Le marketing rencontre l'apprentissage automatique : perspectives d'optimisation des campagnes

ABDOULAYE TANGARA

Student in MSc in Quantitative and Computable Economics

Contacts

LinkedIn

Mon GitHub

Mon Portfolio

abdoulayetangara722@gmail.com

"Si tu veux avoir quelque chose que tu n'as jamais eu, tu fais quelque chose que tu n'as jamais fait"

0.1 Analyse des KPIs et Modélisation de Campagne Marketing

[ ]:
```python
import pandas as pd # Manipulate dataframes
import numpy as np # Operations numériques
from skimpy import skim # Visualisation des données
from tabulate import tabulate # Visualisation des données
from datetime import datetime # Manipulation des dates
import scipy.stats as stat # Test statistiques
import matplotlib.pyplot as plt # Visualisation des données
import seaborn as sns # Visualisation des données
from sklearn.cluster import KMeans # Pour la segmentation
from sklearn.ensemble import RandomForestClassifier, IsolationForest # Modèle␣
 ↪de classification # Detection des outliers
from sklearn.metrics import accuracy_score, classification_report,␣
 ↪confusion_matrix # Métriques de performance
from sklearn.model_selection import train_test_split ,GridSearchCV #␣
 ↪Séparation des données , # Optimisation des hyperparamètres
from sklearn.preprocessing import StandardScaler, OneHotEncoder #␣
 ↪Prétraitement des données
from sklearn.compose import ColumnTransformer # Prétraitement des données
from sklearn.pipeline import Pipeline # Prétraitement des données
import warnings # Gestion des avertissements
warnings.filterwarnings('ignore') # Ignorer les avertissements
```

Chargement et Préparation des Données

[88]:
```python
# Chargement des données
data = pd.read_csv("marketing_campaign.csv", delimiter="\t")
data
```

[88]:
```
         ID  Year_Birth   Education Marital_Status   Income  Kidhome  \
0      5524        1957  Graduation         Single  58138.0        0
1      2174        1954  Graduation         Single  46344.0        1
2      4141        1965  Graduation       Together  71613.0        0
3      6182        1984  Graduation       Together  26646.0        1
4      5324        1981         PhD        Married  58293.0        1
...     ...         ...         ...            ...      ...      ...
2235  10870        1967  Graduation        Married  61223.0        0
2236   4001        1946         PhD       Together  64014.0        2
2237   7270        1981  Graduation       Divorced  56981.0        0
2238   8235        1956      Master       Together  69245.0        0
2239   9405        1954         PhD        Married  52869.0        1

      Teenhome Dt_Customer  Recency  MntWines  ...  NumWebVisitsMonth  \
0            0  04-09-2012       58       635  ...                  7
1            1  08-03-2014       38        11  ...                  5
2            0  21-08-2013       26       426  ...                  4
3            0  10-02-2014       26        11  ...                  6
4            0  19-01-2014       94       173  ...                  5
...        ...         ...      ...       ...  ...                ...
2235         1  13-06-2013       46       709  ...                  5
```

```
2236          1  10-06-2014       56       406  ...                  7
2237          0  25-01-2014       91       908  ...                  6
2238          1  24-01-2014        8       428  ...                  3
2239          1  15-10-2012       40        84  ...                  7

      AcceptedCmp3  AcceptedCmp4  AcceptedCmp5  AcceptedCmp1  AcceptedCmp2  \
0                0             0             0             0             0
1                0             0             0             0             0
2                0             0             0             0             0
3                0             0             0             0             0
4                0             0             0             0             0
...            ...           ...           ...           ...           ...
2235             0             0             0             0             0
2236             0             0             0             1             0
2237             0             1             0             0             0
2238             0             0             0             0             0
2239             0             0             0             0             0

      Complain  Z_CostContact  Z_Revenue  Response
0            0              3         11         1
1            0              3         11         0
2            0              3         11         0
3            0              3         11         0
4            0              3         11         0
...        ...            ...        ...       ...
2235         0              3         11         0
2236         0              3         11         0
2237         0              3         11         0
2238         0              3         11         0
2239         0              3         11         1

[2240 rows x 29 columns]
```

[89]:
```python
# Preparation des données
# Information générales sur les données
print(f"Le nombre de lignes de la base de données est de : {data.shape[0]}")
print(f"\n Le nombre de colonnes de la base de données est de : {data.
↪shape[1]}\n")
print(f"La taille de la base de données est de : {data.size} \n")

info_data = pd.DataFrame({
    'variable' : data.columns.to_list(),
    'Type du variable' : data.dtypes.values,
    'NumMissing' : data.isnull().sum().values,
    '%Missing' : np.round((data.isnull().sum().values/data.shape[0])*100,3)
})

print(f"\n Les informations sur le dataframe: \n {info_data}")
```

```
Le nombre de lignes de la base de données est de : 2240

 Le nombre de colonnes de la base de données est de : 29

La taille de la base de données est de : 64960


 Les informations sur le dataframe:
                variable Type du variable  NumMissing  %Missing
0                    ID            int64           0     0.000
1            Year_Birth            int64           0     0.000
2             Education           object           0     0.000
3        Marital_Status           object           0     0.000
4                Income          float64          24     1.071
5               Kidhome            int64           0     0.000
6              Teenhome            int64           0     0.000
7           Dt_Customer           object           0     0.000
8               Recency            int64           0     0.000
9              MntWines            int64           0     0.000
10            MntFruits            int64           0     0.000
11      MntMeatProducts            int64           0     0.000
12      MntFishProducts            int64           0     0.000
13     MntSweetProducts            int64           0     0.000
14         MntGoldProds            int64           0     0.000
15    NumDealsPurchases            int64           0     0.000
16      NumWebPurchases            int64           0     0.000
17  NumCatalogPurchases            int64           0     0.000
18    NumStorePurchases            int64           0     0.000
19    NumWebVisitsMonth            int64           0     0.000
20         AcceptedCmp3            int64           0     0.000
21         AcceptedCmp4            int64           0     0.000
22         AcceptedCmp5            int64           0     0.000
23         AcceptedCmp1            int64           0     0.000
24         AcceptedCmp2            int64           0     0.000
25             Complain            int64           0     0.000
```

```
26        Z_CostContact            int64          0      0.000
27            Z_Revenue            int64          0      0.000
28             Response            int64          0      0.000
```

[90]:

```
# Traitement de certaines variables
data["Dt_Customer"] = pd.to_datetime(data["Dt_Customer"],format= "%d-%m-%Y")
data.drop(columns=["ID","Z_Revenue","Z_CostContact"], inplace=True)

col_cat = ['AcceptedCmp3', 'AcceptedCmp4', 'AcceptedCmp5', 'AcceptedCmp1',
       'AcceptedCmp2', 'Complain', 'Response']
for i in col_cat:
    data[i] = data[i].astype("category")

if "Year_Birth" in data.columns:
    data["Age"] = datetime.now().year - data["Year_Birth"]
    del data["Year_Birth"]

# Gestion des valeurs manquantes
data["Income"] = data["Income"].fillna(data["Income"].mean())
```

1.2 Analyse Statistique et KPIs des Campagnes

[91]:

```
skim(data)

print('Description des variables qualitatives \n')
data.describe(include=["object","category"]).T

# Selection des variables
var_quanti = data.select_dtypes(include=["int64","float64"]).columns
var_quali = data.select_dtypes(include=["object","category"]).columns

for val in var_quali:
    print(f"La variable {val} contient les modalités suivantes : \n ␣
 ↪{data[val].unique()} \n")
```

```
Description des variables qualitatives

La variable Education contient les modalités suivantes :
  ['Graduation' 'PhD' 'Master' 'Basic' '2n Cycle']

La variable Marital_Status contient les modalités suivantes :
  ['Single' 'Together' 'Married' 'Divorced' 'Widow' 'Alone' 'Absurd' 'YOLO']

La variable AcceptedCmp3 contient les modalités suivantes :
  [0, 1]
Categories (2, int64): [0, 1]

La variable AcceptedCmp4 contient les modalités suivantes :
  [0, 1]
Categories (2, int64): [0, 1]

La variable AcceptedCmp5 contient les modalités suivantes :
```

```
  [0, 1]
Categories (2, int64): [0, 1]

La variable AcceptedCmp1 contient les modalités suivantes :
  [0, 1]
Categories (2, int64): [0, 1]

La variable AcceptedCmp2 contient les modalités suivantes :
  [0, 1]
Categories (2, int64): [0, 1]

La variable Complain contient les modalités suivantes :
  [0, 1]
Categories (2, int64): [0, 1]

La variable Response contient les modalités suivantes :
  [1, 0]
Categories (2, int64): [0, 1]
```

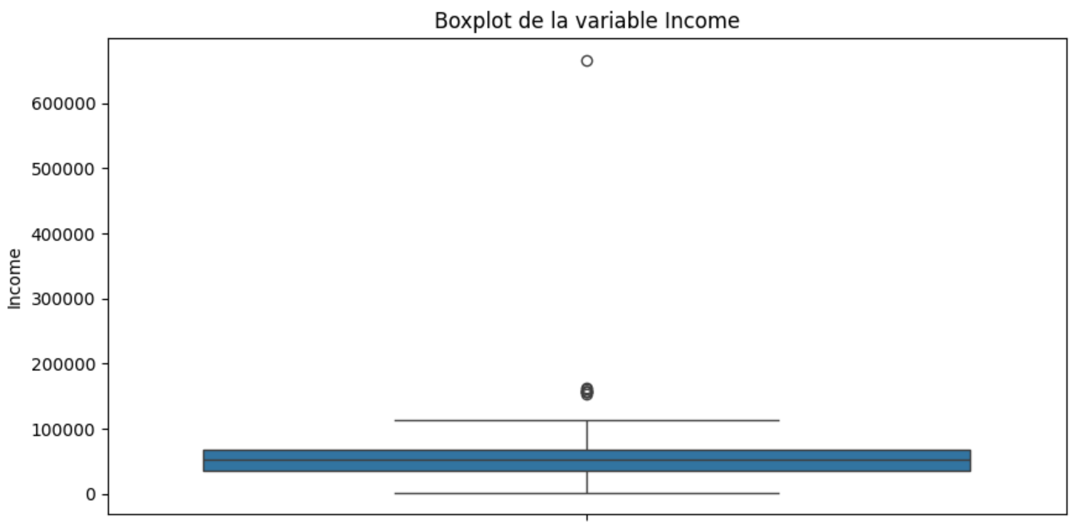
[92]:
```python
def UnivariateAnalysis(data, var_quanti, var_quali):

    # Les variables quantitatives
    # Boxplot
    for col in var_quanti:
        plt.figure(figsize=(10,5))
        sns.boxplot(data[col])

        plt.title(f"Boxplot de la variable {col}")
        plt.show()

    # Histogramme
    for var in var_quanti:
        if var in data.columns:
            pvalue, statistic = stat.shapiro(data[var]) # Hypothèse nulle␣
↪(H0) : Les données suivent une distribution normale.

            if pvalue > 0.05 :
                text0  = f"La variable {var} ne suit pas une loi normale avec␣
↪p-value = {np.round(pvalue,2)} et statistique = {np.round(statistic, 2)}"
            else:
                text0 = f"La variable {var} suit une loi normale avec p-value␣
↪= {np.round(pvalue,2)} et statistique = {np.round(statistic, 2)}"
            plt.figure(figsize=(10,5))
            sns.histplot(data[var], kde=True)
            plt.title(text0)
            plt.show()
        # Diagramme en barres

    # Les variables qualitatives
    for col in var_quali:

        # Informations
        print(f"\n la fréquance des modalités de la variable --{col}-- \n␣
↪{data[col].value_counts(normalize=True)*100}")

        # Visualisation
        plt.figure(figsize=(10,5))
        sns.countplot(data[col])
        plt.title(f"Analyse de la variable {col}")
        plt.show()
```

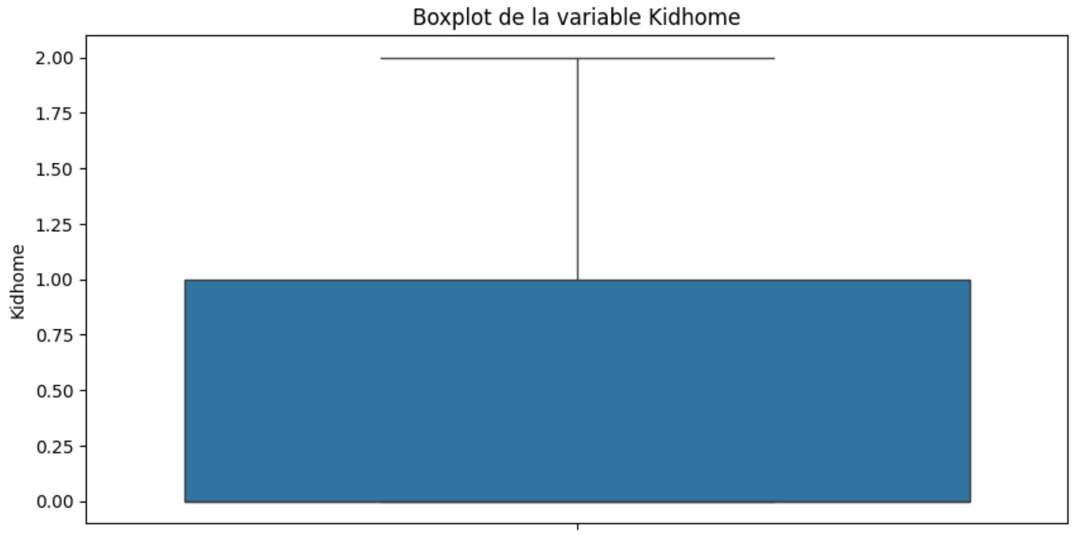
```
[93]: def AnalyseBivariee(data, var_quanti, var_quali):

    # Analyse de la relation entre les variables quantitatives et␣
↪quantitatives
    print(f"\n Analyse de la relation entre les variables quantitatives et␣
↪quantitatives \n")
    corr = data[var_quanti].corr()
    mask = np.triu(np.ones_like(corr, dtype=bool))
    plt.figure(figsize=(10,8))
    sns.heatmap(np.round(corr,1), annot=True, cmap="coolwarm", mask=mask,␣
↪vmax=1, vmin=-1, linewidths=0.5)
    plt.title("Matrice de corrélation des variables quantitatives")
    plt.show()

    plt.figure(figsize=(10,5))
    sns.pairplot(data[var_quanti])
    plt.show()

    # Analyse en série temporelle des variables
    ts_var = ['Income', 'MntWines', 'MntFruits','MntMeatProducts',␣
↪'MntFishProducts', 'MntSweetProducts','MntGoldProds', 'NumDealsPurchases',␣
↪'NumWebPurchases',
       'NumCatalogPurchases', 'NumStorePurchases', 'NumWebVisitsMonth']
    for i in ts_var:
        fig, ax = plt.subplots(figsize=(10,5))
        ax.scatter(x = data["Dt_Customer"], y = data[i])
        plt.title(f'Evoluation du {i}')
        plt.xlabel("Date")
        plt.ylabel(i)
        plt.show()


    # Analyse de la relation entre les variables qualitatives et qualitatives
    print(f"\Analyse de la relation entre les variables qualitatives et␣
↪qualitatives \n")
    for i in var_quali:
        for j in var_quali:
            if i != j and var_quali.get_loc(i) < var_quali.get_loc(j):
                table = pd.crosstab(data[i], data[j])
                print(f"Table de contingence entre les variables {i} et {j}␣
↪\n {table}")
                plt.figure(figsize=(10,5))
                sns.heatmap(table, annot=True, cmap="coolwarm", linewidths=1.
↪5, fmt="d")
                plt.title(f"Table de contingence entre les variables {i} et␣
↪{j}")
                plt.show()

    # Boucle pour analyser chaque paire (quantitative, qualitative)          ␣
↪
```

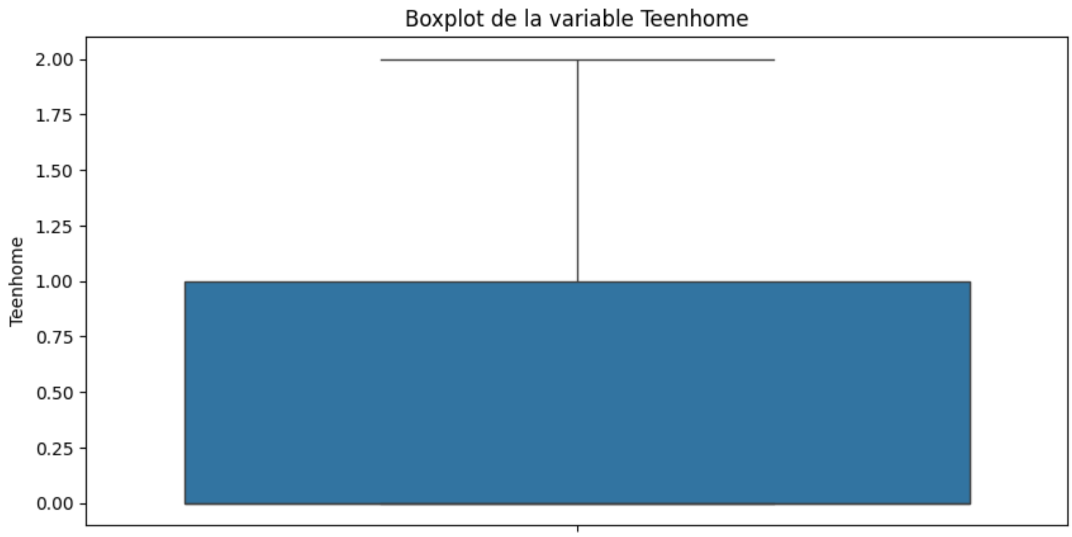
```
    for i in var_quanti:
        for j in var_quali:
            result = data.groupby(j)[i].mean()
            print(f"\n La moyenne de la variable {i} par rapport à la␣
↪variable {j} \n {np.round(result,3).sort_values(ascending=False)}")
            try:
                # Application du test ANOVA
                statistic, pvalue = stat.f_oneway(*[data[i][data[j] == level]␣
↪for level in data[j].unique()])

                # Déterminer la relation entre les variables
                if pvalue < 0.05:
                    text1 = f"La variable {i} et la variable {j} sont␣
↪dépendantes (p-value = {pvalue:.3f})"
                else:
                    text1 = f"La variable {i} et la variable {j} sont␣
↪indépendantes (p-value = {pvalue:.3f})"

                # Affichage des boîtes à moustaches
                plt.figure(figsize=(10, 5))
                sns.boxplot(x=data[j], y=data[i])
                plt.title(text1)
                plt.xticks(rotation=45)
                plt.show()

            except Exception as e:
                print(f"Erreur pour la combinaison {i} et {j} : {e}")
```

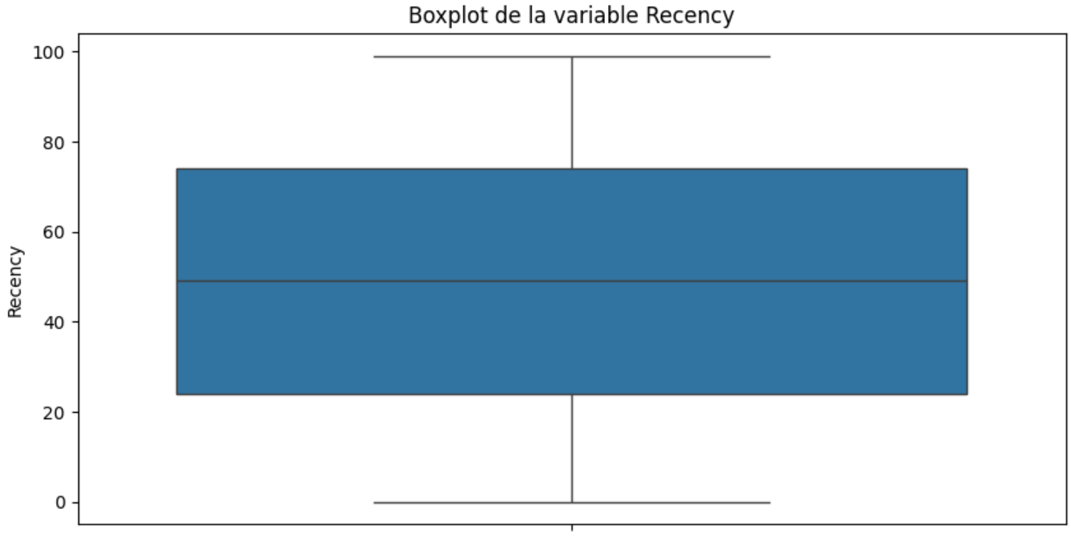
Analyse univariée

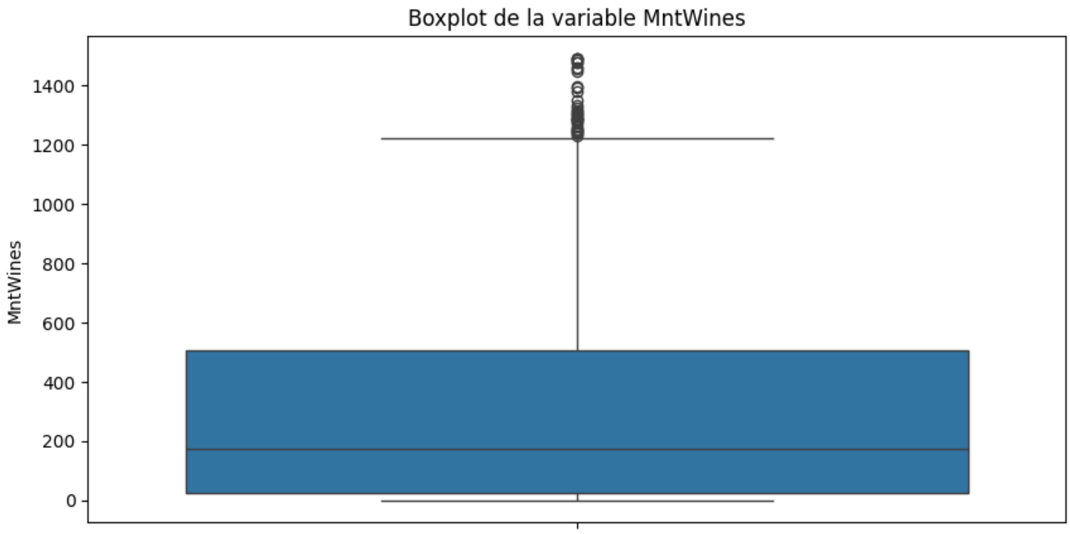
[94]:
```
# Analyse univariée
UnivariateAnalysis(data, var_quanti,var_quali)
```

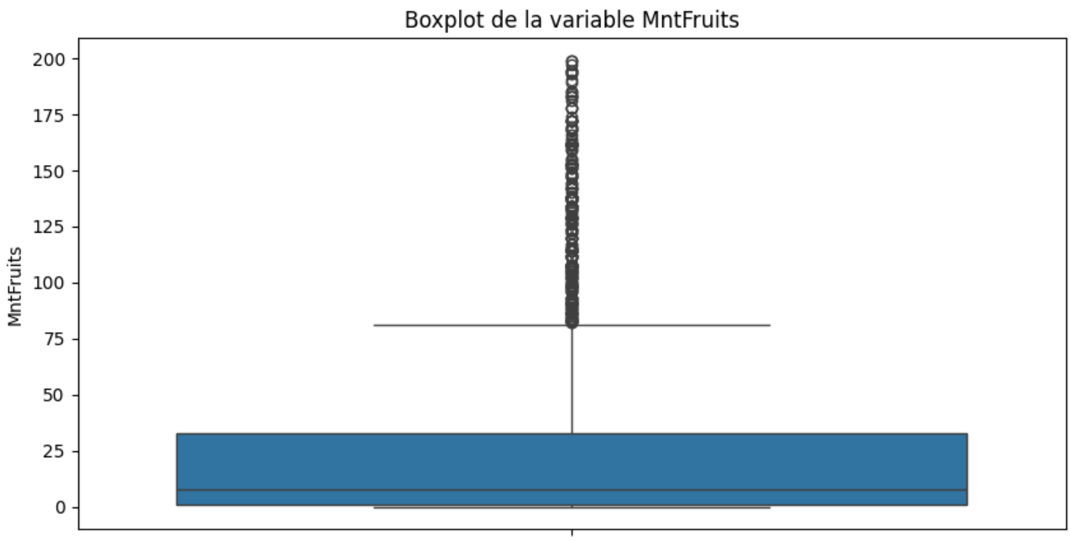


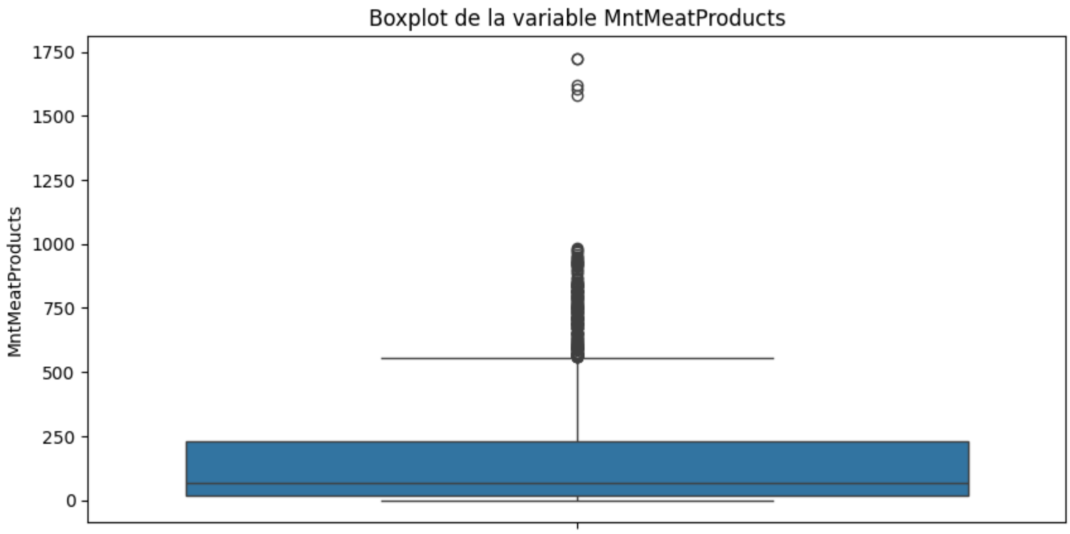

Boxplot de la variable Kidhome
Kidhome
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00

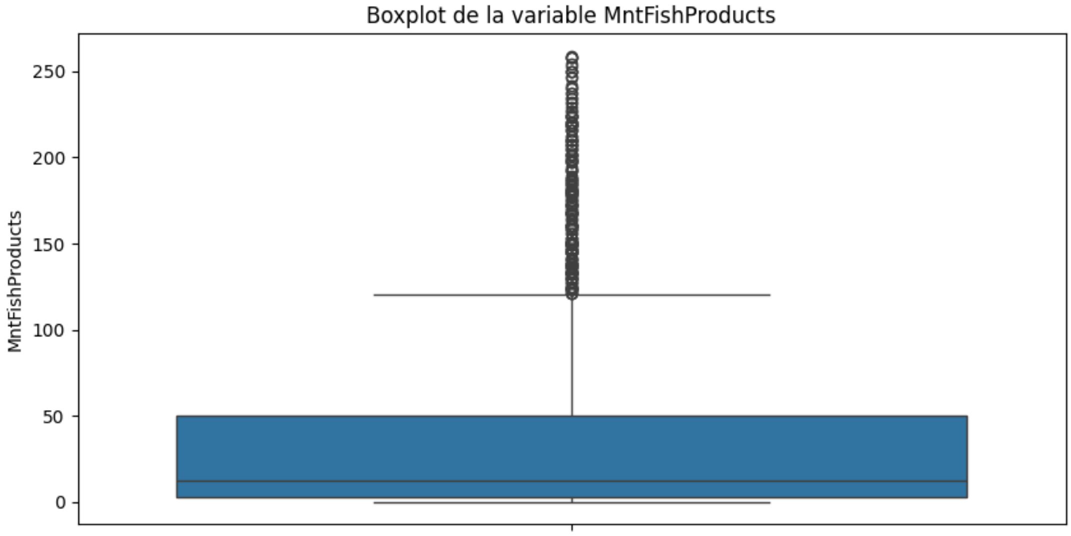

Boxplot de la variable Teenhome
Teenhome
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00

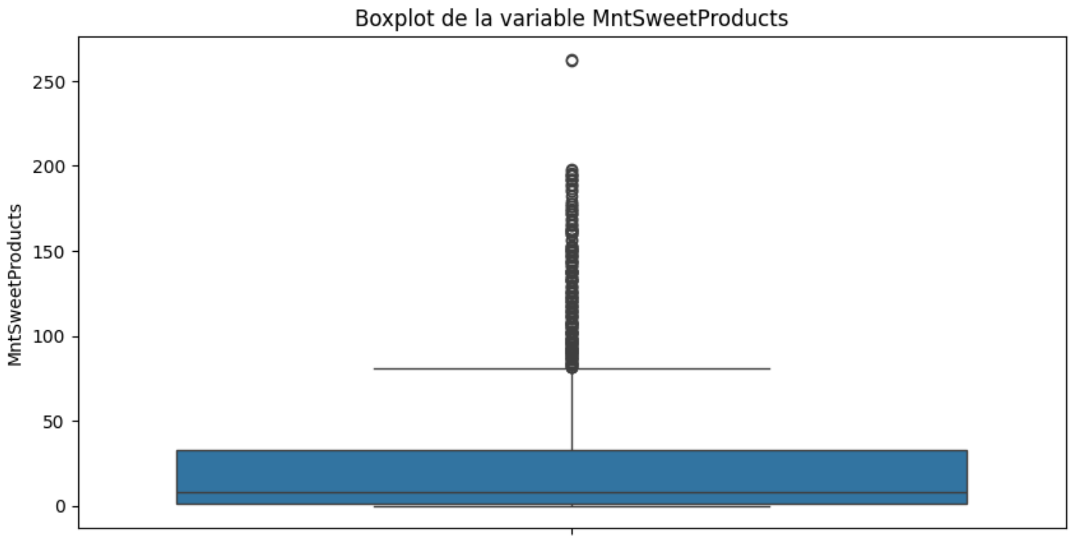

Boxplot de la variable Recency
Recency
100
80
60
40
20
0

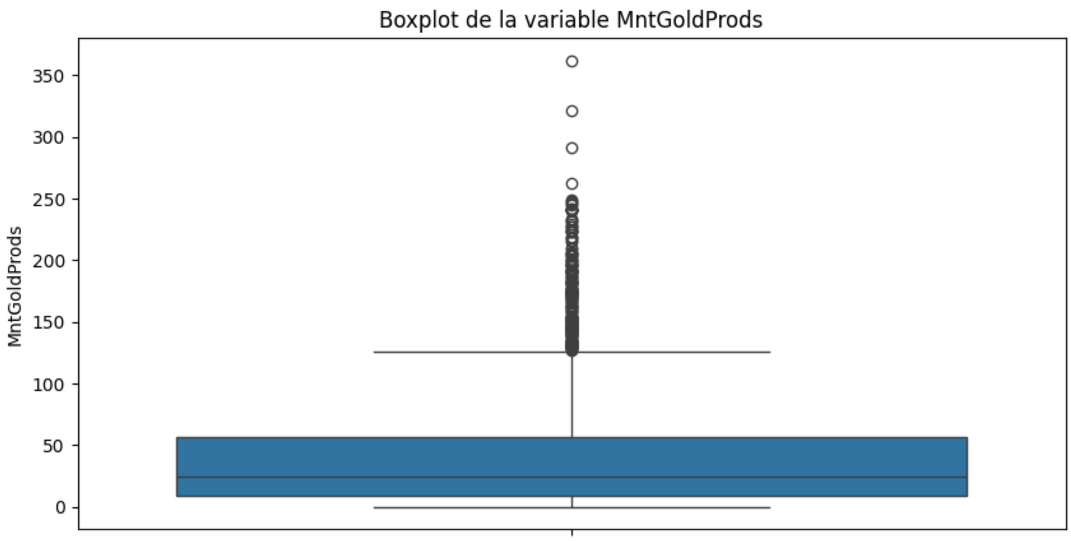

Boxplot de la variable MntWines
MntWines
1400
1200
1000
800
600
400
200
0

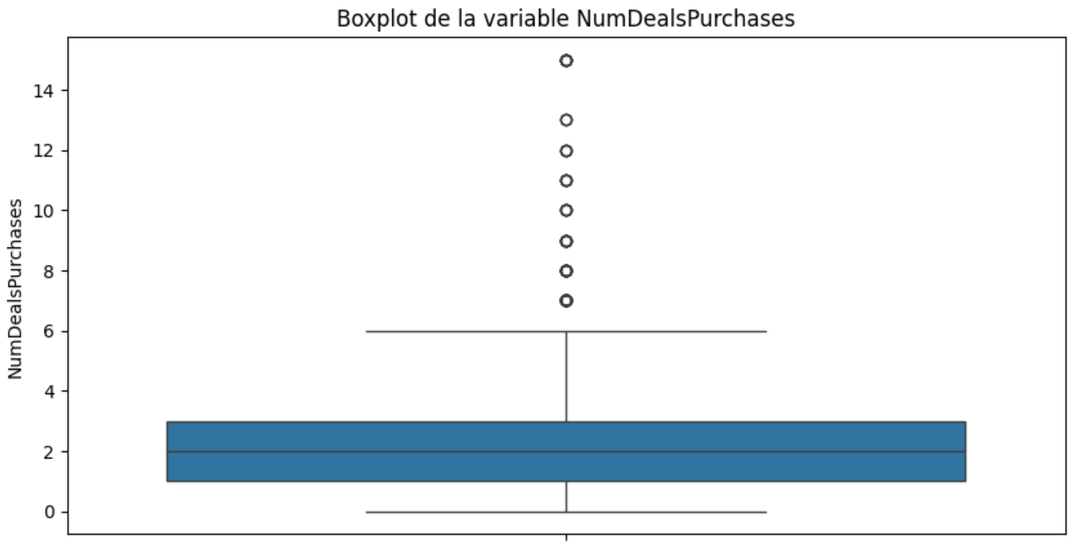

Boxplot de la variable MntFruits
MntFruits
0
25
50
75
100
125
150
175
200

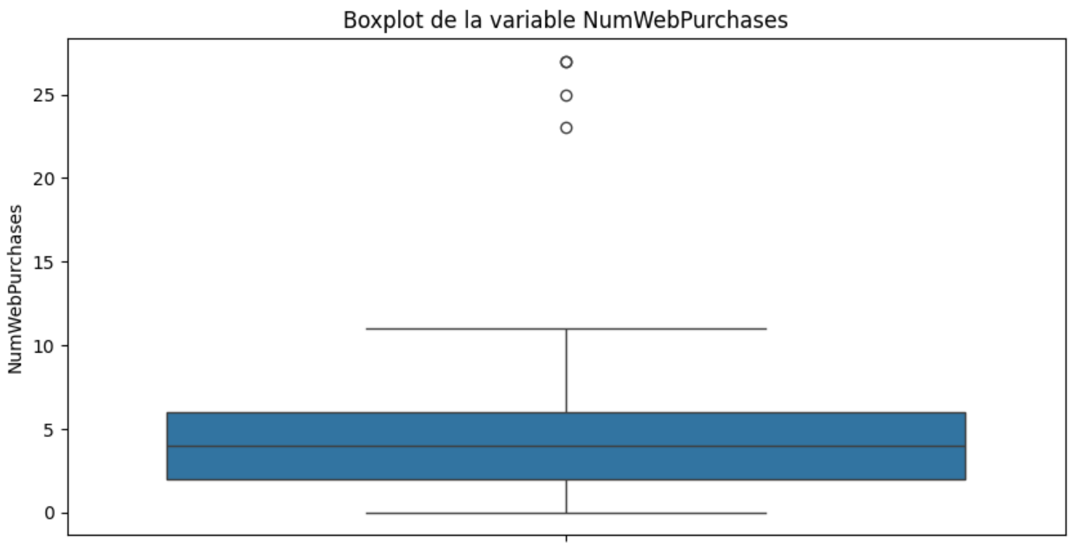

Boxplot de la variable MntMeatProducts
MntMeatProducts
0
250
500
750
1000
1250
1500
1750

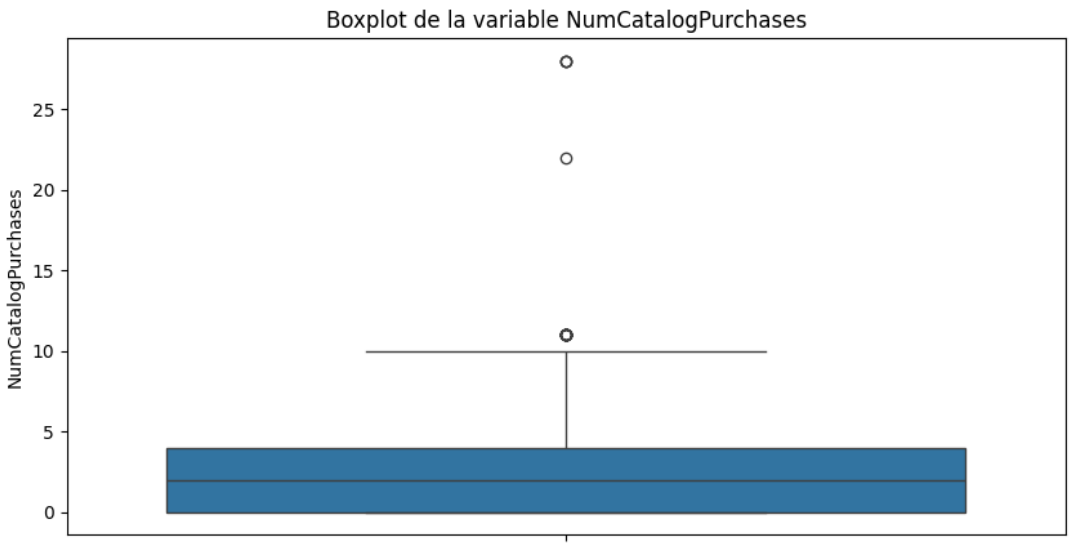

Boxplot de la variable MntFishProducts
MntFishProducts
250
200
150
100
50
0

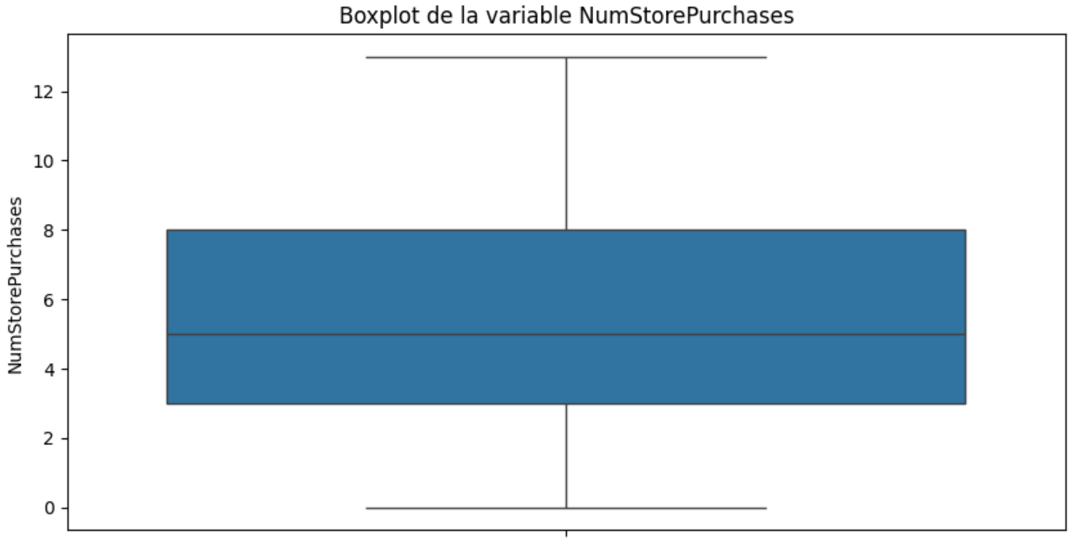

Boxplot de la variable MntSweetProducts
MntSweetProducts
250
200
150
100
50
0

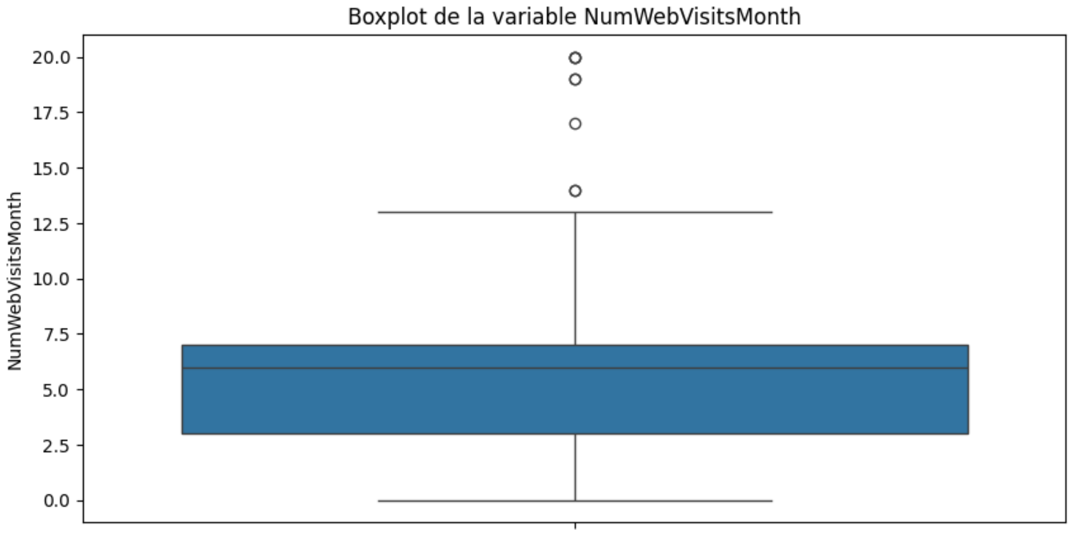

Boxplot de la variable MntGoldProds
MntGoldProds
350
300
250
200
150
100
50
0

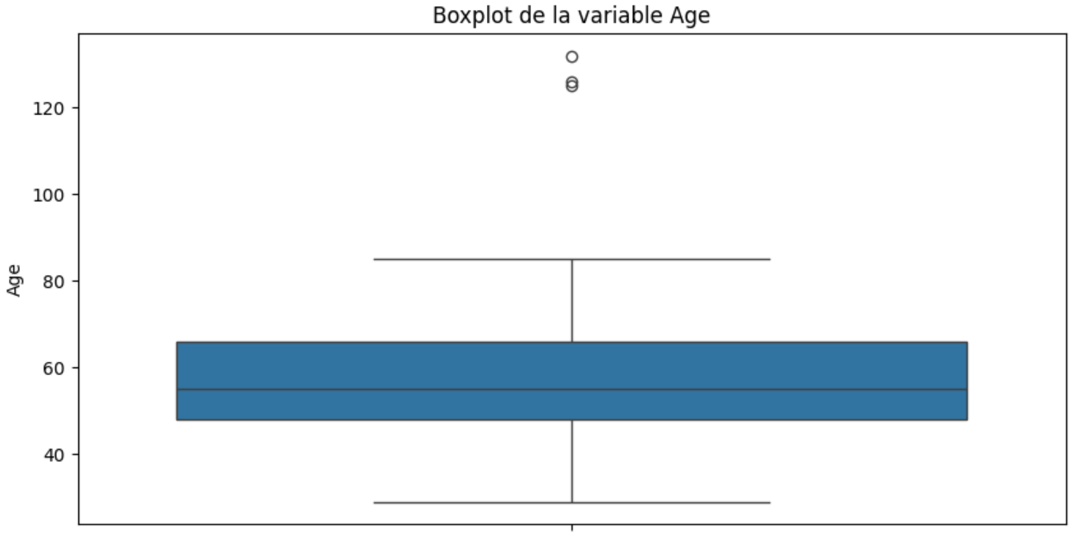

Boxplot de la variable NumDealsPurchases
NumDealsPurchases
14
12
10
8
6
4
2
0

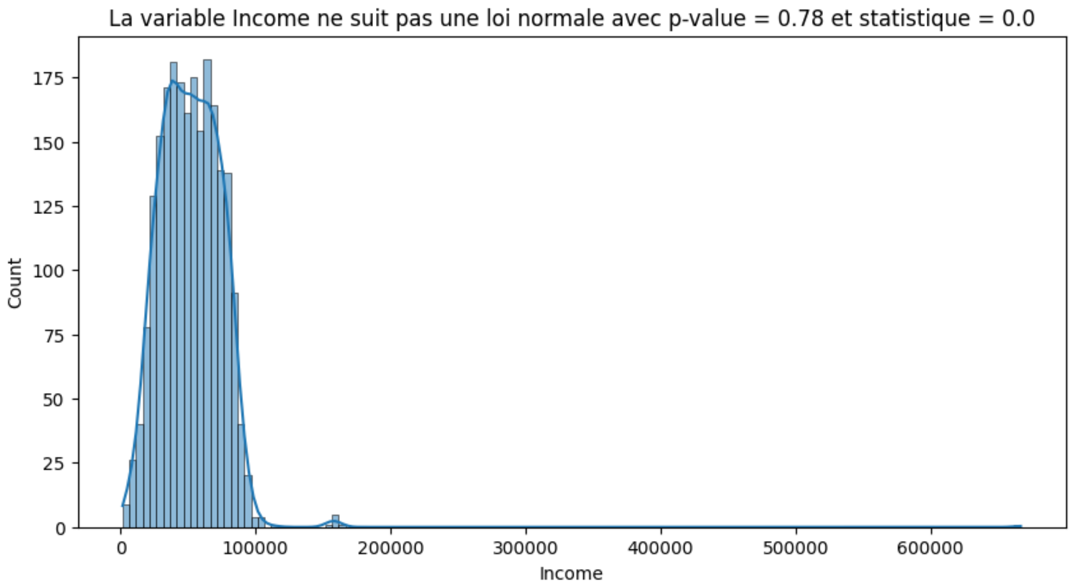

Boxplot de la variable NumWebPurchases
NumWebPurchases
0
5
10
15
20
25

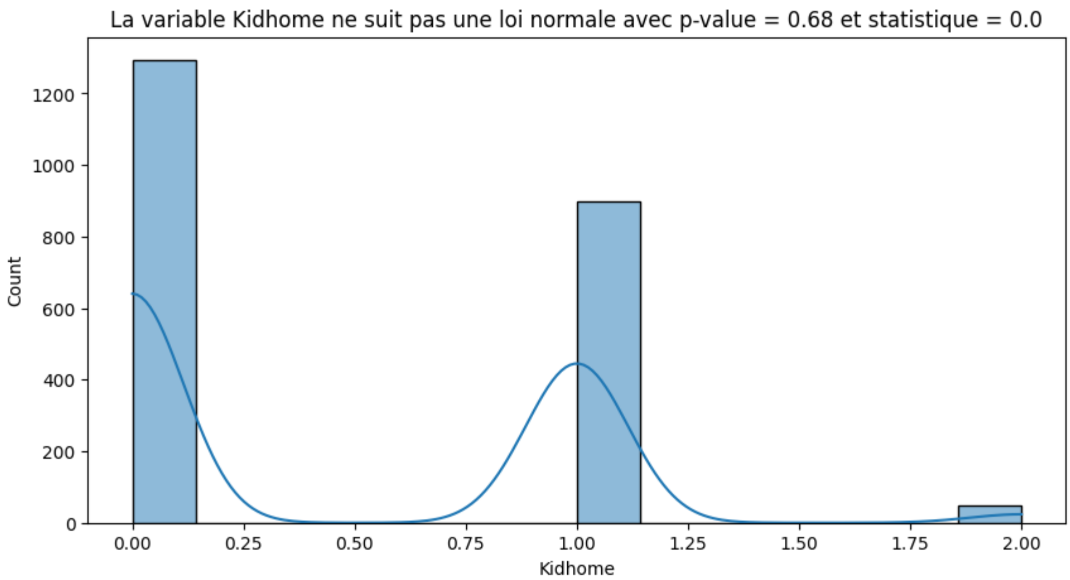

Boxplot de la variable NumCatalogPurchases
NumCatalogPurchases
0
5
10
15
20
25

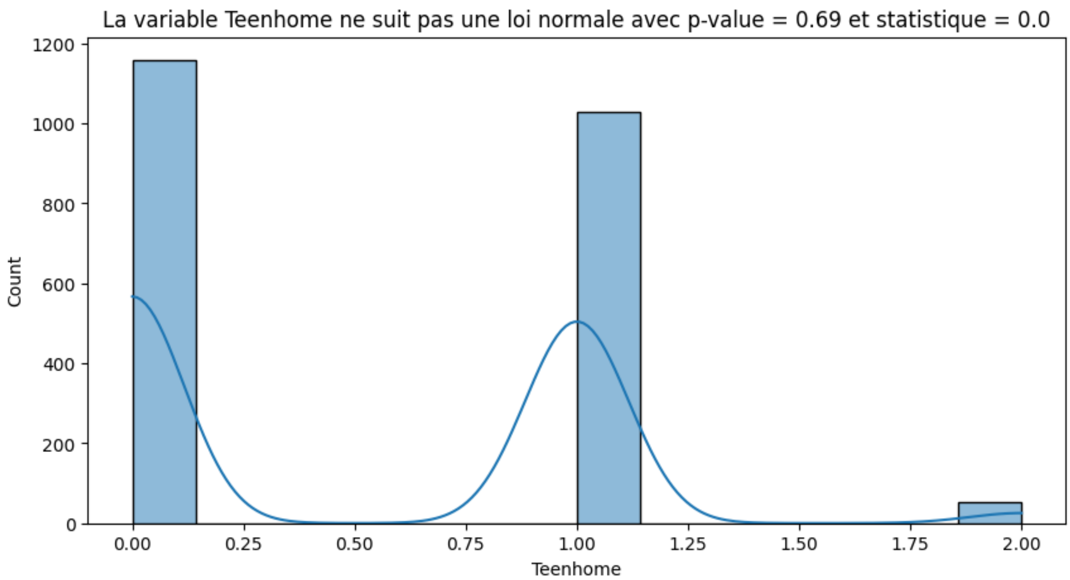

Boxplot de la variable NumStorePurchases
NumStorePurchases
0
2
4
6
8
10
12

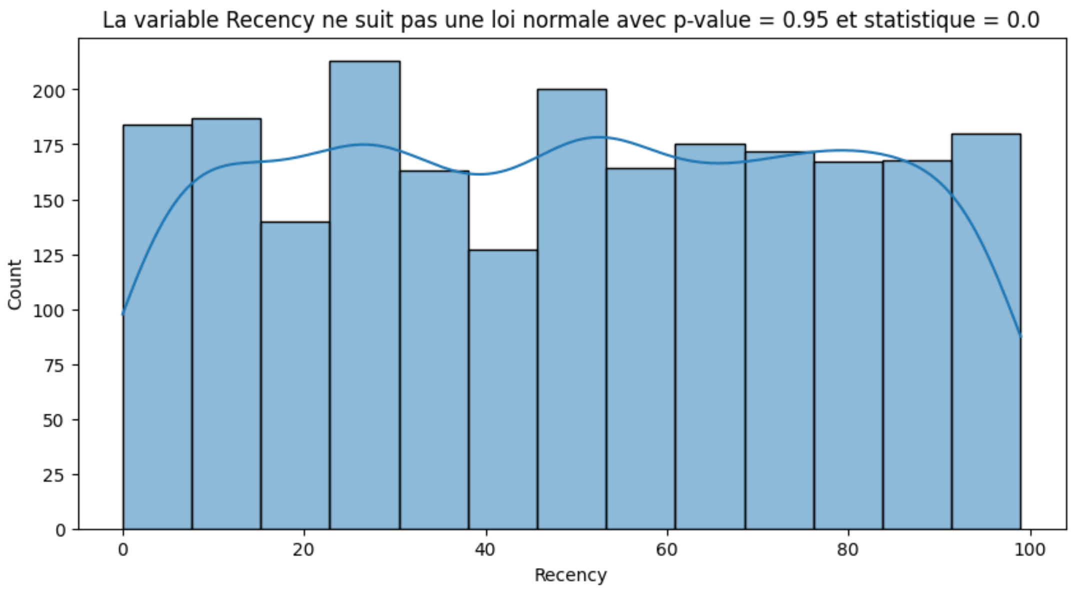

Boxplot de la variable NumWebVisitsMonth
NumWebVisitsMonth
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

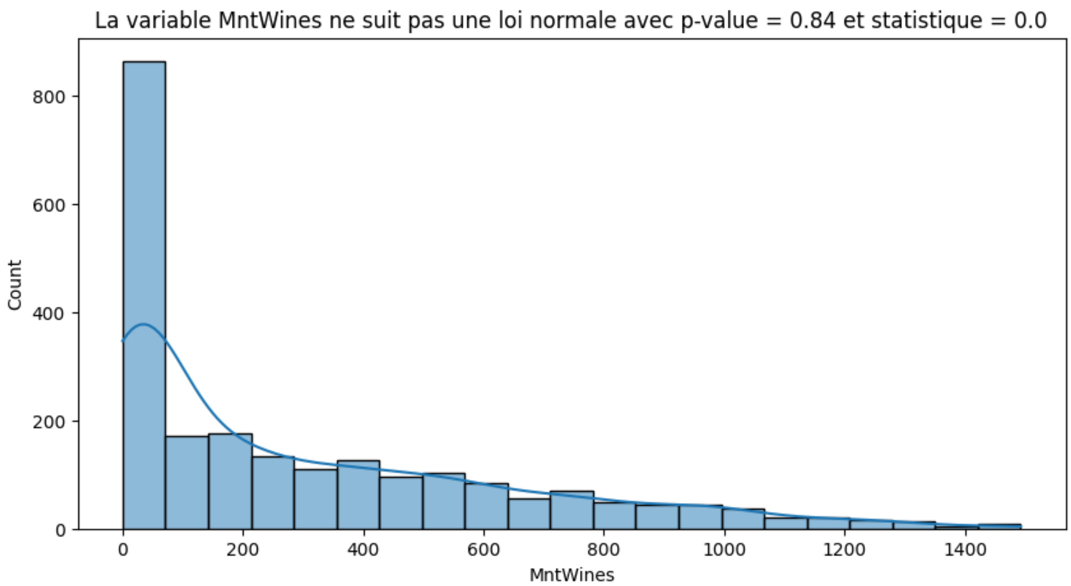

Boxplot de la variable Age
Age
120
100
80
60
40

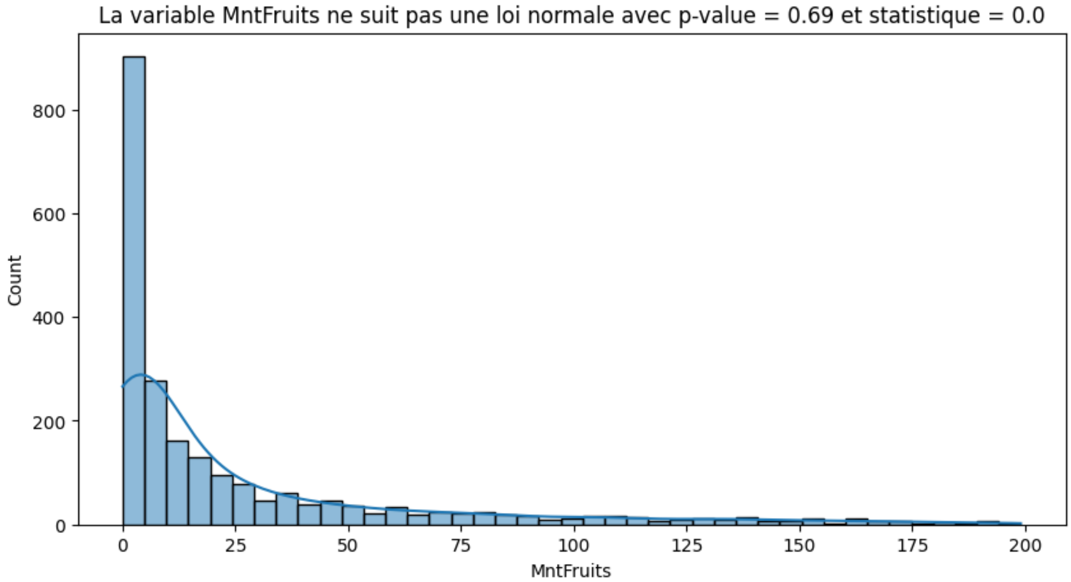

La variable Income ne suit pas une loi normale avec p-value = 0.78 et statistique = 0.0
Count
175
150
125
100
75
50
25
0
0
100000
200000
300000
400000
500000
600000
Income

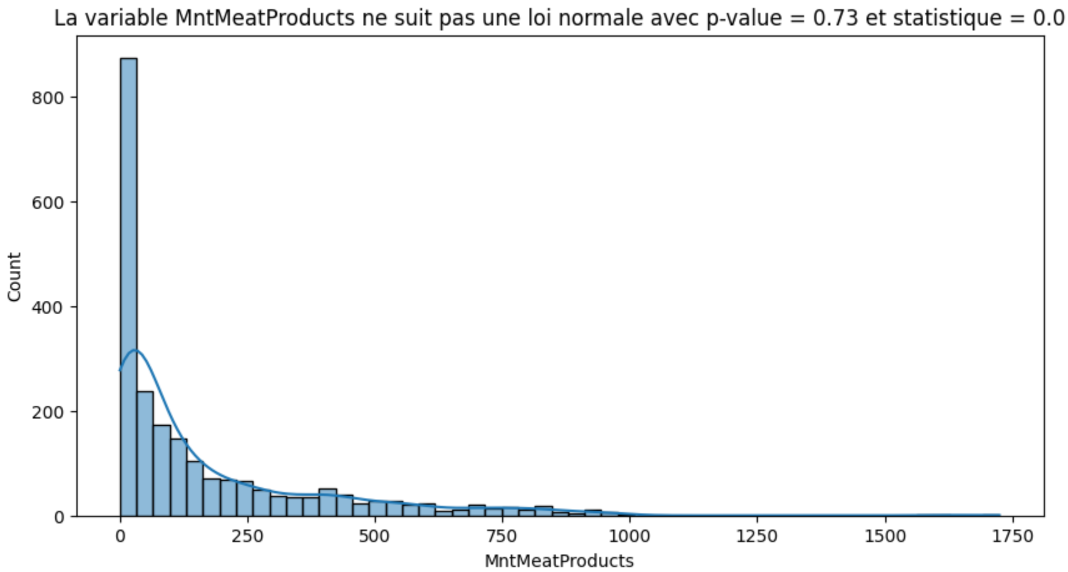

La variable Kidhome ne suit pas une loi normale avec p-value = 0.68 et statistique = 0.0
Count
0
200
400
600
800
1000
1200
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
Kidhome

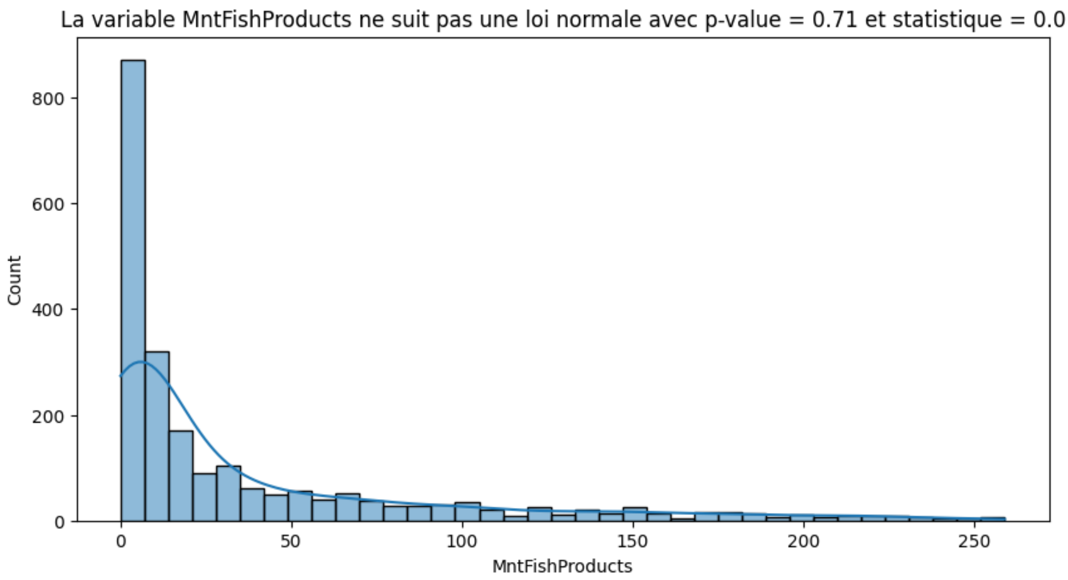

La variable Teenhome ne suit pas une loi normale avec p-value = 0.69 et statistique = 0.0
Count
0
200
400
600
800
1000
1200
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
Teenhome

