has 25 unique values
GeographicField9A has 26 unique values
GeographicField9B has 26 unique values
GeographicField10A has 1 unique values
GeographicField10B has 2 unique values
GeographicField11A has 26 unique values
GeographicField11B has 25 unique values
GeographicField12A has 26 unique values
GeographicField12B has 24 unique values
GeographicField13A has 25 unique values
GeographicField13B has 25 unique values
GeographicField14A has 2 unique values
GeographicField14B has 20 unique values
GeographicField15A has 26 unique values
GeographicField15B has 24 unique values
GeographicField16A has 26 unique values
GeographicField16B has 26 unique values
GeographicField17A has 26 unique values
GeographicField17B has 26 unique values
GeographicField18A has 2 unique values
GeographicField18B has 25 unique values
GeographicField19A has 26 unique values
GeographicField19B has 26 unique values
GeographicField20A has 26 unique values
GeographicField20B has 26 unique values
GeographicField21A has 2 unique values
GeographicField21B has 23 unique values
```

```
GeographicField22A has 2 unique values
GeographicField22B has 12 unique values
GeographicField23A has 2 unique values
GeographicField23B has 26 unique values
GeographicField24A has 25 unique values
GeographicField24B has 25 unique values
GeographicField25A has 25 unique values
GeographicField25B has 25 unique values
GeographicField26A has 26 unique values
GeographicField26B has 25 unique values
GeographicField27A has 25 unique values
GeographicField27B has 24 unique values
GeographicField28A has 26 unique values
GeographicField28B has 26 unique values
GeographicField29A has 26 unique values
GeographicField29B has 26 unique values
GeographicField30A has 26 unique values
GeographicField30B has 26 unique values
GeographicField31A has 26 unique values
GeographicField31B has 26 unique values
GeographicField32A has 26 unique values
GeographicField32B has 26 unique values
GeographicField33A has 26 unique values
GeographicField33B has 26 unique values
GeographicField34A has 26 unique values
GeographicField34B has 26 unique values
GeographicField35A has 26 unique values
GeographicField35B has 26 unique values
GeographicField36A has 26 unique values
GeographicField36B has 26 unique values
GeographicField37A has 26 unique values
GeographicField37B has 26 unique values
GeographicField38A has 26 unique values
GeographicField38B has 26 unique values
GeographicField39A has 26 unique values
GeographicField39B has 26 unique values
GeographicField40A has 26 unique values
GeographicField40B has 26 unique values
GeographicField41A has 26 unique values
GeographicField41B has 26 unique values
GeographicField42A has 26 unique values
GeographicField42B has 26 unique values
GeographicField43A has 26 unique values
GeographicField43B has 26 unique values
GeographicField44A has 26 unique values
GeographicField44B has 26 unique values
GeographicField45A has 26 unique values
GeographicField45B has 26 unique values
```

```
GeographicField46A has 26 unique values
GeographicField46B has 26 unique values
GeographicField47A has 26 unique values
GeographicField47B has 25 unique values
GeographicField48A has 26 unique values
GeographicField48B has 26 unique values
GeographicField49A has 26 unique values
GeographicField49B has 26 unique values
GeographicField50A has 26 unique values
GeographicField50B has 26 unique values
GeographicField51A has 26 unique values
GeographicField51B has 26 unique values
GeographicField52A has 26 unique values
GeographicField52B has 26 unique values
GeographicField53A has 26 unique values
GeographicField53B has 26 unique values
GeographicField54A has 26 unique values
GeographicField54B has 26 unique values
GeographicField55A has 26 unique values
GeographicField55B has 26 unique values
GeographicField56A has 2 unique values
GeographicField56B has 25 unique values
GeographicField57A has 26 unique values
GeographicField57B has 26 unique values
GeographicField58A has 26 unique values
GeographicField58B has 26 unique values
GeographicField59A has 26 unique values
GeographicField59B has 26 unique values
GeographicField60A has 2 unique values
GeographicField60B has 26 unique values
GeographicField61A has 2 unique values
GeographicField61B has 25 unique values
GeographicField62A has 2 unique values
GeographicField62B has 19 unique values
GeographicField63 has 3 unique values
GeographicField64 has 4 unique values
```

[50]:
```
numerical_features = get_numerical_features(dataset)
categorical_features = get_categorical_features(dataset)
```

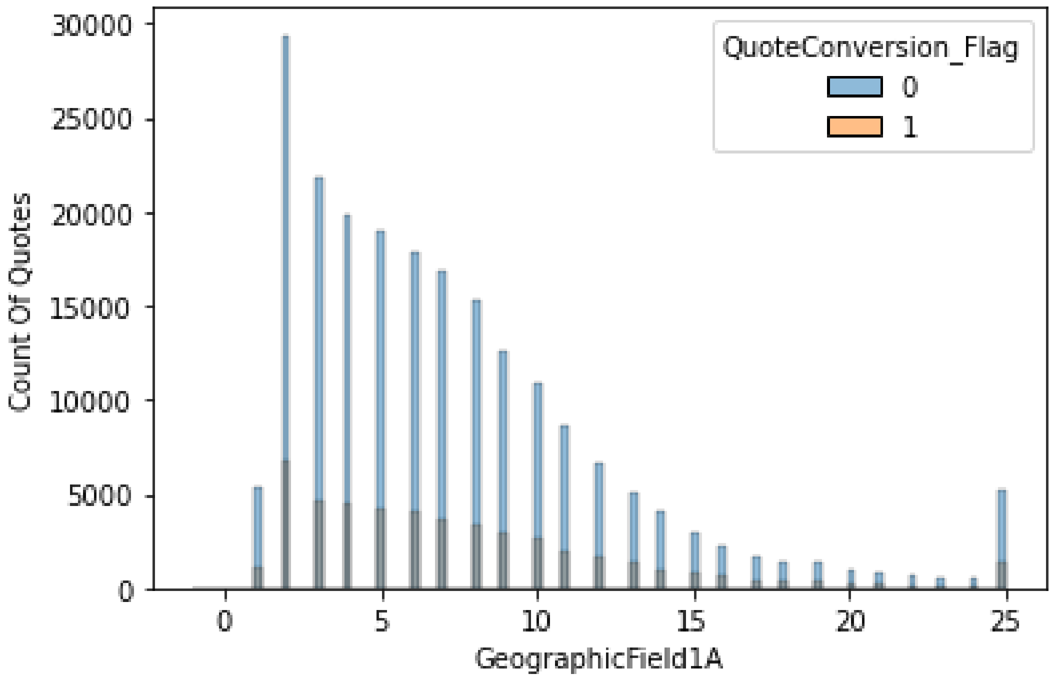
[51]:
```
print('Numerical Features : ', numerical_features)
print('Categorical Features : ', categorical_features)
```

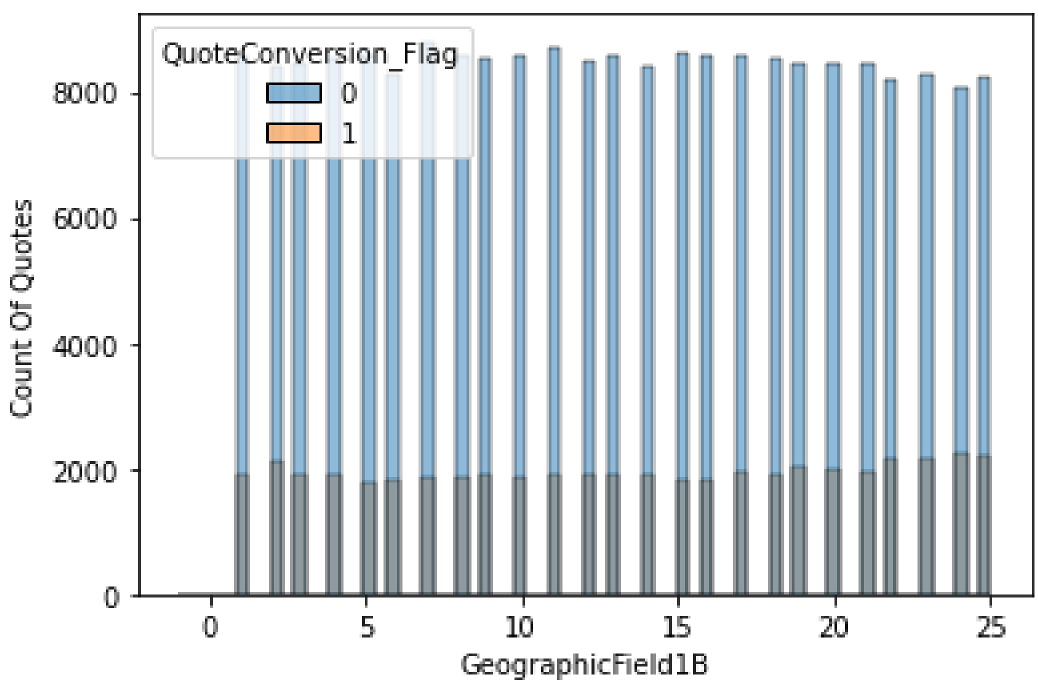
```
Numerical Features :  ['GeographicField1A', 'GeographicField1B',
'GeographicField2A', 'GeographicField2B', 'GeographicField3A',
'GeographicField3B', 'GeographicField4A', 'GeographicField4B',
'GeographicField5A', 'GeographicField5B', 'GeographicField6A',
'GeographicField6B', 'GeographicField7A', 'GeographicField7B',
```

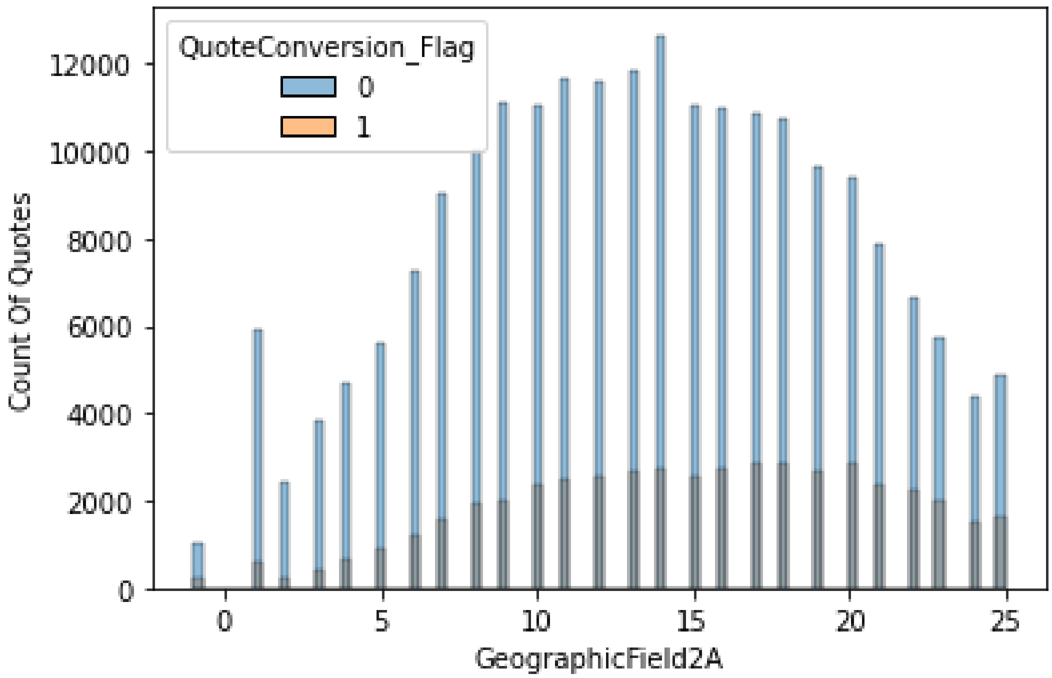
```
'GeographicField8A', 'GeographicField8B', 'GeographicField9A',
'GeographicField9B', 'GeographicField10A', 'GeographicField10B',
'GeographicField11A', 'GeographicField11B', 'GeographicField12A',
'GeographicField12B', 'GeographicField13A', 'GeographicField13B',
'GeographicField14A', 'GeographicField14B', 'GeographicField15A',
'GeographicField15B', 'GeographicField16A', 'GeographicField16B',
'GeographicField17A', 'GeographicField17B', 'GeographicField18A',
'GeographicField18B', 'GeographicField19A', 'GeographicField19B',
'GeographicField20A', 'GeographicField20B', 'GeographicField21A',
'GeographicField21B', 'GeographicField22A', 'GeographicField22B',
'GeographicField23A', 'GeographicField23B', 'GeographicField24A',
'GeographicField24B', 'GeographicField25A', 'GeographicField25B',
'GeographicField26A', 'GeographicField26B', 'GeographicField27A',
'GeographicField27B', 'GeographicField28A', 'GeographicField28B',
'GeographicField29A', 'GeographicField29B', 'GeographicField30A',
'GeographicField30B', 'GeographicField31A', 'GeographicField31B',
'GeographicField32A', 'GeographicField32B', 'GeographicField33A',
'GeographicField33B', 'GeographicField34A', 'GeographicField34B',
'GeographicField35A', 'GeographicField35B', 'GeographicField36A',
'GeographicField36B', 'GeographicField37A', 'GeographicField37B',
'GeographicField38A', 'GeographicField38B', 'GeographicField39A',
'GeographicField39B', 'GeographicField40A', 'GeographicField40B',
'GeographicField41A', 'GeographicField41B', 'GeographicField42A',
'GeographicField42B', 'GeographicField43A', 'GeographicField43B',
'GeographicField44A', 'GeographicField44B', 'GeographicField45A',
'GeographicField45B', 'GeographicField46A', 'GeographicField46B',
'GeographicField47A', 'GeographicField47B', 'GeographicField48A',
'GeographicField48B', 'GeographicField49A', 'GeographicField49B',
'GeographicField50A', 'GeographicField50B', 'GeographicField51A',
'GeographicField51B', 'GeographicField52A', 'GeographicField52B',
'GeographicField53A', 'GeographicField53B', 'GeographicField54A',
'GeographicField54B', 'GeographicField55A', 'GeographicField55B',
'GeographicField56A', 'GeographicField56B', 'GeographicField57A',
'GeographicField57B', 'GeographicField58A', 'GeographicField58B',
'GeographicField59A', 'GeographicField59B', 'GeographicField60A',
'GeographicField60B', 'GeographicField61A', 'GeographicField61B',
'GeographicField62A', 'GeographicField62B', 'QuoteConversion_Flag']
Categorical Features :  ['GeographicField63', 'GeographicField64']
```

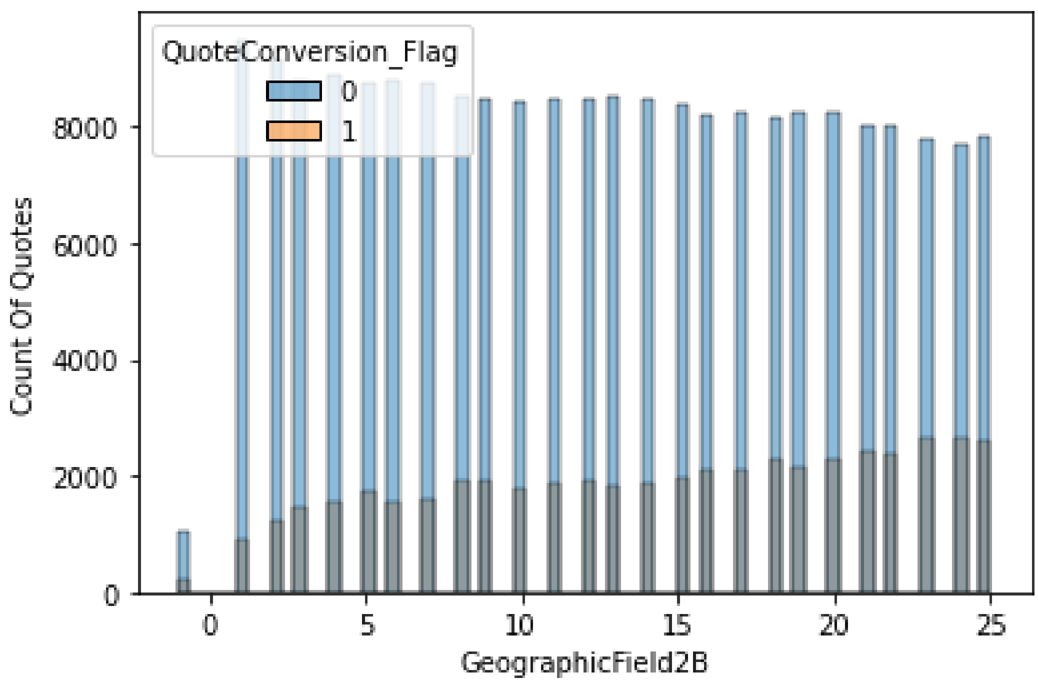
DISCRETE NUMERICAL FEATURES

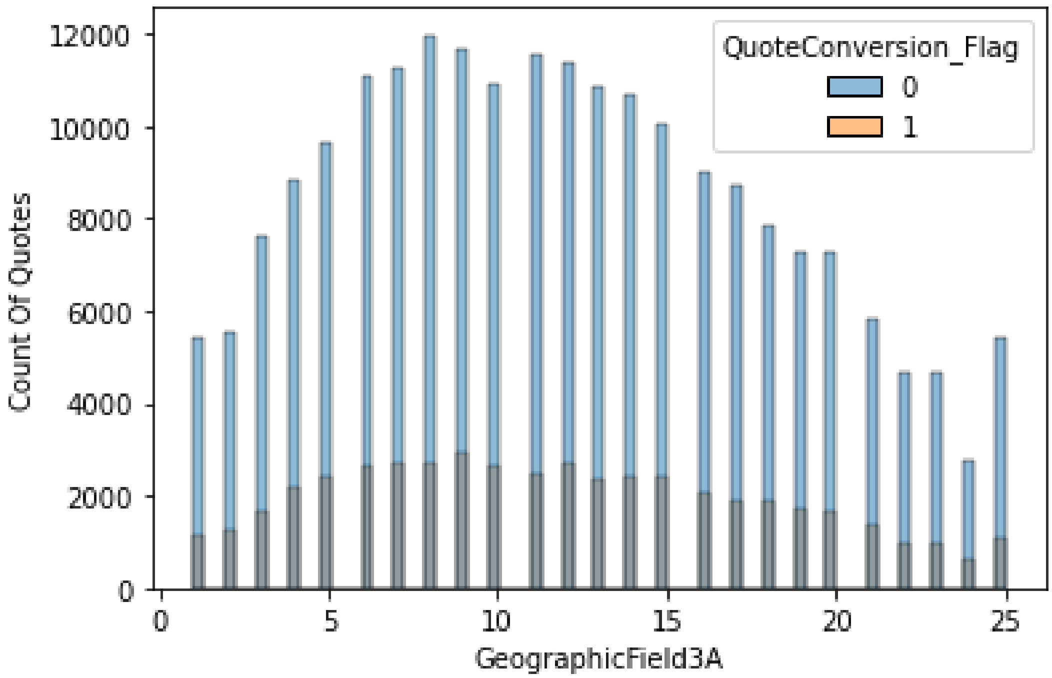
[52]:
```
for feature in get_numerical_features(dataset):
    if feature != 'QuoteConversion_Flag' and len(dataset[feature].unique()) <
 ↪30:
        sns.histplot(data = dataset, x = feature, hue = 'QuoteConversion_Flag')
        plt.xlabel(feature)
        plt.ylabel('Count Of Quotes')
        plt.show()
```

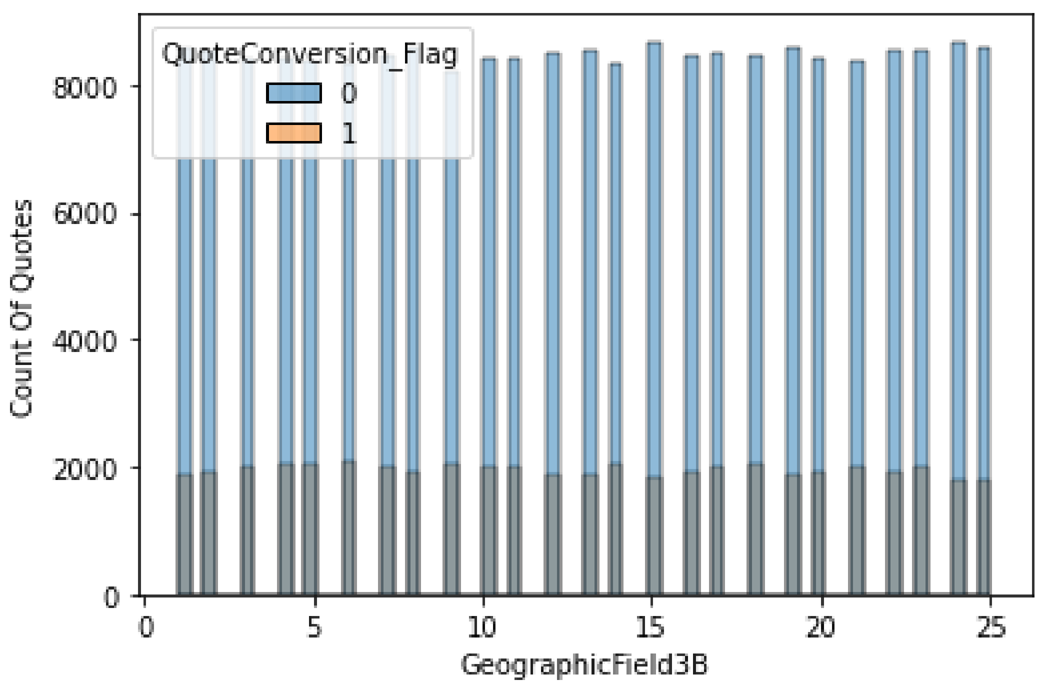

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField1A

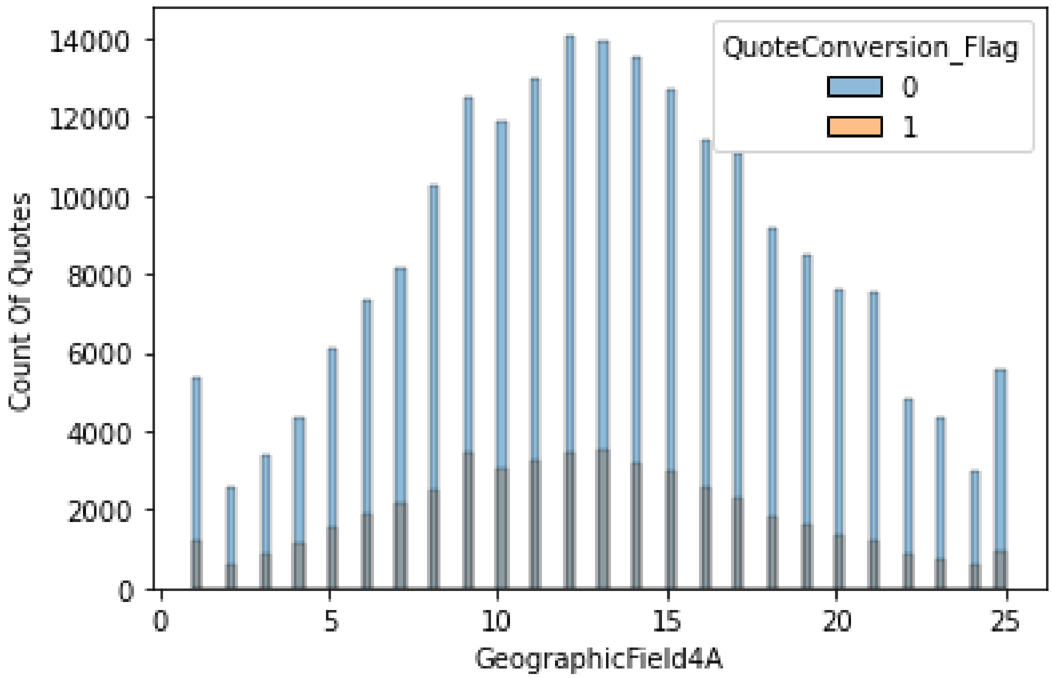

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField1B

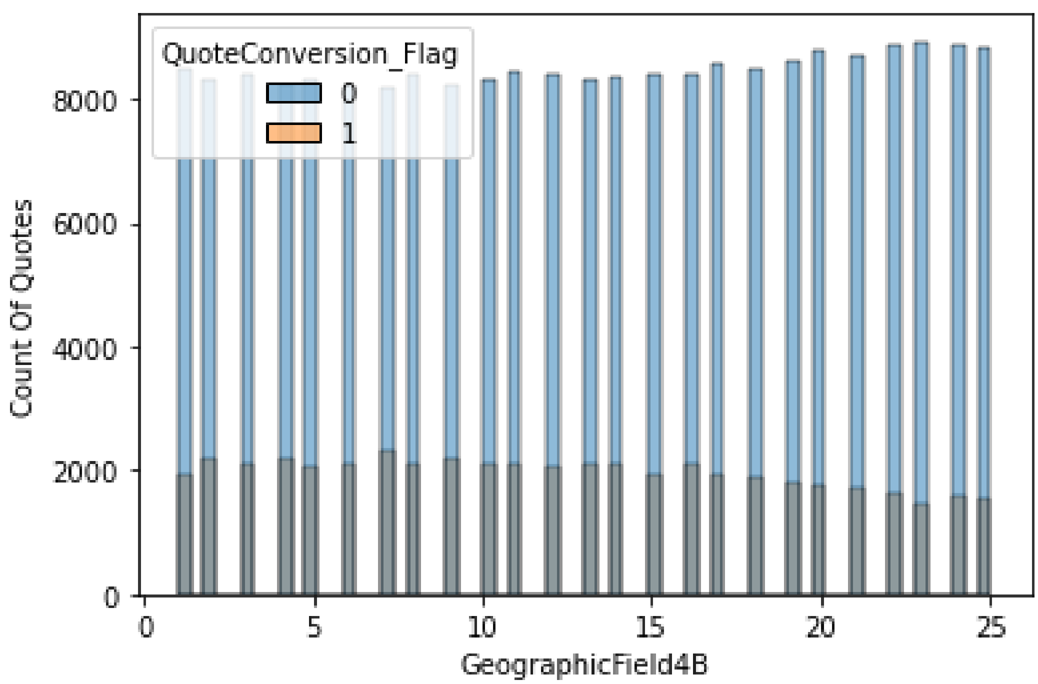

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField2A
0
5
10
15
20
25
0
2000
4000
6000
8000
10000
12000

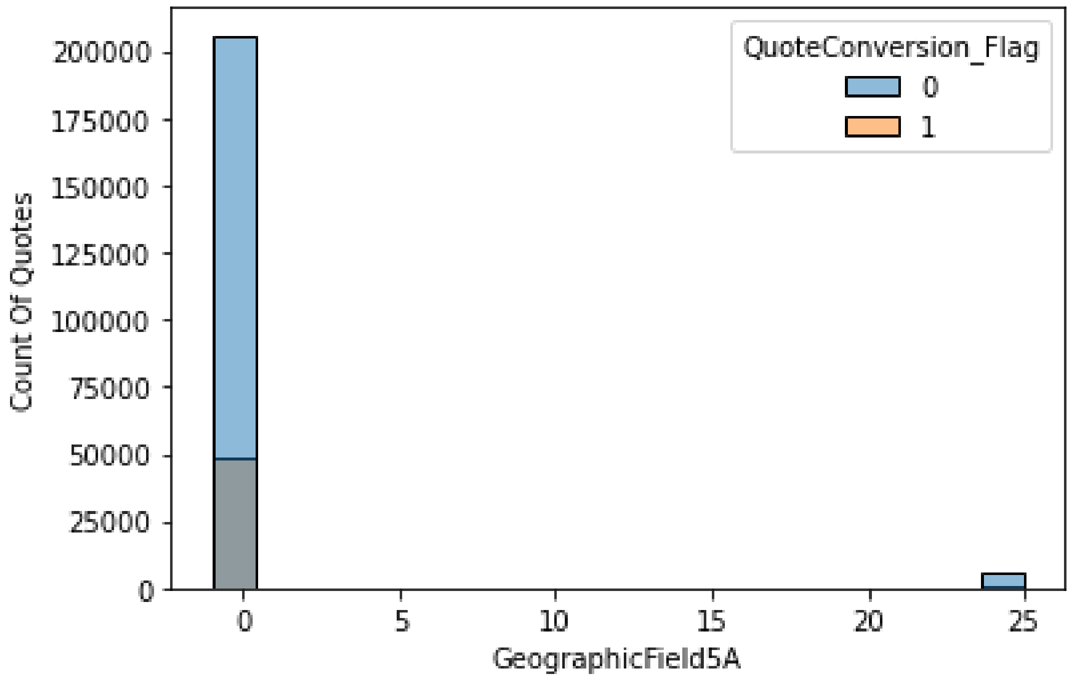

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField2B
0
5
10
15
20
25
0
2000
4000
6000
8000

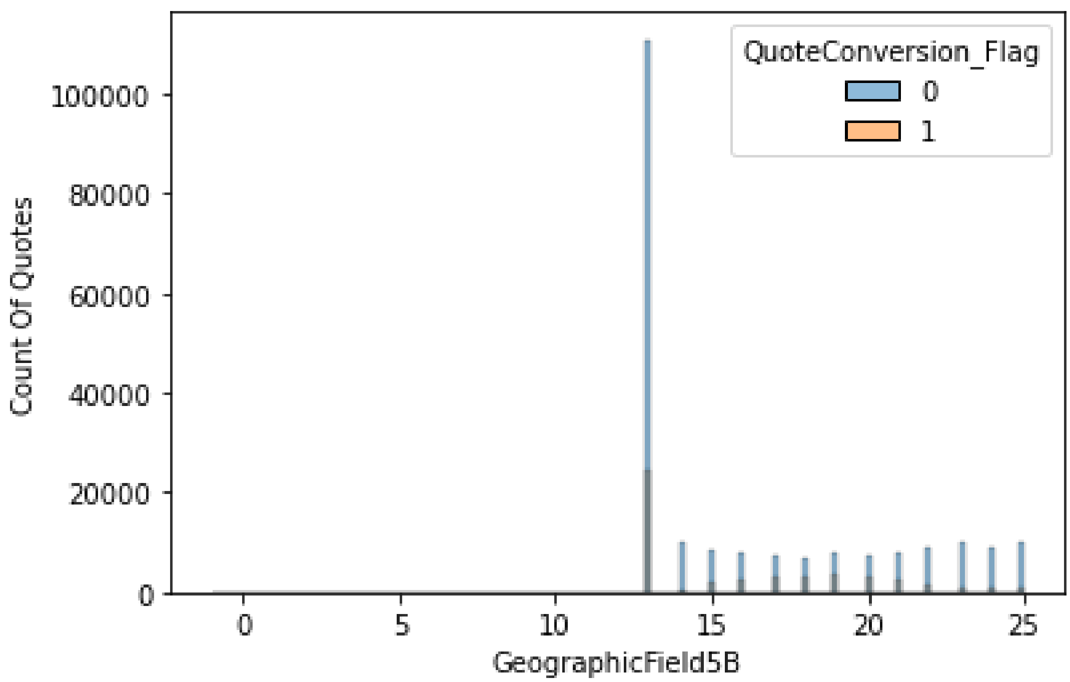

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField3A
0
2000
4000
6000
8000
10000
12000
0
5
10
15
20
25

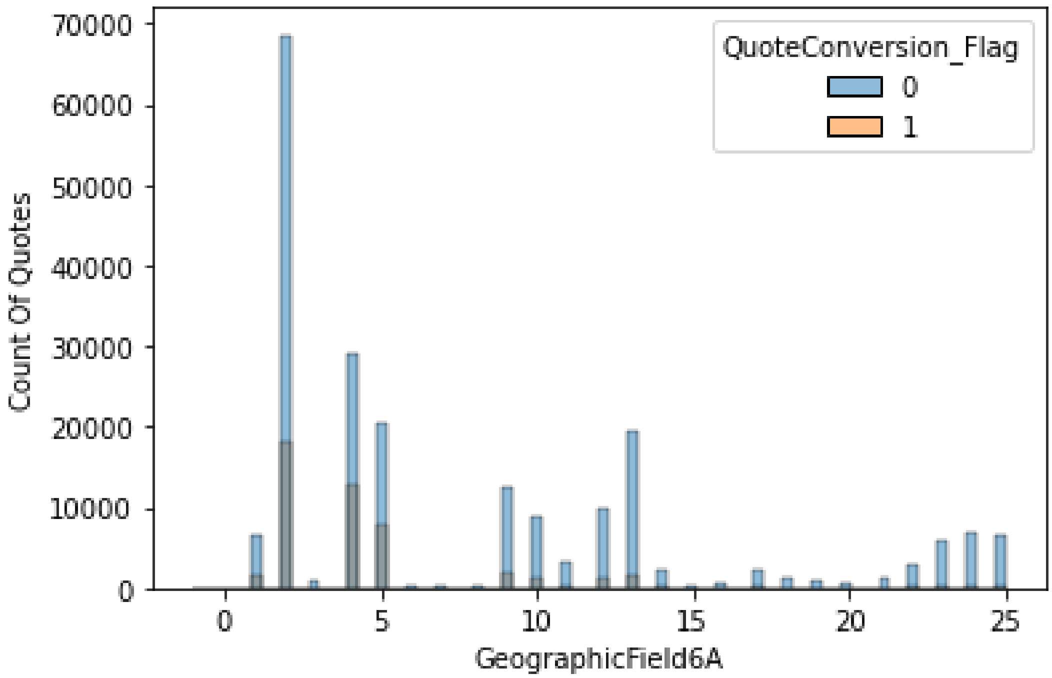

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField3B
0
2000
4000
6000
8000
0
5
10
15
20
25

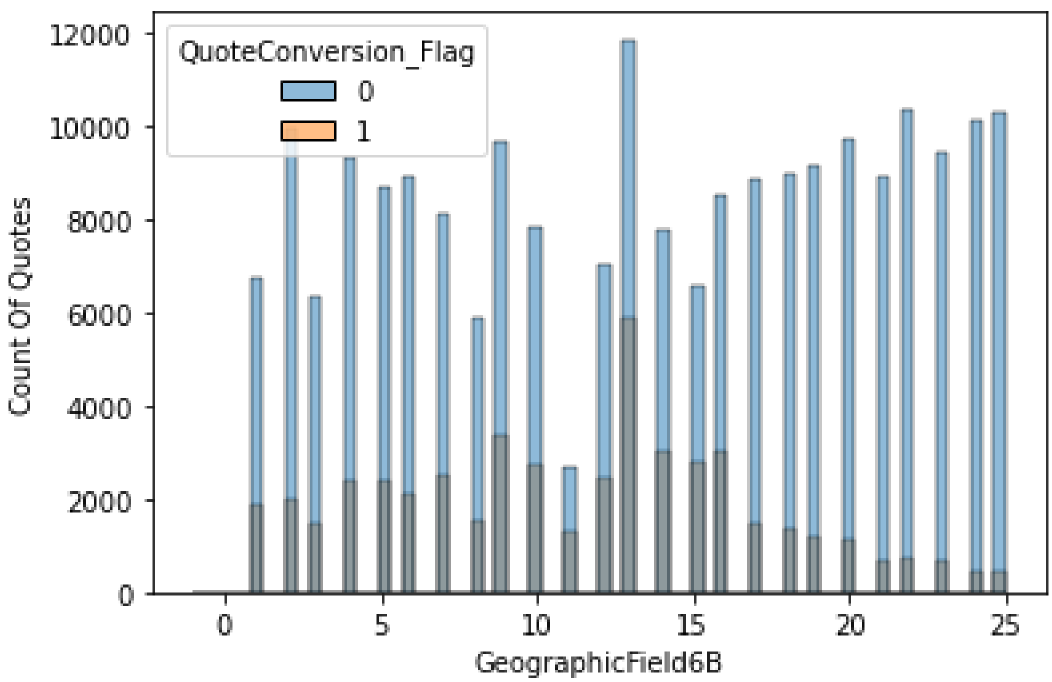

QuoteConversion_Flag
0
1
Count Of Quotes
0
2000
4000
6000
8000
10000
12000
14000
0
5
10
15
20
25
GeographicField4A

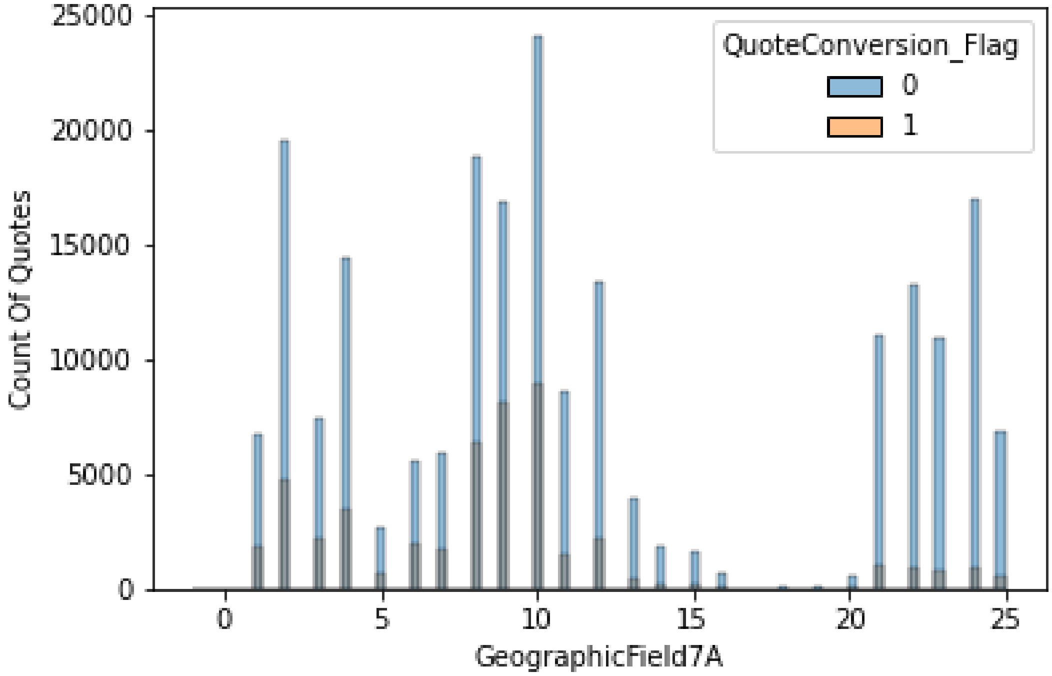

QuoteConversion_Flag
0
1
Count Of Quotes
0
2000
4000
6000
8000
0
5
10
15
20
25
GeographicField4B

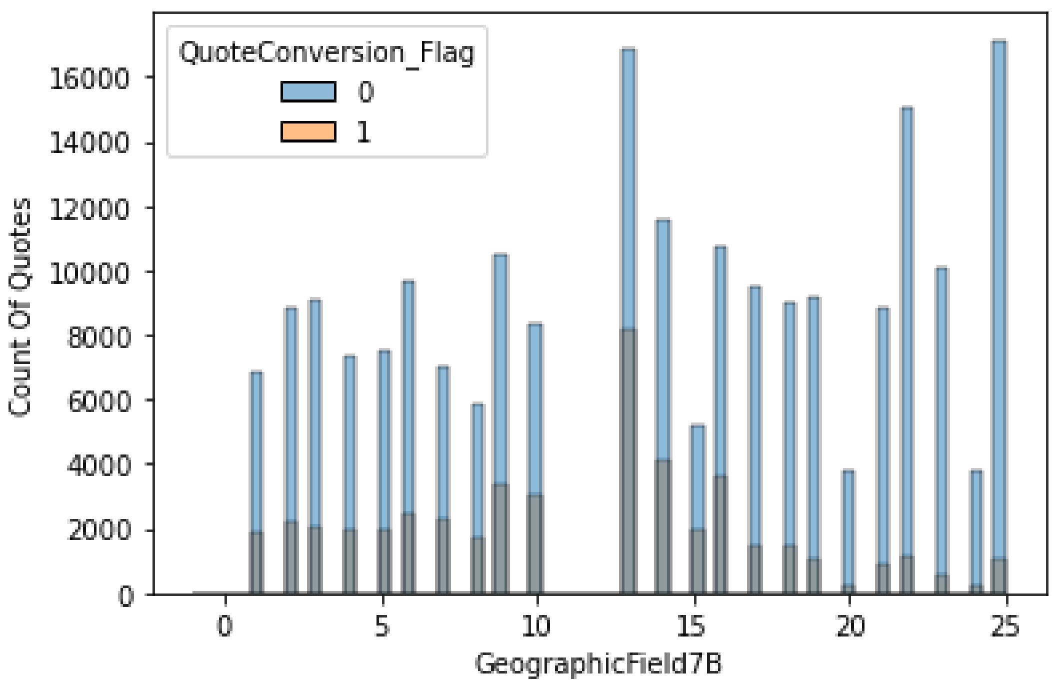