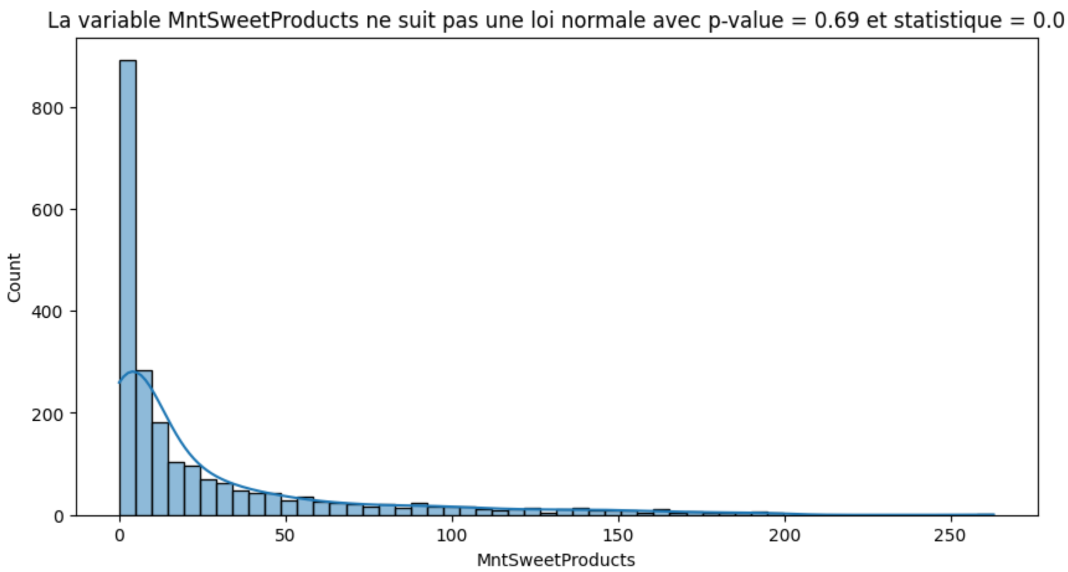

La variable Recency ne suit pas une loi normale avec p-value = 0.95 et statistique = 0.0
Count
Recency

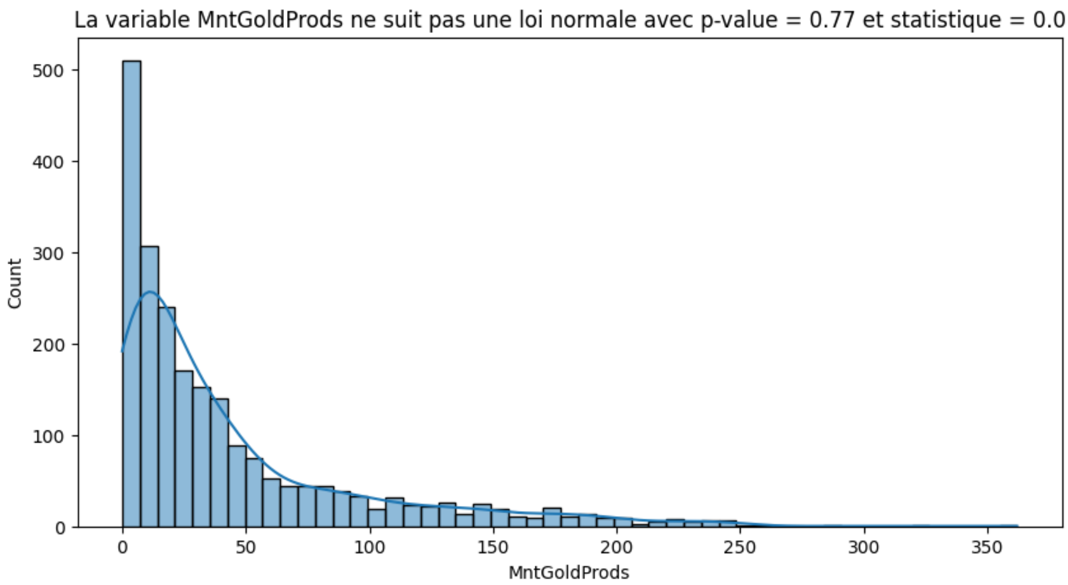

La variable MntWines ne suit pas une loi normale avec p-value = 0.84 et statistique = 0.0
Count
MntWines

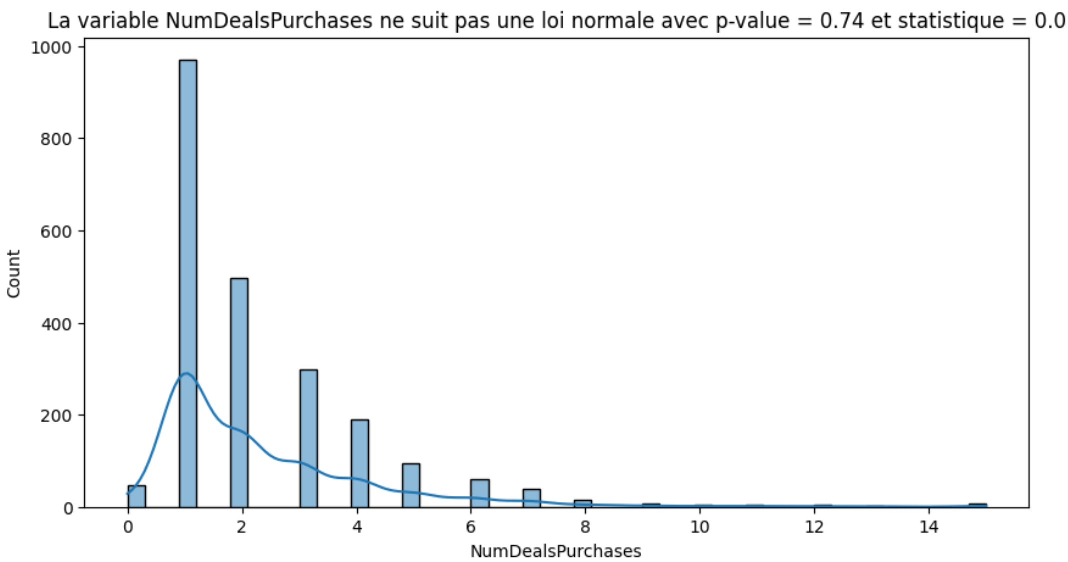

La variable MntFruits ne suit pas une loi normale avec p-value = 0.69 et statistique = 0.0
Count
0
200
400
600
800
0
25
50
75
100
125
150
175
200
MntFruits

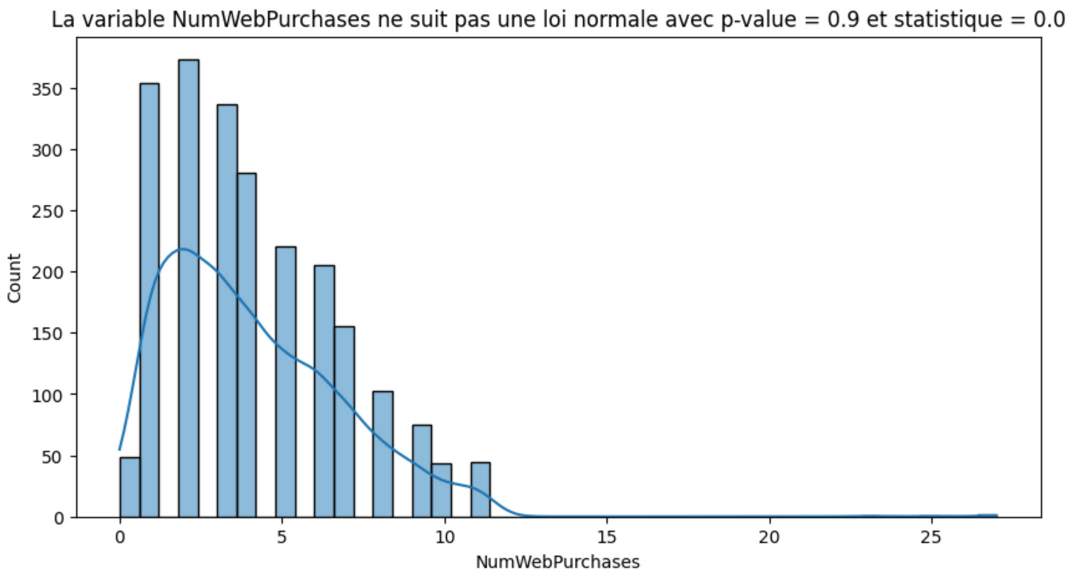

La variable MntMeatProducts ne suit pas une loi normale avec p-value = 0.73 et statistique = 0.0
Count
0
200
400
600
800
0
250
500
750
1000
1250
1500
1750
MntMeatProducts

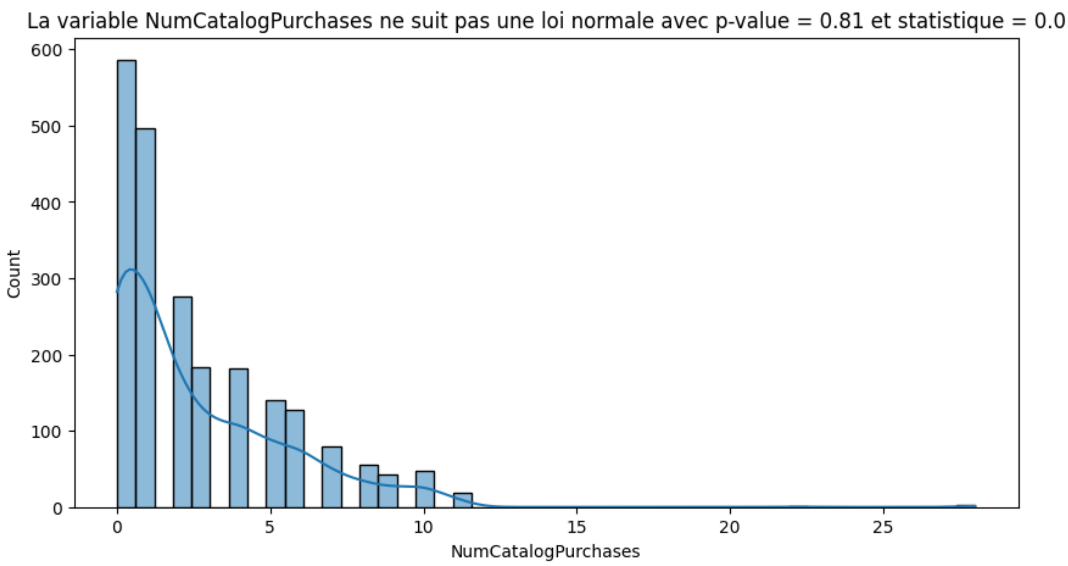

La variable MntFishProducts ne suit pas une loi normale avec p-value = 0.71 et statistique = 0.0
Count
MntFishProducts
0
200
400
600
800
0
50
100
150
200
250

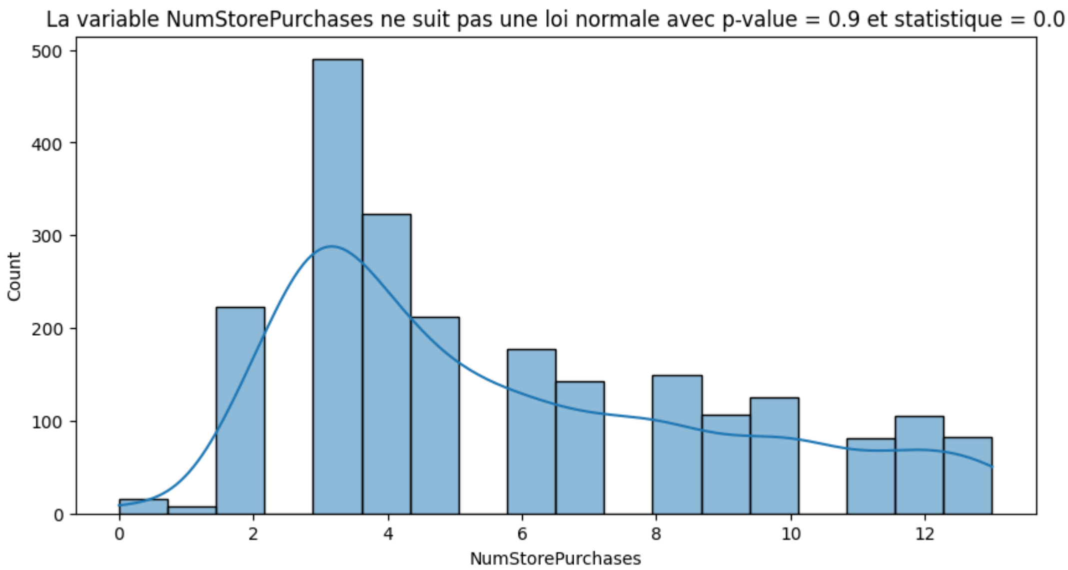

La variable MntSweetProducts ne suit pas une loi normale avec p-value = 0.69 et statistique = 0.0
Count
MntSweetProducts
0
200
400
600
800
0
50
100
150
200
250

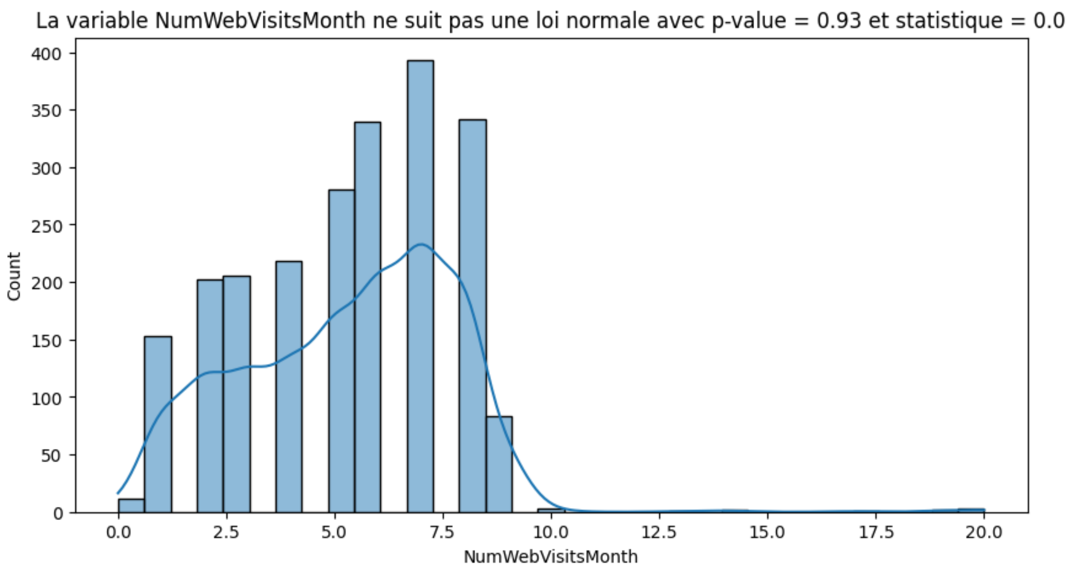

La variable MntGoldProds ne suit pas une loi normale avec p-value = 0.77 et statistique = 0.0
Count
MntGoldProds
0
50
100
150
200
250
300
350
0
100
200
300
400
500

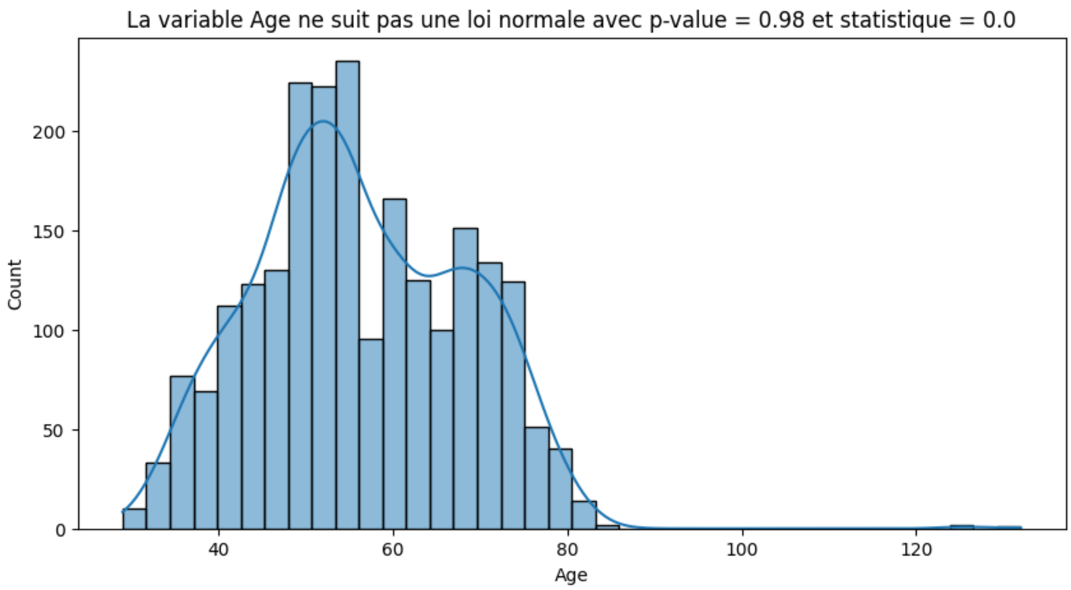

La variable NumDealsPurchases ne suit pas une loi normale avec p-value = 0.74 et statistique = 0.0
Count
NumDealsPurchases
0
2
4
6
8
10
12
14
0
200
400
600
800
1000


La variable NumWebPurchases ne suit pas une loi normale avec p-value = 0.9 et statistique = 0.0
Count
NumWebPurchases

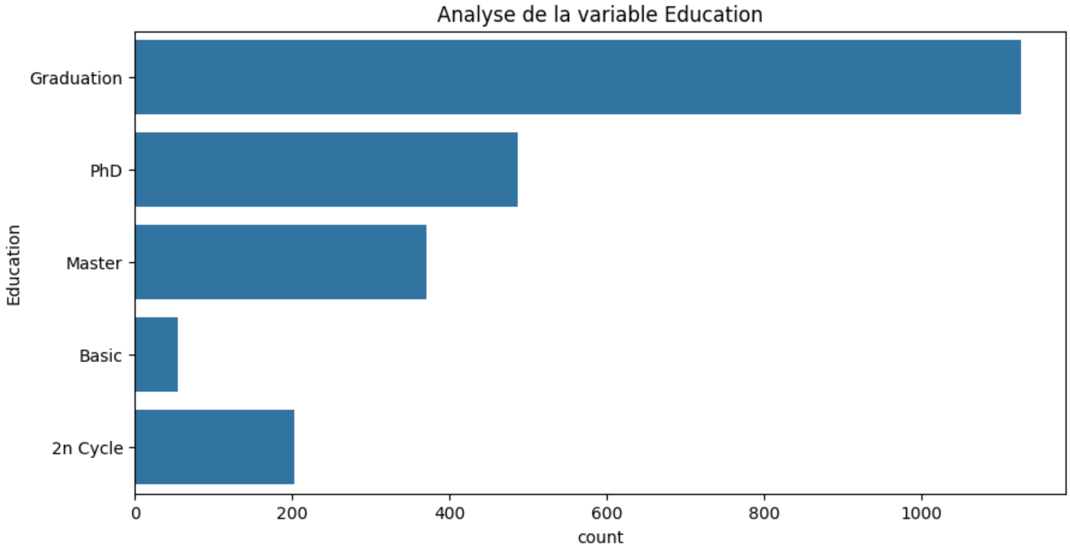

La variable NumCatalogPurchases ne suit pas une loi normale avec p-value = 0.81 et statistique = 0.0
Count
NumCatalogPurchases


La variable NumStorePurchases ne suit pas une loi normale avec p-value = 0.9 et statistique = 0.0
Count
NumStorePurchases

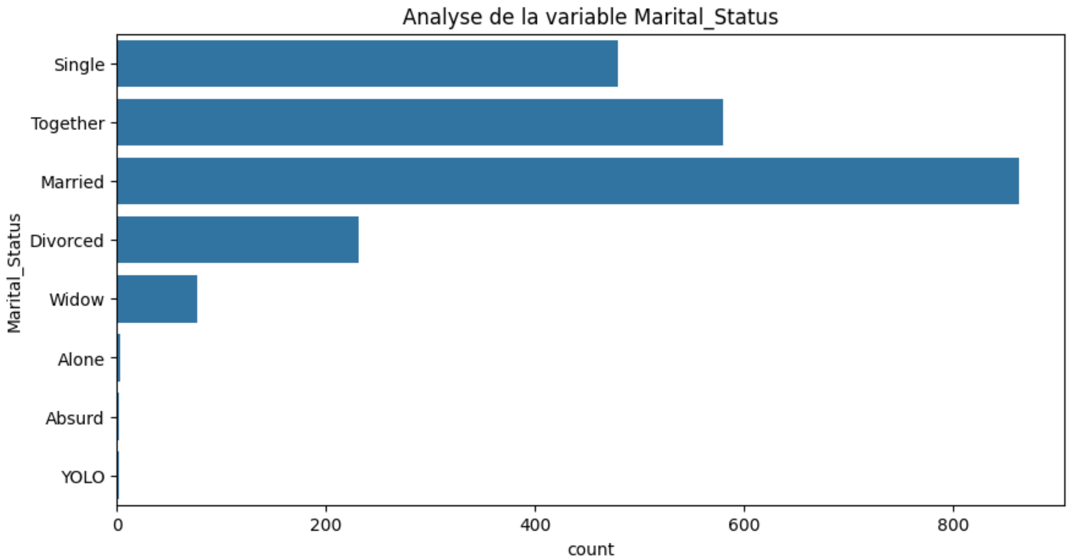

La variable NumWebVisitsMonth ne suit pas une loi normale avec p-value = 0.93 et statistique = 0.0
Count
NumWebVisitsMonth


La variable Age ne suit pas une loi normale avec p-value = 0.98 et statistique = 0.0
Count
200
150
100
50
0
40
60
80
100
120
Age

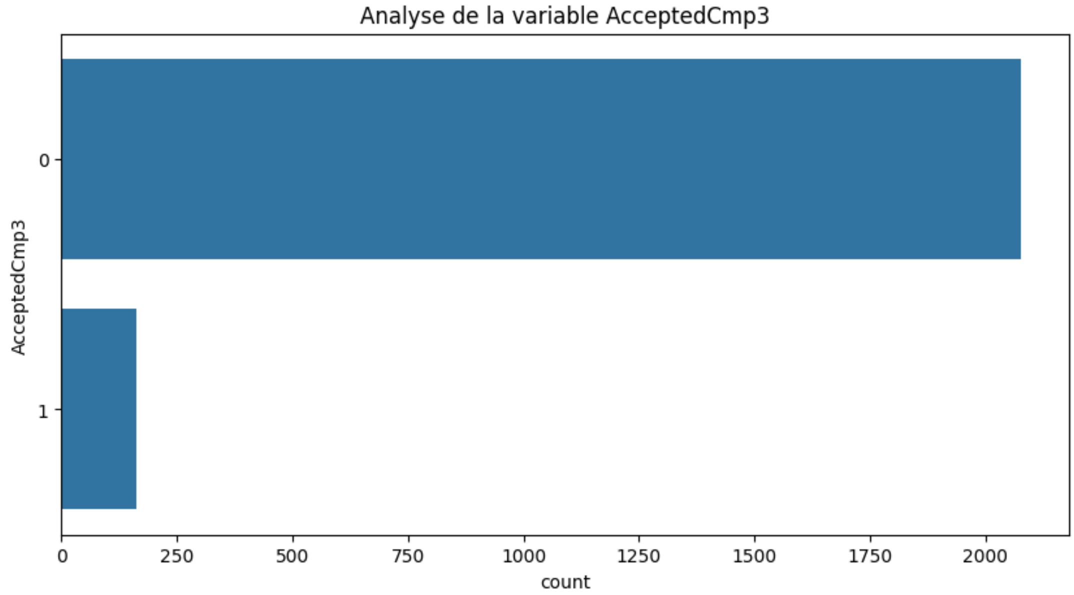
```
la fréquance des modalités de la variable --Education--
Education
Graduation    50.312500
PhD           21.696429
Master        16.517857
2n Cycle       9.062500
Basic          2.410714
Name: proportion, dtype: float64
```



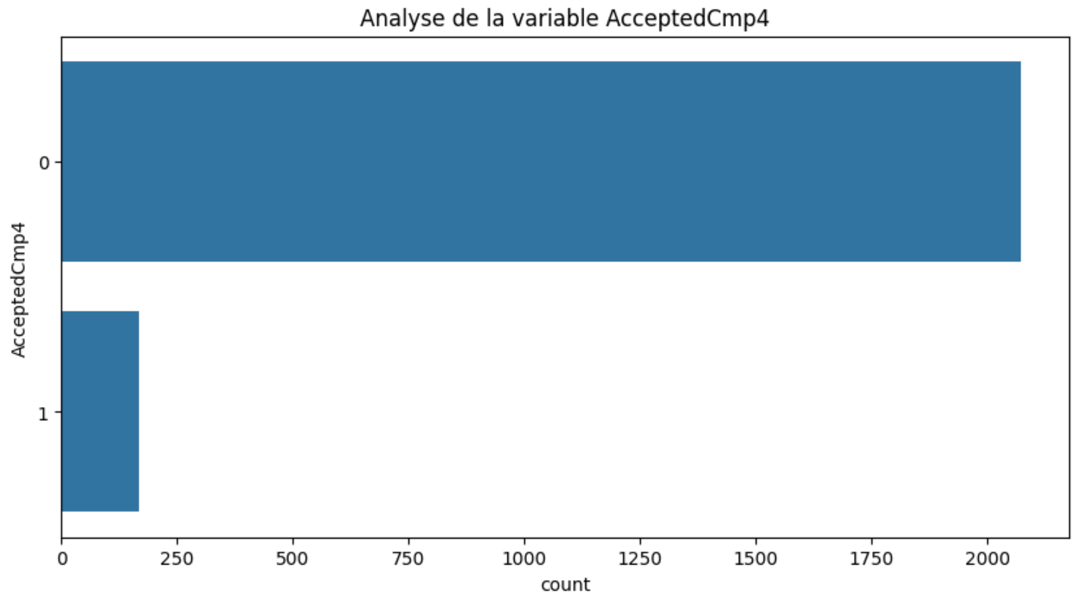
```
 la fréquance des modalités de la variable --Marital_Status--
 Marital_Status
Married     38.571429
Together    25.892857
Single      21.428571
Divorced    10.357143
Widow        3.437500
Alone        0.133929
Absurd       0.089286
YOLO         0.089286
Name: proportion, dtype: float64
```




```
 la fréquance des modalités de la variable --AcceptedCmp3--
 AcceptedCmp3
0    92.723214
1     7.276786
Name: proportion, dtype: float64
```

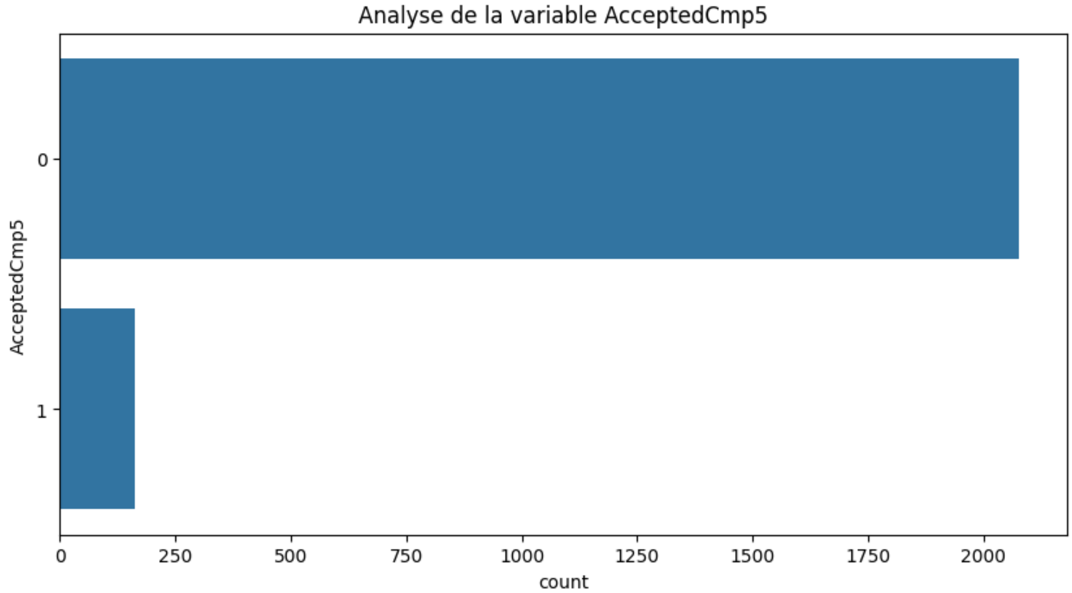



```
 la fréquance des modalités de la variable --AcceptedCmp4--
 AcceptedCmp4
0    92.544643
1     7.455357
Name: proportion, dtype: float64
```

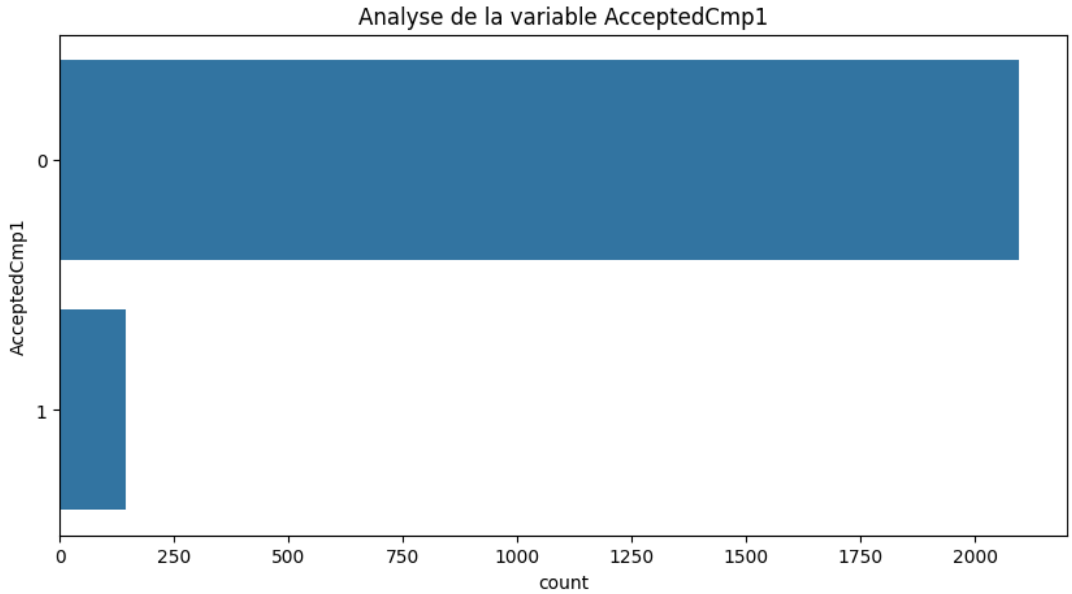



```
 la fréquance des modalités de la variable --AcceptedCmp5--
 AcceptedCmp5
0    92.723214
```

```
1     7.276786
Name: proportion, dtype: float64
```

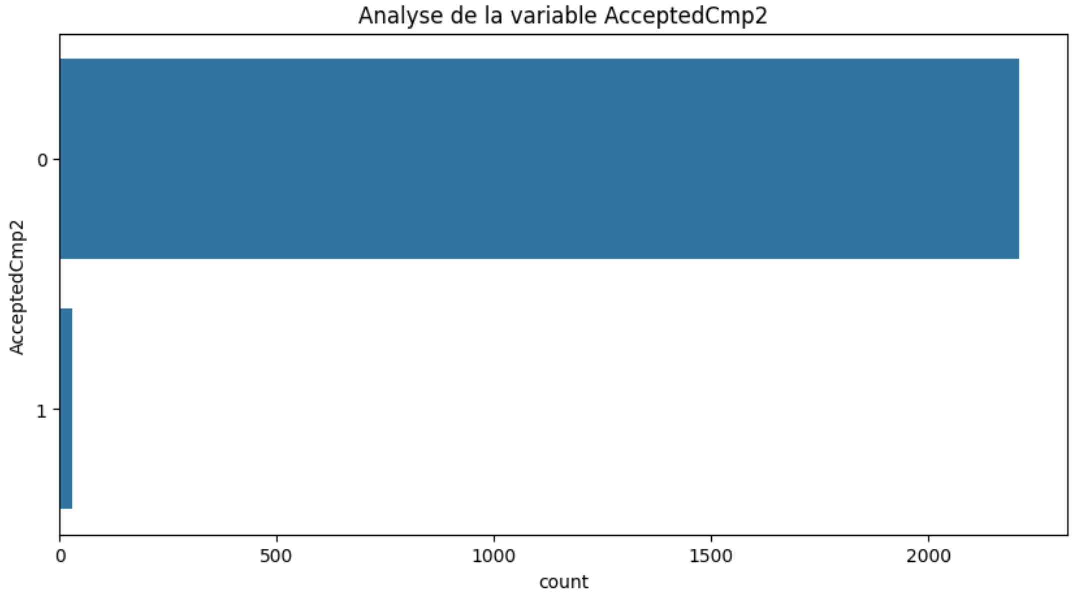



```
 la fréquance des modalités de la variable --AcceptedCmp1--
 AcceptedCmp1
0    93.571429
1     6.428571
Name: proportion, dtype: float64
```

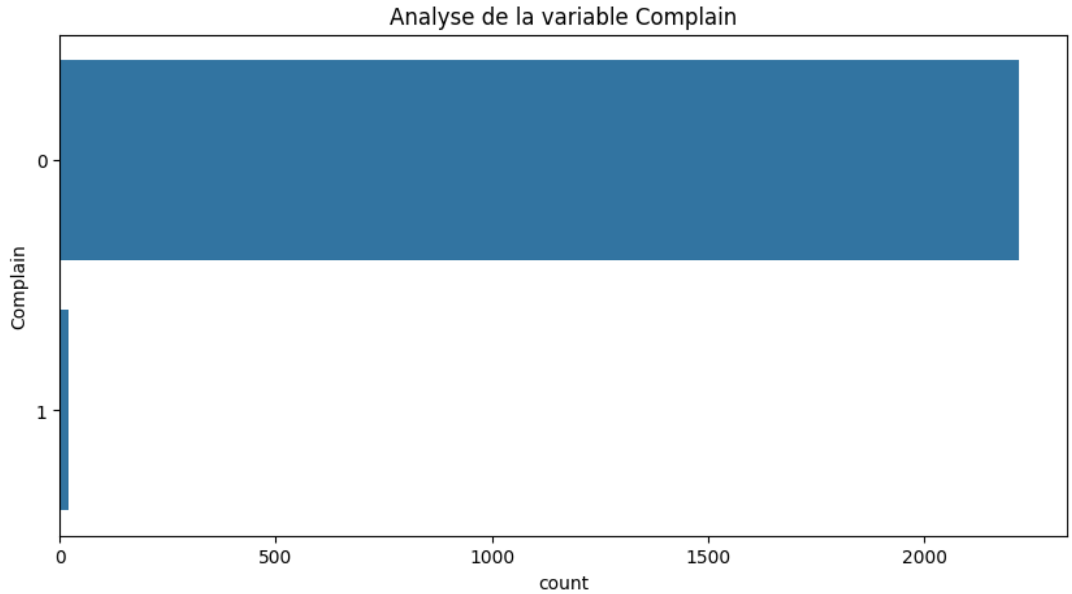



```
 la fréquance des modalités de la variable --AcceptedCmp2--
```

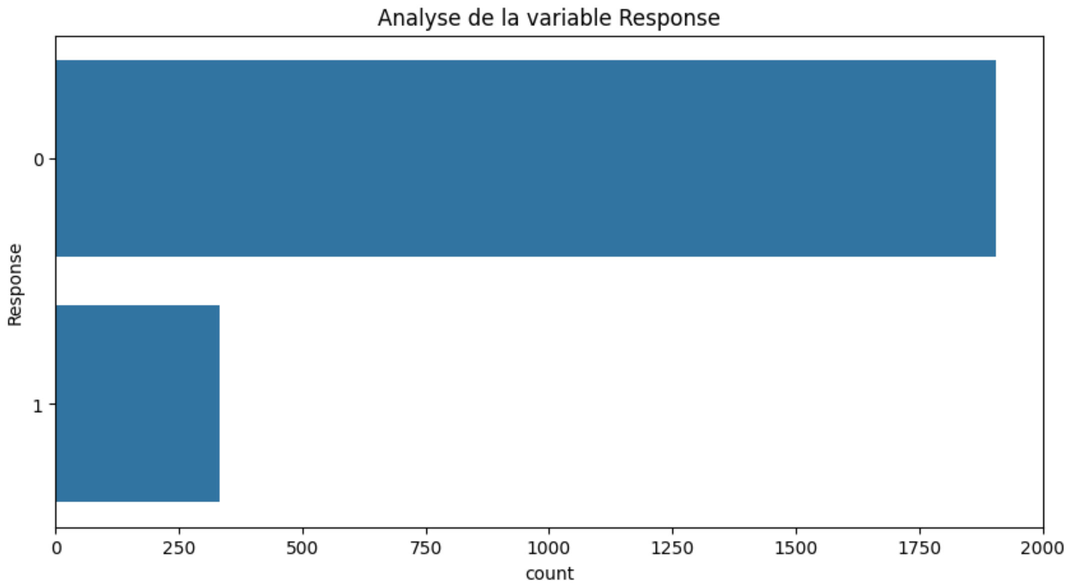
```
 AcceptedCmp2
0    98.660714
1     1.339286
Name: proportion, dtype: float64
```




```
 la fréquance des modalités de la variable --Complain--
 Complain
0    99.0625
1     0.9375
Name: proportion, dtype: float64
```



```
 la fréquance des modalités de la variable --Response--
 Response
0    85.089286
1    14.910714
Name: proportion, dtype: float64
```



Resultat de l'analyse univariée

- Les resultats nous indique la presence des outilers pour certaines variables telle que le revenu du client, le montant de ses dépenses en : viande, en poisson etc.
- Aucune variable n'est normale.

Gestion des outliers

[95]:
```python
# Detection des outliers
def detect_outliers(df):
    outliers_dict = {}  # Dictionnaire pour stocker les colonnes et leurs␣
 ↪valeurs aberrantes
    for var in df.columns:
        if df[var].dtype in ['int64', 'float64']:  # Vérifiez que la colonne␣
 ↪est numérique
            q1 = df[var].quantile(0.25)
            q3 = df[var].quantile(0.75)
            iqr = q3 - q1
            vmin = q1 - 1.5 * iqr
            vmax = q3 + 1.5 * iqr
            # Vérifiez s'il y a des valeurs aberrantes dans la colonne
            outliers = df[(df[var] < vmin) | (df[var] > vmax)]
            if not outliers.empty:  # Si la colonne contient des valeurs␣
 ↪aberrantes
                outliers_dict[var] = outliers[var].tolist()
```

```
                list_var = outliers_dict.keys()
    return list_var

# Exemple d'utilisation
outliers = detect_outliers(data)
print("Variables avec des valeurs aberrantes :", outliers)
```

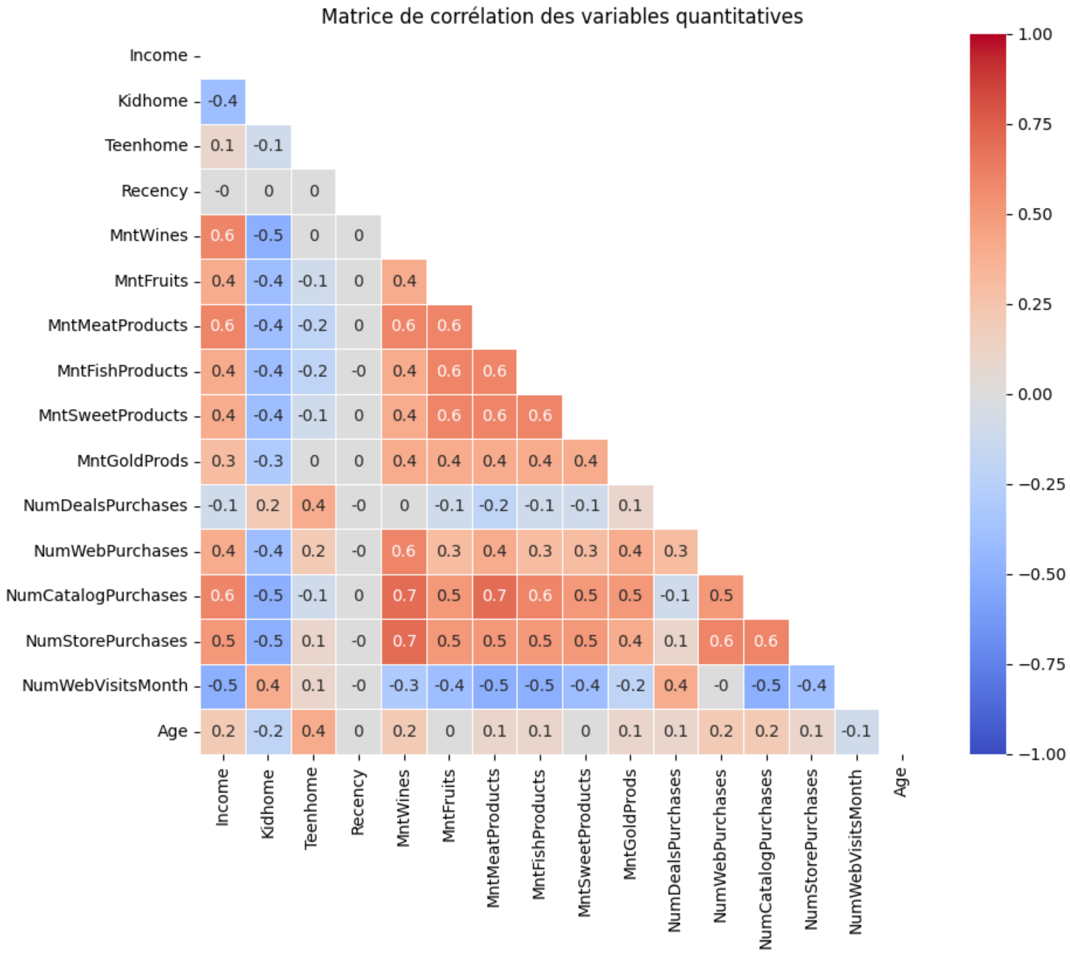
```
Variables avec des valeurs aberrantes : dict_keys(['Income', 'MntWines',
'MntFruits', 'MntMeatProducts', 'MntFishProducts', 'MntSweetProducts',
'MntGoldProds', 'NumDealsPurchases', 'NumWebPurchases', 'NumCatalogPurchases',
'NumWebVisitsMonth', 'Age'])
```

[96]:
```
# Traitement des outliers
def Traitement_Outliers(data, features, contamination=0.05):
    """
    Cette fonction détecte et gère les outliers en utilisant IsolationForest.

    Parameters:
    data (pd.DataFrame): Le dataframe contenant les données.
    features (list): La liste des colonnes à analyser pour les outliers.
    contamination (float): La proportion de données à considérer comme␣
 ↪outliers.

    Returns:
    pd.DataFrame: Le dataframe sans les outliers.
    """
    # Initialiser le modèle IsolationForest
    iso_forest = IsolationForest(contamination=contamination, random_state=42)

    # Ajuster le modèle sur les données
    iso_forest.fit(data[features])

    # Prédire les outliers
    data['outlier'] = iso_forest.predict(data[features])

    # Garder uniquement les données non-outliers
    data_cleaned = data[data['outlier'] == 1].drop(columns=['outlier'])

    return data_cleaned

# Utilisation de la fonction
features_to_check = ['Income', 'MntWines', 'MntFruits', 'MntMeatProducts',
                     'MntFishProducts', 'MntSweetProducts', 'MntGoldProds',
                     'NumDealsPurchases', 'NumWebPurchases',␣
 ↪'NumCatalogPurchases',
                     'NumWebVisitsMonth', 'Age']
data = Traitement_Outliers(data, features_to_check)
```

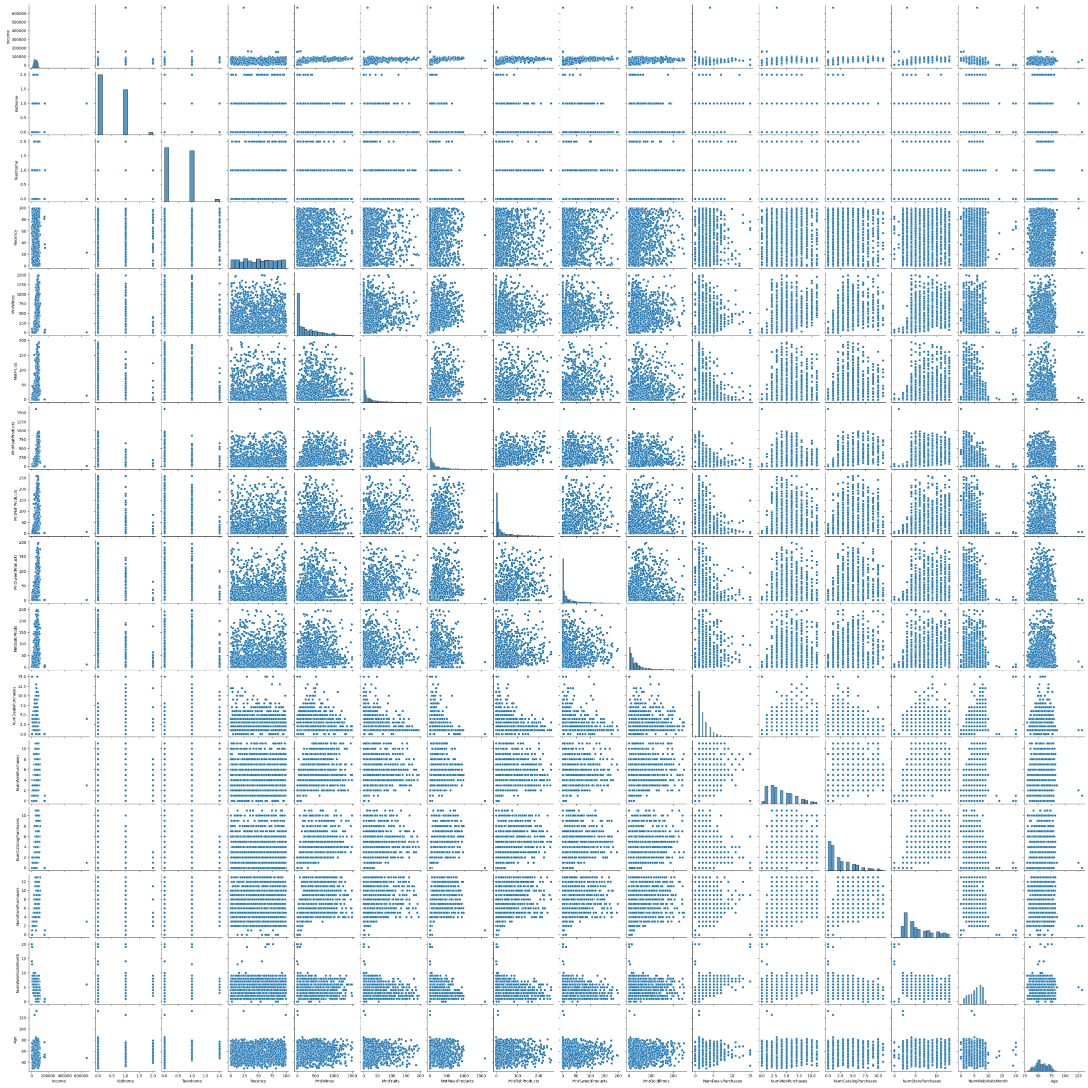
Analyse Bivariée

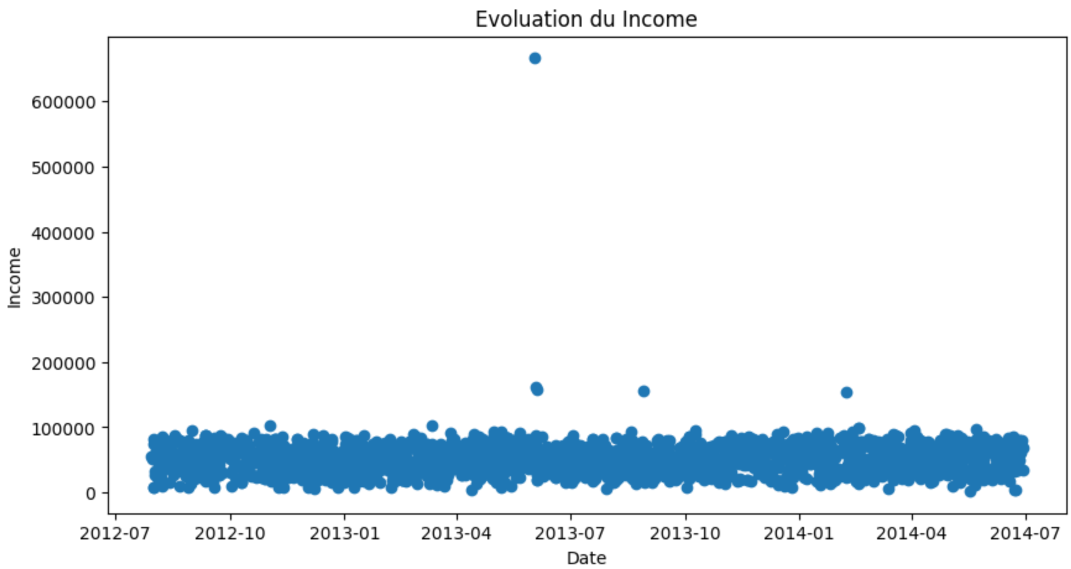
[106]:
```
AnalyseBivariee(data, var_quanti, var_quali)
```

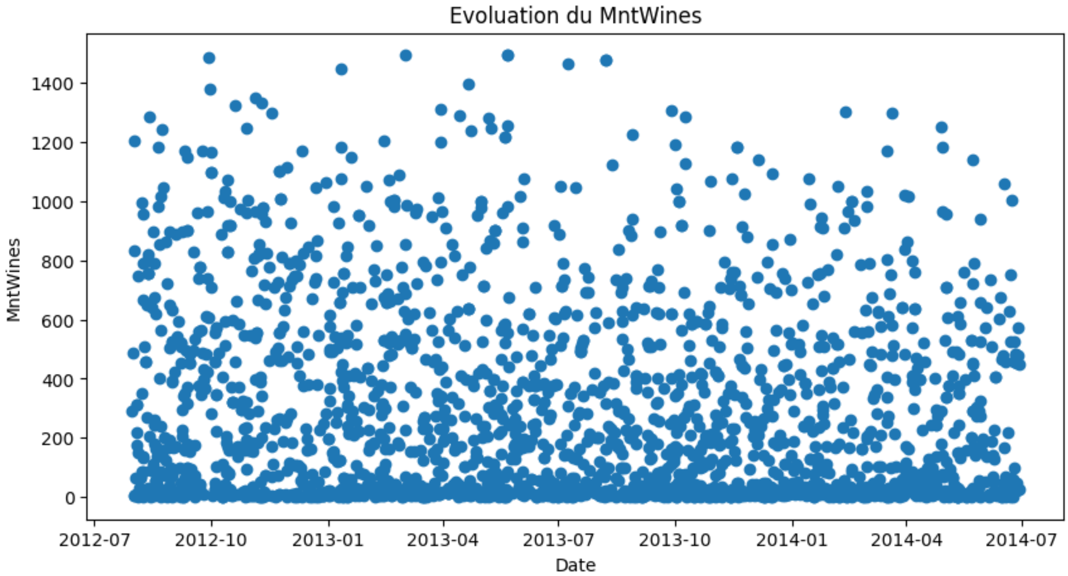
Analyse de la relation entre les variables quantitatives et quantitatives

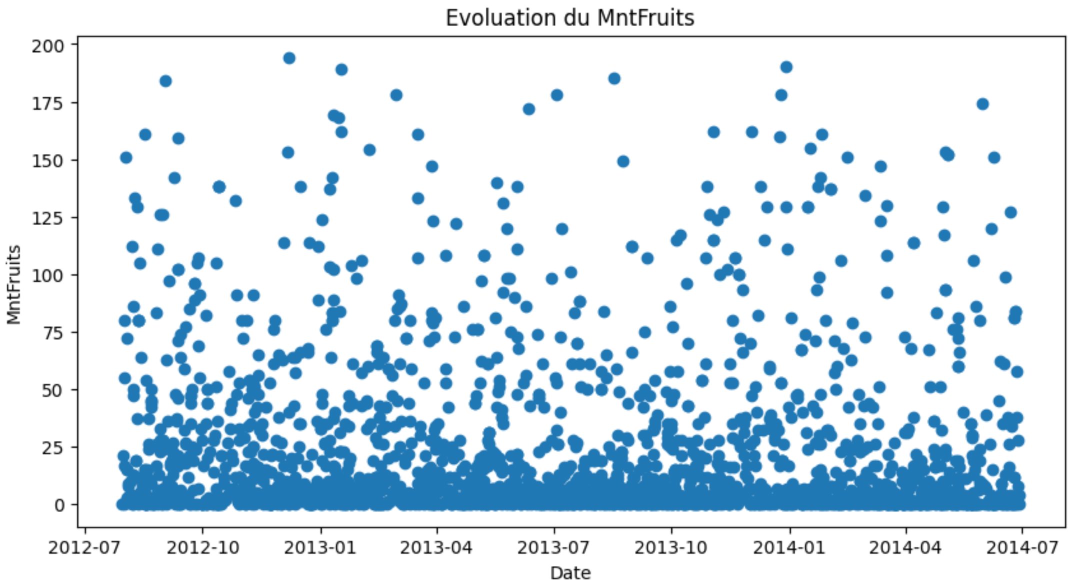


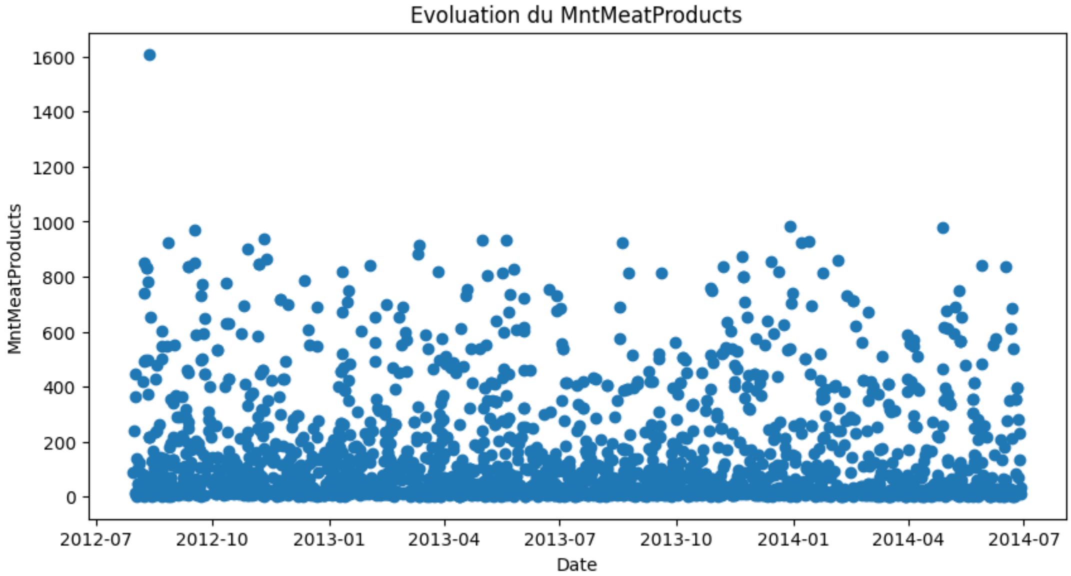
<Figure size 1000x500 with 0 Axes>

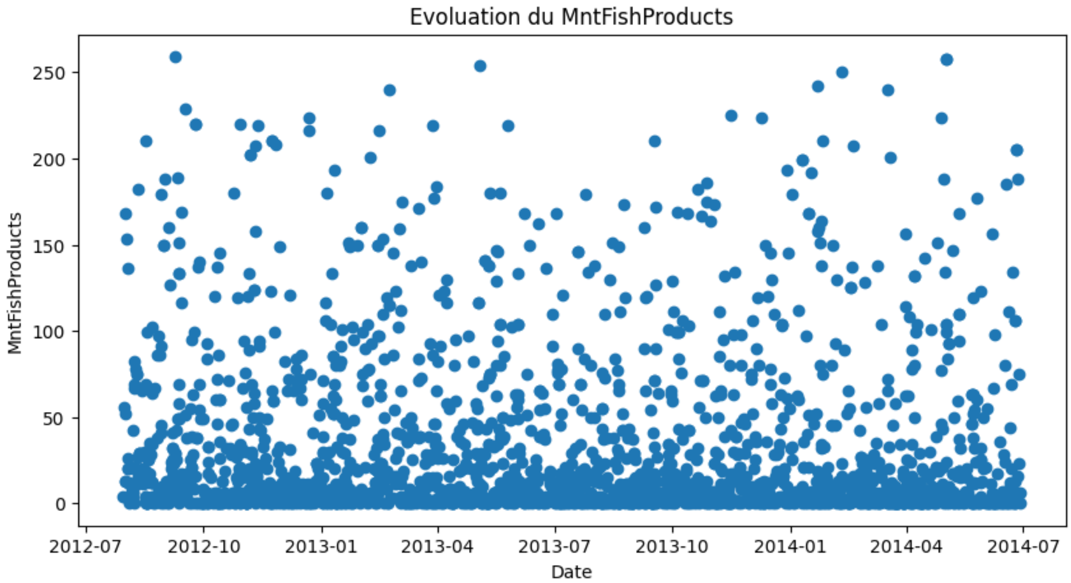


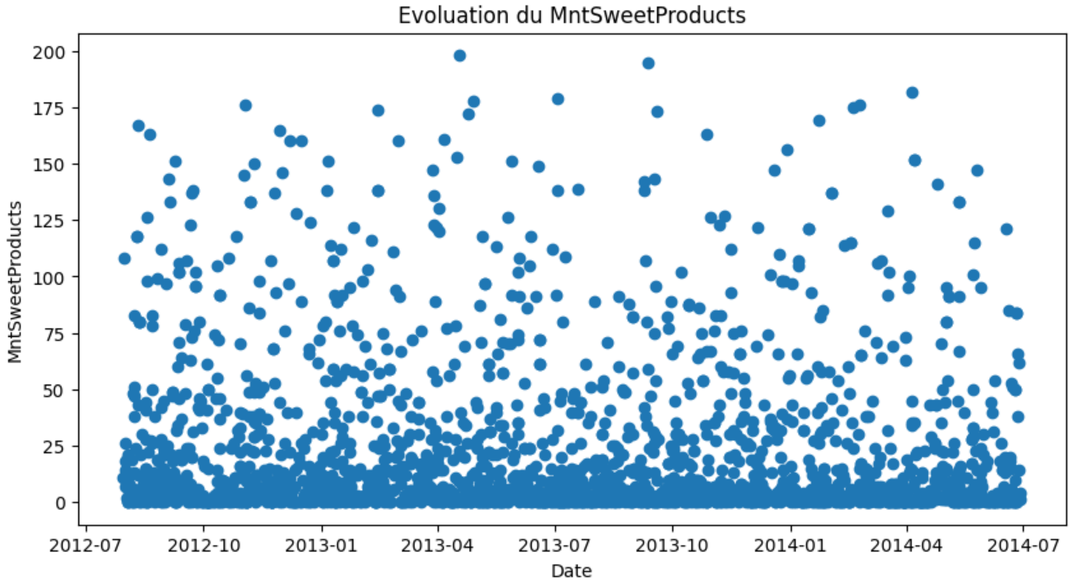

Evoluation du Income
Income
Date
0
100000
200000
300000
400000
500000
600000
2012-07
2012-10
2013-01
2013-04
2013-07
2013-10
2014-01
2014-04
2014-07