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField5A
200000
175000
150000
125000
100000
75000
50000
25000
0
0
5
10
15
20
25

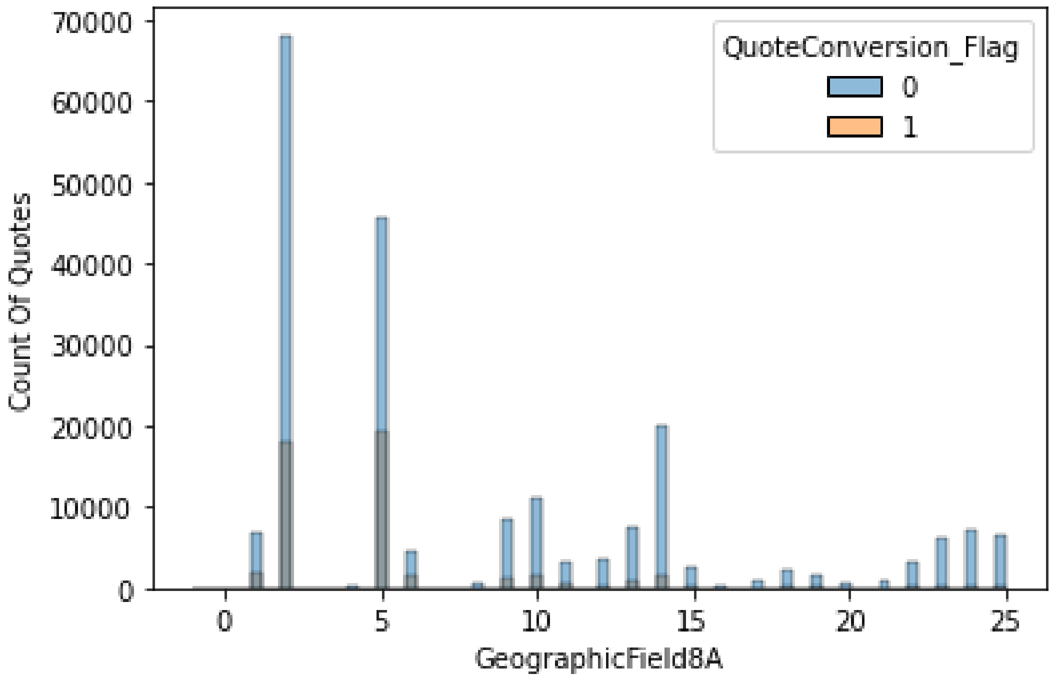

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField5B
100000
80000
60000
40000
20000
0
0
5
10
15
20
25


QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField6A

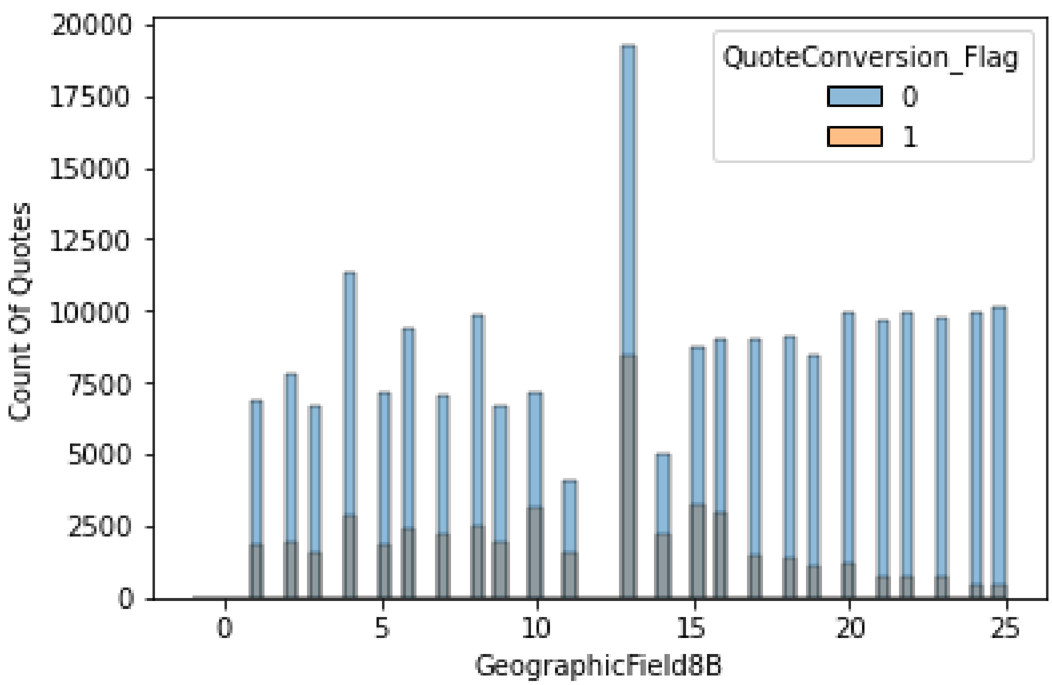

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField6B


QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField7A
25000
20000
15000
10000
5000
0
0
5
10
15
20
25

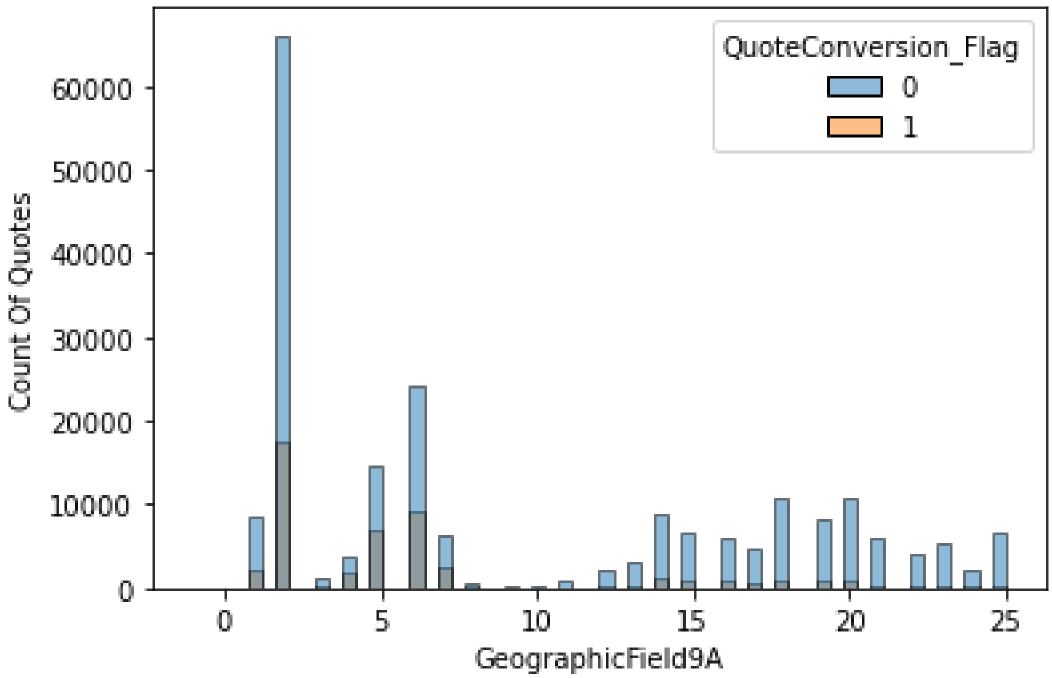

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField7B
16000
14000
12000
10000
8000
6000
4000
2000
0
0
5
10
15
20
25

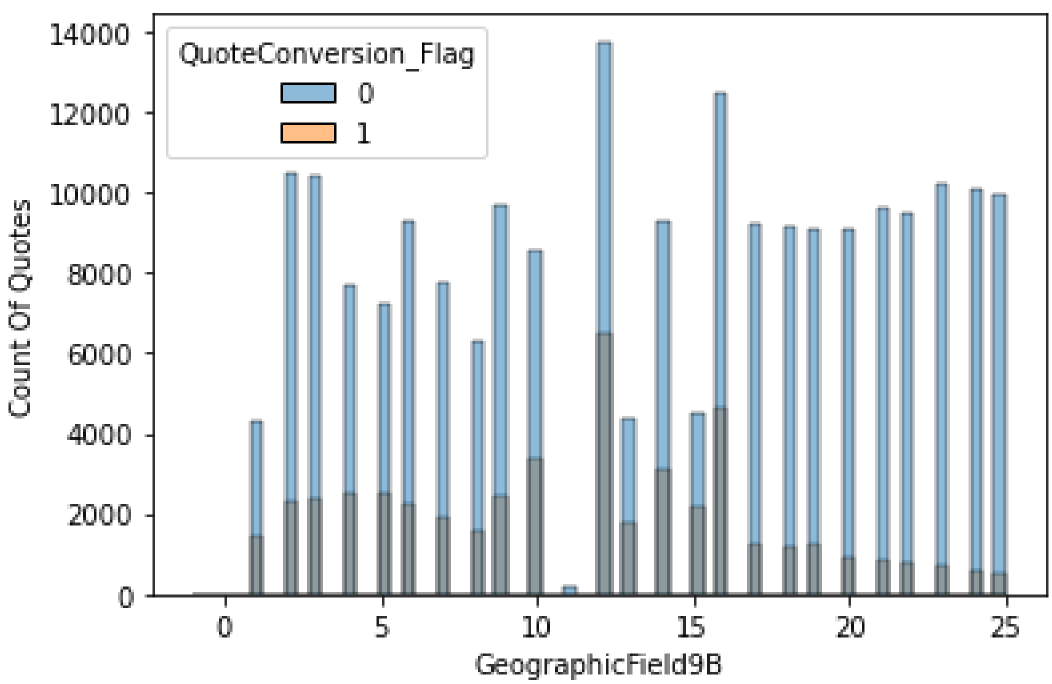


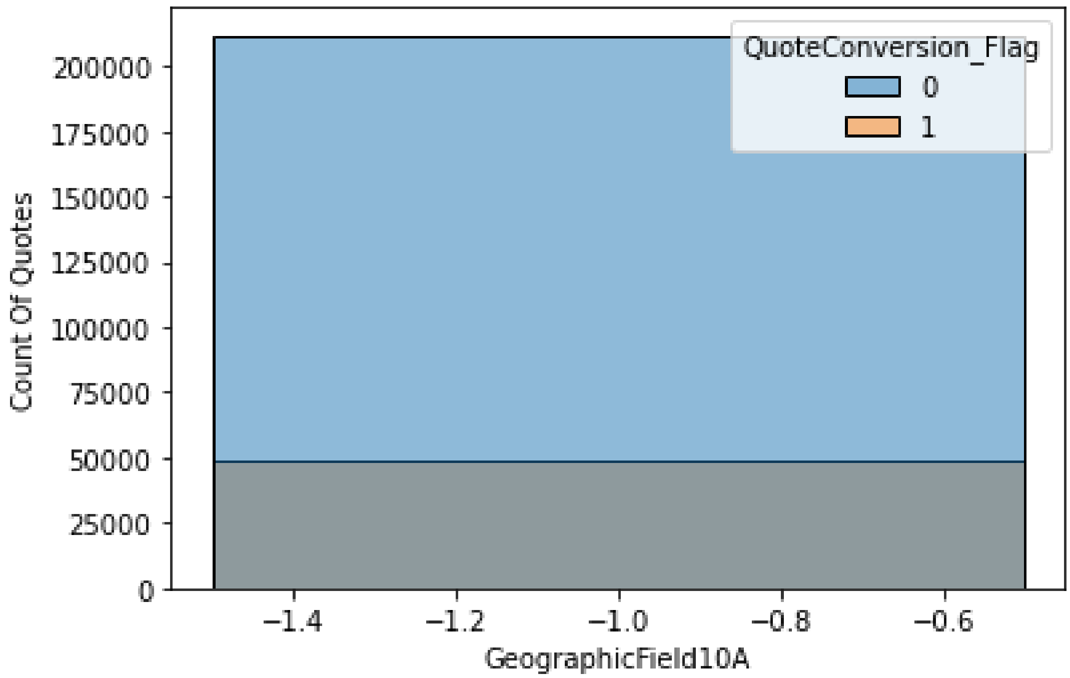
QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField8A
70000
60000
50000
40000
30000
20000
10000
0
0
5
10
15
20
25

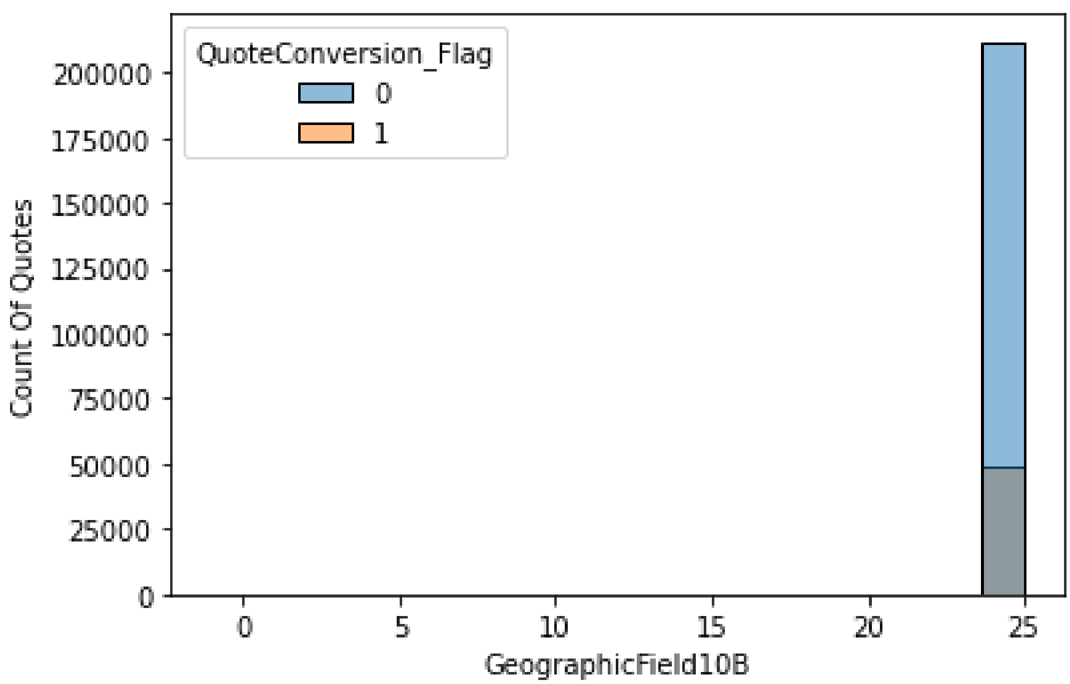


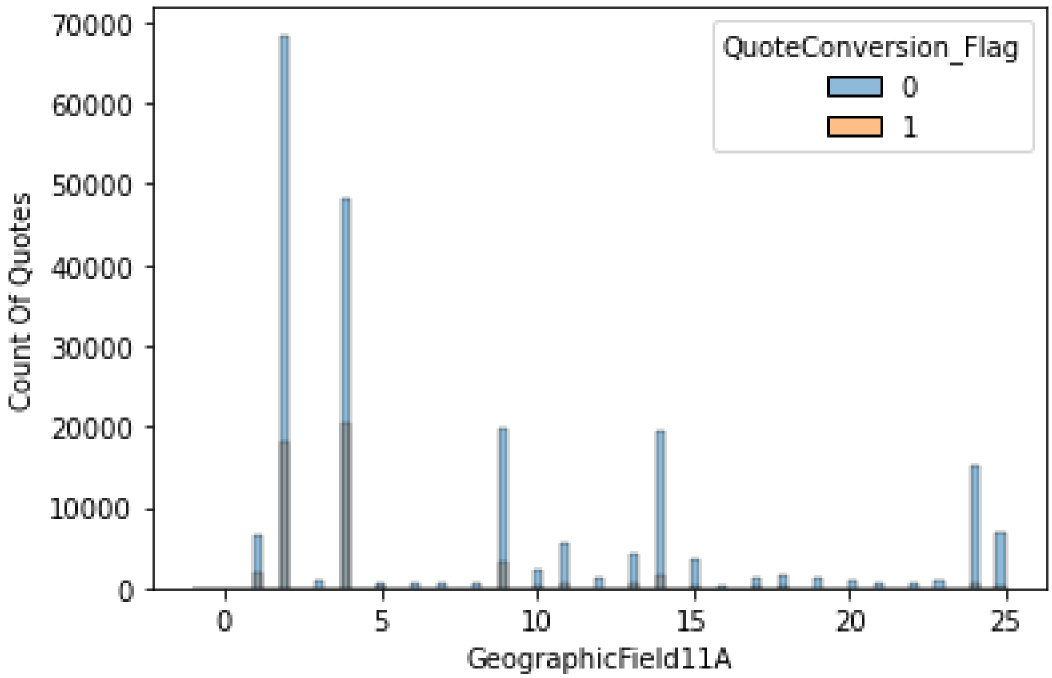
QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField8B
20000
17500
15000
12500
10000
7500
5000
2500
0
0
5
10
15
20
25

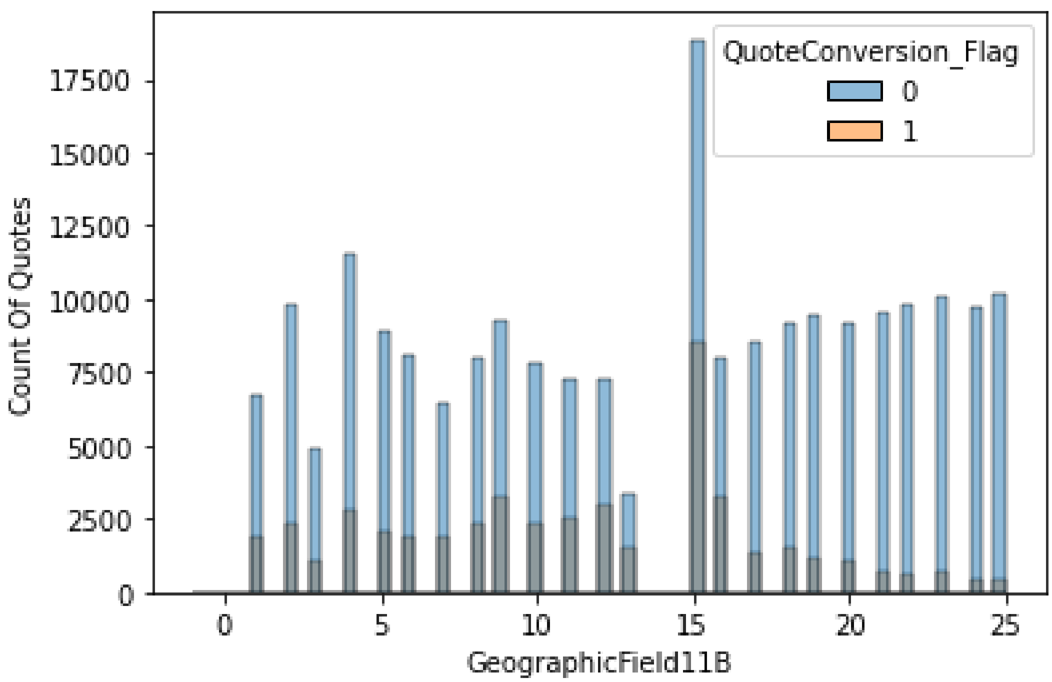

QuoteConversion_Flag
0
1
60000
50000
40000
30000
20000
10000
0
Count Of Quotes
0
5
10
15
20
25
GeographicField9A

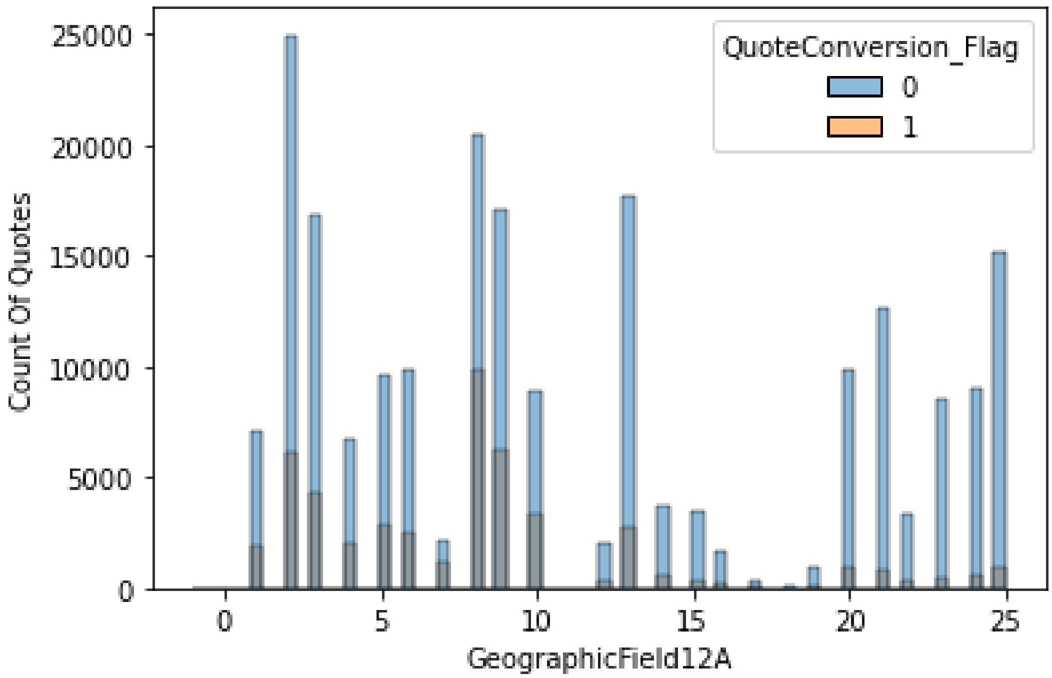

QuoteConversion_Flag
0
1
14000
12000
10000
8000
6000
4000
2000
0
Count Of Quotes
0
5
10
15
20
25
GeographicField9B

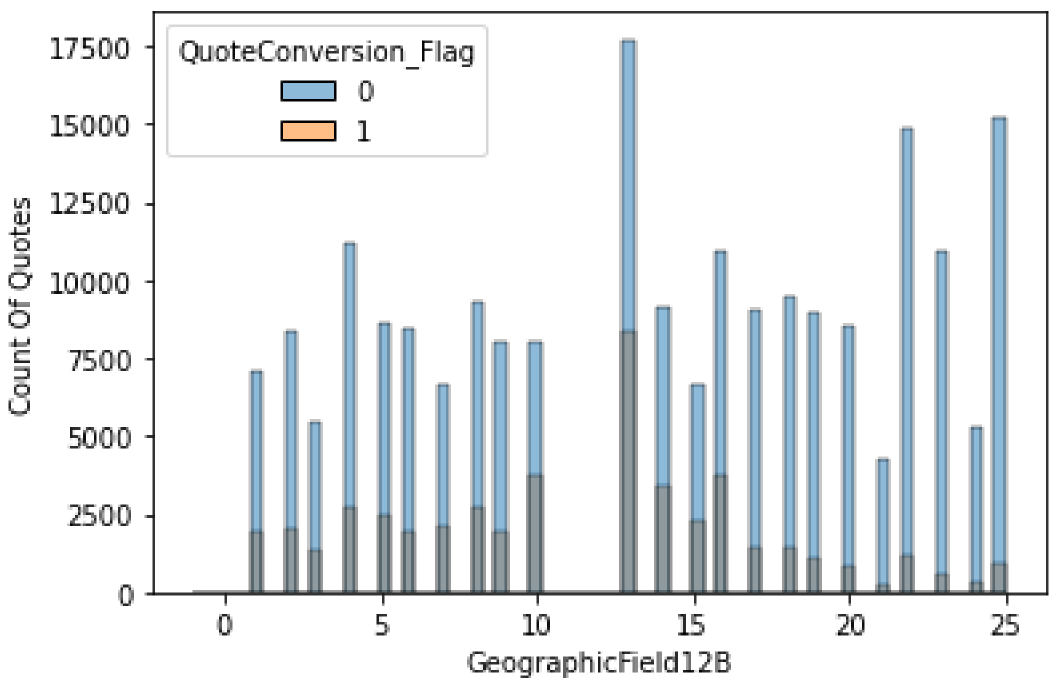

QuoteConversion_Flag
0
1
200000
175000
150000
125000
100000
75000
50000
25000
0
Count Of Quotes
−1.4
−1.2
−1.0
−0.8
−0.6
GeographicField10A

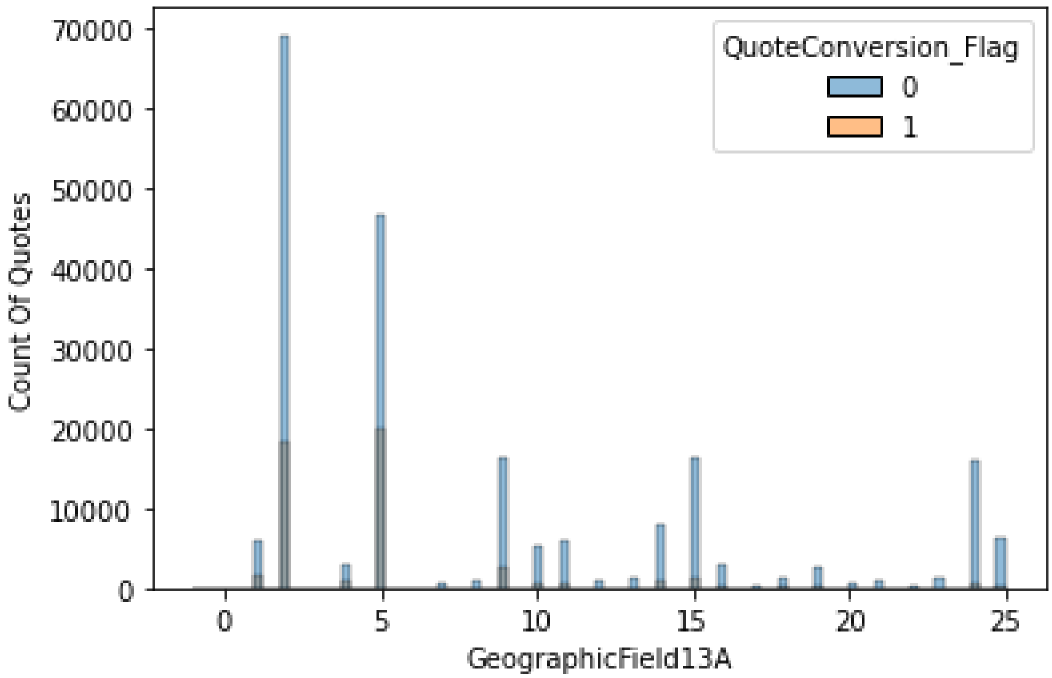

QuoteConversion_Flag
0
1
200000
175000
150000
125000
100000
75000
50000
25000
0
Count Of Quotes
0
5
10
15
20
25
GeographicField10B

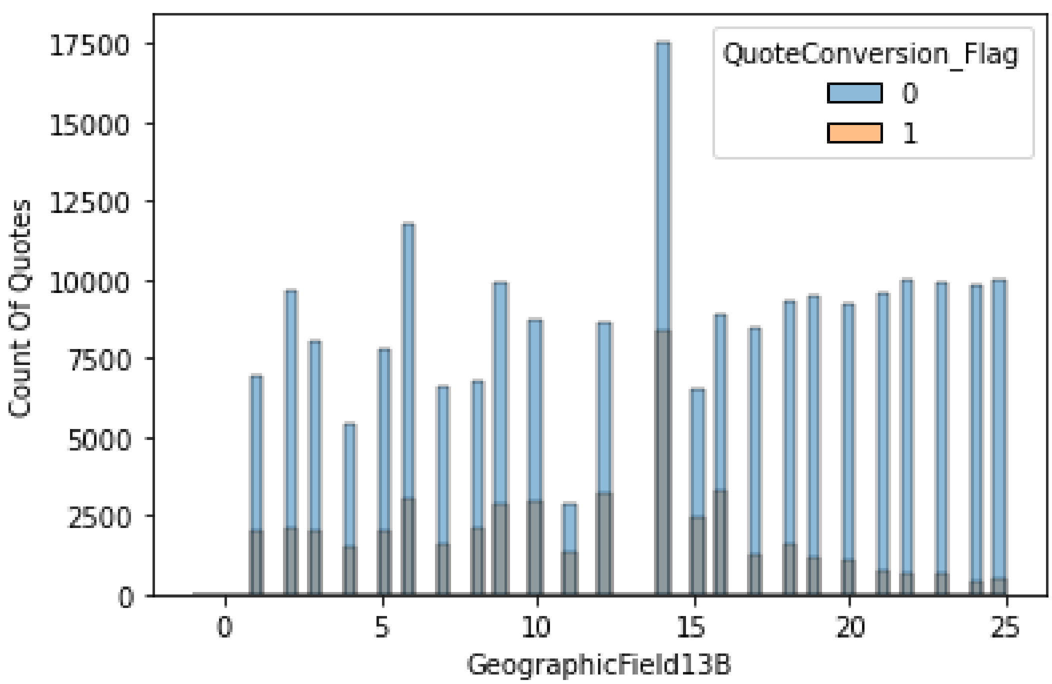

QuoteConversion_Flag
0
1
70000
60000
50000
40000
30000
20000
10000
0
Count Of Quotes
0
5
10
15
20
25
GeographicField11A

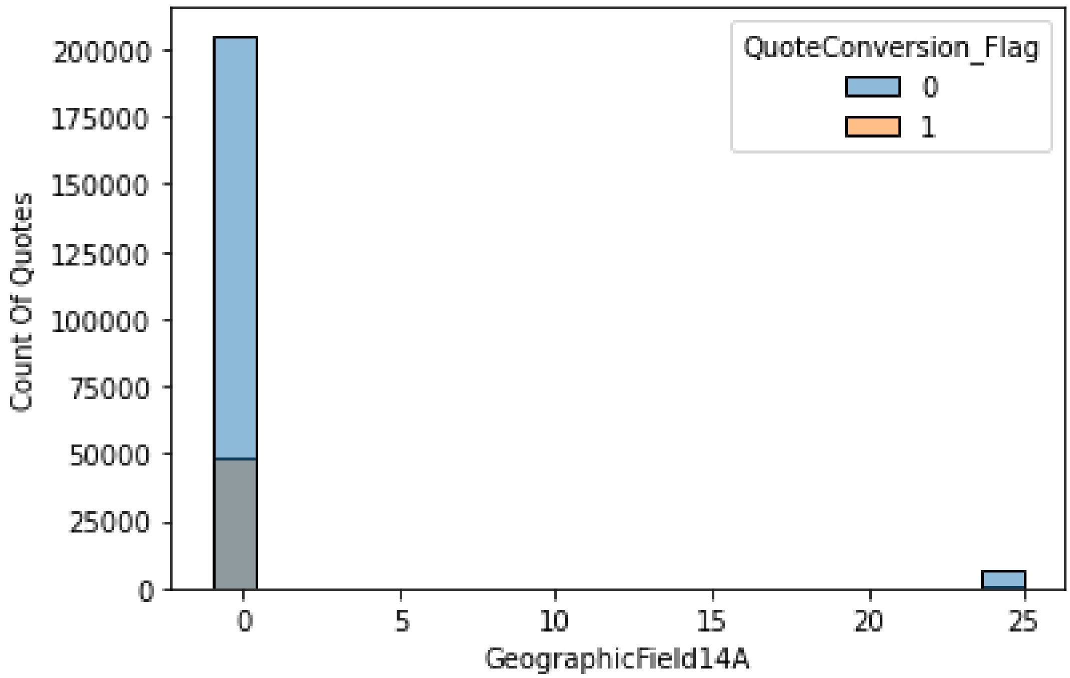

QuoteConversion_Flag
0
1
17500
15000
12500
10000
7500
5000
2500
0
Count Of Quotes
0
5
10
15
20
25
GeographicField11B


QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField12A
0
5000
10000
15000
20000
25000
0
5
10
15
20
25

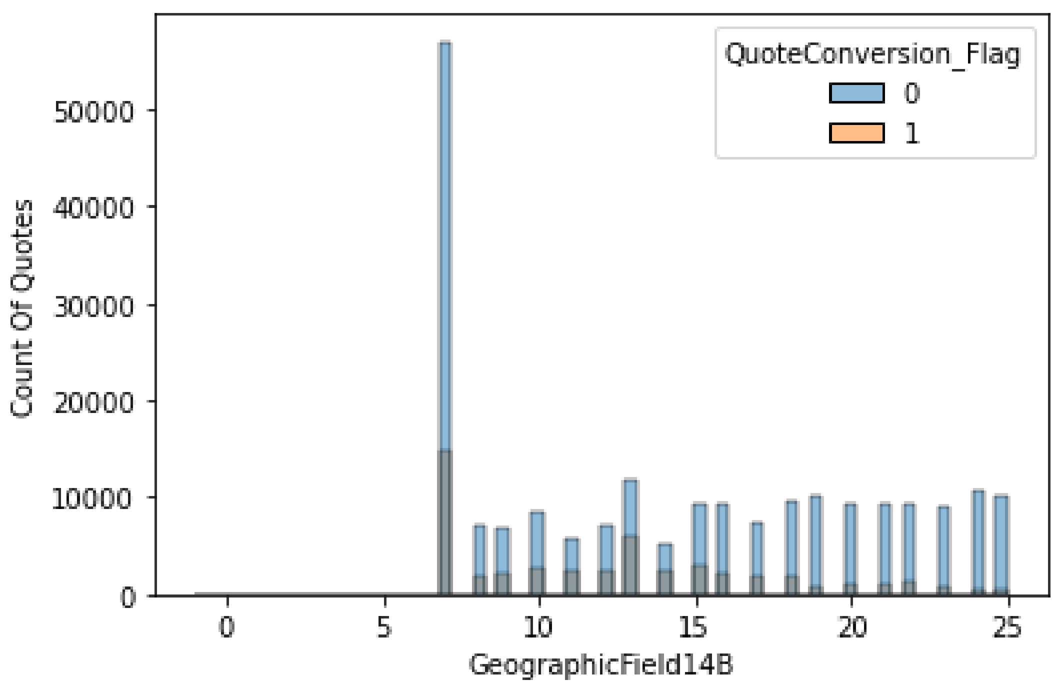

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField12B
0
2500
5000
7500
10000
12500
15000
17500
0
5
10
15
20
25


QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField13A

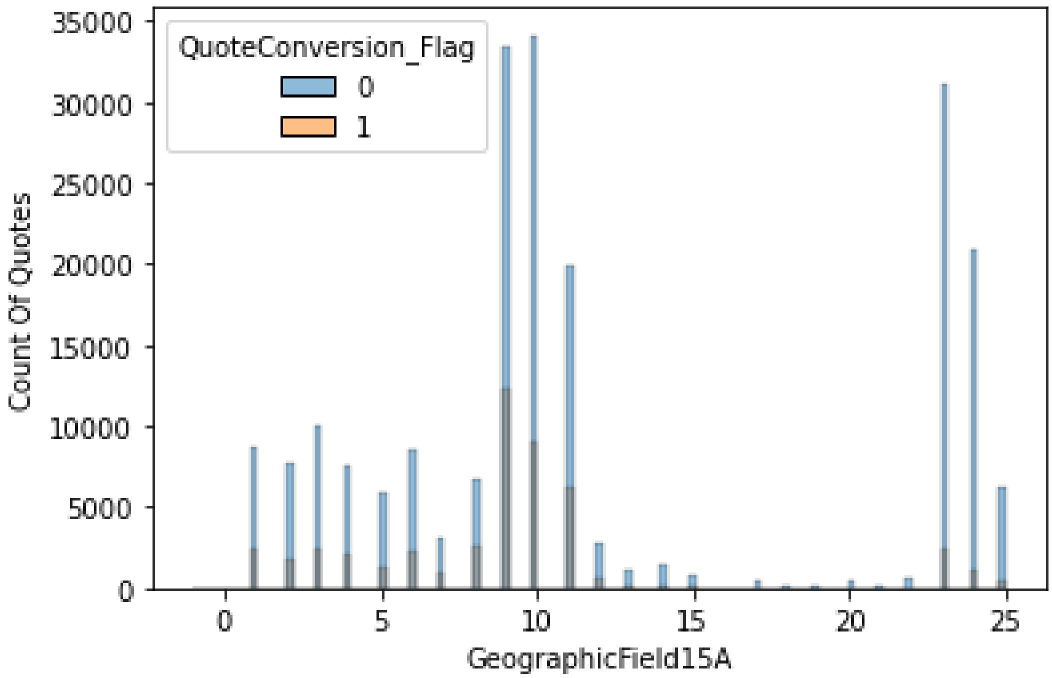

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField13B

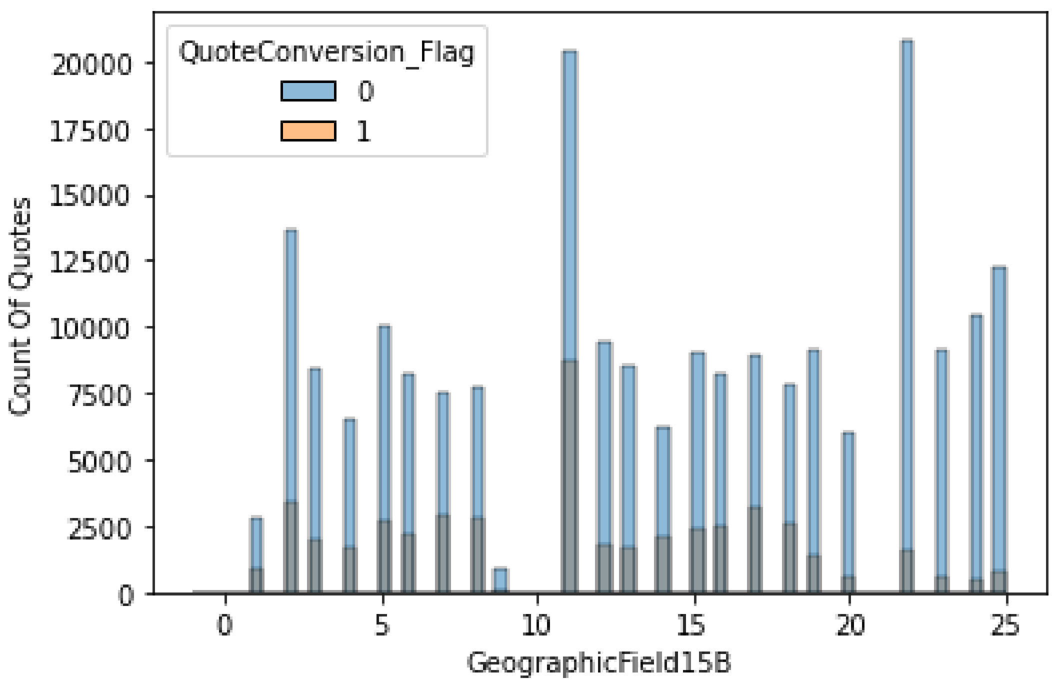


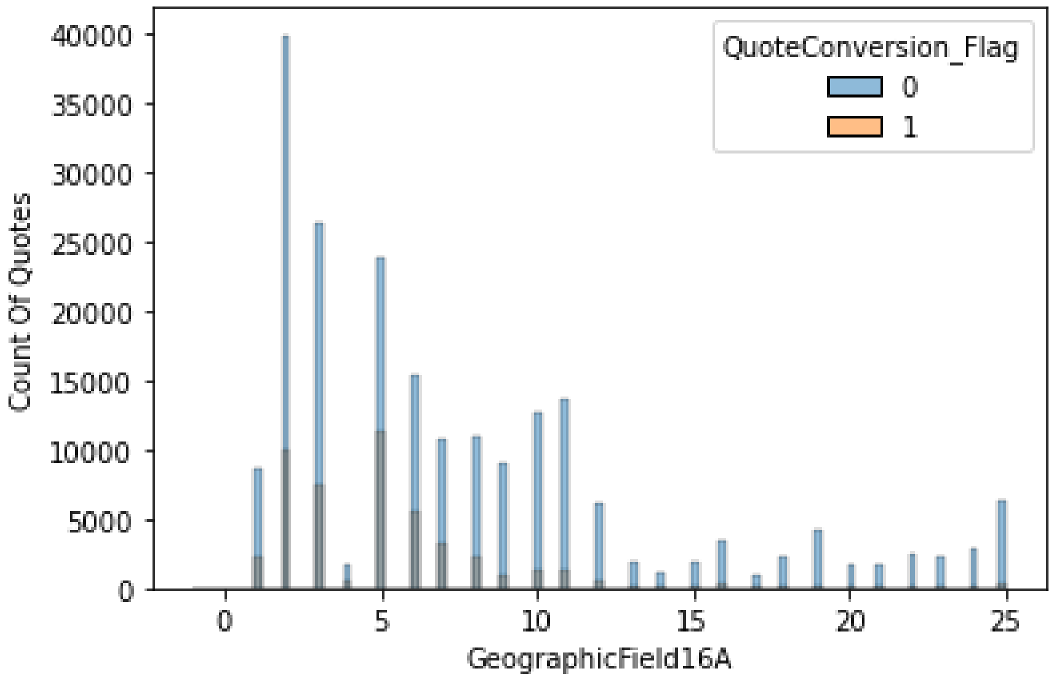
QuoteConversion_Flag
0
1
200000
175000
150000
125000
100000
75000
50000
25000
0
Count Of Quotes
0
5
10
15
20
25
GeographicField14A

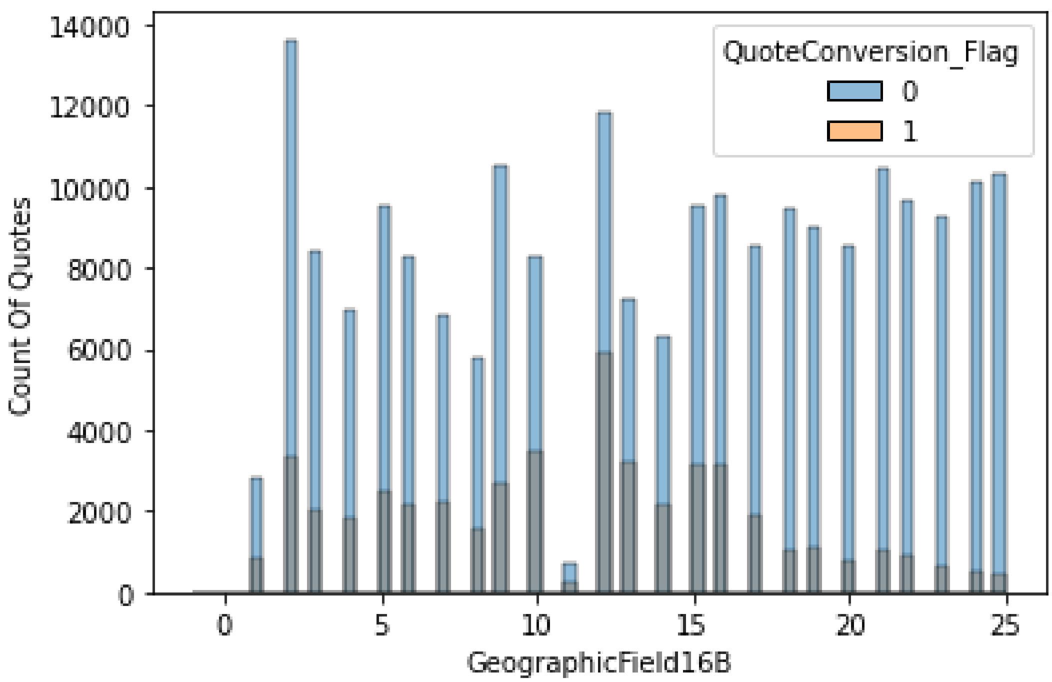


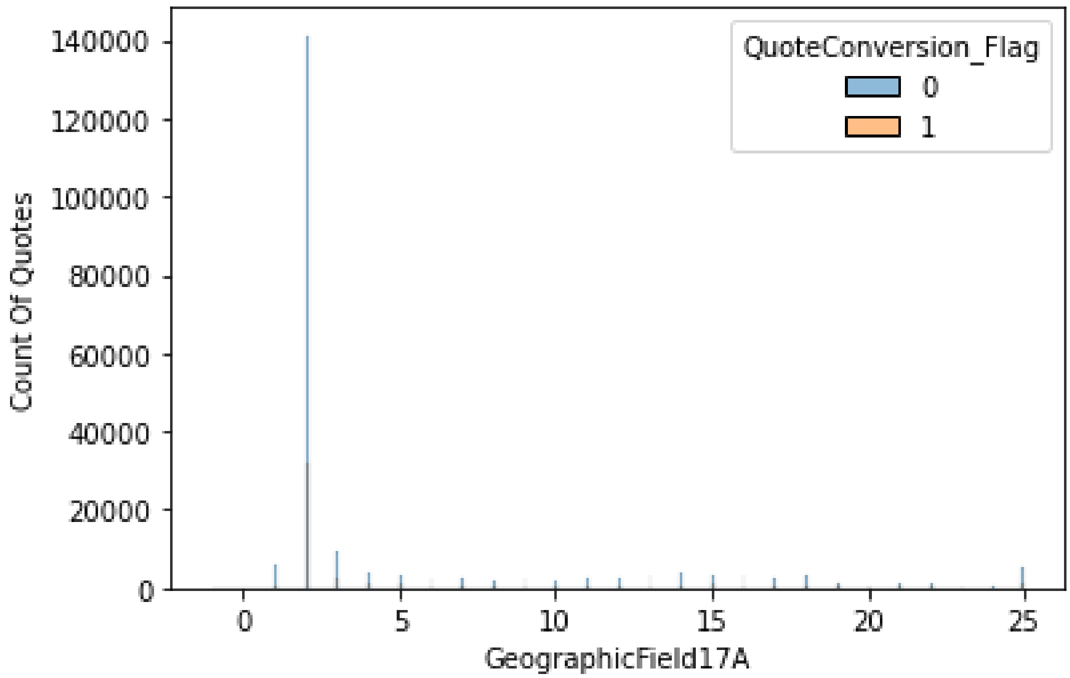
QuoteConversion_Flag
0
1
50000
40000
30000
20000
10000
0
Count Of Quotes
0
5
10
15
20
25
GeographicField14B

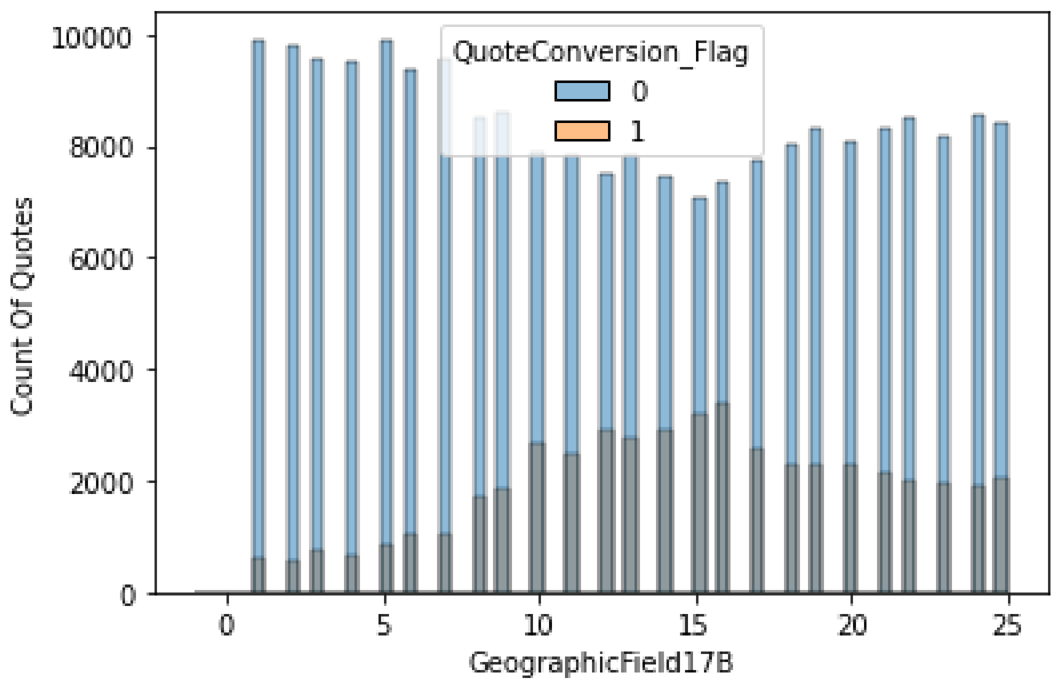

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField15A
35000
30000
25000
20000
15000
10000
5000
0
0
5
10
15
20
25

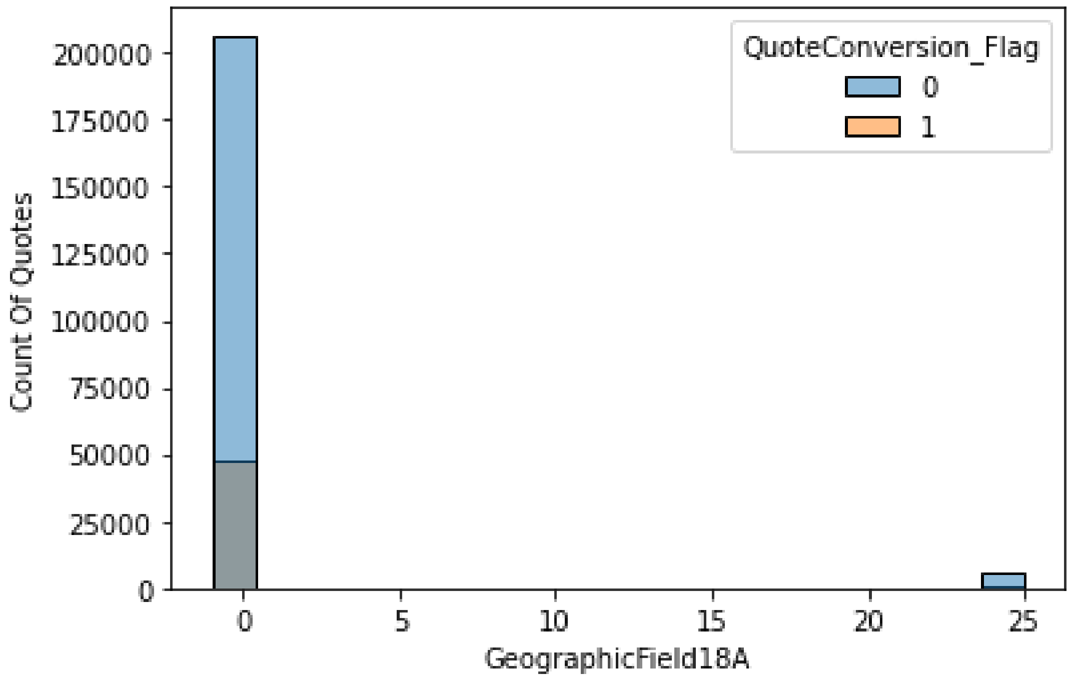

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField15B
20000
17500
15000
12500
10000
7500
5000
2500
0
0
5
10
15
20
25

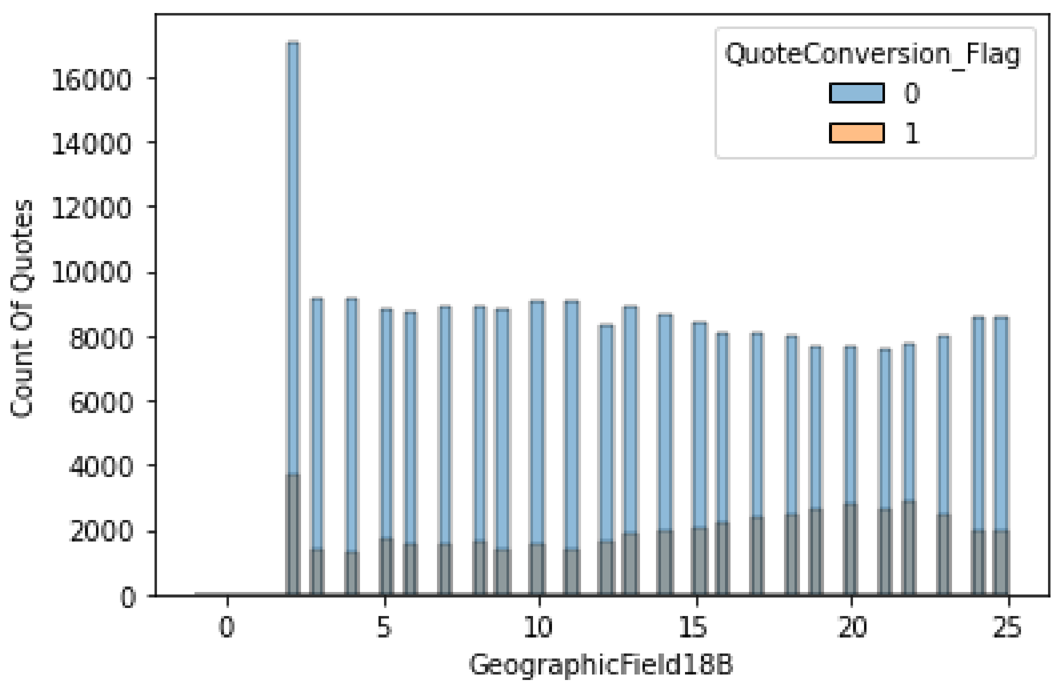

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField16A
0
5000
10000
15000
20000
25000
30000
35000
40000
0
5
10
15
20
25

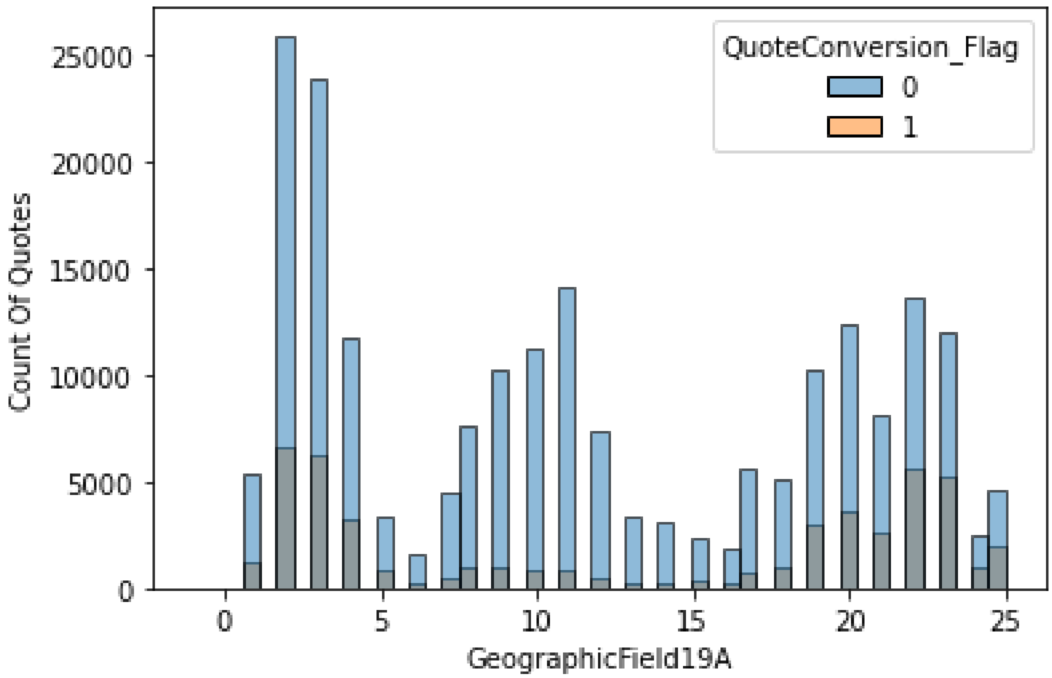

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField16B
0
2000
4000
6000
8000
10000
12000
14000
0
5
10
15
20
25

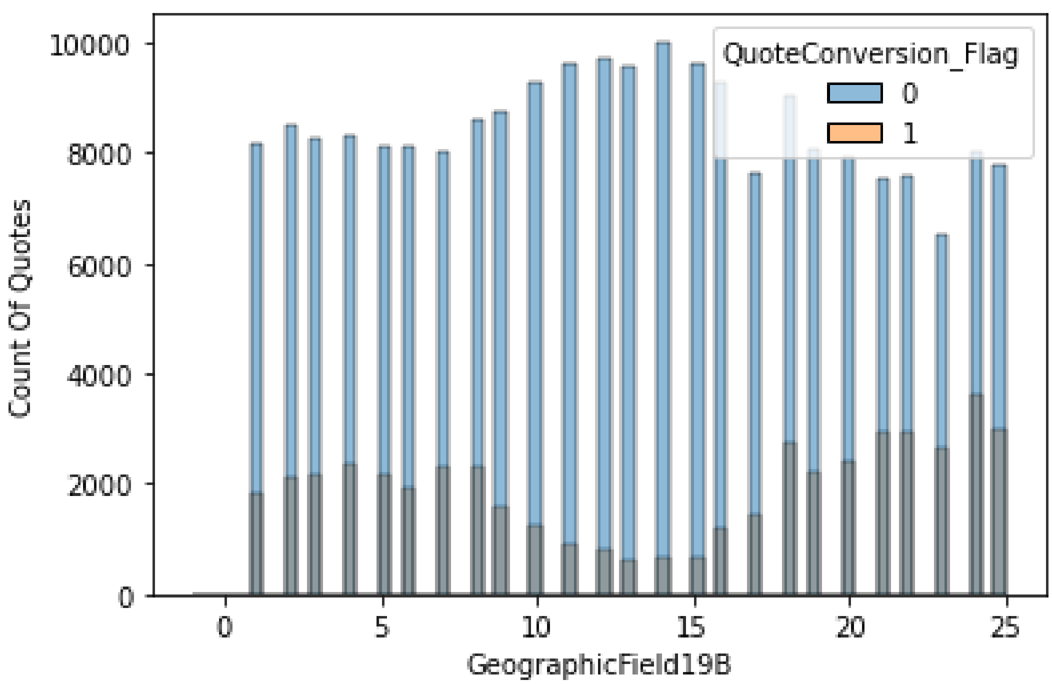

QuoteConversion_Flag
0
1
Count Of Quotes
140000
120000
100000
80000
60000
40000
20000
0
0
5
10
15
20
25
GeographicField17A

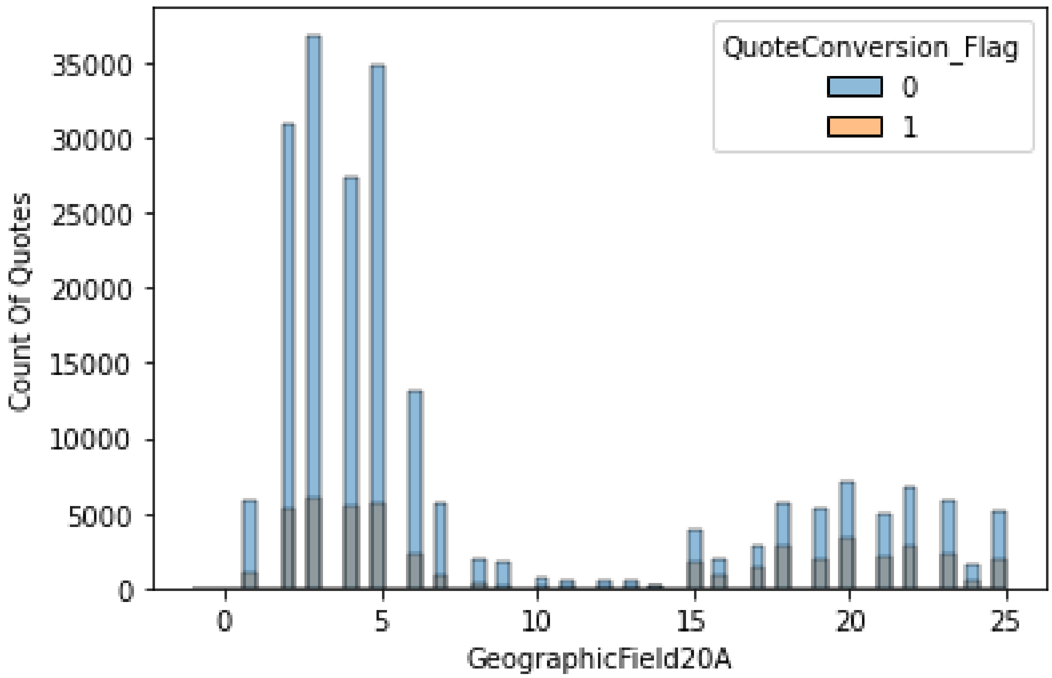

QuoteConversion_Flag
0
1
Count Of Quotes
10000
8000
6000
4000
2000
0
0
5
10
15
20
25
GeographicField17B

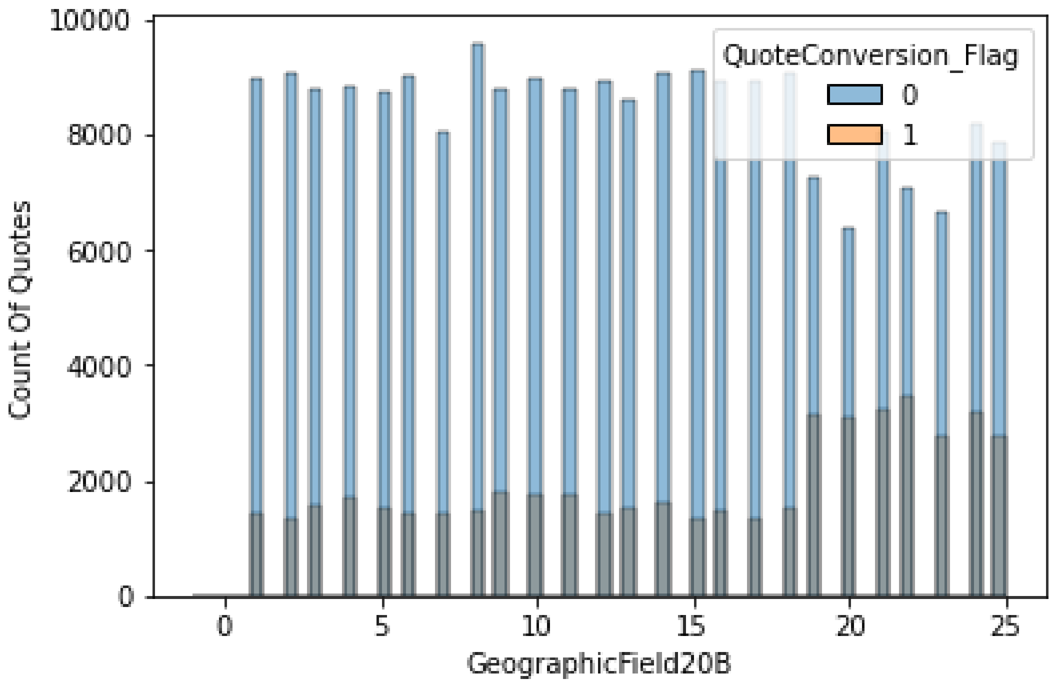

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField18A
0
25000
50000
75000
100000
125000
150000
175000
200000
0
5
10
15
20
25

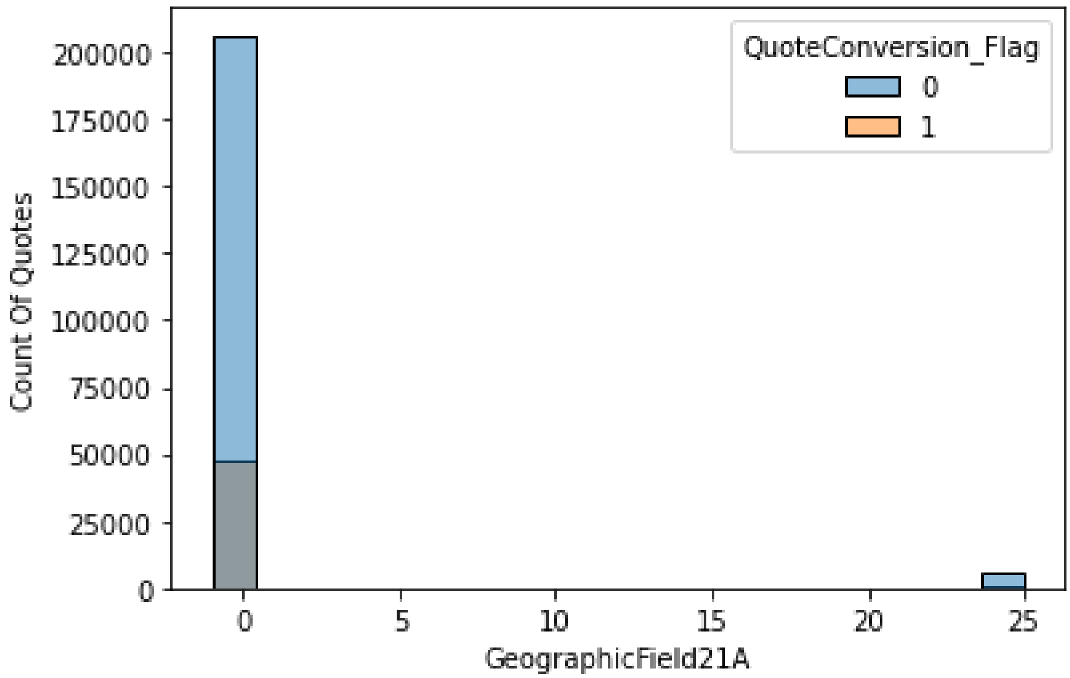

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField18B
0
2000
4000
6000
8000
10000
12000
14000
16000
0
5
10
15
20
25

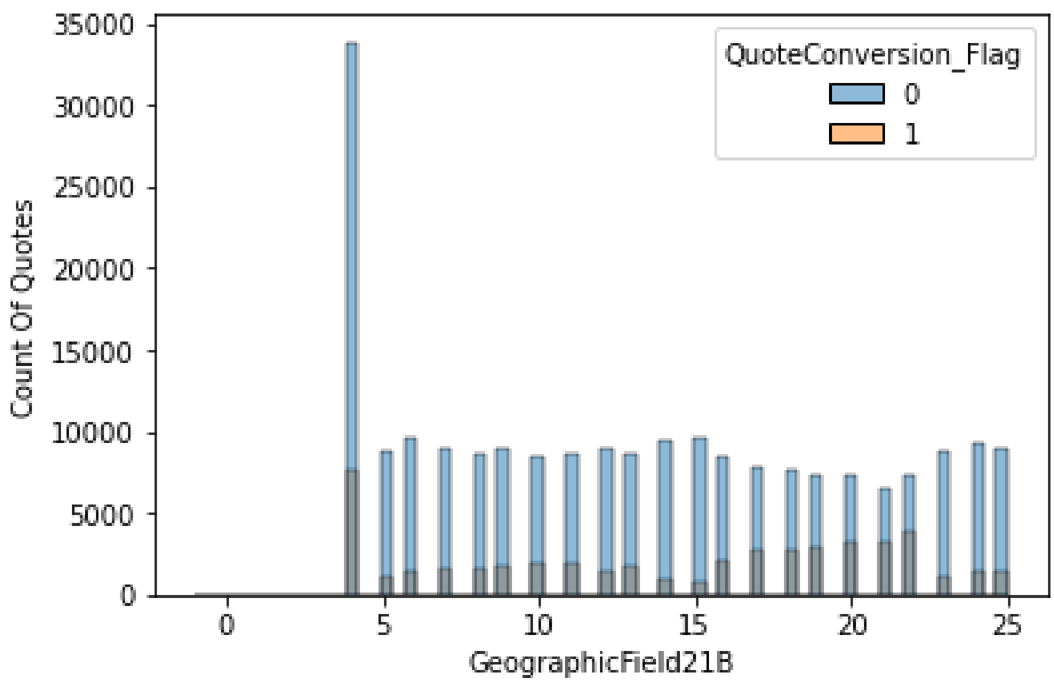

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField19A
0
5000
10000
15000
20000
25000
0
5
10
15
20
25

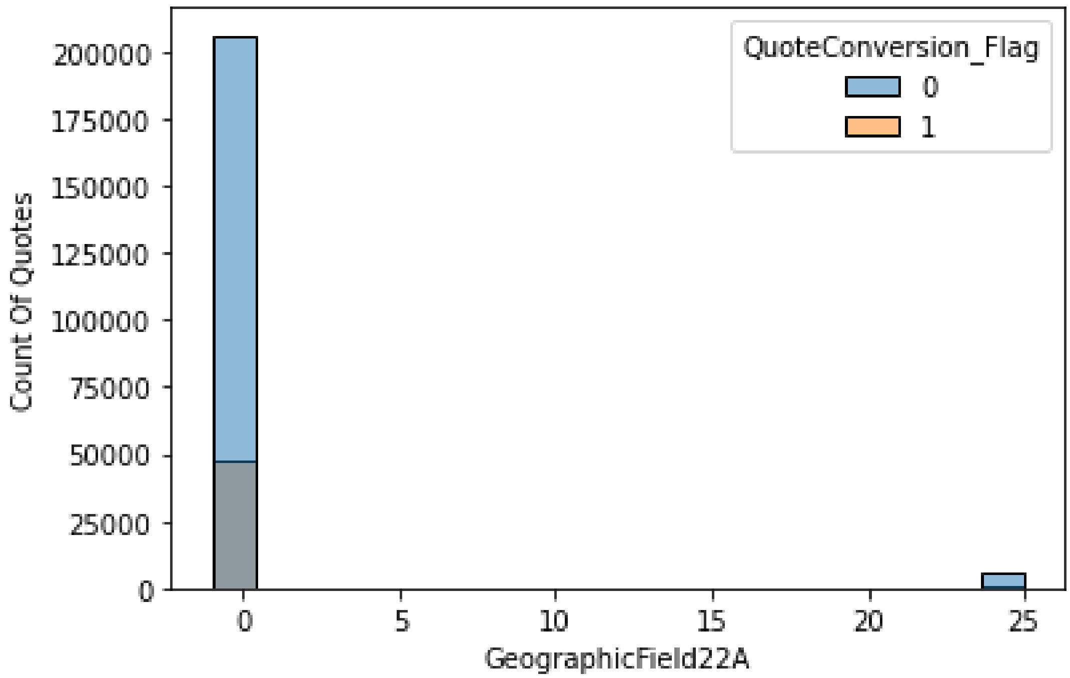

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField19B
0
2000
4000
6000
8000
10000
0
5
10
15
20
25

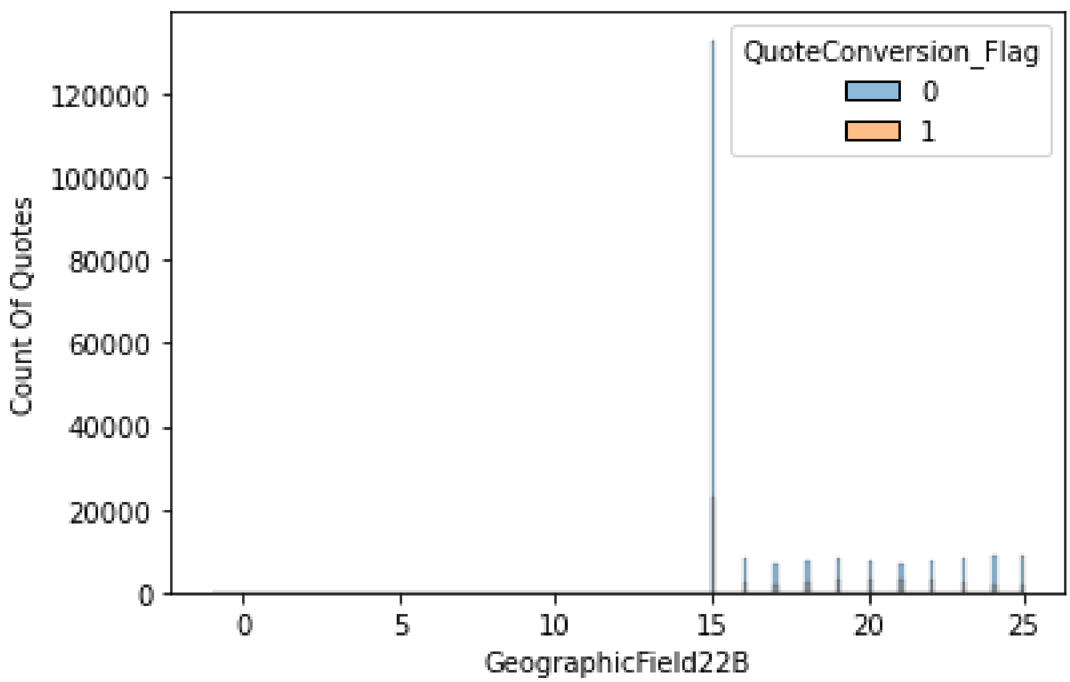

QuoteConversion_Flag
0
1
Count Of Quotes
35000
30000
25000
20000
15000
10000
5000
0
0
5
10
15
20
25
GeographicField20A

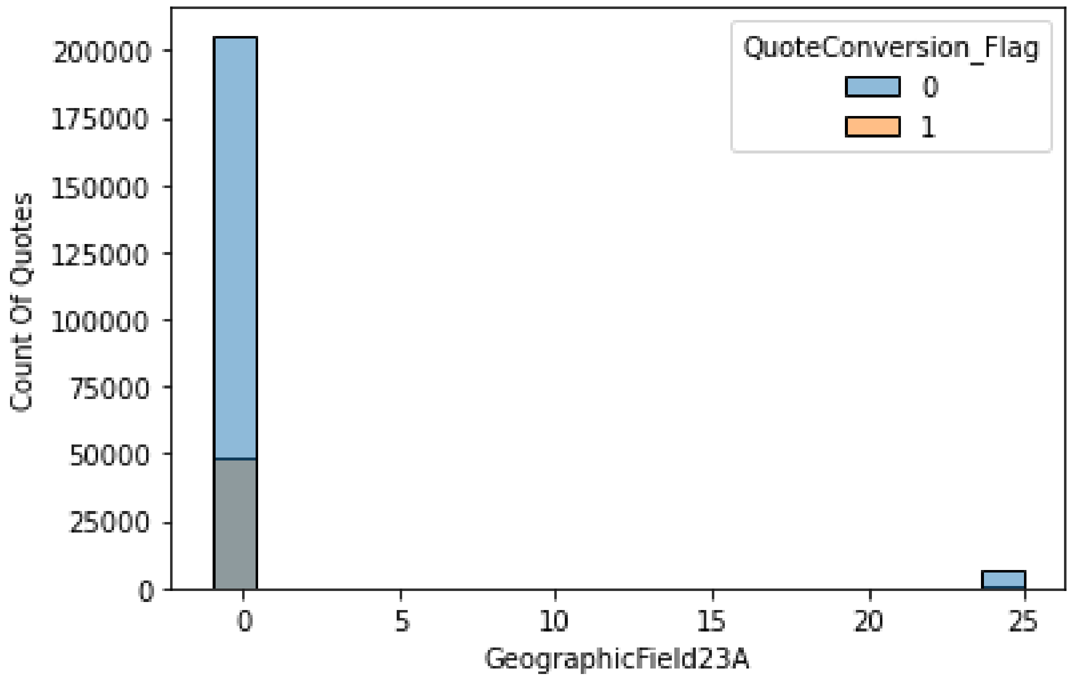

QuoteConversion_Flag
0
1
Count Of Quotes
10000
8000
6000
4000
2000
0
0
5
10
15
20
25
GeographicField20B

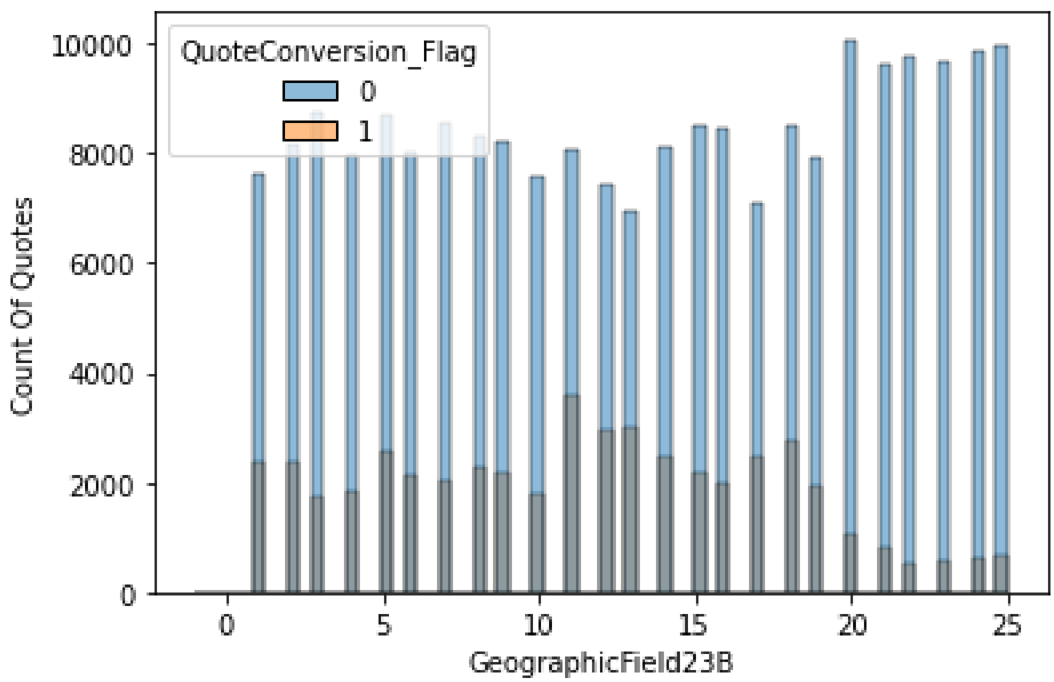

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField21A
200000
175000
150000
125000
100000
75000
50000
25000
0
0
5
10
15
20
25

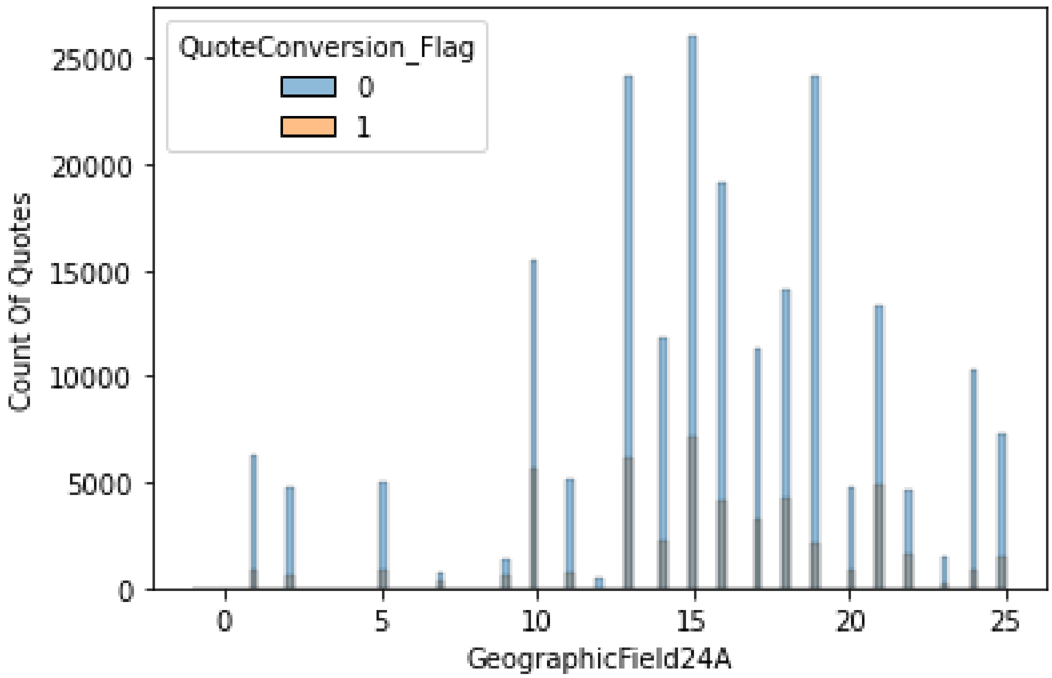

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField21B
35000
30000
25000
20000
15000
10000
5000
0
0
5
10
15
20
25

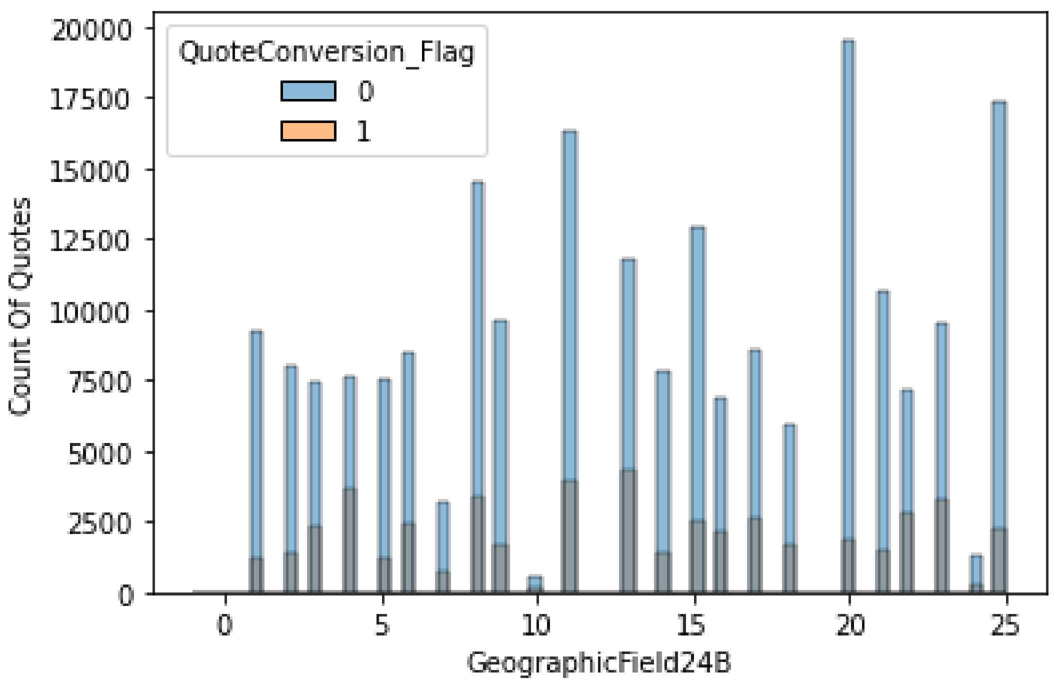

QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
0
5
10
15
20
25
GeographicField22A

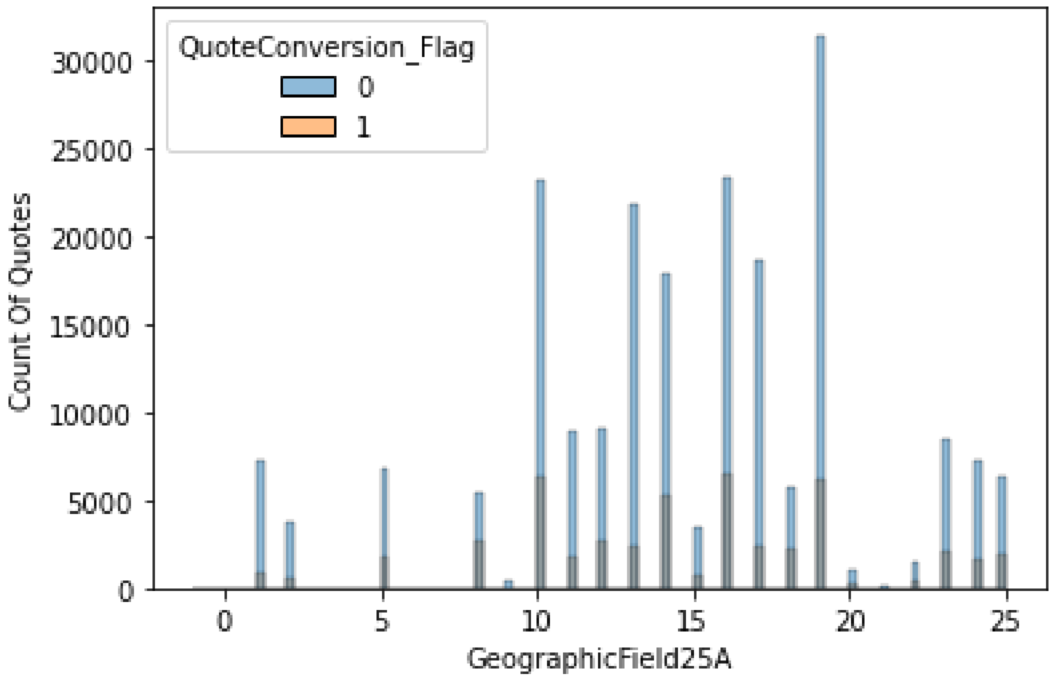

QuoteConversion_Flag
0
1
Count Of Quotes
120000
100000
80000
60000
40000
20000
0
0
5
10
15
20
25
GeographicField22B

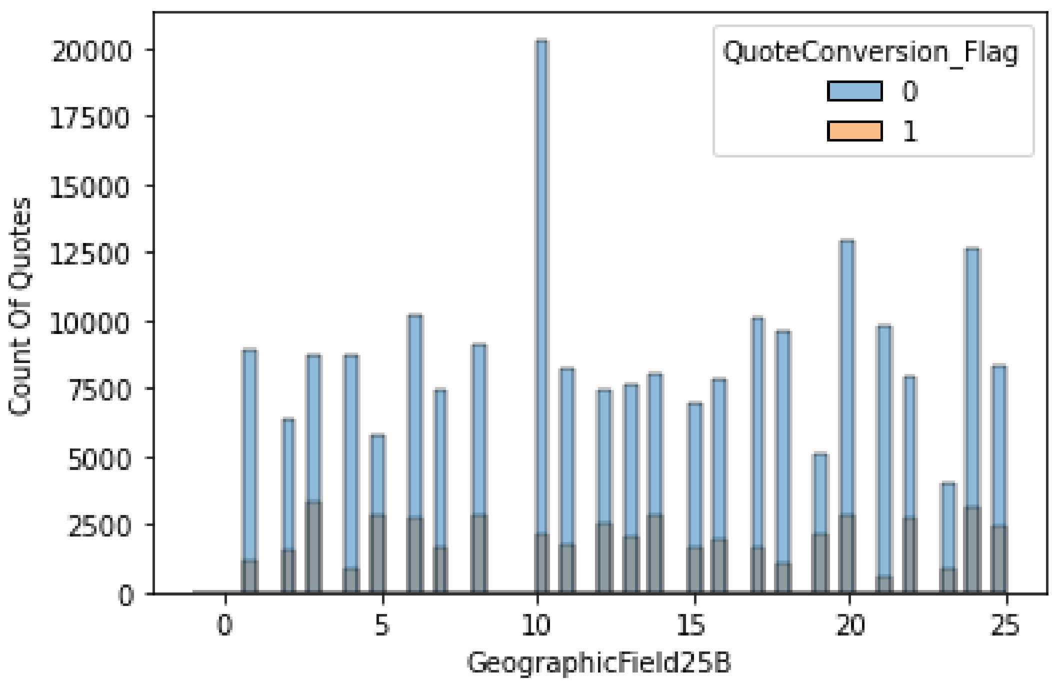

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField23A
0
25000
50000
75000
100000
125000
150000
175000
200000
5
10
15
20
25

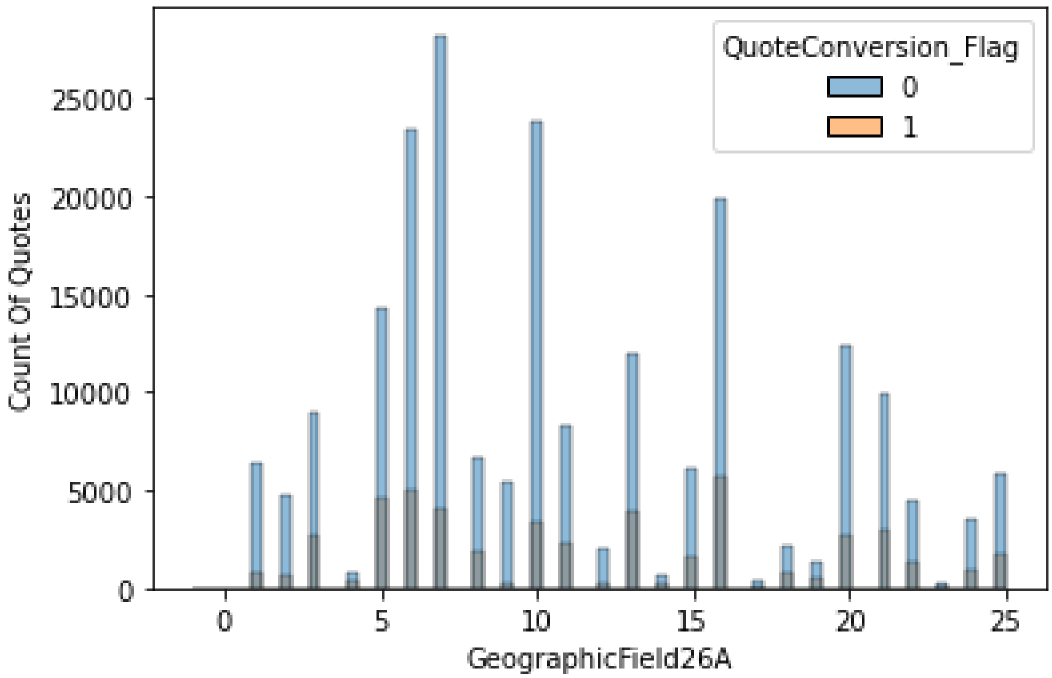

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField23B
0
2000
4000
6000
8000
10000
5
10
15
20
25

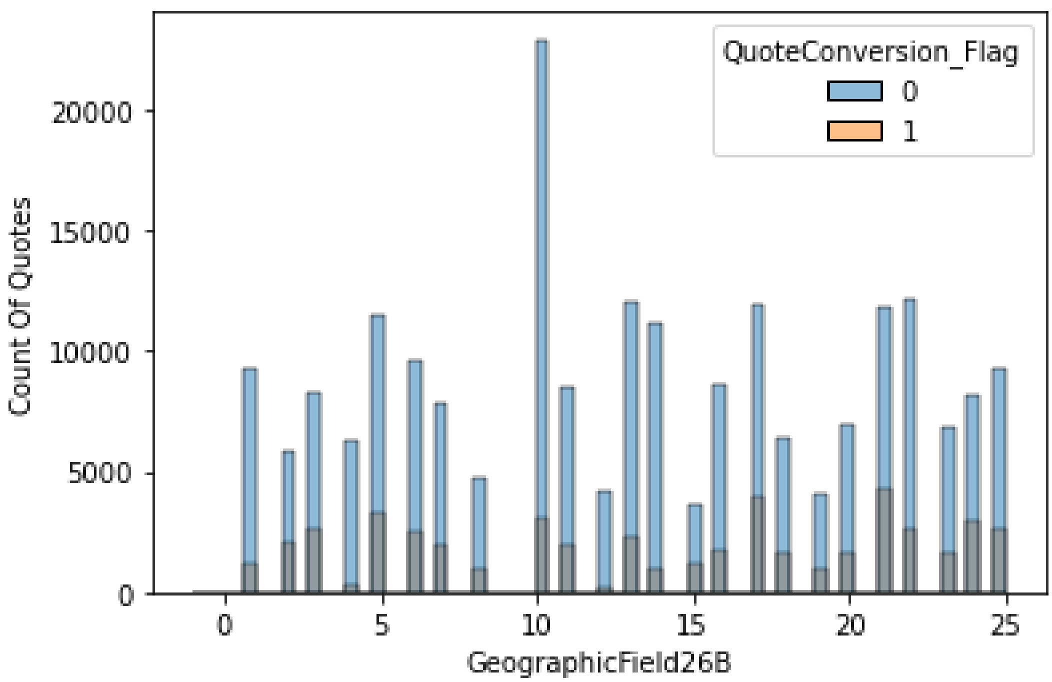

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField24A
0
5000
10000
15000
20000
25000
0
5
10
15
20
25

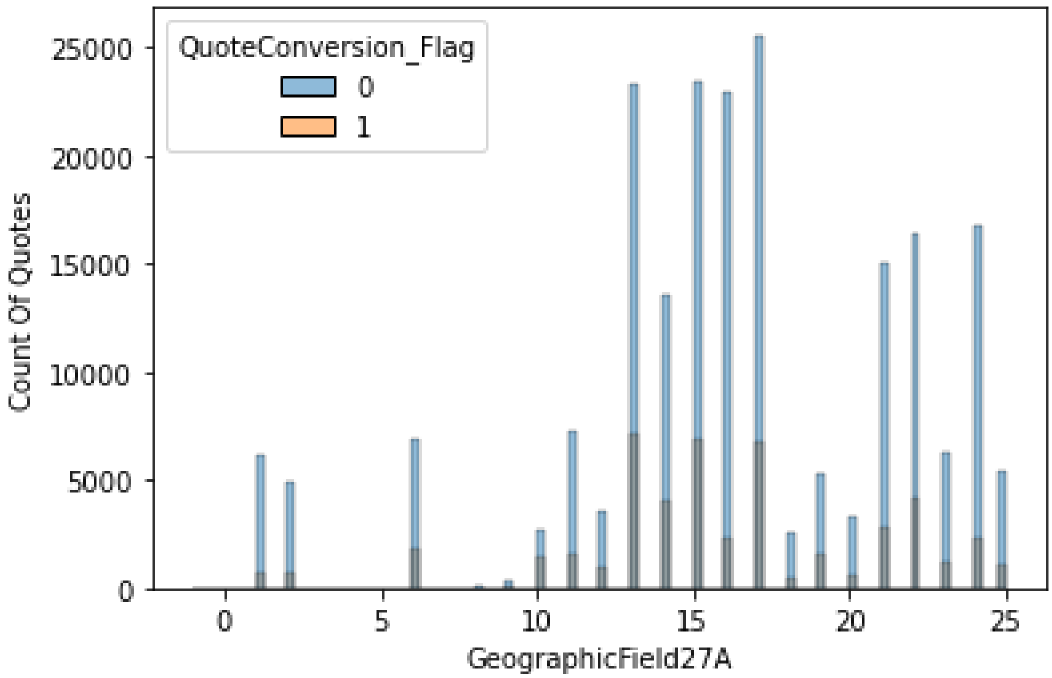

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField24B
0
2500
5000
7500
10000
12500
15000
17500
20000
0
5
10
15
20
25

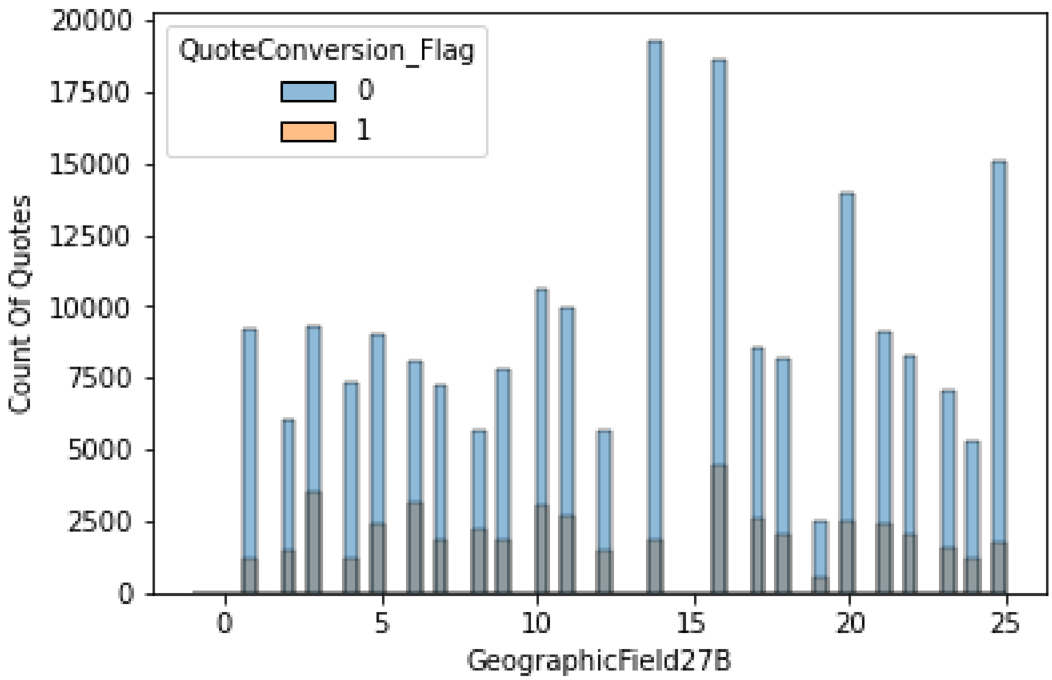

QuoteConversion_Flag
0
1
Count Of Quotes
30000
25000
20000
15000
10000
5000
0
0
5
10
15
20
25
GeographicField25A

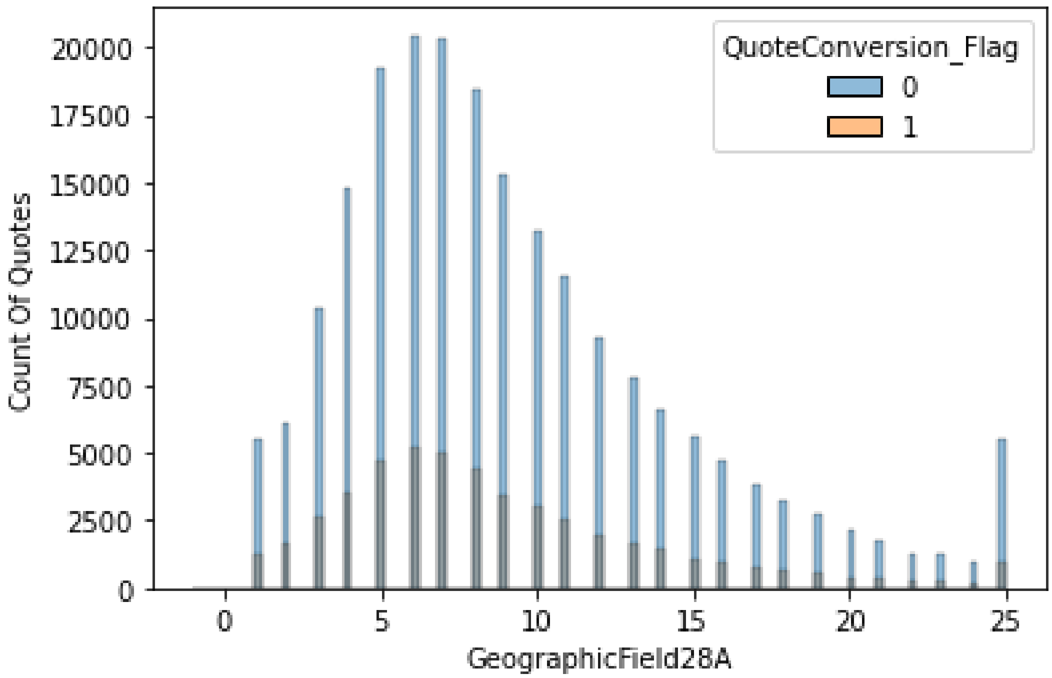

QuoteConversion_Flag
0
1
Count Of Quotes
20000
17500
15000
12500
10000
7500
5000
2500
0
0
5
10
15
20
25
GeographicField25B

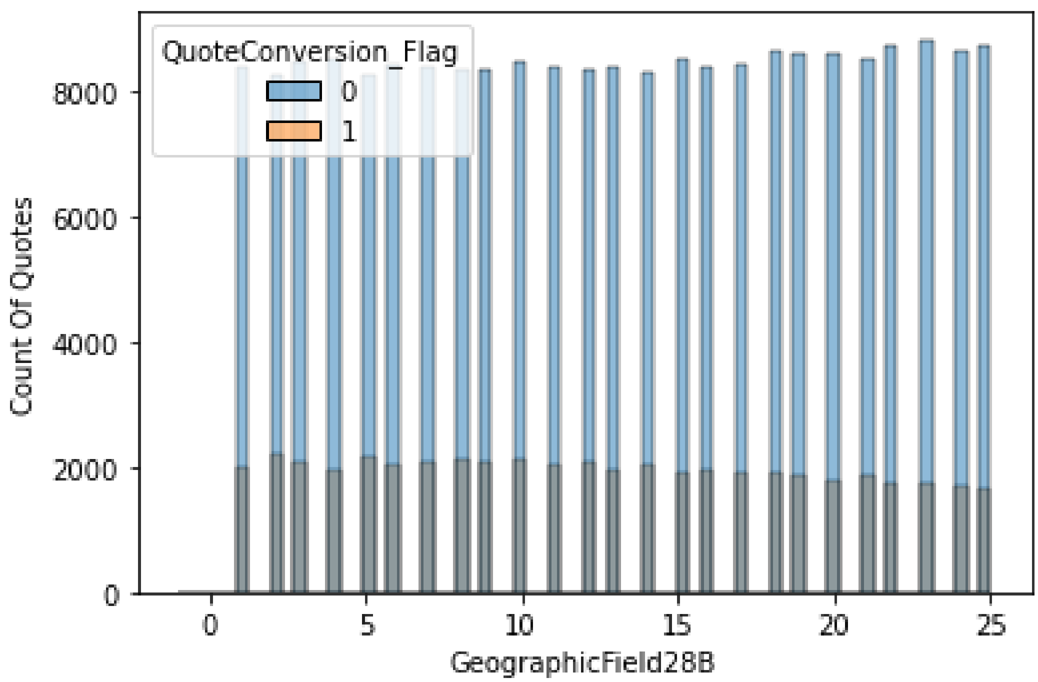

QuoteConversion_Flag
0
1
25000
20000
15000
10000
5000
0
Count Of Quotes
0
5
10
15
20
25
GeographicField26A

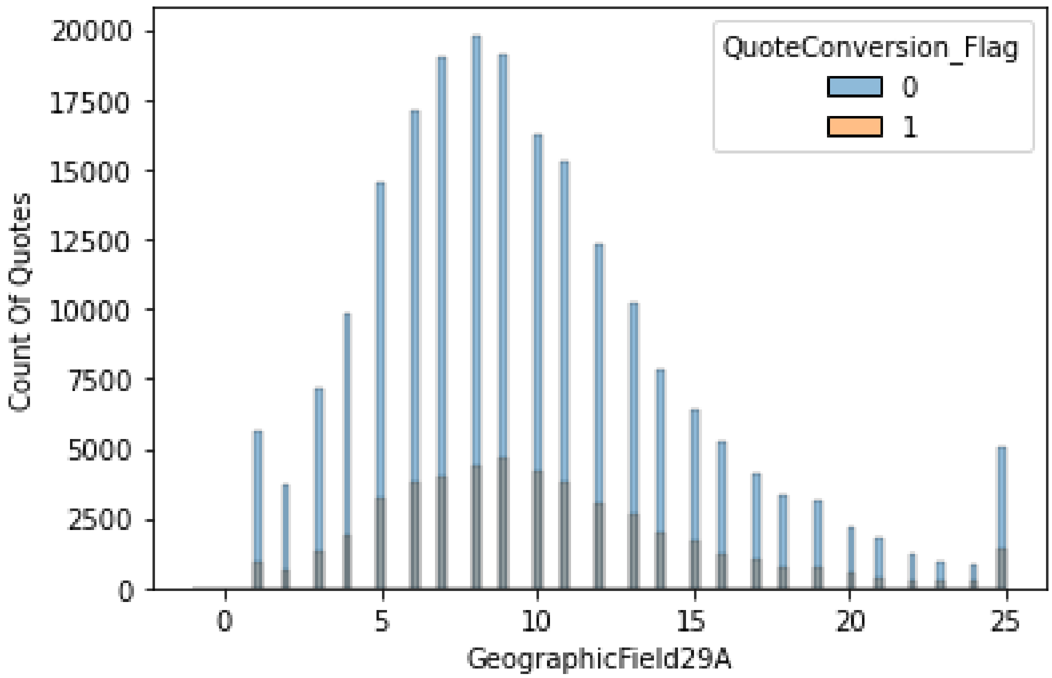

QuoteConversion_Flag
0
1
20000
15000
10000
5000
0
Count Of Quotes
0
5
10
15
20
25
GeographicField26B


QuoteConversion_Flag
0
1
Count Of Quotes
25000
20000
15000
10000
5000
0
0
5
10
15
20
25
GeographicField27A

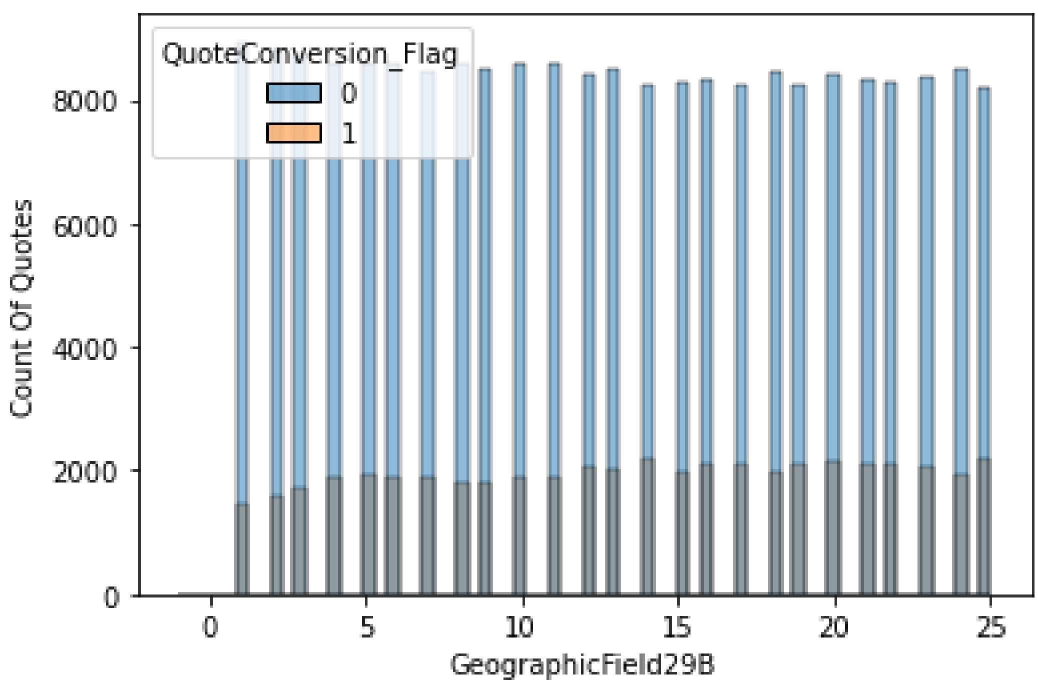

QuoteConversion_Flag
0
1
Count Of Quotes
20000
17500
15000
12500
10000
7500
5000
2500
0
0
5
10
15
20
25
GeographicField27B


QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField28A
0
5
10
15
20
25
0
2500
5000
7500
10000
12500
15000
17500
20000

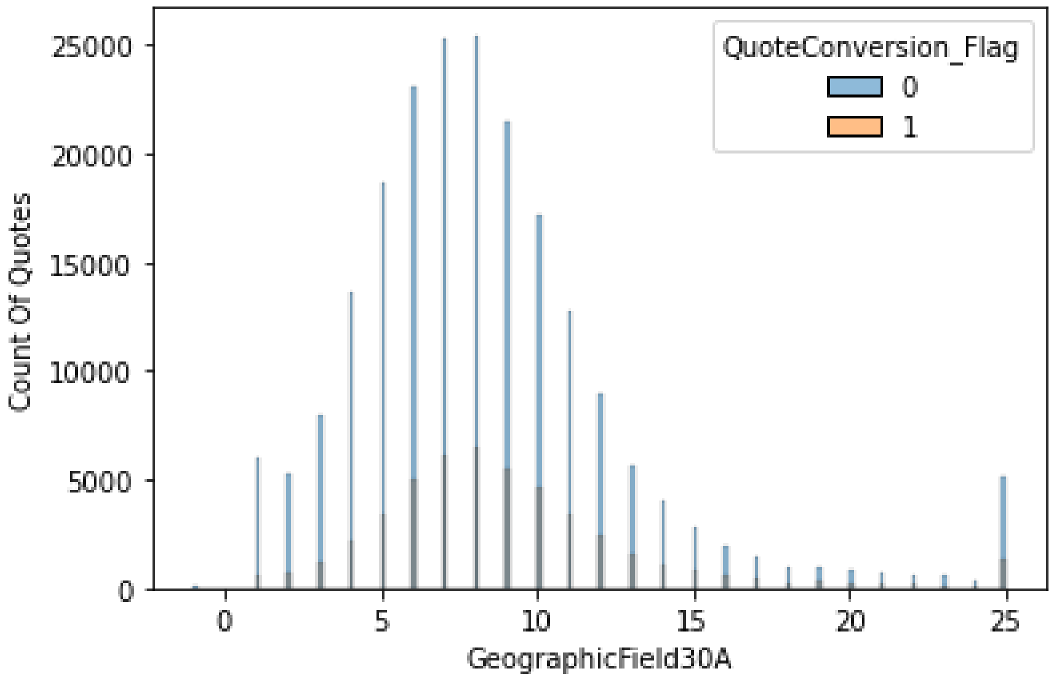

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField28B
0
5
10
15
20
25
0
2000
4000
6000
8000

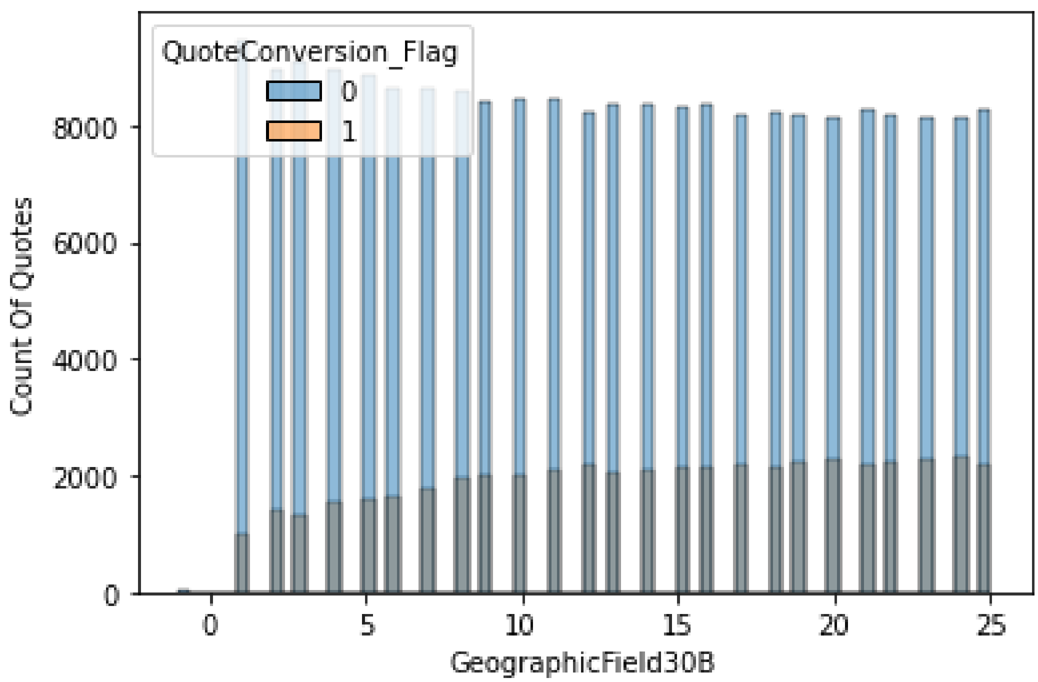


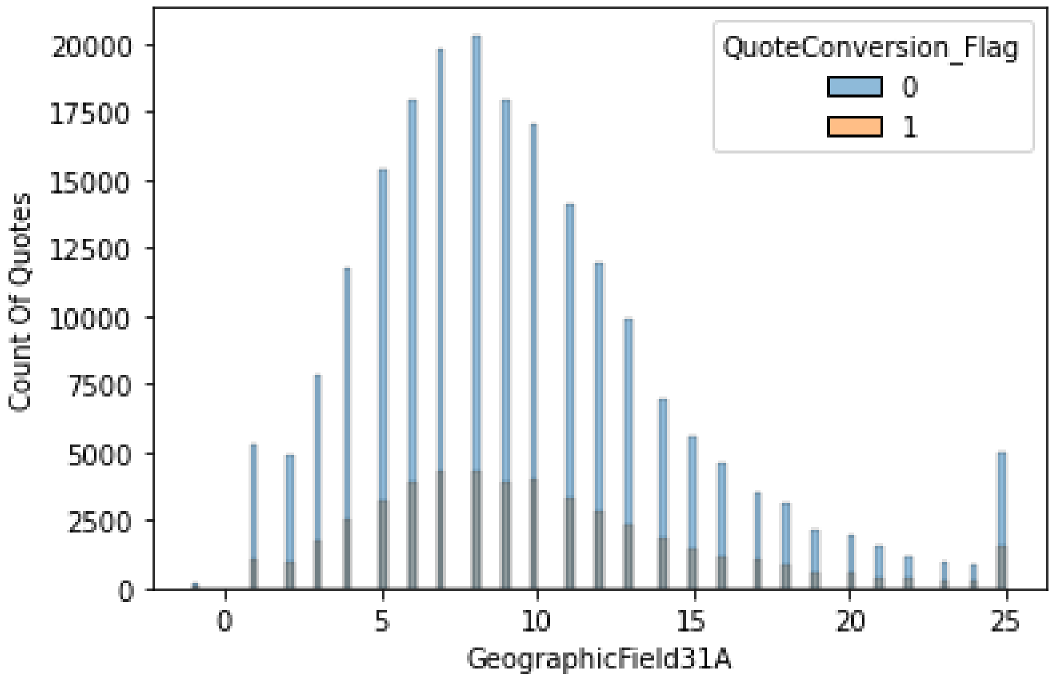
QuoteConversion_Flag
0
1
Count Of Quotes
20000
17500
15000
12500
10000
7500
5000
2500
0
0
5
10
15
20
25
GeographicField29A

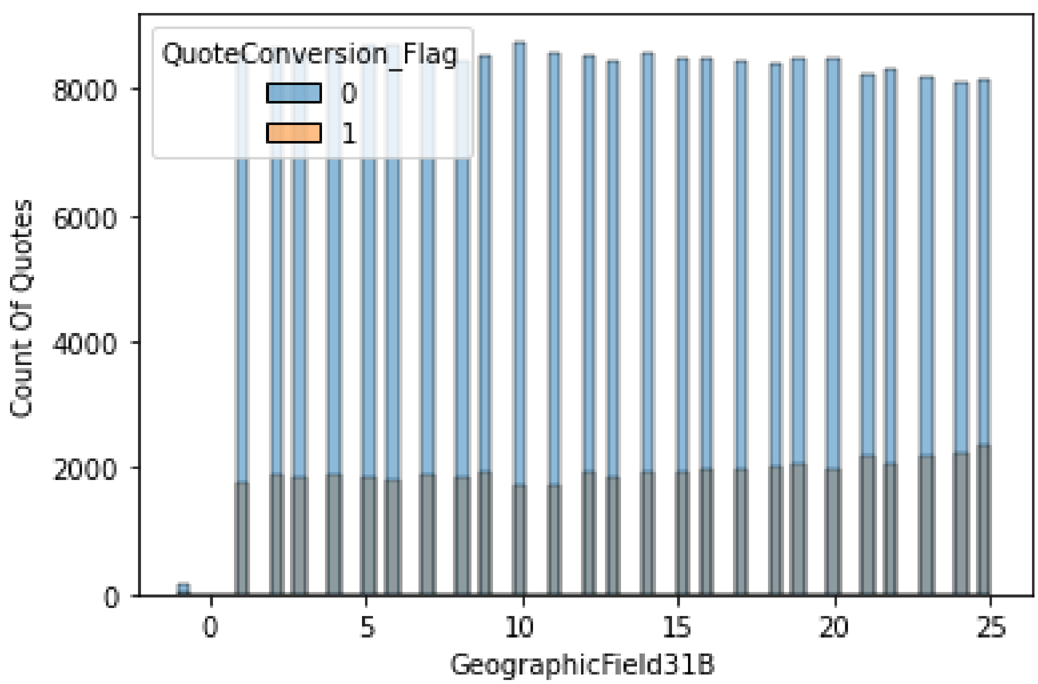


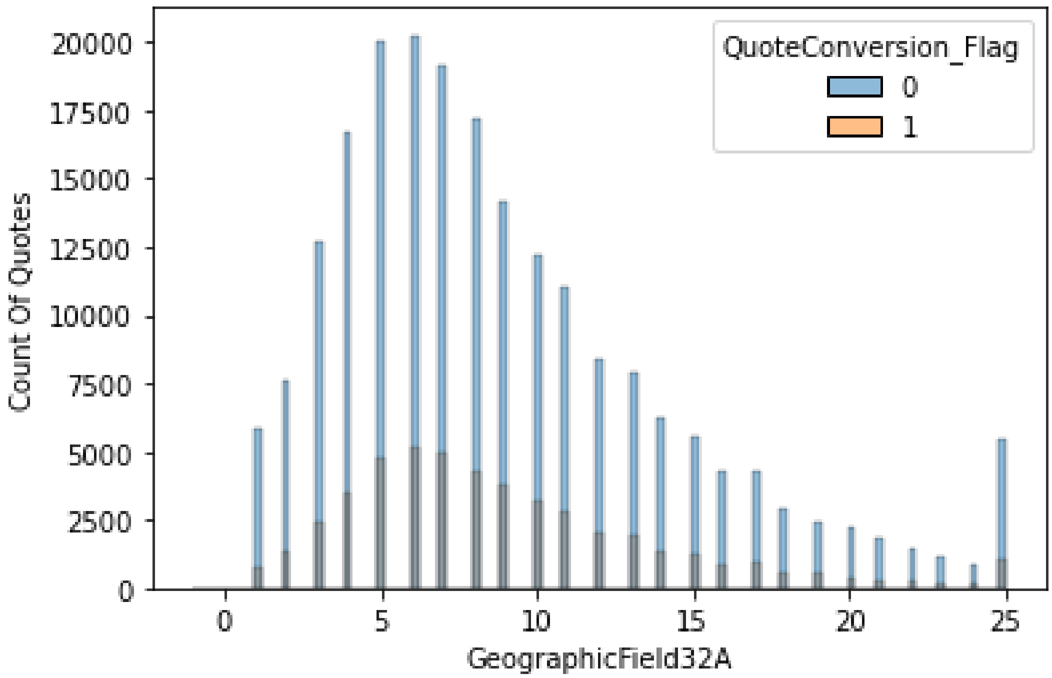
QuoteConversion_Flag
0
1
Count Of Quotes
8000
6000
4000
2000
0
0
5
10
15
20
25
GeographicField29B

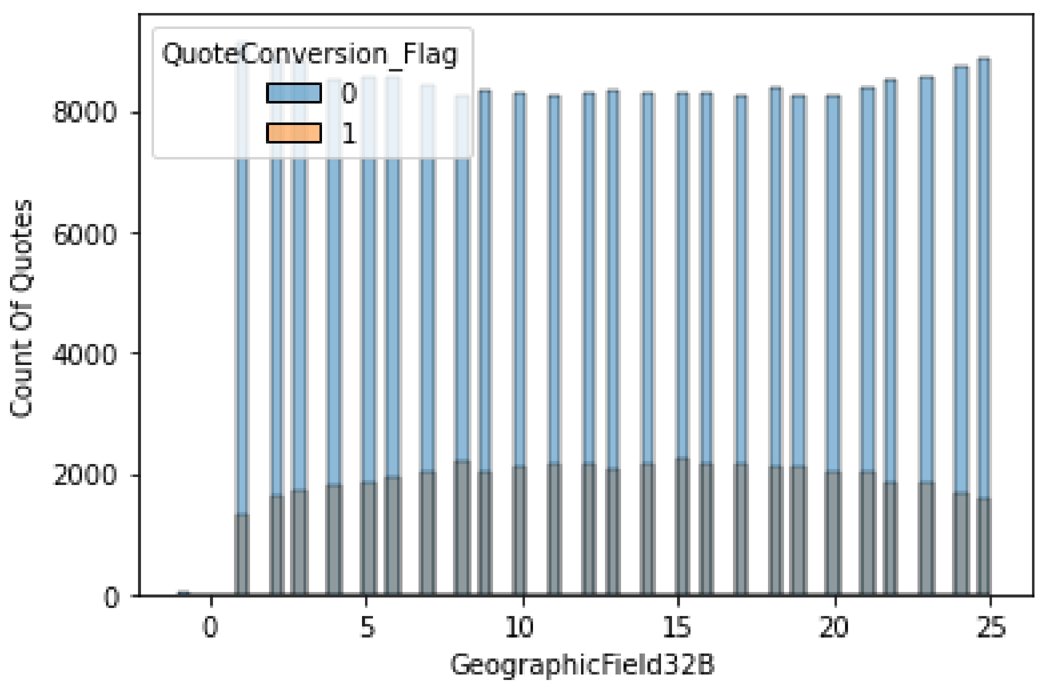

QuoteConversion_Flag
0
1
Count Of Quotes
25000
20000
15000
10000
5000
0
0
5
10
15
20
25
GeographicField30A

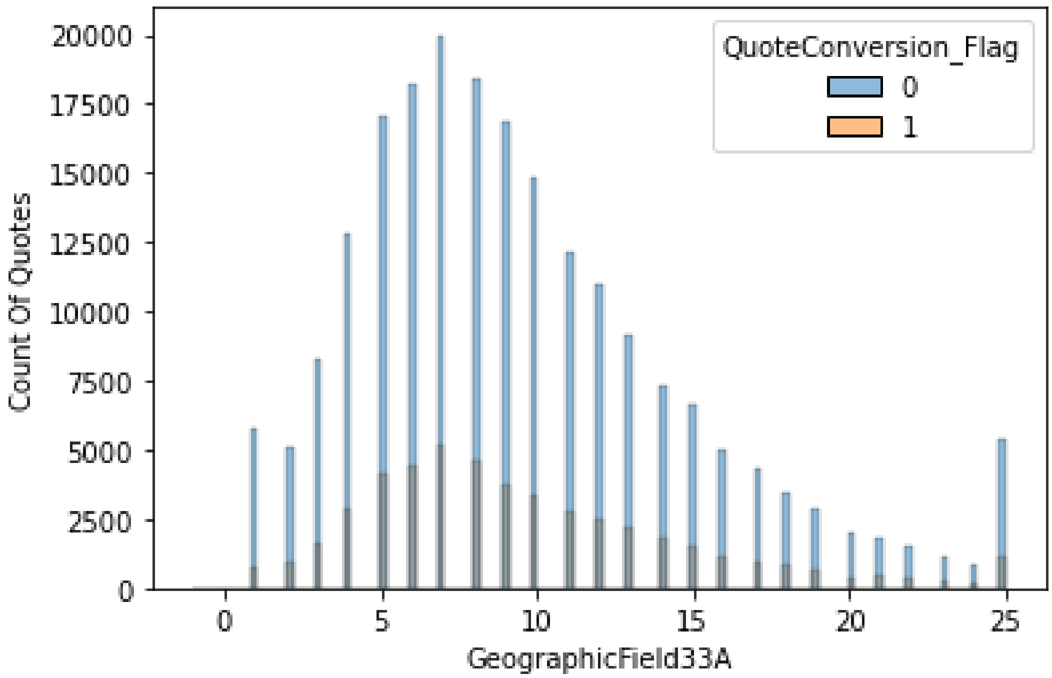

QuoteConversion_Flag
0
1
Count Of Quotes
8000
6000
4000
2000
0
0
5
10
15
20
25
GeographicField30B

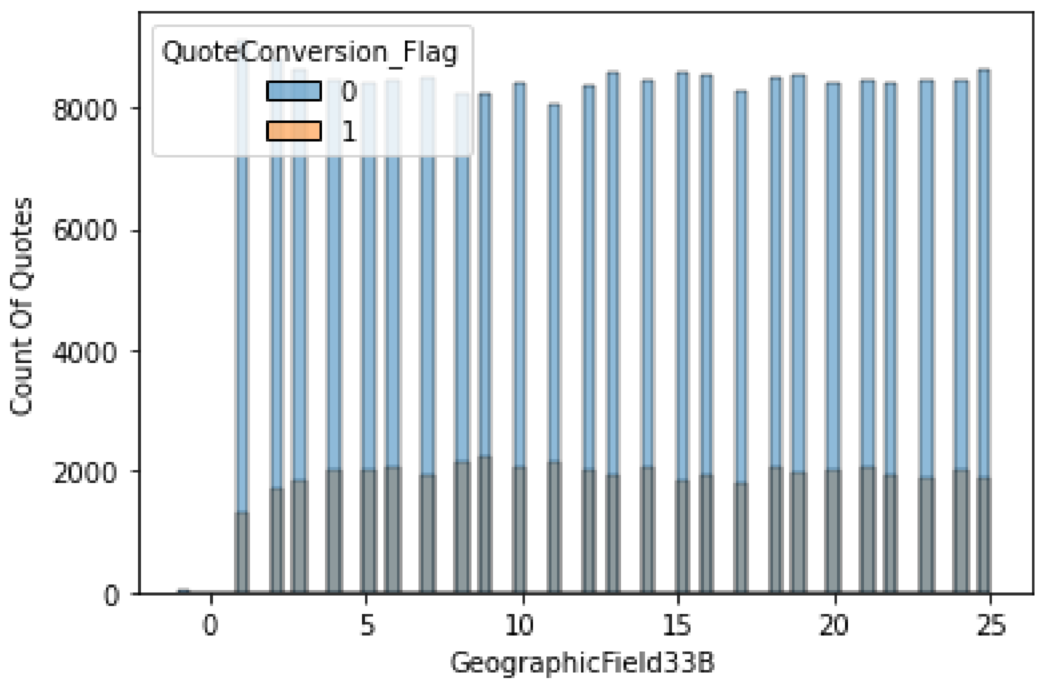

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField31A
0
2500
5000
7500
10000
12500
15000
17500
20000
0
5
10
15
20
25

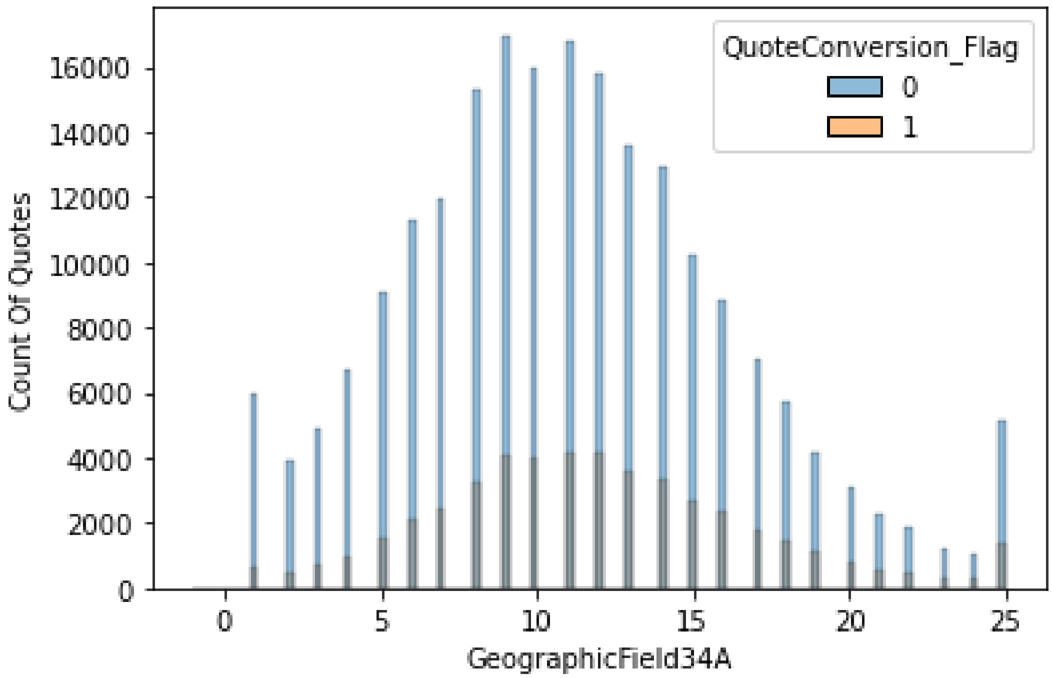

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField31B
0
2000
4000
6000
8000
0
5
10
15
20
25

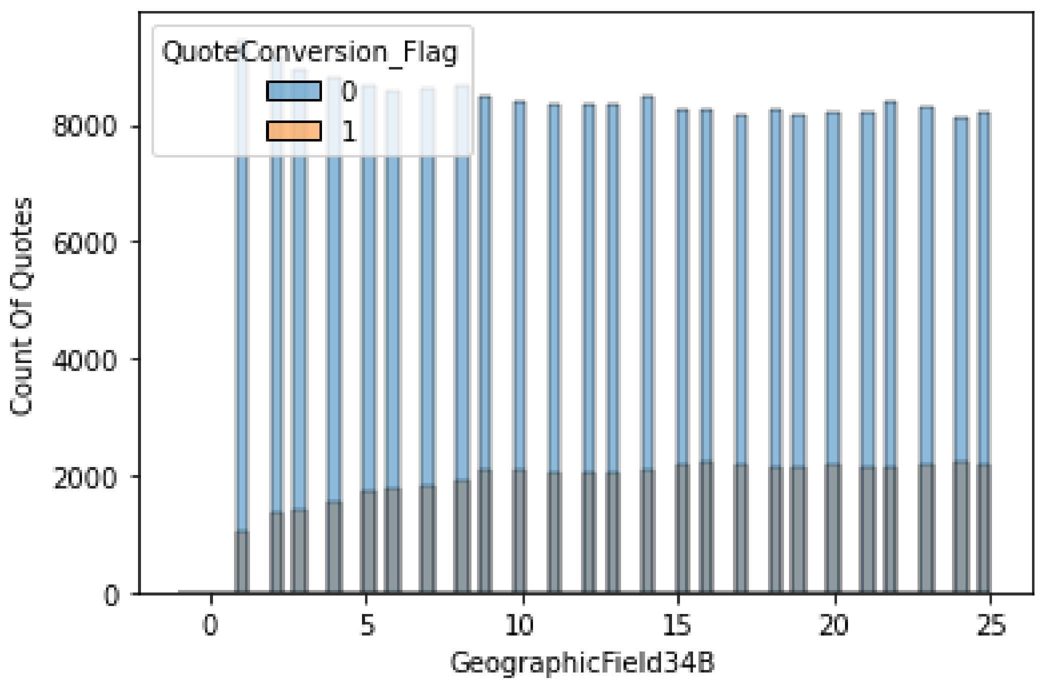

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField32A
0
2500
5000
7500
10000
12500
15000
17500
20000
0
5
10
15
20
25

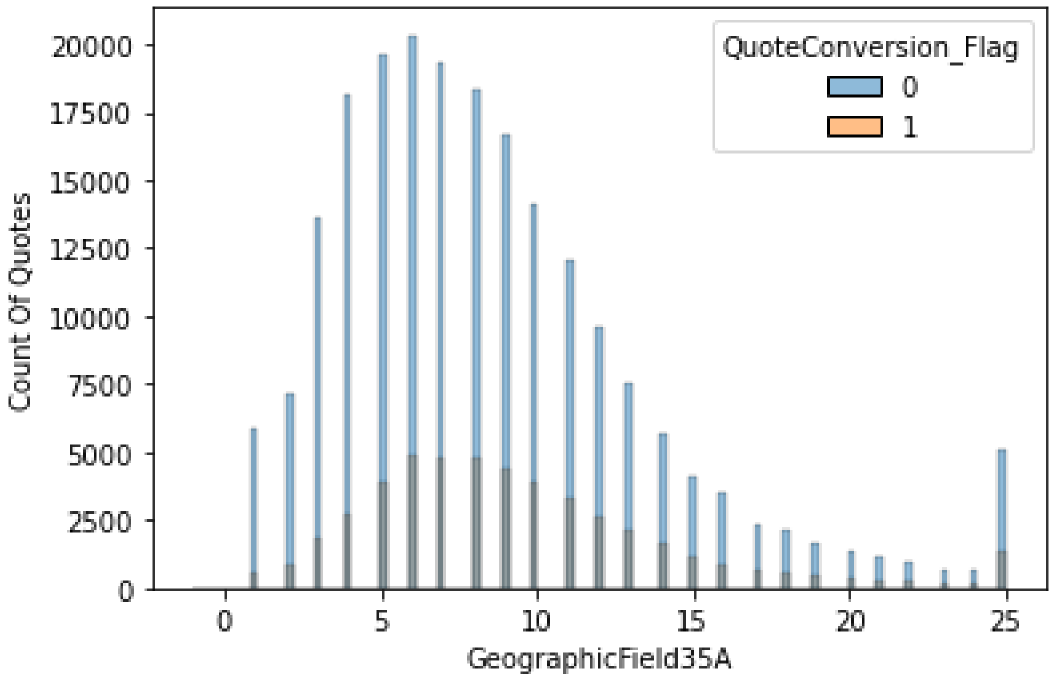

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField32B
0
2000
4000
6000
8000
0
5
10
15
20
25

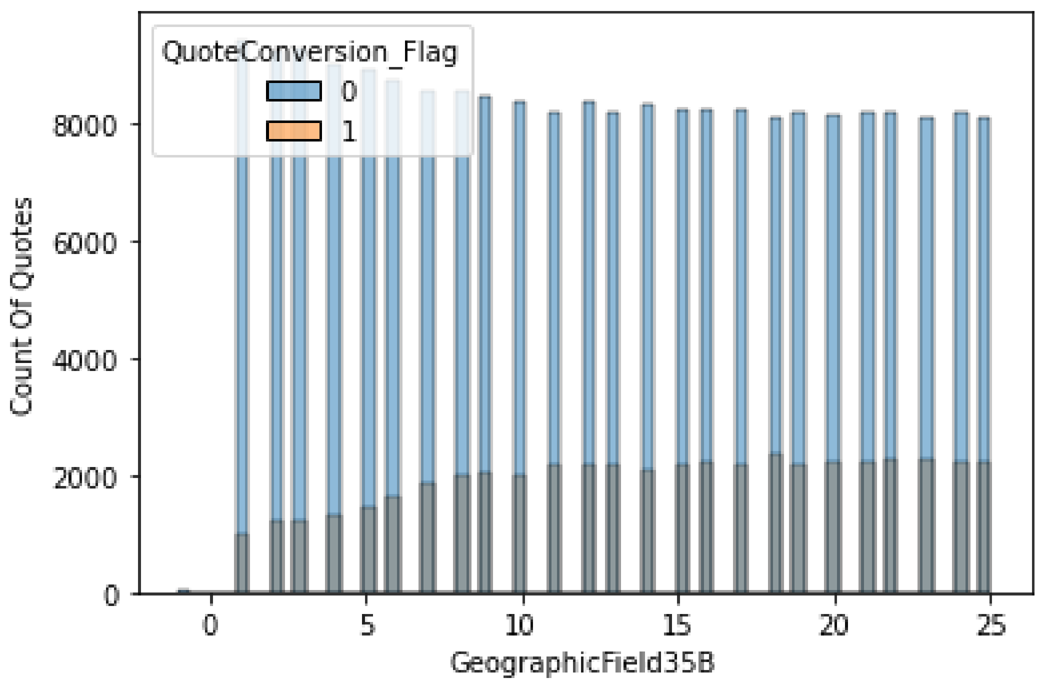

QuoteConversion_Flag
0
1
20000
17500
15000
12500
10000
7500
5000
2500
0
Count Of Quotes
0
5
10
15
20
25
GeographicField33A

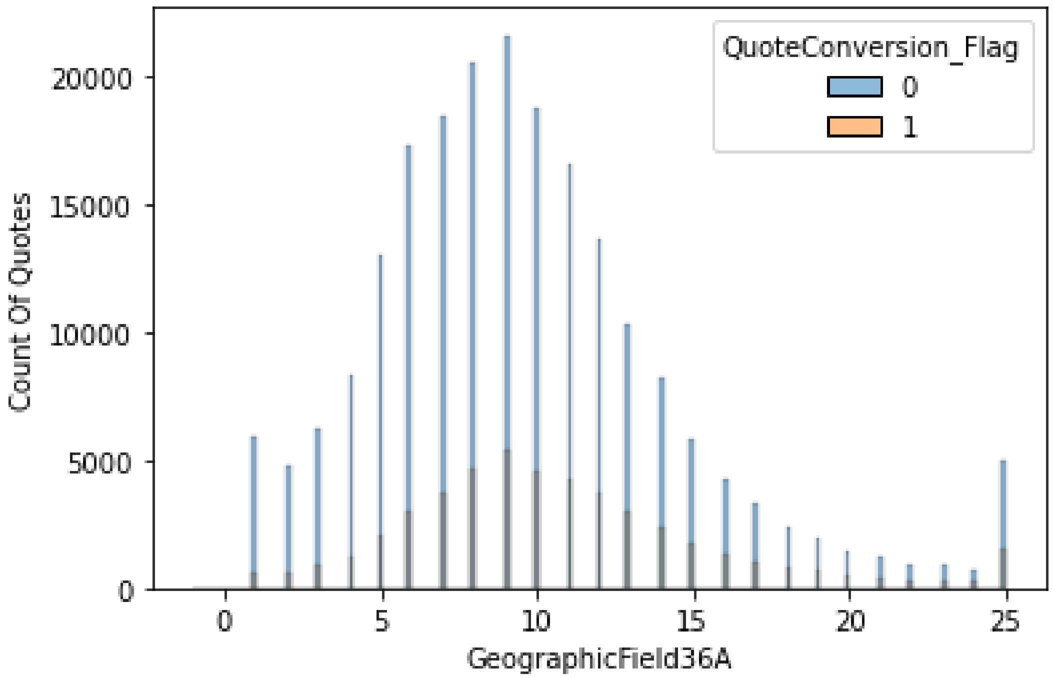

QuoteConversion_Flag
0
1
8000
6000
4000
2000
0
Count Of Quotes
0
5
10
15
20
25
GeographicField33B

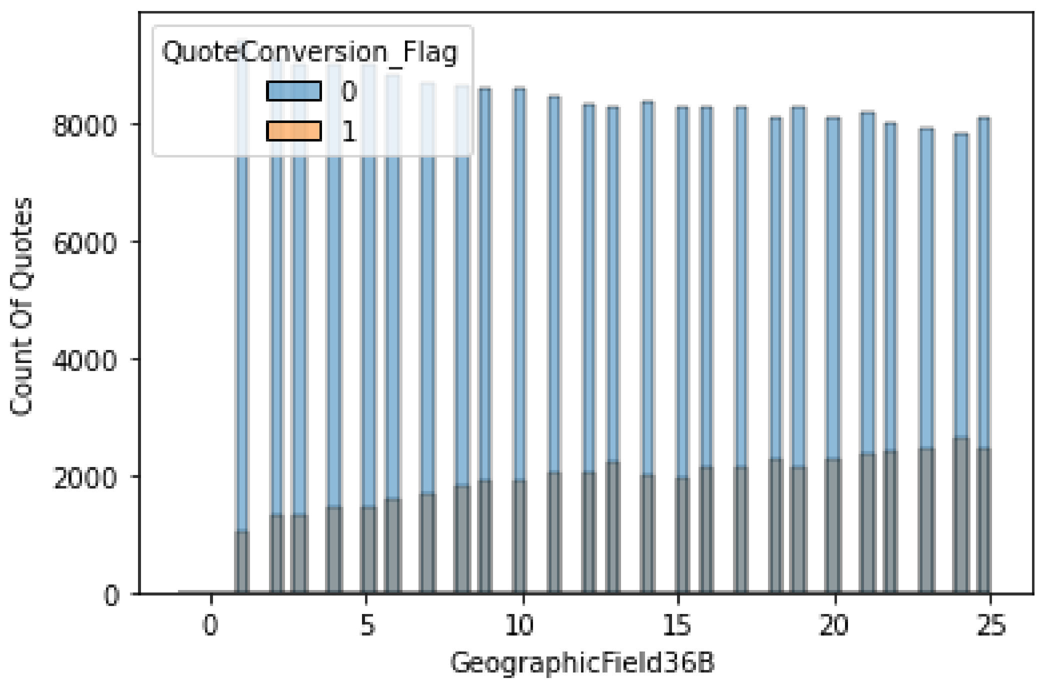

QuoteConversion_Flag
0
1
Count Of Quotes
0
2000
4000
6000
8000
10000
12000
14000
16000
0
5
10
15
20
25
GeographicField34A

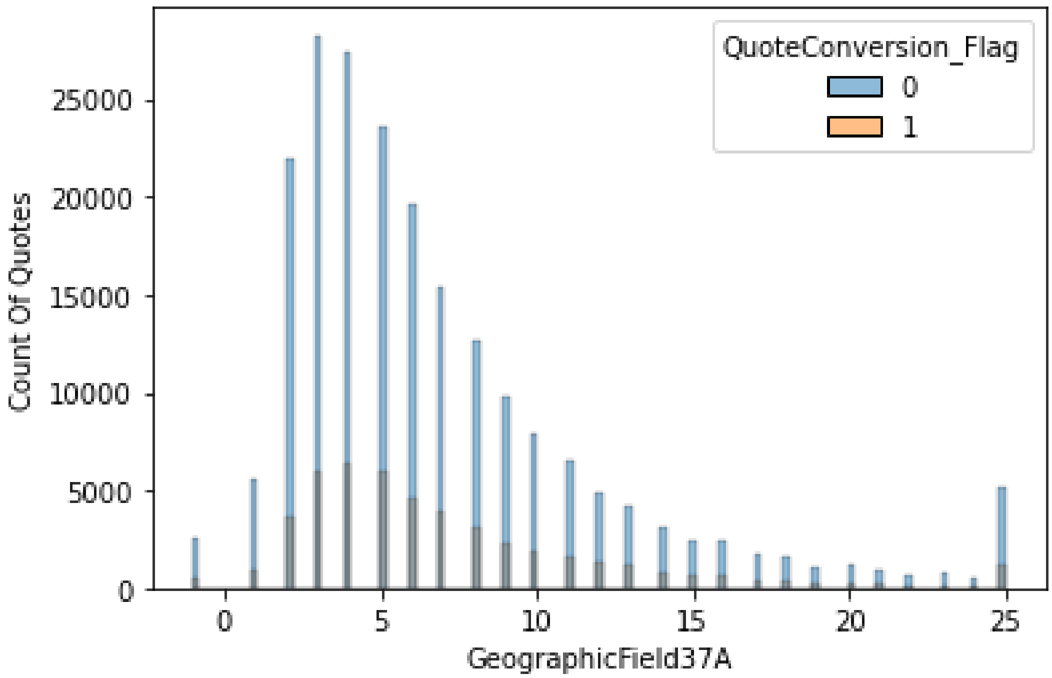

QuoteConversion_Flag
0
1
Count Of Quotes
0
2000
4000
6000
8000
0
5
10
15
20
25
GeographicField34B

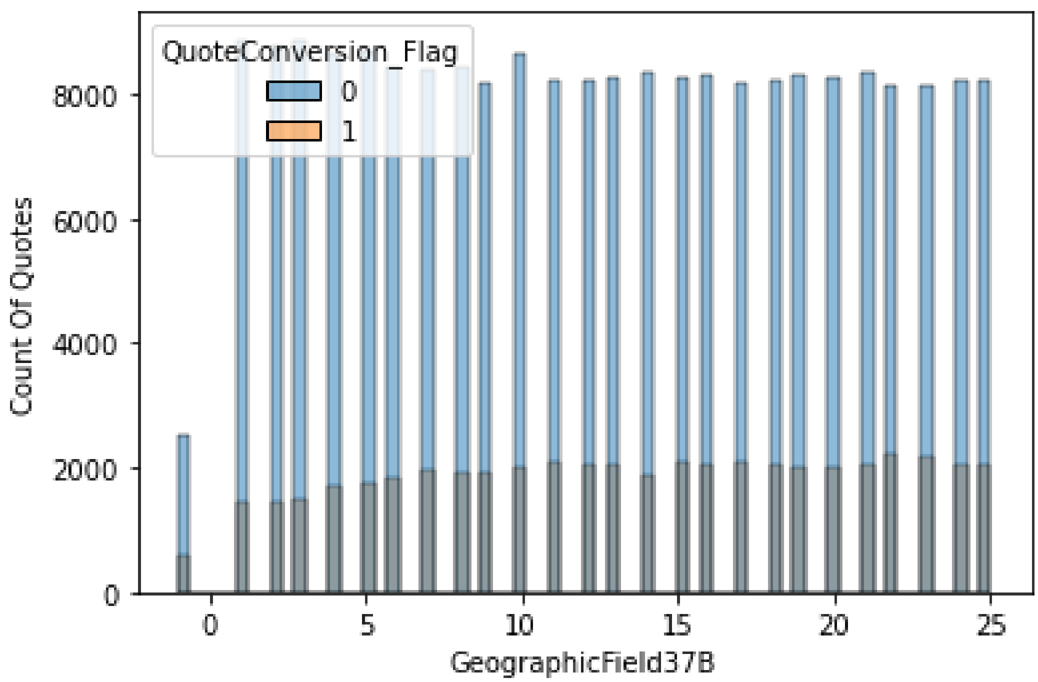

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField35A
0
5
10
15
20
25
0
2500
5000
7500
10000
12500
15000
17500
20000

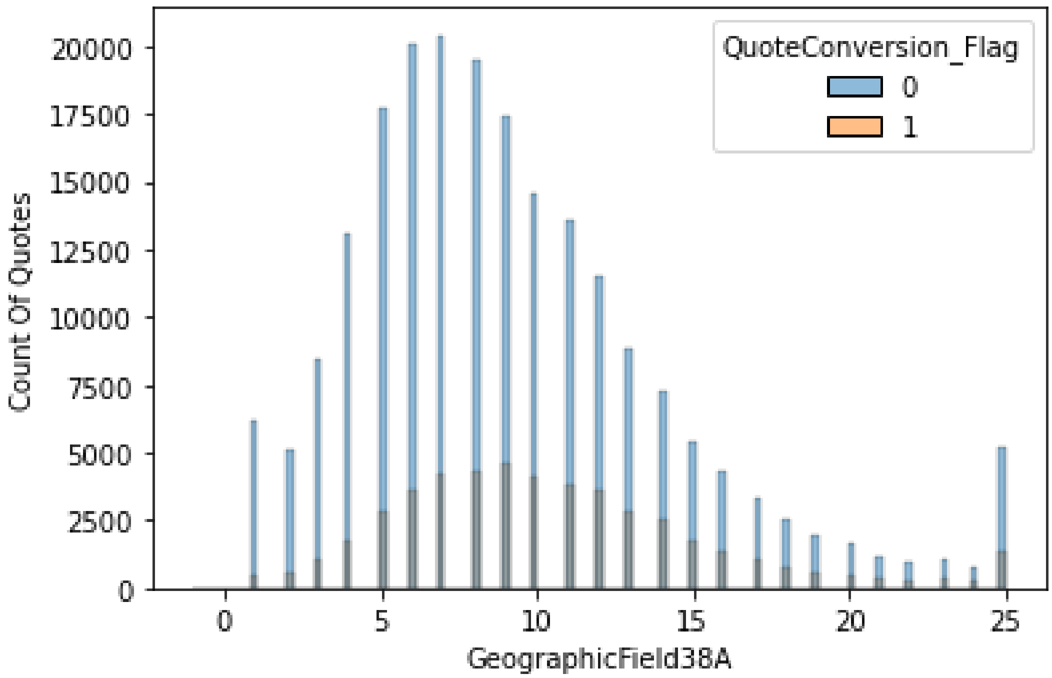

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField35B
0
5
10
15
20
25
0
2000
4000
6000
8000

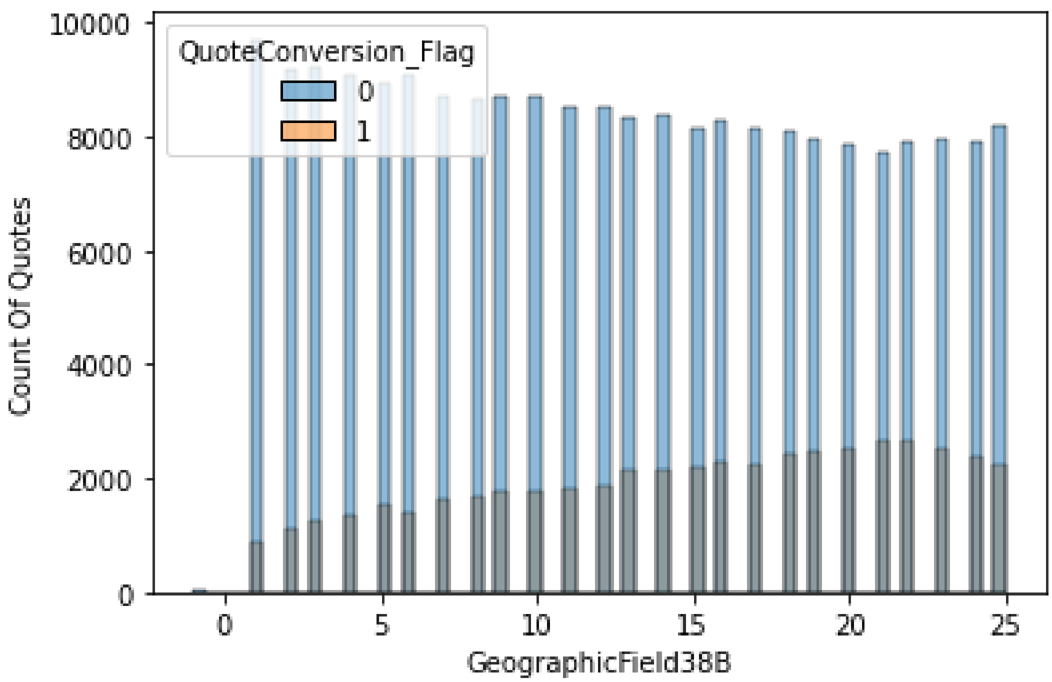

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField36A
0
5000
10000
15000
20000
0
5
10
15
20
25

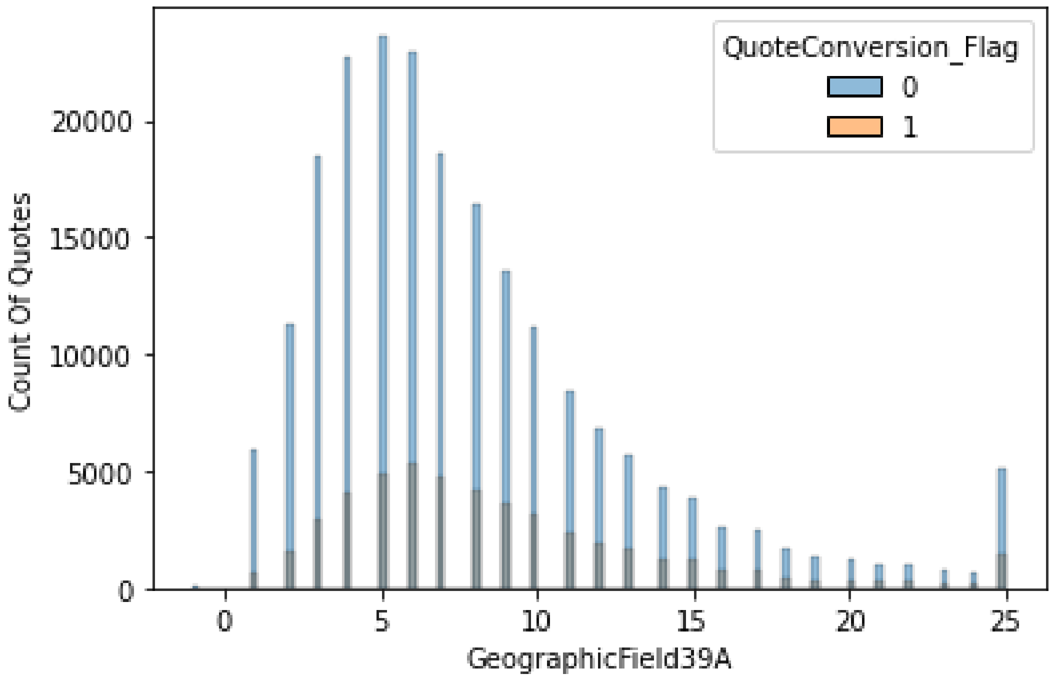

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField36B
0
2000
4000
6000
8000
0
5
10
15
20
25

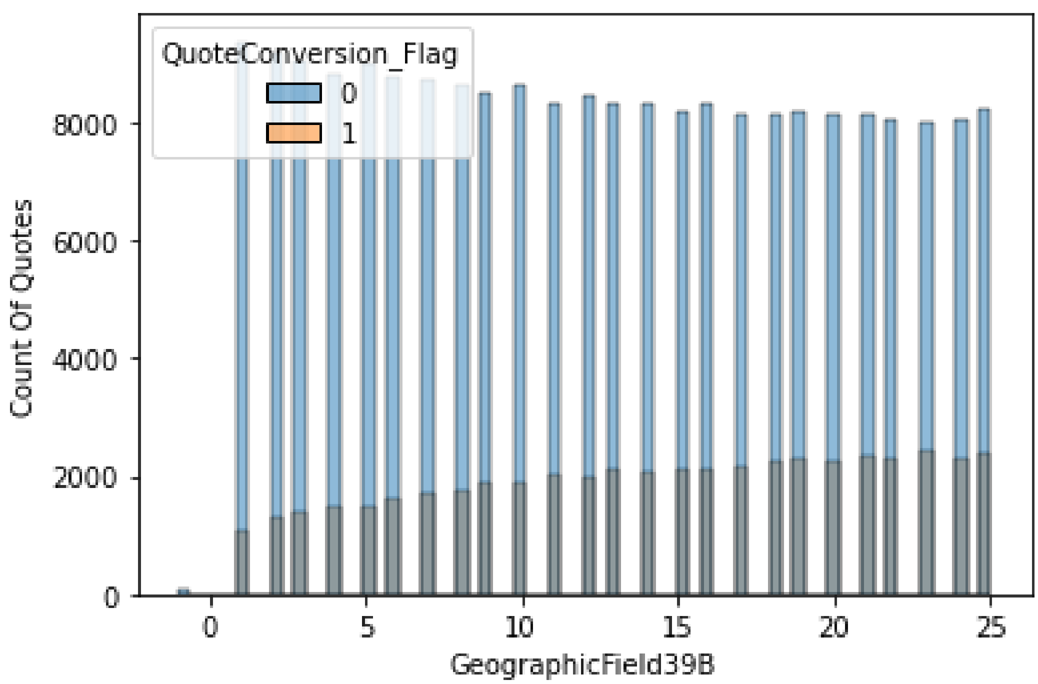

QuoteConversion_Flag
0
1
Count Of Quotes
25000
20000
15000
10000
5000
0
0
5
10
15
20
25
GeographicField37A

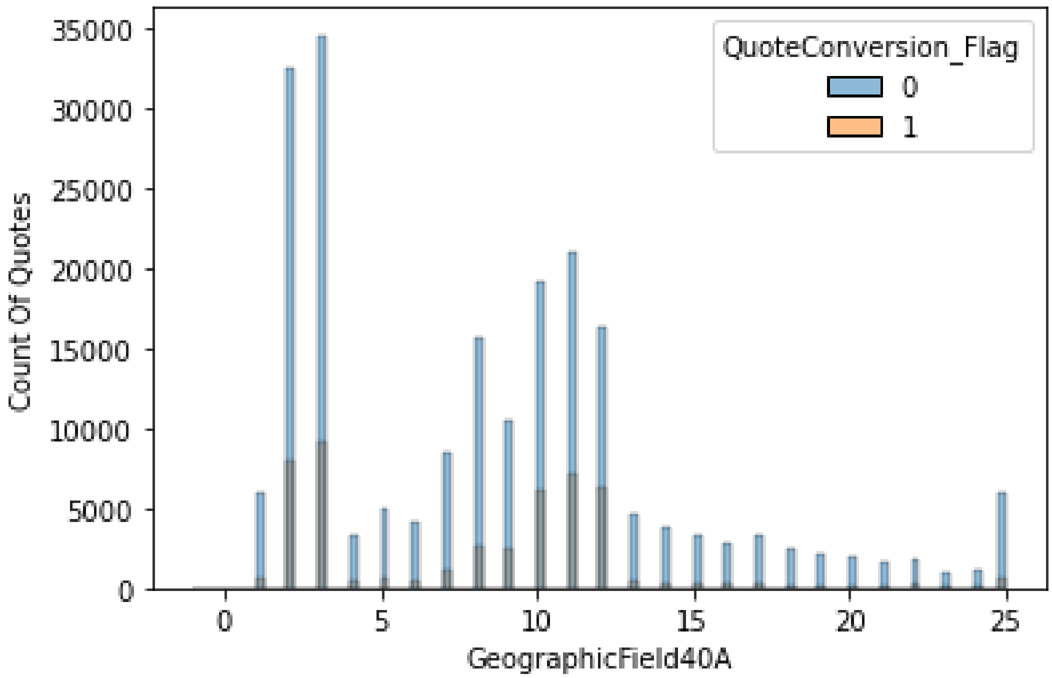

QuoteConversion_Flag
0
1
Count Of Quotes
8000
6000
4000
2000
0
0
5
10
15
20
25
GeographicField37B

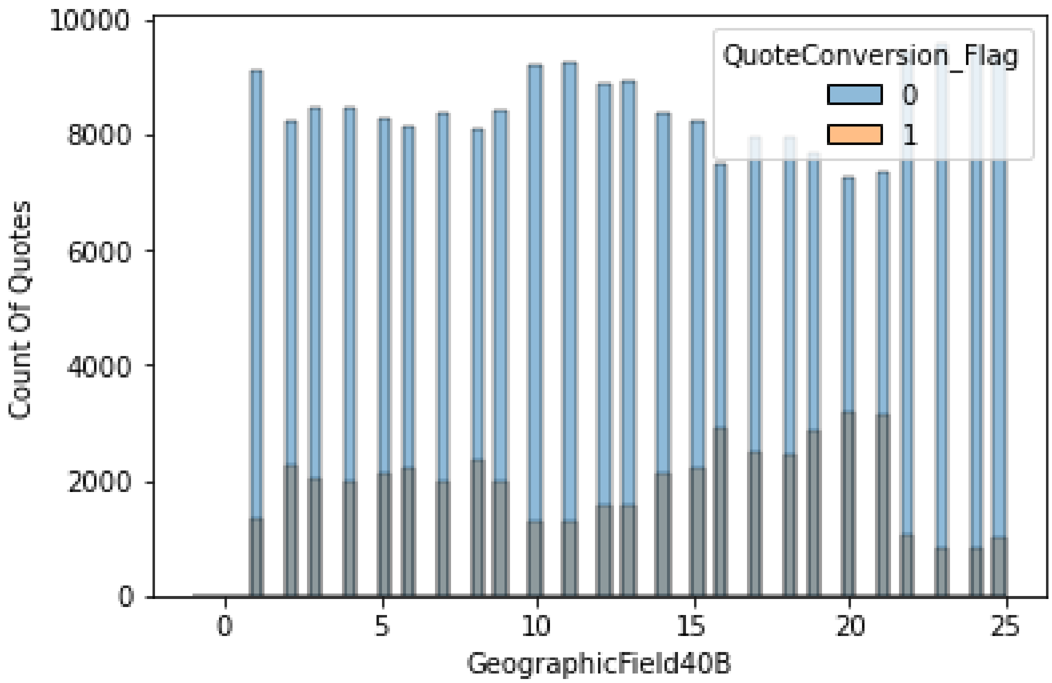

QuoteConversion_Flag
0
1
20000
17500
15000
12500
10000
7500
5000
2500
0
Count Of Quotes
0
5
10
15
20
25
GeographicField38A

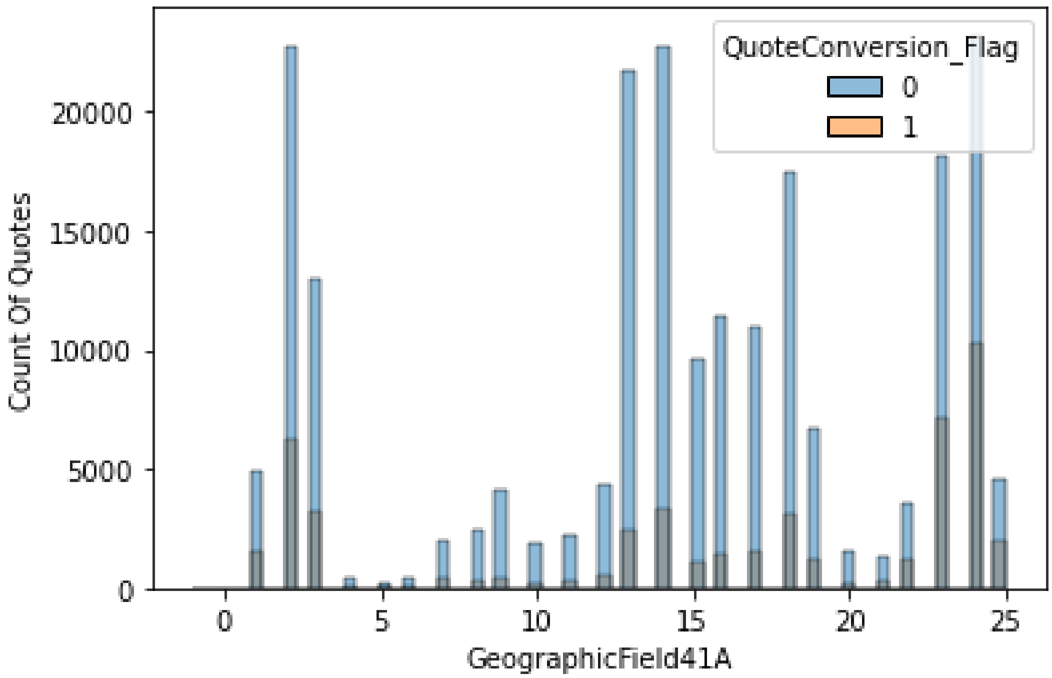

QuoteConversion_Flag
0
1
10000
8000
6000
4000
2000
0
Count Of Quotes
0
5
10
15
20
25
GeographicField38B

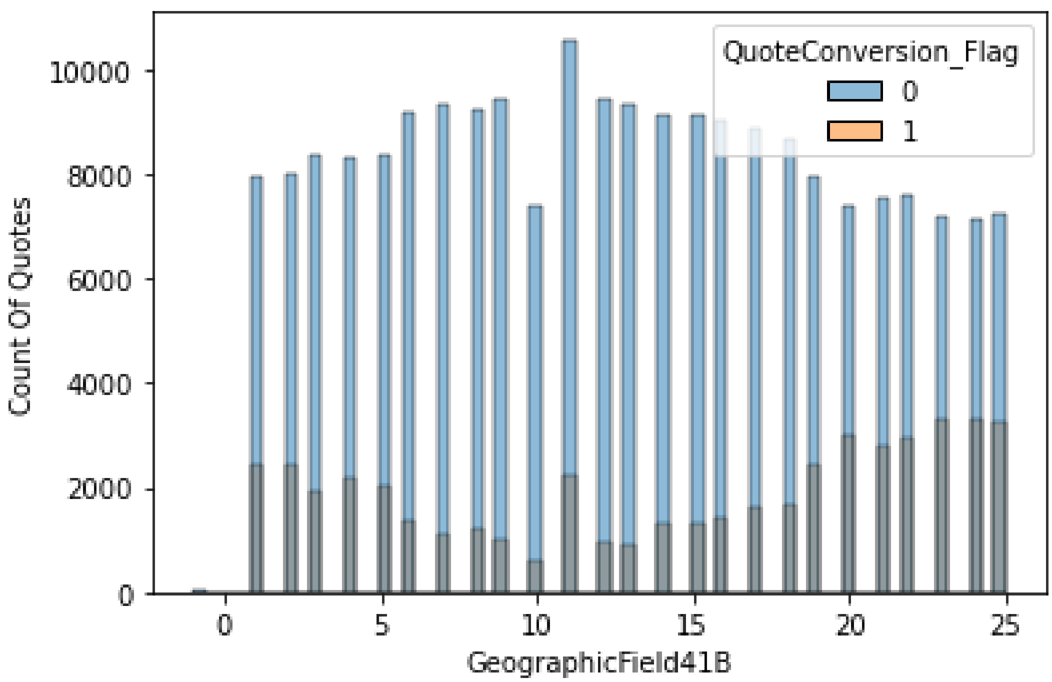

QuoteConversion_Flag
0
1
Count Of Quotes
0
5000
10000
15000
20000
0
5
10
15
20
25
GeographicField39A

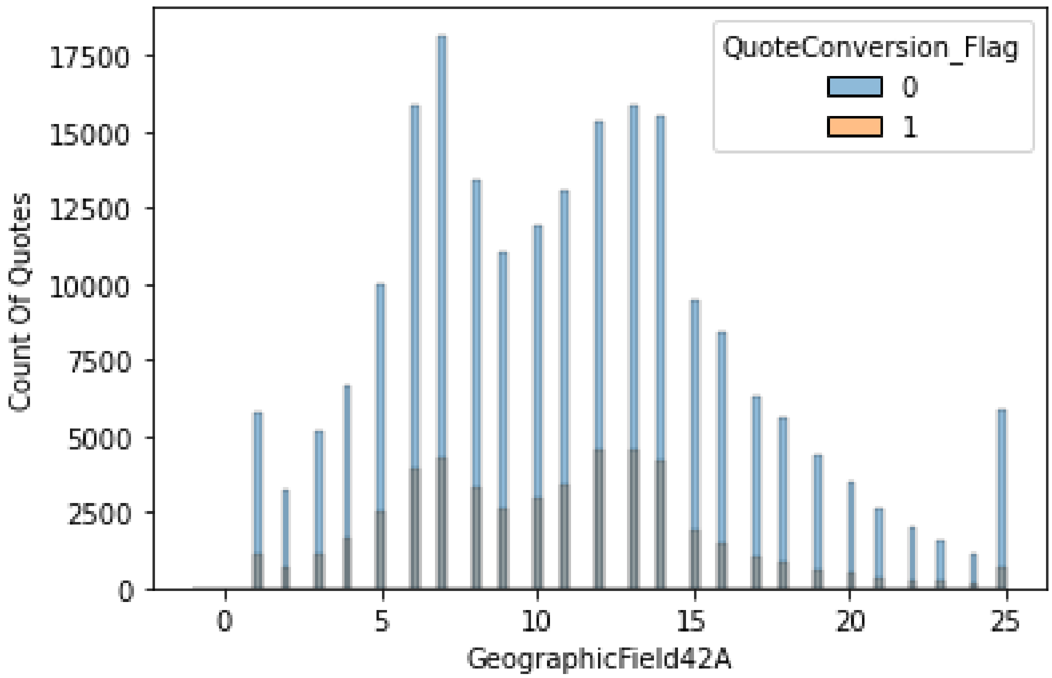

QuoteConversion_Flag
0
1
Count Of Quotes
0
2000
4000
6000
8000
0
5
10
15
20
25
GeographicField39B

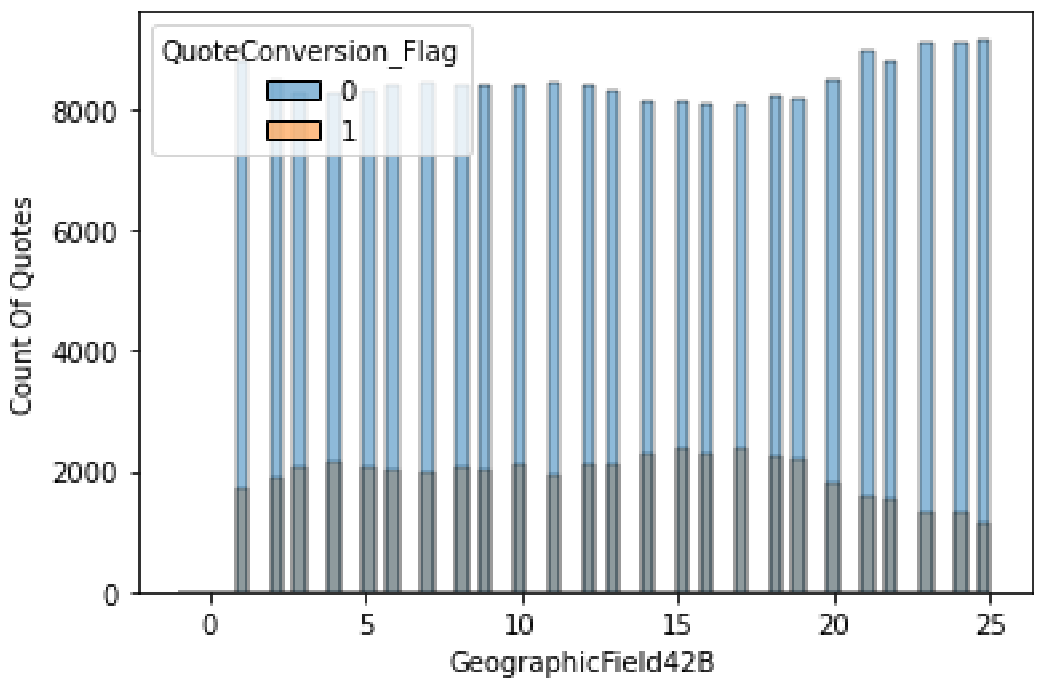

QuoteConversion_Flag
0
1
Count Of Quotes
35000
30000
25000
20000
15000
10000
5000
0
0
5
10
15
20
25
GeographicField40A

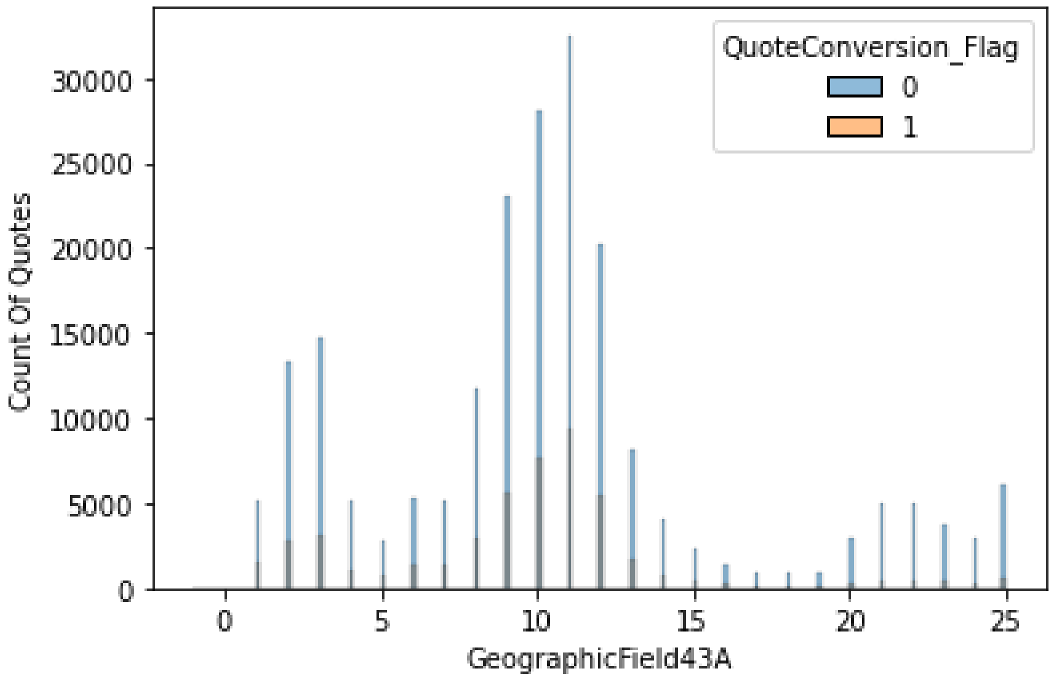

QuoteConversion_Flag
0
1
Count Of Quotes
10000
8000
6000
4000
2000
0
0
5
10
15
20
25
GeographicField40B

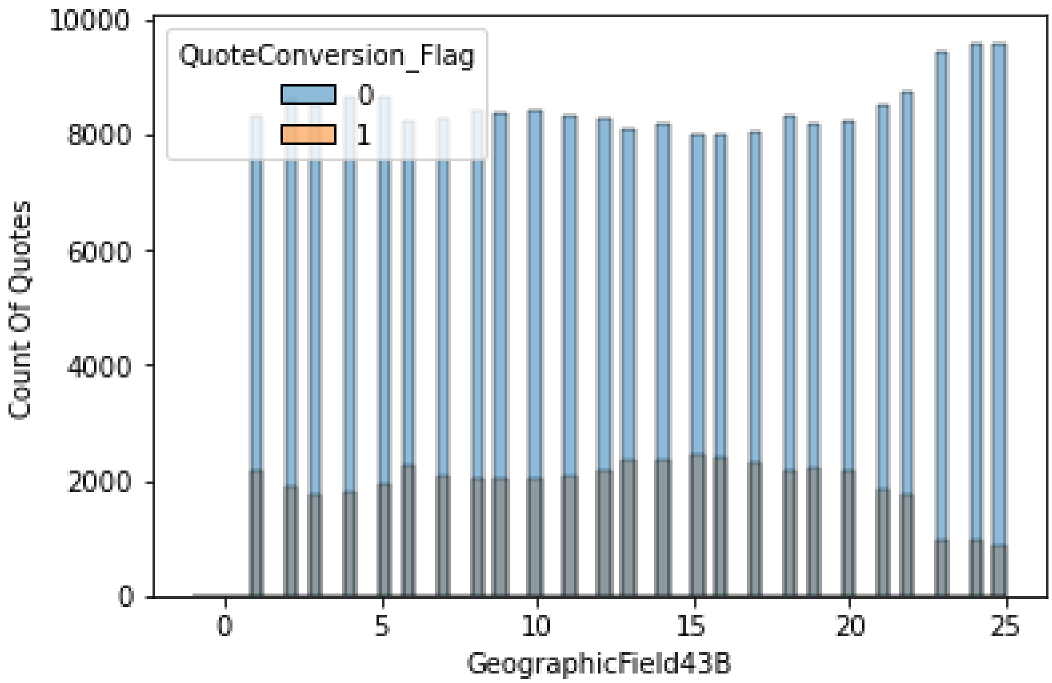

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField41A

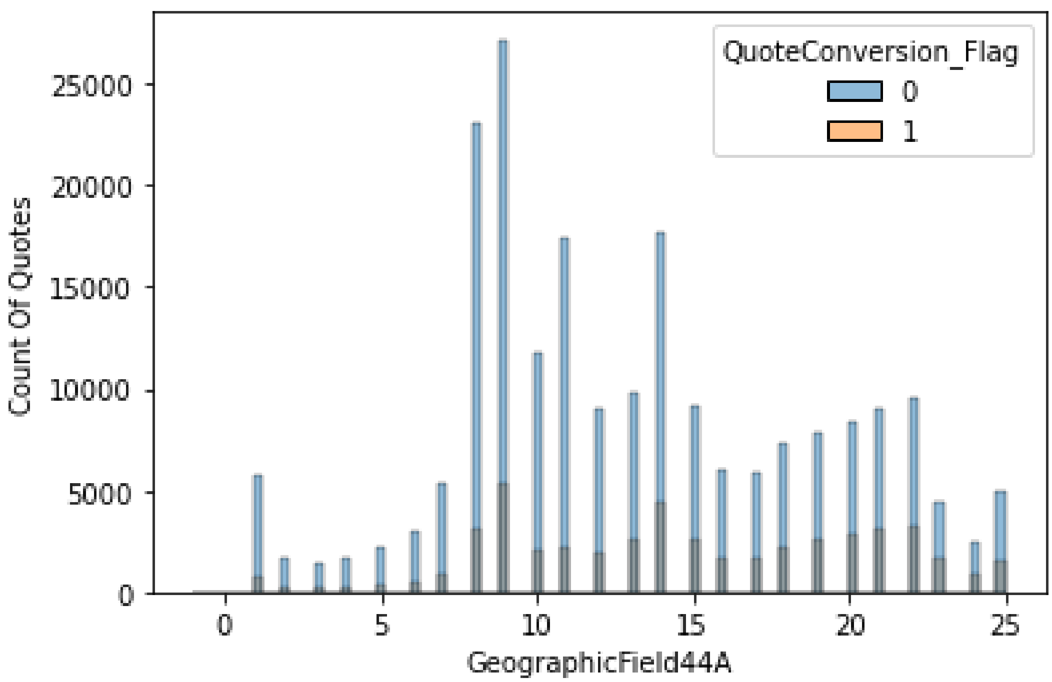

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField41B

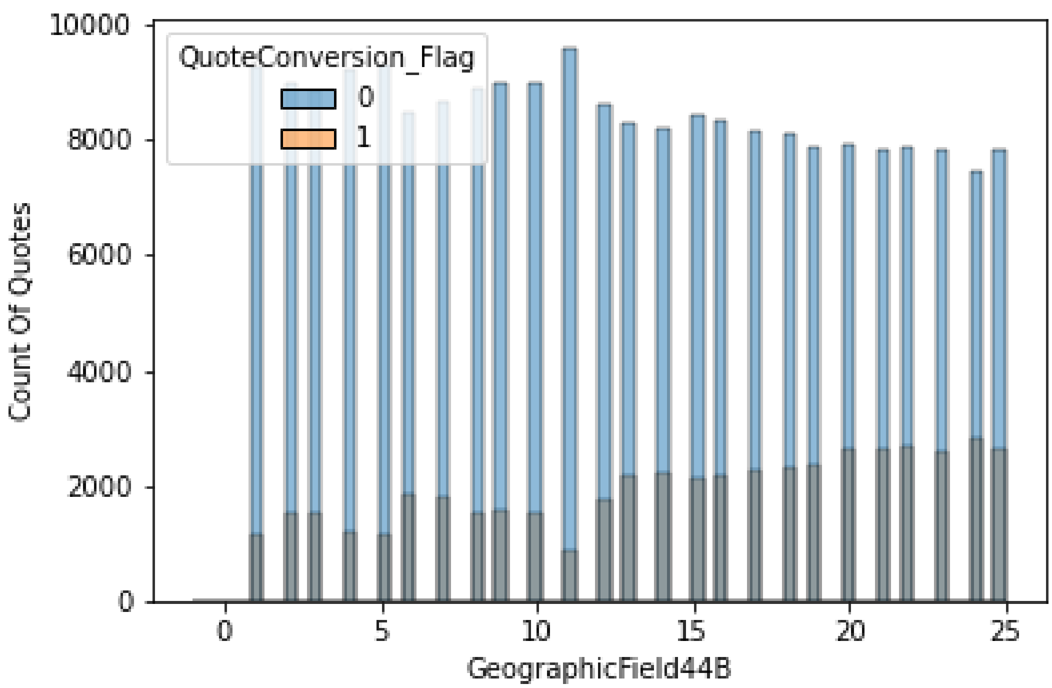

QuoteConversion_Flag
0
1
Count Of Quotes
0
2500
5000
7500
10000
12500
15000
17500
0
5
10
15
20
25
GeographicField42A

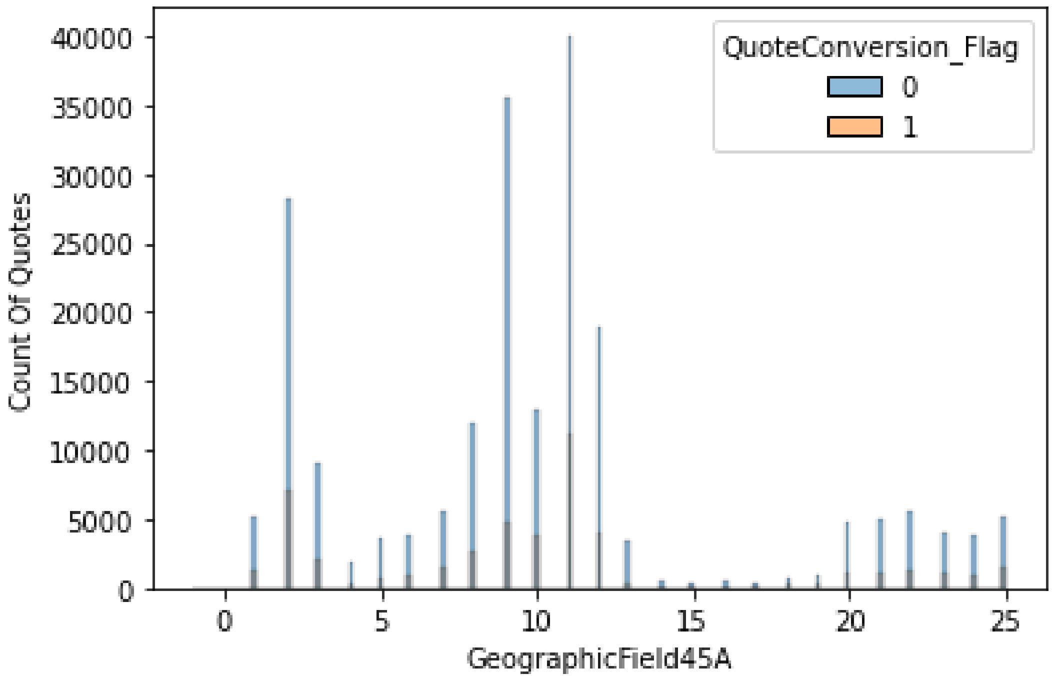

QuoteConversion_Flag
0
1
Count Of Quotes
0
2000
4000
6000
8000
0
5
10
15
20
25
GeographicField42B

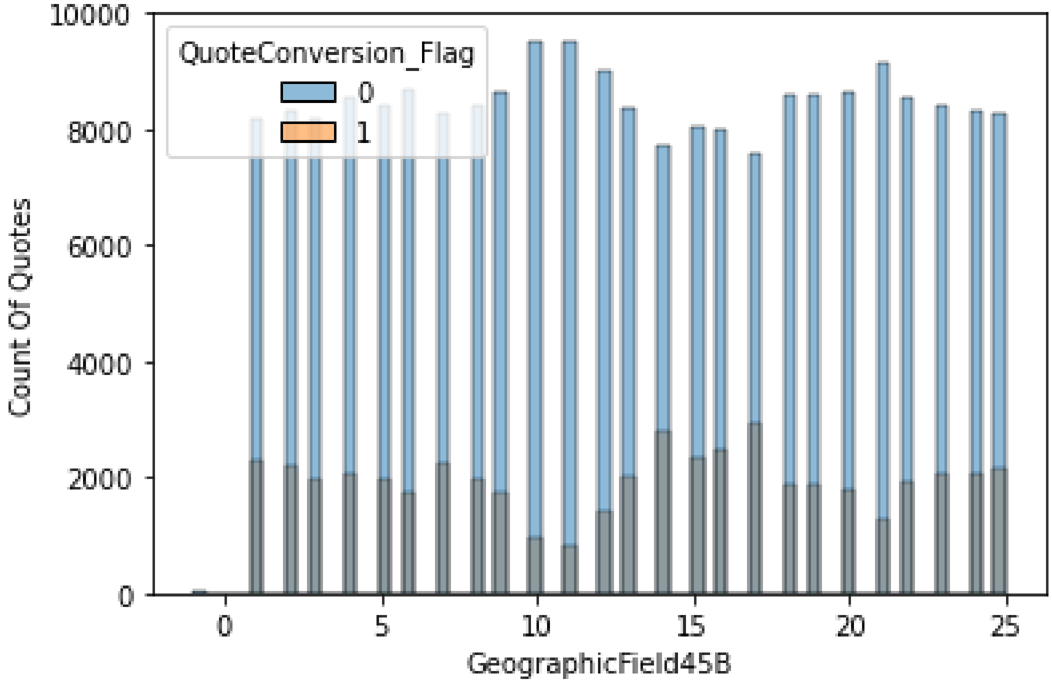

QuoteConversion_Flag
0
1
Count Of Quotes
30000
25000
20000
15000
10000
5000
0
0
5
10
15
20
25
GeographicField43A

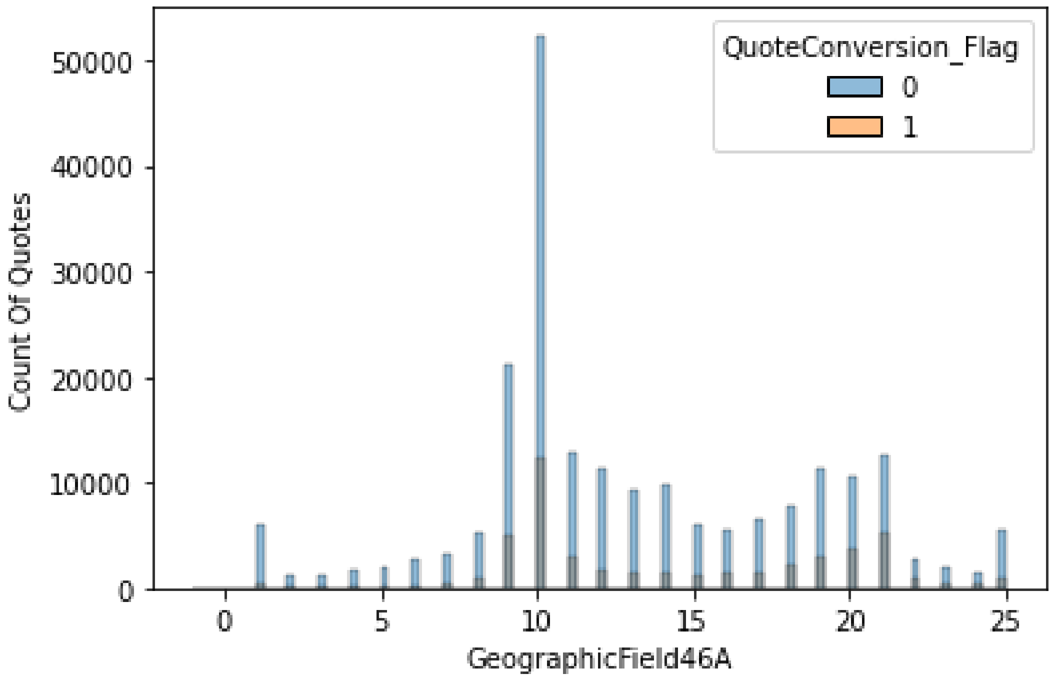

QuoteConversion_Flag
0
1
Count Of Quotes
10000
8000
6000
4000
2000
0
0
5
10
15
20
25
GeographicField43B

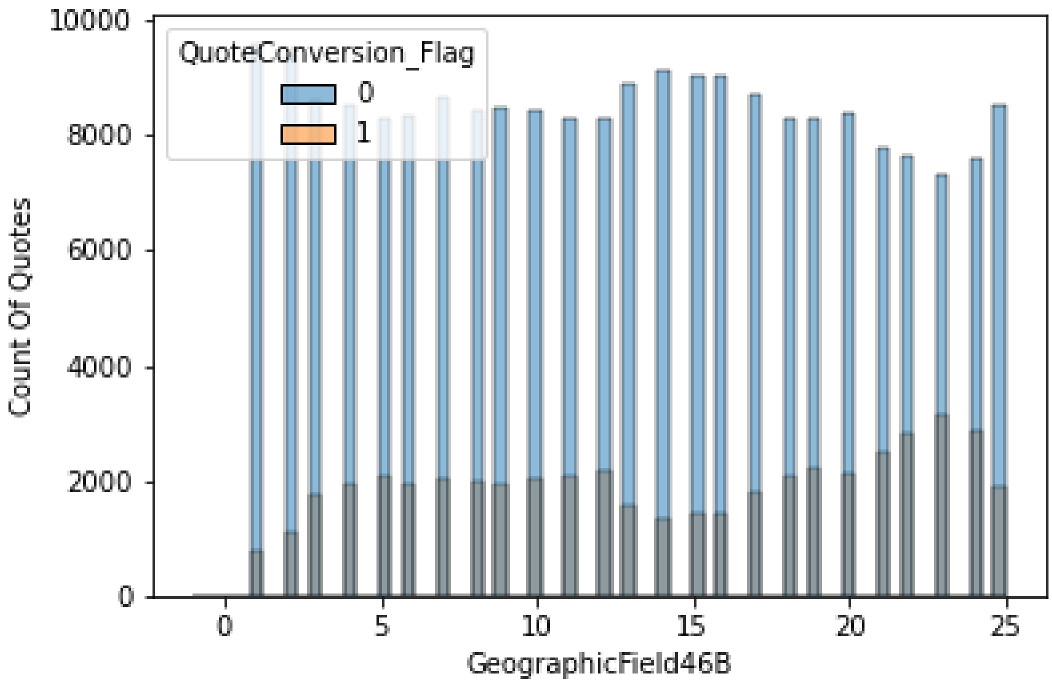

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField44A

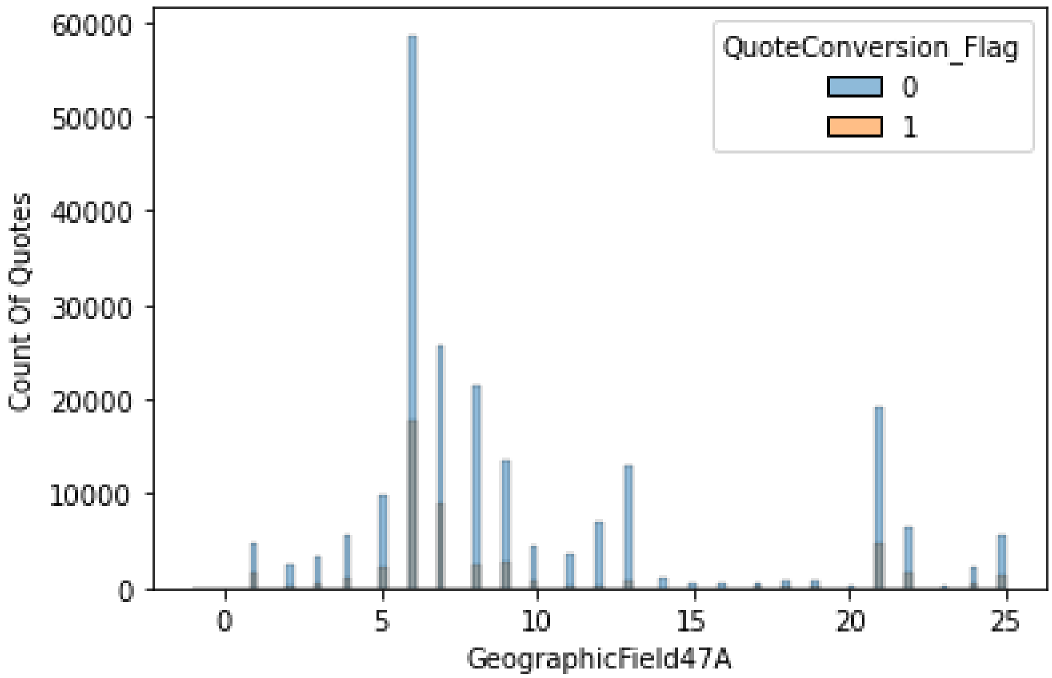

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField44B

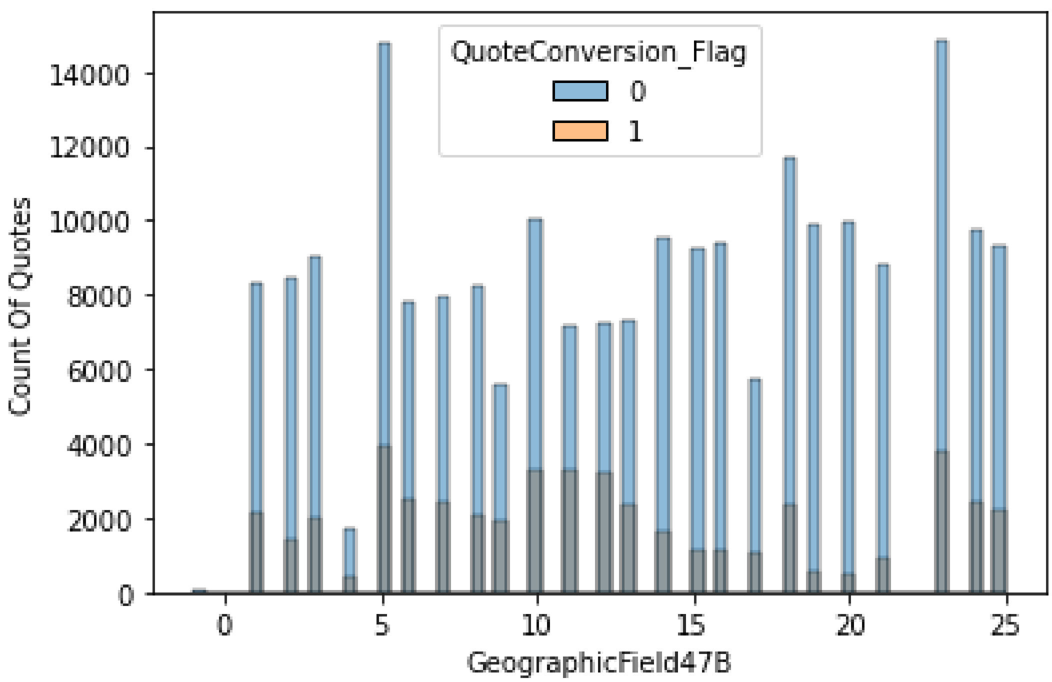

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField45A
40000
35000
30000
25000
20000
15000
10000
5000
0
0
5
10
15
20
25

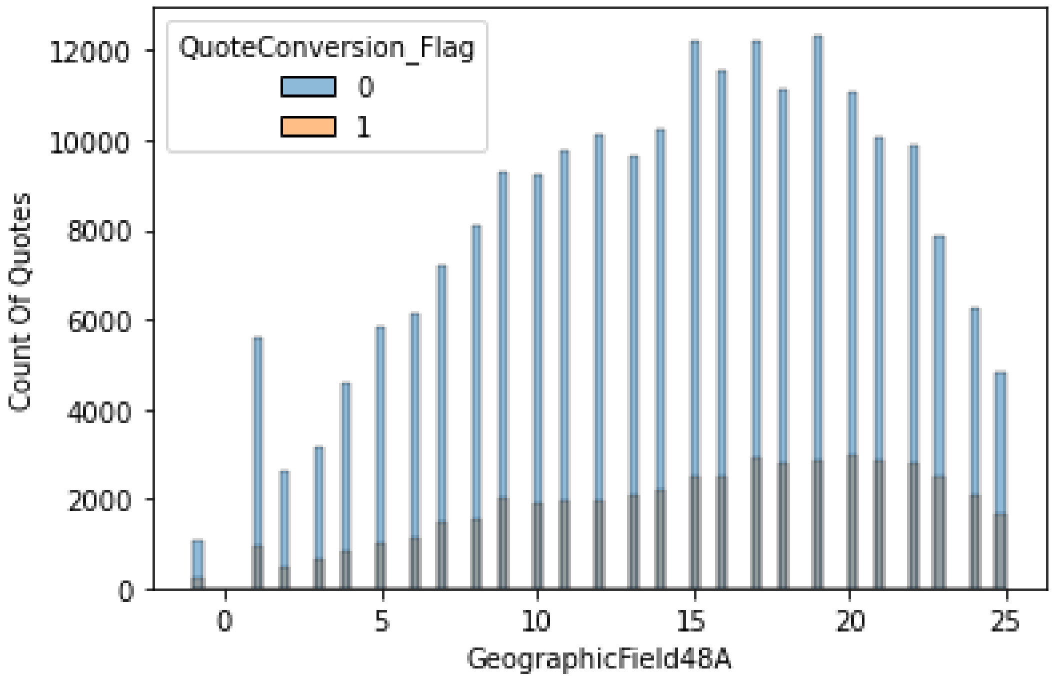

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField45B
10000
8000
6000
4000
2000
0
0
5
10
15
20
25

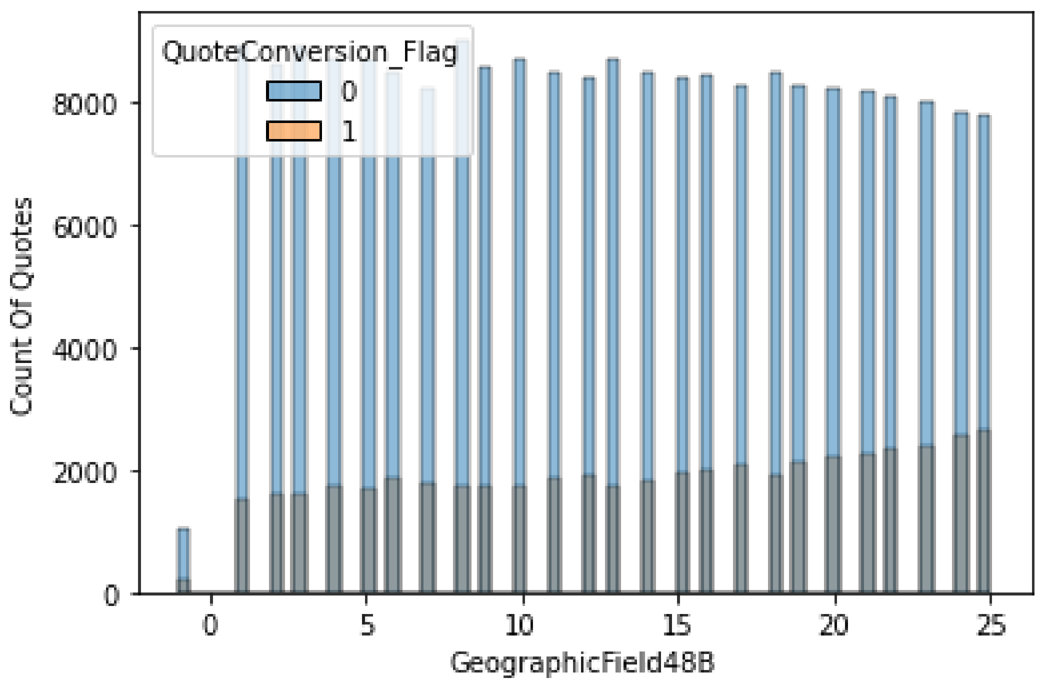