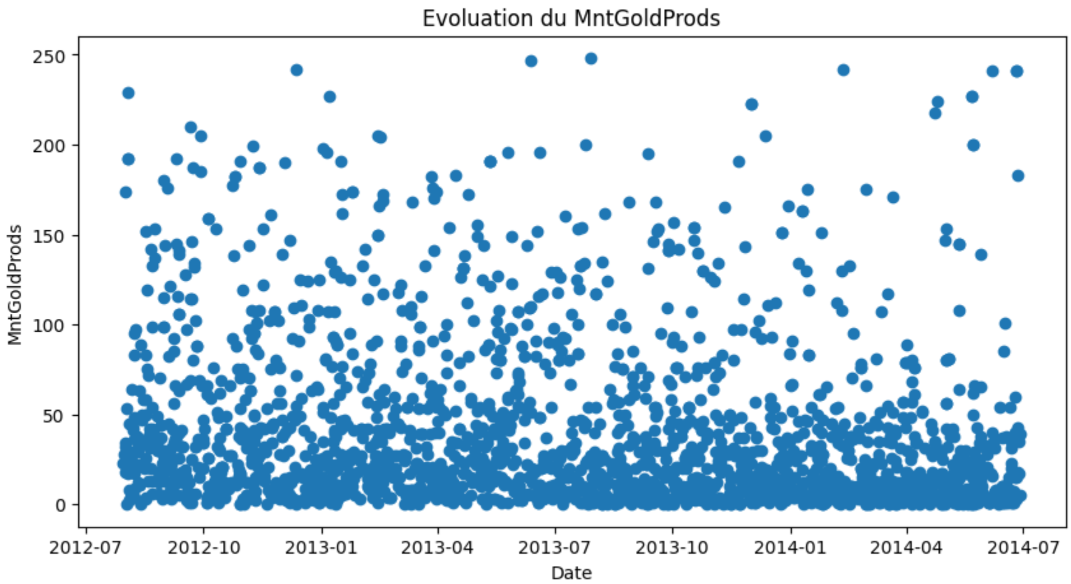

Evoluation du MntWines
MntWines
Date
0
200
400
600
800
1000
1200
1400
2012-07
2012-10
2013-01
2013-04
2013-07
2013-10
2014-01
2014-04
2014-07

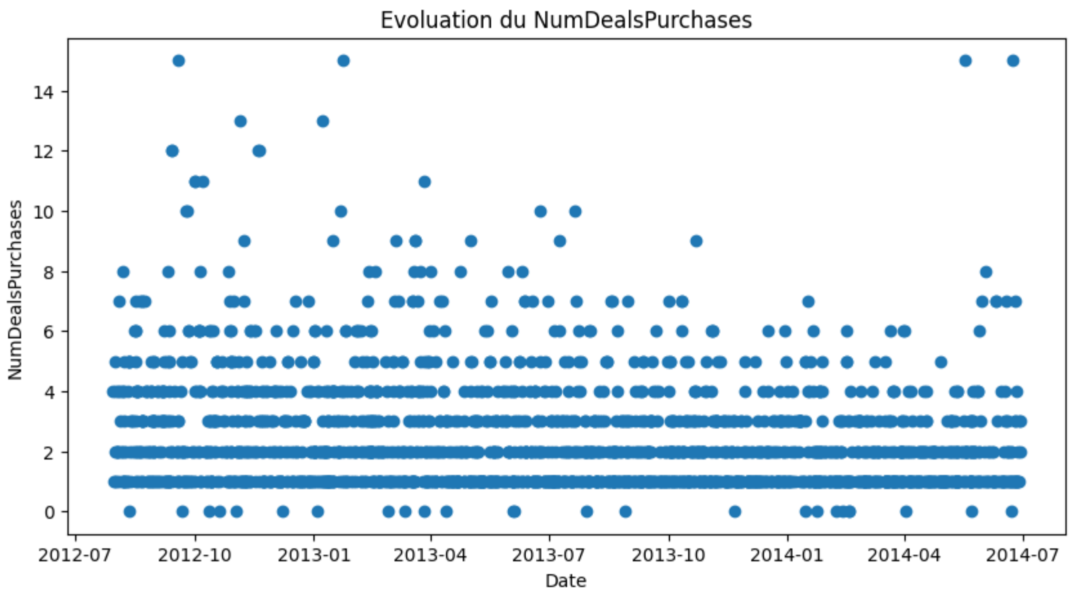

Evoluation du MntFruits
MntFruits
Date
0
25
50
75
100
125
150
175
200
2012-07
2012-10
2013-01
2013-04
2013-07
2013-10
2014-01
2014-04
2014-07

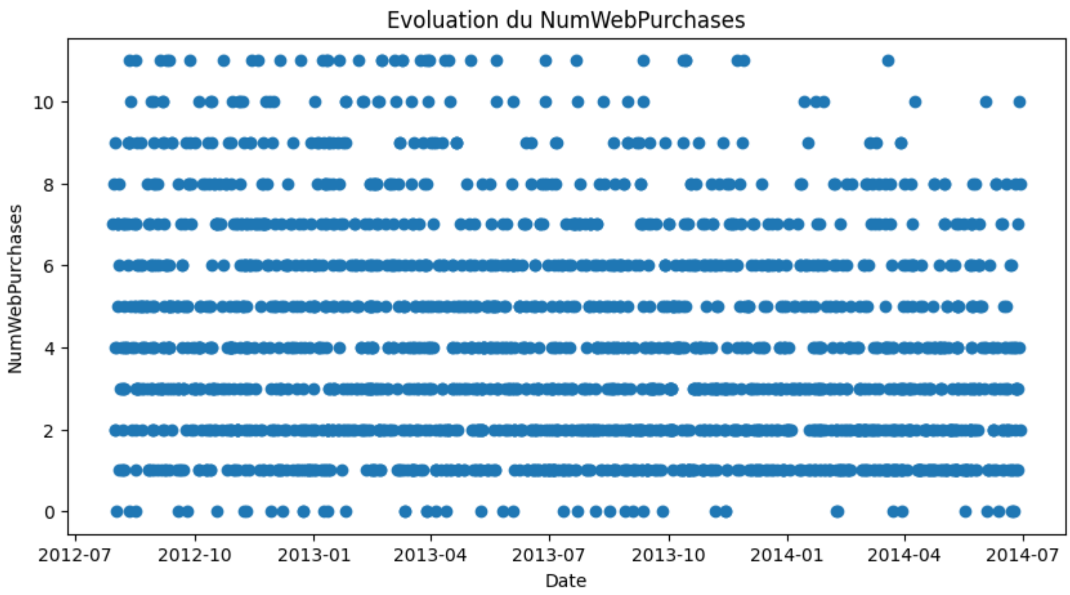

Evoluation du MntMeatProducts
MntMeatProducts
0
200
400
600
800
1000
1200
1400
1600
2012-07
2012-10
2013-01
2013-04
2013-07
2013-10
2014-01
2014-04
2014-07
Date

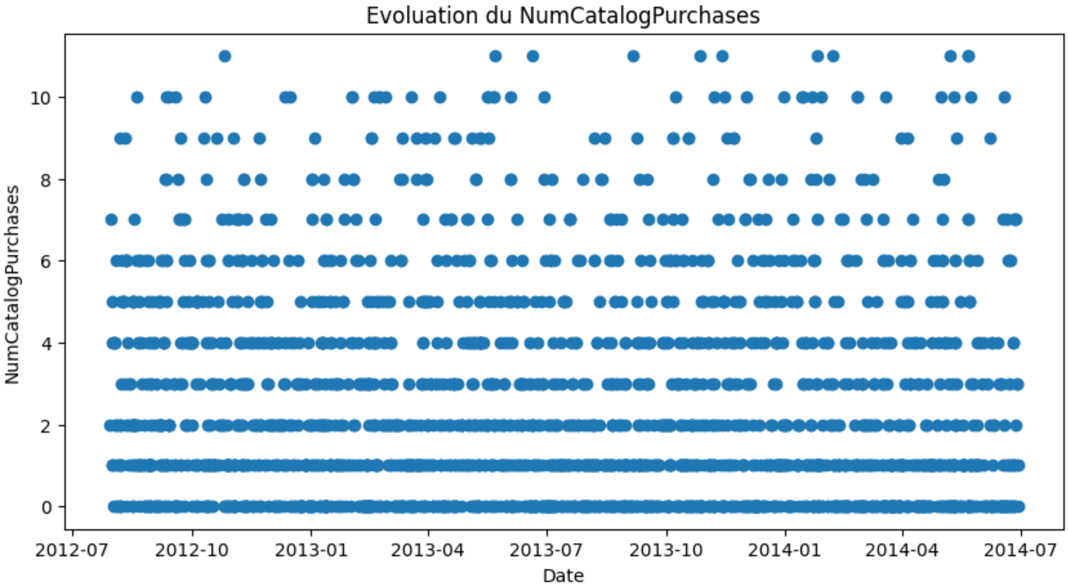

Evoluation du MntFishProducts
MntFishProducts
0
50
100
150
200
250
2012-07
2012-10
2013-01
2013-04
2013-07
2013-10
2014-01
2014-04
2014-07
Date

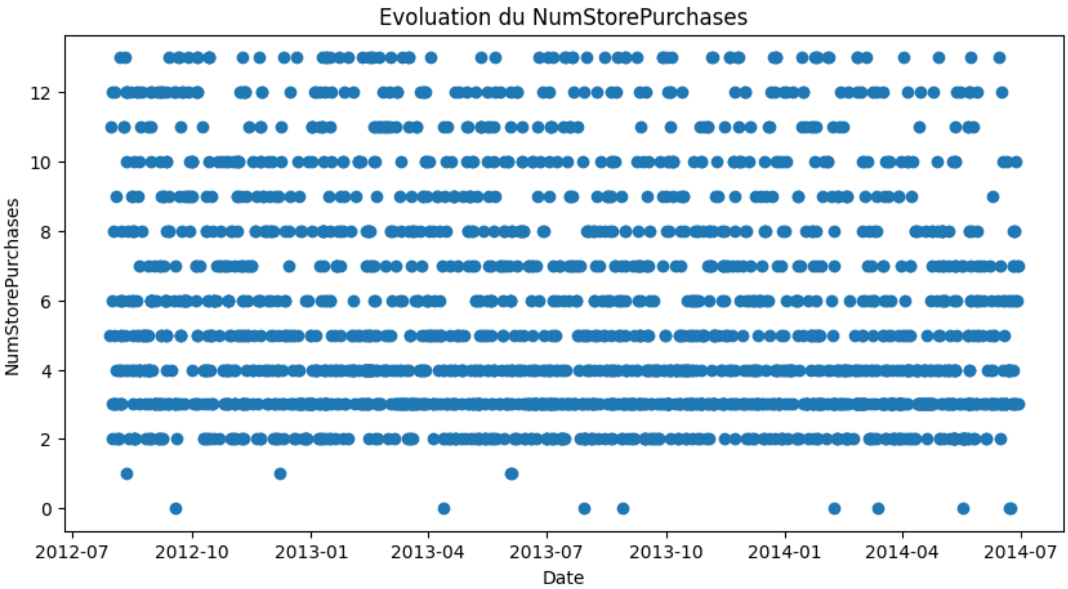

Evoluation du MntSweetProducts
MntSweetProducts
200
175
150
125
100
75
50
25
0
2012-07
2012-10
2013-01
2013-04
2013-07
2013-10
2014-01
2014-04
2014-07
Date

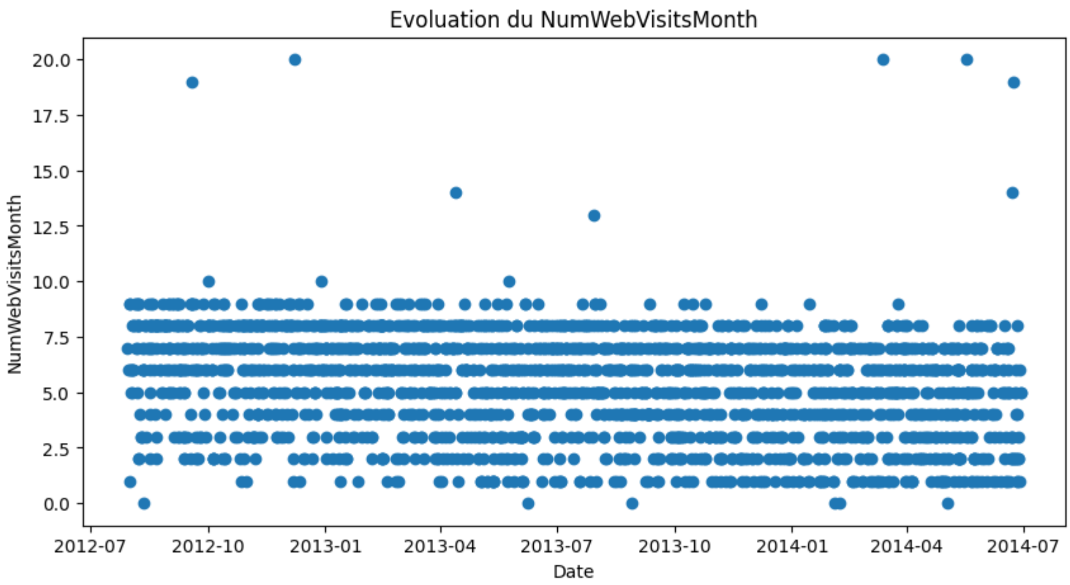

Evoluation du MntGoldProds
MntGoldProds
250
200
150
100
50
0
2012-07
2012-10
2013-01
2013-04
2013-07
2013-10
2014-01
2014-04
2014-07
Date


Evoluation du NumDealsPurchases
NumDealsPurchases
0
2
4
6
8
10
12
14
2012-07
2012-10
2013-01
2013-04
2013-07
2013-10
2014-01
2014-04
2014-07
Date

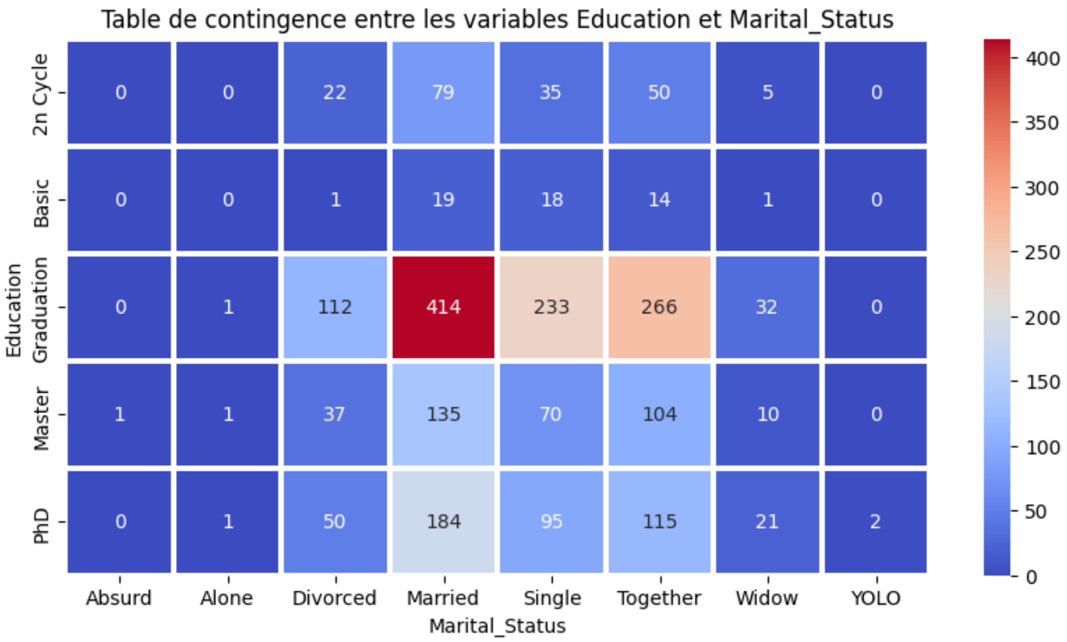

Evoluation du NumWebPurchases
NumWebPurchases
0
2
4
6
8
10
2012-07
2012-10
2013-01
2013-04
2013-07
2013-10
2014-01
2014-04
2014-07
Date


Evoluation du NumCatalogPurchases
NumCatalogPurchases
Date
2012-07
2012-10
2013-01
2013-04
2013-07
2013-10
2014-01
2014-04
2014-07

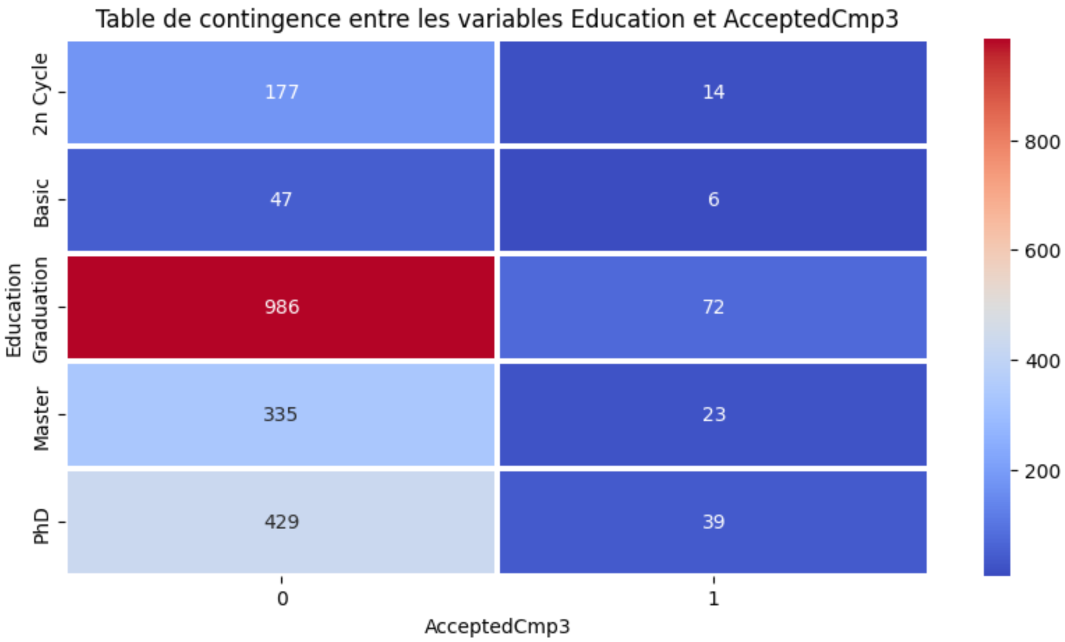

Evoluation du NumStorePurchases
NumStorePurchases
Date
2012-07
2012-10
2013-01
2013-04
2013-07
2013-10
2014-01
2014-04
2014-07


Evoluation du NumWebVisitsMonth
NumWebVisitsMonth
Date
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
2012-07
2012-10
2013-01
2013-04
2013-07
2013-10
2014-01
2014-04
2014-07

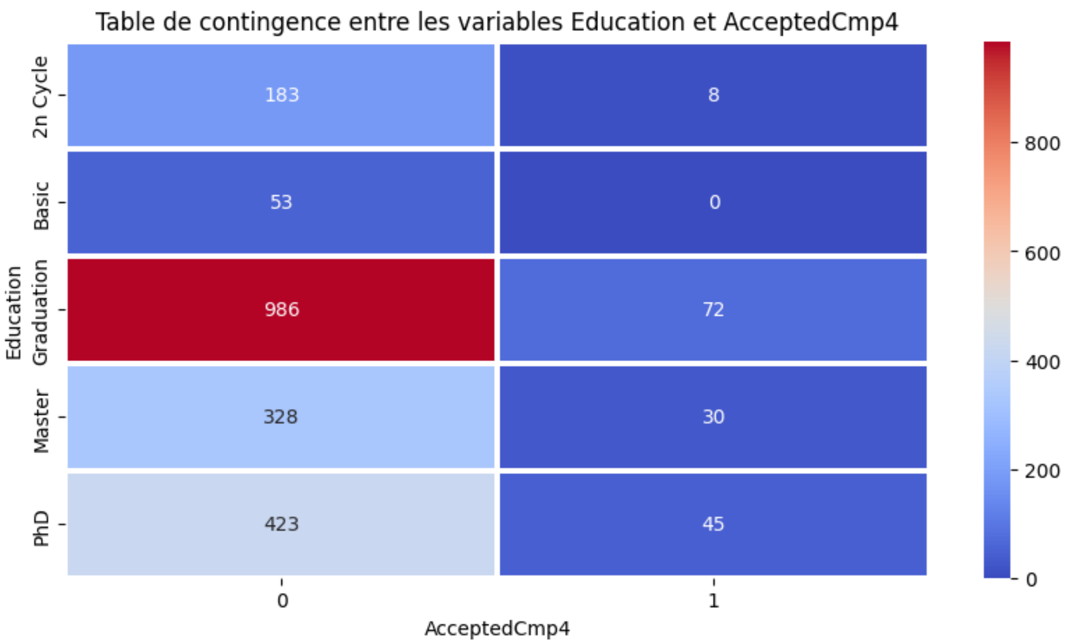
```
\Analyse de la relation entre les variables qualitatives et qualitatives

Table de contingence entre les variables Education et Marital_Status
 Marital_Status  Absurd  Alone  Divorced  Married  Single  Together  Widow  \
Education
2n Cycle              0      0        22       79      35        50      5
Basic                 0      0         1       19      18        14      1
Graduation            0      1       112      414     233       266     32
Master                1      1        37      135      70       104     10
PhD                   0      1        50      184      95       115     21

Marital_Status  YOLO
Education
2n Cycle           0
Basic              0
Graduation         0
Master             0
PhD                2
```



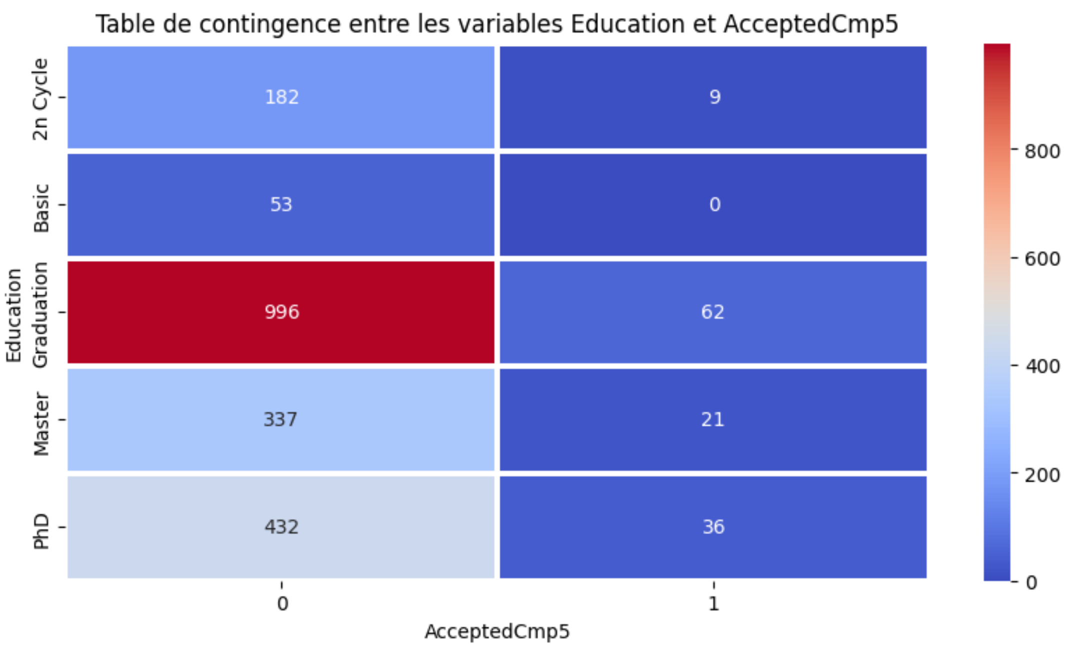
```
Table de contingence entre les variables Education et AcceptedCmp3
 AcceptedCmp3    0   1
Education
2n Cycle       177  14
Basic           47   6
Graduation     986  72
Master         335  23
PhD            429  39
```



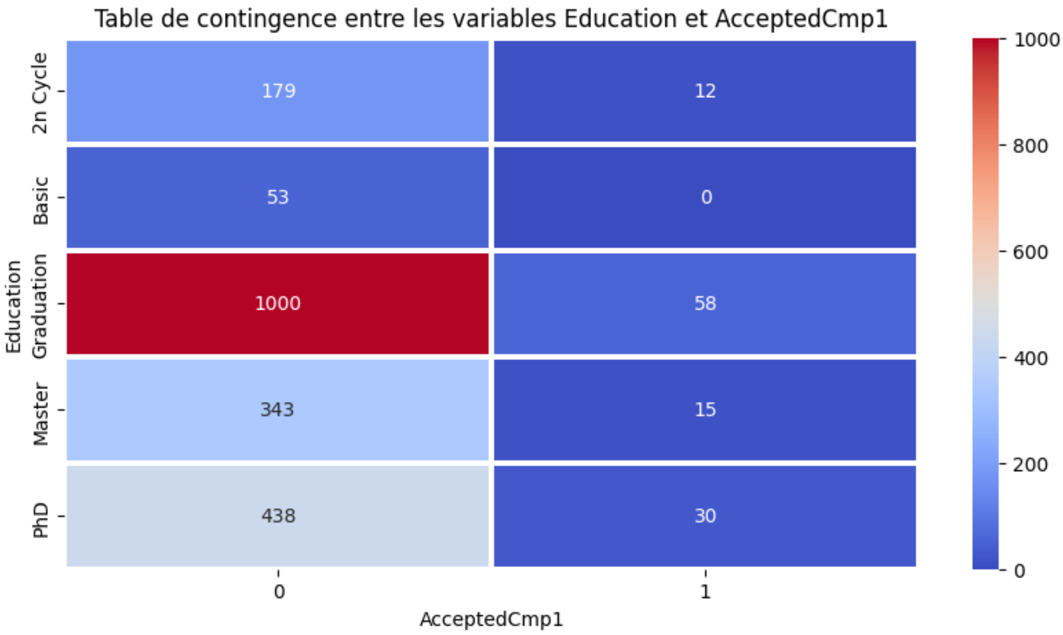

```
Table de contingence entre les variables Education et AcceptedCmp4
 AcceptedCmp4    0   1
Education
2n Cycle       183   8
Basic           53   0
Graduation     986  72
Master         328  30
PhD            423  45
```



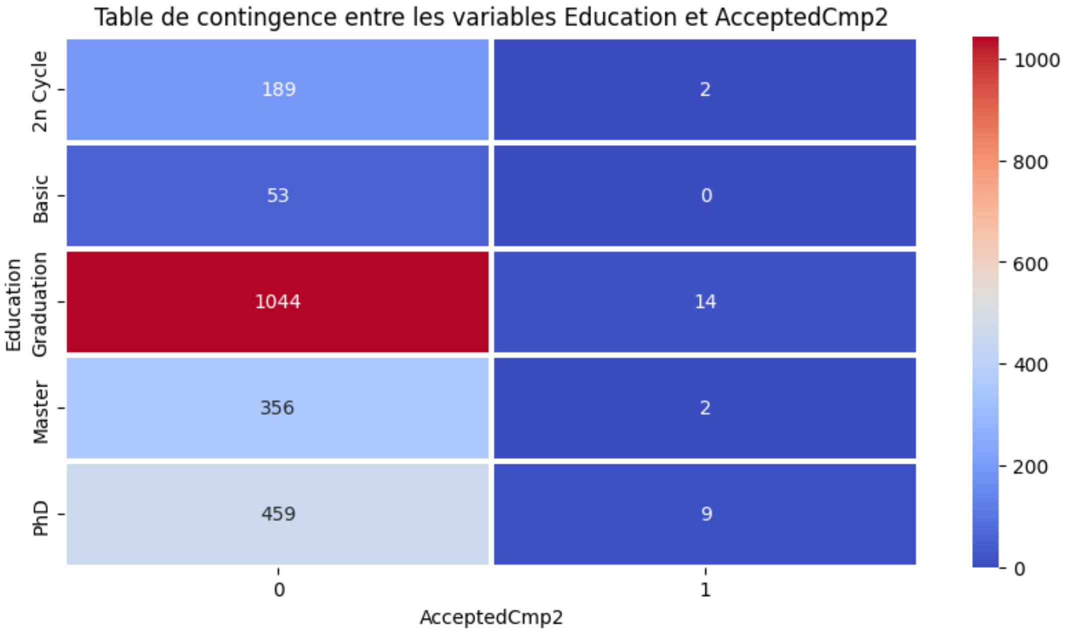

```
Table de contingence entre les variables Education et AcceptedCmp5
 AcceptedCmp5    0   1
Education
2n Cycle       182   9
Basic           53   0
Graduation     996  62
Master         337  21
PhD            432  36
```



```
Table de contingence entre les variables Education et AcceptedCmp1
 AcceptedCmp1     0   1
Education
2n Cycle        179  12
Basic            53   0
Graduation     1000  58
Master          343  15
PhD             438  30
```




```
Table de contingence entre les variables Education et AcceptedCmp2
 AcceptedCmp2     0   1
Education
2n Cycle        189   2
Basic            53   0
Graduation     1044  14
Master          356   2
PhD             459   9
```




```
Table de contingence entre les variables Education et Complain
 Complain       0   1
Education
2n Cycle      187   4
Basic          53   0
Graduation   1044  14
Master        356   2
PhD           467   1
```

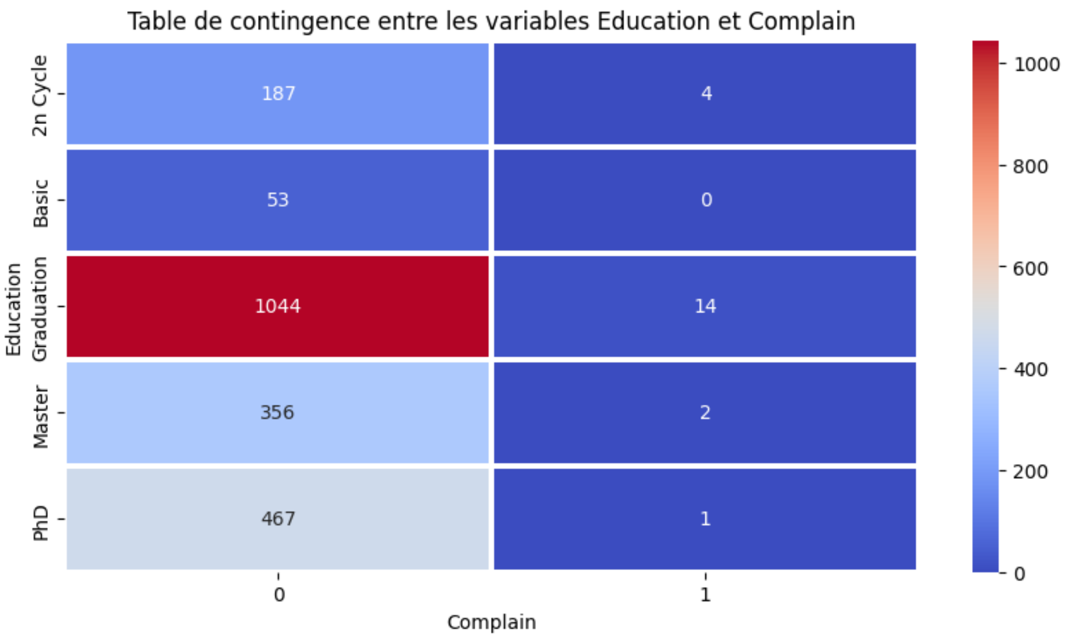


Table de contingence entre les variables Education et Response

Response	0	1
Education		
2n Cycle	171	20
Basic	51	2
Graduation	930	128
Master	308	50
PhD	370	98

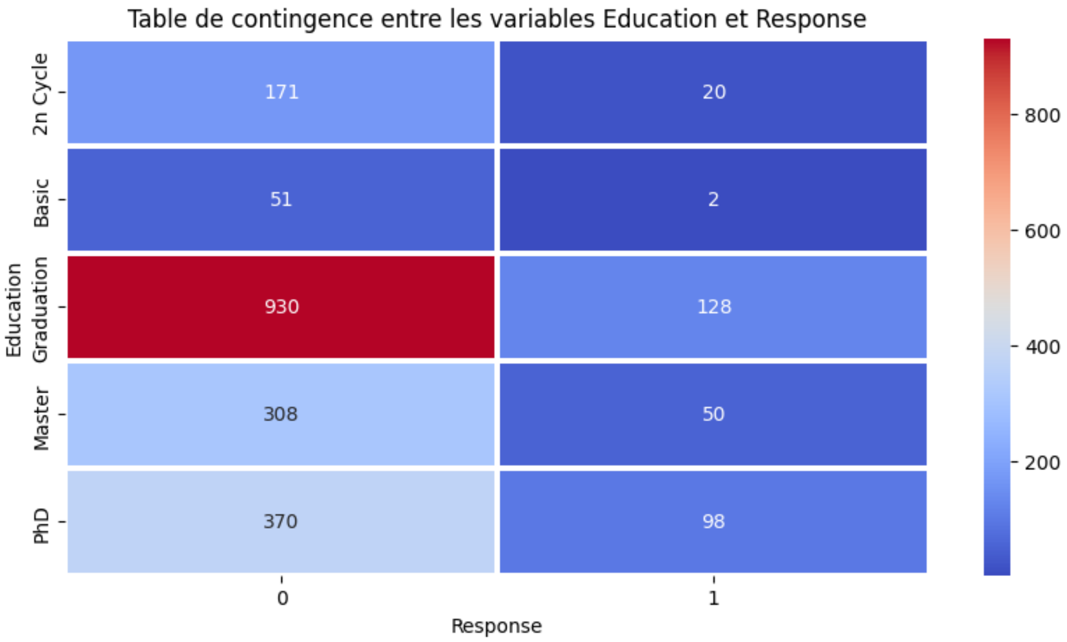
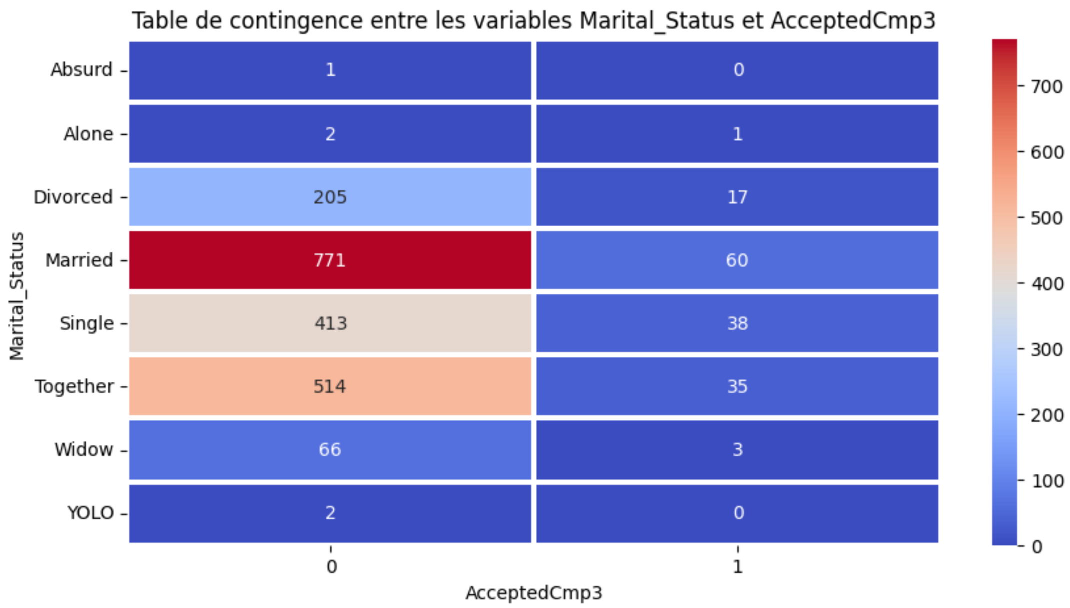
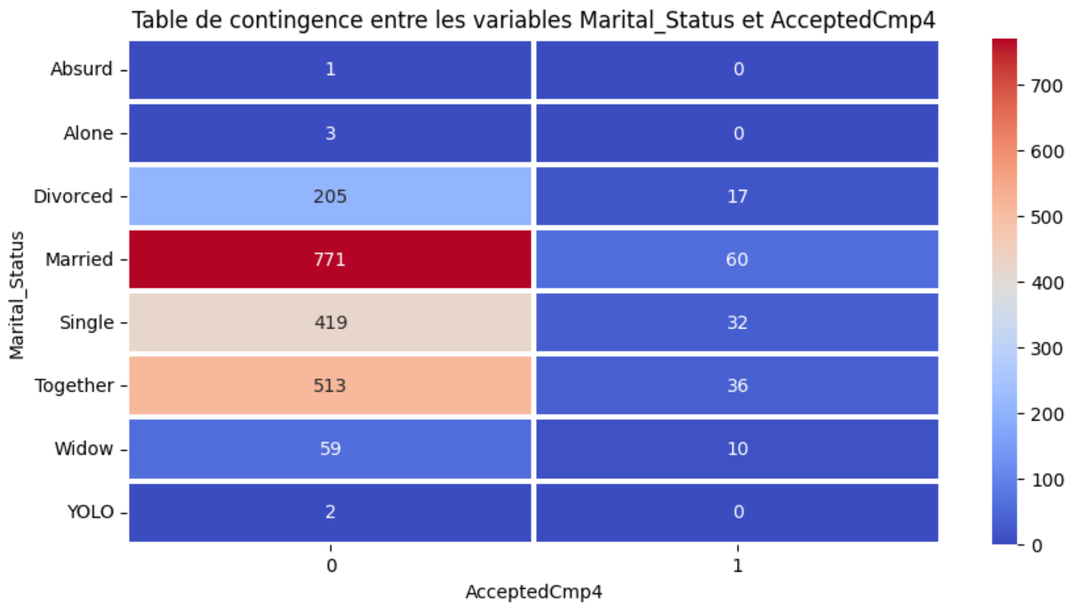


Table de contingence entre les variables Marital_Status et AcceptedCmp3

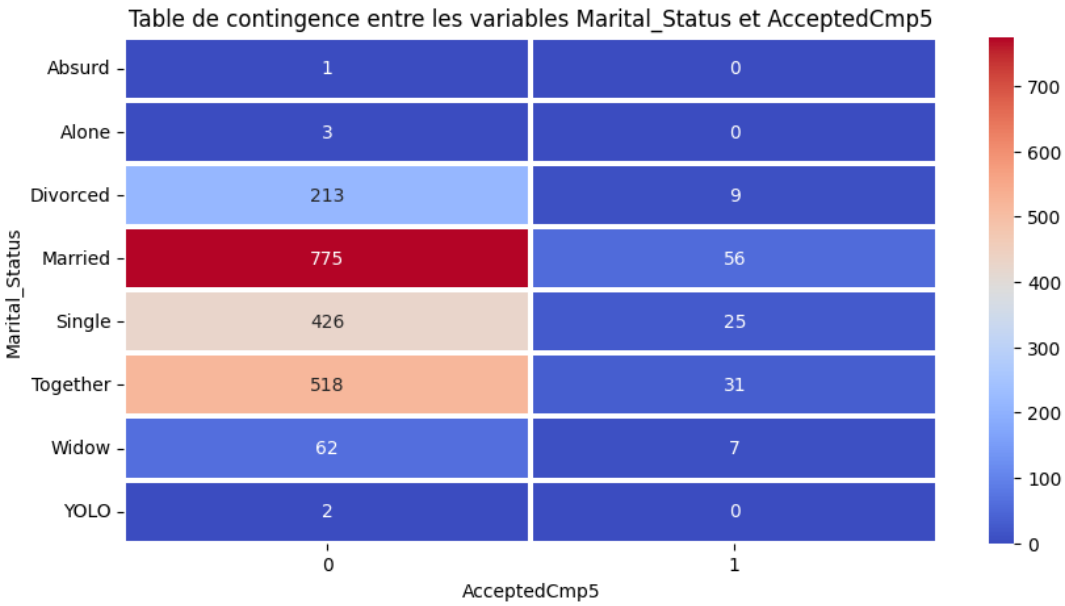
AcceptedCmp3	0	1
Marital_Status		
Absurd	1	0
Alone	2	1
Divorced	205	17
Married	771	60
Single	413	38
Together	514	35
Widow	66	3
YOLO	2	0



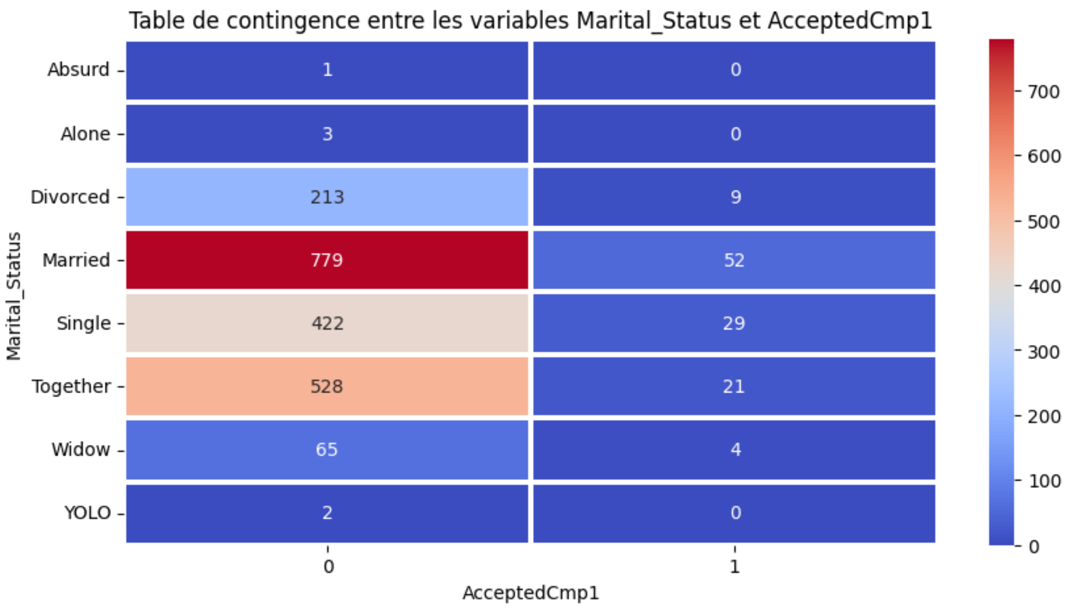

```
Table de contingence entre les variables Marital_Status et AcceptedCmp4
 AcceptedCmp4      0   1
Marital_Status
Absurd           1   0
Alone            3   0
Divorced       205  17
Married        771  60
Single         419  32
Together       513  36
Widow           59  10
YOLO             2   0
```



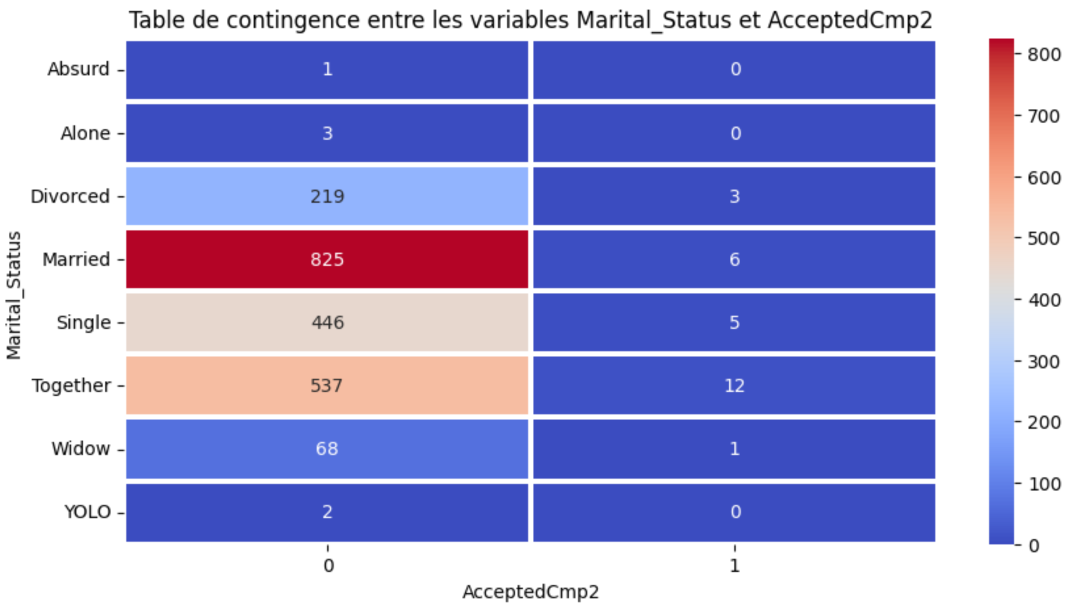
```
Table de contingence entre les variables Marital_Status et AcceptedCmp5
 AcceptedCmp5      0   1
Marital_Status
Absurd            1   0
Alone             3   0
Divorced        213   9
Married         775  56
Single          426  25
Together        518  31
Widow            62   7
YOLO              2   0
```



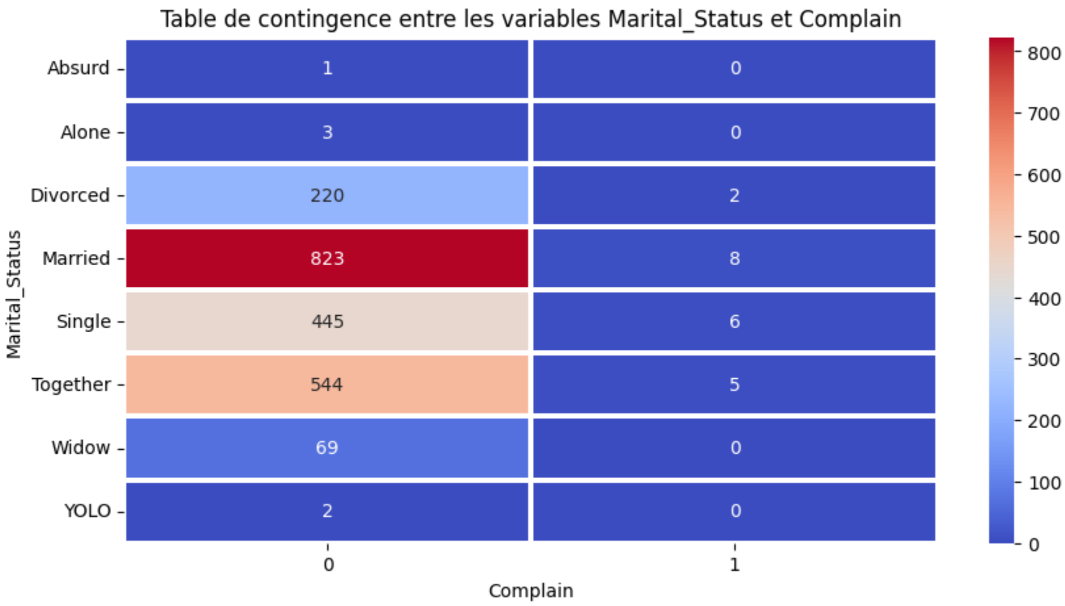

```
Table de contingence entre les variables Marital_Status et AcceptedCmp1
 AcceptedCmp1      0   1
Marital_Status
Absurd            1   0
Alone             3   0
Divorced        213   9
Married         779  52
Single          422  29
Together        528  21
Widow            65   4
YOLO              2   0
```



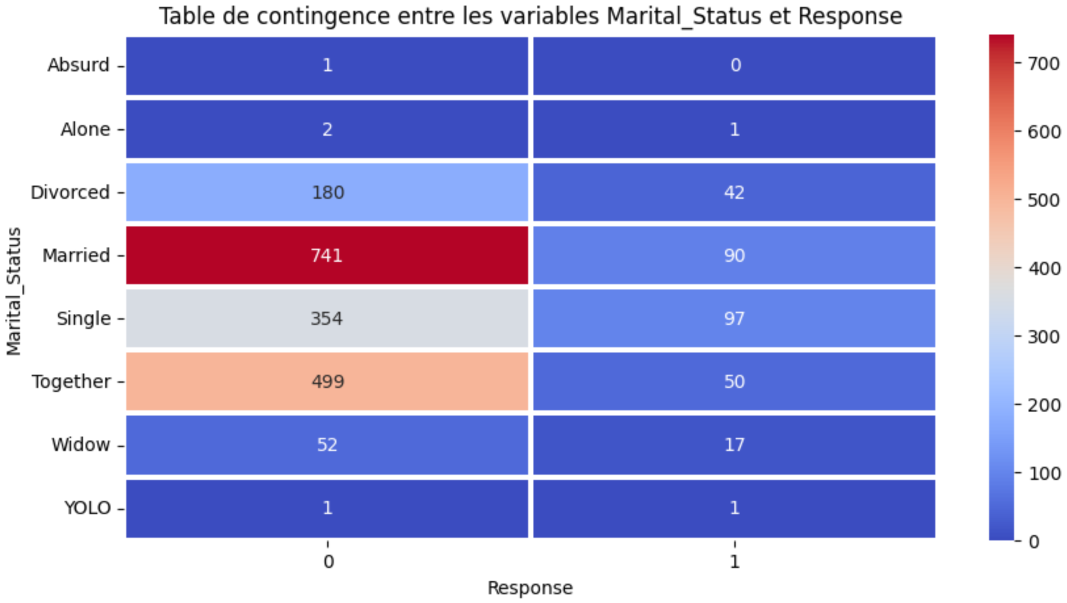

```
Table de contingence entre les variables Marital_Status et AcceptedCmp2
 AcceptedCmp2      0    1
Marital_Status
Absurd             1    0
Alone              3    0
Divorced         219    3
Married          825    6
Single           446    5
Together         537   12
Widow             68    1
YOLO               2    0
```



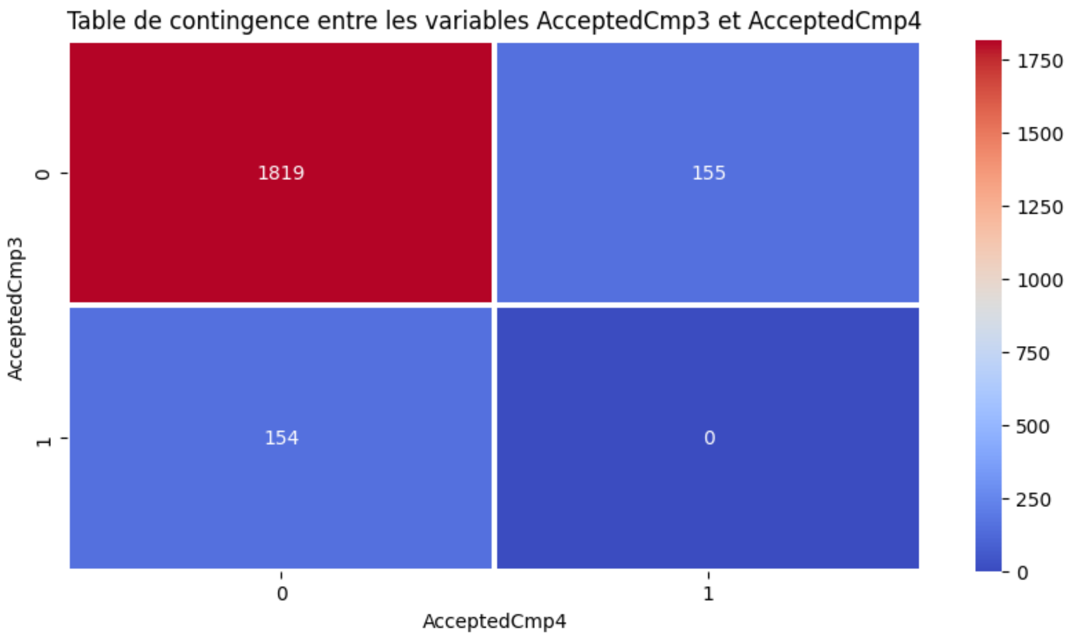
```
Table de contingence entre les variables Marital_Status et Complain
 Complain          0  1
Marital_Status
Absurd           1  0
Alone            3  0
Divorced       220  2
Married        823  8
Single         445  6
Together       544  5
Widow           69  0
YOLO             2  0
```




```
Table de contingence entre les variables Marital_Status et Response
 Response          0   1
Marital_Status
Absurd           1   0
Alone            2   1
Divorced       180  42
Married        741  90
Single         354  97
Together       499  50
Widow           52  17
YOLO             1   1
```

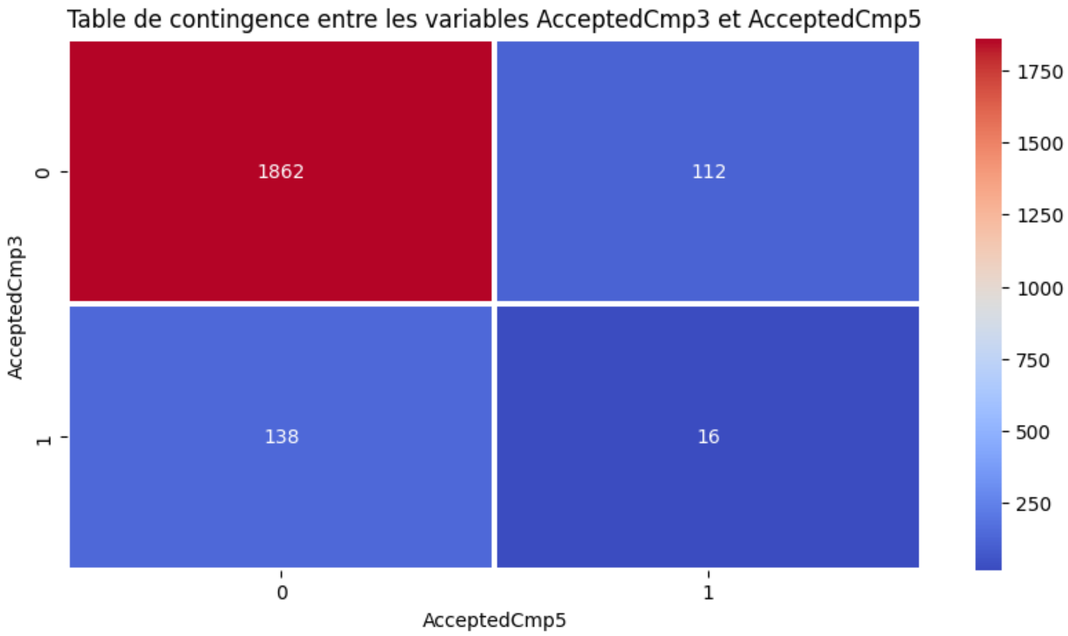



```
Table de contingence entre les variables AcceptedCmp3 et AcceptedCmp4
 AcceptedCmp4     0    1
AcceptedCmp3
0              1819  155
1               154    0
```

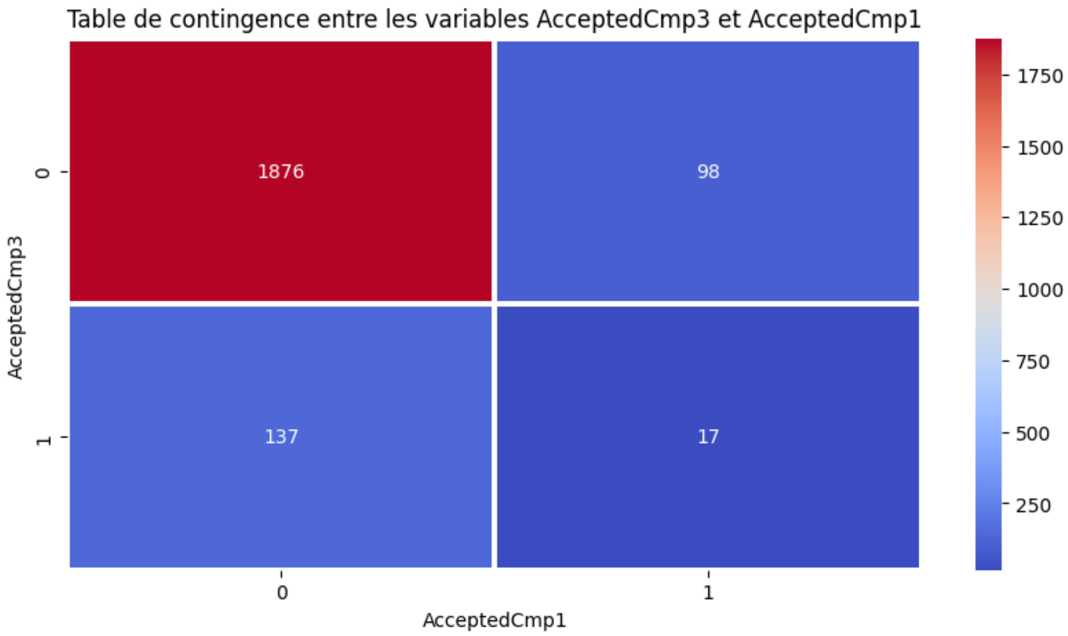



```
Table de contingence entre les variables AcceptedCmp3 et AcceptedCmp5
 AcceptedCmp5     0    1
AcceptedCmp3
0              1862  112
```

1   138   16

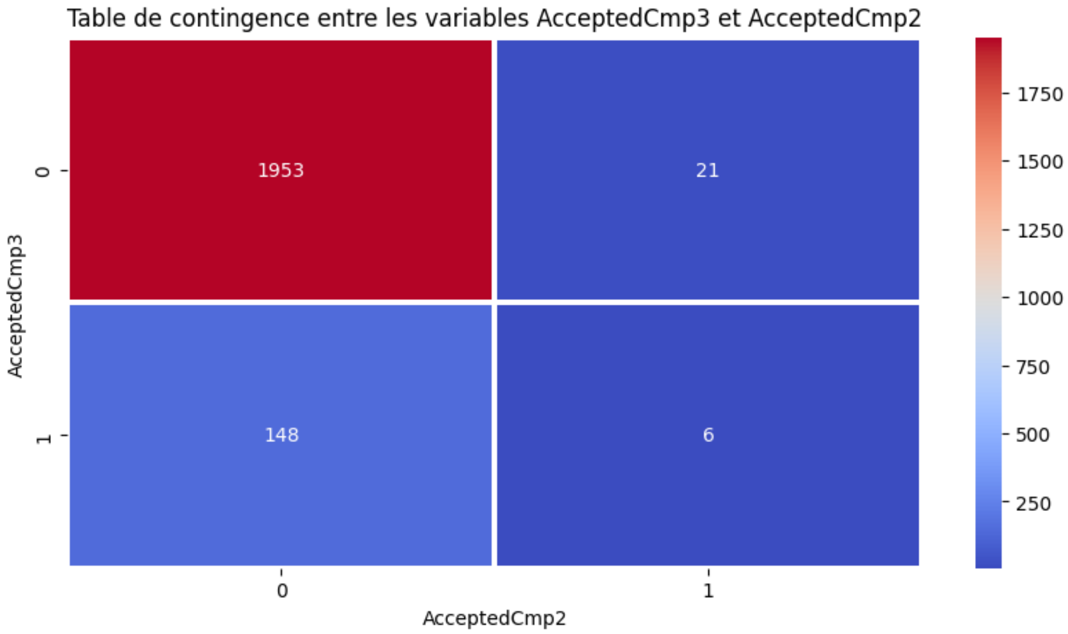



```
Table de contingence entre les variables AcceptedCmp3 et AcceptedCmp1
 AcceptedCmp1     0   1
AcceptedCmp3
0               1876  98
1                137  17
```

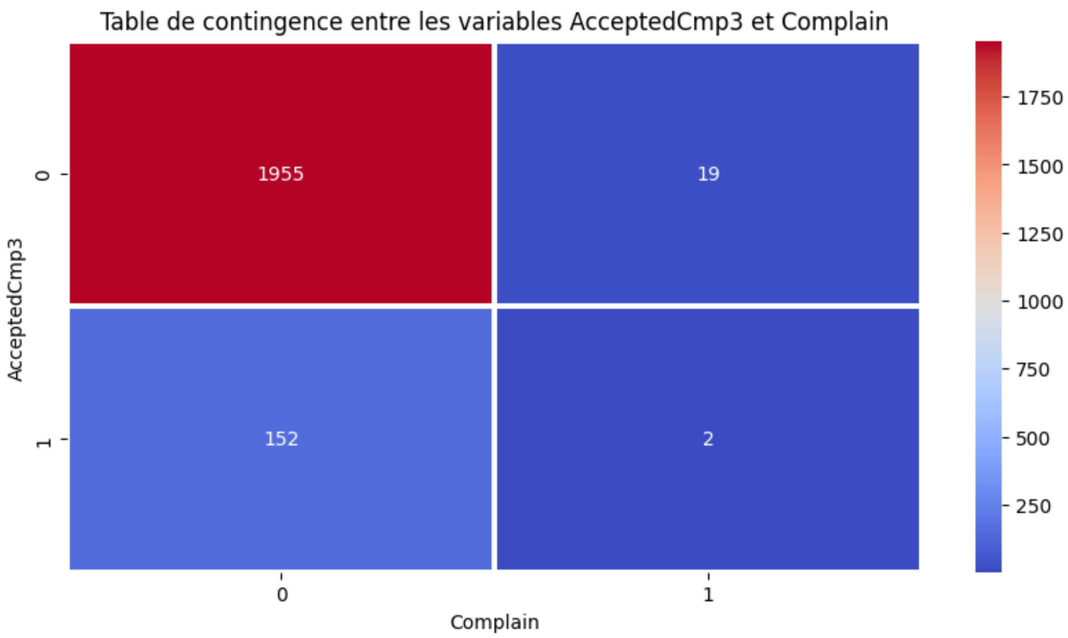



```
Table de contingence entre les variables AcceptedCmp3 et AcceptedCmp2
 AcceptedCmp2     0   1
```

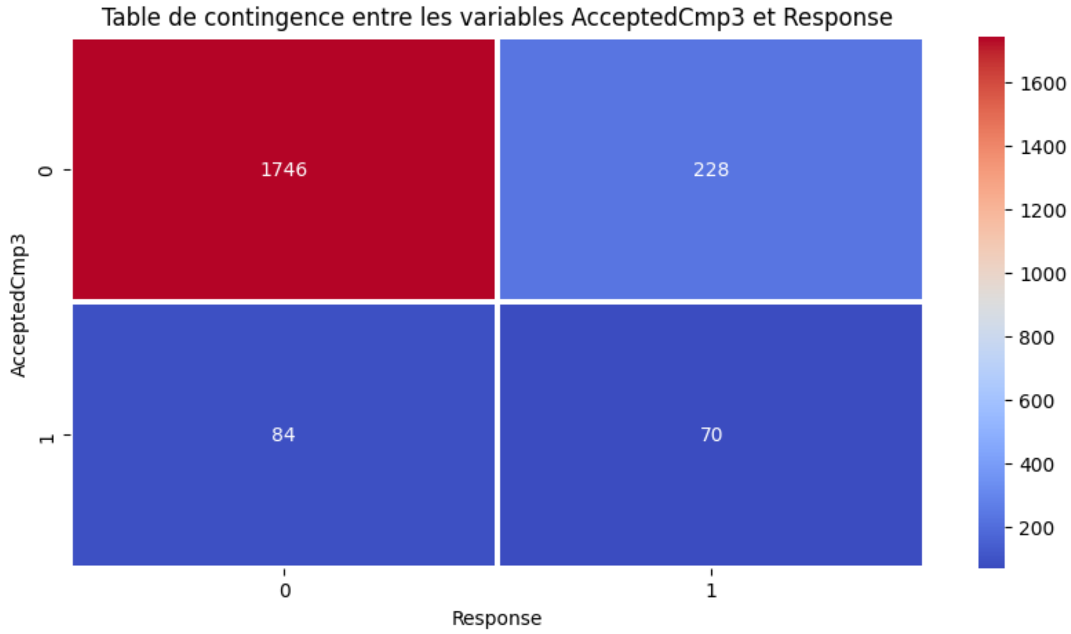
```
AcceptedCmp3
0               1953   21
1                148    6
```



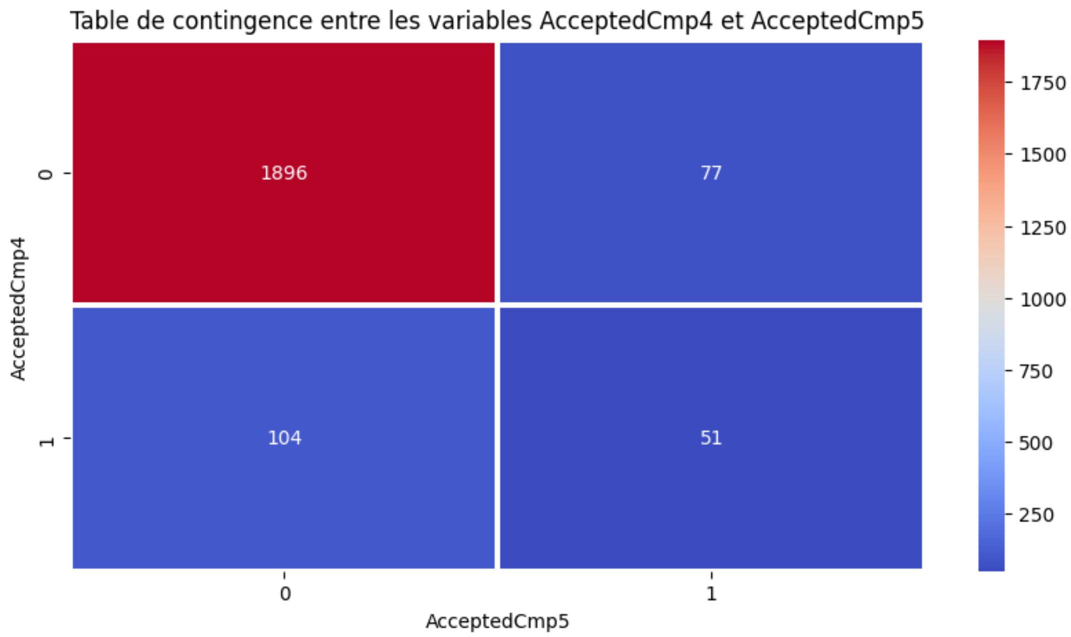

```
Table de contingence entre les variables AcceptedCmp3 et Complain
 Complain          0    1
AcceptedCmp3
0               1955   19
1                152    2
```



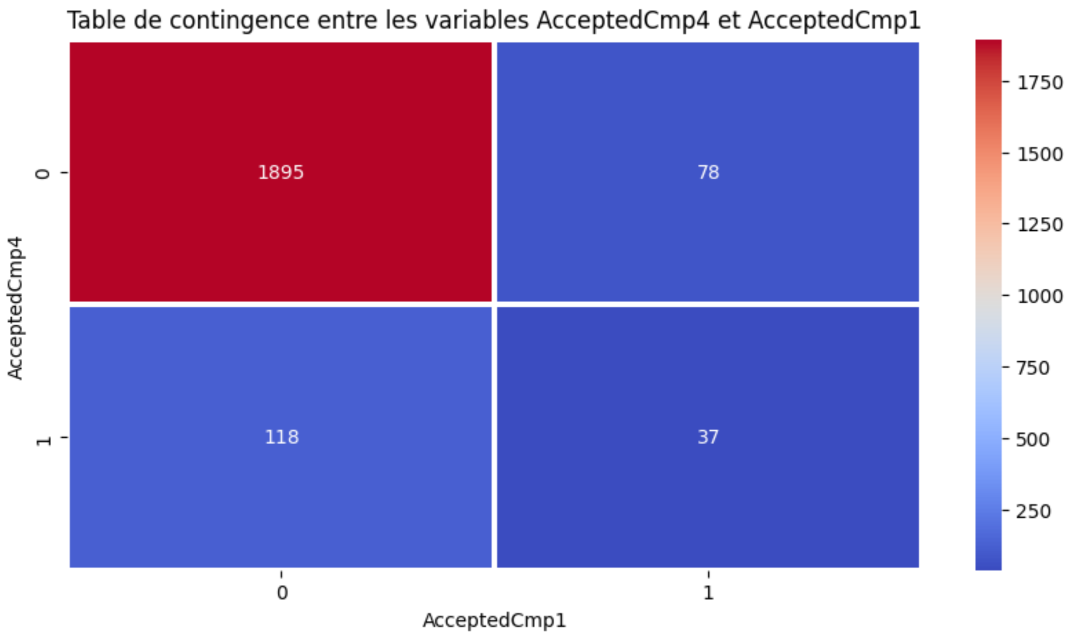
```
Table de contingence entre les variables AcceptedCmp3 et Response
 Response          0    1
AcceptedCmp3
0               1746  228
1                 84   70
```



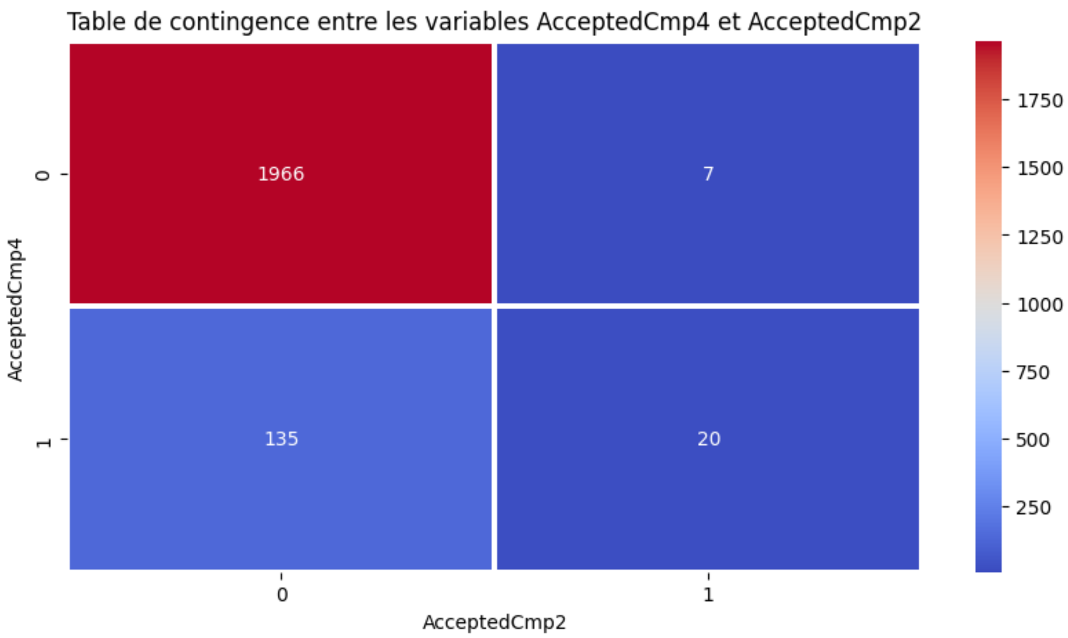

```
Table de contingence entre les variables AcceptedCmp4 et AcceptedCmp5
 AcceptedCmp5      0   1
AcceptedCmp4
0               1896  77
1                104  51
```



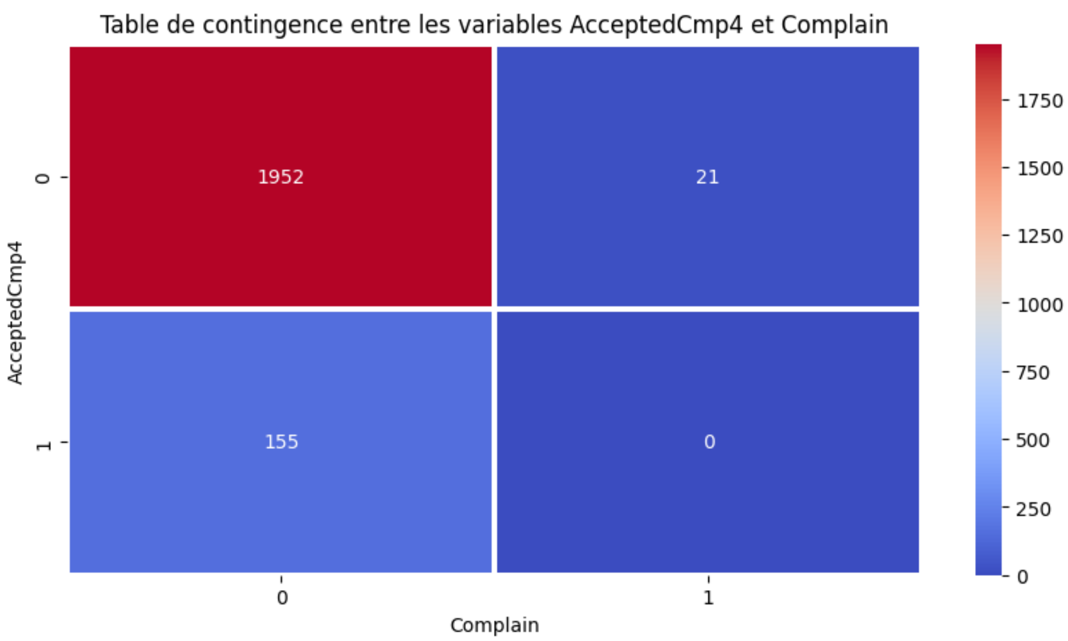
```
Table de contingence entre les variables AcceptedCmp4 et AcceptedCmp1
 AcceptedCmp1     0   1
AcceptedCmp4
0              1895  78
1               118  37
```




```
Table de contingence entre les variables AcceptedCmp4 et AcceptedCmp2
 AcceptedCmp2     0   1
AcceptedCmp4
0              1966   7
1               135  20
```

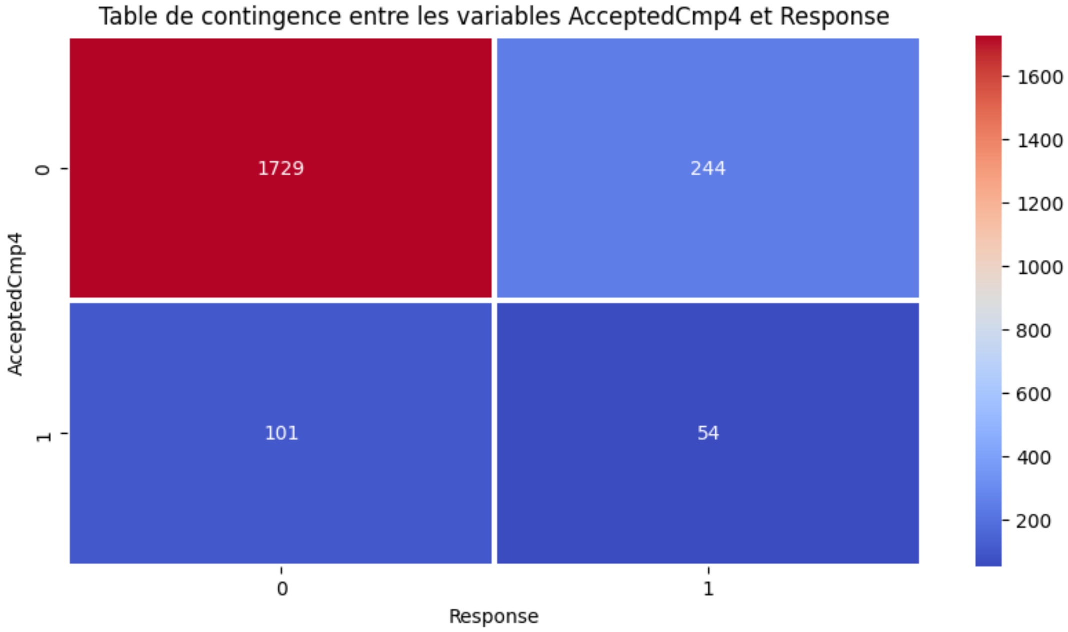



```
Table de contingence entre les variables AcceptedCmp4 et Complain
 Complain          0   1
AcceptedCmp4
0               1952  21
1                155   0
```

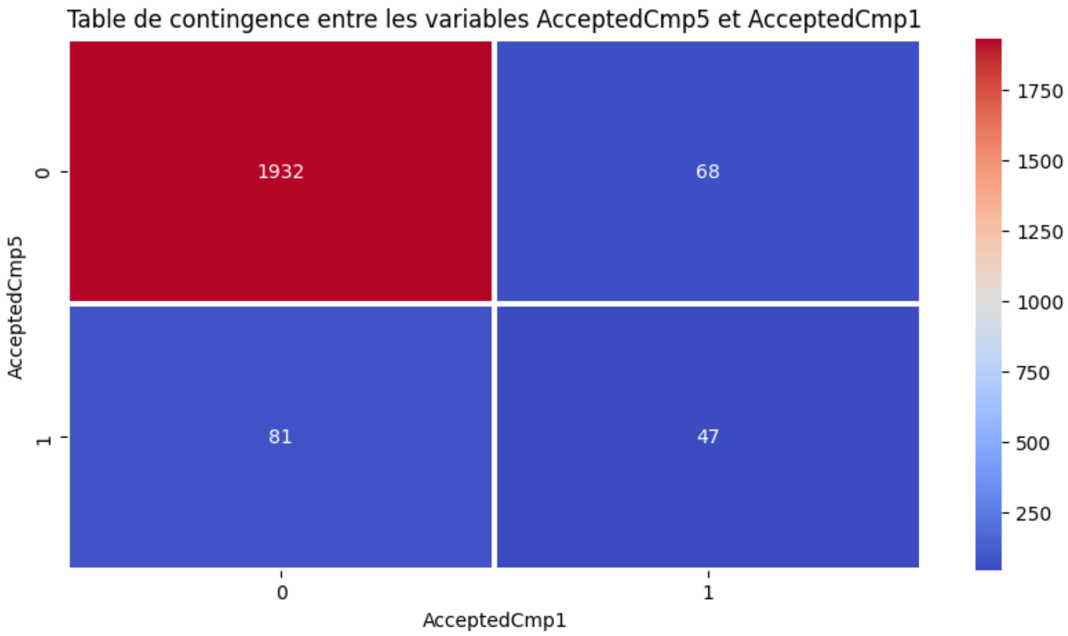



```
Table de contingence entre les variables AcceptedCmp4 et Response
 Response          0    1
AcceptedCmp4
```

```
0               1729  244
1                101   54
```

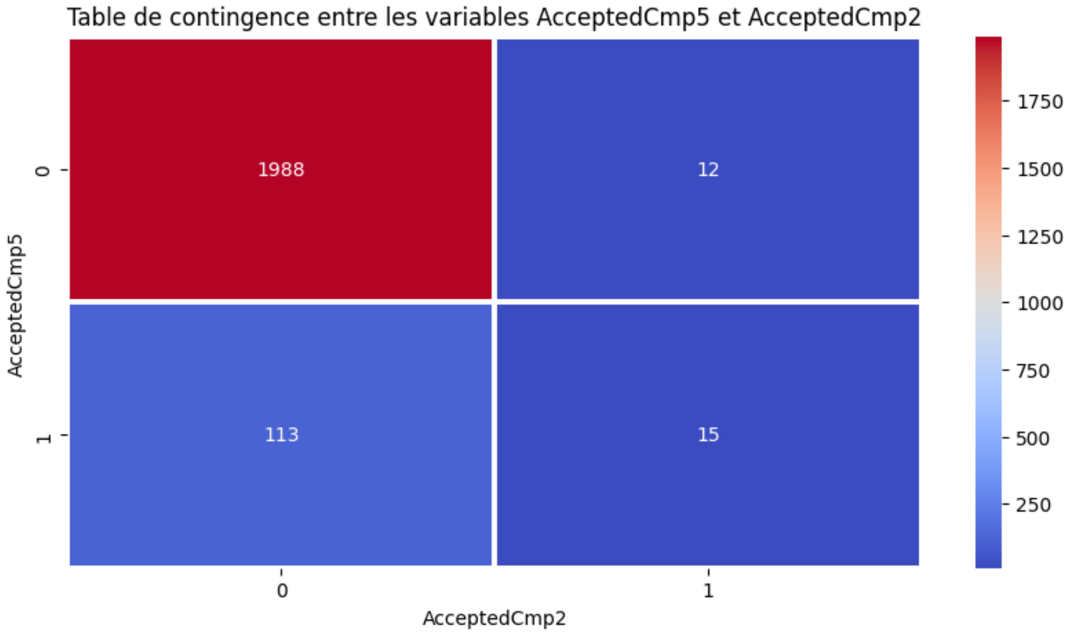



```
Table de contingence entre les variables AcceptedCmp5 et AcceptedCmp1
 AcceptedCmp1     0   1
AcceptedCmp5
0               1932  68
1                 81  47
```

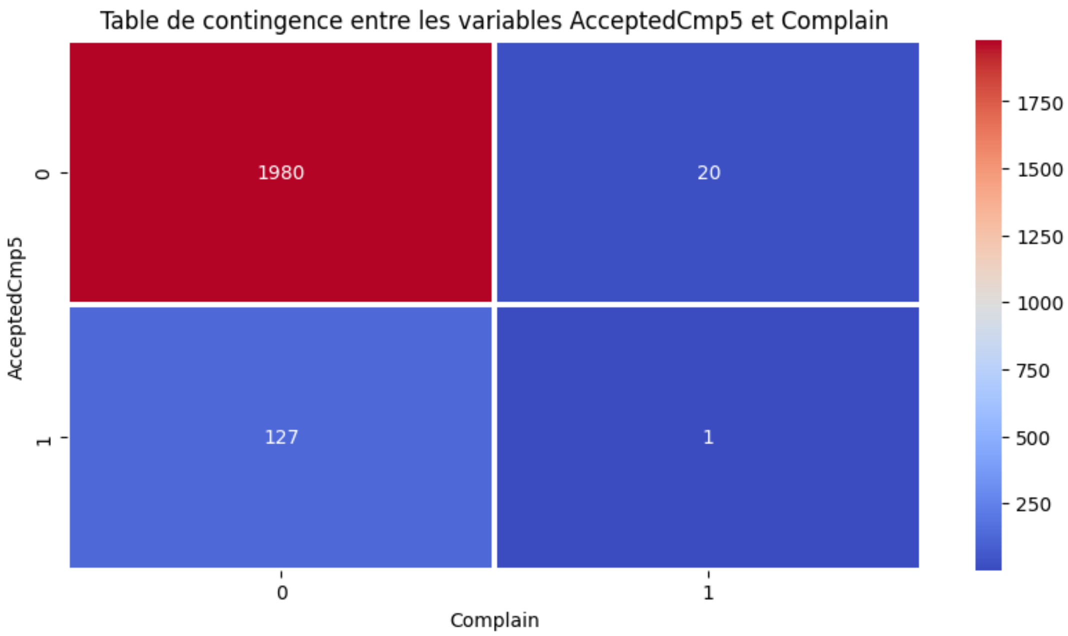



```
Table de contingence entre les variables AcceptedCmp5 et AcceptedCmp2
```

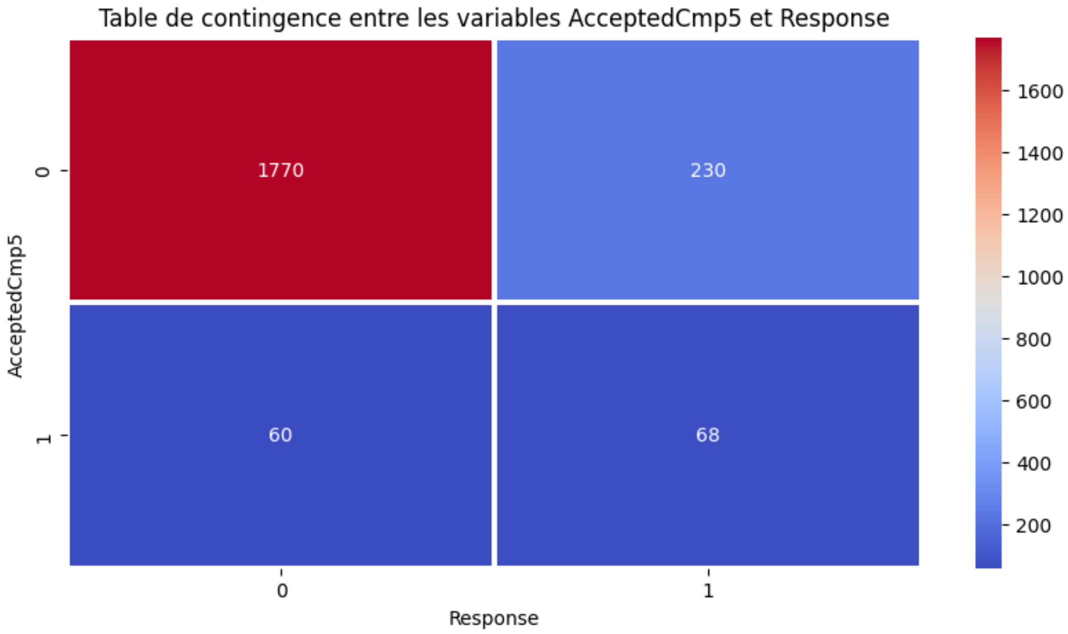
```
 AcceptedCmp2      0   1
AcceptedCmp5
0               1988  12
1                113  15
```



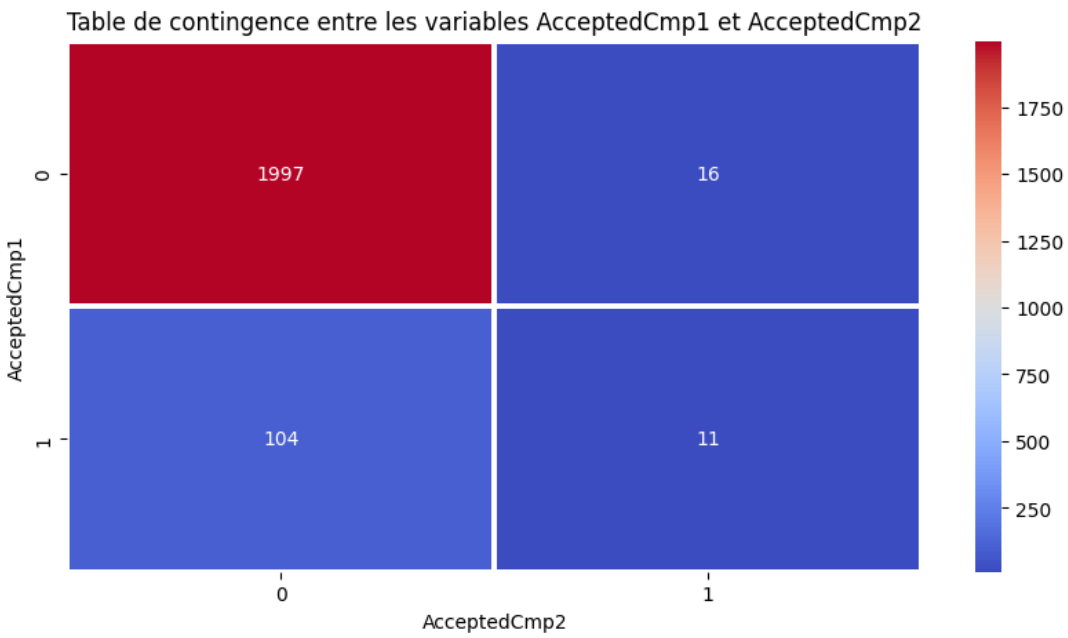

```
Table de contingence entre les variables AcceptedCmp5 et Complain
 Complain          0   1
AcceptedCmp5
0               1980  20
1                127   1
```



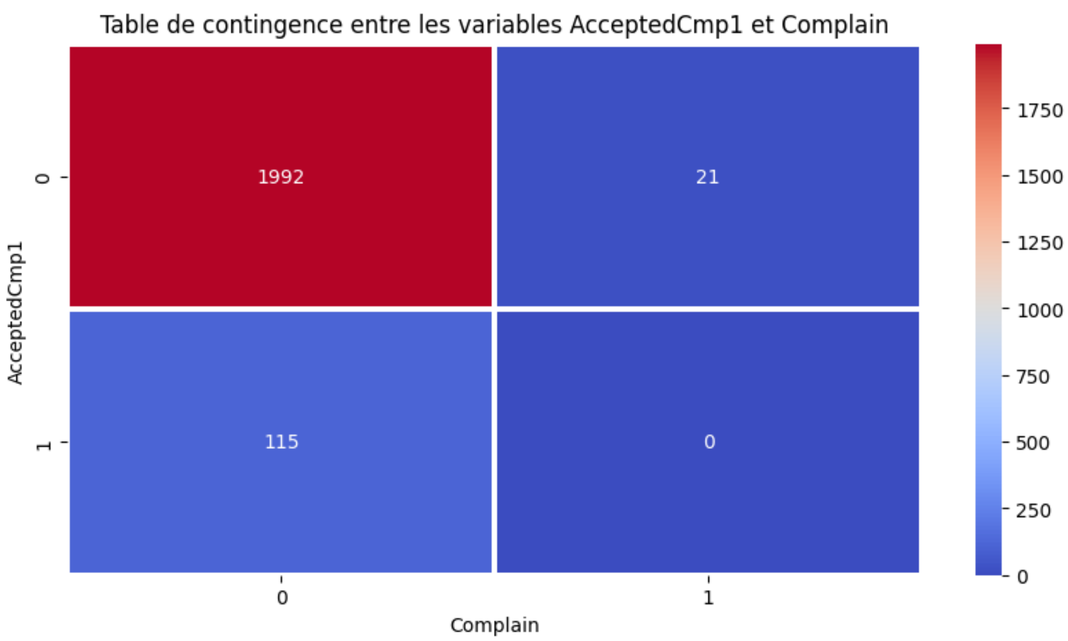
```
Table de contingence entre les variables AcceptedCmp5 et Response
 Response          0    1
AcceptedCmp5
0               1770  230
1                 60   68
```




```
Table de contingence entre les variables AcceptedCmp1 et AcceptedCmp2
 AcceptedCmp2      0   1
AcceptedCmp1
0               1997  16
1                104  11
```

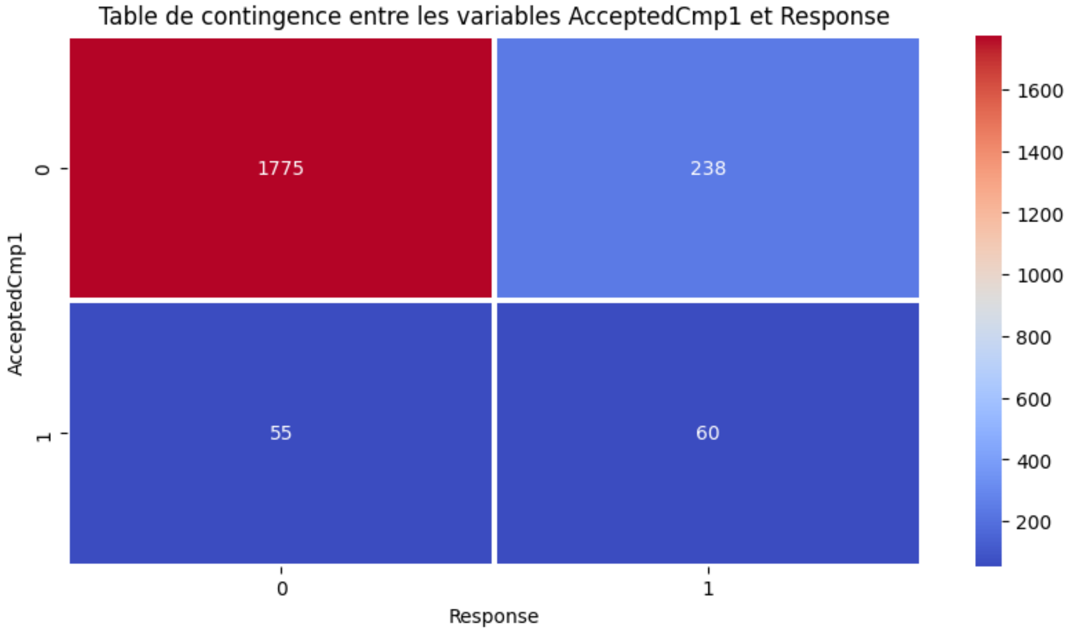



```
Table de contingence entre les variables AcceptedCmp1 et Complain
 Complain          0   1
AcceptedCmp1
0               1992  21
1                115   0
```

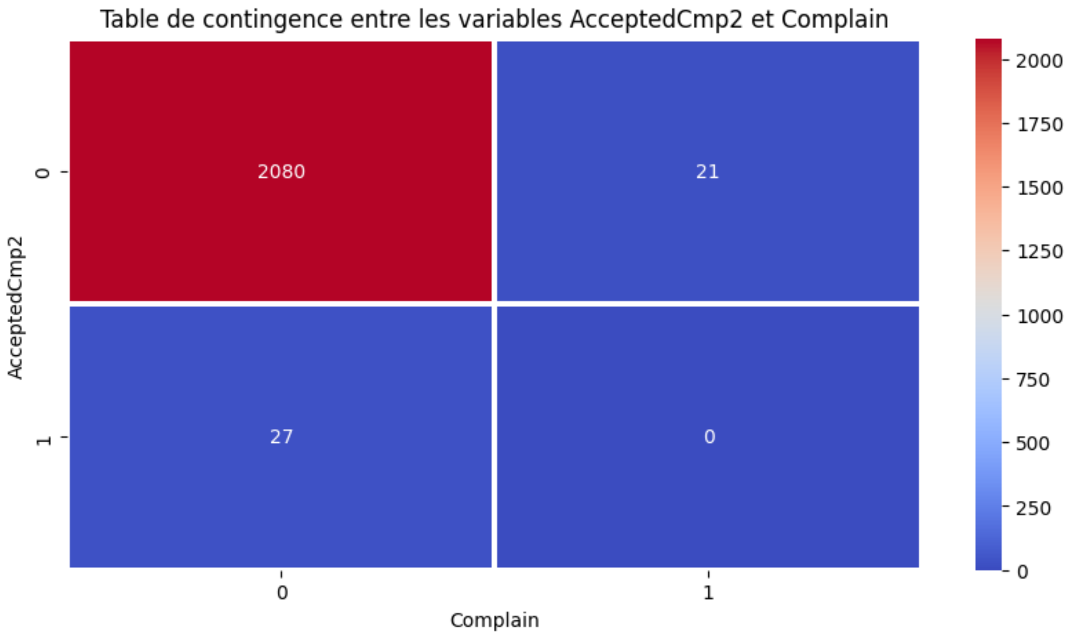



```
Table de contingence entre les variables AcceptedCmp1 et Response
 Response          0    1
AcceptedCmp1
```

```
0                1775   238
1                  55    60
```

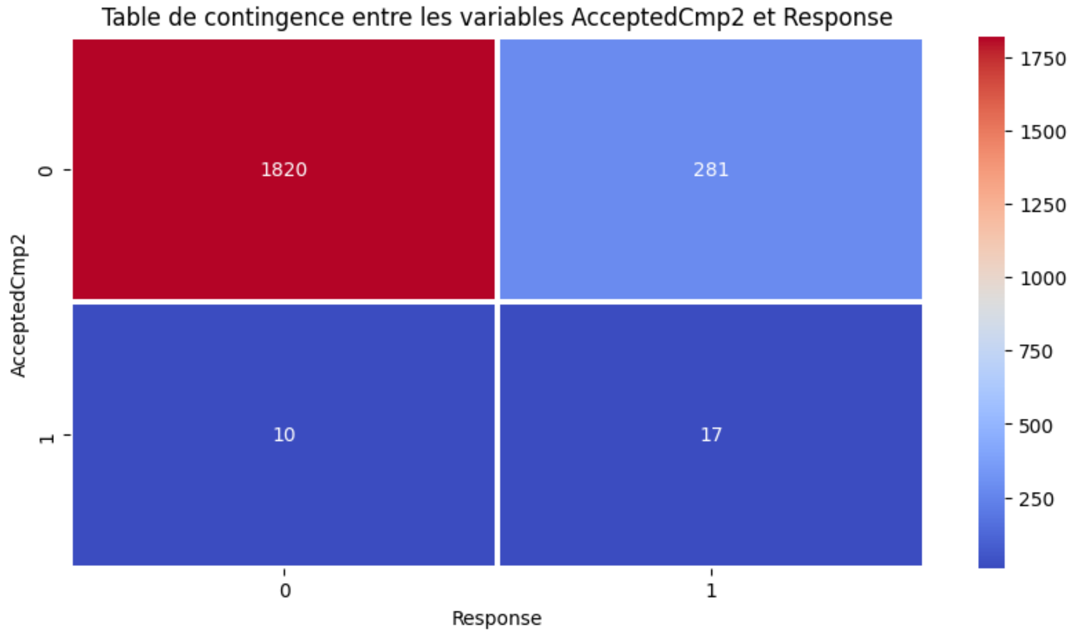



```
Table de contingence entre les variables AcceptedCmp2 et Complain
 Complain          0    1
AcceptedCmp2
0               2080   21
1                 27    0
```

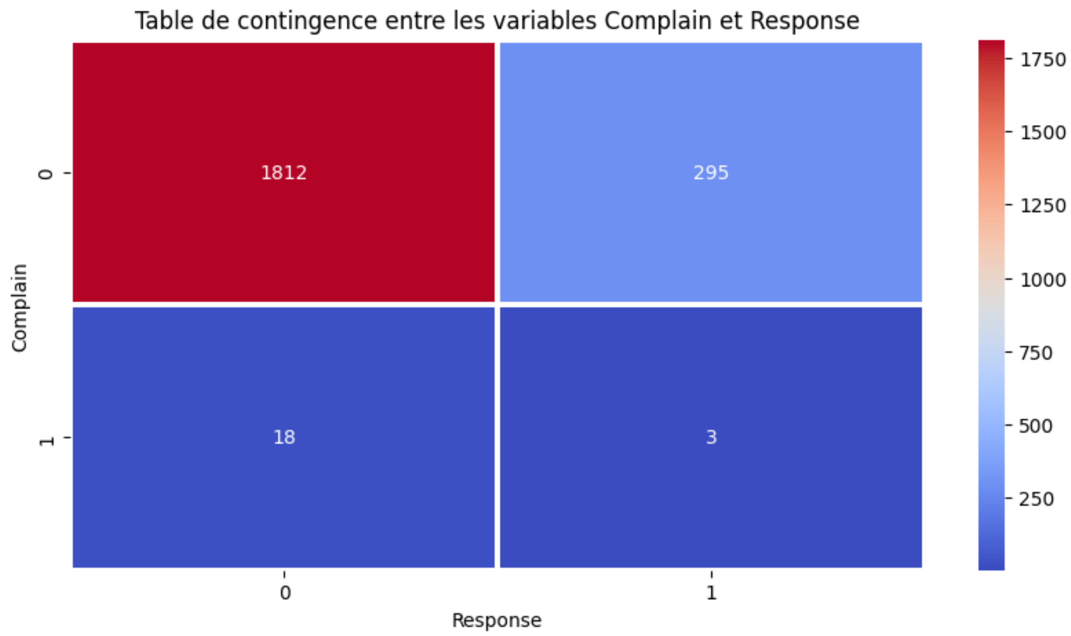



```
Table de contingence entre les variables AcceptedCmp2 et Response
```

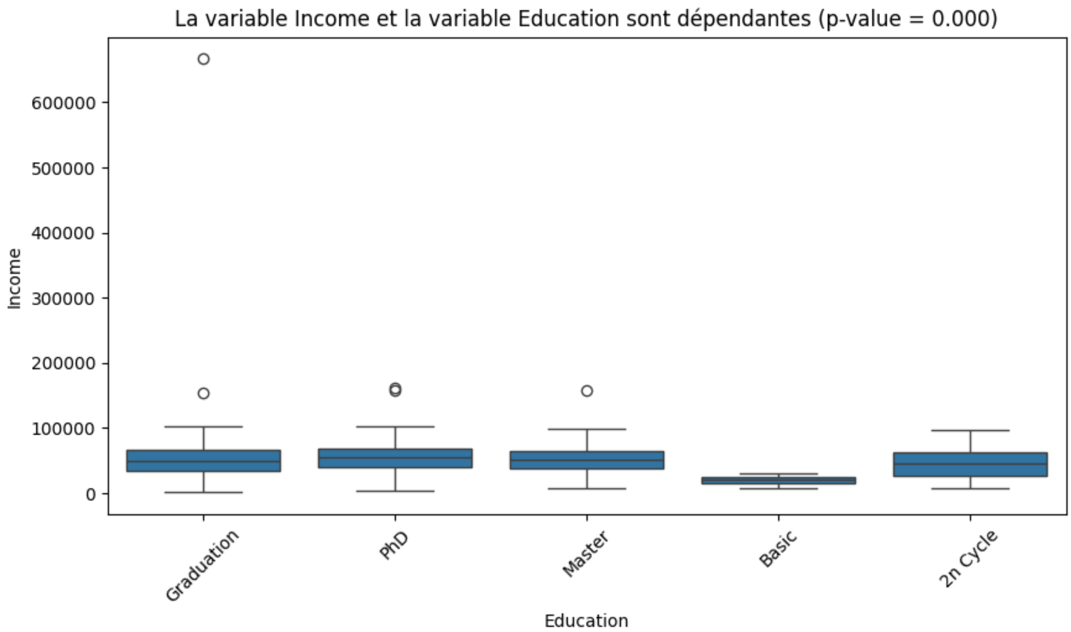
```
 Response          0    1
AcceptedCmp2
0              1820  281
1                10   17
```



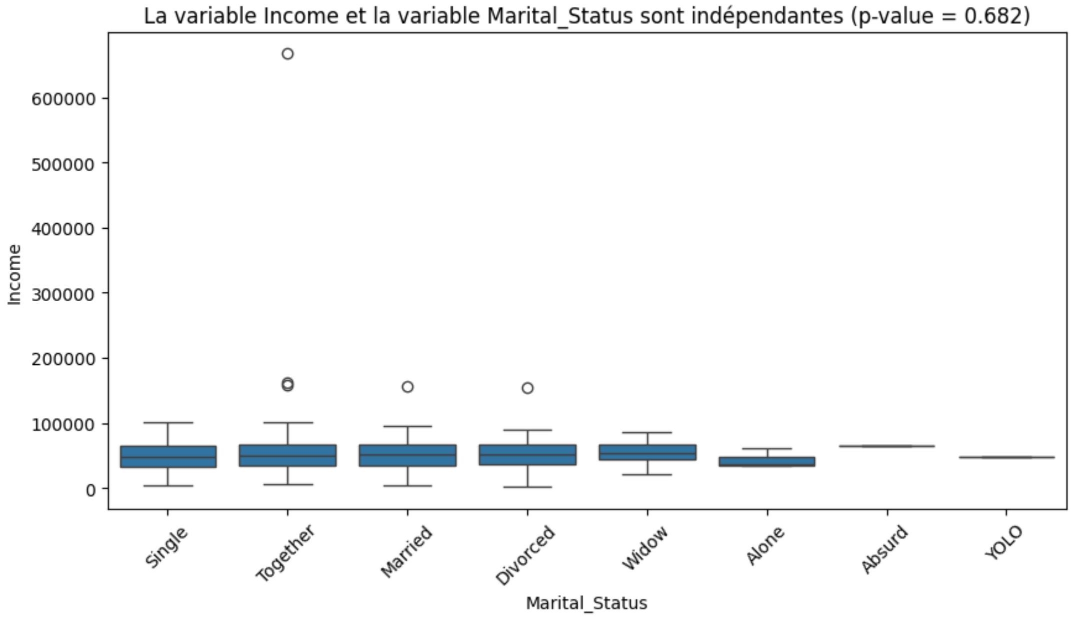

```
Table de contingence entre les variables Complain et Response
 Response      0    1
Complain
0          1812  295
1            18    3
```



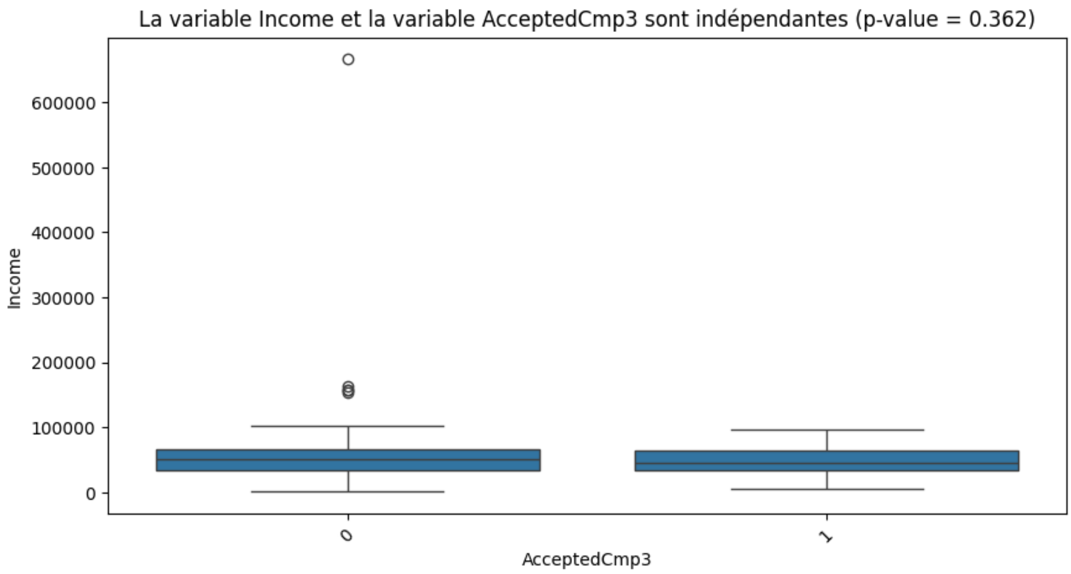
```
 La moyenne de la variable Income par rapport à la variable Education
 Education
PhD           55108.336
Master        52154.797
Graduation    51140.898
2n Cycle      46445.815
Basic         20039.491
Name: Income, dtype: float64
```




```
 La moyenne de la variable Income par rapport à la variable Marital_Status
 Marital_Status
Absurd      65487.000
Widow       55029.033
Divorced    51775.734
Together    51655.085
Married     50809.266
Single      49526.290
YOLO        48432.000
Alone       43789.000
Name: Income, dtype: float64
```

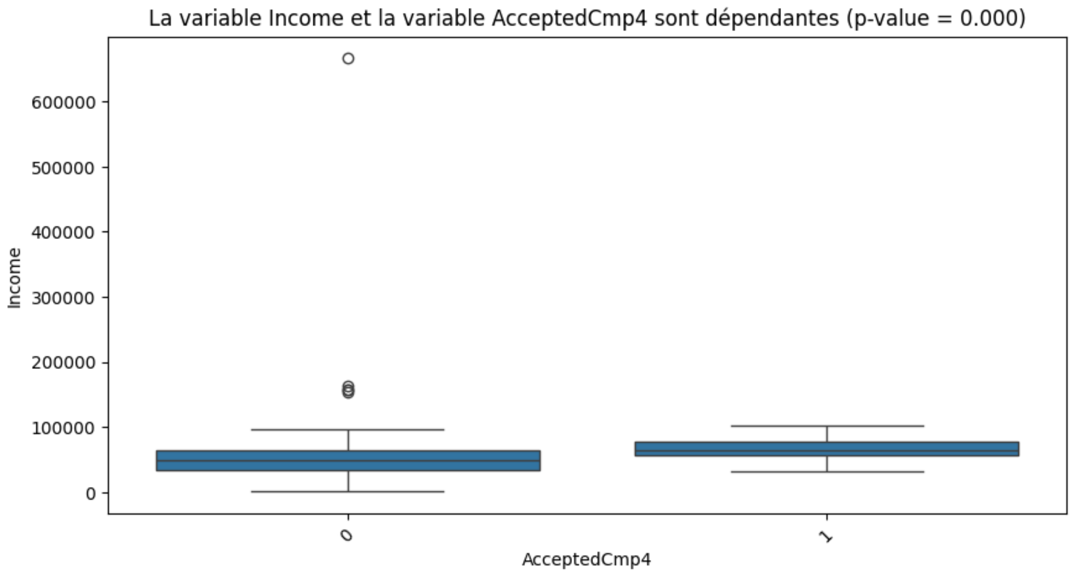



```
 La moyenne de la variable Income par rapport à la variable AcceptedCmp3
 AcceptedCmp3
0    51122.883
1    49258.844
Name: Income, dtype: float64
```

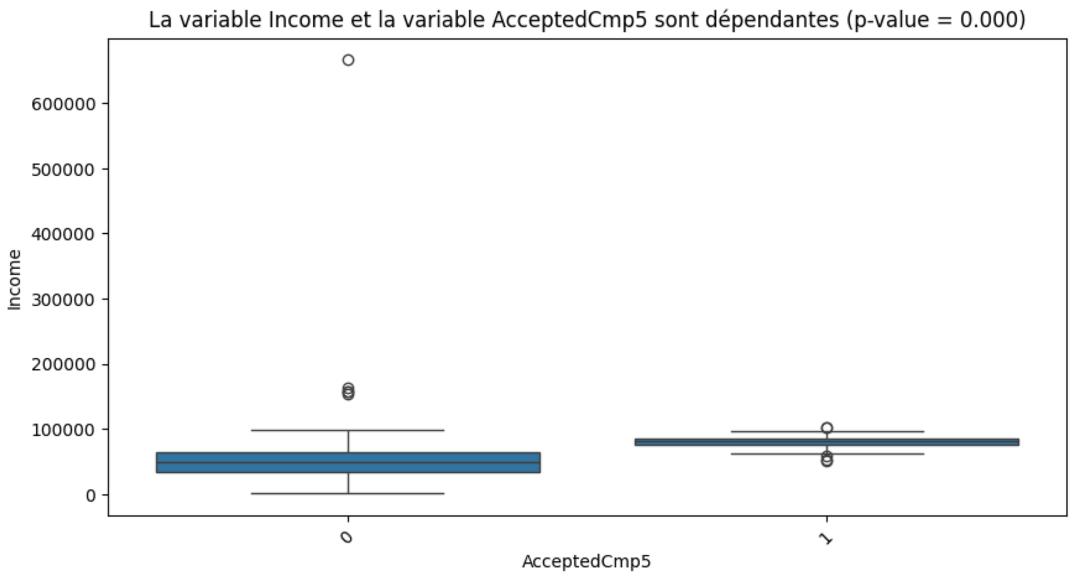



```
 La moyenne de la variable Income par rapport à la variable AcceptedCmp4
 AcceptedCmp4
1    67094.844
```

```
0    49722.621
Name: Income, dtype: float64
```

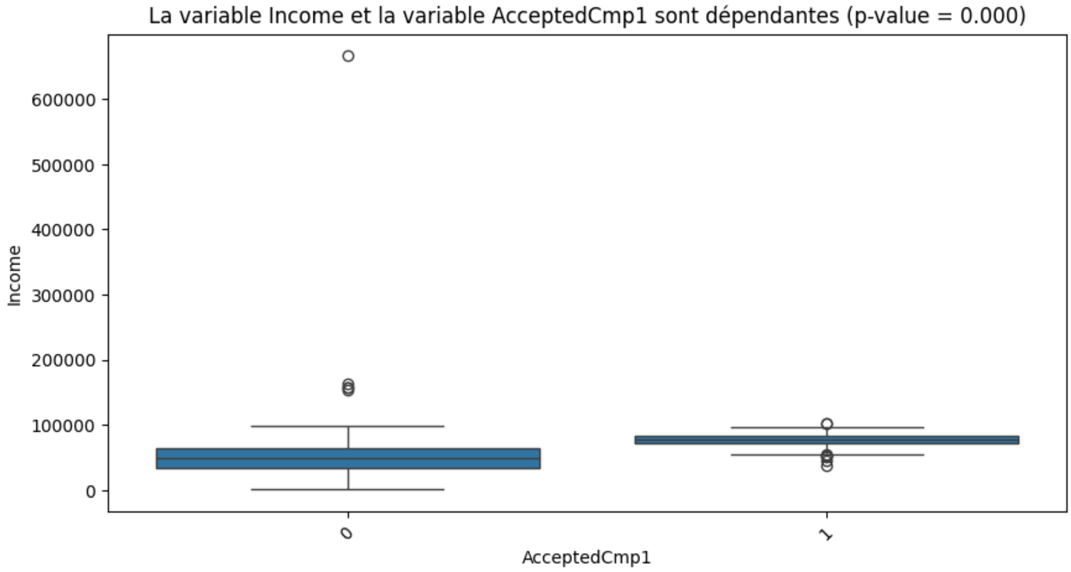



```
 La moyenne de la variable Income par rapport à la variable AcceptedCmp5
 AcceptedCmp5
1    80865.236
0    49075.841
Name: Income, dtype: float64
```

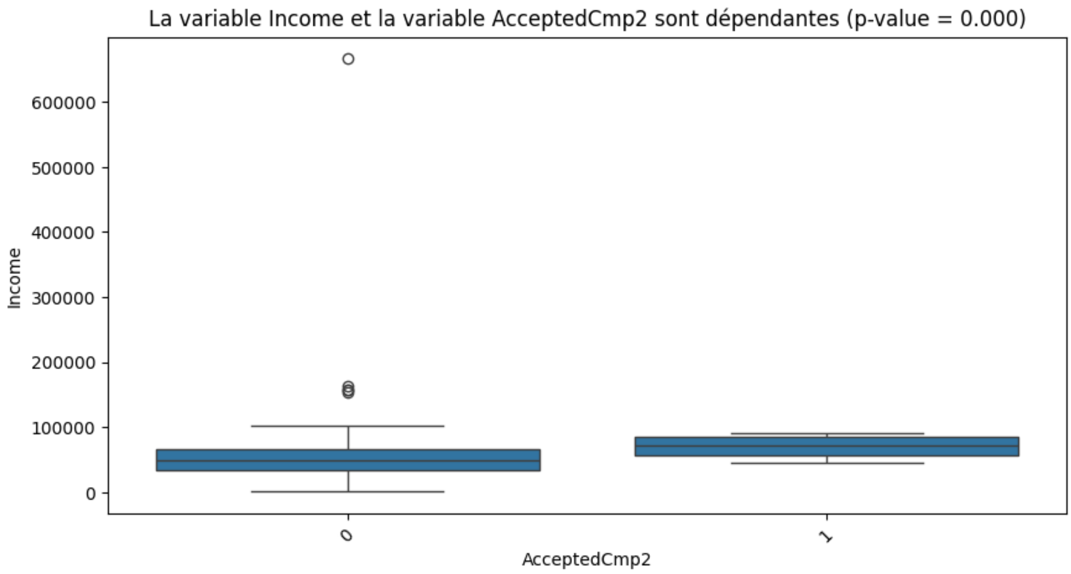



```
 La moyenne de la variable Income par rapport à la variable AcceptedCmp1
 AcceptedCmp1
```

```
1    77006.439
0    49501.586
Name: Income, dtype: float64
```

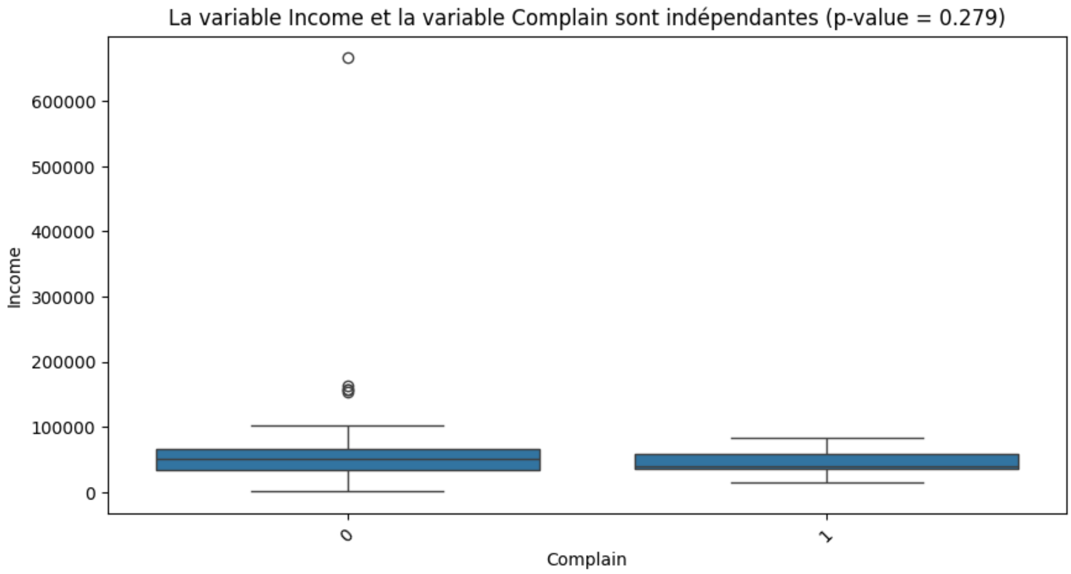



```
 La moyenne de la variable Income par rapport à la variable AcceptedCmp2
 AcceptedCmp2
1    70985.963
0    50730.991
Name: Income, dtype: float64
```

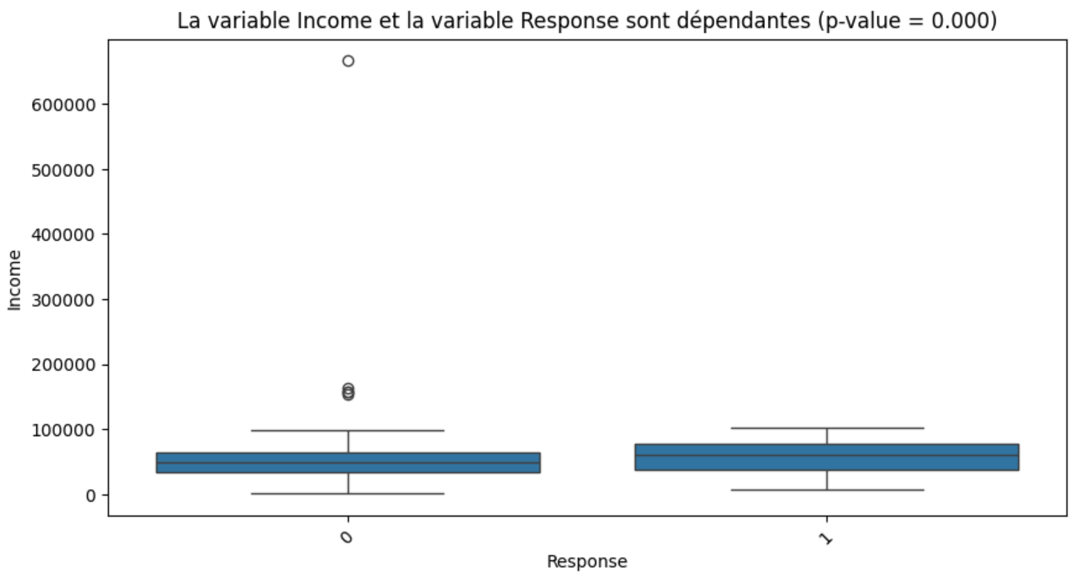



```
 La moyenne de la variable Income par rapport à la variable Complain
```

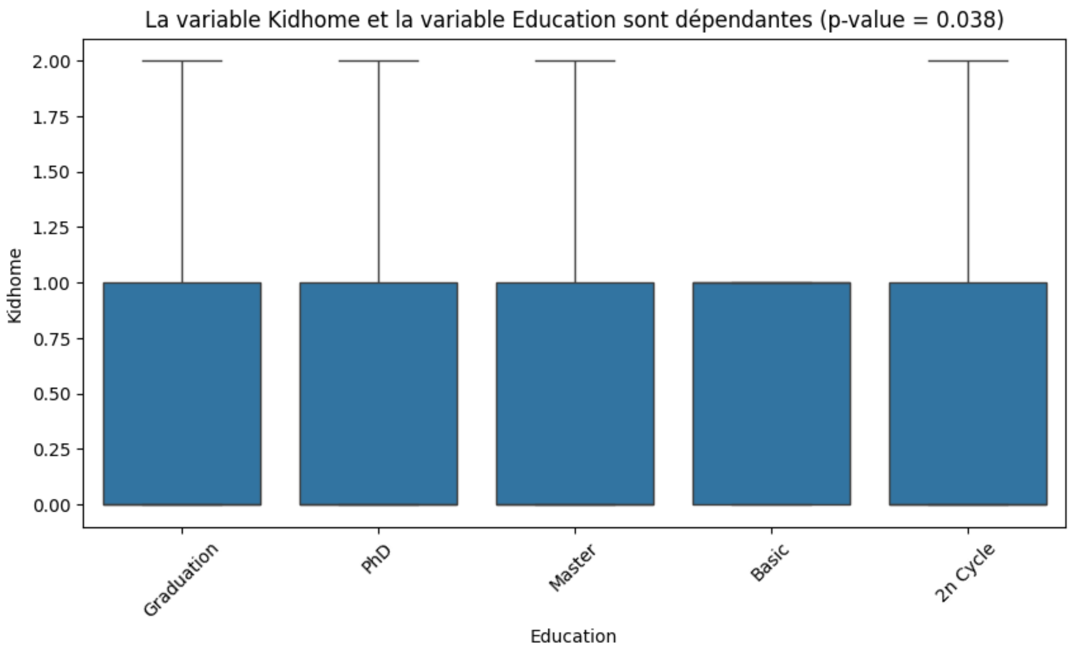
```
 Complain
0    51045.251
1    45242.286
Name: Income, dtype: float64
```