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField46A
0
10000
20000
30000
40000
50000
0
5
10
15
20
25

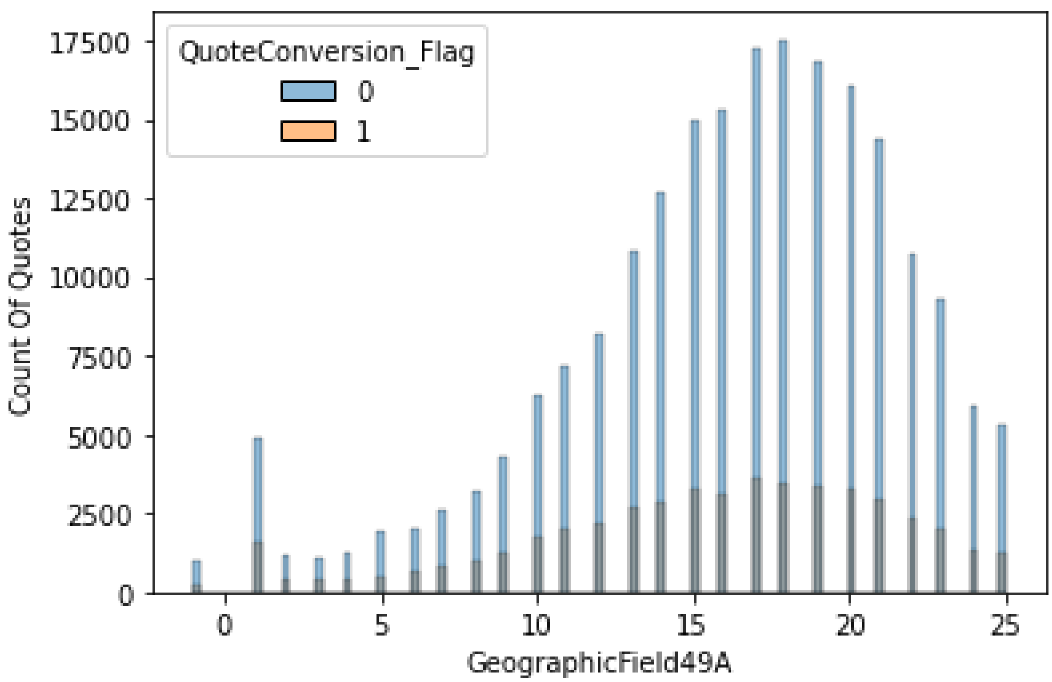

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField46B
0
2000
4000
6000
8000
10000
0
5
10
15
20
25

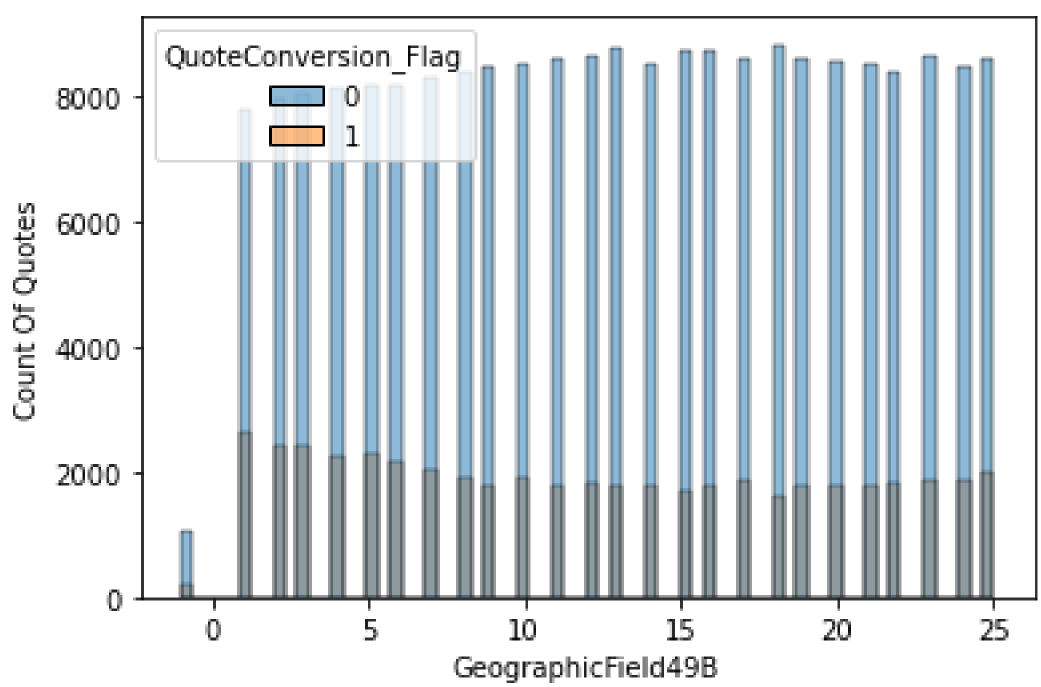

QuoteConversion_Flag
0
1
Count Of Quotes
60000
50000
40000
30000
20000
10000
0
0
5
10
15
20
25
GeographicField47A

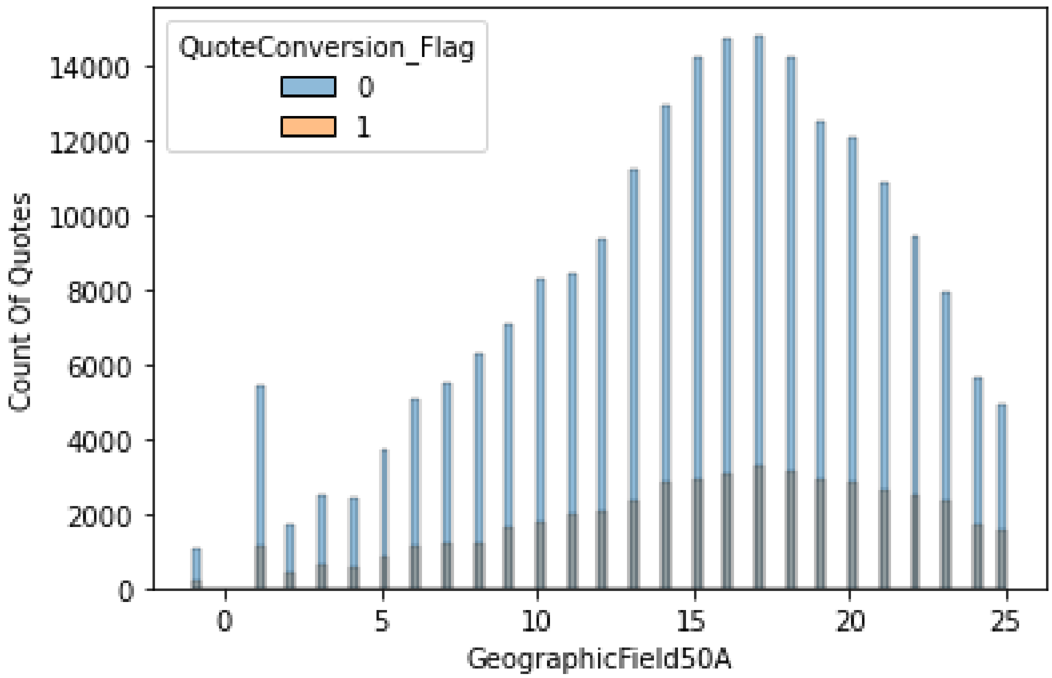

QuoteConversion_Flag
0
1
Count Of Quotes
14000
12000
10000
8000
6000
4000
2000
0
0
5
10
15
20
25
GeographicField47B


QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField48A
0
2000
4000
6000
8000
10000
12000
0
5
10
15
20
25

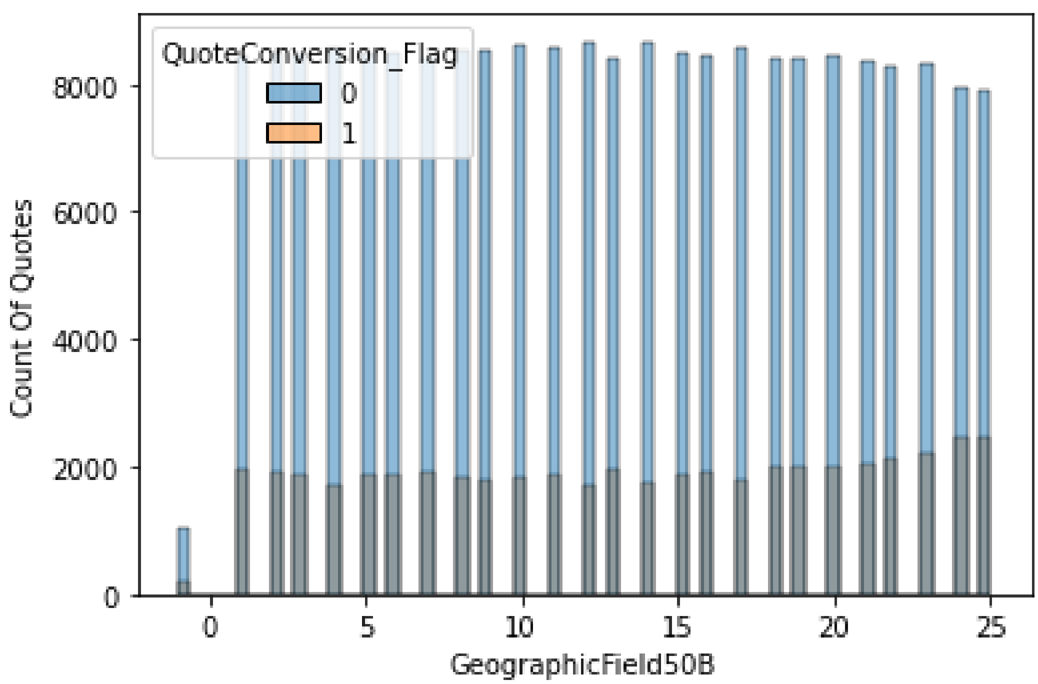

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField48B
0
2000
4000
6000
8000
0
5
10
15
20
25


QuoteConversion_Flag
0
1
17500
15000
12500
10000
7500
5000
2500
0
Count Of Quotes
0
5
10
15
20
25
GeographicField49A

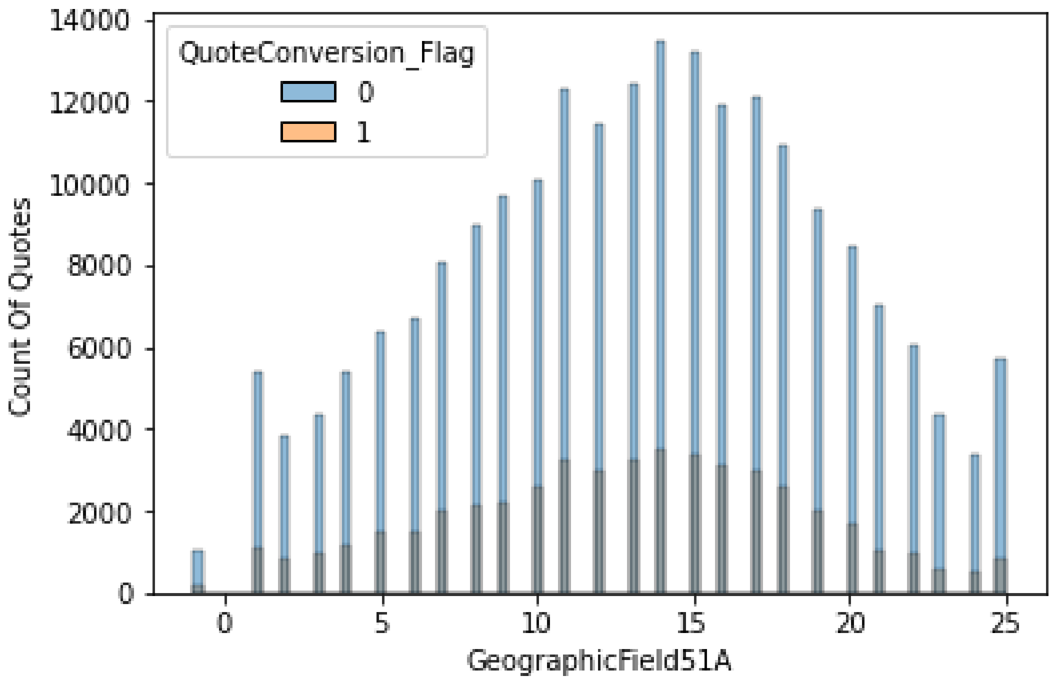

QuoteConversion_Flag
0
1
8000
6000
4000
2000
0
Count Of Quotes
0
5
10
15
20
25
GeographicField49B

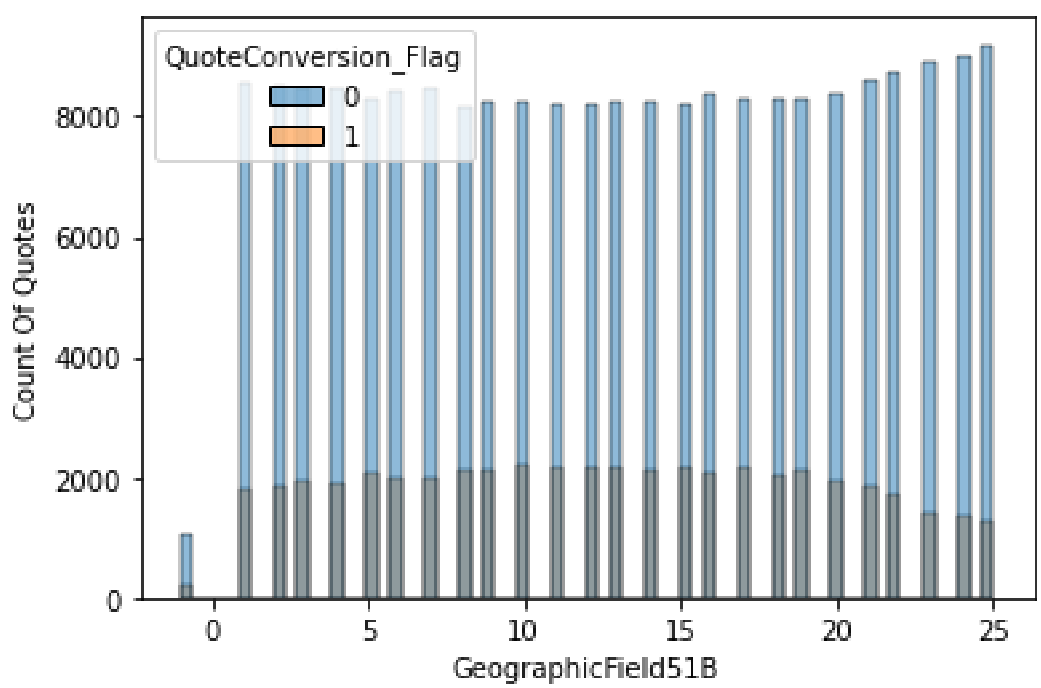


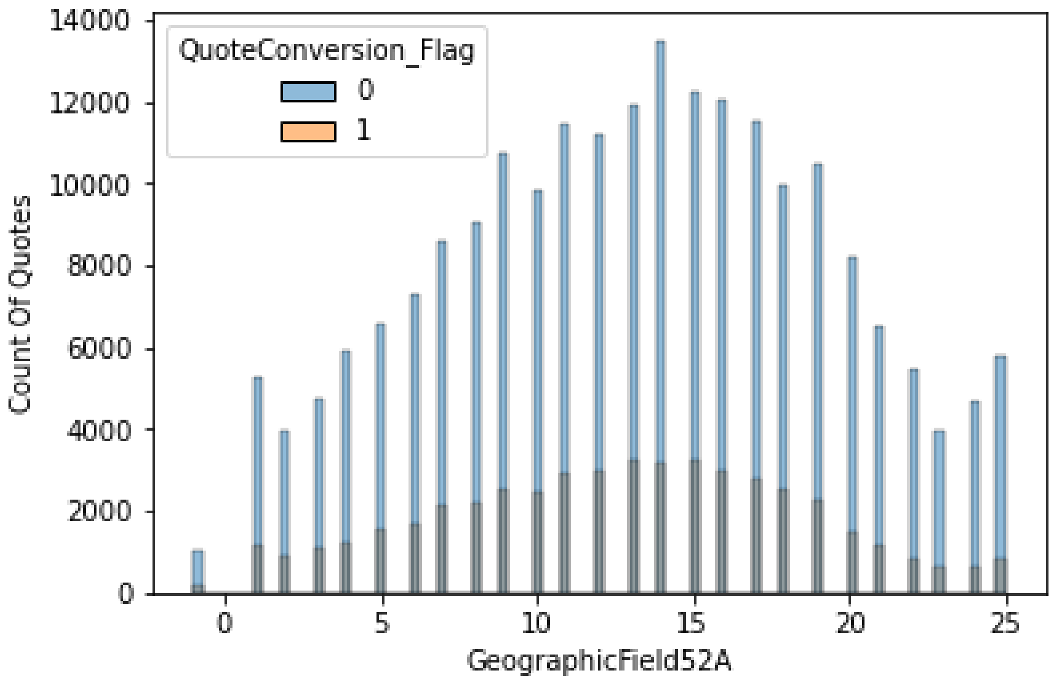
QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField50A

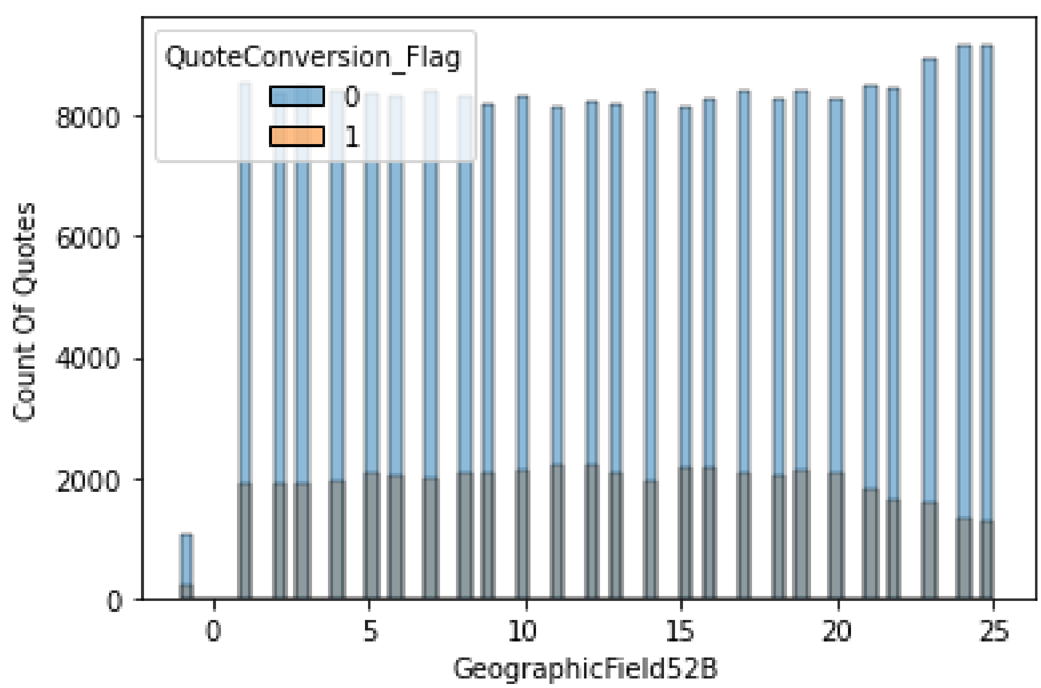


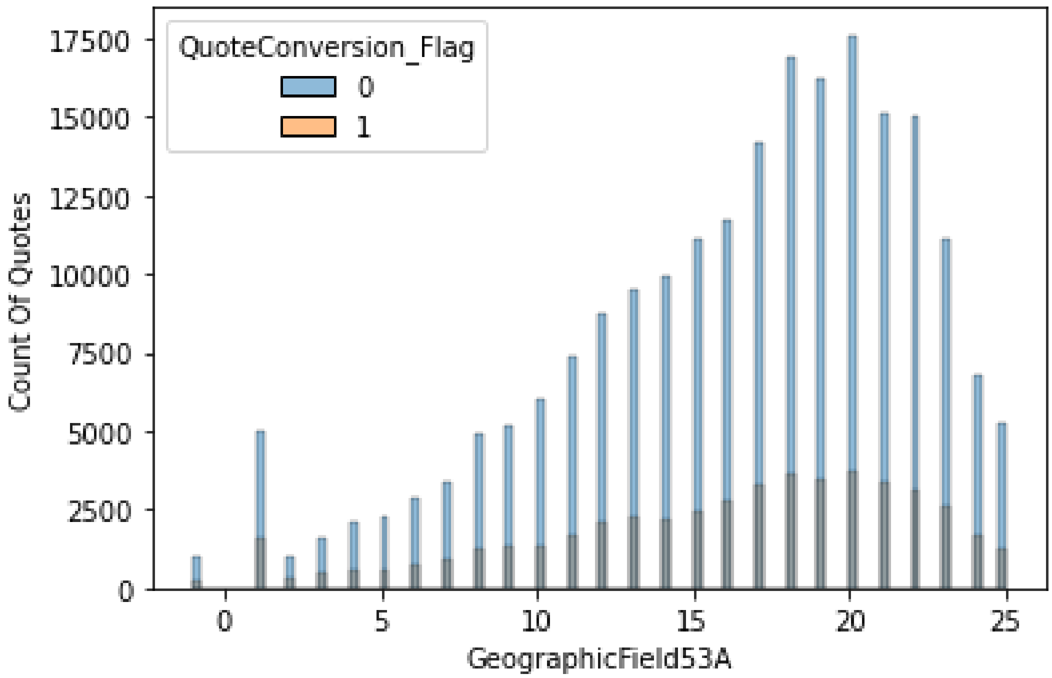
QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField50B

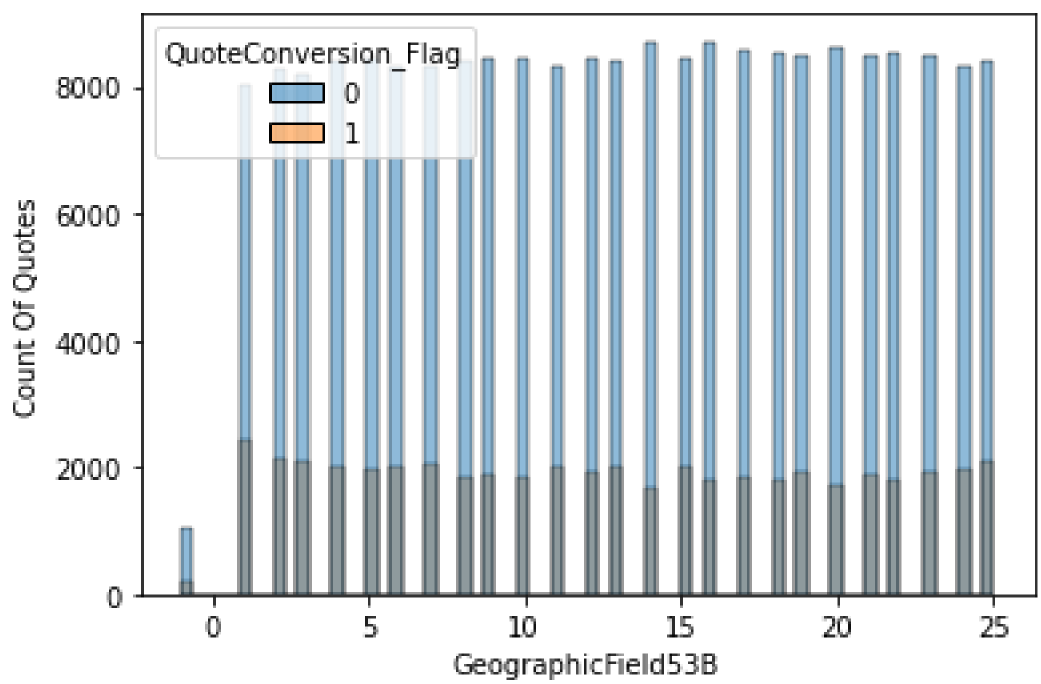

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField51A
0
2000
4000
6000
8000
10000
12000
14000
0
5
10
15
20
25

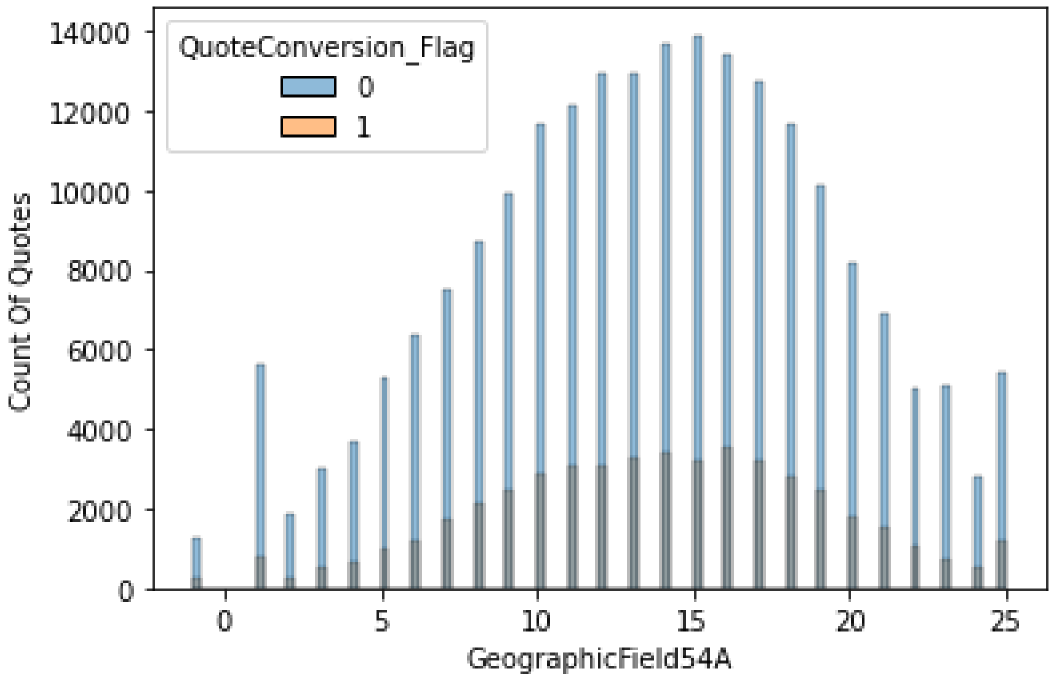

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField51B
0
2000
4000
6000
8000
0
5
10
15
20
25

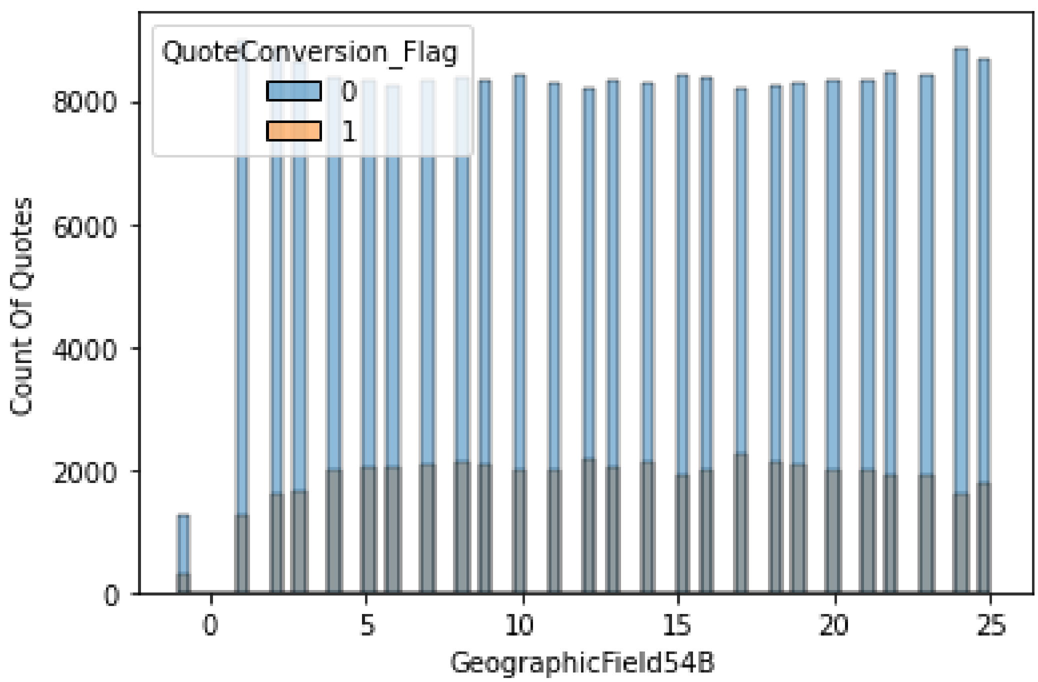

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField52A
0
2000
4000
6000
8000
10000
12000
14000
0
5
10
15
20
25

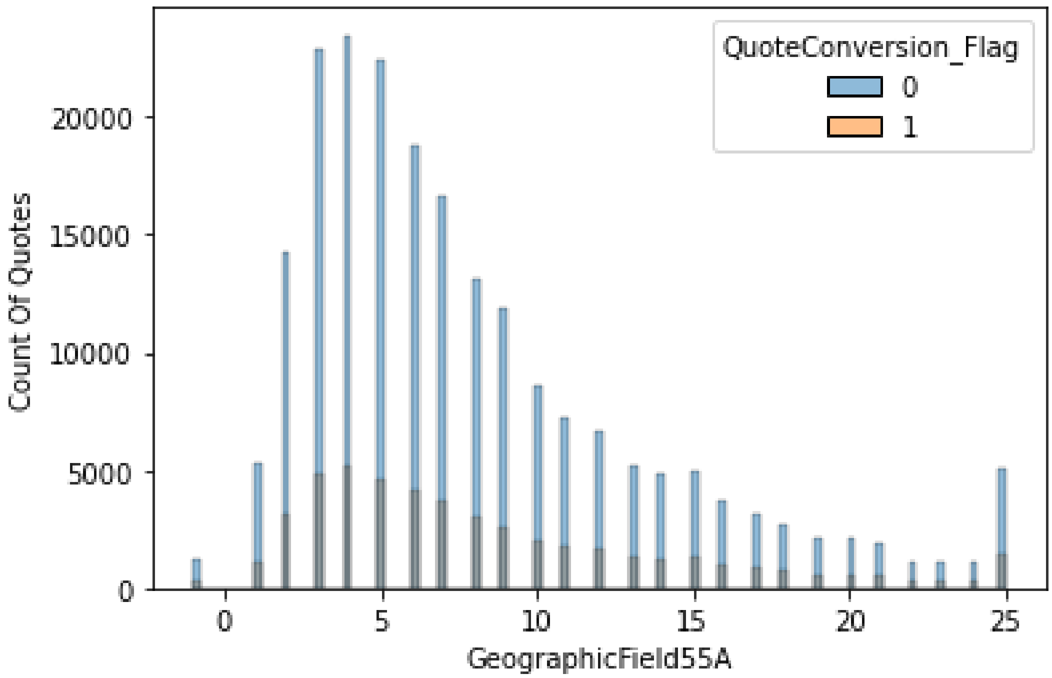

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField52B
0
2000
4000
6000
8000
0
5
10
15
20
25

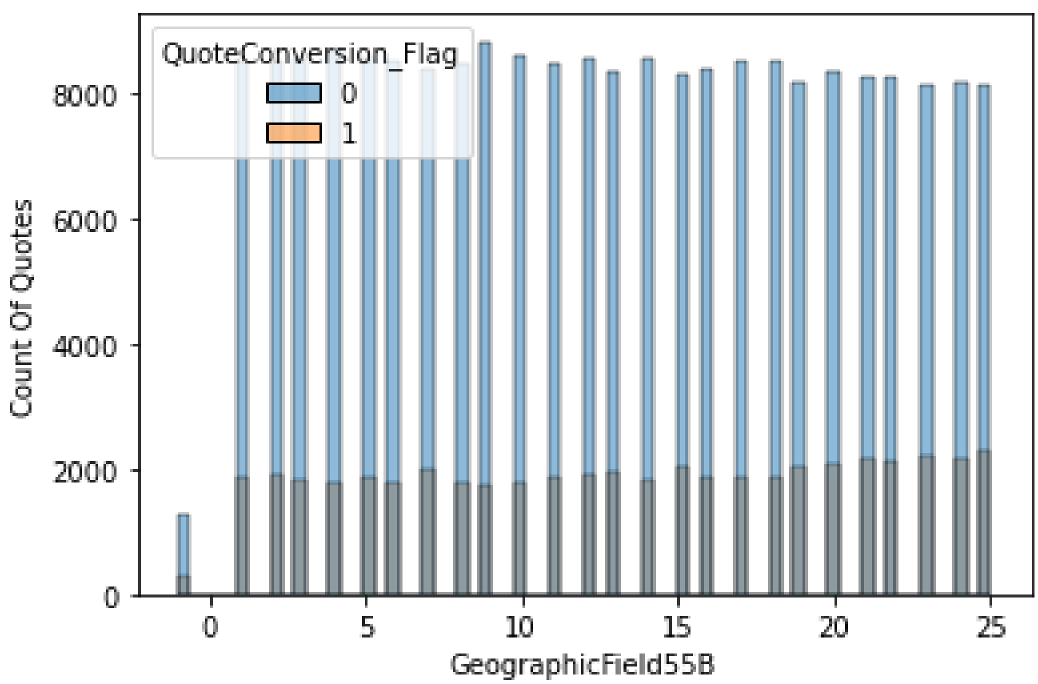

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField53A
0
2500
5000
7500
10000
12500
15000
17500
0
5
10
15
20
25

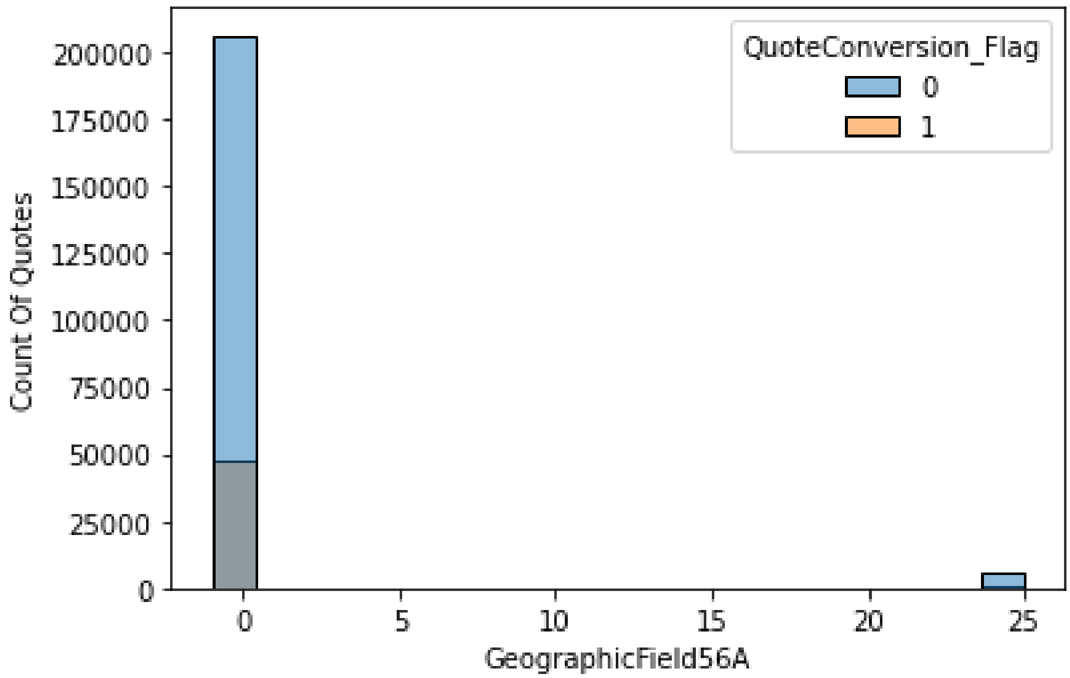

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField53B
0
2000
4000
6000
8000
0
5
10
15
20
25


QuoteConversion_Flag
0
1
Count Of Quotes
0
2000
4000
6000
8000
10000
12000
14000
0
5
10
15
20
25
GeographicField54A

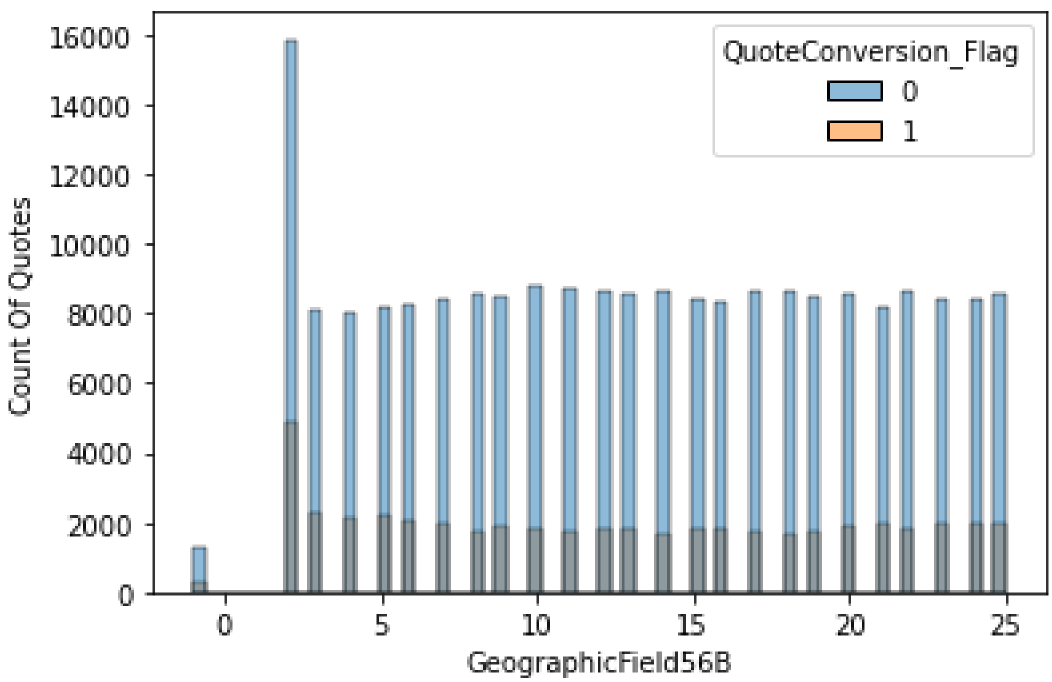

QuoteConversion_Flag
0
1
Count Of Quotes
0
2000
4000
6000
8000
0
5
10
15
20
25
GeographicField54B


QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField55A
0
5000
10000
15000
20000
0
5
10
15
20
25

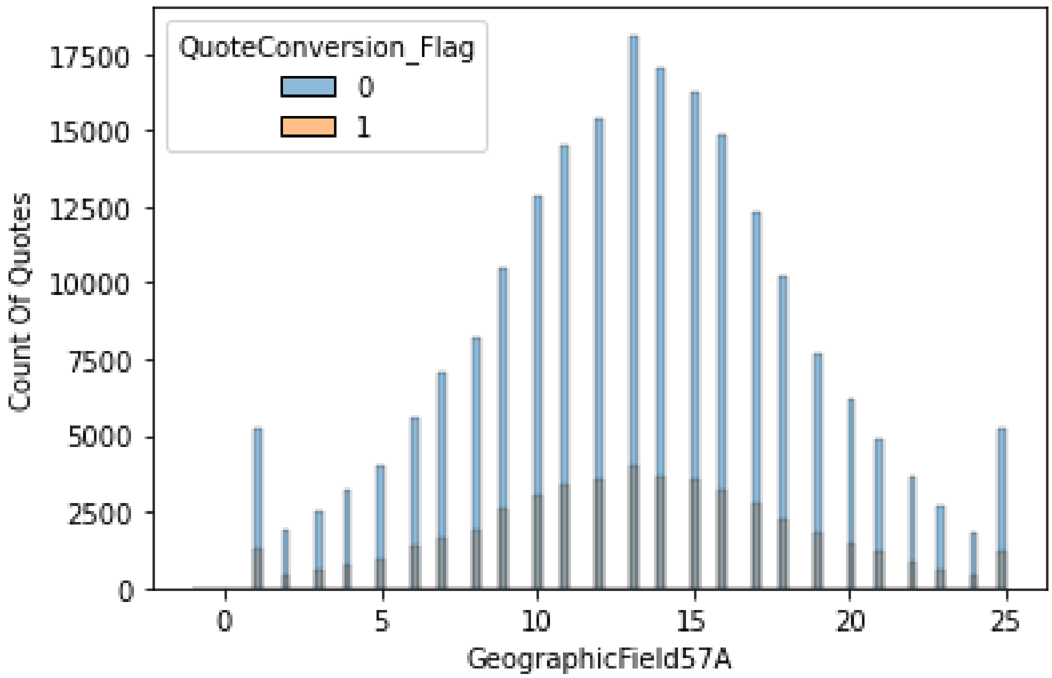

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField55B
0
2000
4000
6000
8000
0
5
10
15
20
25

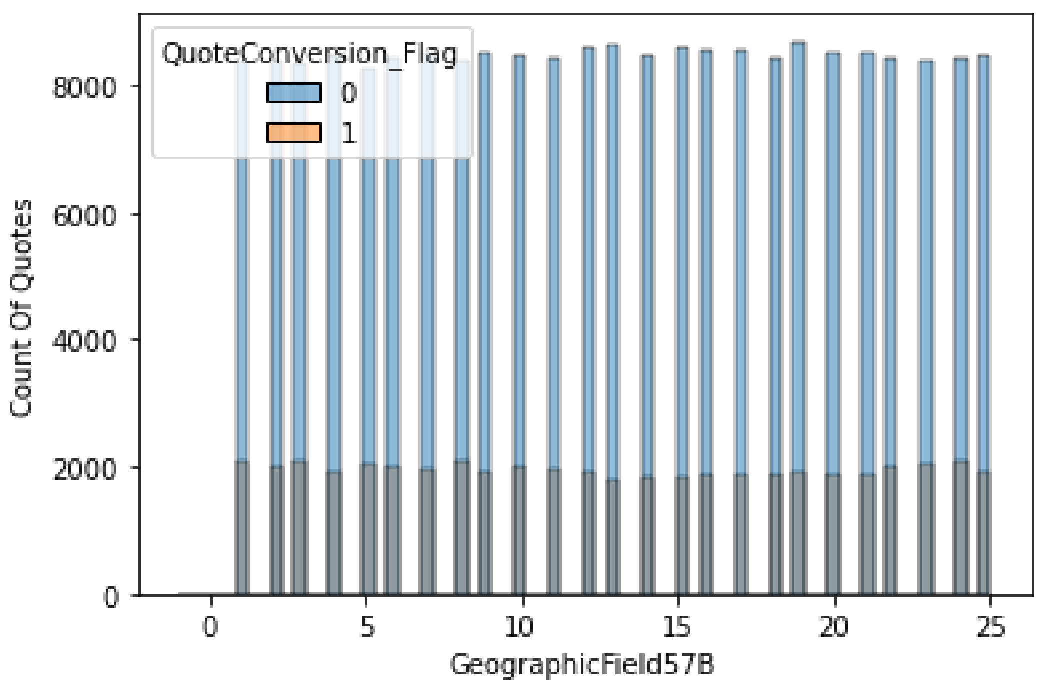


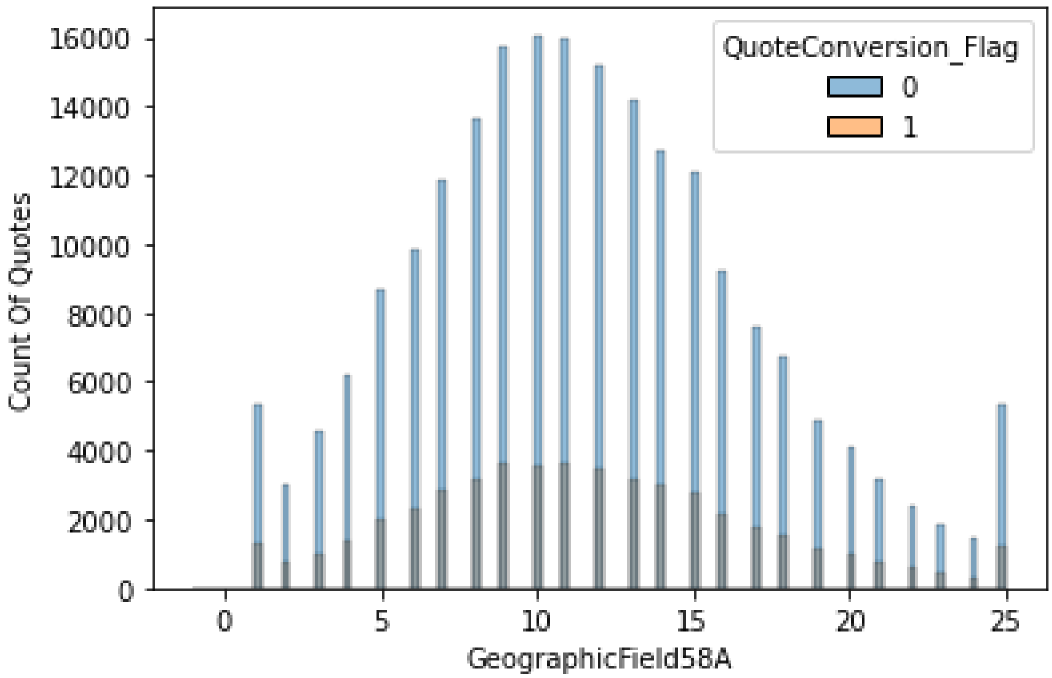
QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
0
5
10
15
20
25
GeographicField56A

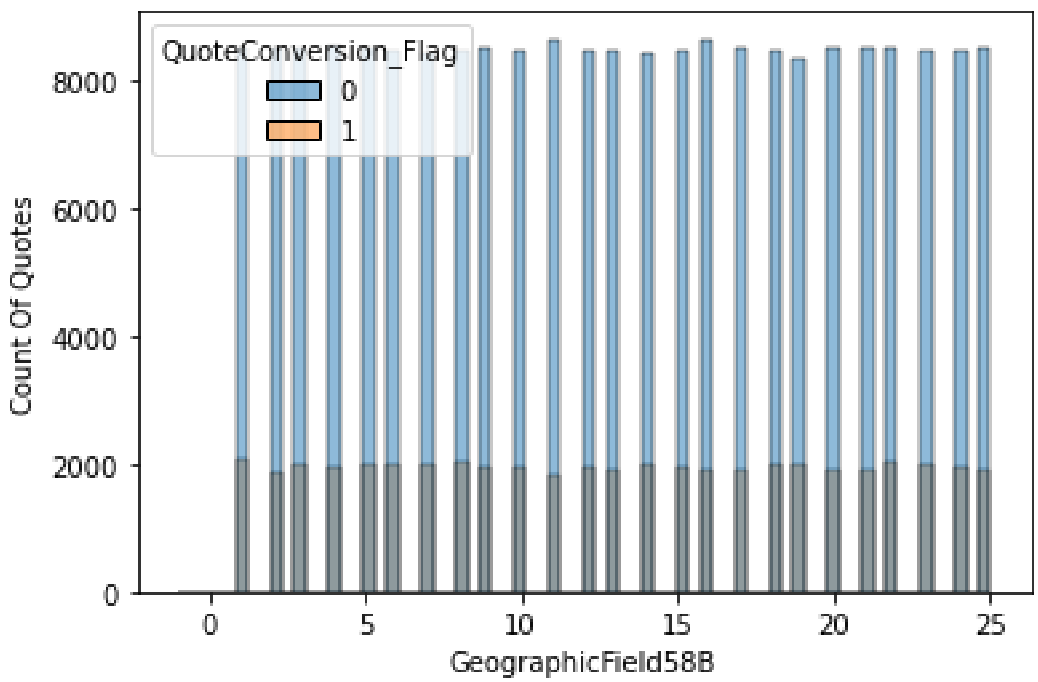


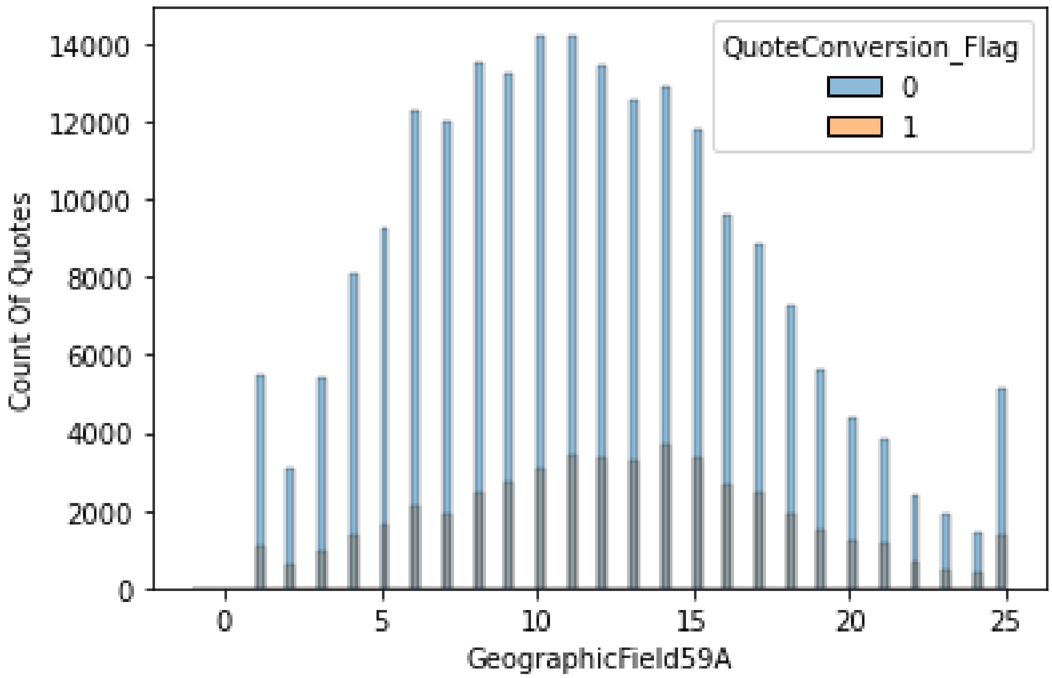
QuoteConversion_Flag
0
1
Count Of Quotes
16000
14000
12000
10000
8000
6000
4000
2000
0
0
5
10
15
20
25
GeographicField56B

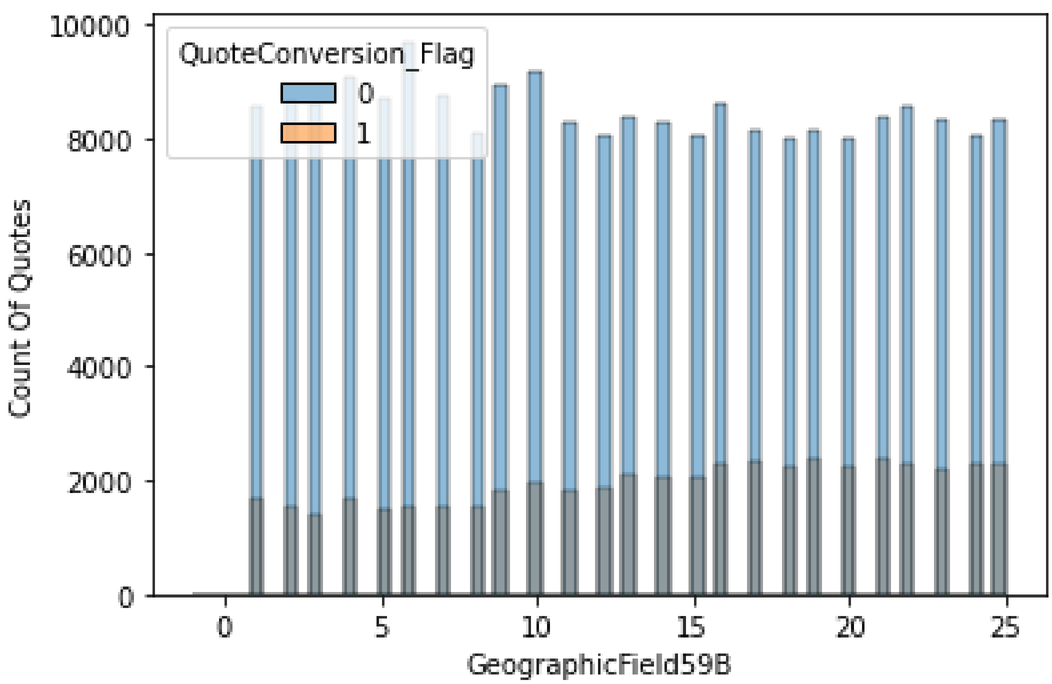

QuoteConversion_Flag
0
1
Count Of Quotes
17500
15000
12500
10000
7500
5000
2500
0
0
5
10
15
20
25
GeographicField57A

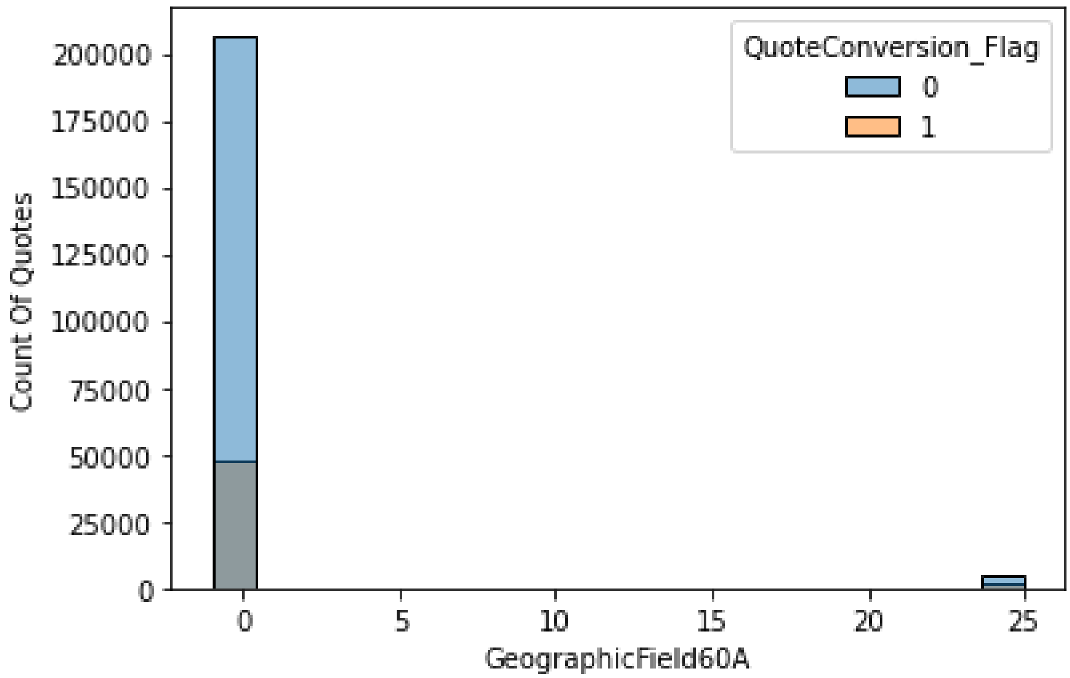

QuoteConversion_Flag
0
1
Count Of Quotes
8000
6000
4000
2000
0
0
5
10
15
20
25
GeographicField57B

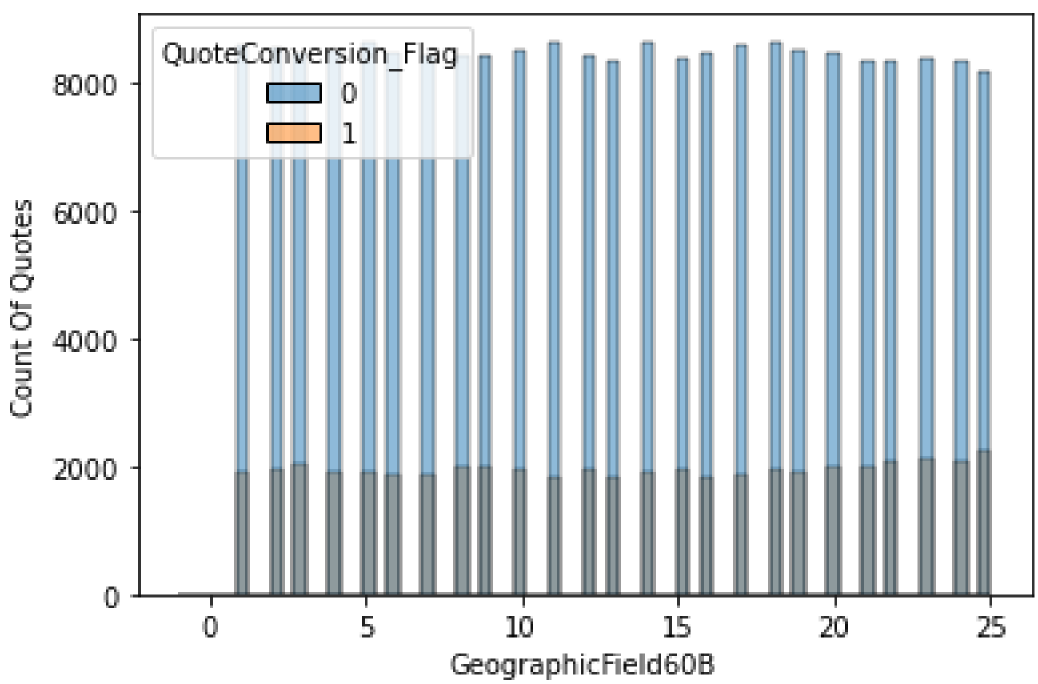

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField58A
0
5
10
15
20
25
0
2000
4000
6000
8000
10000
12000
14000
16000

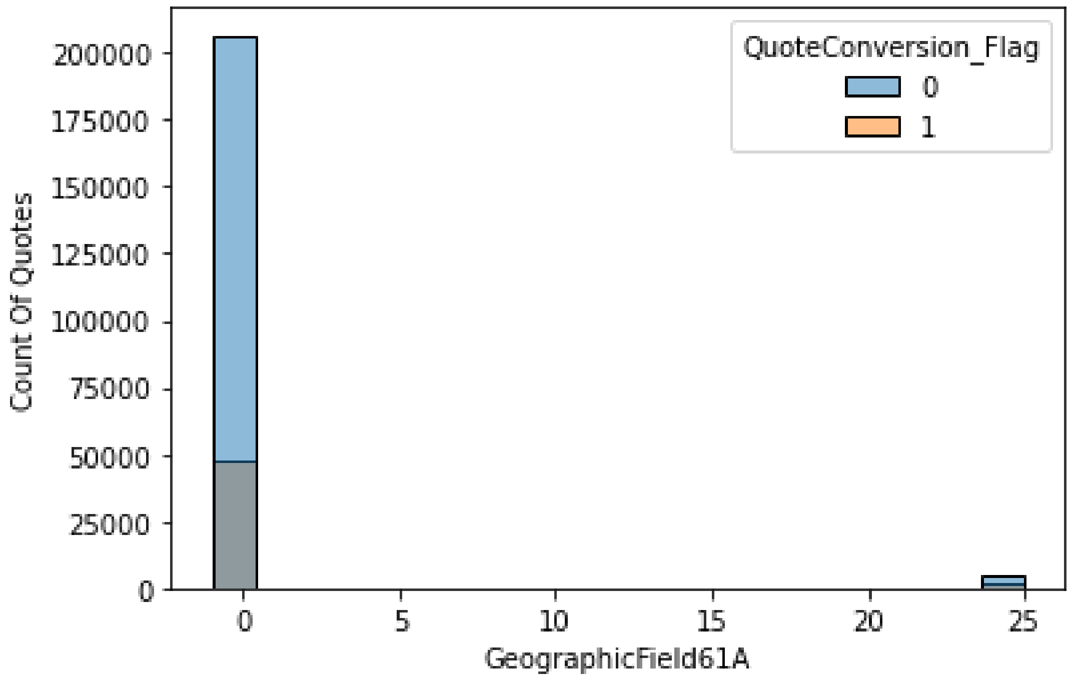

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField58B
0
5
10
15
20
25
0
2000
4000
6000
8000

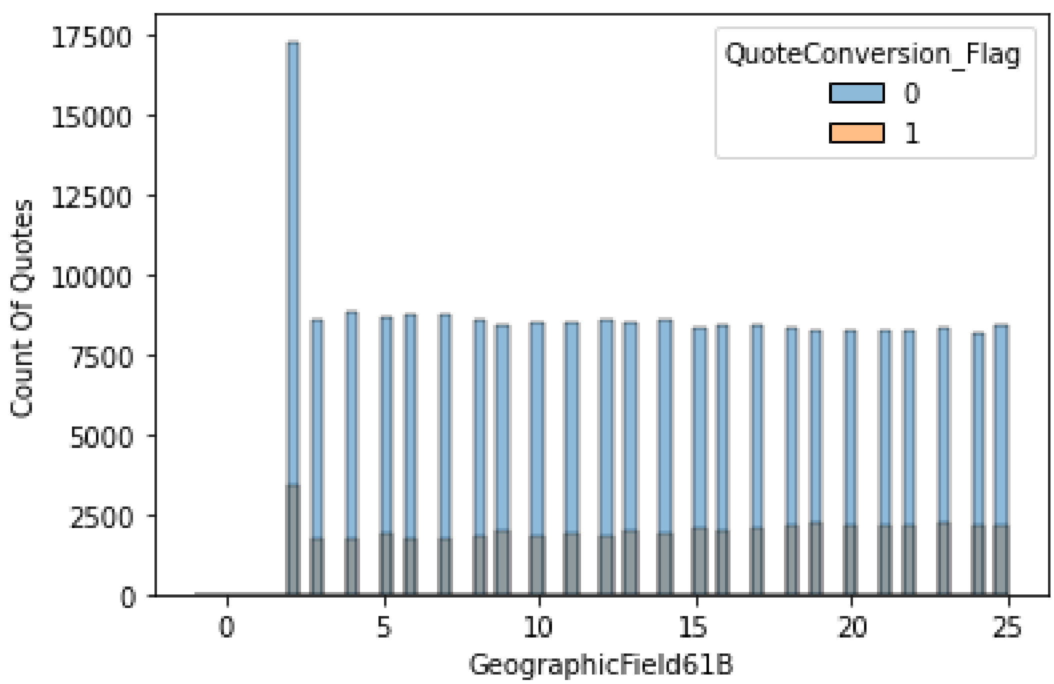

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField59A
0
2000
4000
6000
8000
10000
12000
14000
0
5
10
15
20
25

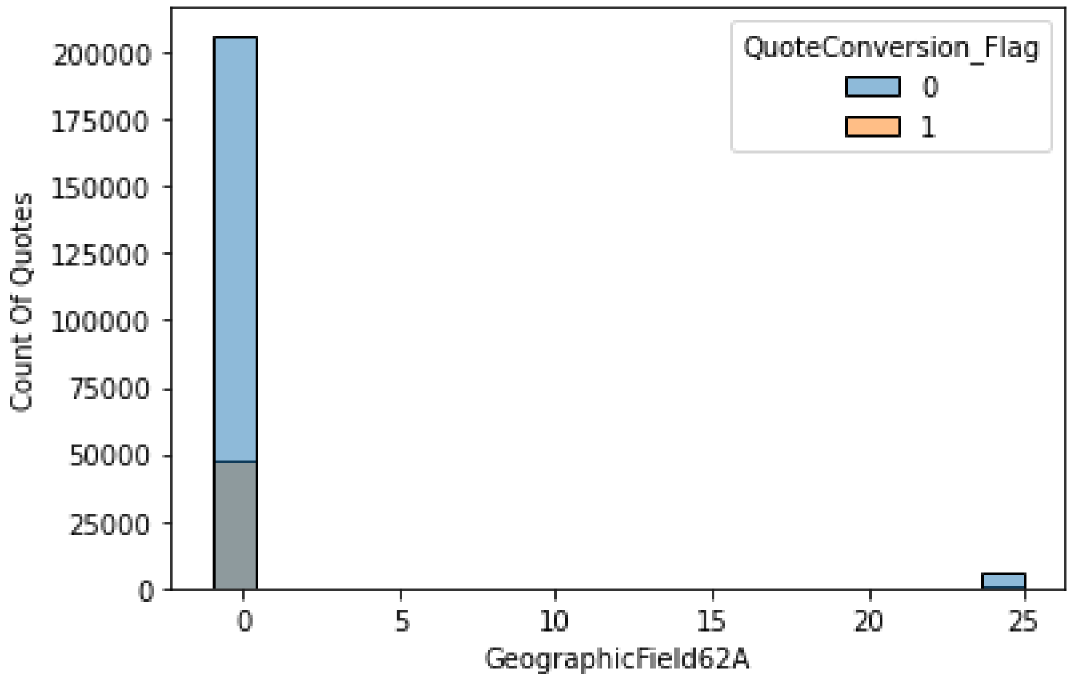

QuoteConversion_Flag
0
1
Count Of Quotes
GeographicField59B
0
2000
4000
6000
8000
10000
0
5
10
15
20
25

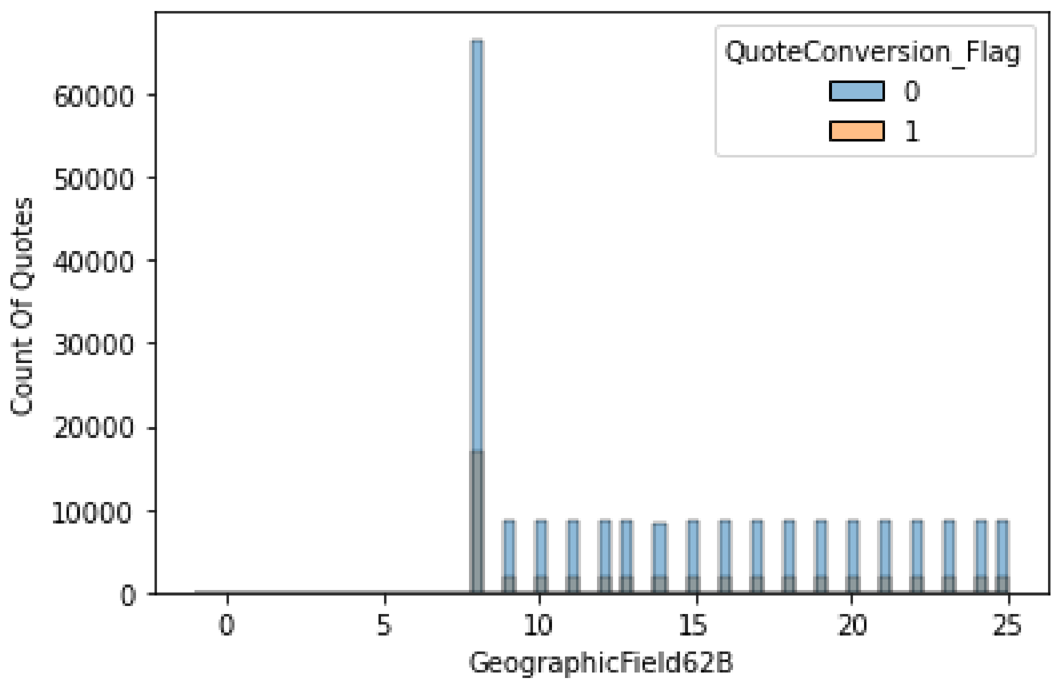

QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
0
5
10
15
20
25
GeographicField60A


QuoteConversion_Flag
0
1
Count Of Quotes
8000
6000
4000
2000
0
0
5
10
15
20
25
GeographicField60B


QuoteConversion_Flag
0
1
Count Of Quotes
200000
175000
150000
125000
100000
75000
50000
25000
0
0
5
10
15
20
25
GeographicField61A

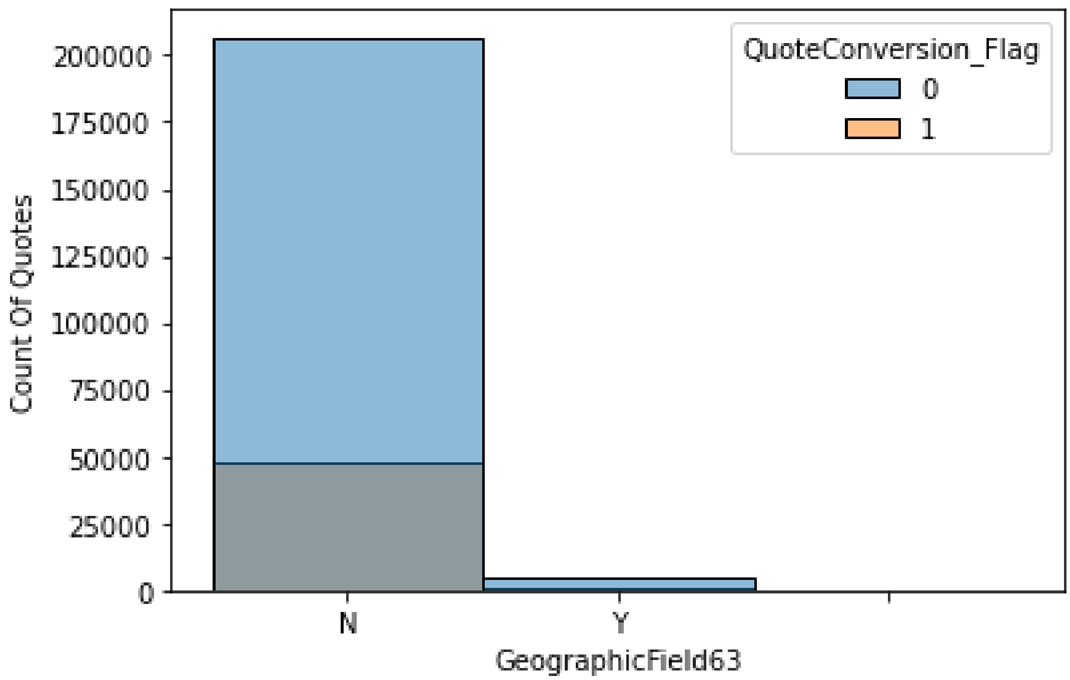

QuoteConversion_Flag
0
1
Count Of Quotes
17500
15000
12500
10000
7500
5000
2500
0
0
5
10
15
20
25
GeographicField61B

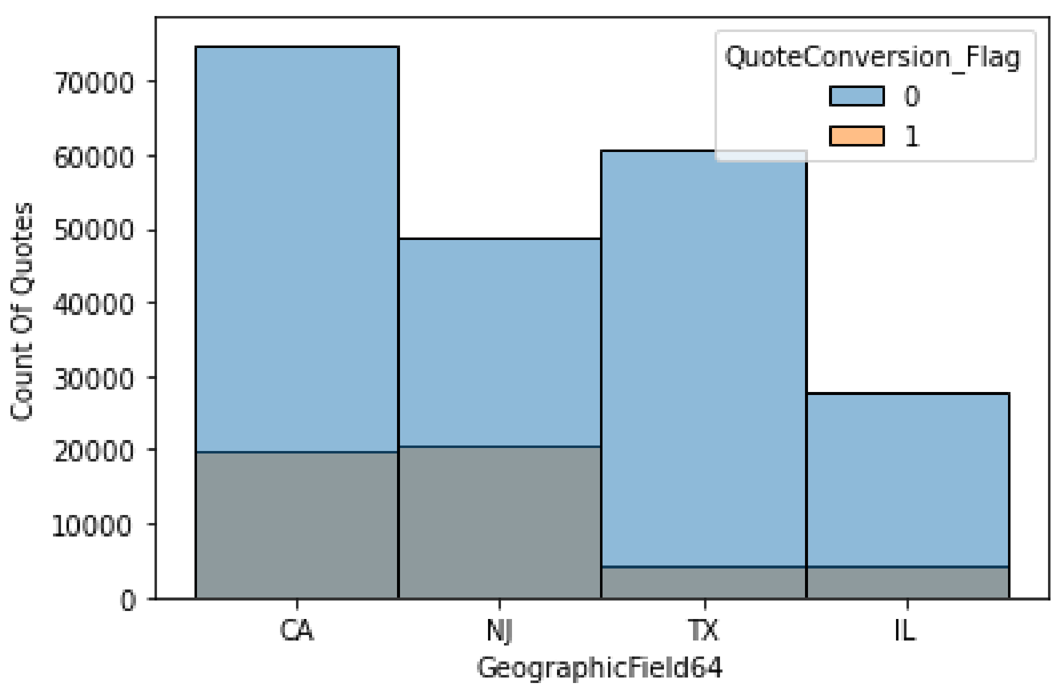




CATEGORICAL FEATURES

[53]:
```
for feature in get_categorical_features(dataset):
        sns.histplot(data = dataset, x = feature, hue = 'QuoteConversion_Flag')
        plt.xlabel(feature)
        plt.ylabel('Count Of Quotes')
        plt.show()
```



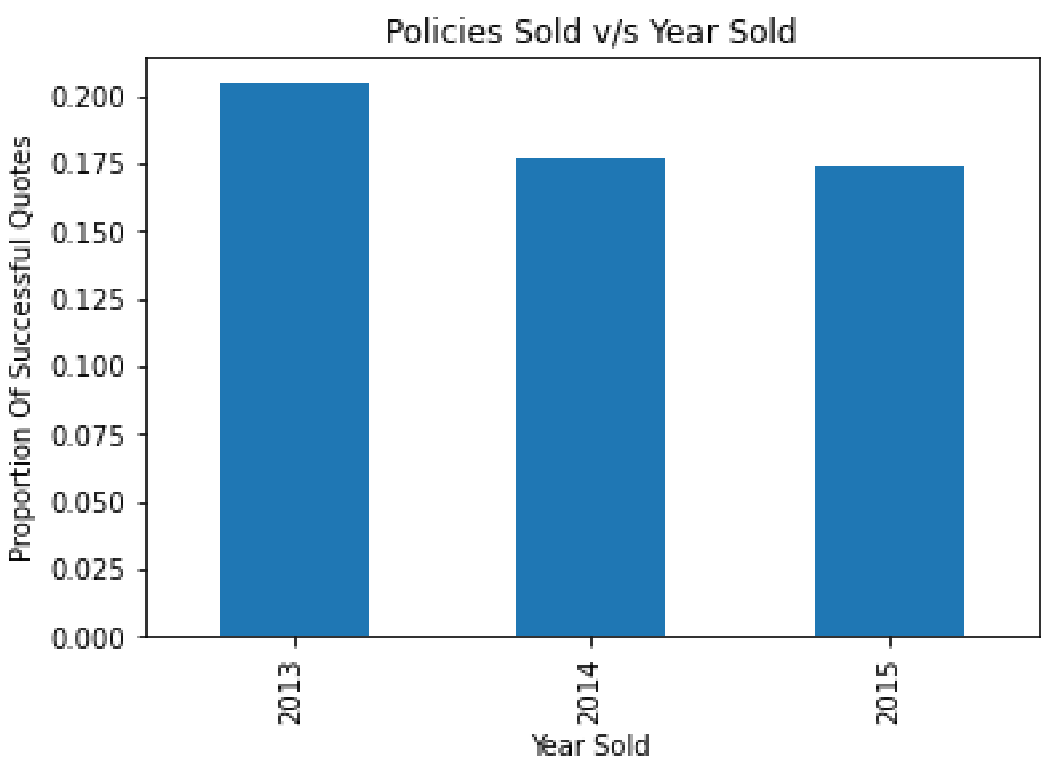


OBSERVATIONS FOR PERSONAL, PROPERTY & GEOGRAPHIC FEATURES: * Similar observations like done for Field, Coverage and Sales features can be done using the distribution curves plotted but since the nature of the features being used is anonymised they dont provide much insight into what exactly the features are depicting.

ORIGNAL_QUOTE_DATE : TEMPORAL VARIABLE

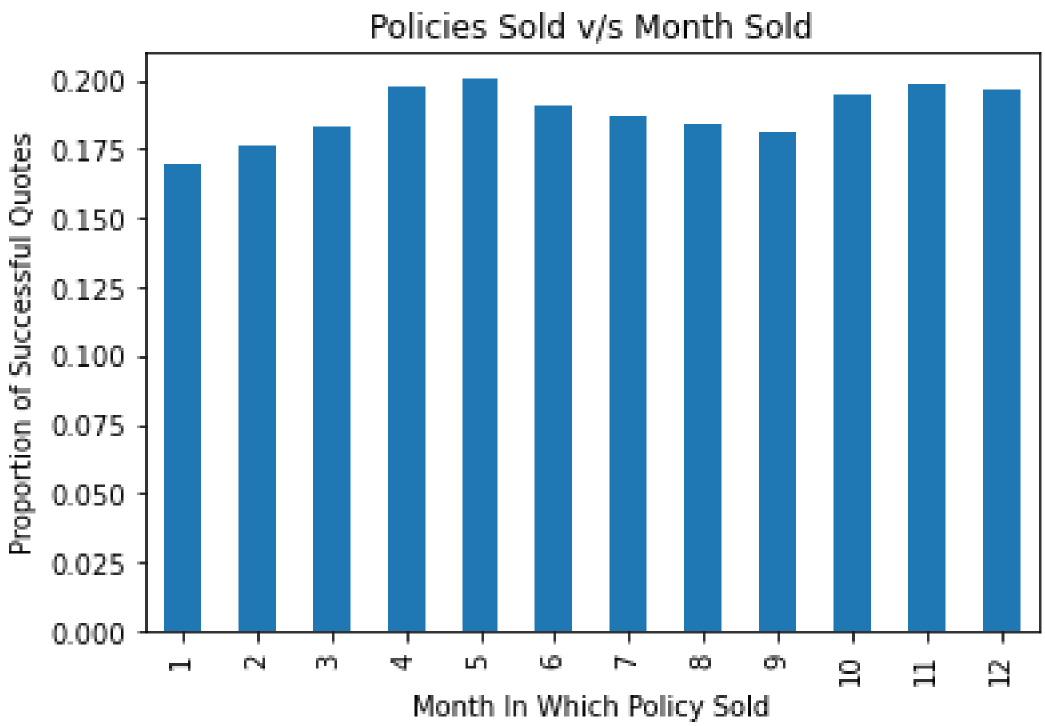
[54]:
```
data['Original_Quote_Date'] = pd.to_datetime(data['Original_Quote_Date'],␣
↪format='%Y-%m-%d')
data['Original_Quote_Day'] = data['Original_Quote_Date'].apply(lambda x: x.day)
data['Original_Quote_Month'] = data['Original_Quote_Date'].apply(lambda x: x.
↪month)
data['Original_Quote_Quater'] = data['Original_Quote_Date'].apply(lambda x:␣
↪math.ceil(x.month/3))
data['Original_Quote_Year'] = data['Original_Quote_Date'].apply(lambda x: x.
↪year)
```

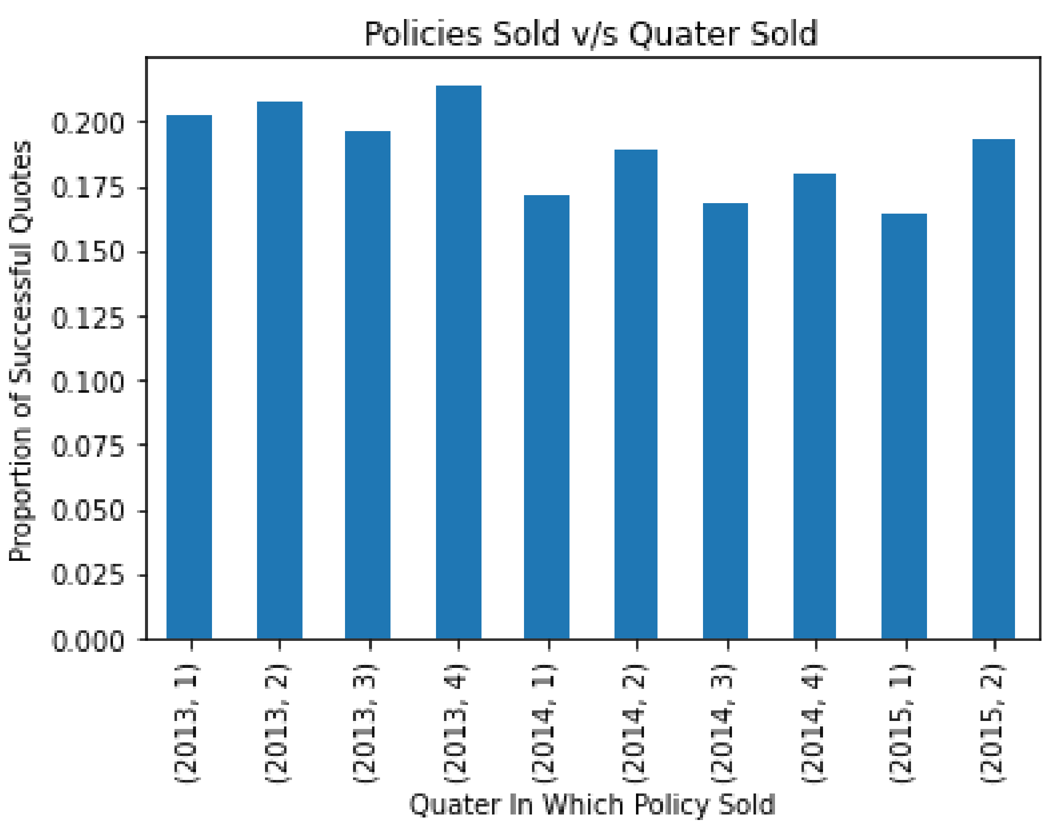
[55]:
```
(data.groupby(['Original_Quote_Year'])['QuoteConversion_Flag'].sum()/data.
↪groupby(['Original_Quote_Year'])['QuoteConversion_Flag'].count()).plot.bar()
plt.xlabel('Year Sold')
plt.ylabel('Proportion Of Successful Quotes')
plt.title('Policies Sold v/s Year Sold')
plt.show()
```

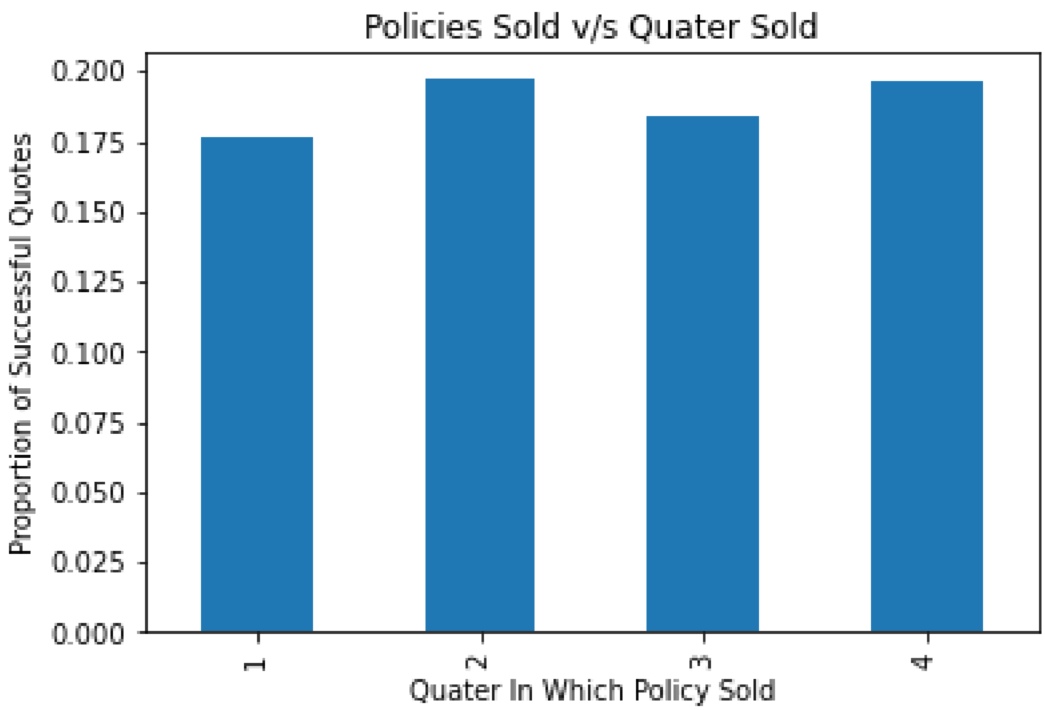


[56]:

```
(data.groupby(['Original_Quote_Month'])['QuoteConversion_Flag'].sum()/data.
 ↪groupby(['Original_Quote_Month'])['QuoteConversion_Flag'].count()).plot.bar()
plt.xlabel('Month In Which Policy Sold')
plt.ylabel('Proportion of Successful Quotes')
plt.title('Policies Sold v/s Month Sold')
plt.show()

(data.
 ↪groupby(['Original_Quote_Year','Original_Quote_Quater'])['QuoteConversion_Flag'].
 ↪sum()/data.
 ↪groupby(['Original_Quote_Year','Original_Quote_Quater'])['QuoteConversion_Flag'].
 ↪count()).plot.bar()
plt.xlabel('Quater In Which Policy Sold')
plt.ylabel('Proportion of Successful Quotes')
plt.title('Policies Sold v/s Quater Sold')
plt.show()

(data.groupby(['Original_Quote_Quater'])['QuoteConversion_Flag'].sum()/data.
 ↪groupby(['Original_Quote_Quater'])['QuoteConversion_Flag'].count()).plot.
 ↪bar()
plt.xlabel('Quater In Which Policy Sold')
plt.ylabel('Proportion of Successful Quotes')
plt.title('Policies Sold v/s Quater Sold')
plt.show()
```


Policies Sold v/s Month Sold
Proportion of Successful Quotes
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
0.200
1
2
3
4
5
6
7
8
9
10
11
12
Month In Which Policy Sold


Policies Sold v/s Quater Sold
Proportion of Successful Quotes
0.000
0.025
0.050
0.075
0.100
0.125
0.150
0.175
0.200
(2013, 1)
(2013, 2)
(2013, 3)
(2013, 4)
(2014, 1)
(2014, 2)
(2014, 3)
(2014, 4)
(2015, 1)
(2015, 2)
Quater In Which Policy Sold



OBSERVATIONS: 1. There is a decline in proportion of successful conversions with each passing year. 2. 2nd and 4th Quater i.e middle & fag end year months witness a rise in proportion of successful quotes in comparision to other two Quaters.

5. FEATURE ENGINEERING & SELECTION

HANDLING IMBALANCE IN DATASET

[57]:
```
unsuccessful_count, successful_count = data['QuoteConversion_Flag'].
 ↪value_counts()
df_successful = data[data['QuoteConversion_Flag'] == 1]
df_unsuccessful = data[data['QuoteConversion_Flag'] == 0]
```

[58]:
```
df_successful.shape,df_unsuccessful.shape
```

[58]:
```
((48894, 305), (211859, 305))
```

UNDERSAMPLING MAJORITY & UPSAMPLING MINORITY CLASS

[59]:
```
unsuccessful_undersampled = df_unsuccessful.sample(2 * successful_count)
successful_upsampled = df_successful.sample(int(0.46 * unsuccessful_count),␣
 ↪replace = True, ignore_index = True)
data = pd.concat([unsuccessful_undersampled, successful_upsampled], axis = 0)
```