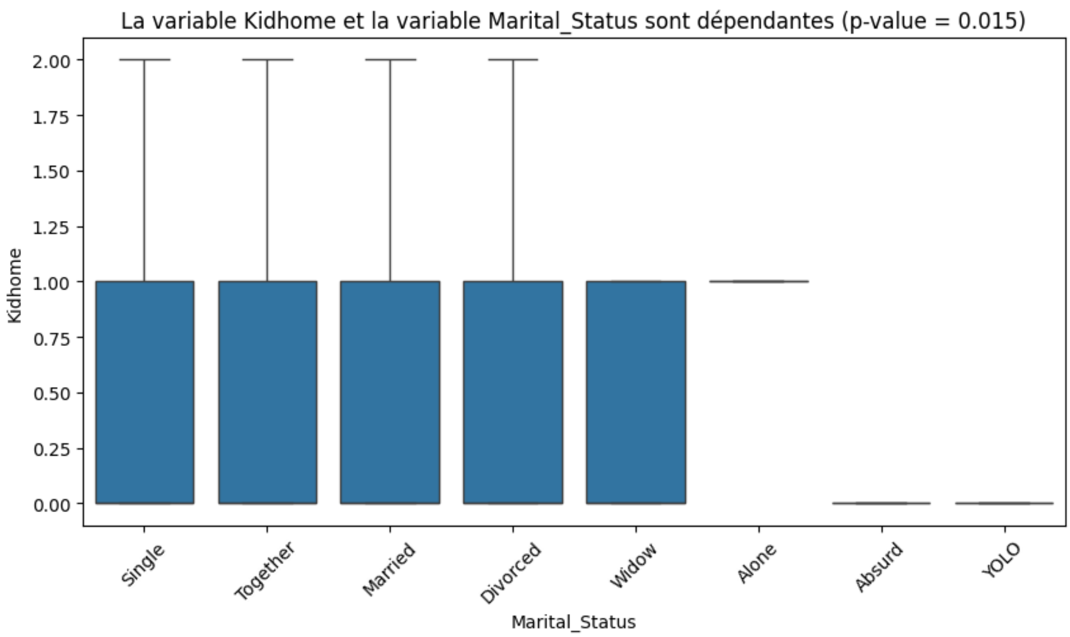

```
 La moyenne de la variable Income par rapport à la variable Response
 Response
1    57685.065
0    49897.423
Name: Income, dtype: float64
```



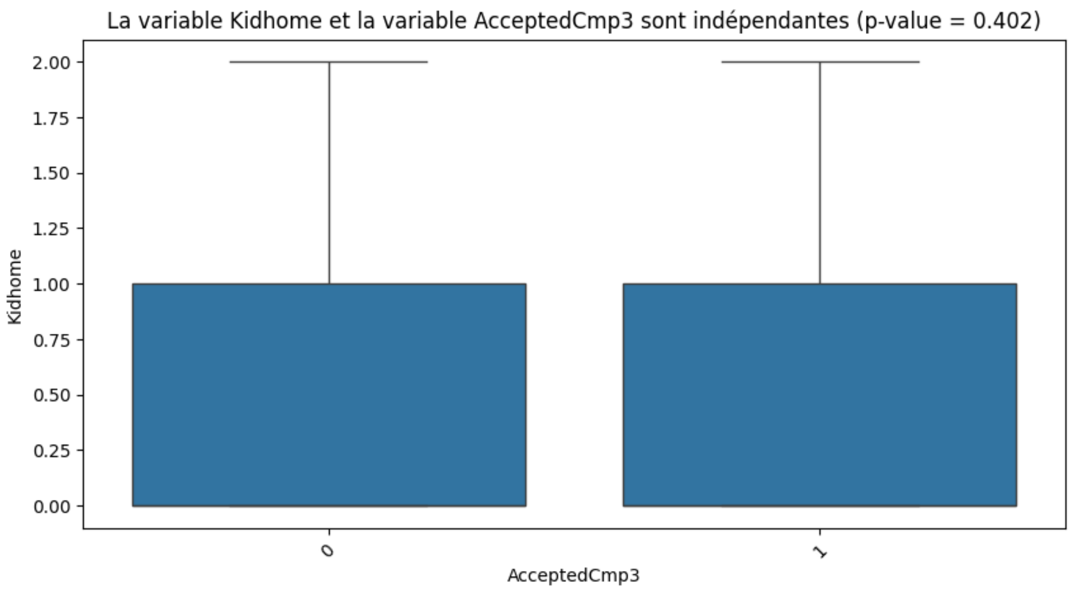
```
 La moyenne de la variable Kidhome par rapport à la variable Education
 Education
Basic         0.642
2n Cycle      0.497
Graduation    0.470
Master        0.469
PhD           0.417
Name: Kidhome, dtype: float64
```




```
 La moyenne de la variable Kidhome par rapport à la variable Marital_Status
 Marital_Status
Alone       1.000
Single      0.492
Married     0.472
Together    0.472
Divorced    0.428
Widow       0.261
Absurd      0.000
YOLO        0.000
Name: Kidhome, dtype: float64
```

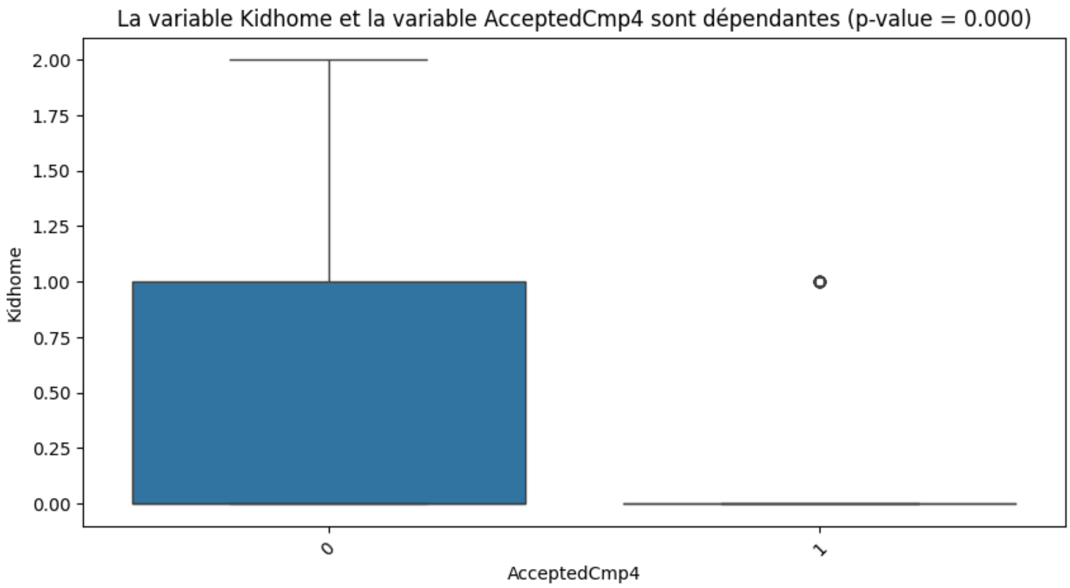



```
 La moyenne de la variable Kidhome par rapport à la variable AcceptedCmp3
 AcceptedCmp3
1    0.500
0    0.462
Name: Kidhome, dtype: float64
```

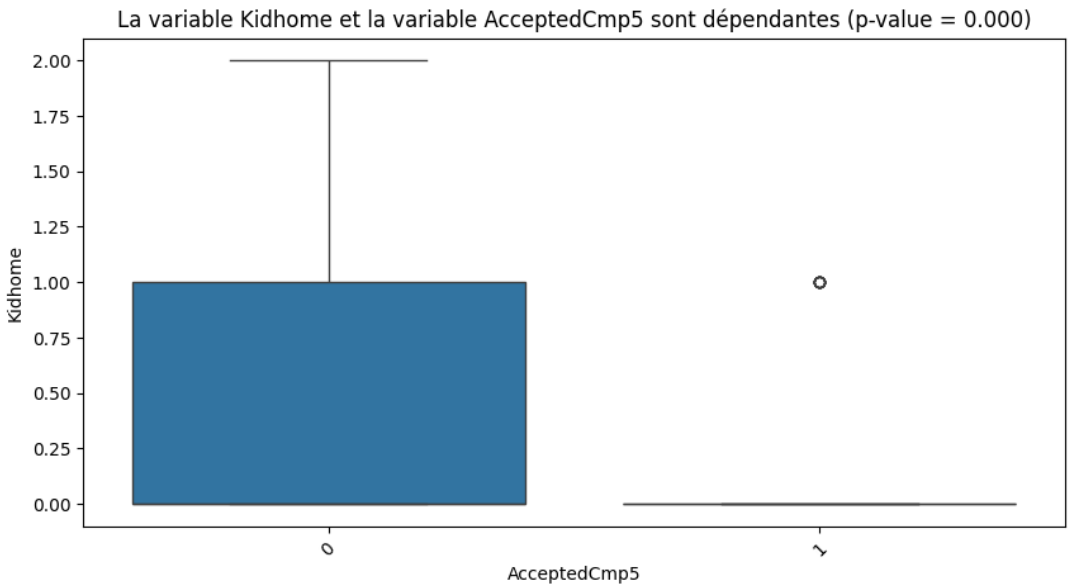



```
 La moyenne de la variable Kidhome par rapport à la variable AcceptedCmp4
 AcceptedCmp4
0    0.490
```

```
1    0.148
Name: Kidhome, dtype: float64
```

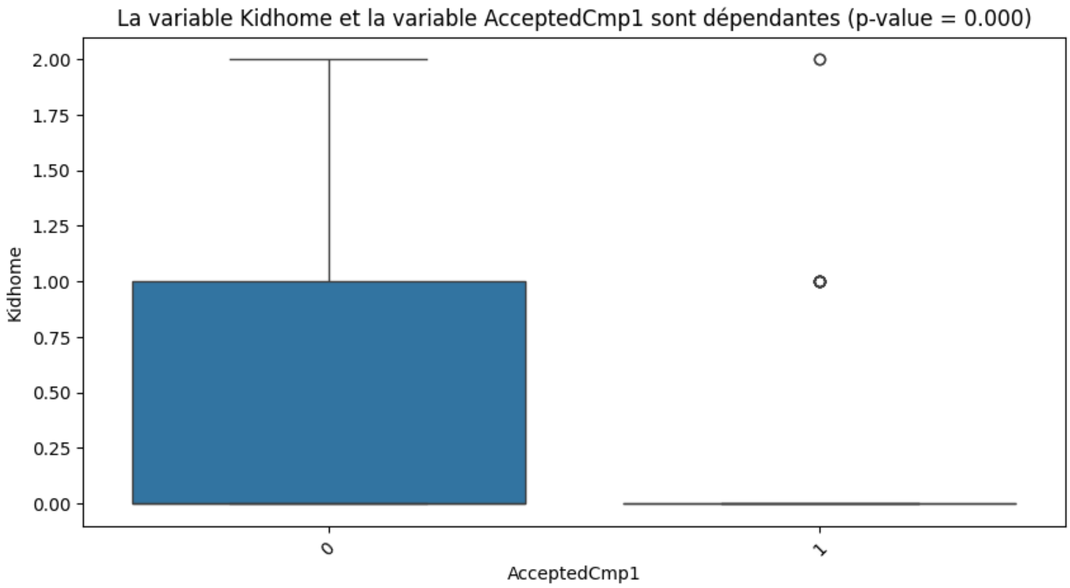



```
 La moyenne de la variable Kidhome par rapport à la variable AcceptedCmp5
 AcceptedCmp5
0    0.490
1    0.062
Name: Kidhome, dtype: float64
```

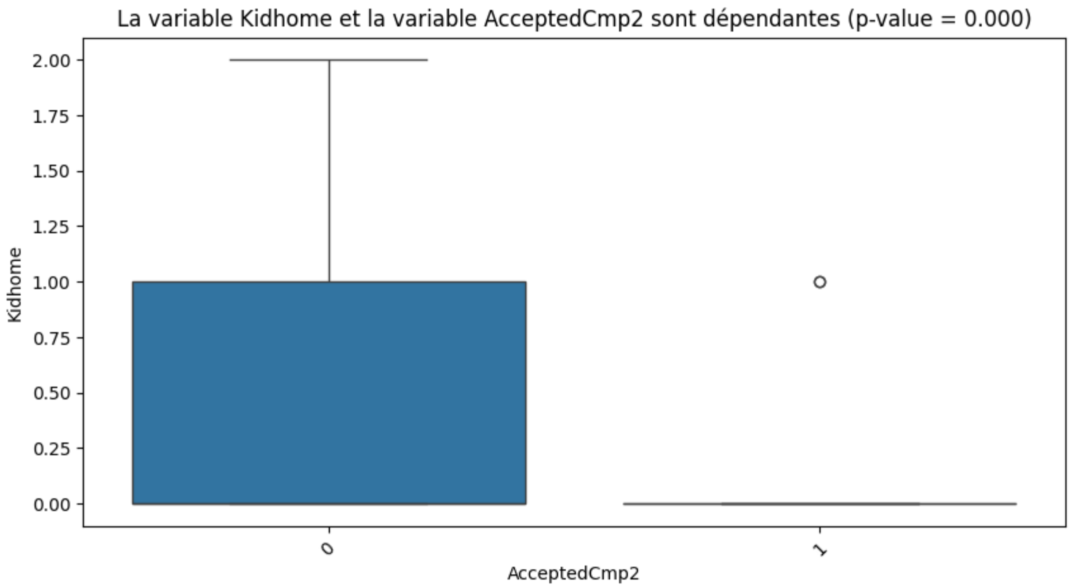



```
 La moyenne de la variable Kidhome par rapport à la variable AcceptedCmp1
```

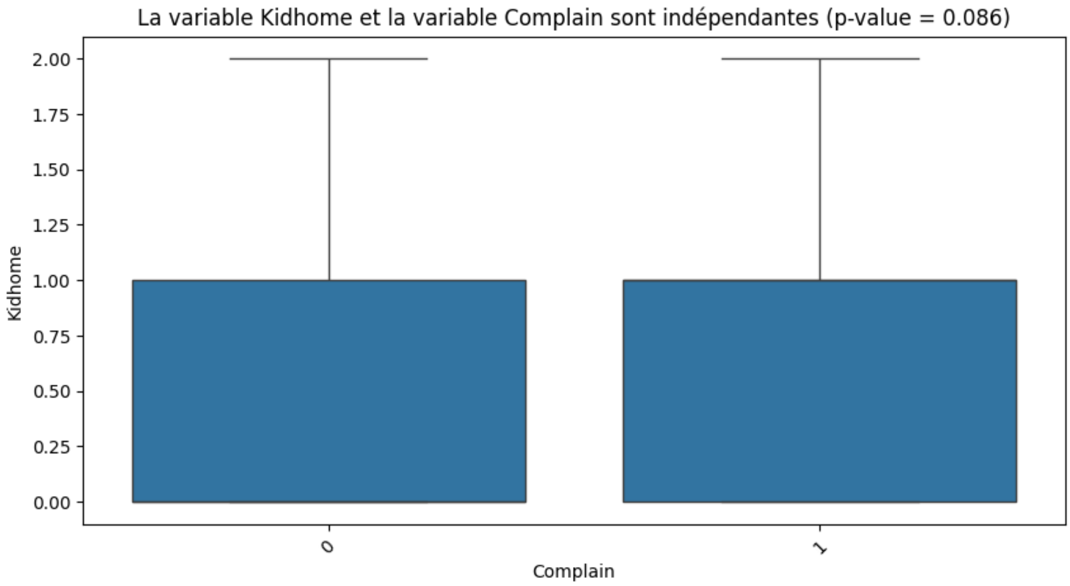
```
 AcceptedCmp1
0    0.485
1    0.104
Name: Kidhome, dtype: float64
```



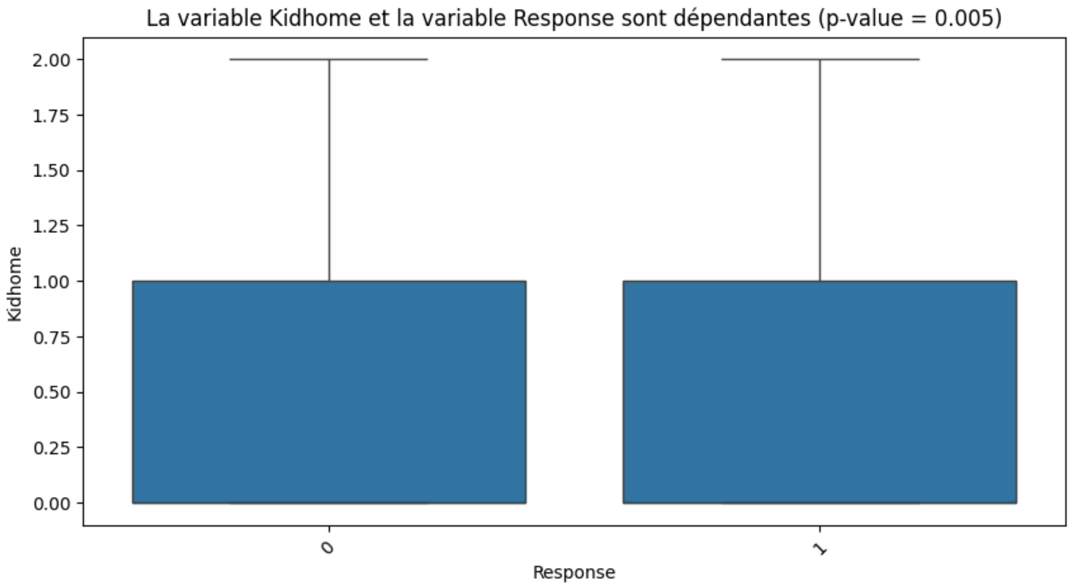

```
 La moyenne de la variable Kidhome par rapport à la variable AcceptedCmp2
 AcceptedCmp2
0    0.470
1    0.074
Name: Kidhome, dtype: float64
```



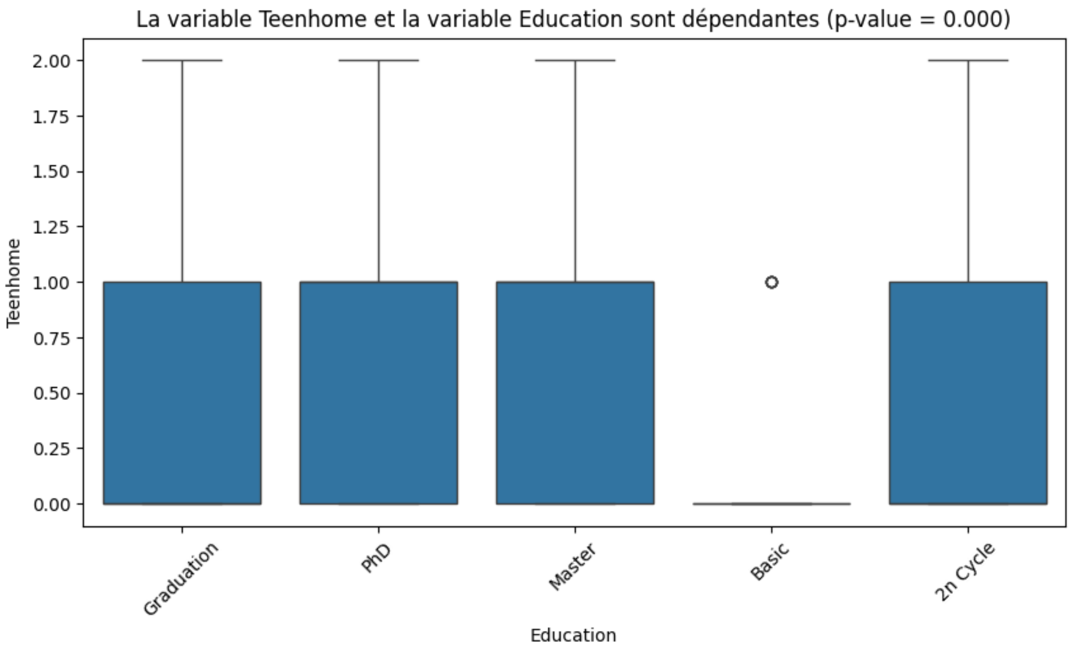
```
 La moyenne de la variable Kidhome par rapport à la variable Complain
 Complain
1    0.667
0    0.463
Name: Kidhome, dtype: float64
```



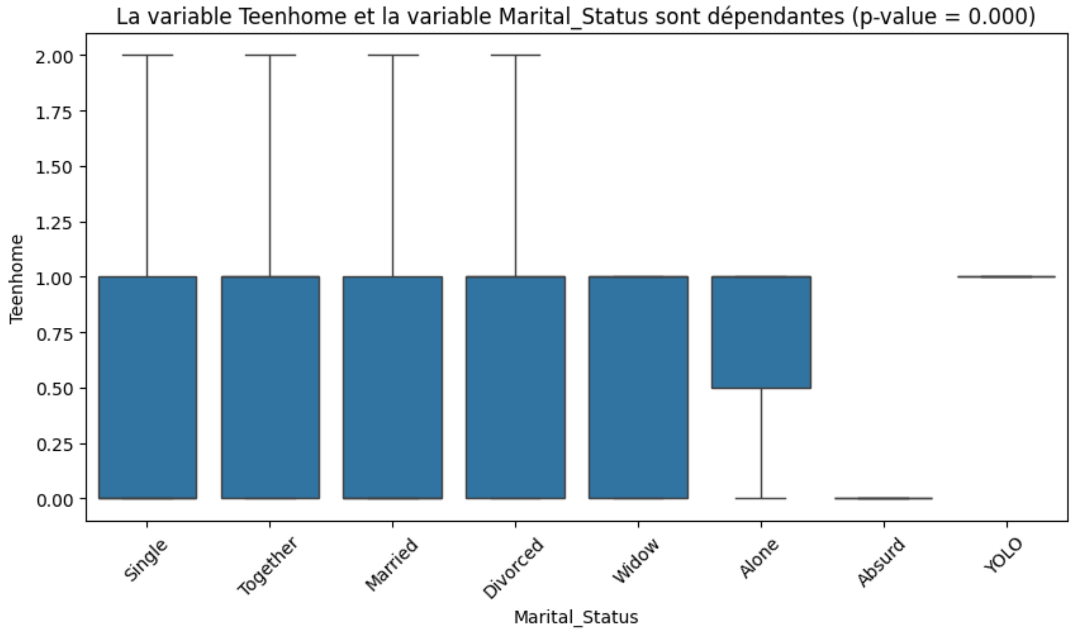

```
 La moyenne de la variable Kidhome par rapport à la variable Response
 Response
0    0.478
1    0.383
Name: Kidhome, dtype: float64
```



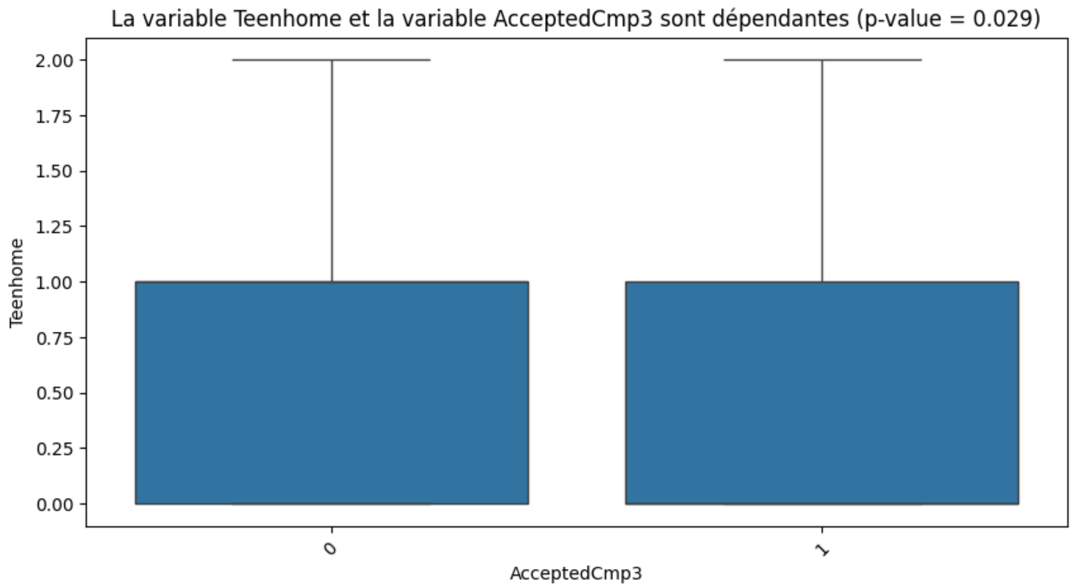
```
 La moyenne de la variable Teenhome par rapport à la variable Education
 Education
PhD            0.607
Master         0.547
Graduation     0.516
2n Cycle       0.403
Basic          0.094
Name: Teenhome, dtype: float64
```




```
 La moyenne de la variable Teenhome par rapport à la variable Marital_Status
 Marital_Status
YOLO         1.000
Alone        0.667
Widow        0.667
Divorced     0.599
Together     0.545
Married      0.526
Single       0.419
Absurd       0.000
Name: Teenhome, dtype: float64
```

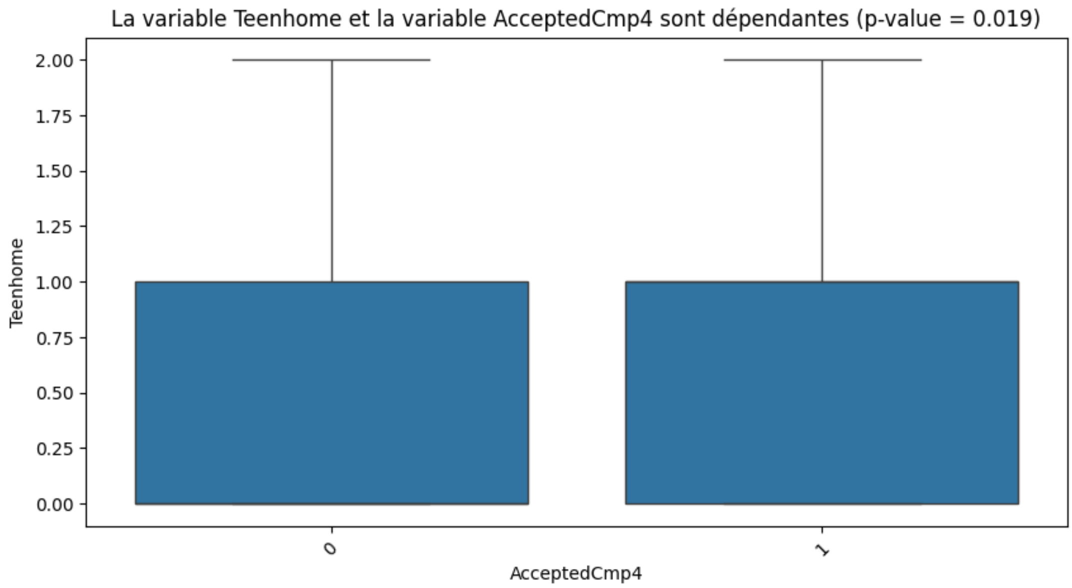



```
 La moyenne de la variable Teenhome par rapport à la variable AcceptedCmp3
 AcceptedCmp3
0    0.528
1    0.429
Name: Teenhome, dtype: float64
```

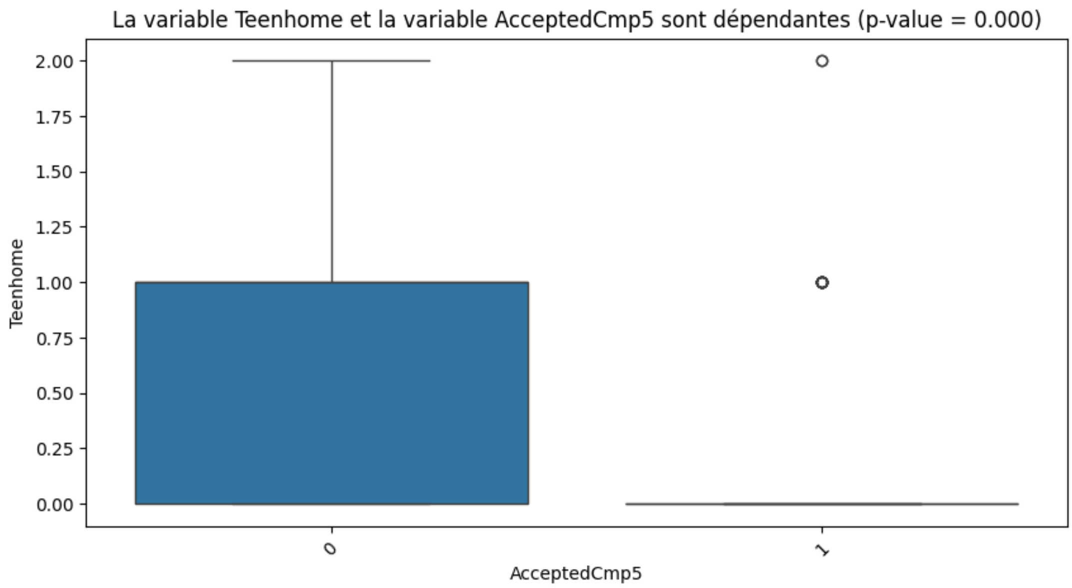



```
 La moyenne de la variable Teenhome par rapport à la variable AcceptedCmp4
 AcceptedCmp4
1    0.619
```

```
0    0.513
Name: Teenhome, dtype: float64
```

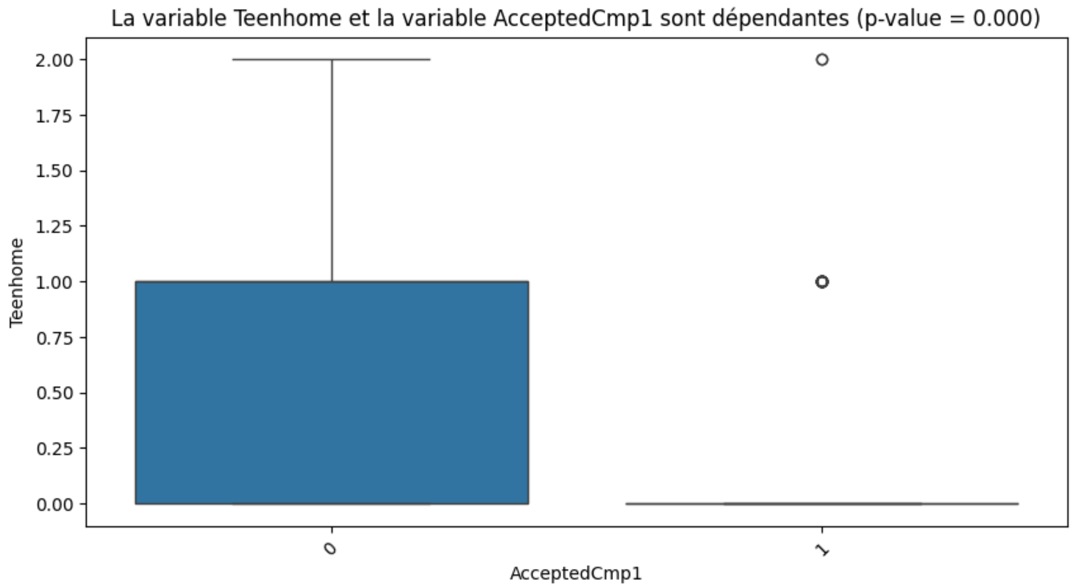



```
 La moyenne de la variable Teenhome par rapport à la variable AcceptedCmp5
 AcceptedCmp5
0    0.544
1    0.148
Name: Teenhome, dtype: float64
```

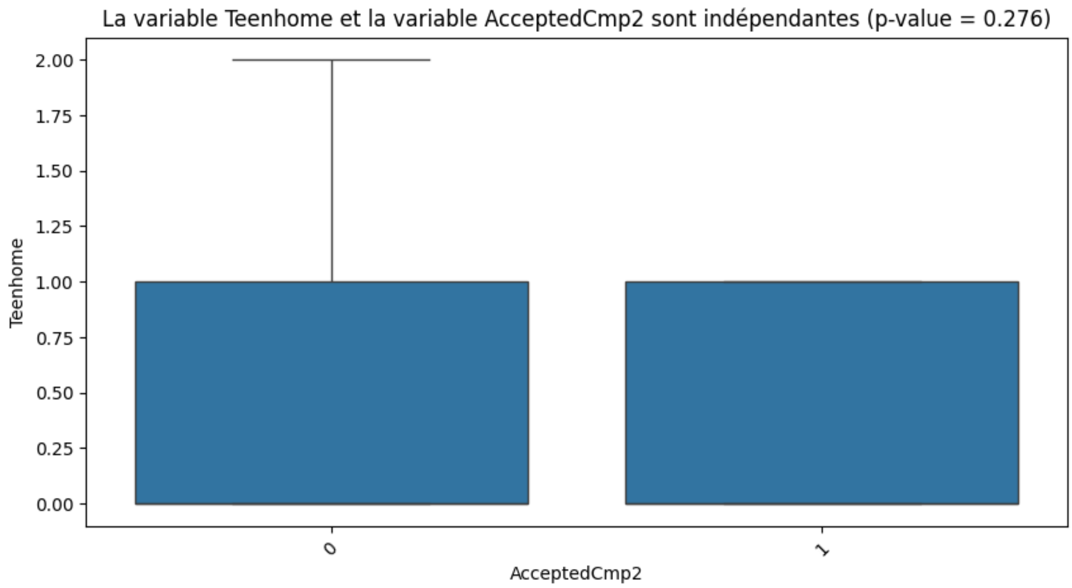



```
 La moyenne de la variable Teenhome par rapport à la variable AcceptedCmp1
```

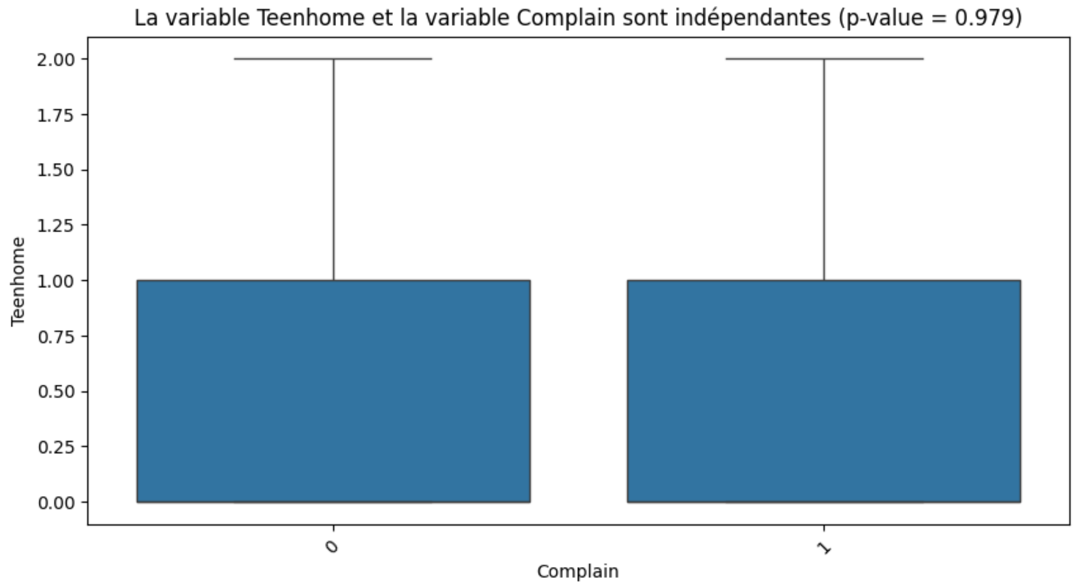
```
 AcceptedCmp1
0    0.536
1    0.261
Name: Teenhome, dtype: float64
```



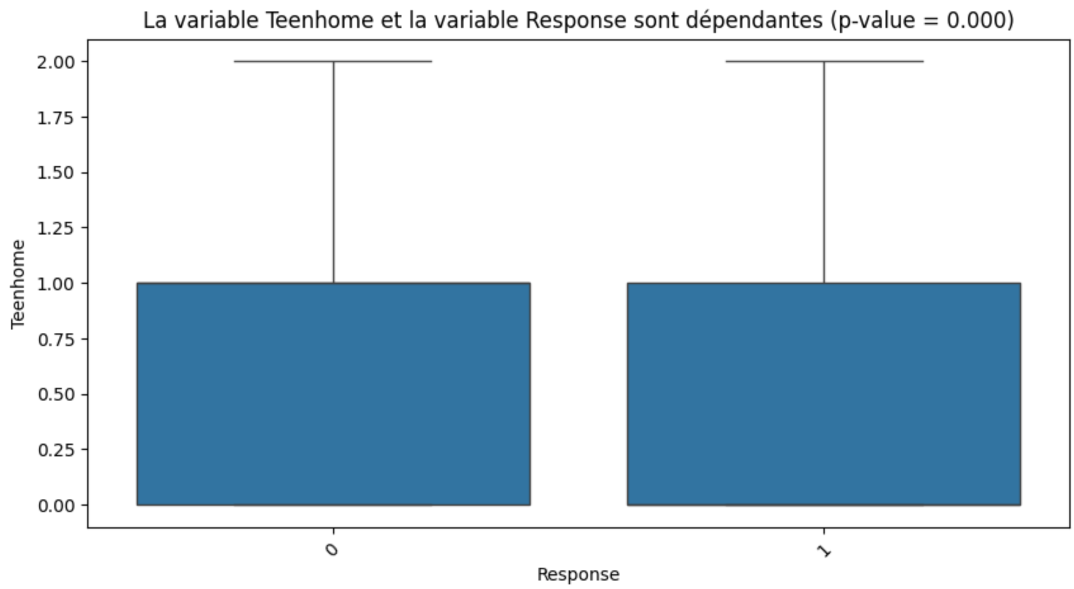

```
 La moyenne de la variable Teenhome par rapport à la variable AcceptedCmp2
 AcceptedCmp2
0    0.522
1    0.407
Name: Teenhome, dtype: float64
```



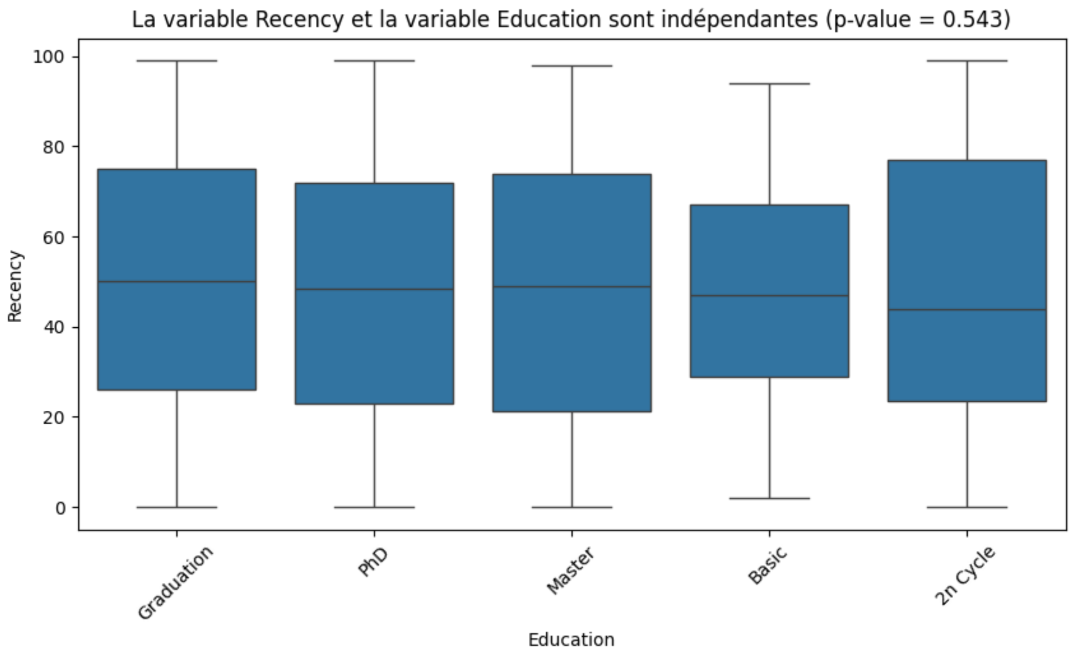
```
 La moyenne de la variable Teenhome par rapport à la variable Complain
 Complain
1    0.524
0    0.521
Name: Teenhome, dtype: float64
```



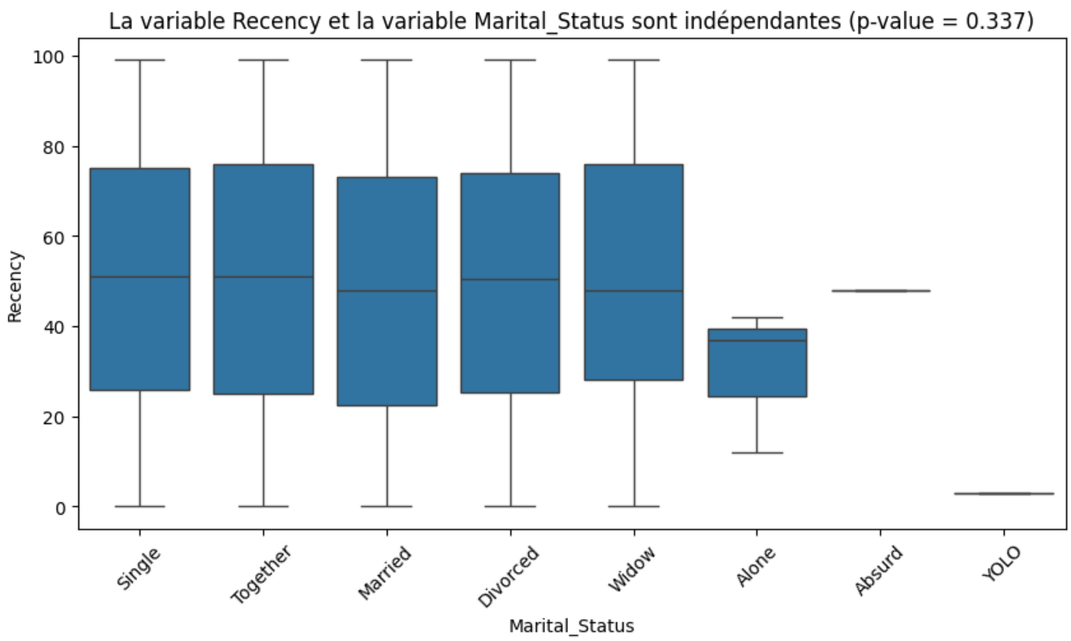

```
 La moyenne de la variable Teenhome par rapport à la variable Response
 Response
0    0.552
1    0.329
Name: Teenhome, dtype: float64
```



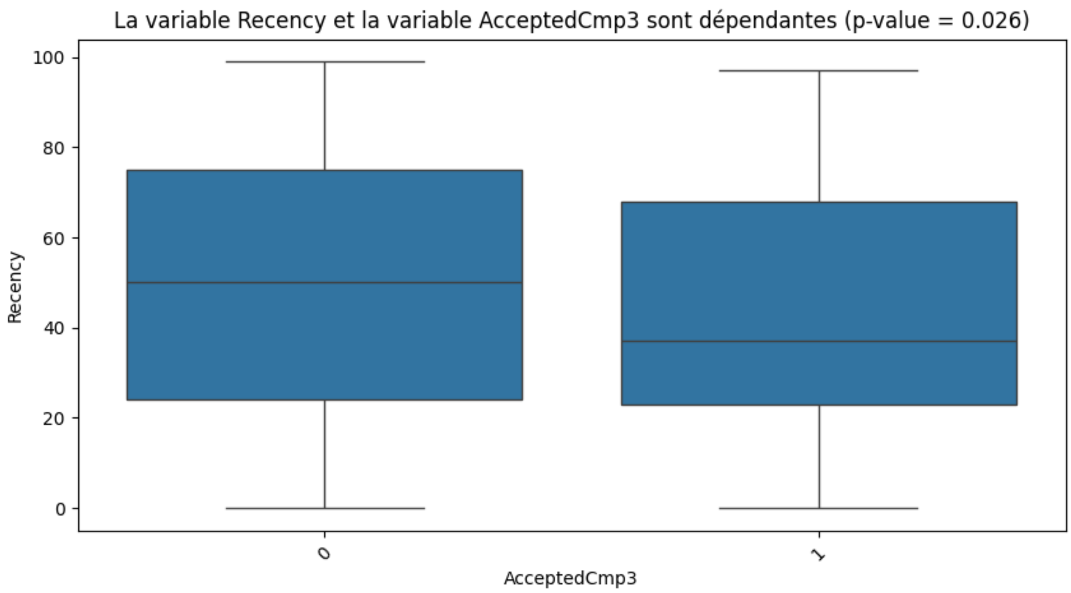
```
La moyenne de la variable Recency par rapport à la variable Education
 Education
Graduation    50.200
2n Cycle      48.241
PhD           48.150
Master        47.749
Basic         47.623
Name: Recency, dtype: float64
```




```
La moyenne de la variable Recency par rapport à la variable Marital_Status
 Marital_Status
Together    50.035
Single      49.860
Widow       49.217
Divorced    49.203
Married     48.205
Absurd      48.000
Alone       30.333
YOLO         3.000
Name: Recency, dtype: float64
```

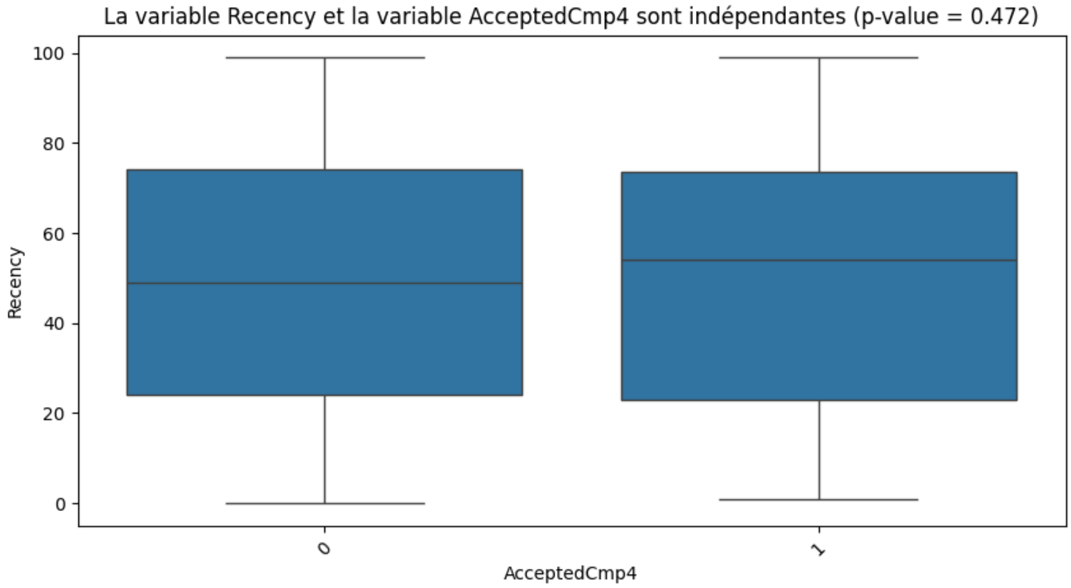



```
 La moyenne de la variable Recency par rapport à la variable AcceptedCmp3
 AcceptedCmp3
0    49.487
1    44.091
Name: Recency, dtype: float64
```

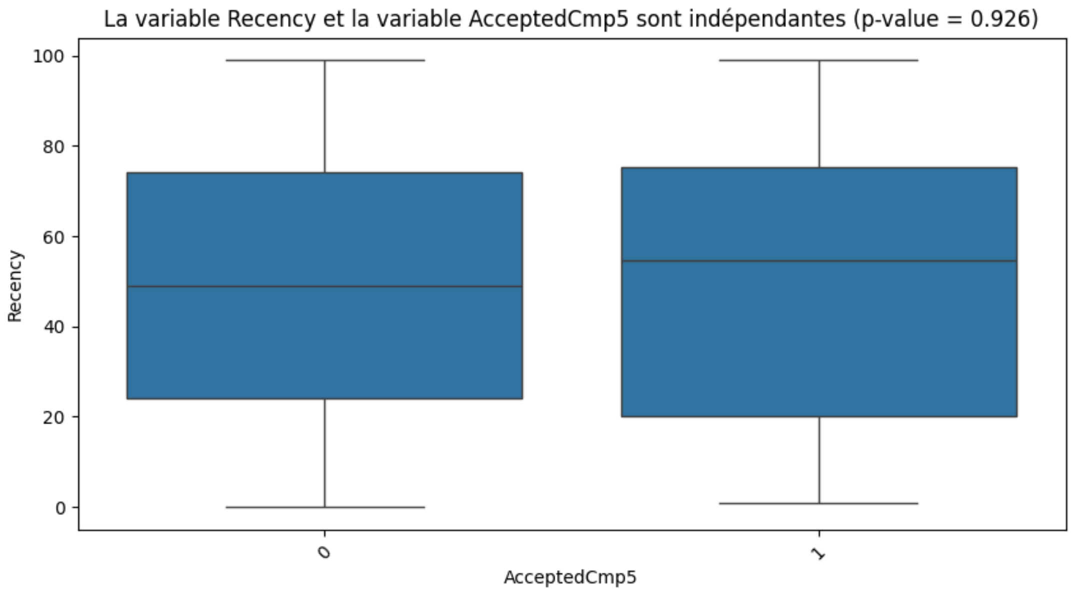



```
 La moyenne de la variable Recency par rapport à la variable AcceptedCmp4
 AcceptedCmp4
```

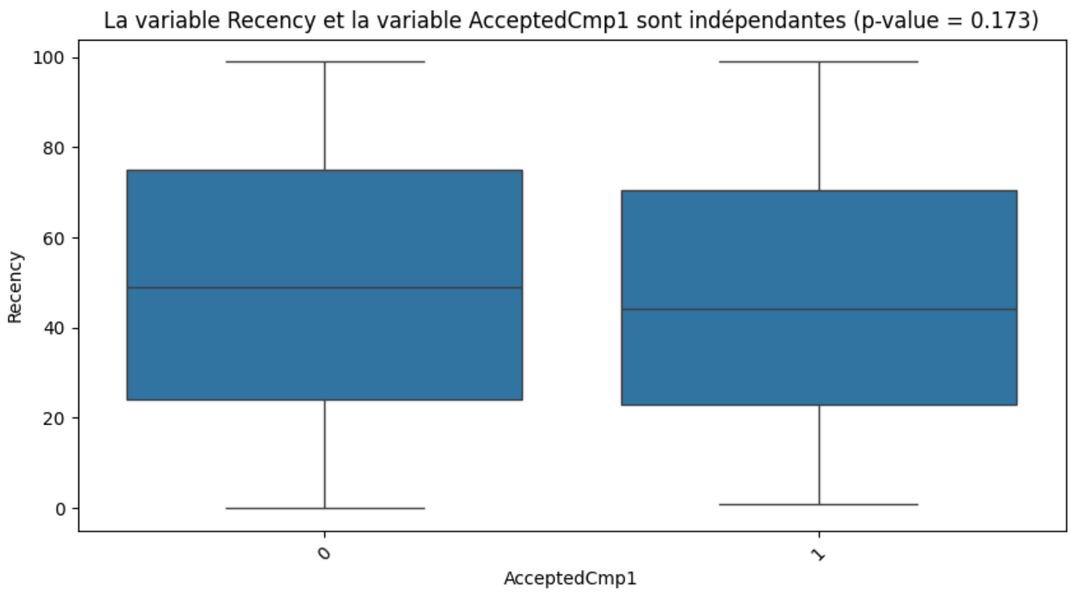
```
1    50.716
0    48.970
Name: Recency, dtype: float64
```



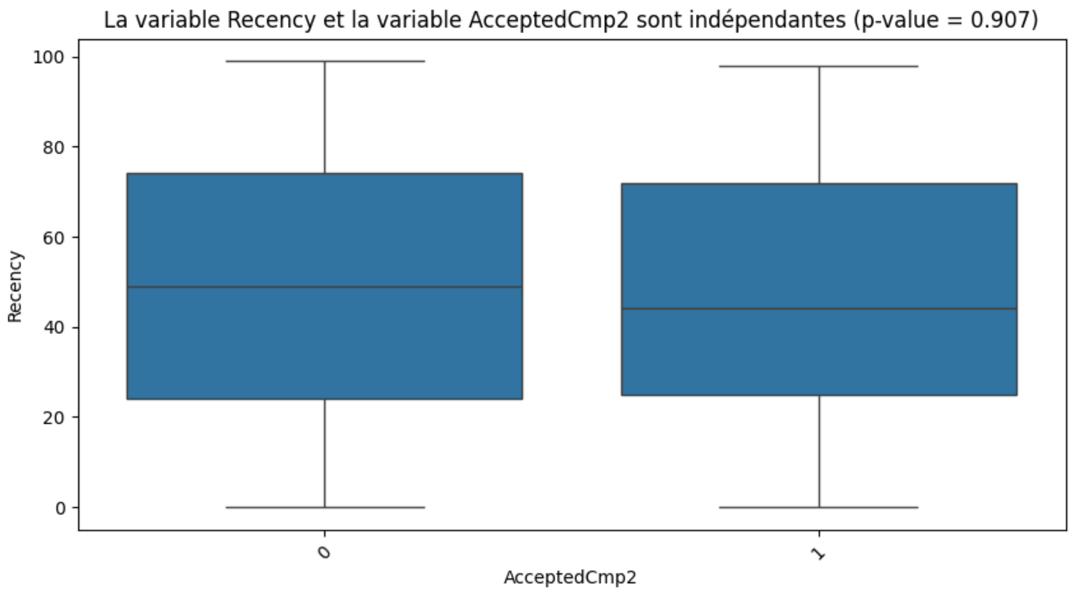

```
 La moyenne de la variable Recency par rapport à la variable AcceptedCmp5
 AcceptedCmp5
1    49.328
0    49.082
Name: Recency, dtype: float64
```



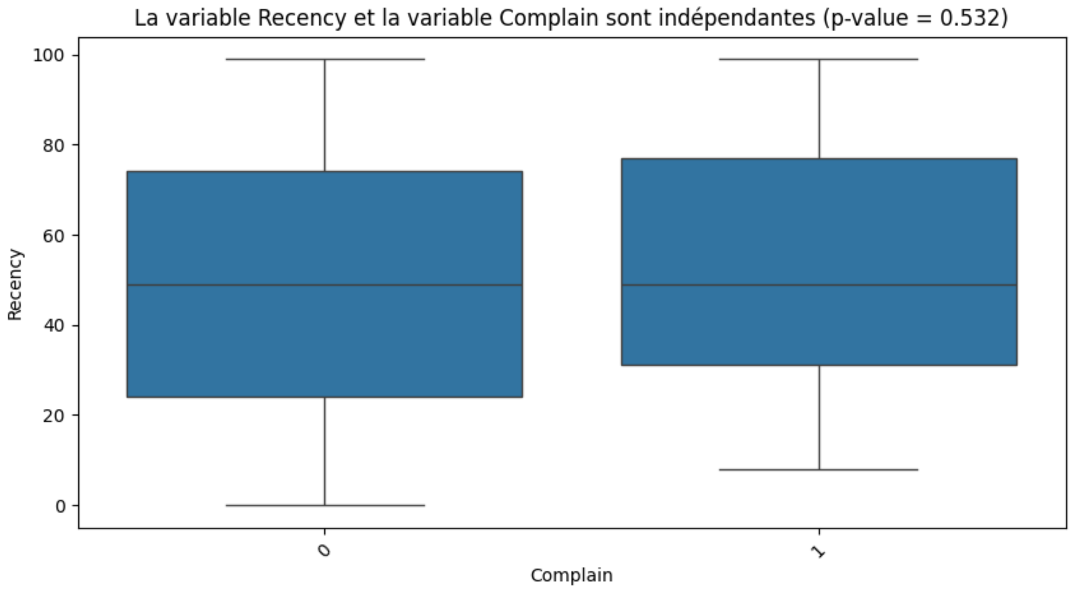
```
 La moyenne de la variable Recency par rapport à la variable AcceptedCmp1
 AcceptedCmp1
0    49.302
1    45.504
Name: Recency, dtype: float64
```



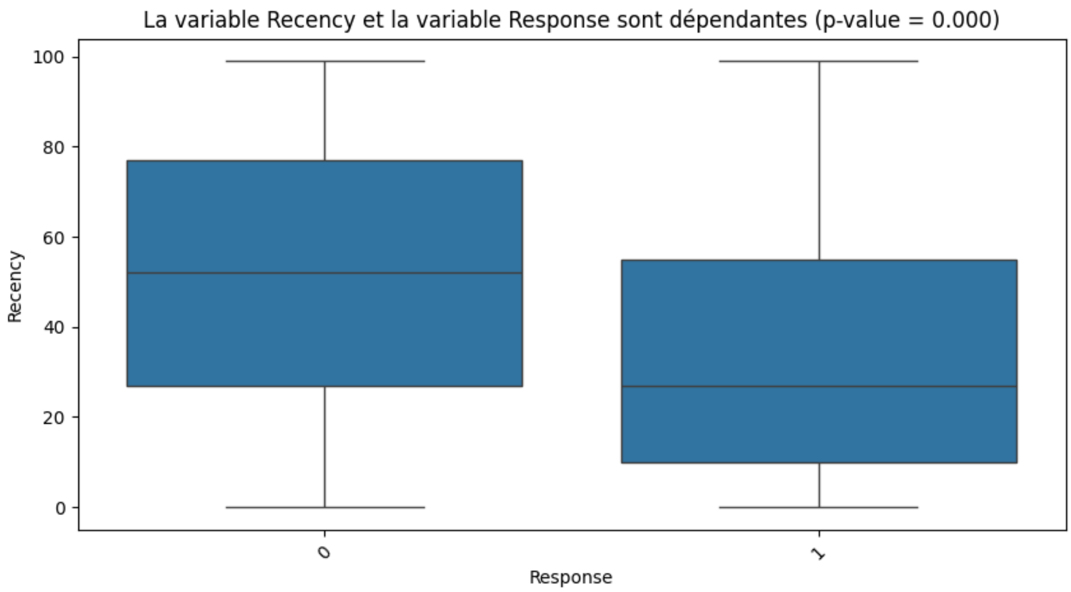

```
 La moyenne de la variable Recency par rapport à la variable AcceptedCmp2
 AcceptedCmp2
0    49.105
1    48.444
Name: Recency, dtype: float64
```



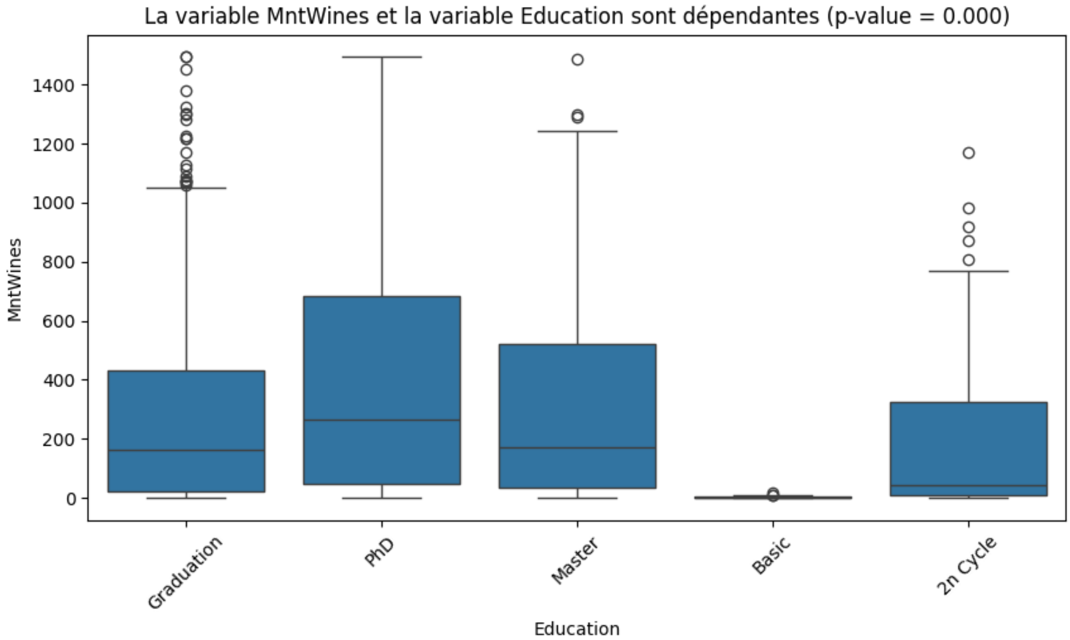
```
 La moyenne de la variable Recency par rapport à la variable Complain
 Complain
1    53.048
0    49.057
Name: Recency, dtype: float64
```



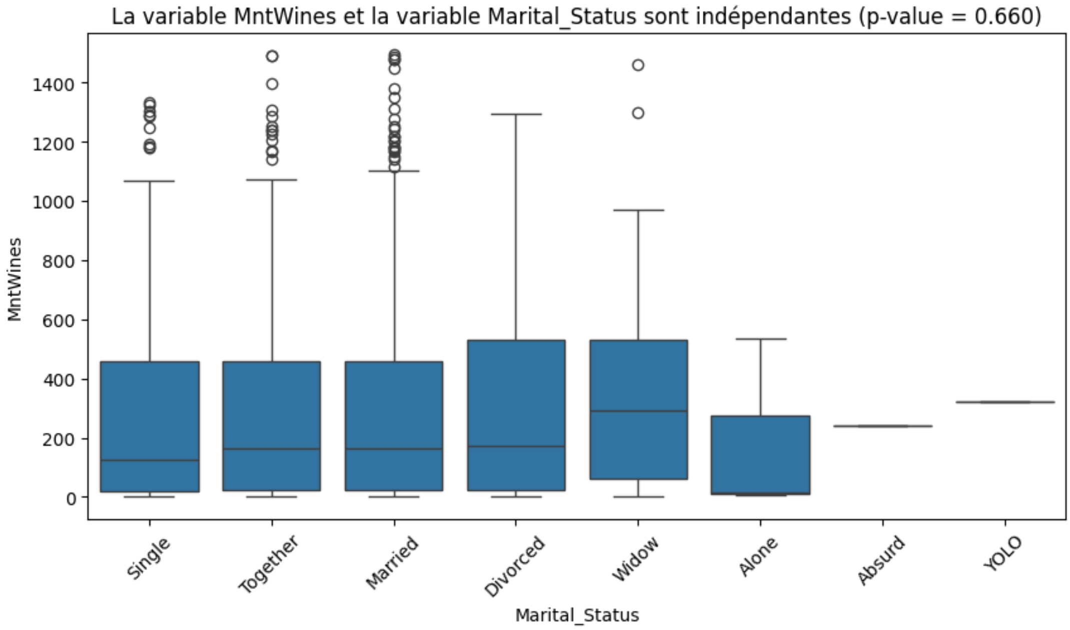

```
 La moyenne de la variable Recency par rapport à la variable Response
 Response
0    51.531
1    34.151
Name: Recency, dtype: float64
```