```
data.shape
```

[59]:
```
(195243, 305)
```

[60]:
```
data['QuoteConversion_Flag'].value_counts()
```

[60]:
```
0    97788
1    97455
Name: QuoteConversion_Flag, dtype: int64
```

[61]:
```
data.head()
```

[61]:
```
        QuoteNumber Original_Quote_Date  QuoteConversion_Flag Field6  Field7  \
46351         76982          2014-06-11                     0      E      23
38529         64060          2015-01-08                     0      B      25
35128         58443          2014-08-14                     0      B      25
45313         75272          2015-05-06                     0      B       2
245213       408586          2015-05-05                     0      F      22

        Field8  Field9 Field10  Field11 Field12  CoverageField1A  \
46351   0.9392  0.0006   1,487   1.3045       N                8
38529   0.9153  0.0007     935   1.0200       N                5
35128   0.9153  0.0007     935   1.0200       N                8
45313   0.9153  0.0007     935   1.0200       N               10
245213  1.0101  0.0040     548   1.2694       N               24

        CoverageField1B  CoverageField2A  CoverageField2B  CoverageField3A  \
46351                14                8               14                7
38529                 6                5                6                7
35128                14                8               14               11
45313                18               11               18               14
245213               25               24               25               21

        CoverageField3B  CoverageField4A  CoverageField4B  CoverageField5A  \
46351                12                8               13               13
38529                11                4                6               13
35128                19                8               13               13
45313                21               10               17               25
245213               24               22               25               13

        CoverageField5B  CoverageField6A  CoverageField6B CoverageField8  \
46351                22               13               23              T
38529                22               13               23              Y
35128                22               13               23              T
45313                25               25               25              Y
245213               22               13               23              Y
```

	CoverageField9	CoverageField11A	CoverageField11B	SalesField1A	\
46351	F	4	6	2	
38529	E	11	21	2	
35128	J	4	6	5	
45313	E	6	13	5	
245213	E	3	4	10	

	SalesField1B	SalesField2A	SalesField2B	SalesField3	SalesField4	\
46351	2	2	1	1	5	
38529	1	3	6	1	4	
35128	15	5	16	0	3	
45313	15	4	12	0	3	
245213	21	3	10	0	5	

	SalesField5	SalesField6	SalesField7	SalesField8	SalesField9	\
46351	5	11	T	6377	0	
38529	3	20	Q	13261	0	
35128	4	11	Q	42564	0	
45313	4	11	T	12738	0	
245213	5	11	P	38907	0	

	SalesField10	SalesField11	SalesField12	SalesField13	SalesField14	\
46351	0	0	0	0	0	
38529	0	0	0	0	0	
35128	1	1	1	0	0	
45313	0	0	0	0	0	
245213	0	0	0	0	0	

	SalesField15	PersonalField1	PersonalField2	PersonalField4A	\
46351	0	1	1	15	
38529	0	0	0	25	
35128	0	1	1	14	
45313	0	1	1	3	
245213	0	1	1	9	

	PersonalField4B	PersonalField5	PersonalField6	PersonalField7	\
46351	20	6	1	N	
38529	25	7	0	N	
35128	19	6	1	N	
45313	3	7	0	N	
245213	12	6	1	N	

	PersonalField8	PersonalField9	PersonalField10A	PersonalField10B	\
46351	1	2	7	14	
38529	1	2	6	11	
35128	1	2	7	14	
45313	1	2	6	10	

```
245213              1              3              -1              -1

        PersonalField11  PersonalField12  PersonalField13  PersonalField14  \
46351                 0                1                2                2
38529                 0                1                2                2
35128                 0                1                2                2
45313                 0                1                2                2
245213                0                5                2                2

        PersonalField15 PersonalField16 PersonalField17 PersonalField18  \
46351                24              ZA              ZE              XR
38529                24              ZA              ZE              XR
35128                24              ZA              ZE              XR
45313                24              ZA              ZE              XR
245213                6              XJ              XV              YI

       PersonalField19  PersonalField22  PersonalField23  PersonalField24  \
46351               XD                1                0                0
38529               XD                1                0                0
35128               XD                1                0                0
45313               XD                1                0                0
245213              XC                1                0                0

        PersonalField25  PersonalField26  PersonalField27  PersonalField28  \
46351                 0                0                1                2
38529                 0                0                1                2
35128                 0                0                1                2
45313                 0                0                1                1
245213                0                0                0                1

        PersonalField29  PersonalField30  PersonalField31  PersonalField32  \
46351                 2                0                0                1
38529                 2                0                0                1
35128                 2                0                0                1
45313                 1                0                0                0
245213                1                0                0                0

        PersonalField33  PersonalField34  PersonalField35  PersonalField36  \
46351                 1                1                0                0
38529                 1                1                0                0
35128                 1                1                0                0
45313                 1                1                0                0
245213                0                1                0                0

        PersonalField37  PersonalField38  PersonalField39  PersonalField40  \
46351                 0                0                0                0
38529                 0                0                0                0
```

```
35128                0                0                0                0
45313                0                0                0                0
245213               0                0                0                0

        PersonalField41  PersonalField42  PersonalField43  PersonalField44  \
46351                0                0                1                0
38529                0                0                1                0
35128                0                0                1                0
45313                0                0                1                0
245213               0                0                1                0

        PersonalField45  PersonalField46  PersonalField47  PersonalField48  \
46351                0                1                1                2
38529                0                1                1                2
35128                0                1                1                2
45313                0                0                1                1
245213               0                0                0                1

        PersonalField49  PersonalField50  PersonalField51  PersonalField52  \
46351                0                0                0                0
38529                0                0                0                0
35128                0                0                0                0
45313                0                0                0                0
245213               0                0                0                0

        PersonalField53  PersonalField54  PersonalField55  PersonalField56  \
46351                1                0                0                0
38529                1                0                0                0
35128                1                0                0                0
45313                1                0                0                0
245213               1                0                0                0

        PersonalField57  PersonalField58  PersonalField59  PersonalField60  \
46351                0                1                0                0
38529                0                1                0                0
35128                0                1                0                0
45313                0                1                0                0
245213               0                1                0                0

        PersonalField61  PersonalField62  PersonalField63  PersonalField64  \
46351                0                0                1                0
38529                0                0                1                0
35128                0                0                1                0
45313                0                0                1                0
245213               0                0                1                0

        PersonalField65  PersonalField66  PersonalField67  PersonalField68  \
```

```
46351                0              0              0              1
38529                0              0              0              1
35128                0              0              0              1
45313                0              0              0              1
245213               0              0              0              1

        PersonalField69  PersonalField70  PersonalField71  PersonalField72  \
46351                0                0                0                0
38529                0                0                0                0
35128                0                0                0                0
45313                0                0                0                0
245213               0                0                0                0

        PersonalField73  PersonalField74  PersonalField75  PersonalField76  \
46351                1                0                0                1
38529                1                0                0                1
35128                1                0                0                1
45313                1                0                0                0
245213               1                0                0                0

        PersonalField77  PersonalField78  PersonalField79  PersonalField80  \
46351                1                2                0                0
38529                1                2                0                0
35128                1                2                0                0
45313                1                1                0                0
245213               0                1                0                0

        PersonalField81  PersonalField82  PersonalField83  PersonalField84  \
46351                0                0                1              2.0
38529                0                0                1              2.0
35128                0                0                1              2.0
45313                0                0                1              2.0
245213               0                0                1            100.0

        PropertyField1A  PropertyField1B  PropertyField2A  PropertyField2B  \
46351               11               16               -1                5
38529                6                9               -1               23
35128               14               18               -1               24
45313                7               10               -1               24
245213              15               20               -1               15

       PropertyField3 PropertyField4 PropertyField5  PropertyField6  \
46351               N              N              Y               0
38529               N              N              Y               0
35128               N              N              Y               0
45313               N              N              Y               0
245213              N              N              Y               0
```

	PropertyField7	PropertyField8	PropertyField9	PropertyField10	\
46351	D	0	0	1	
38529	O	1	0	1	
35128	O	1	0	1	
45313	A	1	0	1	
245213	R	1	0	1	

	PropertyField11A	PropertyField11B	PropertyField12	PropertyField13	\
46351	-1	21	4	2	
38529	-1	21	2	1	
35128	-1	21	2	2	
45313	-1	21	4	2	
245213	-1	21	4	1	

	PropertyField14	PropertyField15	PropertyField16A	PropertyField16B	\
46351	C	1	6	12	
38529	A	4	13	22	
35128	C	4	14	22	
45313	C	4	10	20	
245213	C	4	5	10	

	PropertyField17	PropertyField18	PropertyField19	PropertyField20	\
46351	0	0	0	0	
38529	0	1	1	0	
35128	1	2	0	0	
45313	1	3	0	0	
245213	0	2	1	0	

	PropertyField21A	PropertyField21B	PropertyField22	PropertyField23	\
46351	8	14	2	1	
38529	5	6	2	1	
35128	8	14	2	11	
45313	10	18	2	2	
245213	24	25	2	1	

	PropertyField24A	PropertyField24B	PropertyField25	PropertyField26A	\
46351	9	13	3.0	3	
38529	3	3	2.0	2	
35128	8	11	1.0	14	
45313	10	14	2.0	6	
245213	25	25	2.0	25	

	PropertyField26B	PropertyField27	PropertyField28	PropertyField29	\
46351	2	15	A	0.0	
38529	1	4	B	100.0	
35128	20	10	B	100.0	

```
45313                    7                 4                B            100.0
245213                  25                13                B            100.0

       PropertyField30 PropertyField31 PropertyField32 PropertyField33  \
46351                N               O               N               G
38529                N               O               Y               H
35128                N               O               Y               H
45313                N               O               Y               H
245213               N               N               Y               G

       PropertyField34  PropertyField35 PropertyField36 PropertyField37  \
46351                N                0               N               N
38529                N                2               N               N
35128                Y                2               Y               N
45313                N                2               N               N
245213               Y                2               N               Y

       PropertyField38  PropertyField39A  PropertyField39B  GeographicField1A  \
46351                N                11                12                 11
38529                N                 7                 6                 13
35128                N                 9                10                  6
45313                N                 8                 7                 12
245213               N                25                25                  7

        GeographicField1B  GeographicField2A  GeographicField2B  \
46351                  21                 19                 20
38529                  23                 21                 22
35128                  13                 16                 16
45313                  22                  7                  4
245213                 16                 19                 20

        GeographicField3A  GeographicField3B  GeographicField4A  \
46351                  13                 15                  7
38529                  17                 19                  8
35128                  13                 15                 17
45313                   9                  9                 10
245213                 14                 16                 15

        GeographicField4B  GeographicField5A  GeographicField5B  \
46351                   5                 -1                 13
38529                   5                 -1                 13
35128                  19                 -1                 13
45313                   9                 -1                 13
245213                 16                 -1                 17

        GeographicField6A  GeographicField6B  GeographicField7A  \
46351                   9                 17                 12
```

38529	2	5	3
35128	2	3	4
45313	2	3	4
245213	5	16	10

	GeographicField7B	GeographicField8A	GeographicField8B \
46351	18	9	17
38529	4	2	4
35128	5	2	5
45313	5	2	5
245213	16	6	16

	GeographicField9A	GeographicField9B	GeographicField10A \
46351	14	17	-1
38529	2	9	-1
35128	2	2	-1
45313	2	2	-1
245213	7	16	-1

	GeographicField10B	GeographicField11A	GeographicField11B \
46351	25	9	18
38529	25	1	1
35128	25	2	6
45313	25	2	6
245213	25	4	13

	GeographicField12A	GeographicField12B	GeographicField13A \
46351	13	17	9
38529	2	3	2
35128	3	6	2
45313	3	6	2
245213	10	16	5

	GeographicField13B	GeographicField14A	GeographicField14B \
46351	18	-1	18
38529	1	-1	7
35128	6	-1	7
45313	6	-1	7
245213	15	-1	18

	GeographicField15A	GeographicField15B	GeographicField16A \
46351	10	13	8
38529	4	5	2
35128	5	5	2
45313	5	5	2
245213	11	18	8

	GeographicField16B	GeographicField17A	GeographicField17B	\
46351	17	2	8	
38529	5	10	21	
35128	5	11	22	
45313	5	5	20	
245213	17	2	12	

	GeographicField18A	GeographicField18B	GeographicField19A	\
46351	-1	18	19	
38529	-1	18	14	
35128	-1	23	2	
45313	-1	11	2	
245213	-1	20	23	

	GeographicField19B	GeographicField20A	GeographicField20B	\
46351	18	5	13	
38529	15	22	24	
35128	3	3	5	
45313	3	3	5	
245213	23	23	24	

	GeographicField21A	GeographicField21B	GeographicField22A	\
46351	-1	24	-1	
38529	-1	6	-1	
35128	-1	7	-1	
45313	-1	4	-1	
245213	-1	17	-1	

	GeographicField22B	GeographicField23A	GeographicField23B	\
46351	24	-1	16	
38529	15	-1	5	
35128	15	-1	3	
45313	15	-1	4	
245213	18	-1	17	

	GeographicField24A	GeographicField24B	GeographicField25A	\
46351	5	2	5	
38529	20	21	19	
35128	13	8	14	
45313	13	8	16	
245213	21	22	18	

	GeographicField25B	GeographicField26A	GeographicField26B	\
46351	3	3	3	
38529	22	17	21	
35128	11	11	16	
45313	15	16	20	

```
245213                 19                 13                 17

        GeographicField27A  GeographicField27B  GeographicField28A  \
46351                    6                   3                   7
38529                   20                  19                  12
35128                   14                   9                   9
45313                   15                  11                  11
245213                  22                  21                   8

        GeographicField28B  GeographicField29A  GeographicField29B  \
46351                   10                  15                  22
38529                   20                  16                  23
35128                   14                  14                  21
45313                   19                  20                  24
245213                  14                  11                  16

        GeographicField30A  GeographicField30B  GeographicField31A  \
46351                    8                  13                  14
38529                   15                  24                   9
35128                    8                  14                  15
45313                   10                  19                   8
245213                   6                   7                  13

        GeographicField31B  GeographicField32A  GeographicField32B  \
46351                   21                  10                  17
38529                   13                   4                   4
35128                   22                  15                  22
45313                   11                  24                  25
245213                  21                  14                  22

        GeographicField33A  GeographicField33B  GeographicField34A  \
46351                    9                  13                  17
38529                   14                  21                  13
35128                   14                  21                   9
45313                    9                  14                  11
245213                   9                  15                  12

        GeographicField34B  GeographicField35A  GeographicField35B  \
46351                   22                  15                  23
38529                   16                   8                  14
35128                    9                  16                  23
45313                   13                  10                  17
245213                  14                  16                  23

        GeographicField36A  GeographicField36B  GeographicField37A  \
46351                   12                  19                  14
38529                   18                  24                  16
```

```
35128                   11                 17                 25
45313                   10                 14                  8
245213                  10                 14                  8

        GeographicField37B  GeographicField38A  GeographicField38B  \
46351                   23                  8                 12
38529                   24                 13                 20
35128                   25                 14                 22
45313                   18                 13                 20
245213                  18                 16                 23

        GeographicField39A  GeographicField39B  GeographicField40A  \
46351                   15                 23                  8
38529                    6                 12                  2
35128                    4                  6                  3
45313                   17                 24                  4
245213                  12                 22                 16

        GeographicField40B  GeographicField41A  GeographicField41B  \
46351                   13                 18                 16
38529                    2                 17                 15
35128                    7                  2                  3
45313                    9                  2                  1
245213                  23                 20                 18

        GeographicField42A  GeographicField42B  GeographicField43A  \
46351                   14                 19                 12
38529                    6                  6                  2
35128                    6                  6                  9
45313                    8                 10                  8
245213                  14                 20                 10

        GeographicField43B  GeographicField44A  GeographicField44B  \
46351                   19                 21                 22
38529                    2                  7                  2
35128                    8                 13                 14
45313                    7                 13                 14
245213                  12                 12                 12

        GeographicField45A  GeographicField45B  GeographicField46A  \
46351                    9                  9                 19
38529                   23                 24                 10
35128                    2                  4                 10
45313                    2                  3                 10
245213                   8                  8                 10

        GeographicField46B  GeographicField47A  GeographicField47B  \
```

46351	20	6	8	
38529	8	7	14	
35128	9	21	23	
45313	10	21	23	
245213	7	6	6	

	GeographicField48A	GeographicField48B	GeographicField49A	\
46351	21	21	21	
38529	18	17	20	
35128	15	13	13	
45313	8	5	9	
245213	21	21	19	

	GeographicField49B	GeographicField50A	GeographicField50B	\
46351	20	18	16	
38529	19	18	17	
35128	7	15	12	
45313	3	10	6	
245213	17	19	19	

	GeographicField51A	GeographicField51B	GeographicField52A	\
46351	15	15	15	
38529	10	8	10	
35128	11	9	11	
45313	8	6	7	
245213	16	17	16	

	GeographicField52B	GeographicField53A	GeographicField53B	\
46351	16	22	21	
38529	9	21	20	
35128	9	17	13	
45313	5	9	4	
245213	17	18	14	

	GeographicField54A	GeographicField54B	GeographicField55A	\
46351	8	5	4	
38529	12	10	4	
35128	18	20	9	
45313	20	22	21	
245213	10	8	3	

	GeographicField55B	GeographicField56A	GeographicField56B	\
46351	6	-1	5	
38529	7	-1	10	
35128	18	-1	18	
45313	24	25	25	
245213	4	-1	13	

	GeographicField57A	GeographicField57B	GeographicField58A	\
46351	18	21	3	
38529	1	1	3	
35128	7	3	16	
45313	11	8	20	
245213	15	17	14	

	GeographicField58B	GeographicField59A	GeographicField59B	\
46351	2	13	15	
38529	2	12	15	
35128	20	14	18	
45313	24	11	12	
245213	17	12	13	

	GeographicField60A	GeographicField60B	GeographicField61A	\
46351	-1	17	-1	
38529	-1	17	-1	
35128	-1	20	-1	
45313	-1	11	-1	
245213	-1	20	-1	

	GeographicField61B	GeographicField62A	GeographicField62B	\
46351	3	-1	8	
38529	8	-1	23	
35128	12	-1	11	
45313	16	-1	24	
245213	18	-1	8	

	GeographicField63	GeographicField64	PersonalField84_nan	\
46351	N	IL	0	
38529	N	CA	0	
35128	N	CA	0	
45313	N	CA	0	
245213	N	NJ	1	

	PropertyField29_nan	Original_Quote_Day	Original_Quote_Month	\
46351	0	11	6	
38529	1	8	1	
35128	1	14	8	
45313	1	6	5	
245213	1	5	5	

	Original_Quote_Quater	Original_Quote_Year
46351	2	2014
38529	1	2015
35128	3	2014

```
45313                            2                  2015
245213                           2                  2015
```

[62]:
```
class_labels = data['QuoteConversion_Flag']
data.drop(columns =␣
↪['QuoteNumber','QuoteConversion_Flag','Original_Quote_Date'], axis = 1,␣
↪inplace = True)
```

TRAIN & CV DATA SPLIT

[63]:
```
from sklearn.model_selection import train_test_split

x_train, x_cv, y_train, y_cv =␣
↪train_test_split(data,class_labels,stratify=class_labels,test_size=0.20)
x_train.shape, x_cv.shape, y_train.shape, y_cv.shape
```

[63]:
```
((156194, 302), (39049, 302), (156194,), (39049,))
```

ENCODING THE CATEGORICAL FEATURES

[64]:
```
# def encode_categorical_data(data,data_test):
#     """This function takes dataframe as an input and identifies all the␣
↪categorical variables and returns
#     a one hot encoded feature corresponding to each one of them"""