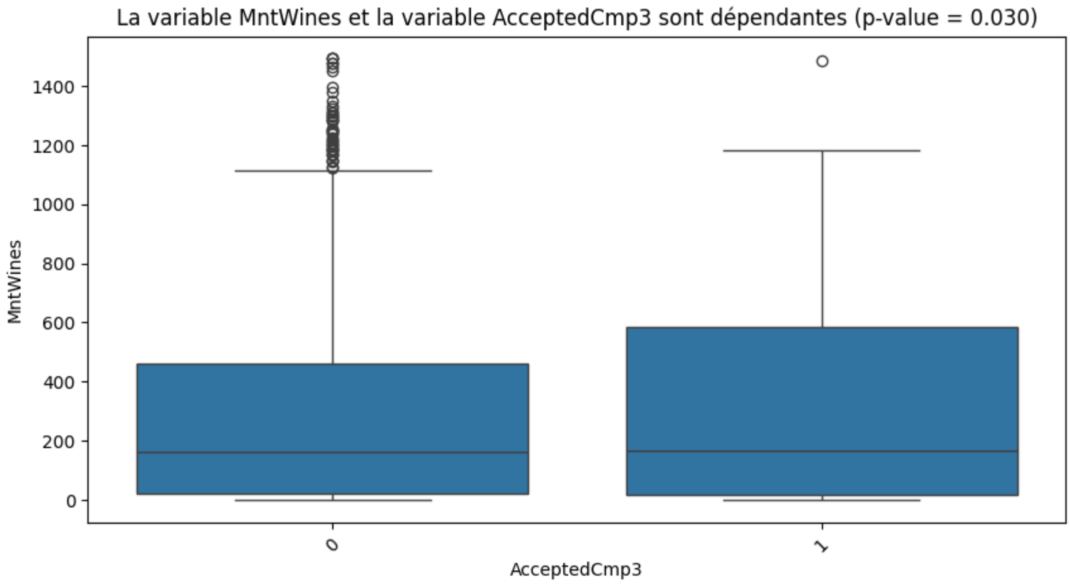
```
 La moyenne de la variable MntWines par rapport à la variable Education
 Education
PhD           395.547
Master        315.243
Graduation    264.464
2n Cycle      180.885
Basic           3.075
Name: MntWines, dtype: float64
```




```
 La moyenne de la variable MntWines par rapport à la variable Marital_Status
 Marital_Status
Widow       360.101
YOLO        322.000
Divorced    303.171
Married     286.815
Together    285.492
Single      274.548
Absurd      240.000
Alone       184.667
Name: MntWines, dtype: float64
```

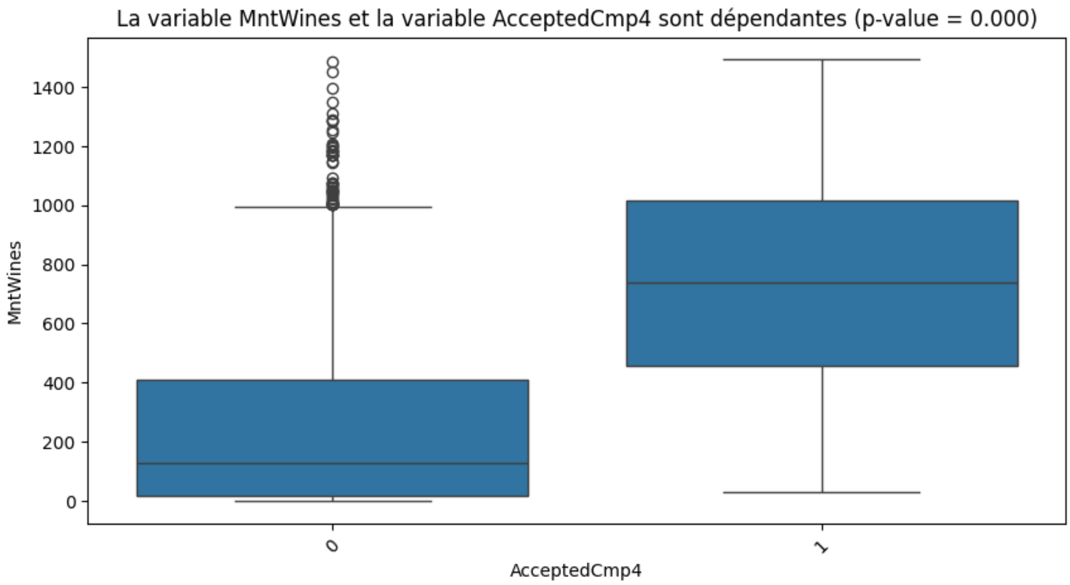



```
 La moyenne de la variable MntWines par rapport à la variable AcceptedCmp3
 AcceptedCmp3
1    342.610
0    283.549
Name: MntWines, dtype: float64
```

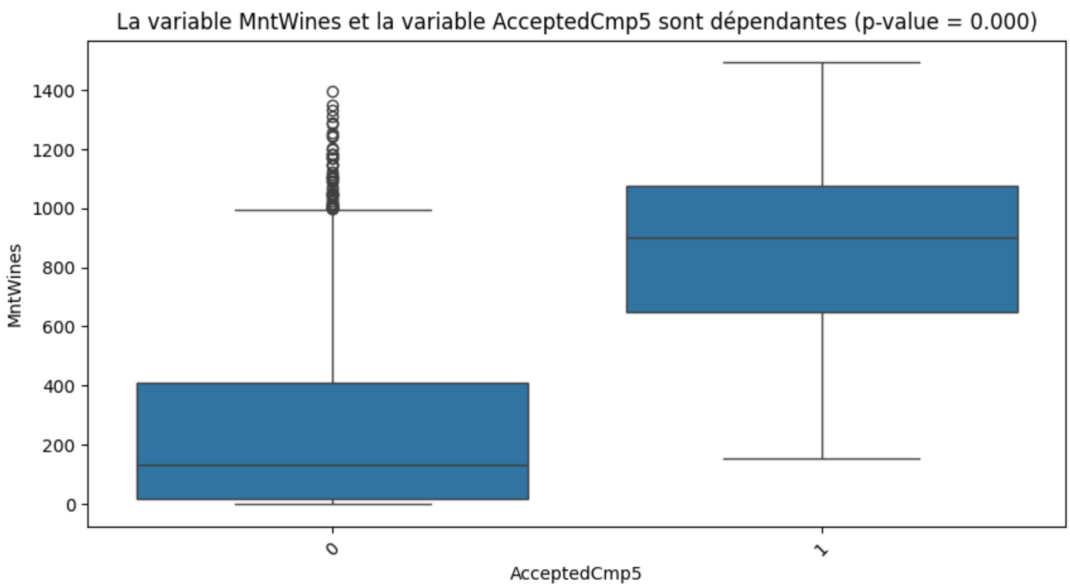



```
 La moyenne de la variable MntWines par rapport à la variable AcceptedCmp4
 AcceptedCmp4
1    736.161
```

```
0    252.602
Name: MntWines, dtype: float64
```

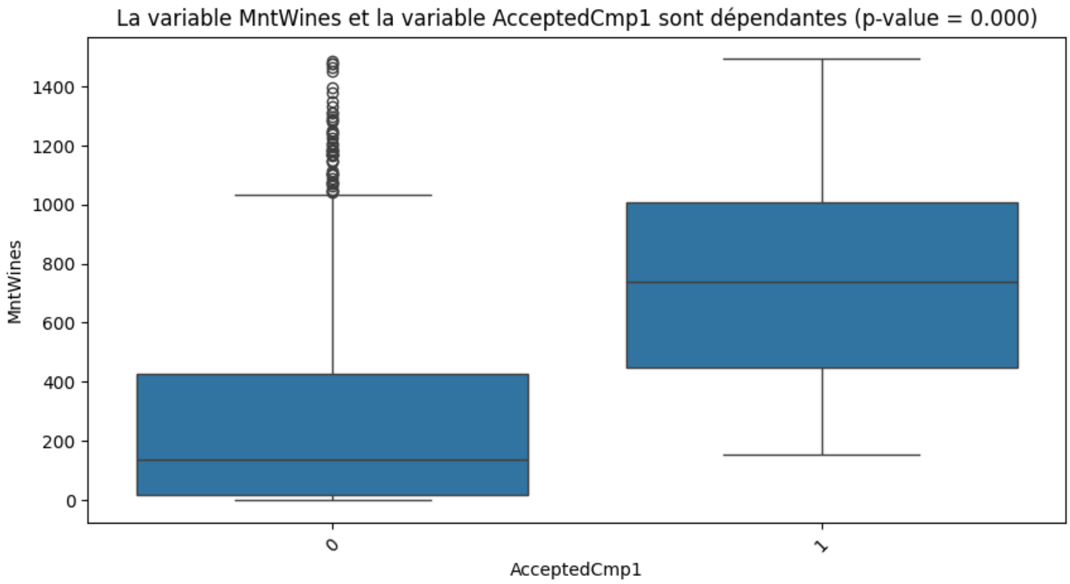



```
 La moyenne de la variable MntWines par rapport à la variable AcceptedCmp5
 AcceptedCmp5
1    861.508
0    251.108
Name: MntWines, dtype: float64
```

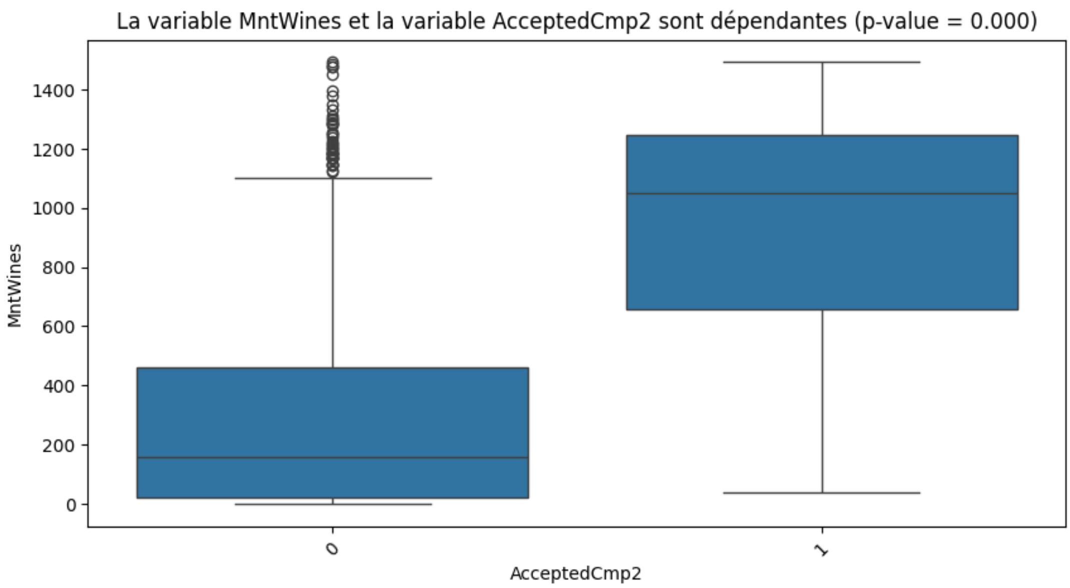



```
 La moyenne de la variable MntWines par rapport à la variable AcceptedCmp1
```

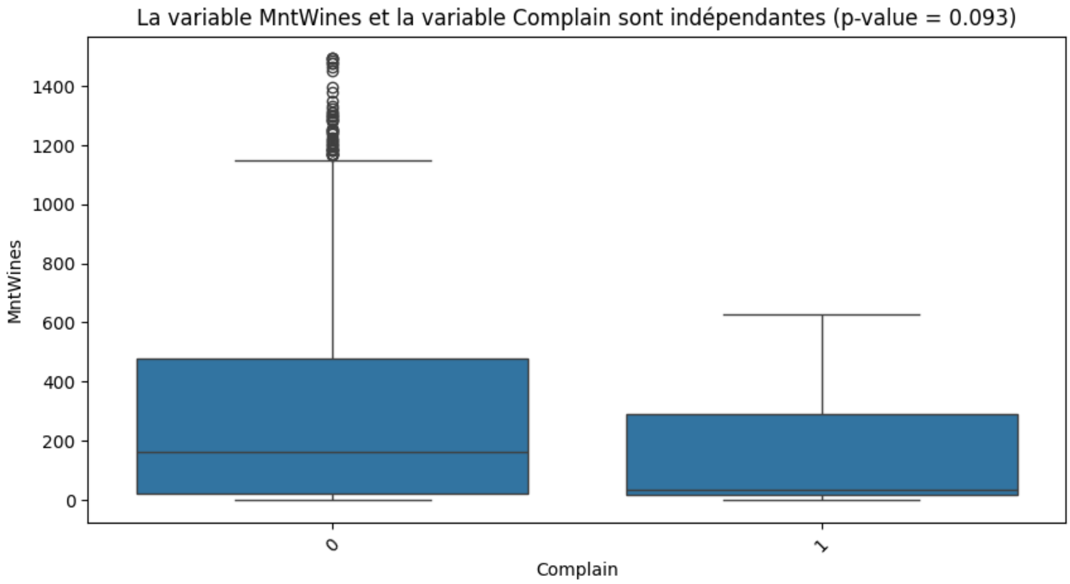
 AcceptedCmp1
1   736.991
0   262.163
Name: MntWines, dtype: float64



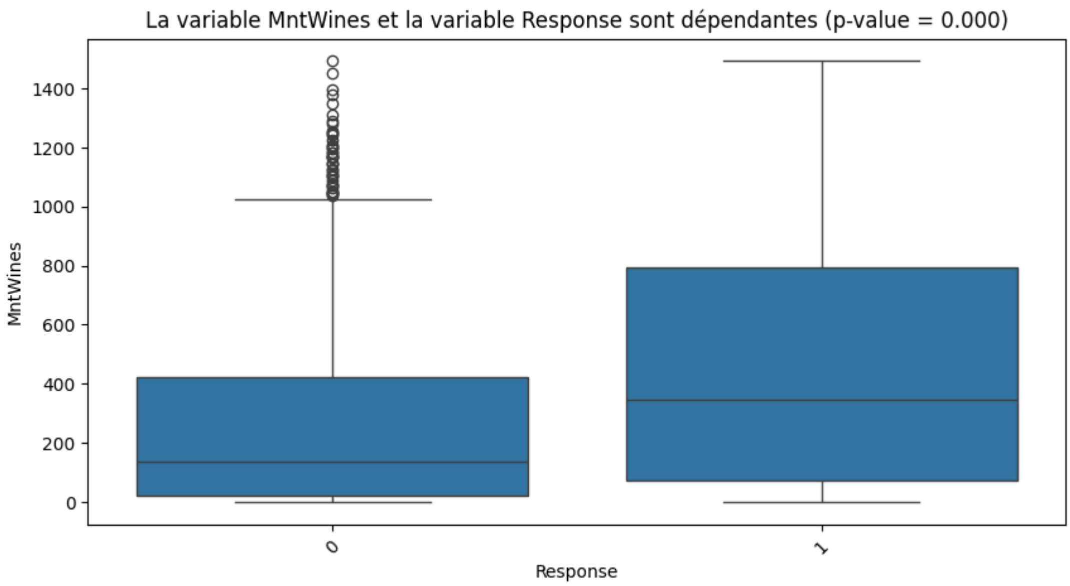
 La moyenne de la variable MntWines par rapport à la variable AcceptedCmp2
 AcceptedCmp2
1   899.778
0   279.959
Name: MntWines, dtype: float64



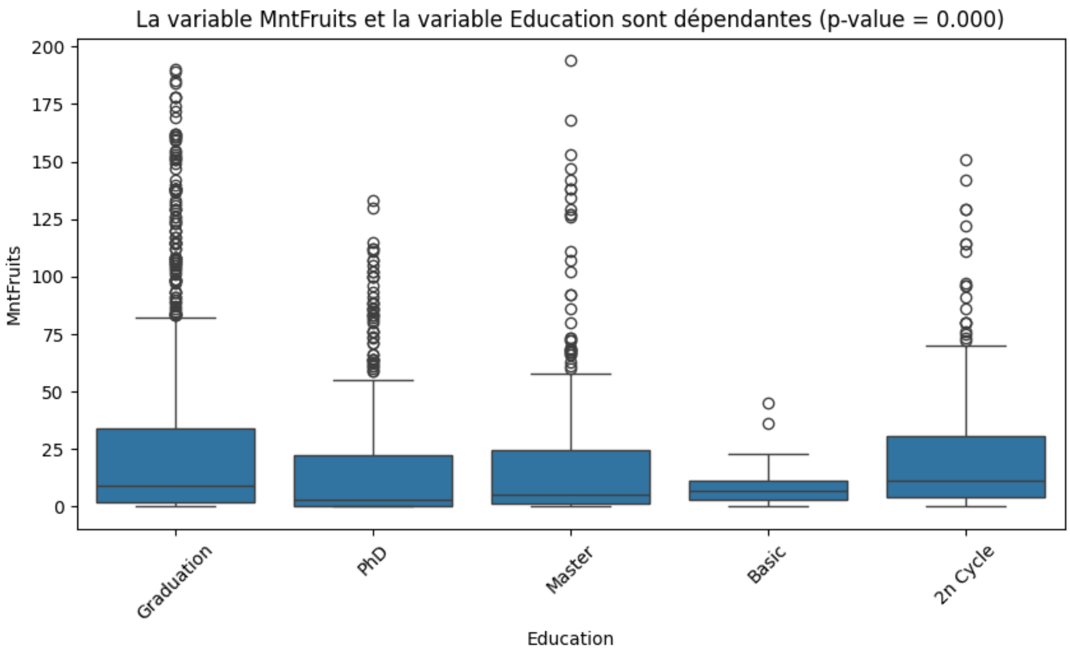
```
 La moyenne de la variable MntWines par rapport à la variable Complain
 Complain
0    289.008
1    169.000
Name: MntWines, dtype: float64
```



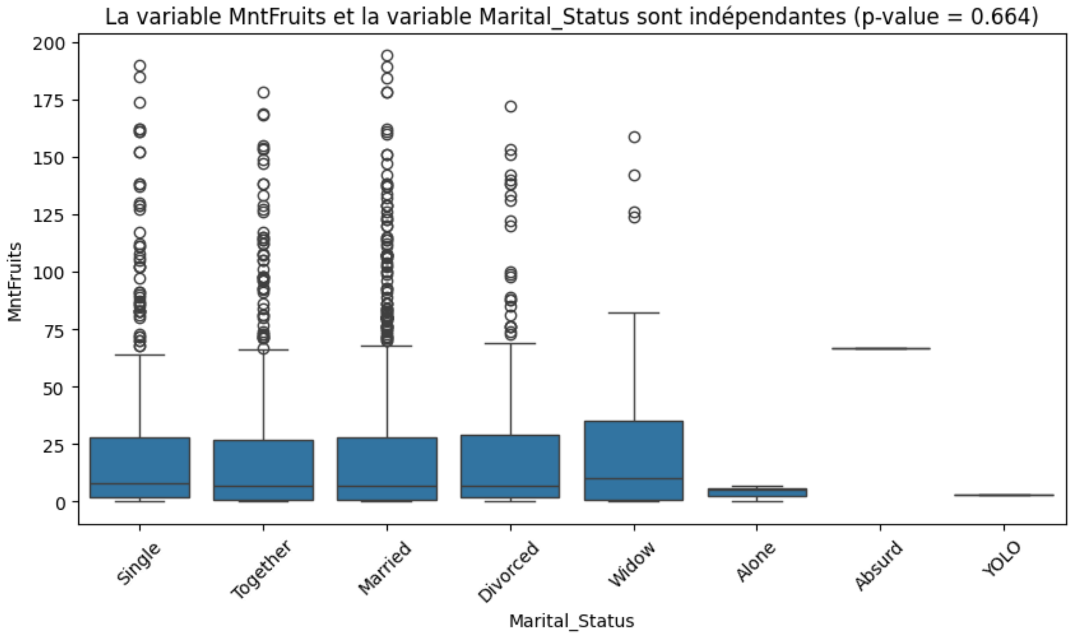

```
 La moyenne de la variable MntWines par rapport à la variable Response
 Response
1    462.517
0    259.376
Name: MntWines, dtype: float64
```



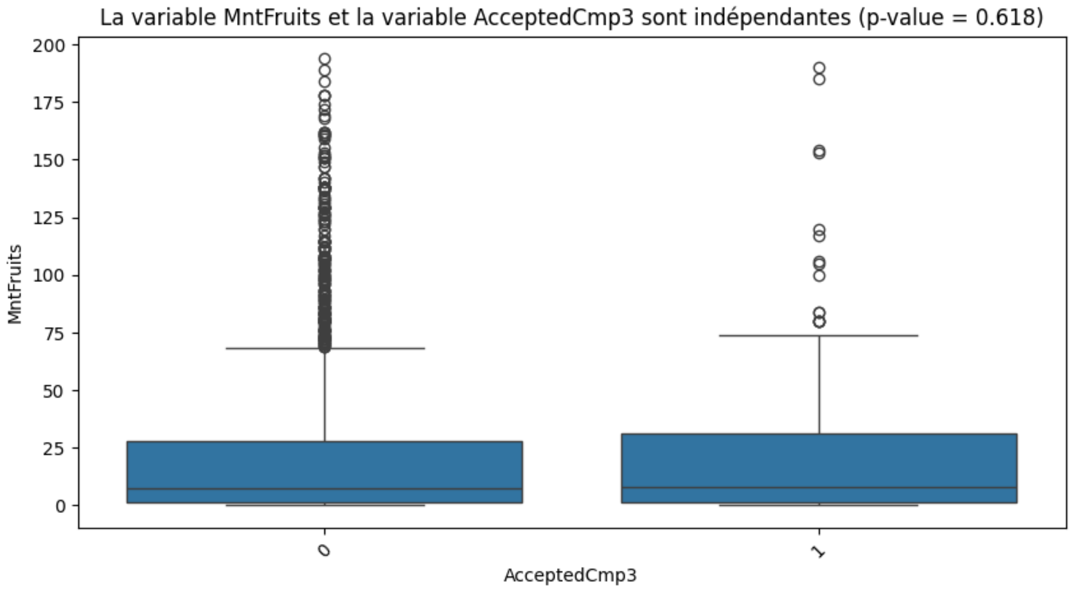
```
 La moyenne de la variable MntFruits par rapport à la variable Education
 Education
Graduation    26.419
2n Cycle      24.293
Master        19.330
PhD           17.620
Basic          9.019
Name: MntFruits, dtype: float64
```




```
 La moyenne de la variable MntFruits par rapport à la variable Marital_Status
 Marital_Status
Absurd      67.000
Widow       24.971
Divorced    24.392
Single      23.162
Married     22.684
Together    21.341
Alone        4.000
YOLO         3.000
Name: MntFruits, dtype: float64
```

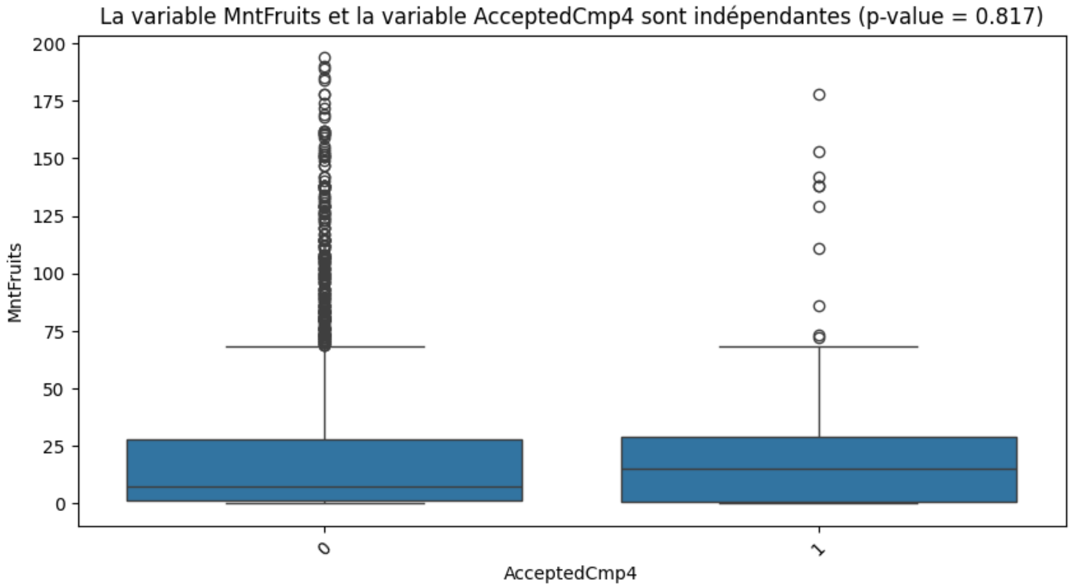



```
 La moyenne de la variable MntFruits par rapport à la variable AcceptedCmp3
 AcceptedCmp3
1    24.000
0    22.563
Name: MntFruits, dtype: float64
```

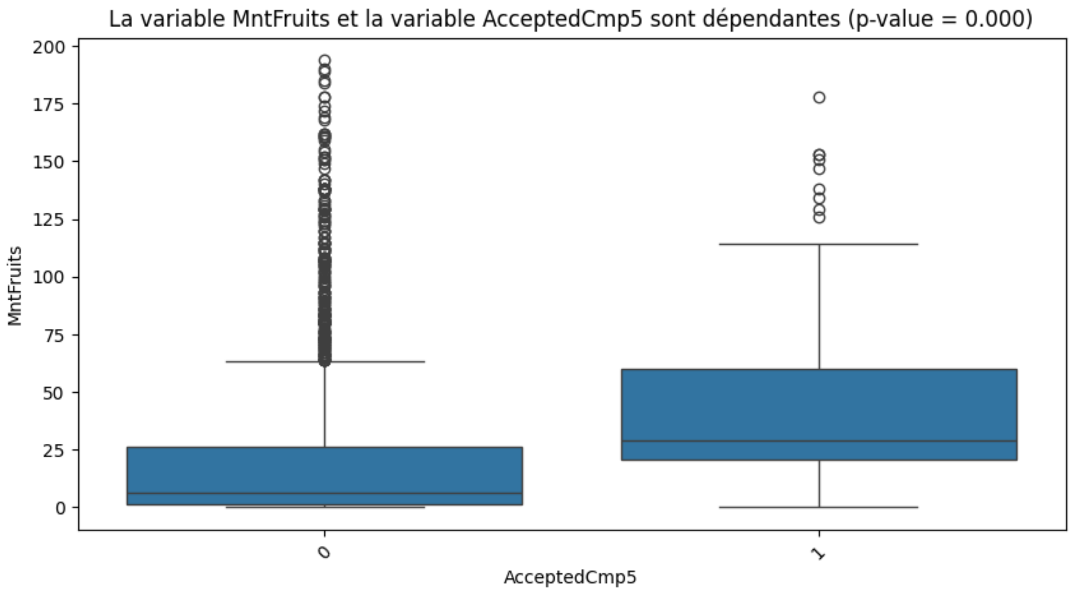



```
 La moyenne de la variable MntFruits par rapport à la variable AcceptedCmp4
 AcceptedCmp4
```

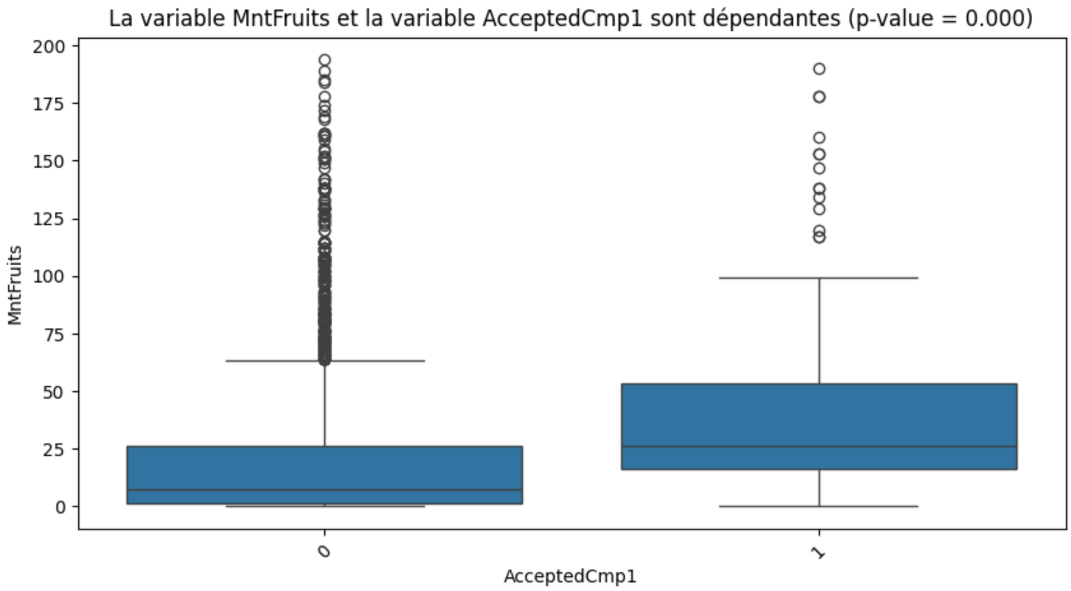
```
1    23.284
0    22.618
Name: MntFruits, dtype: float64
```



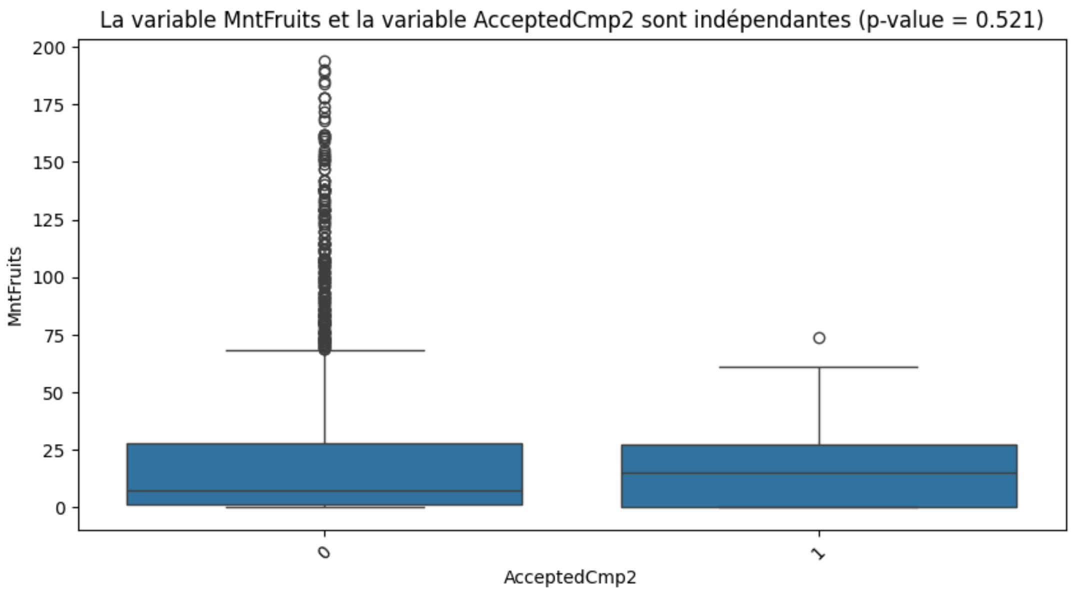

```
 La moyenne de la variable MntFruits par rapport à la variable AcceptedCmp5
 AcceptedCmp5
1    44.516
0    21.268
Name: MntFruits, dtype: float64
```



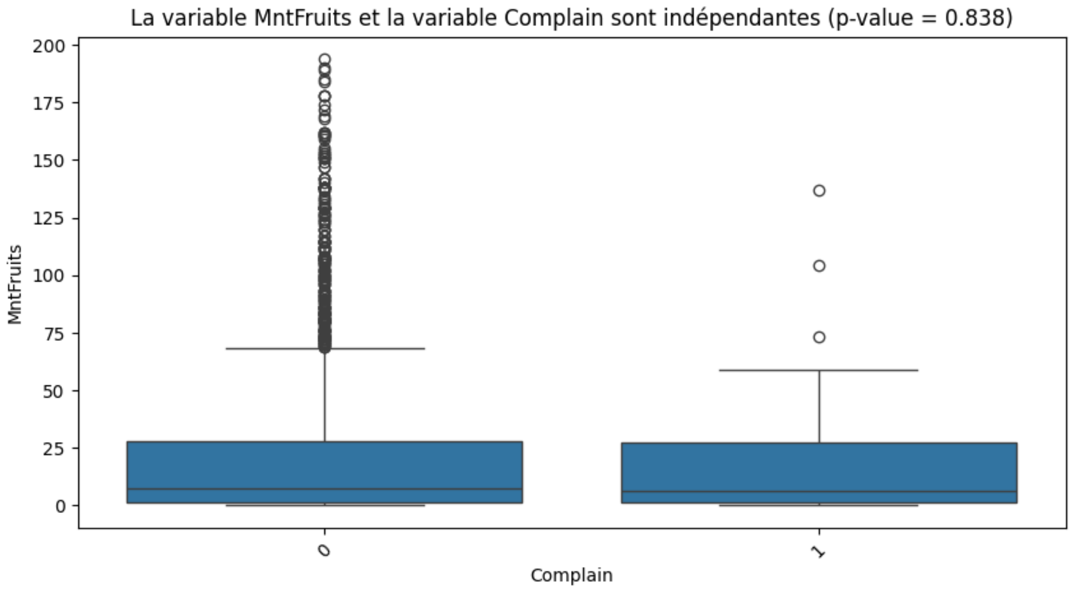
```
 La moyenne de la variable MntFruits par rapport à la variable AcceptedCmp1
 AcceptedCmp1
1    42.296
0    21.545
Name: MntFruits, dtype: float64
```



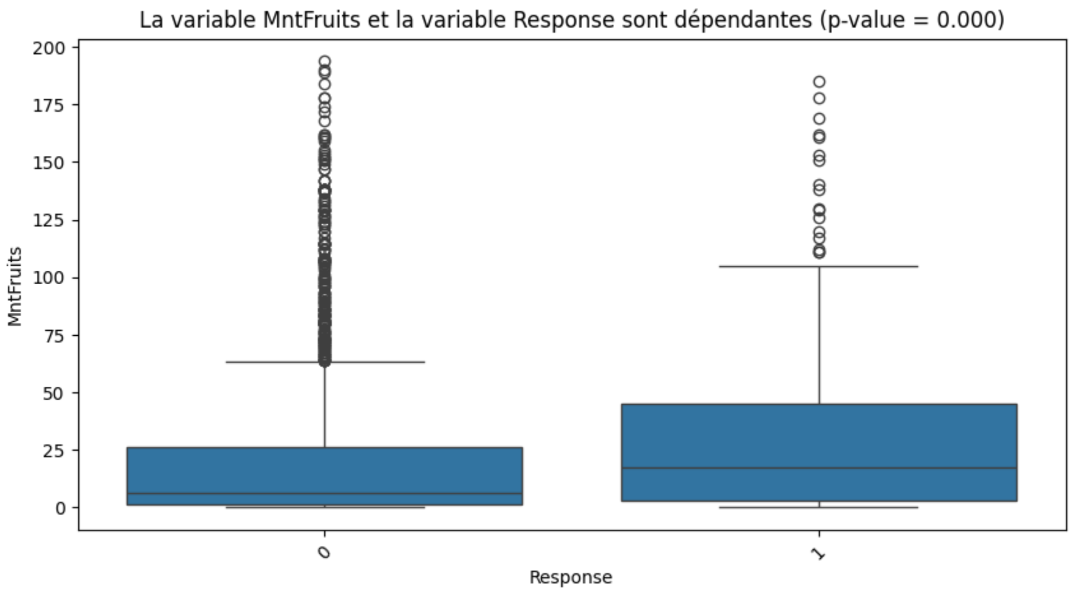

```
 La moyenne de la variable MntFruits par rapport à la variable AcceptedCmp2
 AcceptedCmp2
0    22.721
1    18.444
Name: MntFruits, dtype: float64
```



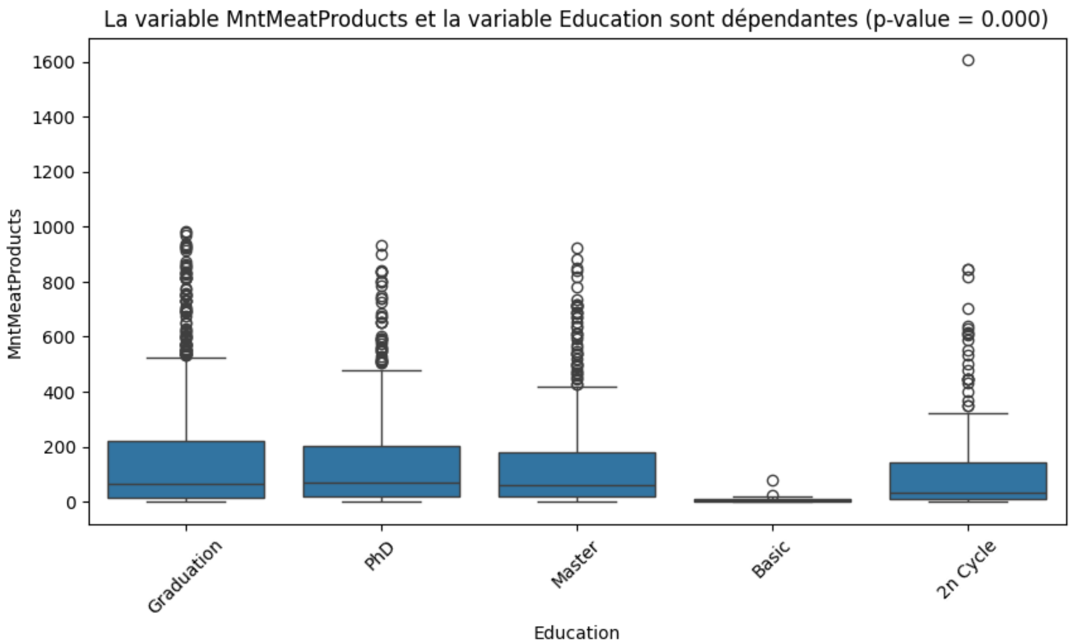
La moyenne de la variable MntFruits par rapport à la variable Complain
Complain
1   24.190
0   22.652
Name: MntFruits, dtype: float64



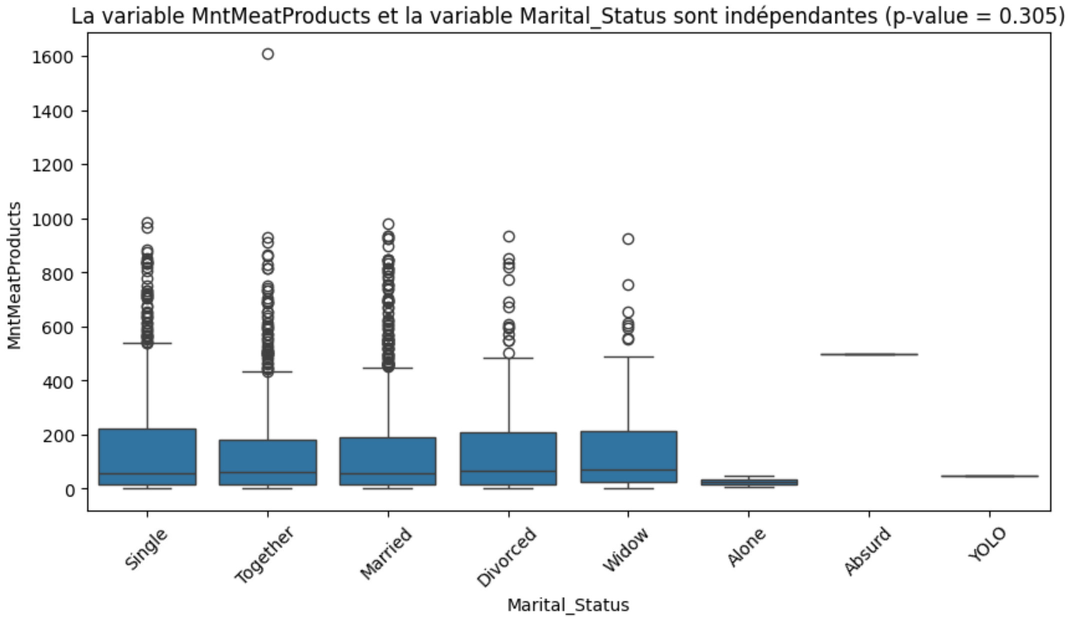
La moyenne de la variable MntFruits par rapport à la variable Response
Response
1   31.017
0   21.307
Name: MntFruits, dtype: float64



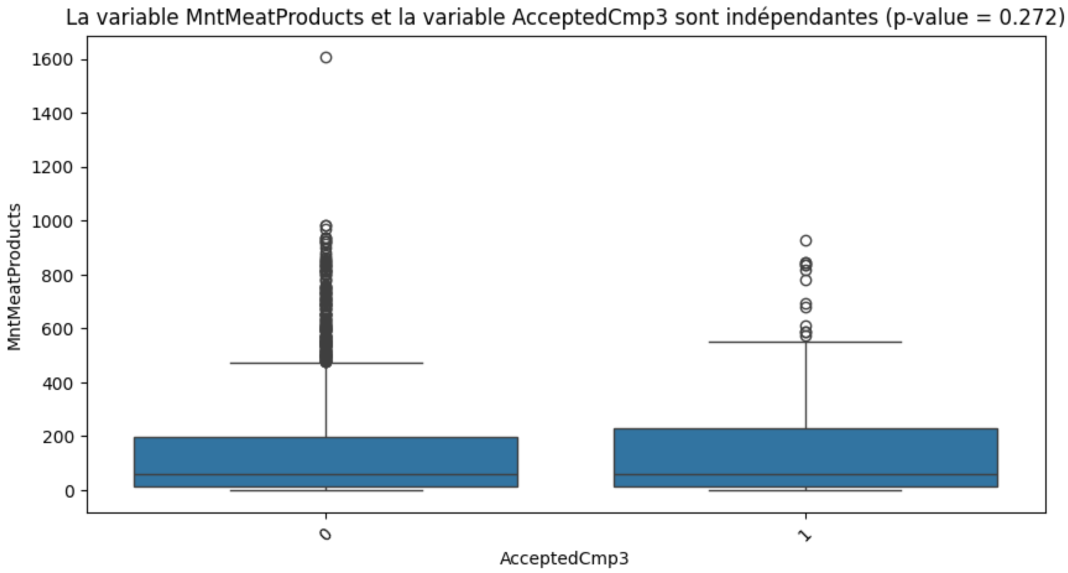
```
 La moyenne de la variable MntMeatProducts par rapport à la variable Education
 Education
Graduation    156.087
PhD           154.472
Master        147.168
2n Cycle      131.162
Basic           9.358
Name: MntMeatProducts, dtype: float64
```




```
 La moyenne de la variable MntMeatProducts par rapport à la variable
Marital_Status
 Marital_Status
Absurd      500.000
Widow       167.493
Single      161.920
Together    146.279
Married     142.810
Divorced    141.545
YOLO         50.000
Alone        26.333
Name: MntMeatProducts, dtype: float64
```

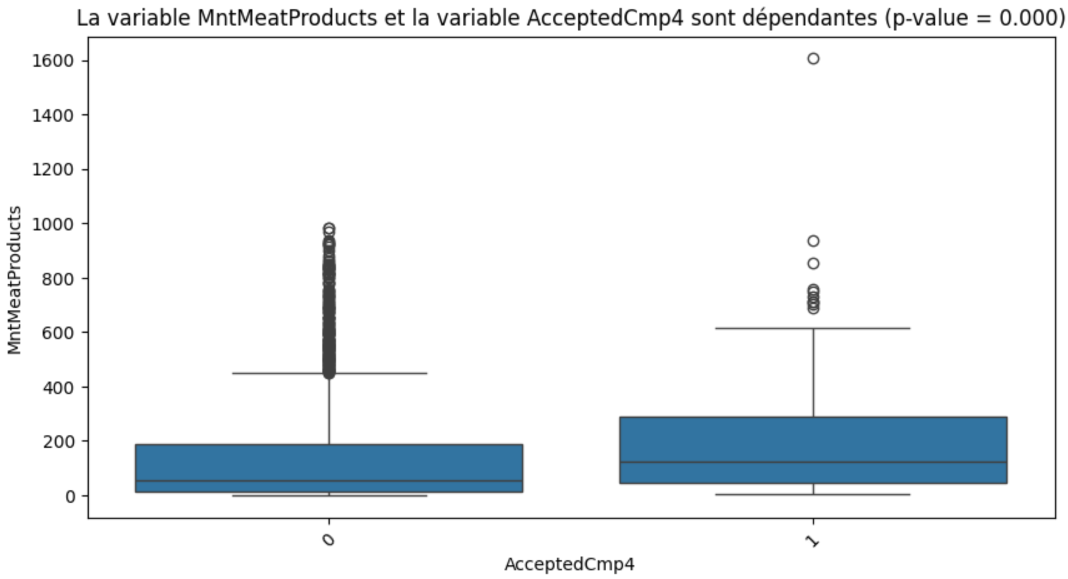



```
 La moyenne de la variable MntMeatProducts par rapport à la variable
AcceptedCmp3
 AcceptedCmp3
1    165.416
0    147.008
Name: MntMeatProducts, dtype: float64
```

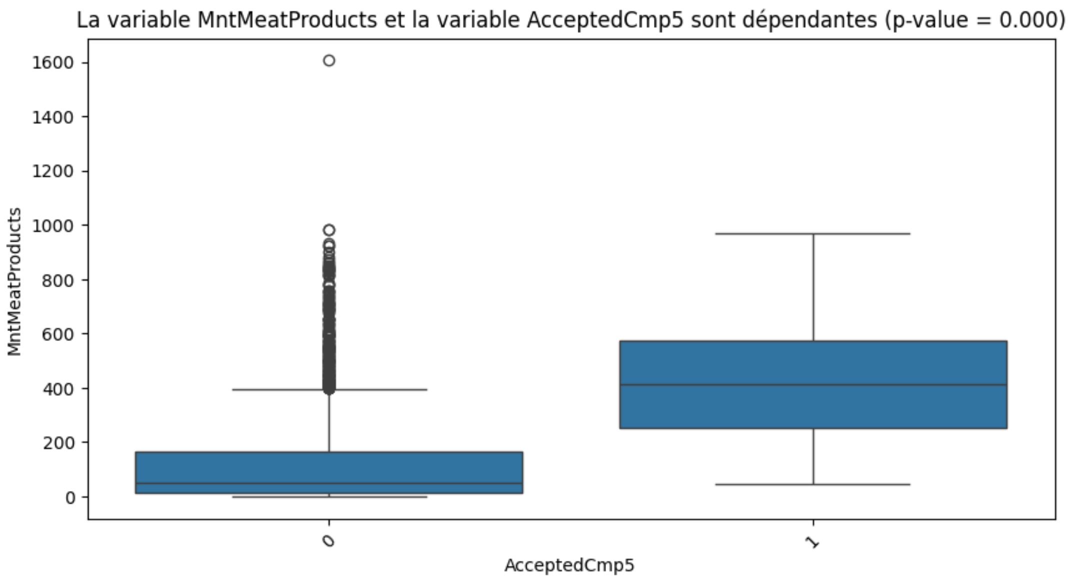



```
 La moyenne de la variable MntMeatProducts par rapport à la variable
AcceptedCmp4
```

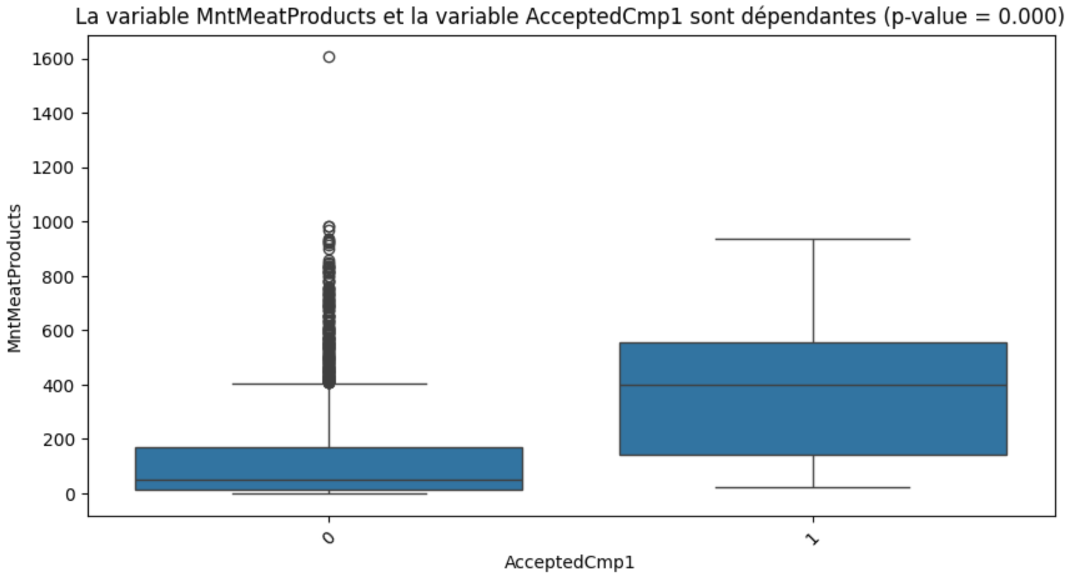
```
 AcceptedCmp4
1    215.600
0    143.056
Name: MntMeatProducts, dtype: float64
```



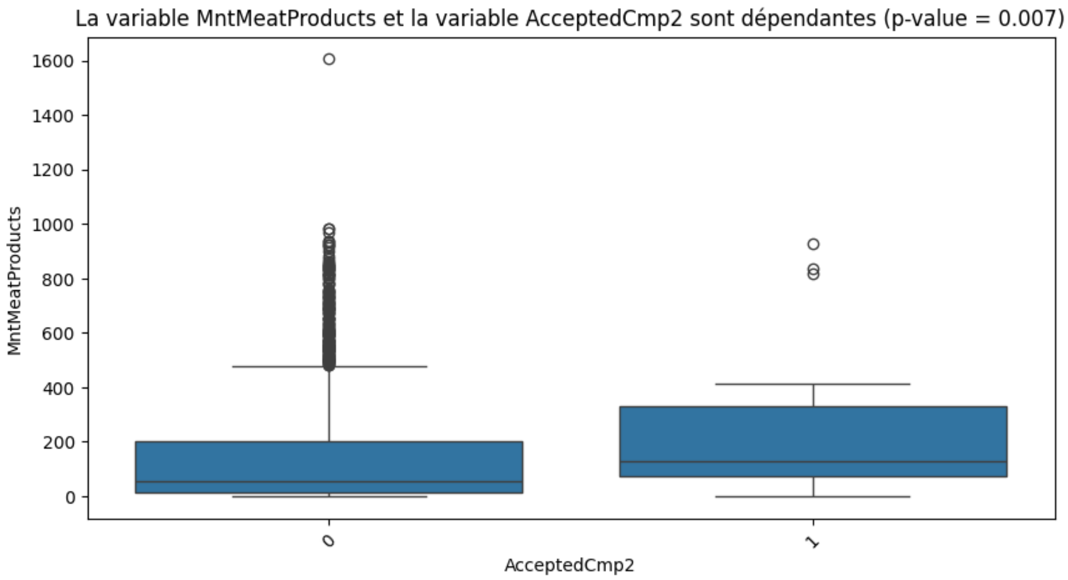

```
 La moyenne de la variable MntMeatProducts par rapport à la variable
AcceptedCmp5
 AcceptedCmp5
1    429.398
0    130.352
Name: MntMeatProducts, dtype: float64
```



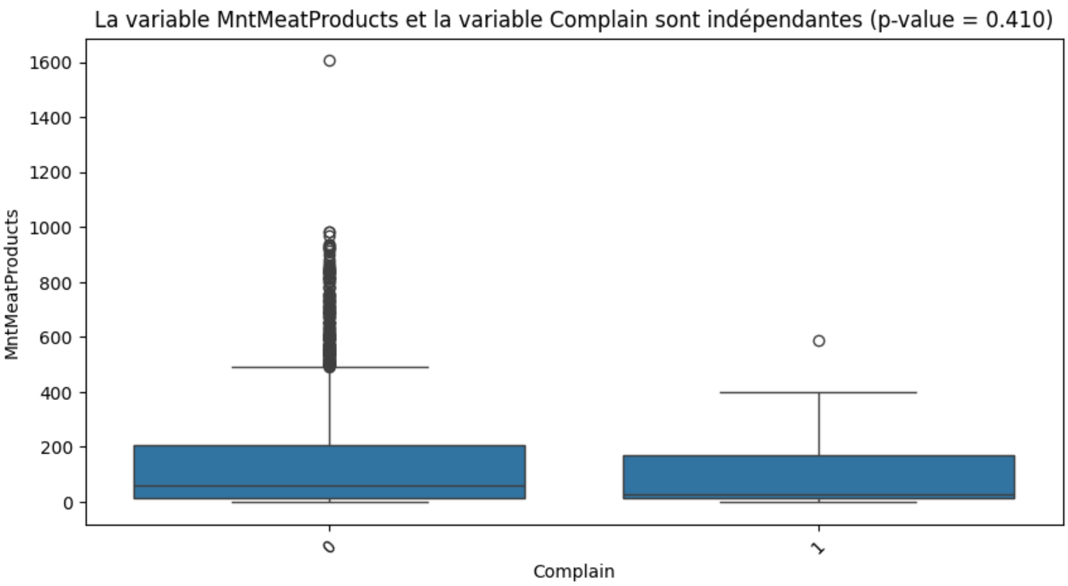
```
 La moyenne de la variable MntMeatProducts par rapport à la variable
AcceptedCmp1
 AcceptedCmp1
1    403.713
0    133.751
Name: MntMeatProducts, dtype: float64
```



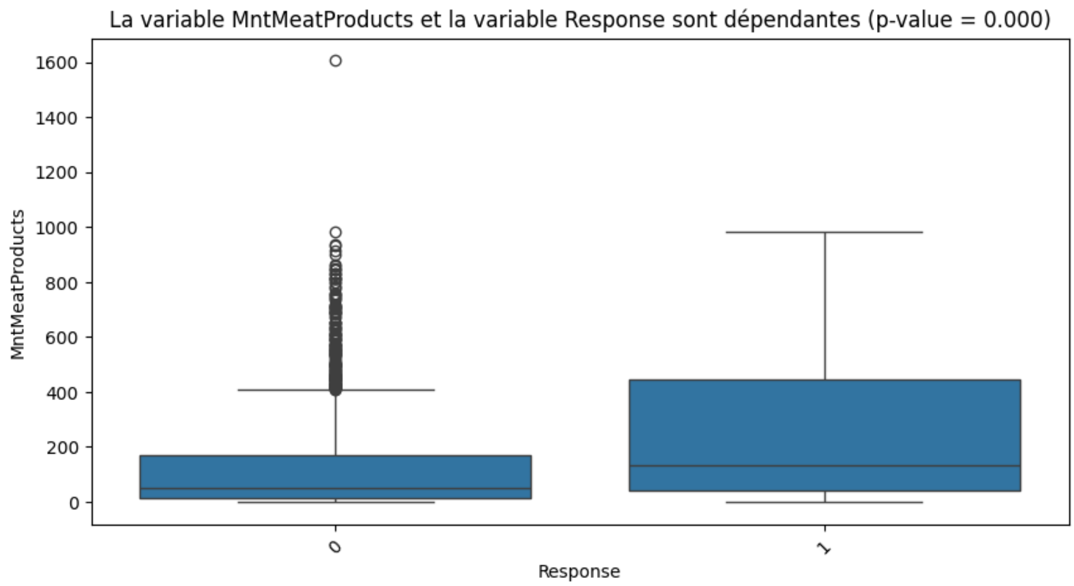

```
 La moyenne de la variable MntMeatProducts par rapport à la variable
AcceptedCmp2
 AcceptedCmp2
1    251.852
0    147.010
Name: MntMeatProducts, dtype: float64
```



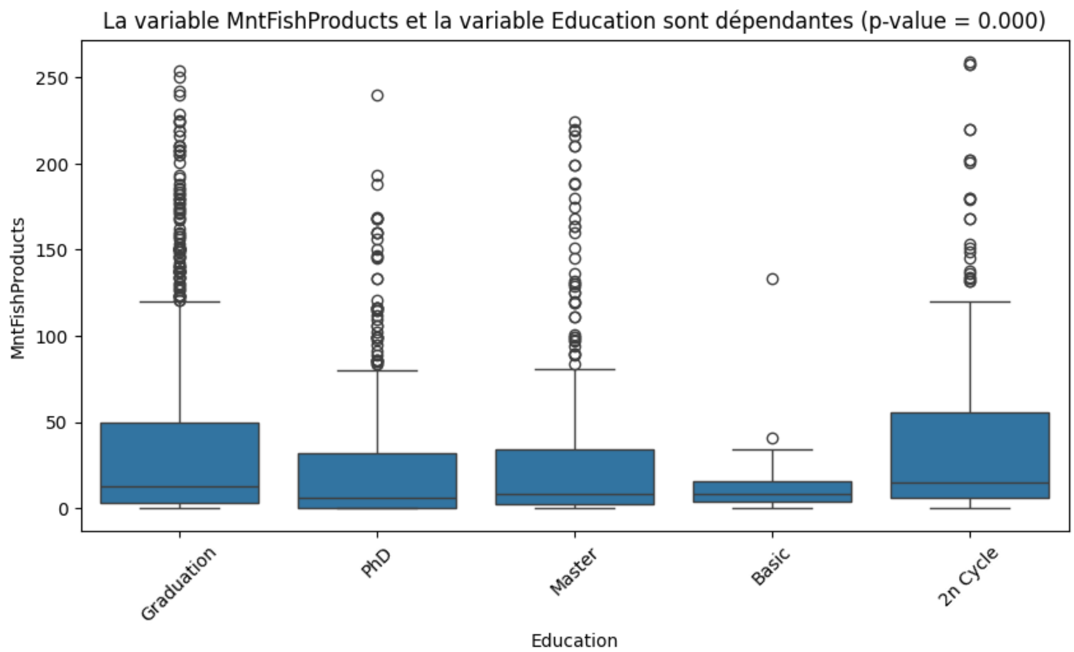
```
 La moyenne de la variable MntMeatProducts par rapport à la variable Complain
 Complain
0    148.697
1    112.476
Name: MntMeatProducts, dtype: float64
```



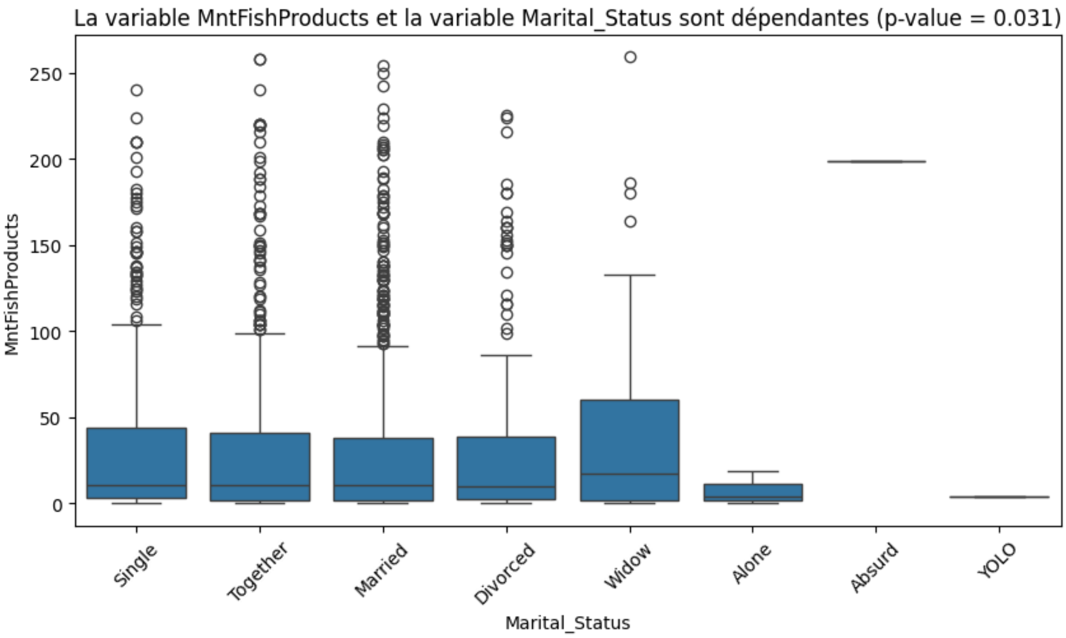

```
 La moyenne de la variable MntMeatProducts par rapport à la variable Response
 Response
1    261.312
0    129.943
Name: MntMeatProducts, dtype: float64
```