#     encoded_data = np.array([])
#     encoded_data_test = np.array([])
#     categorical_columns = []
#     flag = 0
#     for column in data.columns:
#         if type(data[column].iloc[0]) == str:
#             categorical_columns.append(column)
#             onehotencoder = OneHotEncoder(handle_unknown = 'ignore')
#             if flag == 0:
#                 encoded_data = onehotencoder.fit_transform(data[column].
↪values.reshape(-1,1)).toarray()
#                 encoded_data_test = onehotencoder.transform(data_test[column].
↪values.reshape(-1,1)).toarray()
#                 flag = 1
#             else:
#                 encoded_data = np.hstack((encoded_data,onehotencoder.
↪fit_transform(data[column].values.reshape(-1,1)).toarray()))
#                 encoded_data_test = np.
↪hstack((encoded_data_test,onehotencoder.transform(data_test[column].values.
↪reshape(-1,1)).toarray()))
#     return categorical_columns,encoded_data, encoded_data_test
```

[65]:
```python
def get_encoders(data):
    """This function takes dataframe as an input and identifies all the␣
↪categorical variables and returns
    a one hot encoded feature corresponding to each one of them"""

    encoders_dict = {}
    for column in data.columns:
        if type(data[column].iloc[0]) == str:
            onehotencoder = OneHotEncoder(handle_unknown = 'ignore')
            encoders_dict[column] = onehotencoder
            onehotencoder.fit(data[column].values.reshape(-1,1))
    return encoders_dict
```

[66]:
```python
def encode_categorical_data(data, encoders_dict):
    encoded_data = []
    flag = 0
    for column, encoder in encoders_dict.items():
        if flag == 0:
            encoded_data = encoder.transform(data[column].values.reshape(-1,1)).
↪toarray()
            flag = 1
        else:
            encoded_data = np.hstack((encoded_data,encoder.
↪transform(data[column].values.reshape(-1,1)).toarray()))
    return encoded_data
```

[67]:
```python
encoders_dict = get_encoders(x_train)
```

[68]:
```python
encoded_categorical_data = encode_categorical_data(x_train,encoders_dict)
encoded_categorical_data_test = encode_categorical_data(x_cv,encoders_dict)
```

[69]:
```python
# categorical_columns, encoded_categorical_data ,␣
↪encoded_categorical_data_test= encode_categorical_data(x_train, x_cv)
```

[70]:
```python
x_train.drop(labels = list(encoders_dict.keys()), axis = 1 , inplace = True)
x_cv.drop(labels = list(encoders_dict.keys()), axis = 1 , inplace = True)
```

[71]:
```python
x_train = np.hstack((x_train.to_numpy(),encoded_categorical_data))
x_cv = np.hstack((x_cv.to_numpy(),encoded_categorical_data_test))
```

[72]:
```python
x_train.shape, y_train.shape, x_cv.shape, y_train.shape
```

[72]:
```
((156194, 606), (156194,), (39049, 606), (156194,))
```

SAVING ONE HOT ENCODERS

[73]:
```
f = open('encoders_dict.pkl','wb')
pickle.dump(encoders_dict,f)
f.close()
```

DROPPING CONSTANT FEATURES AND FEATURES WITH LOW VARIANCE

[74]:
```
from sklearn.feature_selection import VarianceThreshold
vr = VarianceThreshold(threshold = 0.0)
vr.fit(x_train)
```

[74]:
```
VarianceThreshold()
```

[75]:
```
x_train = x_train[:,vr.get_support()]
x_cv = x_cv[:,vr.get_support()]
x_train.shape , x_cv.shape
```

[75]:
```
((156194, 603), (39049, 603))
```

[76]:
```
f = open('constant_features.pkl','wb')
pickle.dump(vr,f)
f.close()
```

STANDARDIZE DATASET

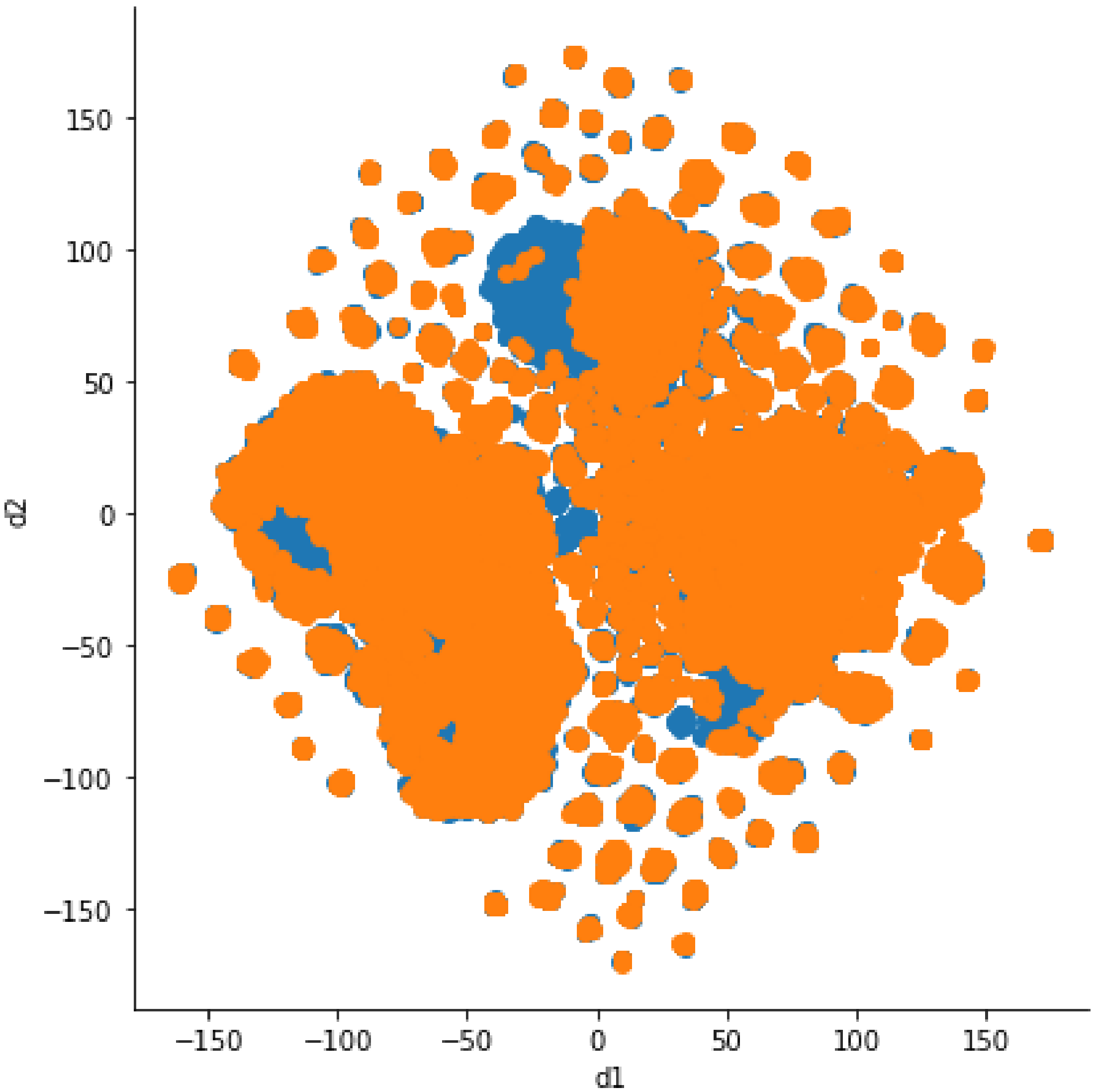
[77]:
```
from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()
x_train = scaler.fit_transform(x_train)
x_cv = scaler.transform(x_cv)
```

[78]:
```
f = open('feature_scaling.pkl','wb')
pickle.dump(scaler,f)
f.close()
```

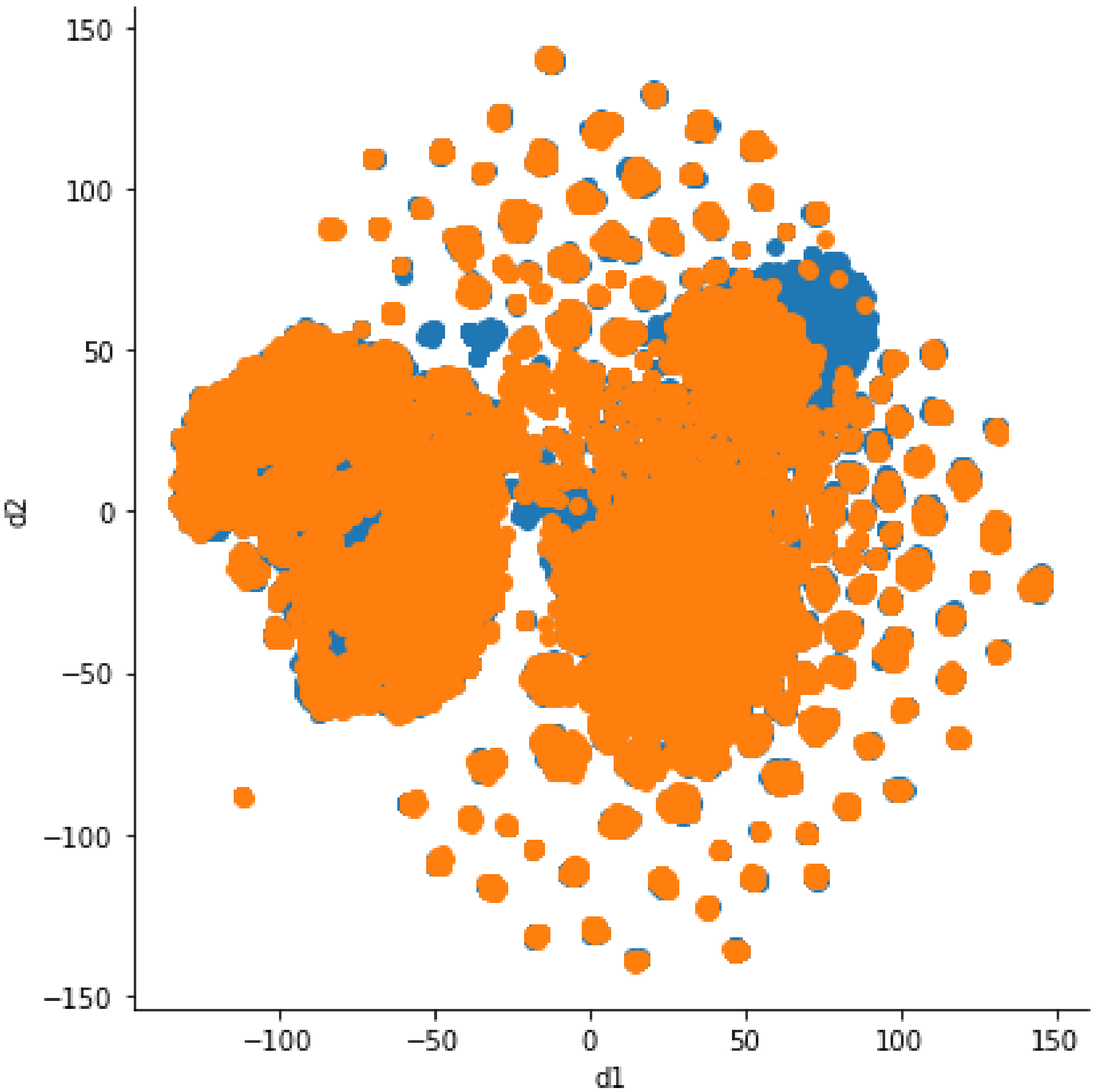
DATA VISUALIZATION : T-SNE

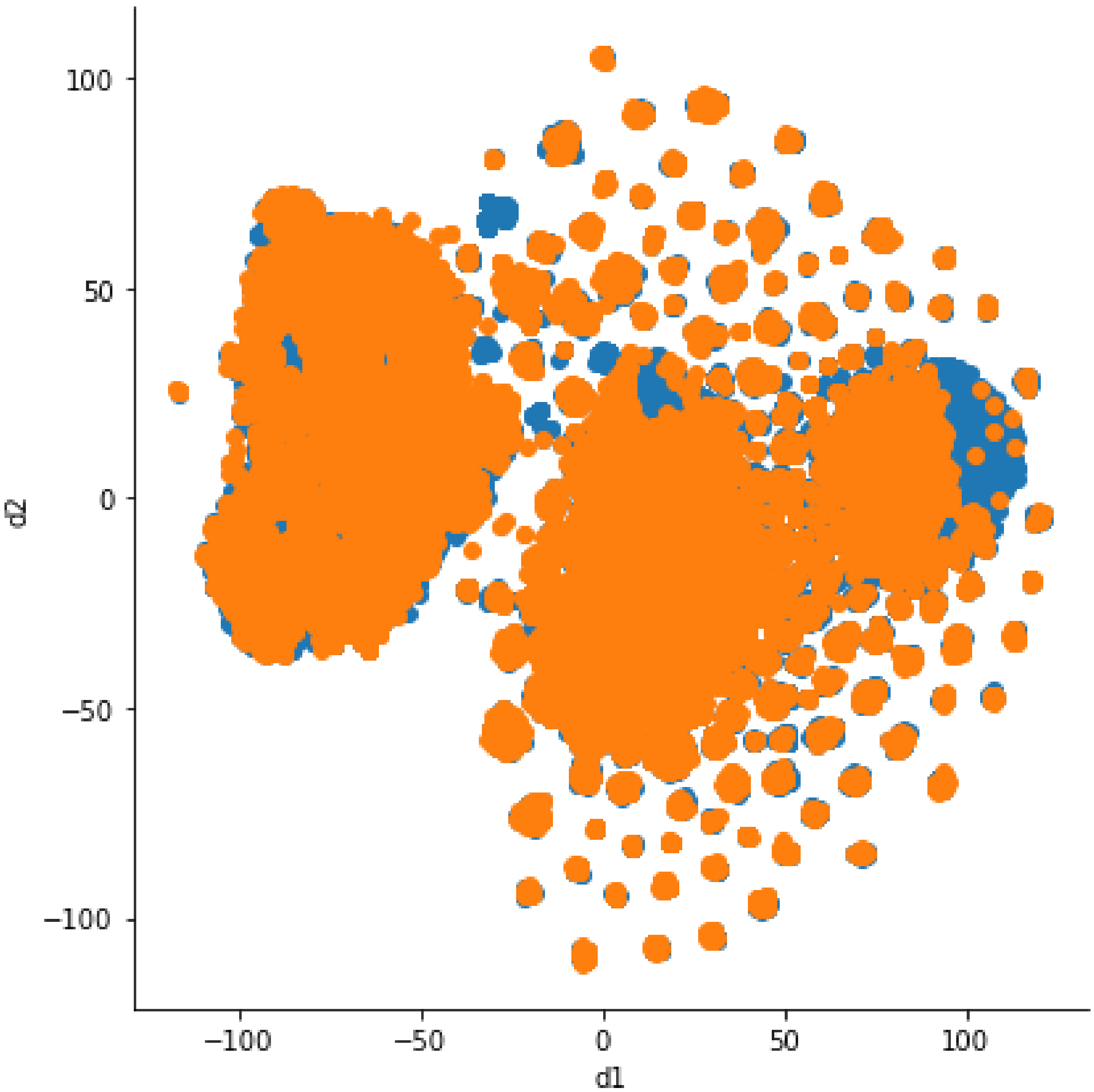
[284]:
```
perplexity = [50,100,200]
```

[285]:
```
from sklearn.manifold import TSNE
import seaborn as sns
for p in perplexity:
    model = TSNE(n_components = 2, random_state = 0, n_iter = 2000, perplexity␣
 ↪= p, learning_rate = 'auto',init = 'random')
    tsne_data = model.fit_transform(x_train)
    tsne_data = np.hstack((tsne_data,y_train.to_numpy().reshape(-1,1)))
    tsne_df = pd.DataFrame(data = tsne_data, columns =('d1','d2','label'))
    sns.FacetGrid(tsne_df,hue = 'label', height = 6).map(plt.scatter, 'd1','d2')
    plt.show()
```



150
100
50
0
−50
−100
−150
d2
−150
−100
−50
0
50
100
150
d1


150
100
50
0
−50
−100
−150
d2
−100
−50
0
50
100
150
d1



MACHINE LEARNING MODELS

RANDOM FOREST CLASSIFIER

[79]:
```
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import roc_auc_score

clf = RandomForestClassifier(n_estimators = 1000, max_depth=10, random_state=0,␣
 ↪n_jobs = -1)
clf.fit(x_train, y_train)
```

[79]:
```
RandomForestClassifier(max_depth=10, n_estimators=1000, n_jobs=-1,
                       random_state=0)
```

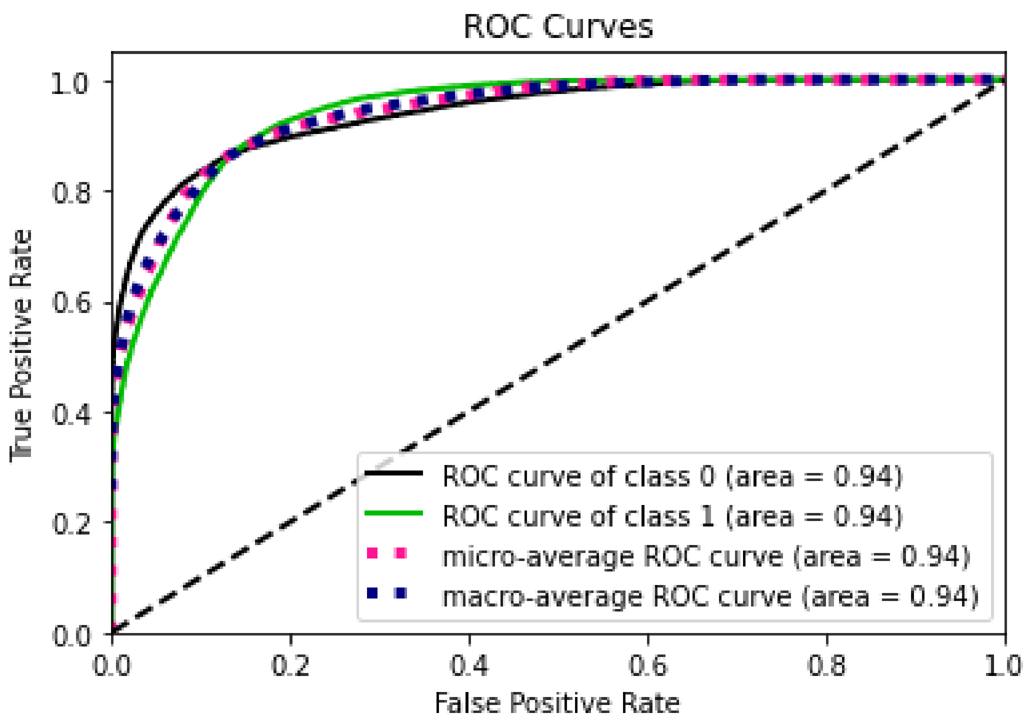
[80]:
```
y_pred_proba = clf.predict_proba(x_train)
y_pred_proba_cv  = clf.predict_proba(x_cv)
auc = roc_auc_score(y_train, y_pred_proba[:,1])
auc_cv = roc_auc_score(y_cv, y_pred_proba_cv[:,1])
```

ROC AUC SCORE

[81]:
```
print('Train AUC = {auc}, CV AUC = {auc_cv}'.format(auc = auc,auc_cv = auc_cv))
```

```
Train AUC = 0.9463742742375456, CV AUC = 0.9415541105551879
```

[82]:
```
skplt.metrics.plot_roc(y_cv, y_pred_proba_cv)
plt.show()
```



TRAIN & CV LOG LOSS

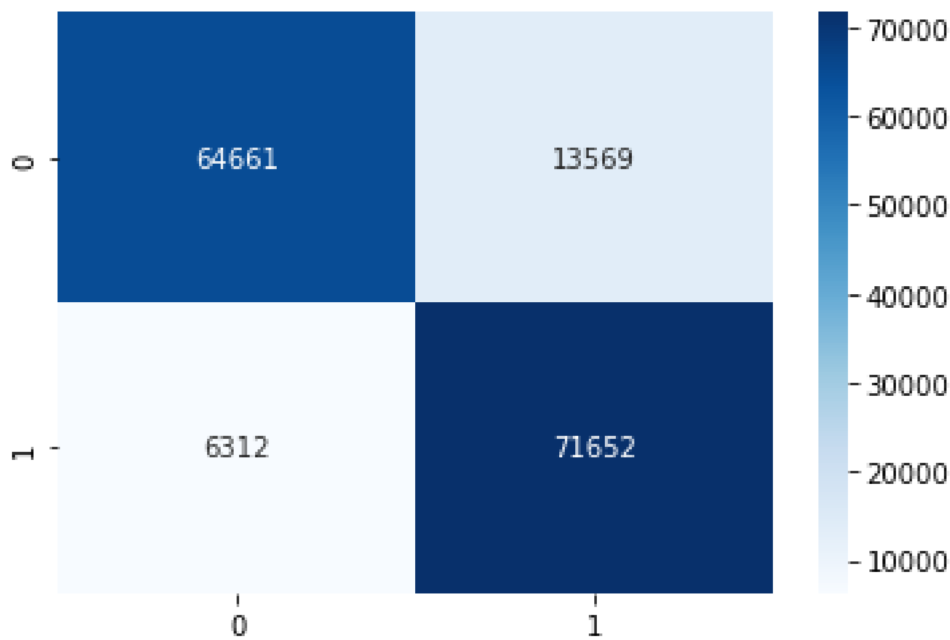
[83]:
```
from sklearn.metrics import log_loss
print('Train Log Loss = {}, CV Log Loss = {}'.
 ↪format(log_loss(y_train,y_pred_proba), log_loss(y_cv,y_pred_proba_cv)))
```

```
Train Log Loss = 0.34767920499841987, CV Log Loss = 0.35524322553748683
```

CONFUSION MATRIX

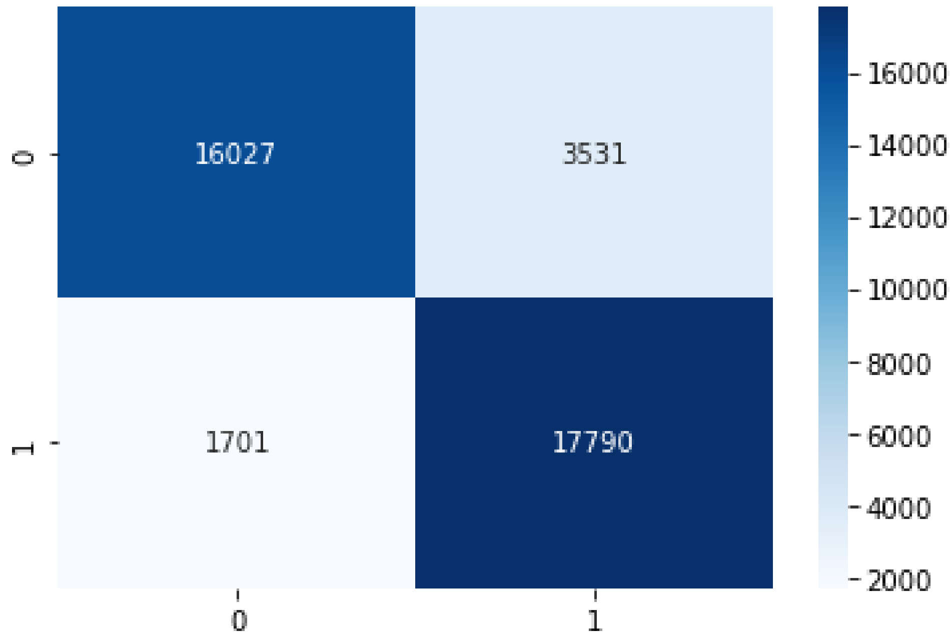
[84]:
```
from sklearn.metrics import confusion_matrix
sns.heatmap(confusion_matrix(y_train, np.argmax(y_pred_proba, axis = -1)),␣
↪annot=True,cmap='Blues', fmt='g')
```

[84]: <AxesSubplot:>



[85]:
```
sns.heatmap(confusion_matrix(y_cv, np.argmax(y_pred_proba_cv, axis = -1)),␣
↪annot=True,cmap='Blues', fmt='g')
```

[85]: <AxesSubplot:>



F1_SCORE

```
[86]: from sklearn.metrics import f1_score
train_f1_score = f1_score(y_train, np.argmax(y_pred_proba, axis = -1))
cv_f1_score = f1_score(y_cv, np.argmax(y_pred_proba_cv, axis = -1))
print('Train F1_Score = {}, CV F1_Score = {}'.
 ↪format(train_f1_score,cv_f1_score))
```

```
Train F1_Score = 0.8781689493519624, CV F1_Score = 0.8718024110555719
```

XGBOOST CLASSIFIER

```
[87]: from xgboost import XGBClassifier
from sklearn.calibration import CalibratedClassifierCV
x_cfl=XGBClassifier(n_estimators=1500, max_depth = 3, learning_rate = 0.03,␣
 ↪colsample_bytree = 0.5, subsample = 1)
x_cfl.fit(x_train,y_train)
c_cfl=CalibratedClassifierCV(x_cfl,method='sigmoid')
c_cfl.fit(x_train,y_train)
```

```
[87]: CalibratedClassifierCV(base_estimator=XGBClassifier(base_score=0.5,
                                                    booster='gbtree',
                                                    callbacks=None,
                                                    colsample_bylevel=1,
                                                    colsample_bynode=1,
                                                    colsample_bytree=0.5,
                                                    early_stopping_rounds=None,
```

```
                                          enable_categorical=False,
                                          eval_metric=None, gamma=0,
                                          gpu_id=-1,
                                          grow_policy='depthwise',
                                          importance_type=None,
                                          interaction_constraints='',
                                          learning_rate=0.03,
                                          max_bin=256,
                                          max_cat_to_onehot=4,
                                          max_delta_step=0,
                                          max_depth=3, max_leaves=0,
                                          min_child_weight=1,
                                          missing=nan,
                                          monotone_constraints='()',
                                          n_estimators=1500, n_jobs=0,
                                          num_parallel_tree=1,
                                          predictor='auto',
                                          random_state=0, reg_alpha=0,
                                          reg_lambda=1, …))
```

[88]:

```
y_pred_proba = c_cfl.predict_proba(x_train)
y_pred_proba_cv  = c_cfl.predict_proba(x_cv)
auc = roc_auc_score(y_train, y_pred_proba[:,1])
auc_cv = roc_auc_score(y_cv, y_pred_proba_cv[:,1])
```

ROC AUC SCORE

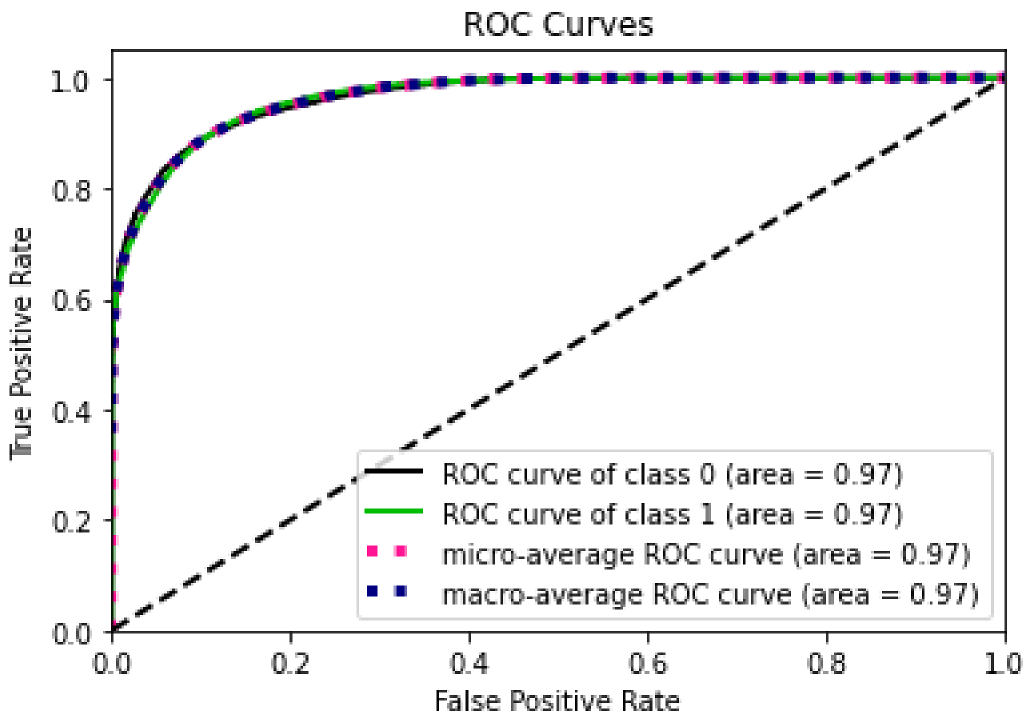
[89]:

```
print('Train AUC = {auc}, CV AUC = {auc_cv}'.format(auc = auc,auc_cv = auc_cv))
```

```
Train AUC = 0.9688238956562764, CV AUC = 0.9658205919336132
```

[90]:

```
skplt.metrics.plot_roc(y_cv, y_pred_proba_cv)
plt.show()
```



TRAIN & CV LOG LOSS

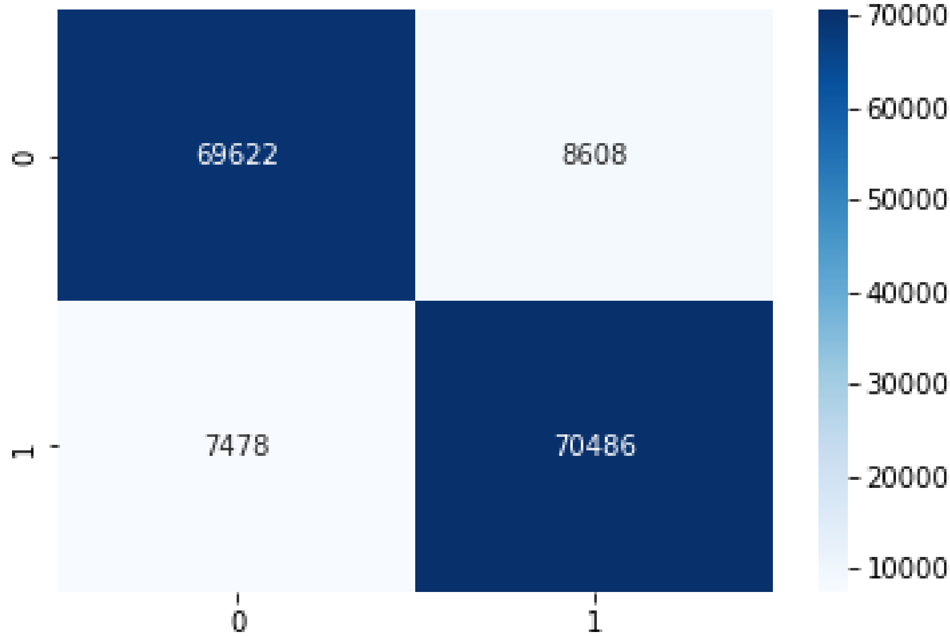
[91]:
```
from sklearn.metrics import log_loss
print('Train Log Loss = {}, CV Log Loss = {}'.
 ↪format(log_loss(y_train,y_pred_proba), log_loss(y_cv,y_pred_proba_cv)))
```

```
Train Log Loss = 0.23931956411446156, CV Log Loss = 0.25061700294124206
```

CONFUSION MATRIX

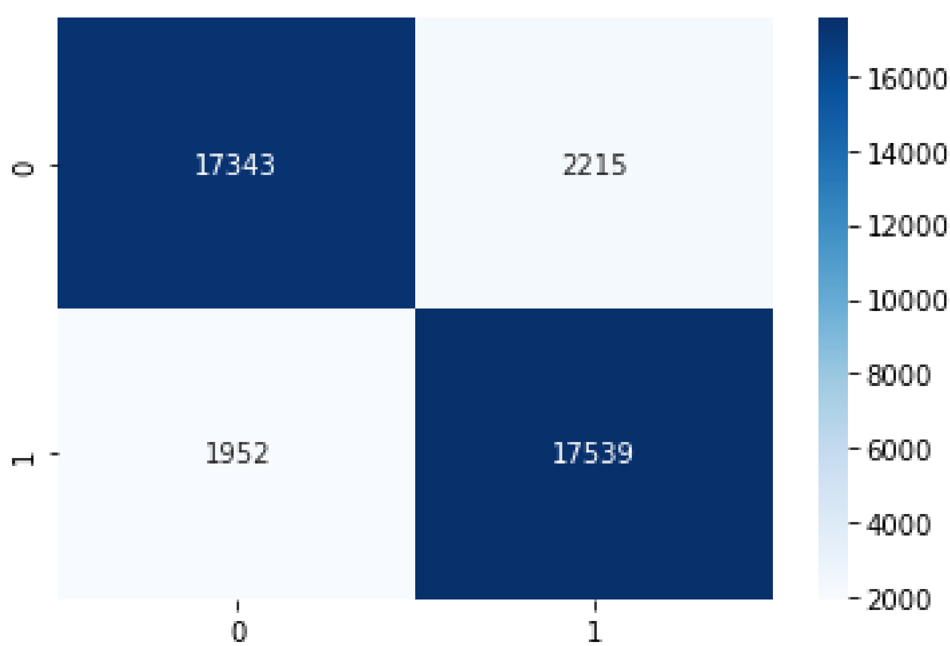
[92]:
```
from sklearn.metrics import confusion_matrix
sns.heatmap(confusion_matrix(y_train, np.argmax(y_pred_proba, axis = -1)),␣
 ↪annot=True,cmap='Blues', fmt='g')
```

[92]:
```
<AxesSubplot:>
```



[93]:
```
sns.heatmap(confusion_matrix(y_cv, np.argmax(y_pred_proba_cv, axis = -1)),␣
↪annot=True,cmap='Blues', fmt='g')
```

[93]: <AxesSubplot:>



F1_SCORE

[94]:
```
from sklearn.metrics import f1_score
train_f1_score = f1_score(y_train, np.argmax(y_pred_proba, axis = -1))
cv_f1_score = f1_score(y_cv, np.argmax(y_pred_proba_cv, axis = -1))
print('Train F1_Score = {}, CV F1_Score = {}'.
 ↪format(train_f1_score,cv_f1_score))
```

```
Train F1_Score = 0.8975792382431969, CV F1_Score = 0.893820868900497
```

INTRODUCING FEATURE INTERACTIONS

[95]:
```
import pandas as pd
from sklearn.ensemble import RandomForestClassifier
from sklearn.feature_selection import SelectFromModel

def imp_features(data, keep, labels):
    '''
    Collect important features using Random Forest Classifier
    '''
    rf = RandomForestClassifier(n_estimators = 1000,max_depth = 10, n_jobs = -1)
    rf.fit(data, labels)
    imp_feature_indx = np.argsort(rf.feature_importances_)[::-1]
    return imp_feature_indx[:keep]
```

[96]:
```
indexes = imp_features(x_train, 50, y_train)
```

[97]:
```
x_train[:,indexes].shape
```

[97]:
```
(156194, 50)
```

[98]:
```
def introduce_feature_interactions(data, columns):
    """This function randomly picks three features from the list of features␣
 ↪provided to it
    and performs various feature interactions on them to generate a set of new␣
 ↪features"""
    new_features = []

    for features in columns:

        feature_1 = data[:,features[0]]
        feature_2 = data[:,features[1]]
        feature_3 = data[:,features[2]]

        new_features.append(feature_1 * feature_2)
        new_features.append(feature_1 + feature_2)
        new_features.append(feature_1 - feature_2)
        new_features.append(feature_1 * feature_2 + feature_3)
```

```
        new_features.append(feature_1 + feature_2 - feature_3)
        new_features.append(feature_1 - feature_2 * feature_3)


    new_features = np.array(new_features)
    return new_features.T
```

[99]:
```
columns = []
for i in range(0,100):
    columns.append(np.random.choice(indexes, size=3, replace=False))
```

[100]:
```
new_features =  introduce_feature_interactions(x_train,columns)
new_features_cv = introduce_feature_interactions(x_cv,columns)
new_features.shape, new_features_cv.shape
```

[100]: ((156194, 600), (39049, 600))

[101]:
```
x_train = np.hstack((x_train,new_features))
x_cv = np.hstack((x_cv,new_features_cv))
x_train.shape, x_cv.shape
```

[101]: ((156194, 1203), (39049, 1203))

[102]:
```
f = open('feature_interactions.pkl','wb')
pickle.dump(columns,f)
f.close()
```

DROPPING HIGHLY CORRELATED FEATURES

[103]:
```
dataset = pd.DataFrame(x_train)
```

[104]:
```
def correlation(dataset, threshold):
    column = set()
    cd = dataset.corr()
    for i in range(len(cd.columns)):
        for j in range(i):
            if abs(cd.iloc[i,j]) >threshold:
                col_name = cd.columns[i]
                column.add(col_name)
    return column
```

[105]:
```
correlated_features = list(correlation(dataset,0.99))
```

[106]:
```
final_features = [feature for feature in dataset.columns if feature not in␣
 ↪correlated_features]
```

[107]:
```
x_train = x_train[:,final_features]
x_cv = x_cv[:,final_features]
x_train.shape , x_cv.shape
```

[107]: ((156194, 1146), (39049, 1146))

[108]:
```
f = open('final_features.pkl','wb')
pickle.dump(final_features,f)
f.close()
```

RANDOM FOREST CLASSIFIER WITH FEATURE INTERACTIONS

[109]:
```
from sklearn.ensemble import RandomForestClassifier
clf = RandomForestClassifier(n_estimators = 1000, max_depth=10, random_state=0,␣
↪n_jobs = -1)
clf.fit(x_train, y_train)
```

[109]:
```
RandomForestClassifier(max_depth=10, n_estimators=1000, n_jobs=-1,
                       random_state=0)
```

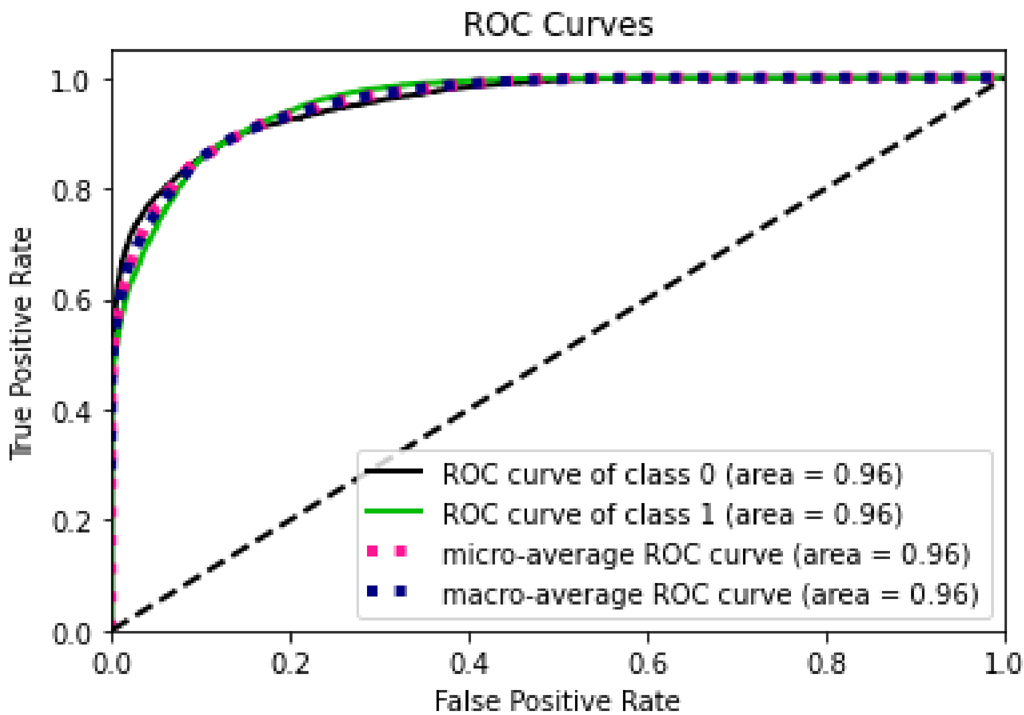
[110]:
```
y_pred_proba = clf.predict_proba(x_train)
y_pred_proba_cv  = clf.predict_proba(x_cv)
auc = roc_auc_score(y_train, y_pred_proba[:,1])
auc_cv = roc_auc_score(y_cv, y_pred_proba_cv[:,1])
```

ROC AUC SCORE

[111]:
```
print('Train AUC = {auc}, CV AUC = {auc_cv}'.format(auc = auc,auc_cv = auc_cv))
```

```
Train AUC = 0.9601581853958523, CV AUC = 0.9563737386451443
```

[112]:
```
skplt.metrics.plot_roc(y_cv, y_pred_proba_cv)
plt.show()
```



TRAIN & CV LOG LOSS

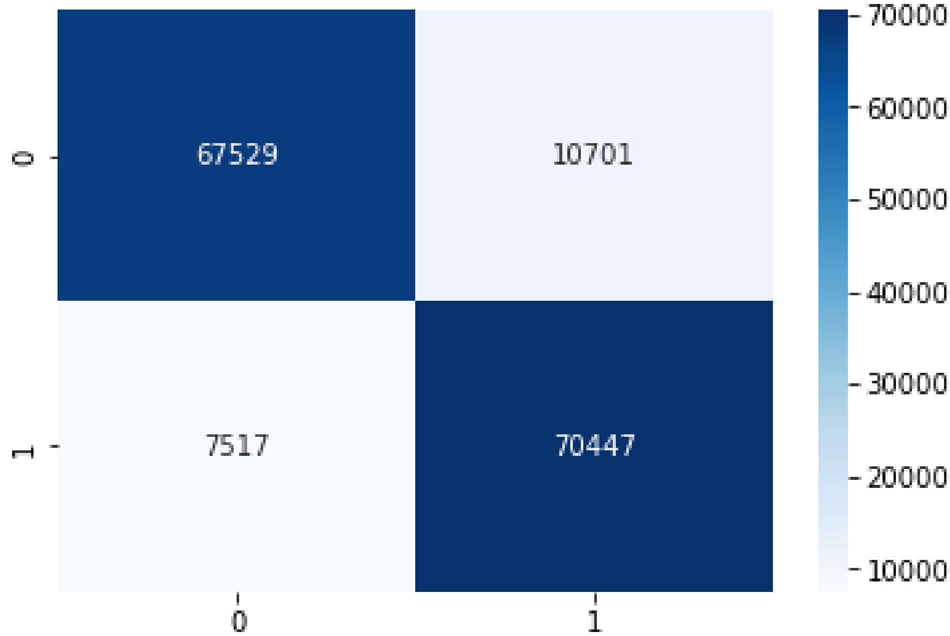
[113]:
```
from sklearn.metrics import log_loss
print('Train Log Loss = {}, CV Log Loss = {}'.
 ↪format(log_loss(y_train,y_pred_proba), log_loss(y_cv,y_pred_proba_cv)))
```

```
Train Log Loss = 0.2593797983885249, CV Log Loss = 0.268504114714604
```

CONFUSION MATRIX

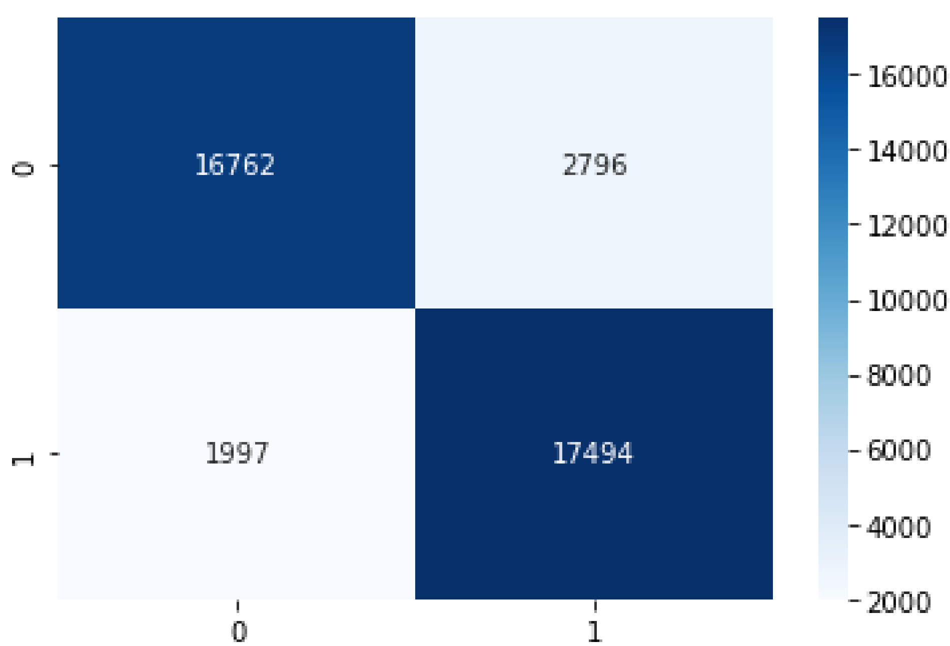
[114]:
```
from sklearn.metrics import confusion_matrix
sns.heatmap(confusion_matrix(y_train, np.argmax(y_pred_proba, axis = -1)),␣
 ↪annot=True,cmap='Blues', fmt='g')
```

[114]: <AxesSubplot:>



[115]:
```
sns.heatmap(confusion_matrix(y_cv, np.argmax(y_pred_proba_cv, axis = -1)),␣
 ↪annot=True,cmap='Blues', fmt='g')
```

[115]: <AxesSubplot:>



F1_SCORE

[116]:
```
from sklearn.metrics import f1_score
train_f1_score = f1_score(y_train, np.argmax(y_pred_proba, axis = -1))
cv_f1_score = f1_score(y_cv, np.argmax(y_pred_proba_cv, axis = -1))
print('Train F1_Score = {}, CV F1_Score = {}'.
 ↪format(train_f1_score,cv_f1_score))
```

```
Train F1_Score = 0.8855020363014732, CV F1_Score = 0.8795153465222091
```

XGBOOST CLASSIFIER WITH FEATURE INTERACTIONS

HYPERPARAMETER TUNING

[117]:
```
from sklearn.model_selection import RandomizedSearchCV
x_cfl=XGBClassifier()
prams={
    'learning_rate':[0.01,0.03,0.05,0.1,0.15,0.2],
     'n_estimators':[200,500,1500],
     'max_depth':[3,5],
    'colsample_bytree':[0.3,0.5,1],
    'subsample':[0.1,0.3,0.5,1]
}
random_cfl1=RandomizedSearchCV(x_cfl,param_distributions=prams, n_iter = 4)
random_cfl1.fit(x_train,y_train)
print (random_cfl1.best_params_)
```

```
{'subsample': 0.5, 'n_estimators': 1500, 'max_depth': 5, 'learning_rate': 0.05,
'colsample_bytree': 0.3}
```

MODEL WITH BEST HYPERPARAMETERS

[127]:
```
from xgboost import XGBClassifier
from sklearn.calibration import CalibratedClassifierCV
x_cfl=XGBClassifier(n_estimators=1500, max_depth = 5, learning_rate = 0.05,␣
 ↪colsample_bytree = 0.3, subsample = 0.5)
x_cfl.fit(x_train,y_train)
c_cfl=CalibratedClassifierCV(x_cfl,method='sigmoid')
c_cfl.fit(x_train,y_train)
```

[127]:
```
CalibratedClassifierCV(base_estimator=XGBClassifier(base_score=0.5,
                                                    booster='gbtree',
                                                    callbacks=None,
                                                    colsample_bylevel=1,
                                                    colsample_bynode=1,
                                                    colsample_bytree=0.3,
                                                    early_stopping_rounds=None,
                                                    enable_categorical=False,
                                                    eval_metric=None, gamma=0,
                                                    gpu_id=-1,
```

```
                                                  grow_policy='depthwise',
                                                  importance_type=None,
                                                  interaction_constraints='',
                                                  learning_rate=0.05,
                                                  max_bin=256,
                                                  max_cat_to_onehot=4,
                                                  max_delta_step=0,
                                                  max_depth=5, max_leaves=0,
                                                  min_child_weight=1,
                                                  missing=nan,
                                                  monotone_constraints='()',
                                                  n_estimators=1500, n_jobs=0,
                                                  num_parallel_tree=1,
                                                  predictor='auto',
                                                  random_state=0, reg_alpha=0,
                                                  reg_lambda=1, …))
```

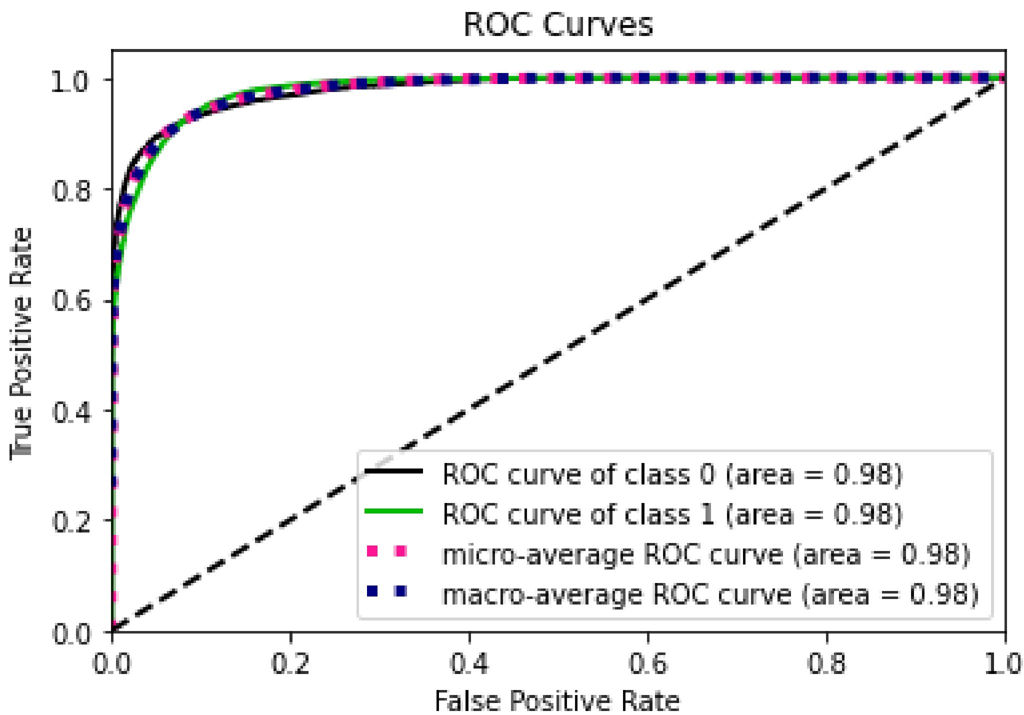
[128]:
```
y_pred_proba = c_cfl.predict_proba(x_train)
y_pred_proba_cv  = c_cfl.predict_proba(x_cv)
auc = roc_auc_score(y_train, y_pred_proba[:,1])
auc_cv = roc_auc_score(y_cv, y_pred_proba_cv[:,1])
```

ROC AUC SCORE

[129]:
```
print('Train AUC = {auc}, CV AUC = {auc_cv}'.format(auc = auc,auc_cv = auc_cv))
```

```
Train AUC = 0.9936127243373905, CV AUC = 0.9790056519146504
```

[130]:
```
skplt.metrics.plot_roc(y_cv, y_pred_proba_cv)
plt.show()
```



TRAIN & CV LOG LOSS

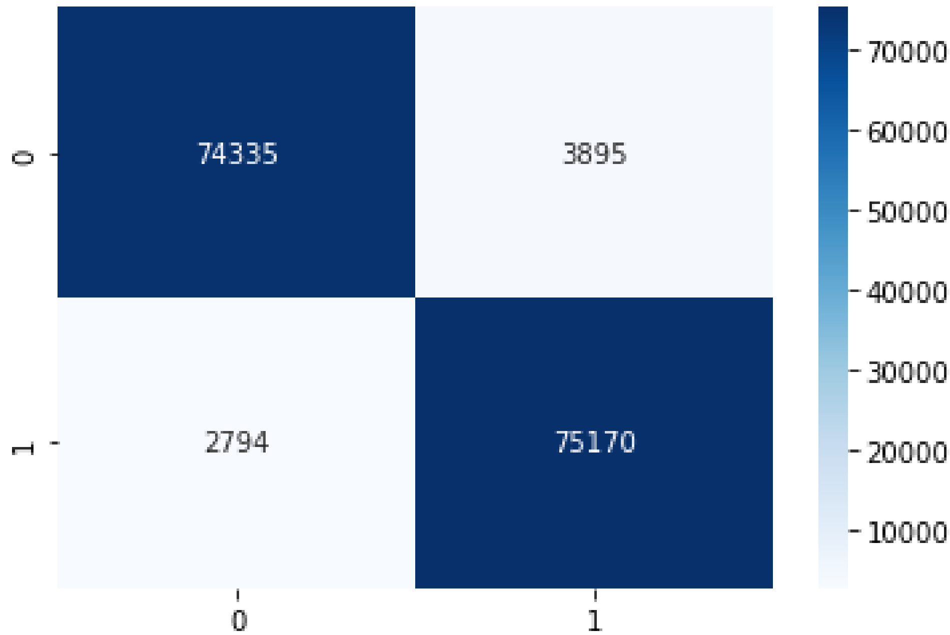
[131]:
```
from sklearn.metrics import log_loss
print('Train Log Loss = {}, CV Log Loss = {}'.
 ↪format(log_loss(y_train,y_pred_proba), log_loss(y_cv,y_pred_proba_cv)))
```

```
Train Log Loss = 0.12016930618142911, CV Log Loss = 0.19465866768431764
```

CONFUSION MATRIX

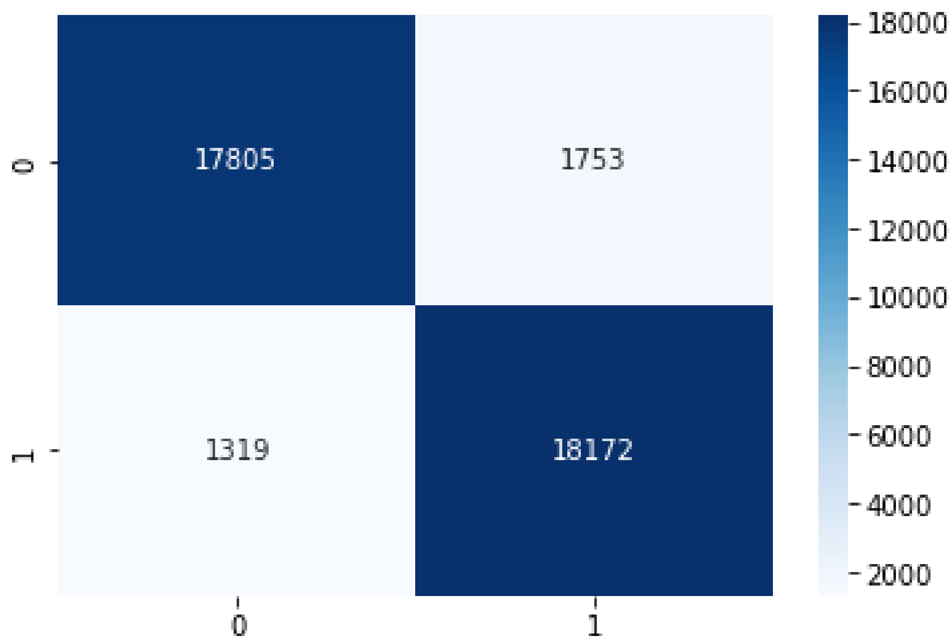
[132]:
```
from sklearn.metrics import confusion_matrix
sns.heatmap(confusion_matrix(y_train, np.argmax(y_pred_proba, axis = -1)),␣
 ↪annot=True,cmap='Blues', fmt='g')
```

[132]:
```
<AxesSubplot:>
```



[133]:
```
sns.heatmap(confusion_matrix(y_cv, np.argmax(y_pred_proba_cv, axis = -1)),
    annot=True,cmap='Blues', fmt='g')
```

[133]: <AxesSubplot:>



F1_SCORE

[134]:
```
from sklearn.metrics import f1_score
train_f1_score = f1_score(y_train, np.argmax(y_pred_proba, axis = -1))
cv_f1_score = f1_score(y_cv, np.argmax(y_pred_proba_cv, axis = -1))
print('Train F1_Score = {}, CV F1_Score = {}'.
 ↪format(train_f1_score,cv_f1_score))
```

```
Train F1_Score = 0.9574027727362461, CV F1_Score = 0.9220621067586767
```

SAVING BEST MODEL

[136]:
```
f = open('best_model_xgboost.pkl','wb')
pickle.dump(x_cfl,f)
f.close()
```

[137]:
```
f = open('best_model_cc.pkl','wb')
pickle.dump(c_cfl,f)
f.close()
```