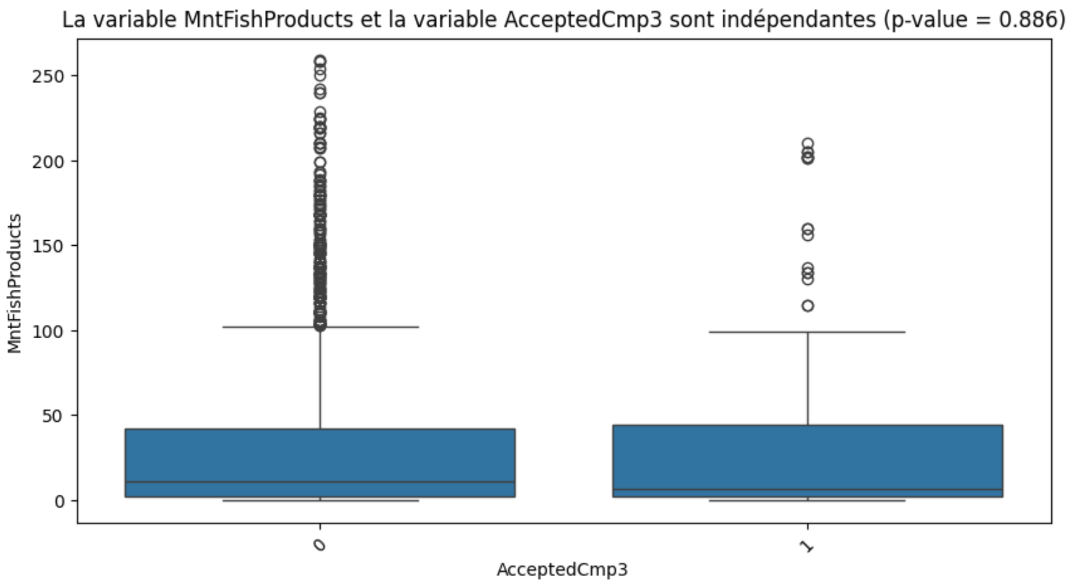

```
 La moyenne de la variable MntFishProducts par rapport à la variable Education
 Education
2n Cycle      42.230
Graduation    36.384
Master        29.321
PhD           24.991
Basic         13.453
Name: MntFishProducts, dtype: float64
```



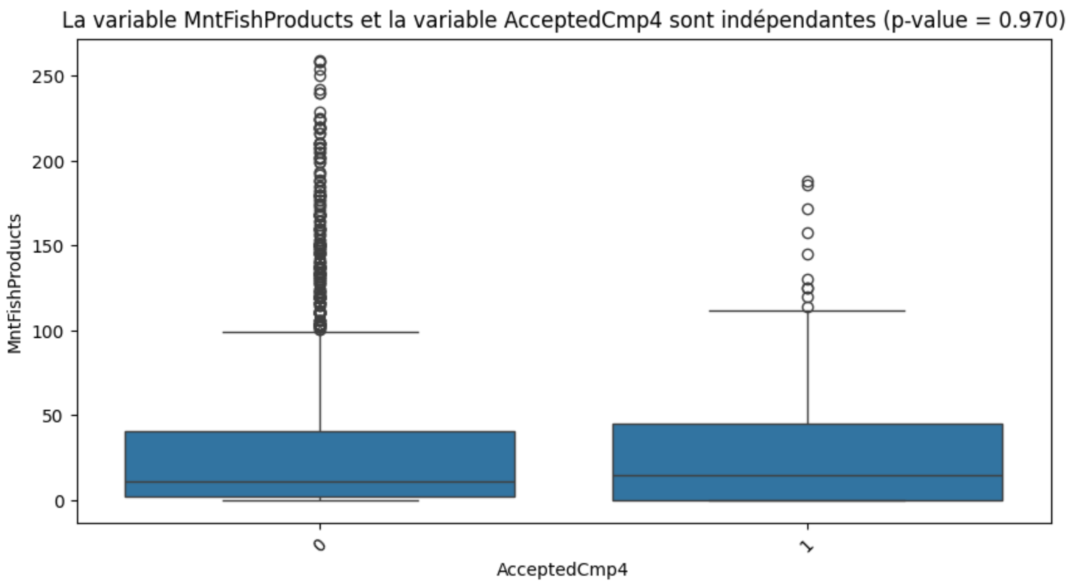
```
 La moyenne de la variable MntFishProducts par rapport à la variable
Marital_Status
 Marital_Status
Absurd      199.000
Widow        40.232
Together     33.184
Single       32.568
Divorced     32.018
Married      31.824
Alone         7.667
YOLO          4.000
Name: MntFishProducts, dtype: float64
```




```
 La moyenne de la variable MntFishProducts par rapport à la variable
AcceptedCmp3
 AcceptedCmp3
1    33.182
0    32.602
Name: MntFishProducts, dtype: float64
```

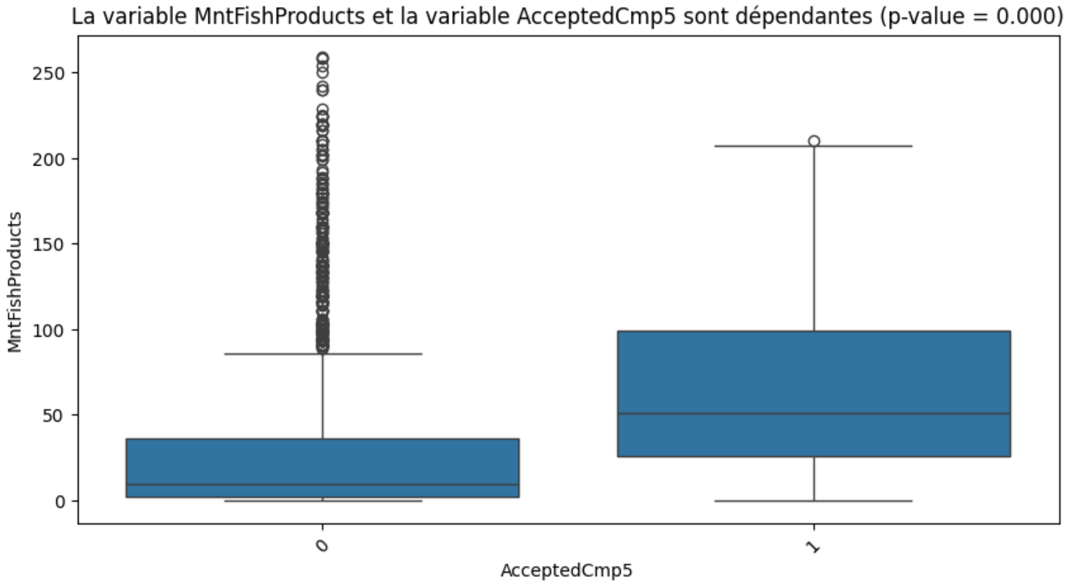



```
 La moyenne de la variable MntFishProducts par rapport à la variable
AcceptedCmp4
 AcceptedCmp4
0    32.655
1    32.503
Name: MntFishProducts, dtype: float64
```

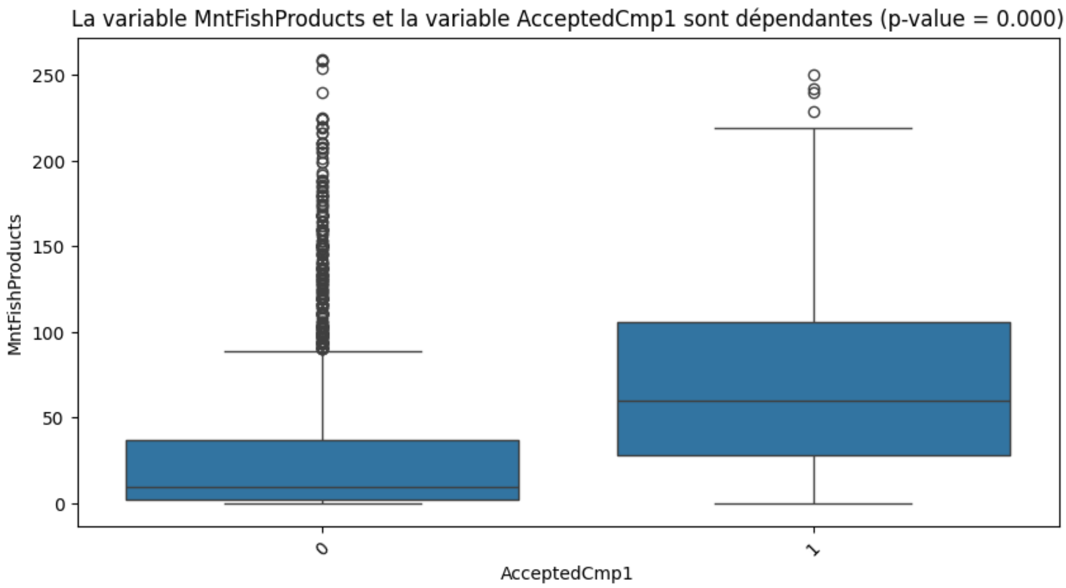



```
 La moyenne de la variable MntFishProducts par rapport à la variable
AcceptedCmp5
 AcceptedCmp5
```

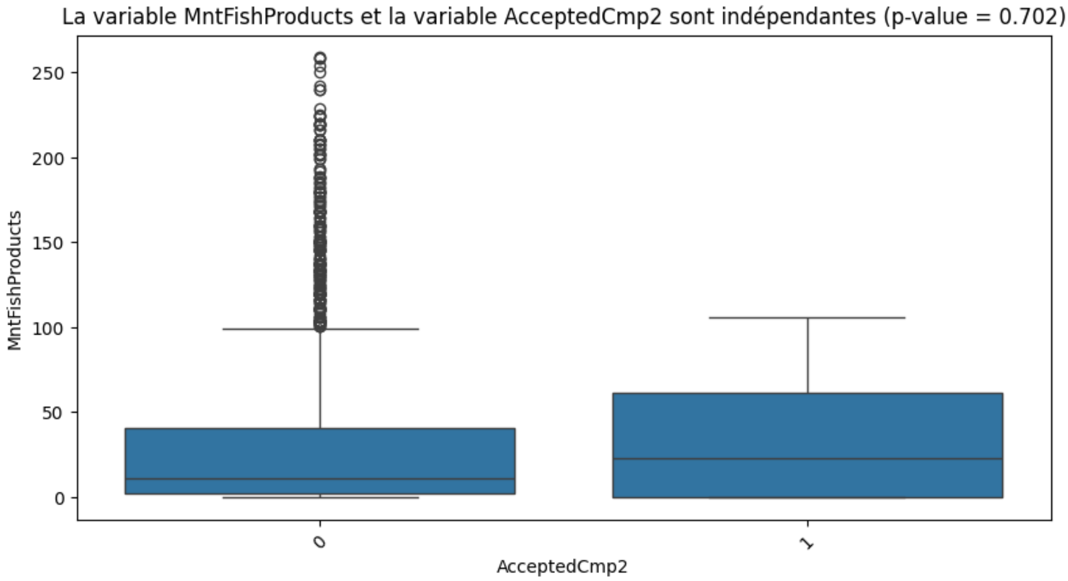
```
1    63.484
0    30.670
Name: MntFishProducts, dtype: float64
```



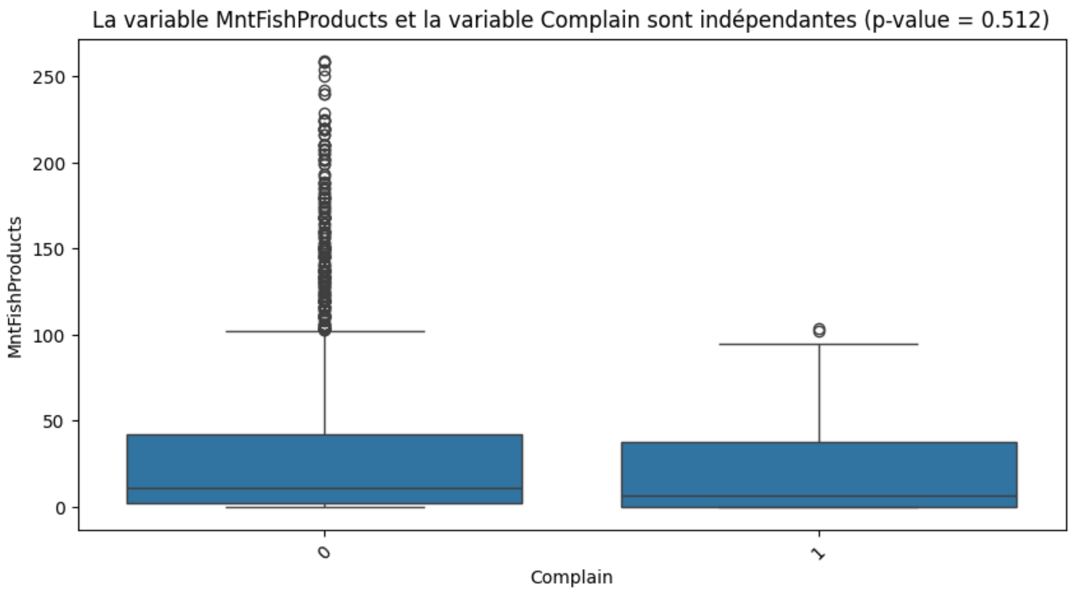

```
 La moyenne de la variable MntFishProducts par rapport à la variable
AcceptedCmp1
 AcceptedCmp1
1    75.243
0    30.210
Name: MntFishProducts, dtype: float64
```



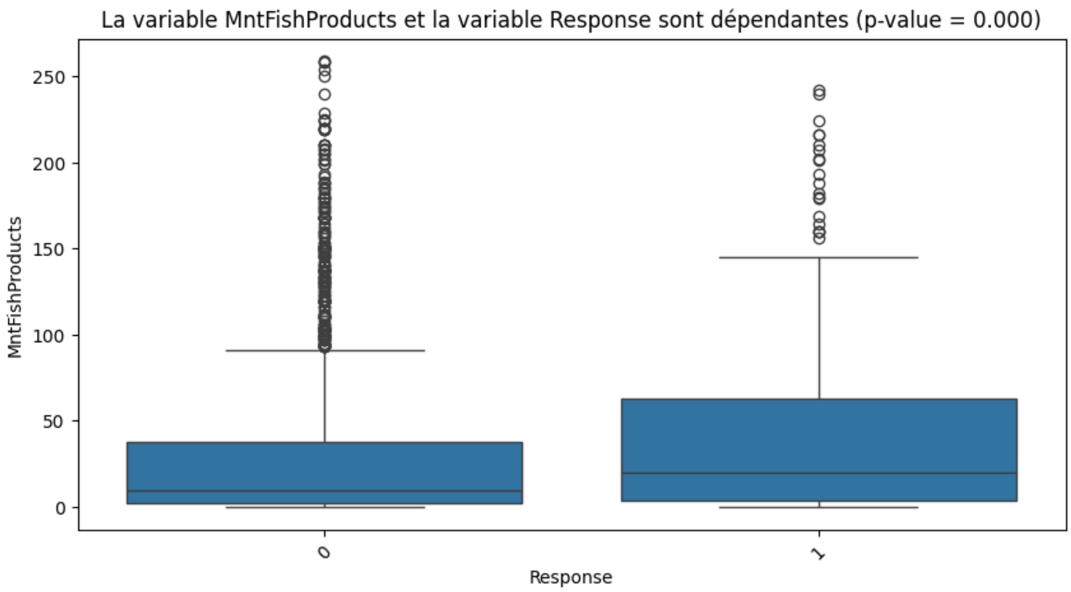
```
 La moyenne de la variable MntFishProducts par rapport à la variable
AcceptedCmp2
 AcceptedCmp2
1    36.185
0    32.598
Name: MntFishProducts, dtype: float64
```



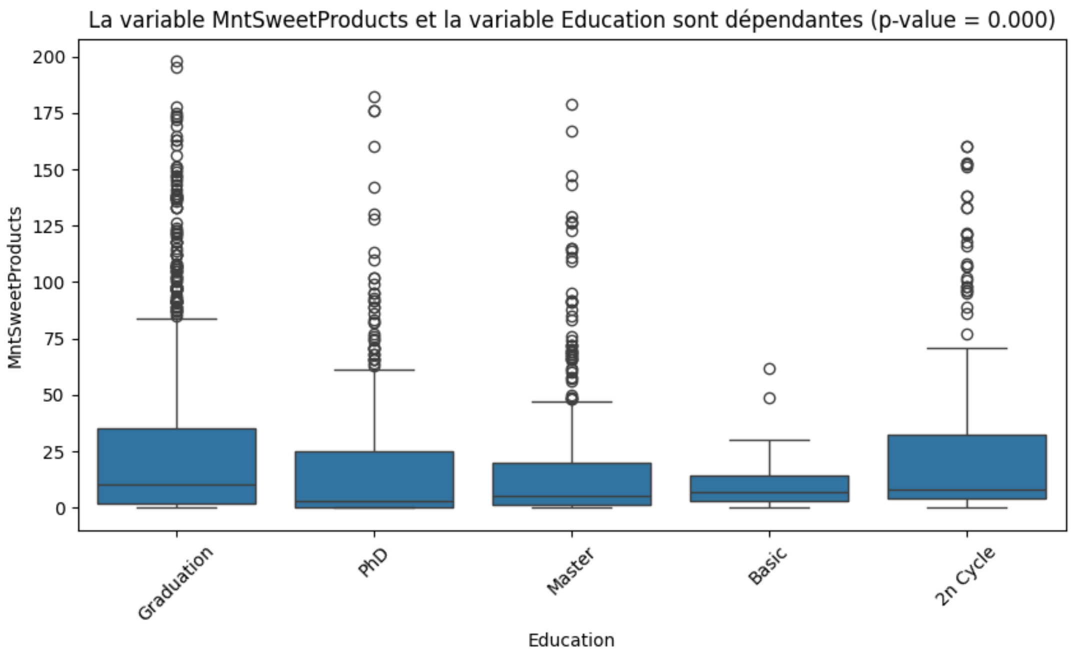

```
 La moyenne de la variable MntFishProducts par rapport à la variable Complain
 Complain
0    32.712
1    25.762
Name: MntFishProducts, dtype: float64
```



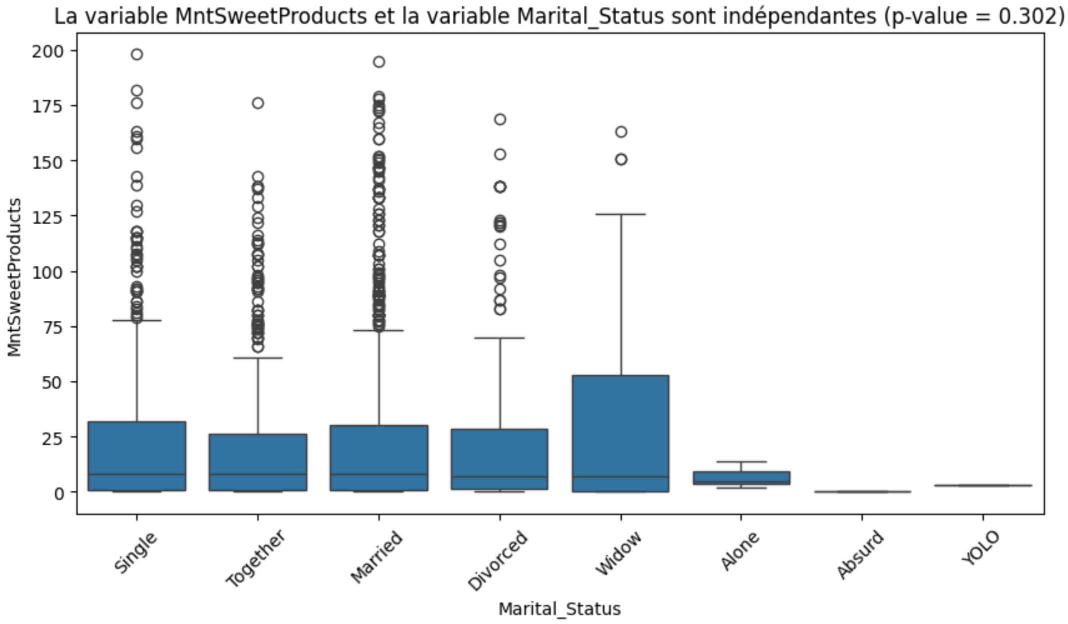
```
 La moyenne de la variable MntFishProducts par rapport à la variable Response
 Response
1    42.815
0    30.987
Name: MntFishProducts, dtype: float64
```




```
 La moyenne de la variable MntSweetProducts par rapport à la variable Education
 Education
2n Cycle      28.262
Graduation    26.607
Master        18.813
PhD           17.906
Basic          9.906
Name: MntSweetProducts, dtype: float64
```

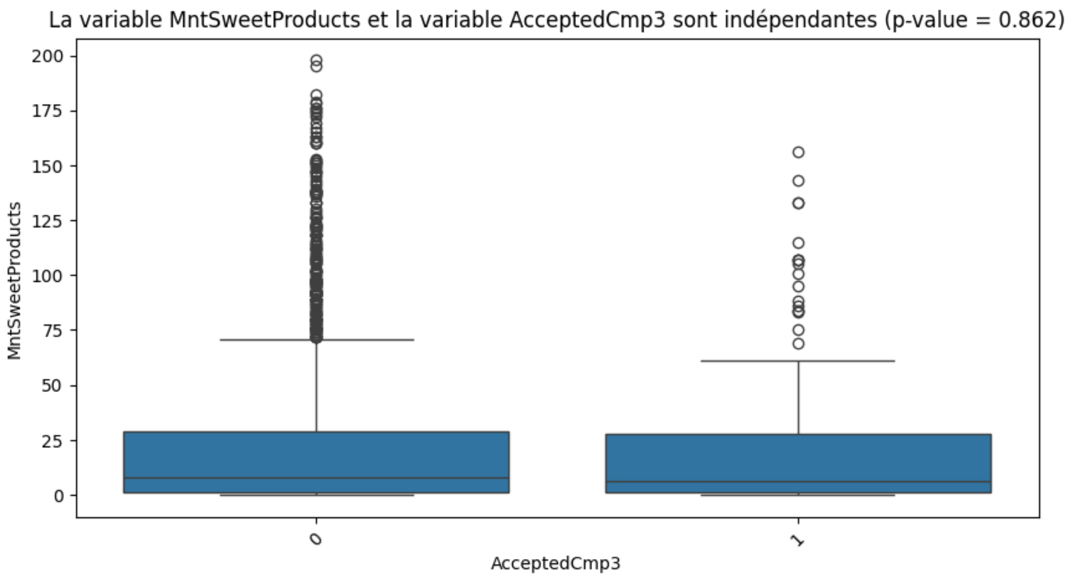



```
 La moyenne de la variable MntSweetProducts par rapport à la variable
Marital_Status
 Marital_Status
Widow       30.565
Married     24.049
Single      23.665
Divorced    22.806
Together    20.639
Alone        7.000
YOLO         3.000
Absurd       0.000
Name: MntSweetProducts, dtype: float64
```



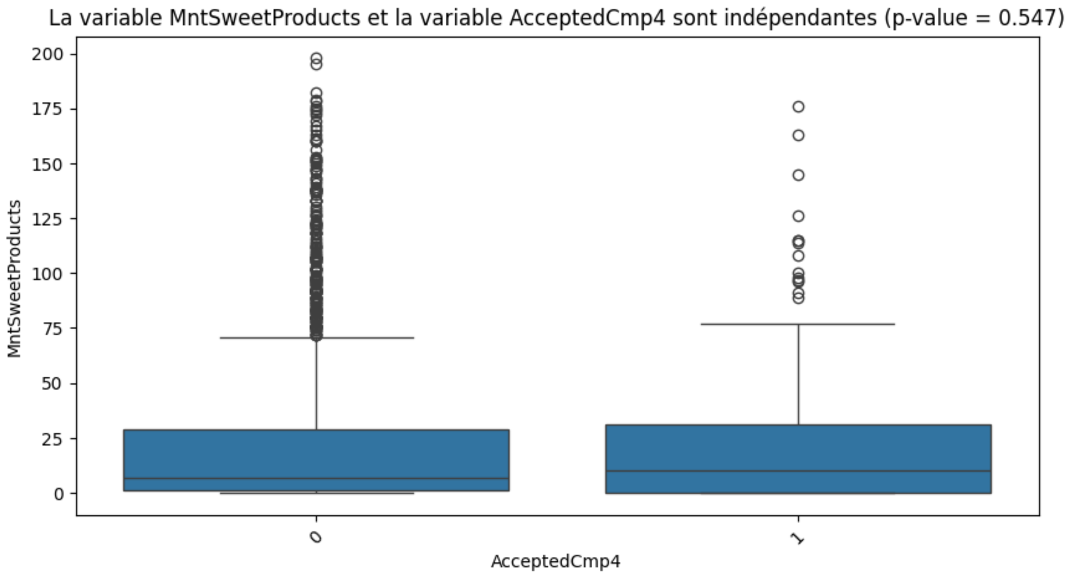
La moyenne de la variable MntSweetProducts par rapport à la variable AcceptedCmp3

```
 AcceptedCmp3
0    23.151
1    22.643
Name: MntSweetProducts, dtype: float64
```

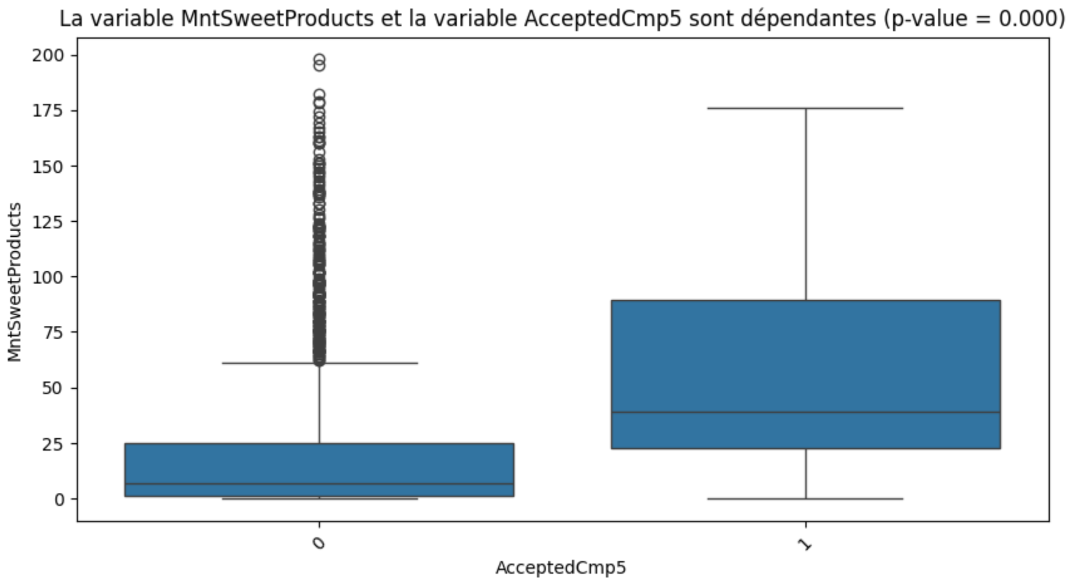


La moyenne de la variable MntSweetProducts par rapport à la variable AcceptedCmp4

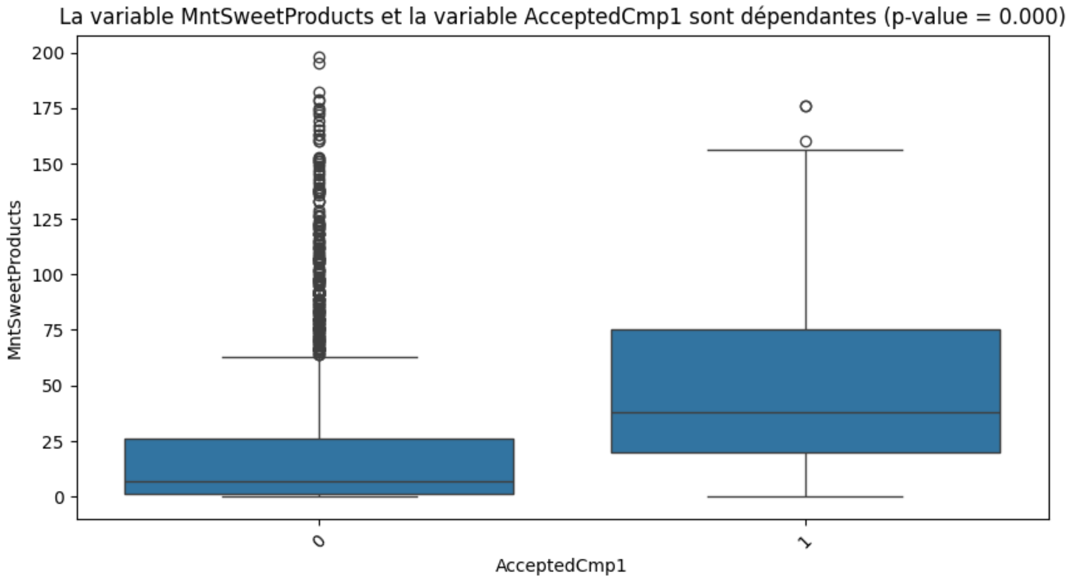
```
 AcceptedCmp4
1    24.742
0    22.987
Name: MntSweetProducts, dtype: float64
```



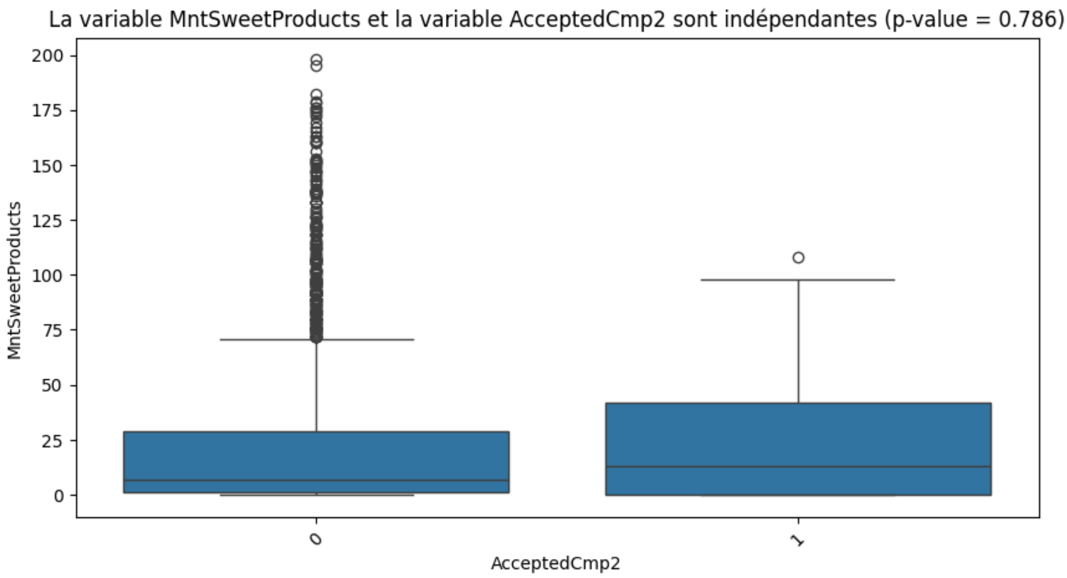

```
 La moyenne de la variable MntSweetProducts par rapport à la variable
AcceptedCmp5
 AcceptedCmp5
1    56.867
0    20.954
Name: MntSweetProducts, dtype: float64
```



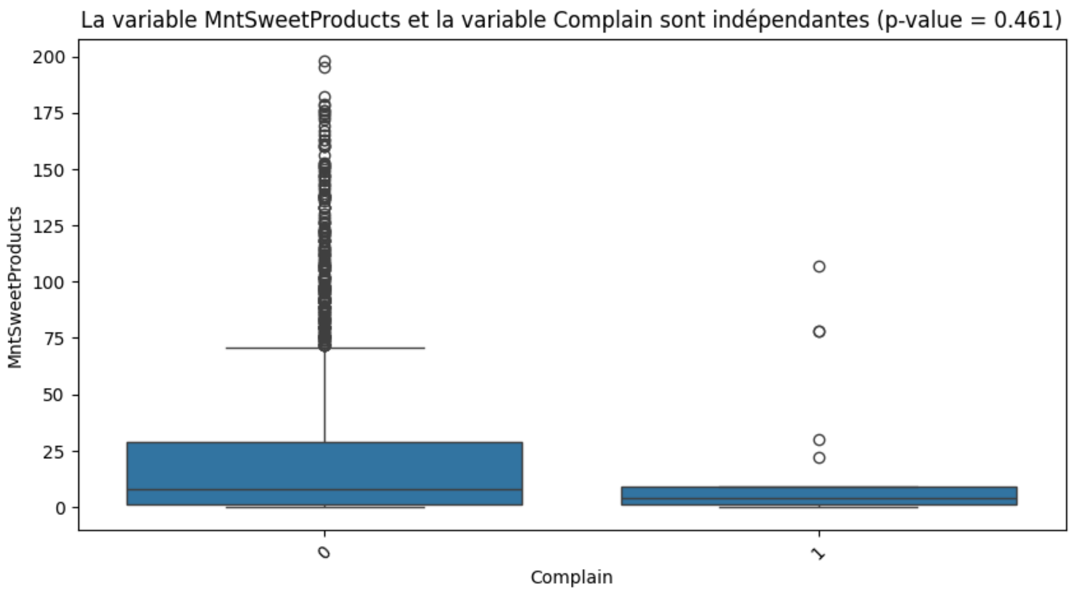
```
 La moyenne de la variable MntSweetProducts par rapport à la variable
AcceptedCmp1
 AcceptedCmp1
1    52.235
0    21.451
Name: MntSweetProducts, dtype: float64
```



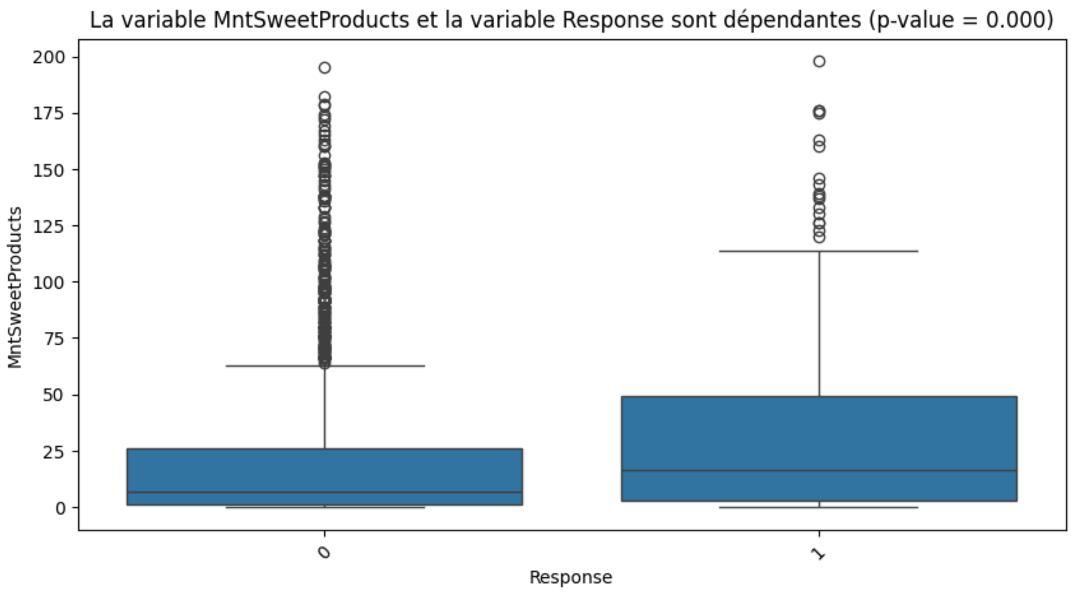

```
 La moyenne de la variable MntSweetProducts par rapport à la variable
AcceptedCmp2
 AcceptedCmp2
1    24.926
0    23.091
Name: MntSweetProducts, dtype: float64
```



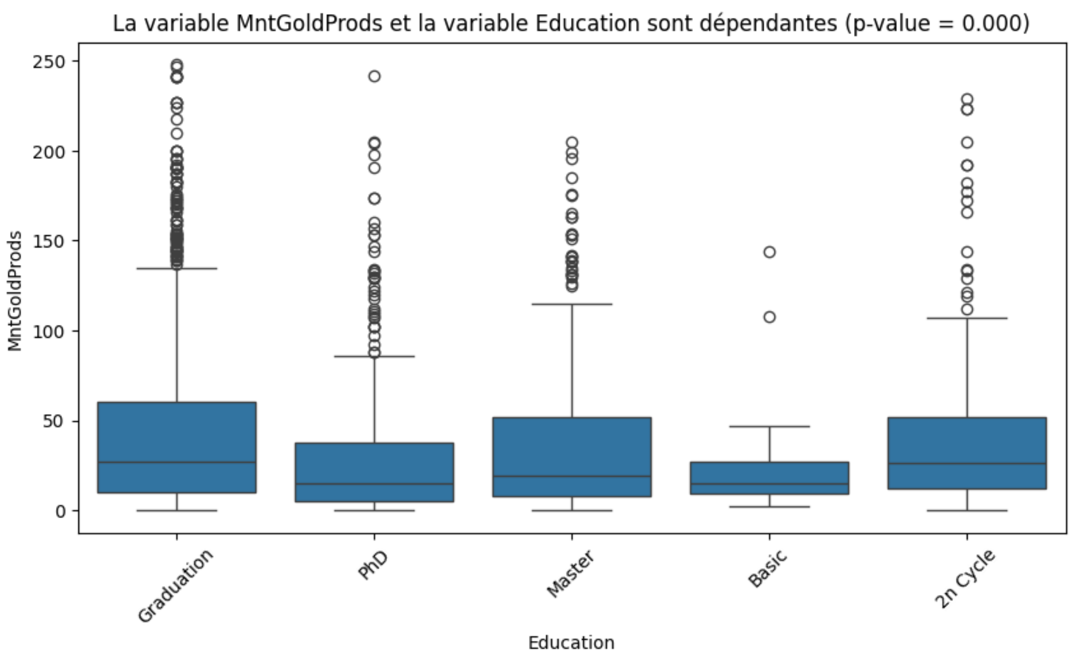
```
 La moyenne de la variable MntSweetProducts par rapport à la variable Complain
 Complain
0    23.170
1    17.524
Name: MntSweetProducts, dtype: float64
```



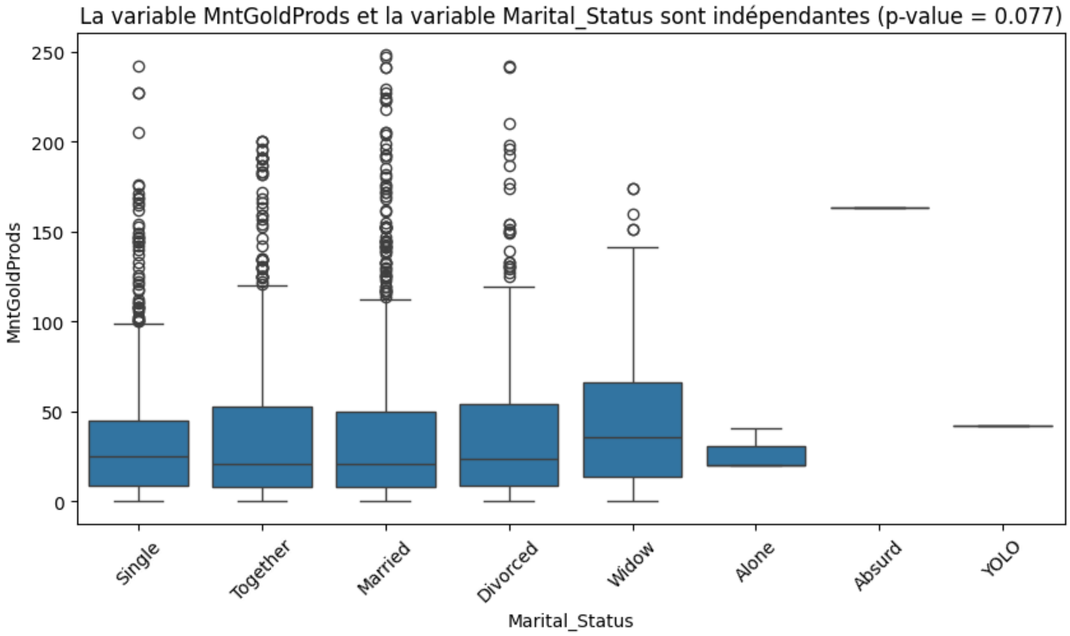

```
 La moyenne de la variable MntSweetProducts par rapport à la variable Response
 Response
1    32.846
0    21.530
Name: MntSweetProducts, dtype: float64
```



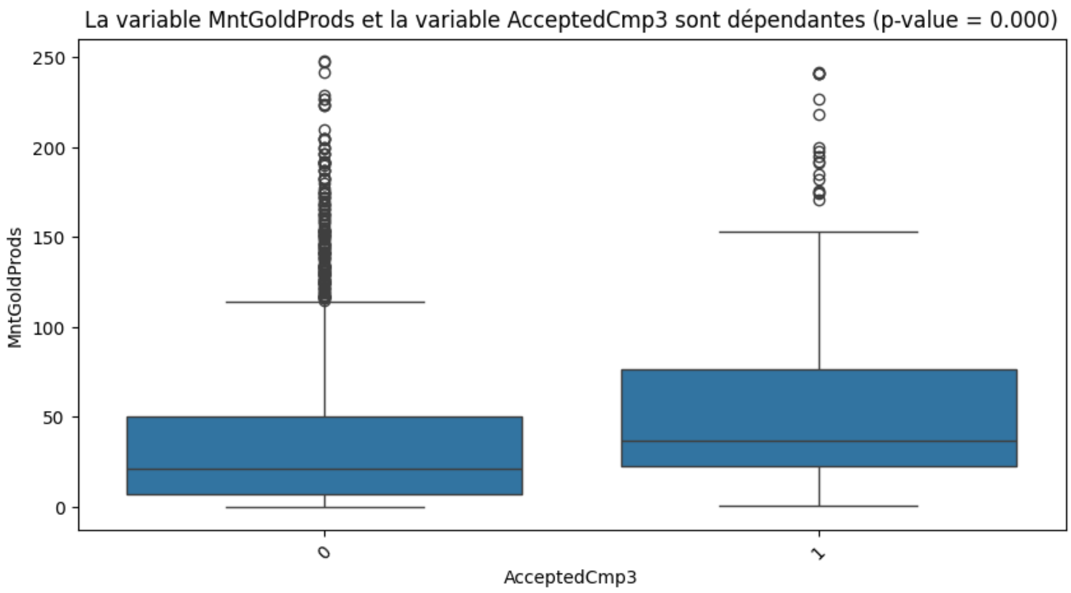

```
 La moyenne de la variable MntGoldProds par rapport à la variable Education
 Education
Graduation    45.638
2n Cycle      41.654
Master        37.497
PhD           29.853
Basic         22.698
Name: MntGoldProds, dtype: float64
```



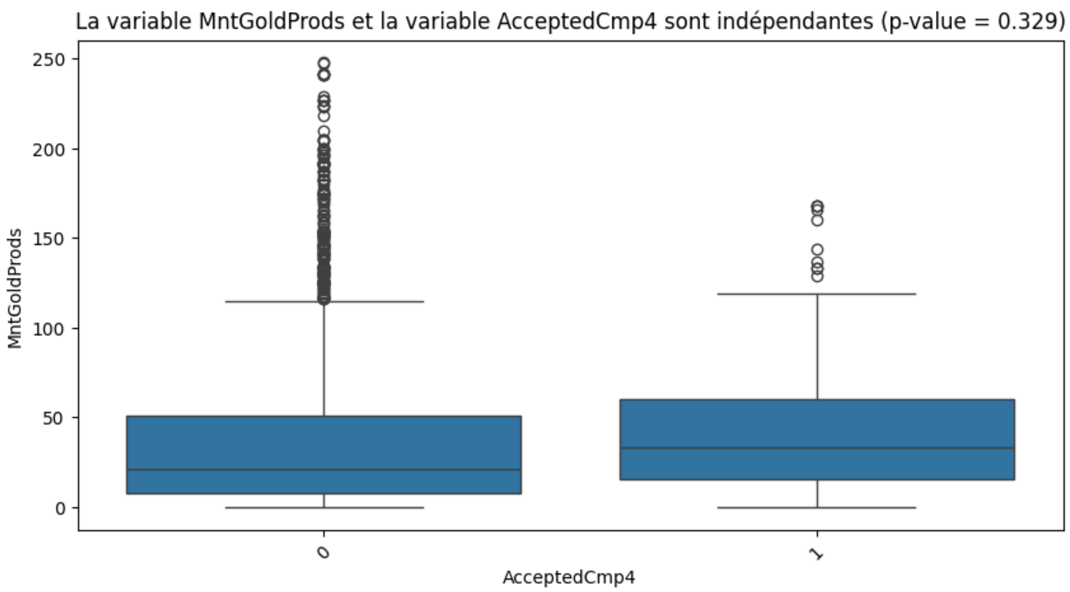
```
 La moyenne de la variable MntGoldProds par rapport à la variable Marital_Status
 Marital_Status
Absurd      163.000
Widow        50.957
Divorced     42.613
YOLO         42.000
Married      39.481
Together     39.102
Single       38.268
Alone        27.000
Name: MntGoldProds, dtype: float64
```




```
 La moyenne de la variable MntGoldProds par rapport à la variable AcceptedCmp3
 AcceptedCmp3
1    62.955
0    38.067
Name: MntGoldProds, dtype: float64
```

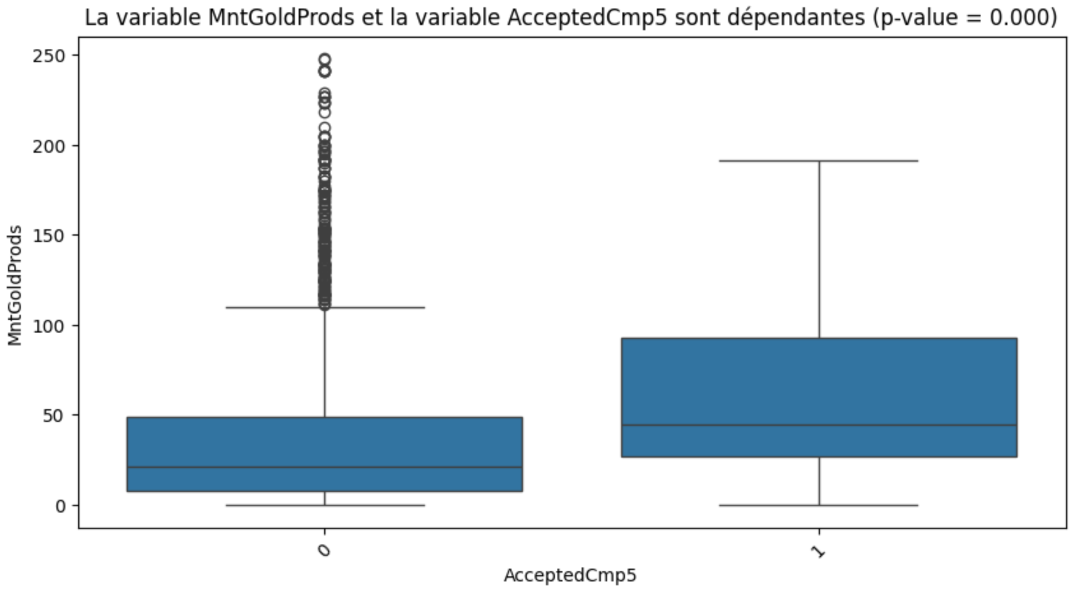



```
 La moyenne de la variable MntGoldProds par rapport à la variable AcceptedCmp4
 AcceptedCmp4
1    43.361
0    39.594
Name: MntGoldProds, dtype: float64
```

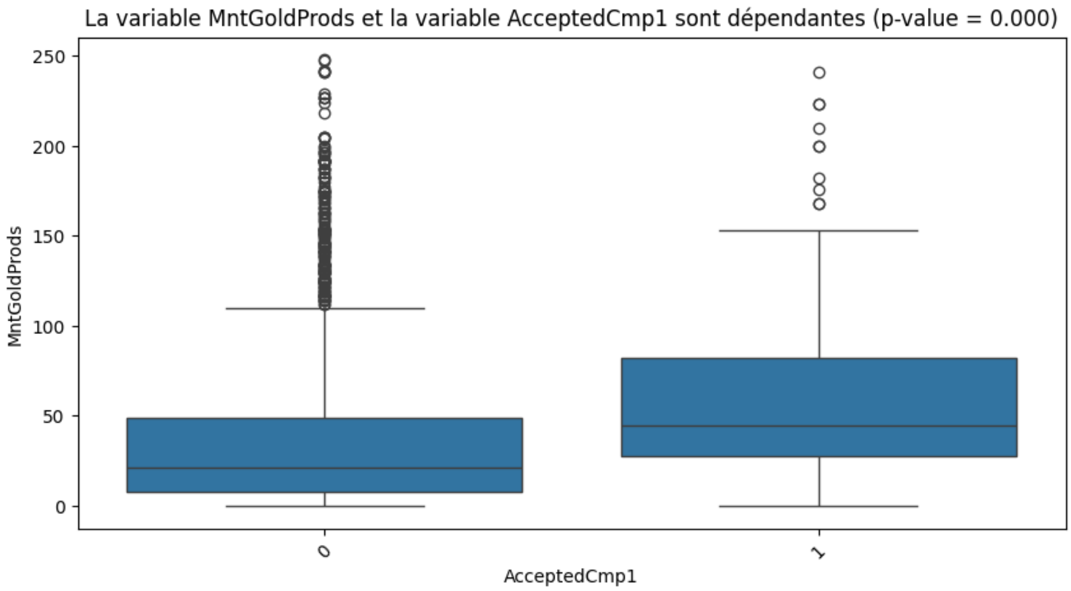



```
 La moyenne de la variable MntGoldProds par rapport à la variable AcceptedCmp5
 AcceptedCmp5
1    62.984
0    38.388
```

Name: MntGoldProds, dtype: float64

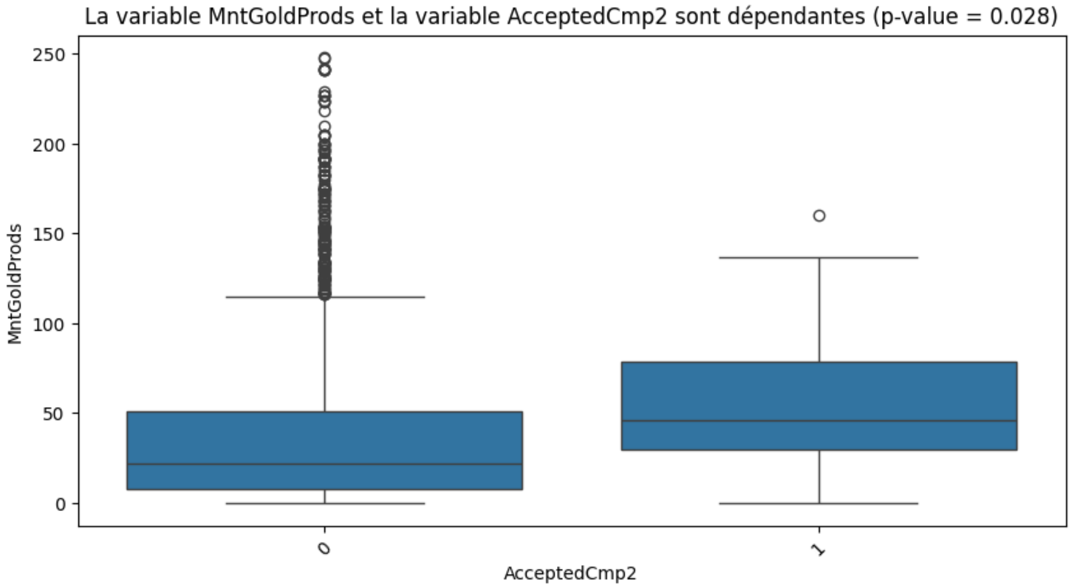


La moyenne de la variable MntGoldProds par rapport à la variable AcceptedCmp1
AcceptedCmp1
1   64.139
0   38.481
Name: MntGoldProds, dtype: float64

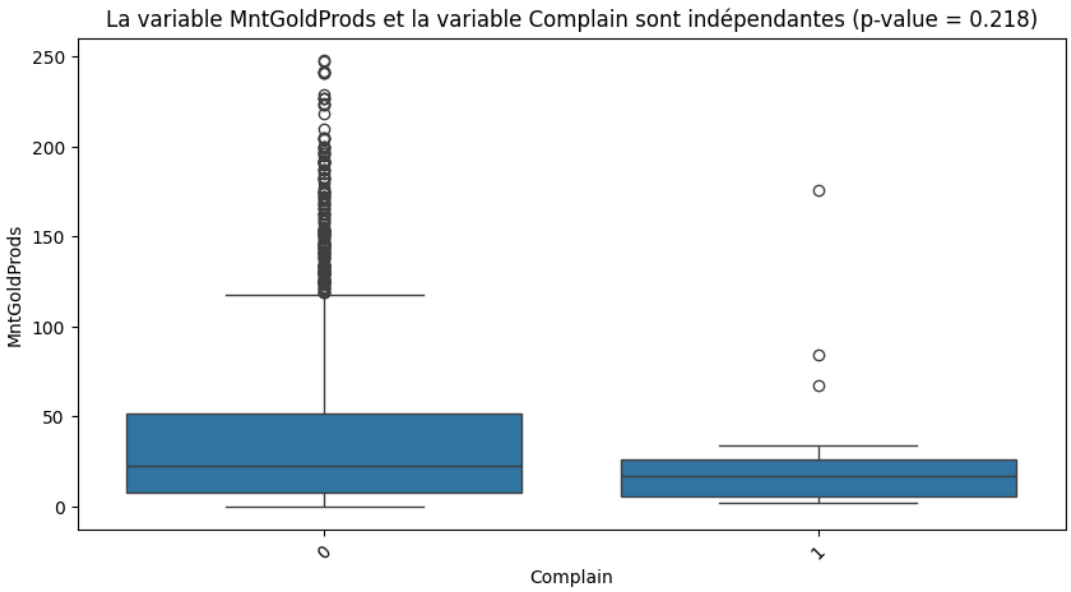


La moyenne de la variable MntGoldProds par rapport à la variable AcceptedCmp2
AcceptedCmp2

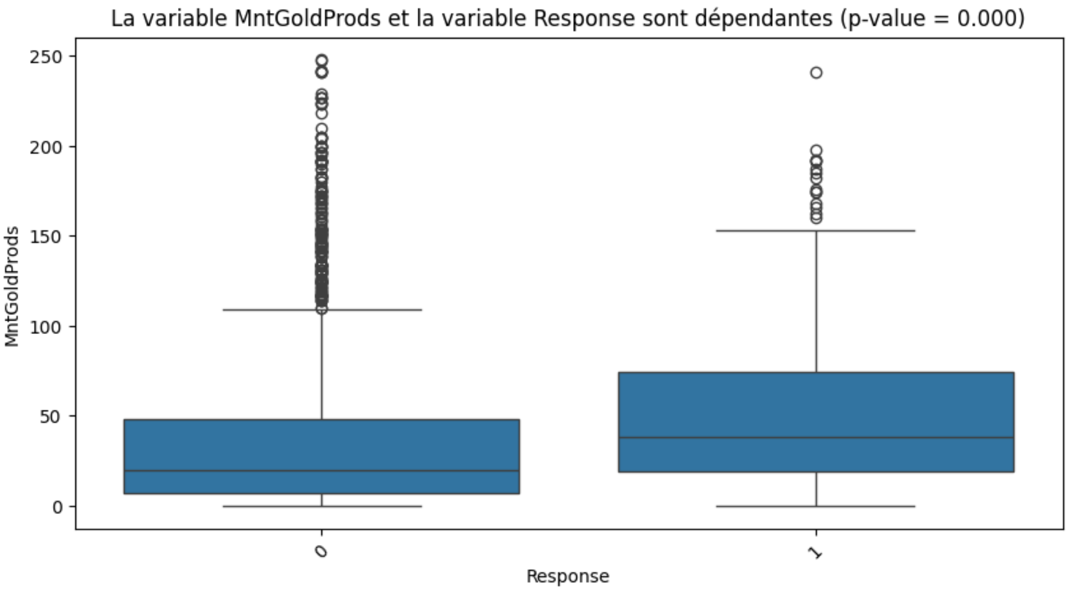
```
1    59.259
0    39.619
Name: MntGoldProds, dtype: float64
```



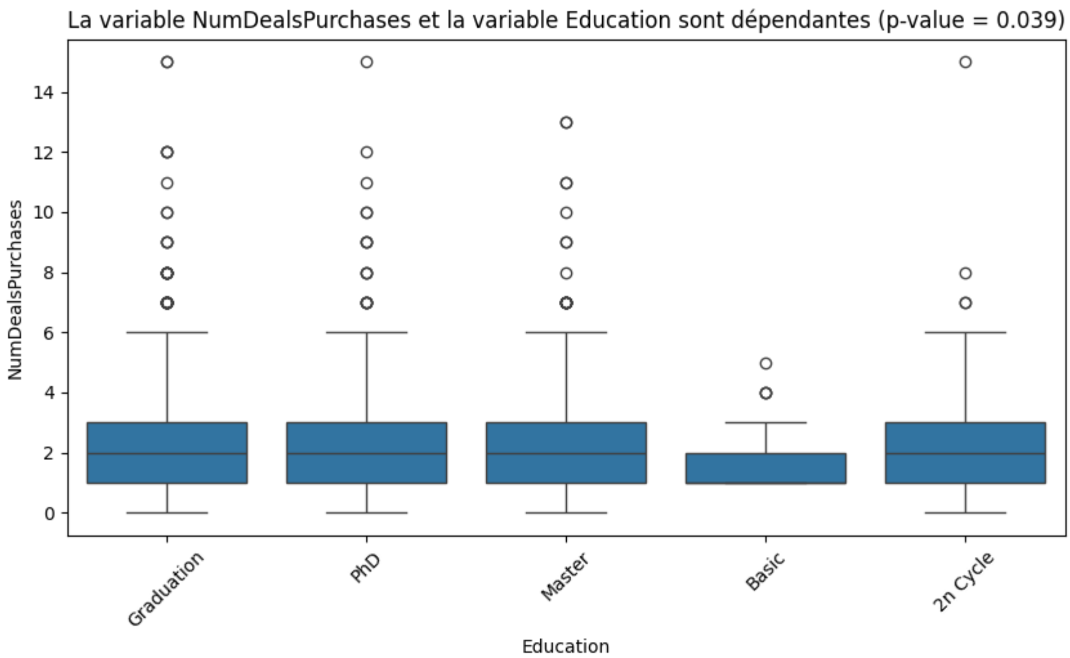

```
 La moyenne de la variable MntGoldProds par rapport à la variable Complain
 Complain
0    39.991
1    27.476
Name: MntGoldProds, dtype: float64
```



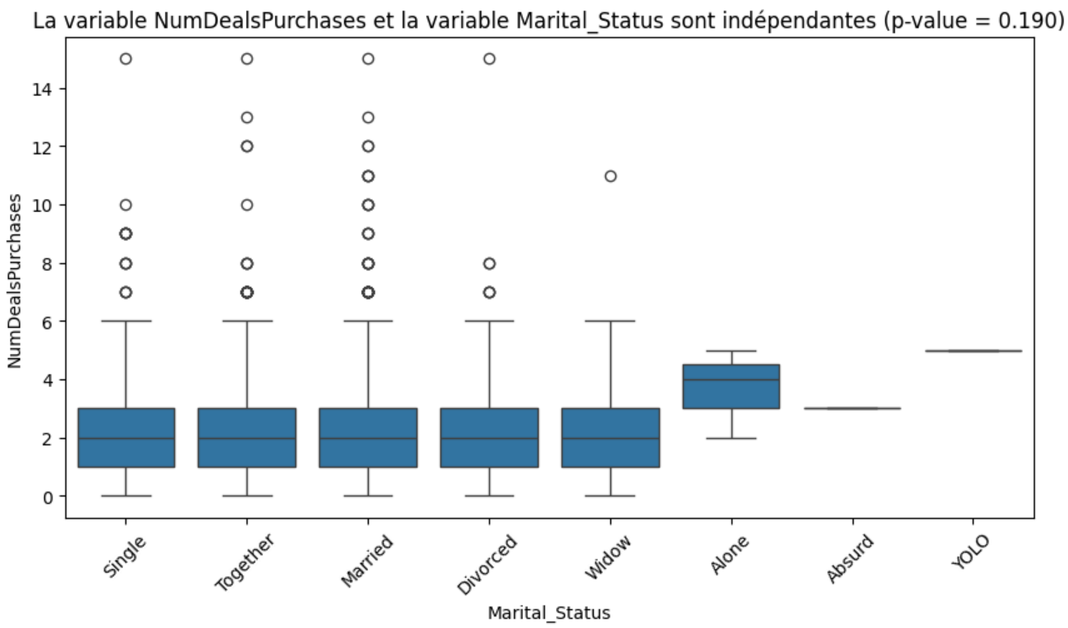
```
 La moyenne de la variable MntGoldProds par rapport à la variable Response
 Response
1    53.436
0    37.658
Name: MntGoldProds, dtype: float64
```




```
 La moyenne de la variable NumDealsPurchases par rapport à la variable Education
 Education
Master        2.475
Graduation    2.358
PhD           2.340
2n Cycle      2.147
Basic         1.717
Name: NumDealsPurchases, dtype: float64
```

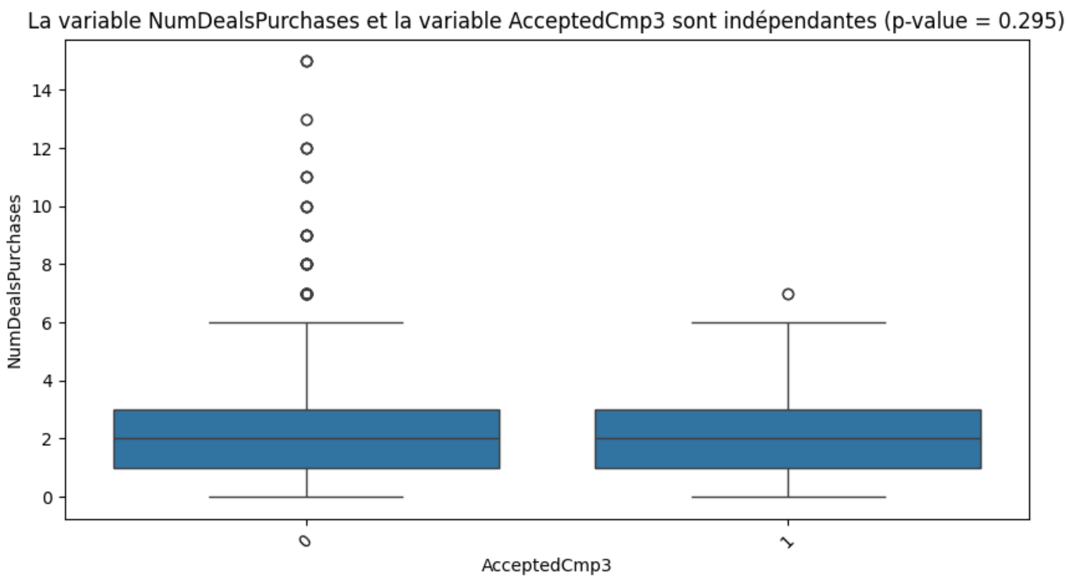



```
 La moyenne de la variable NumDealsPurchases par rapport à la variable
Marital_Status
 Marital_Status
YOLO        5.000
Alone       3.667
Absurd      3.000
Widow       2.406
Divorced    2.405
Married     2.381
Together    2.352
Single      2.180
Name: NumDealsPurchases, dtype: float64
```



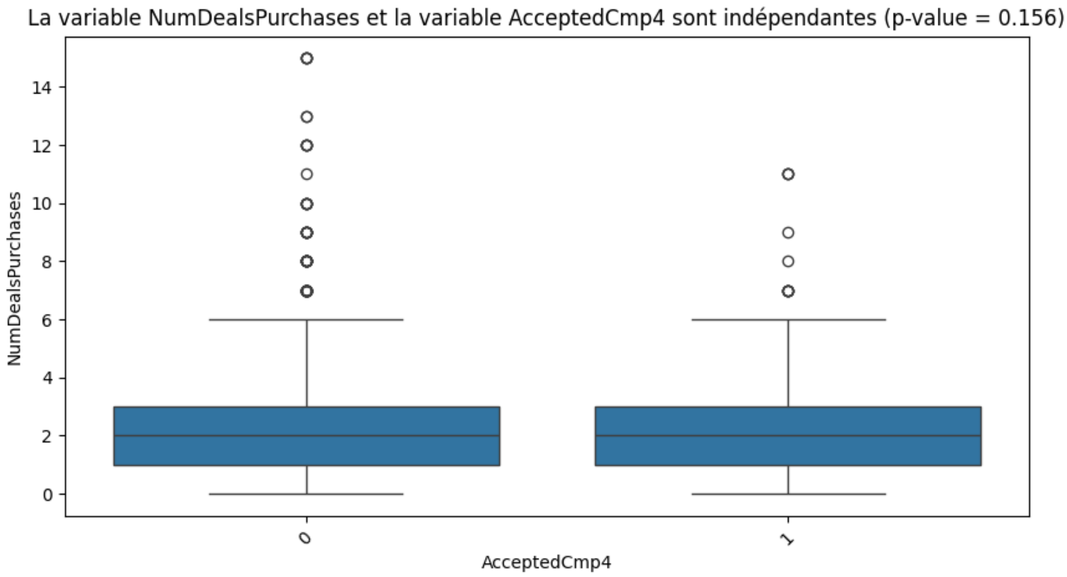
La moyenne de la variable NumDealsPurchases par rapport à la variable AcceptedCmp3

```
 AcceptedCmp3
0    2.351
1    2.188
Name: NumDealsPurchases, dtype: float64
```

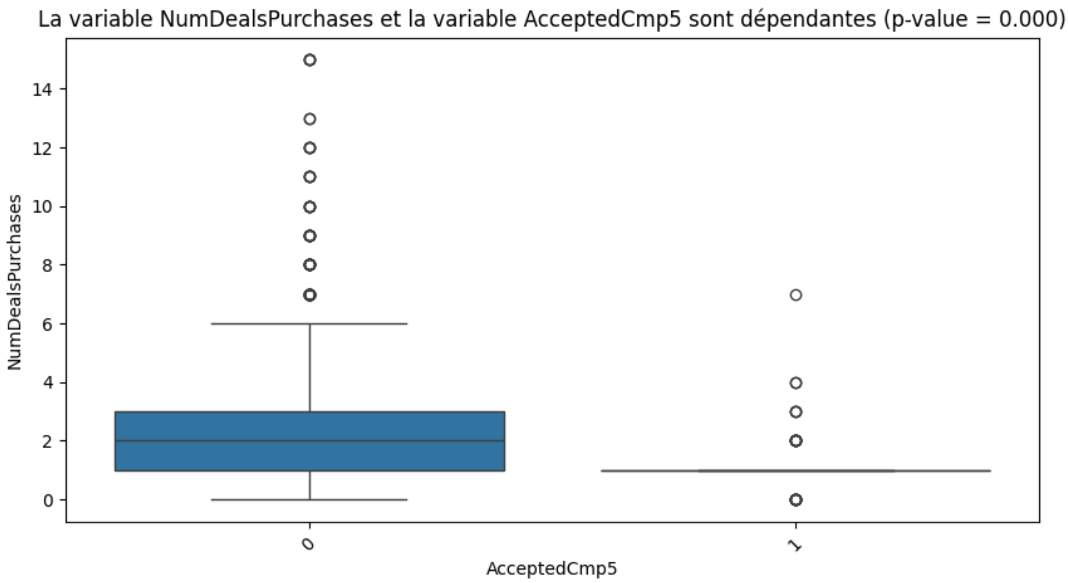


La moyenne de la variable NumDealsPurchases par rapport à la variable AcceptedCmp4

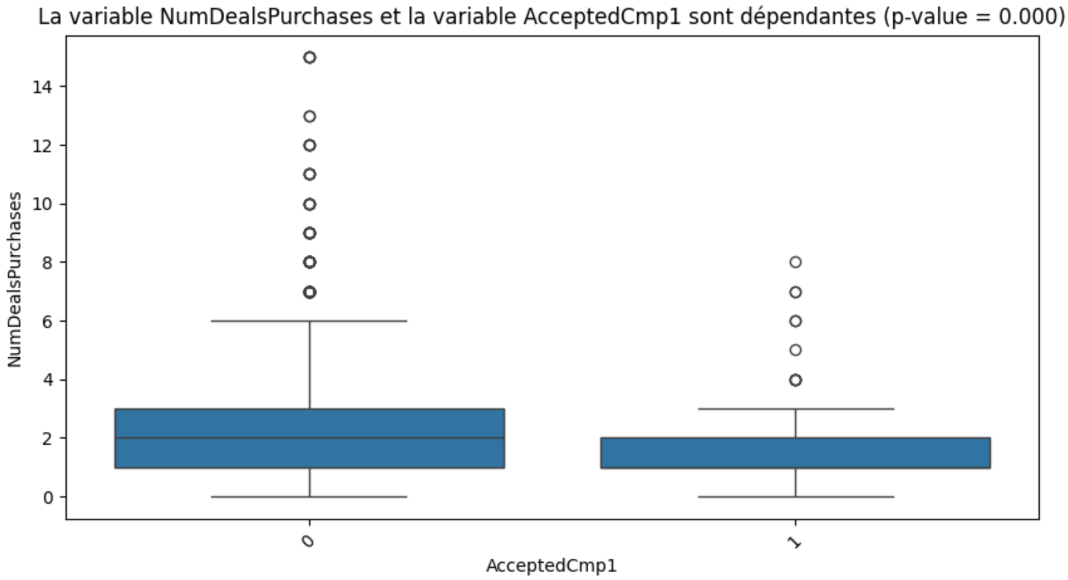
```
 AcceptedCmp4
1    2.542
0    2.323
Name: NumDealsPurchases, dtype: float64
```



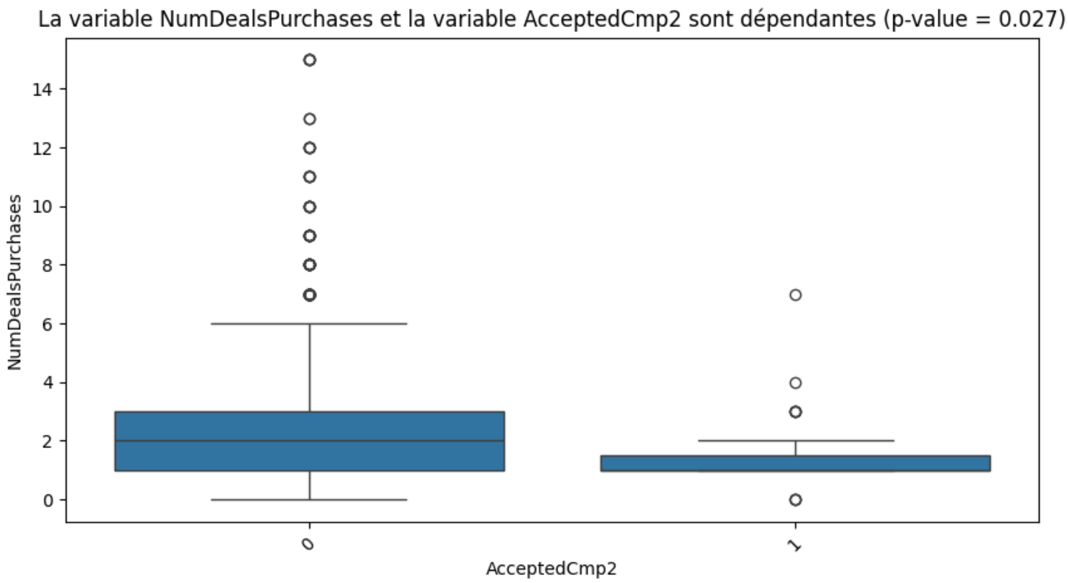

```
 La moyenne de la variable NumDealsPurchases par rapport à la variable
AcceptedCmp5
 AcceptedCmp5
0    2.416
1    1.125
Name: NumDealsPurchases, dtype: float64
```



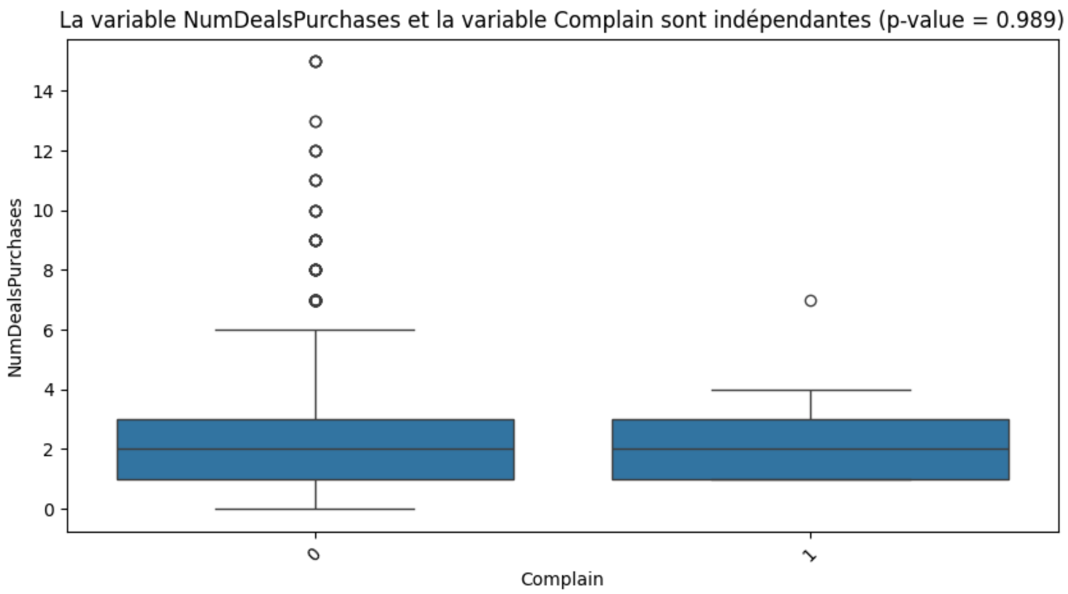
```
 La moyenne de la variable NumDealsPurchases par rapport à la variable
AcceptedCmp1
 AcceptedCmp1
0    2.383
1    1.565
Name: NumDealsPurchases, dtype: float64
```




```
 La moyenne de la variable NumDealsPurchases par rapport à la variable
AcceptedCmp2
 AcceptedCmp2
0    2.349
1    1.556
Name: NumDealsPurchases, dtype: float64
```

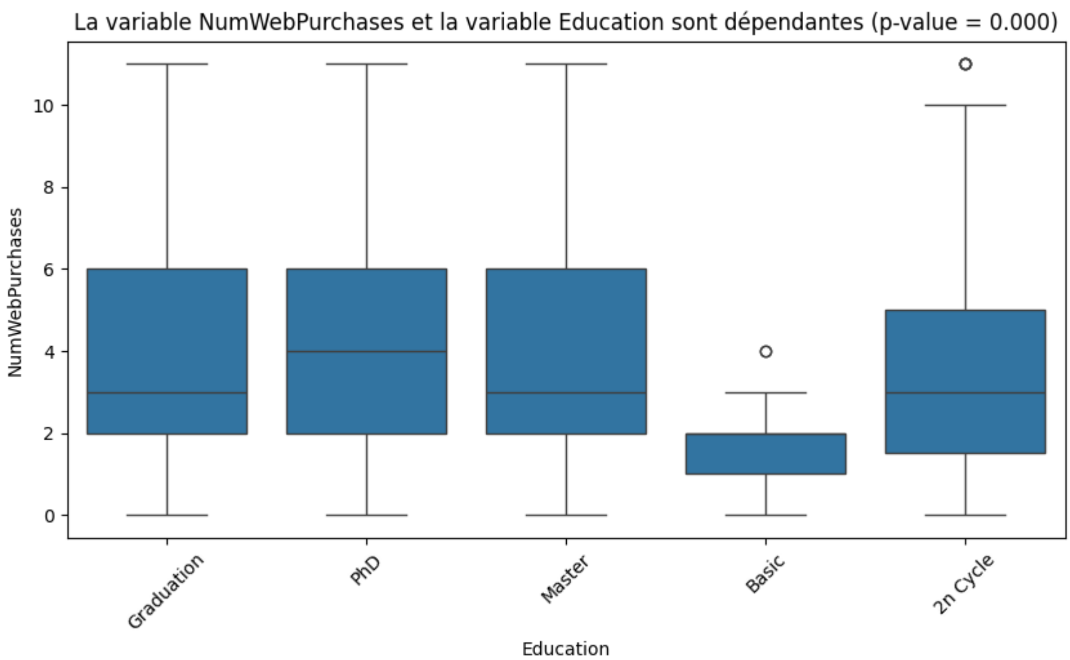


```
 La moyenne de la variable NumDealsPurchases par rapport à la variable Complain
 Complain
0    2.339
1    2.333
Name: NumDealsPurchases, dtype: float64
```

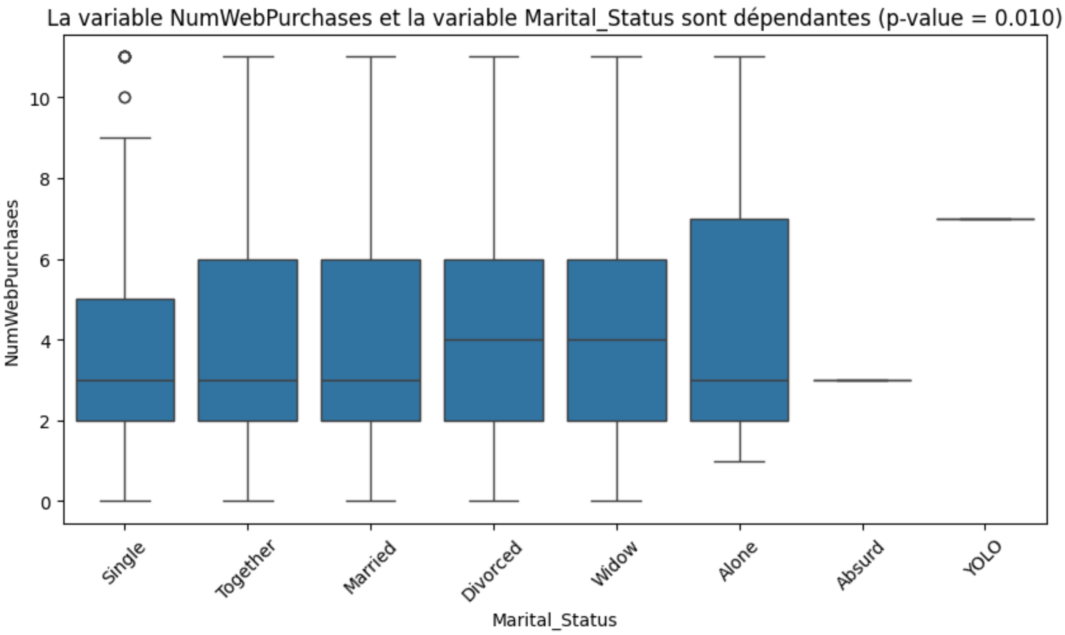



```
 La moyenne de la variable NumDealsPurchases par rapport à la variable Response
 Response
1    2.430
0    2.324
Name: NumDealsPurchases, dtype: float64
```

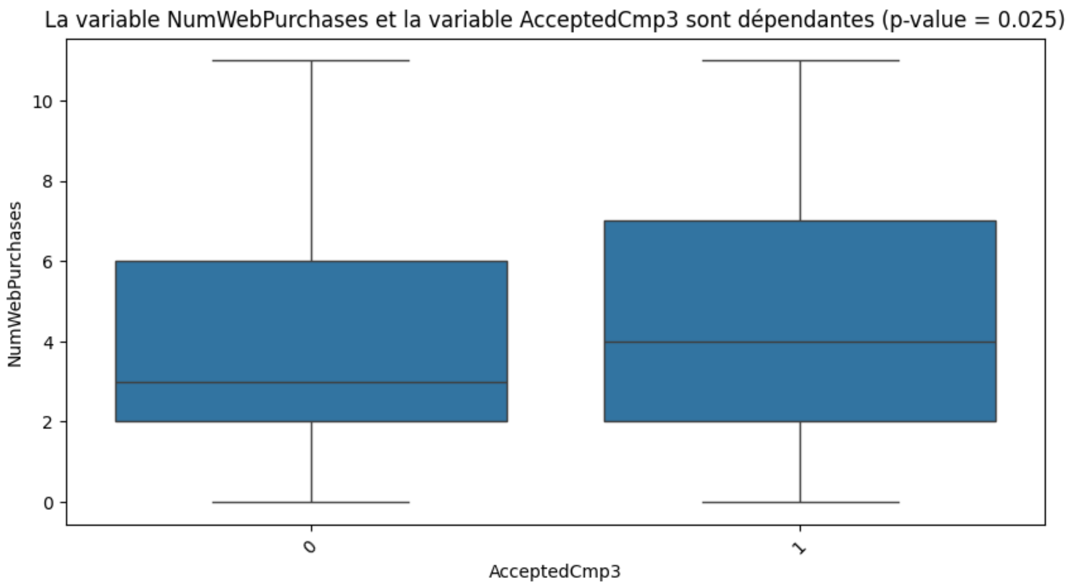

```
 La moyenne de la variable NumWebPurchases par rapport à la variable Education
 Education
PhD           4.348
Graduation    4.000
Master        4.000
2n Cycle      3.581
Basic         1.717
Name: NumWebPurchases, dtype: float64
```



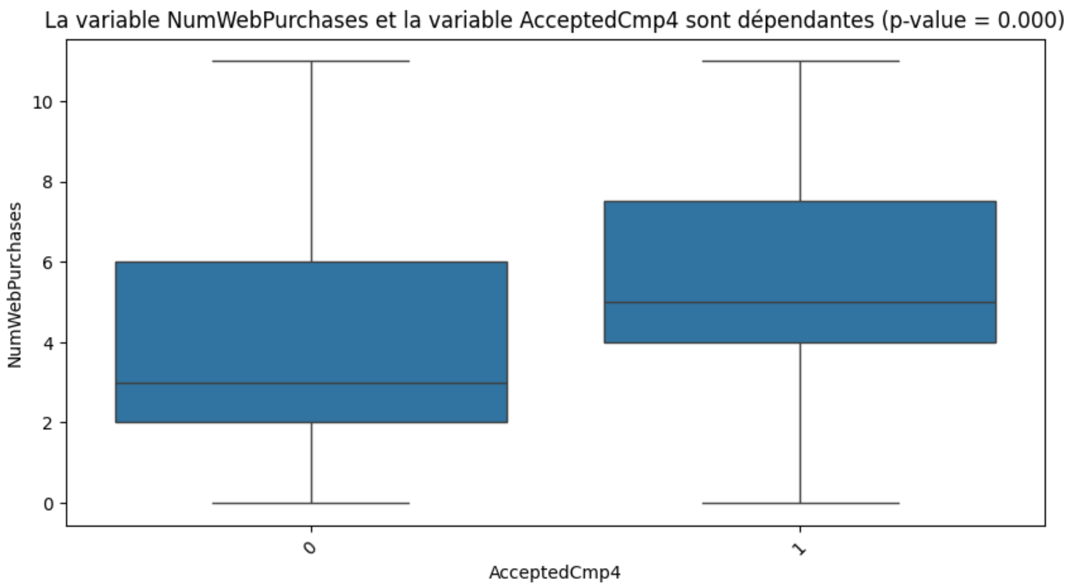
```
 La moyenne de la variable NumWebPurchases par rapport à la variable
Marital_Status
 Marital_Status
YOLO        7.000
Alone       5.000
Widow       4.551
Divorced    4.275
Married     4.046
Together    3.978
Single      3.621
Absurd      3.000
Name: NumWebPurchases, dtype: float64
```




```
 La moyenne de la variable NumWebPurchases par rapport à la variable
AcceptedCmp3
 AcceptedCmp3
1    4.435
0    3.947
Name: NumWebPurchases, dtype: float64
```

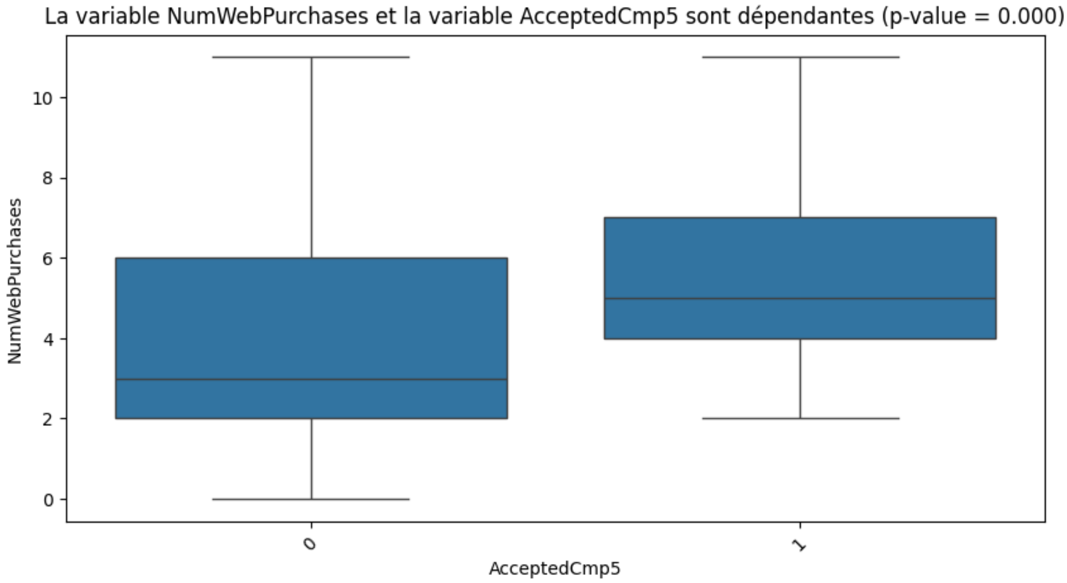



```
 La moyenne de la variable NumWebPurchases par rapport à la variable
AcceptedCmp4
 AcceptedCmp4
1    5.606
0    3.855
Name: NumWebPurchases, dtype: float64
```

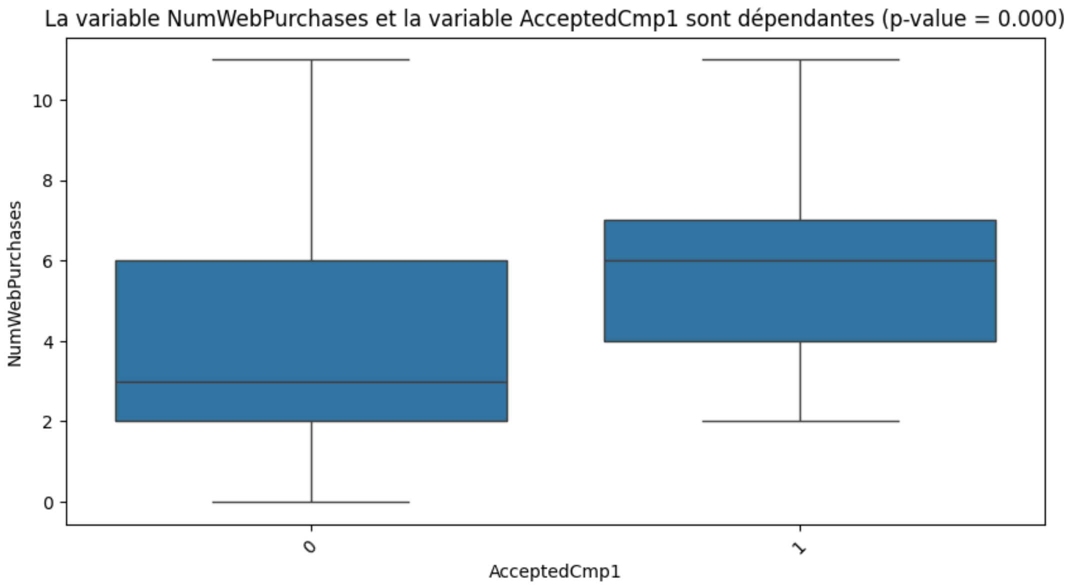



```
 La moyenne de la variable NumWebPurchases par rapport à la variable
AcceptedCmp5
 AcceptedCmp5
```

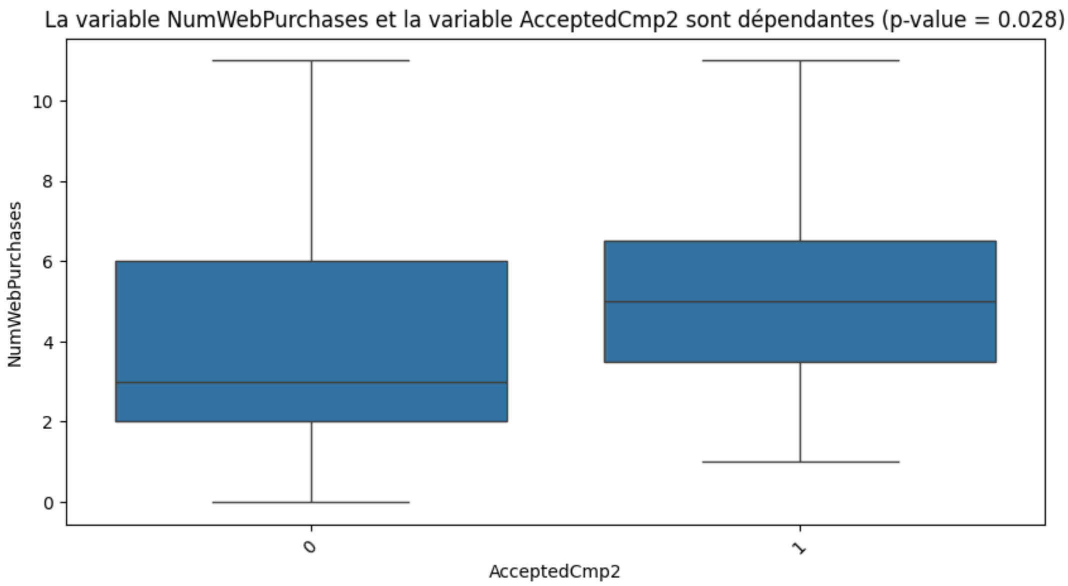
```
1    5.508
0    3.884
Name: NumWebPurchases, dtype: float64
```



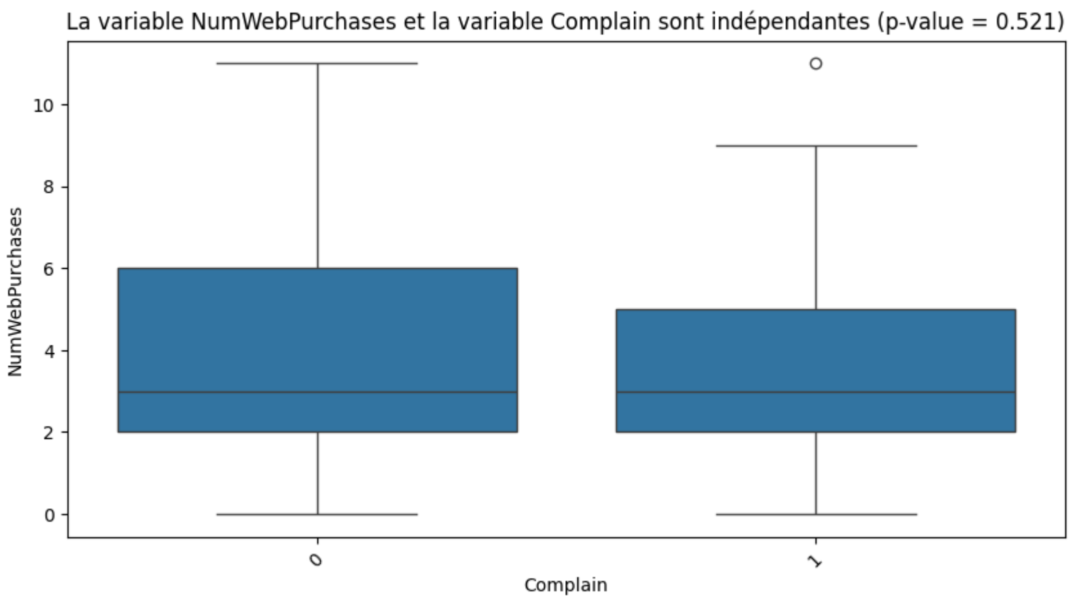

```
 La moyenne de la variable NumWebPurchases par rapport à la variable
AcceptedCmp1
 AcceptedCmp1
1    5.783
0    3.879
Name: NumWebPurchases, dtype: float64
```



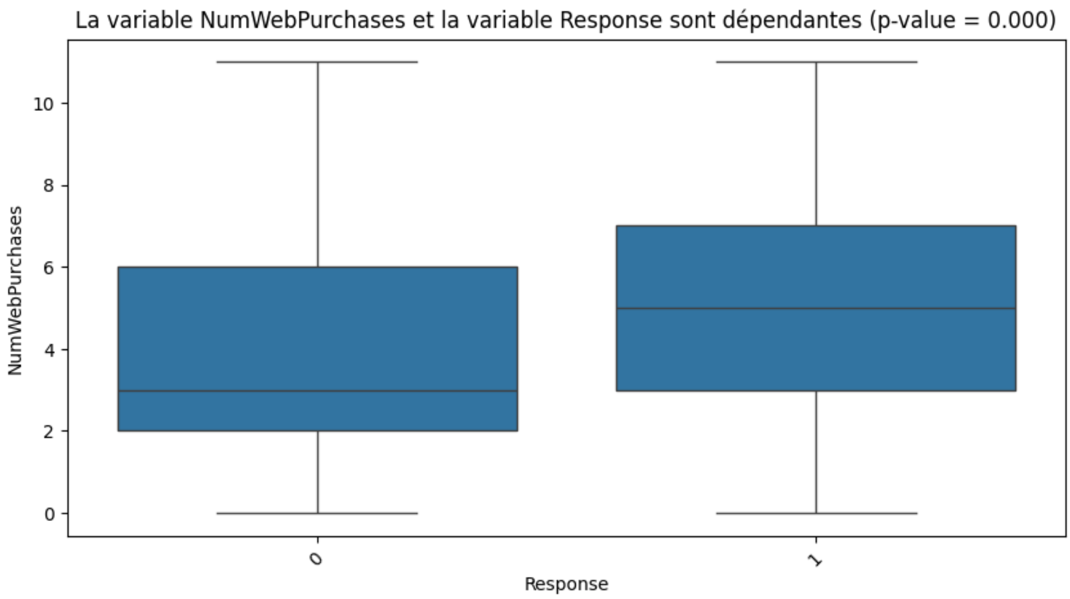
```
 La moyenne de la variable NumWebPurchases par rapport à la variable
AcceptedCmp2
 AcceptedCmp2
1    5.074
0    3.968
Name: NumWebPurchases, dtype: float64
```



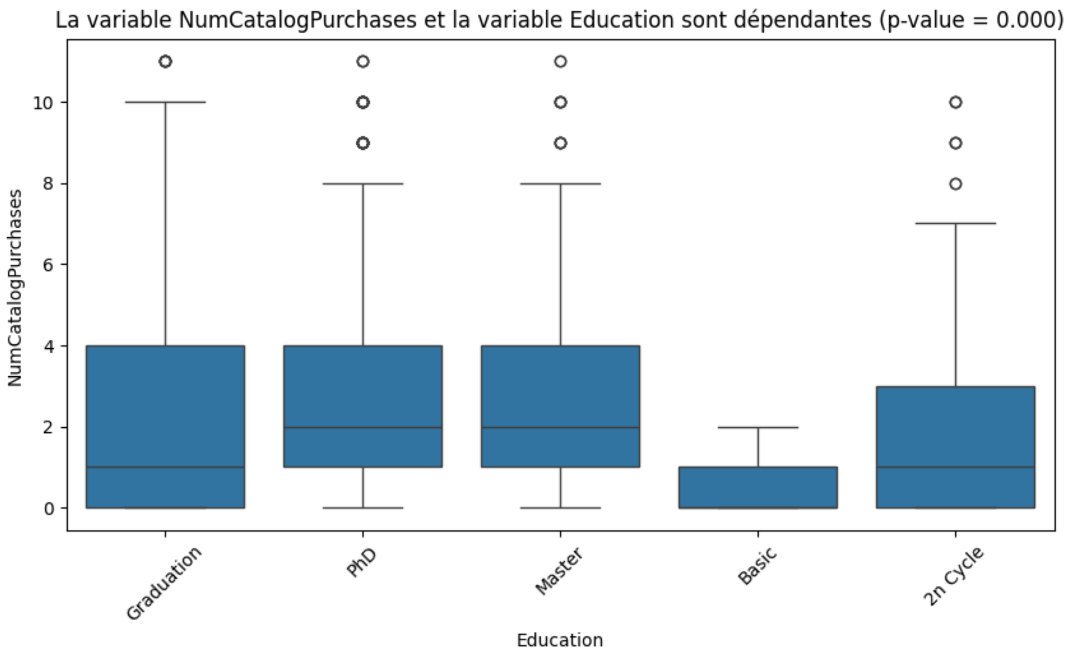

```
 La moyenne de la variable NumWebPurchases par rapport à la variable Complain
 Complain
0    3.986
1    3.619
Name: NumWebPurchases, dtype: float64
```



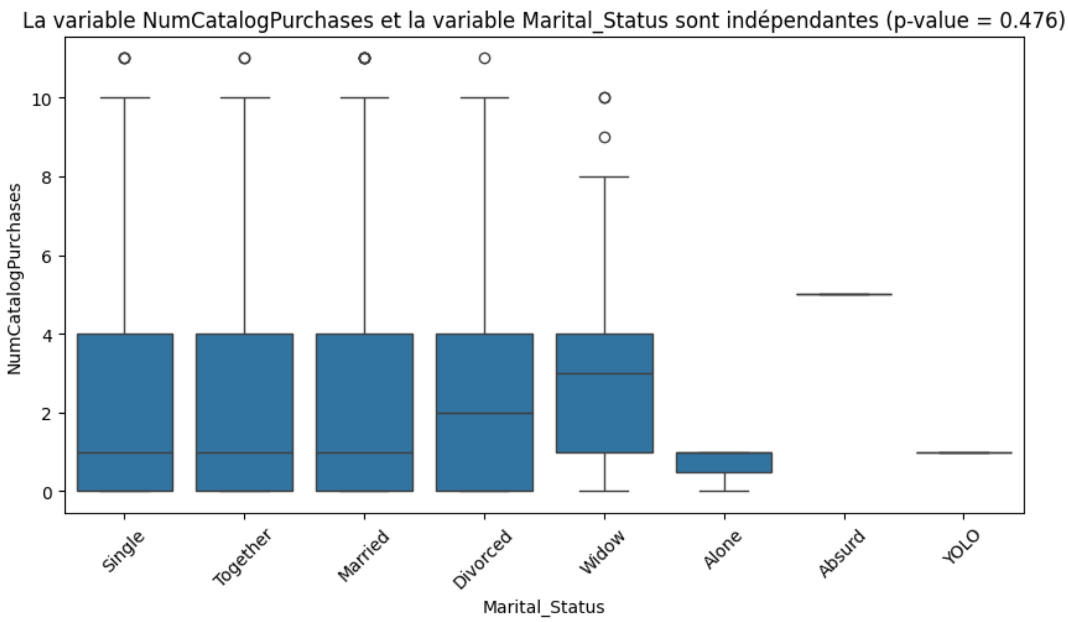
```
 La moyenne de la variable NumWebPurchases par rapport à la variable Response
 Response
1    5.010
0    3.815
Name: NumWebPurchases, dtype: float64
```




```
 La moyenne de la variable NumCatalogPurchases par rapport à la variable
Education
 Education
PhD           2.767
Master        2.480
Graduation    2.425
2n Cycle      2.105
Basic         0.453
Name: NumCatalogPurchases, dtype: float64
```

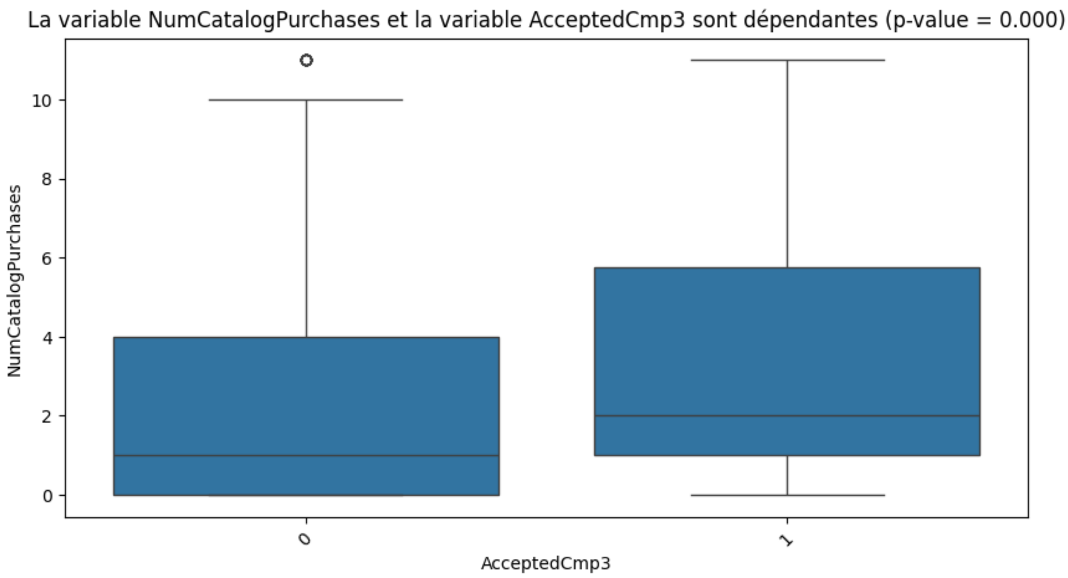



```
 La moyenne de la variable NumCatalogPurchases par rapport à la variable
Marital_Status
 Marital_Status
Absurd      5.000
Widow       2.986
Divorced    2.477
Together    2.450
Single      2.410
Married     2.380
YOLO        1.000
Alone       0.667
Name: NumCatalogPurchases, dtype: float64
```



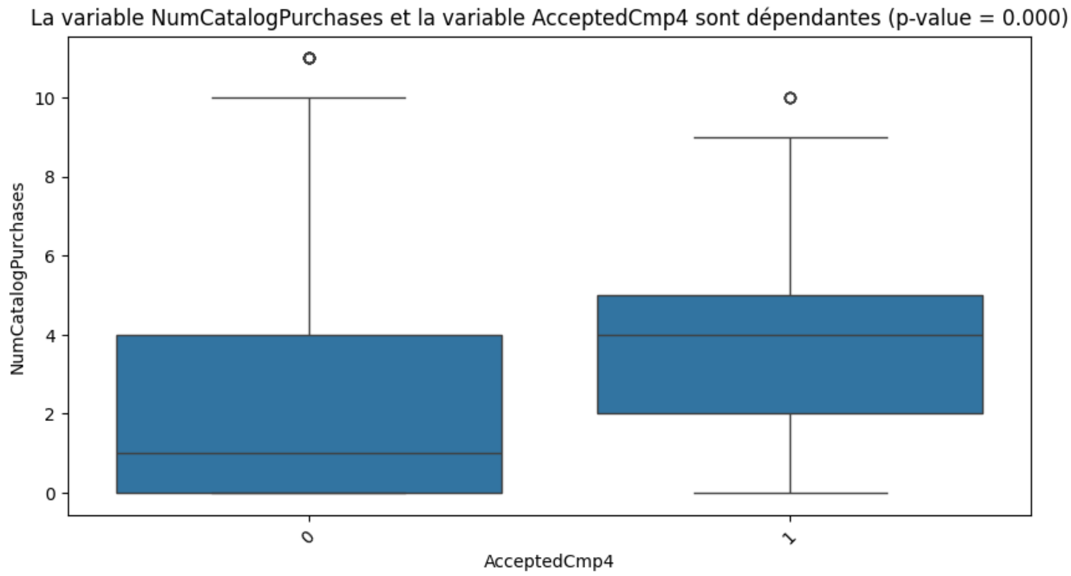
La moyenne de la variable NumCatalogPurchases par rapport à la variable AcceptedCmp3

```
 AcceptedCmp3
1    3.513
0    2.348
Name: NumCatalogPurchases, dtype: float64
```

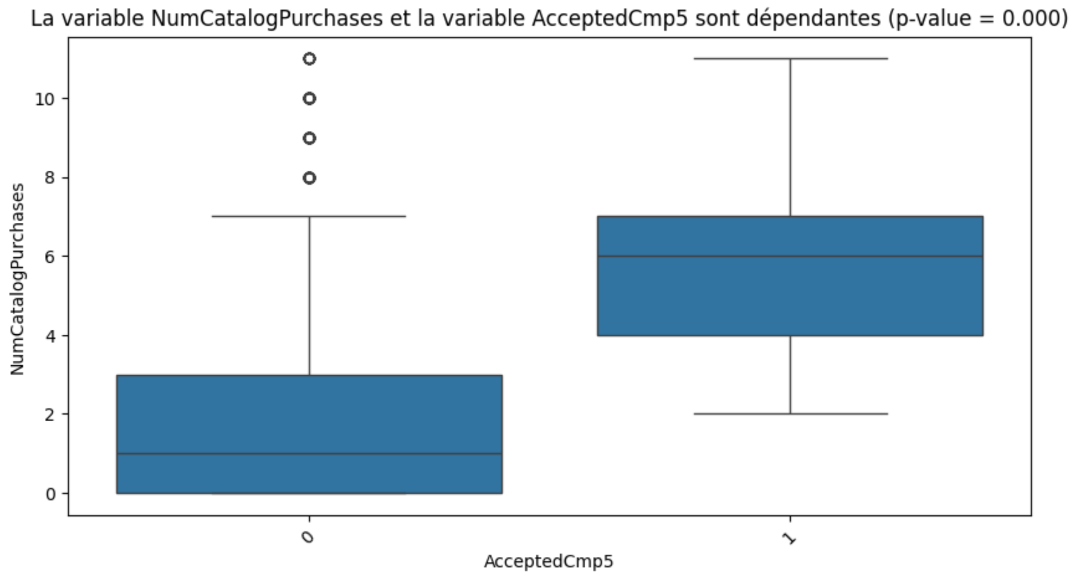


La moyenne de la variable NumCatalogPurchases par rapport à la variable AcceptedCmp4

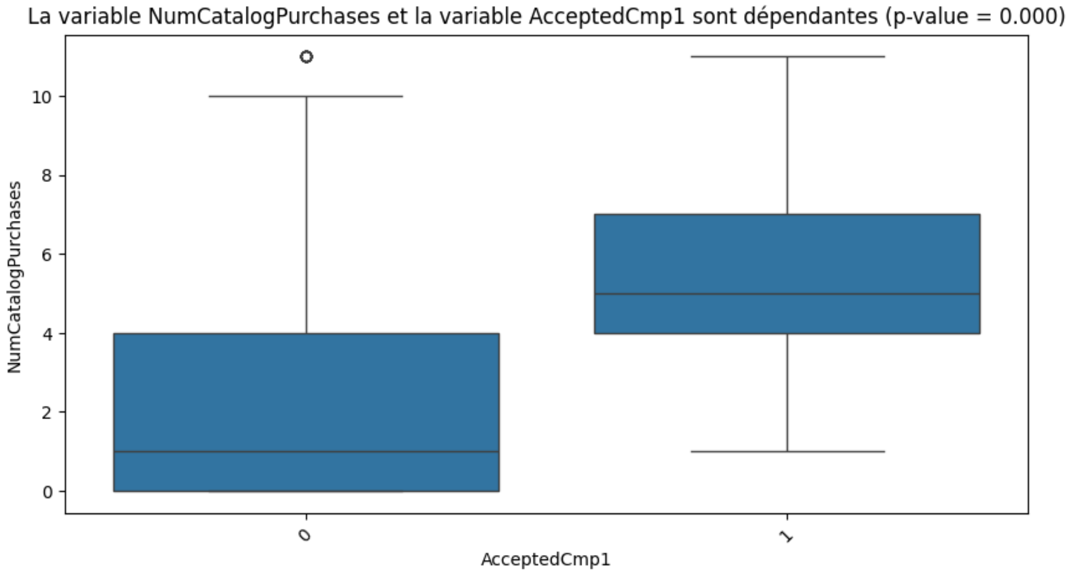
```
 AcceptedCmp4
1    3.897
0    2.317
Name: NumCatalogPurchases, dtype: float64
```



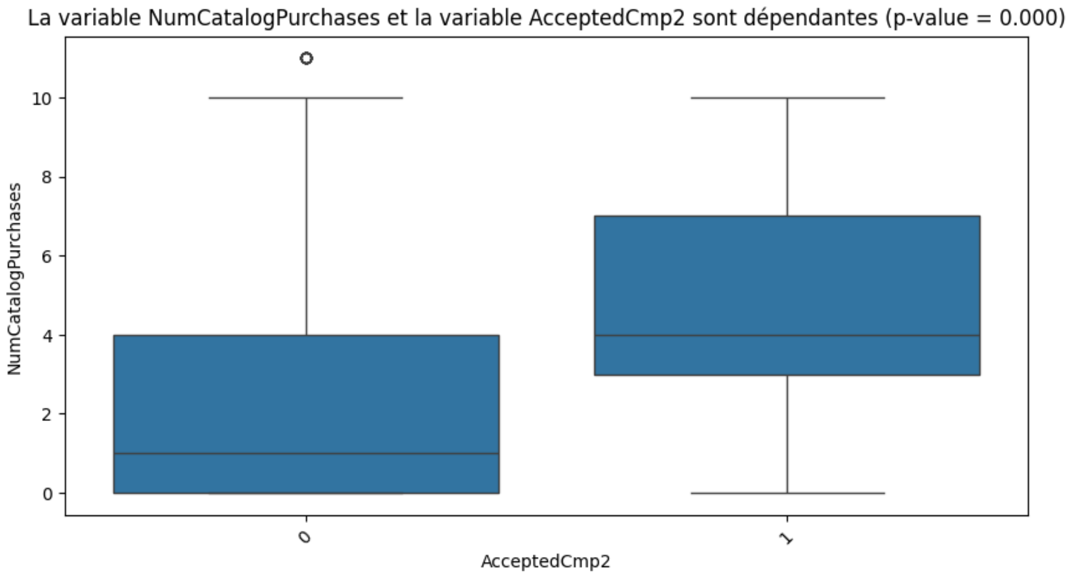

```
 La moyenne de la variable NumCatalogPurchases par rapport à la variable
AcceptedCmp5
 AcceptedCmp5
1    5.648
0    2.226
Name: NumCatalogPurchases, dtype: float64
```



```
 La moyenne de la variable NumCatalogPurchases par rapport à la variable
AcceptedCmp1
 AcceptedCmp1
1    5.704
0    2.245
Name: NumCatalogPurchases, dtype: float64
```

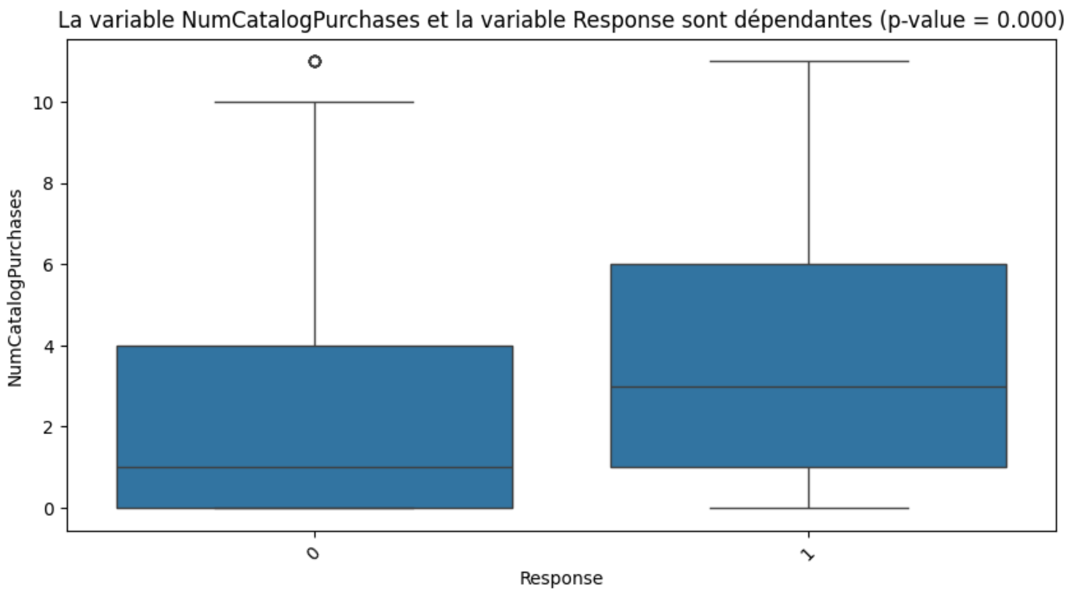



```
 La moyenne de la variable NumCatalogPurchases par rapport à la variable
AcceptedCmp2
 AcceptedCmp2
1    4.926
0    2.400
Name: NumCatalogPurchases, dtype: float64
```

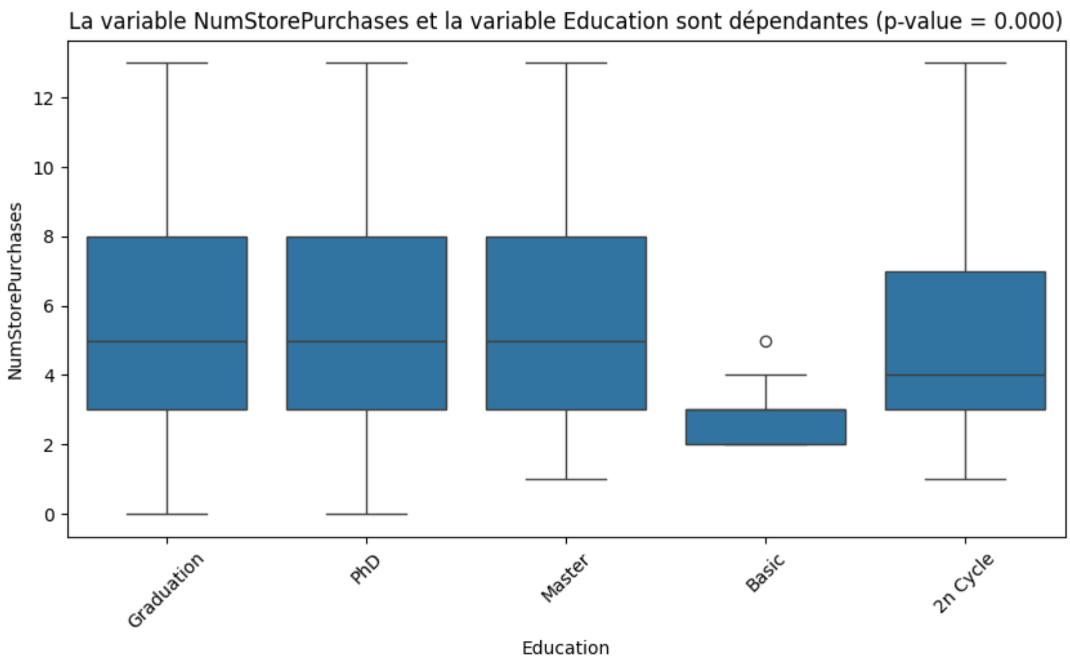


```
 La moyenne de la variable NumCatalogPurchases par rapport à la variable
Complain
 Complain
0    2.436
1    2.048
Name: NumCatalogPurchases, dtype: float64
```

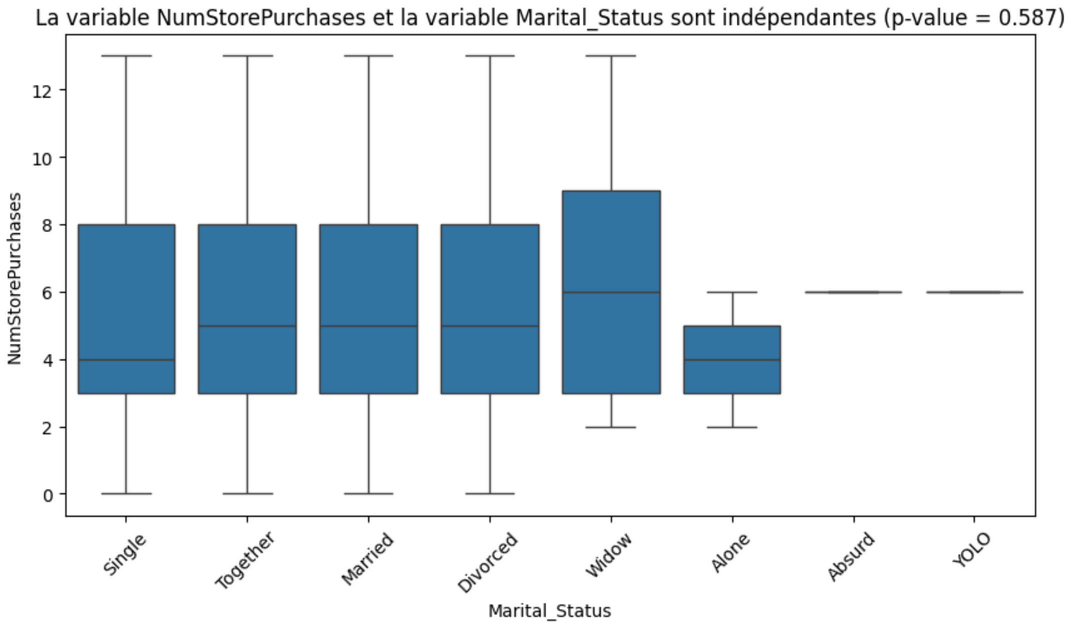

```
 La moyenne de la variable NumCatalogPurchases par rapport à la variable
Response
 Response
1    3.799
0    2.209
Name: NumCatalogPurchases, dtype: float64
```



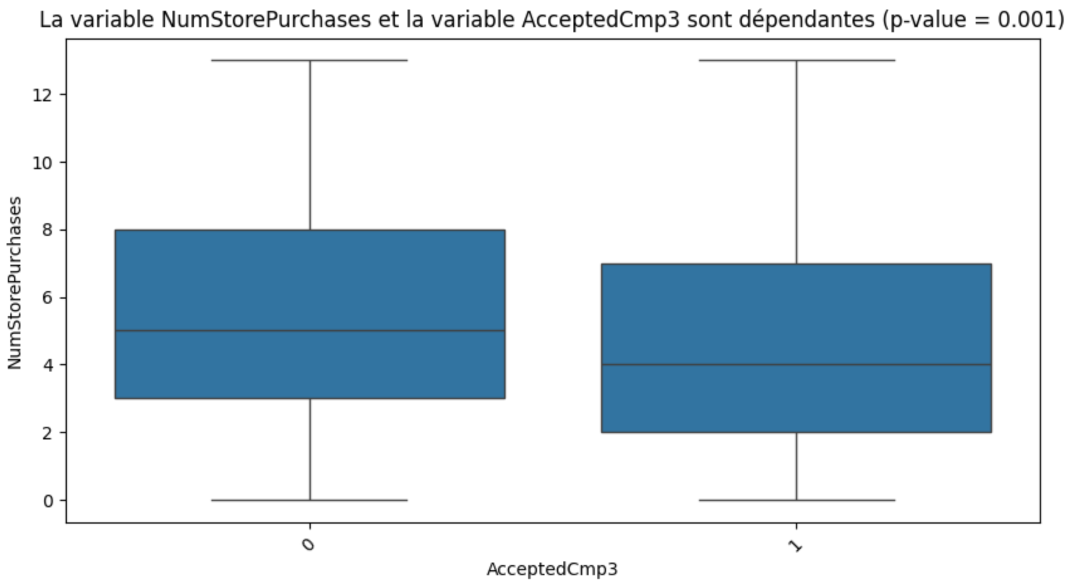

```
 La moyenne de la variable NumStorePurchases par rapport à la variable Education
 Education
PhD           6.058
Master        5.799
Graduation    5.706
2n Cycle      5.304
Basic         2.755
Name: NumStorePurchases, dtype: float64
```



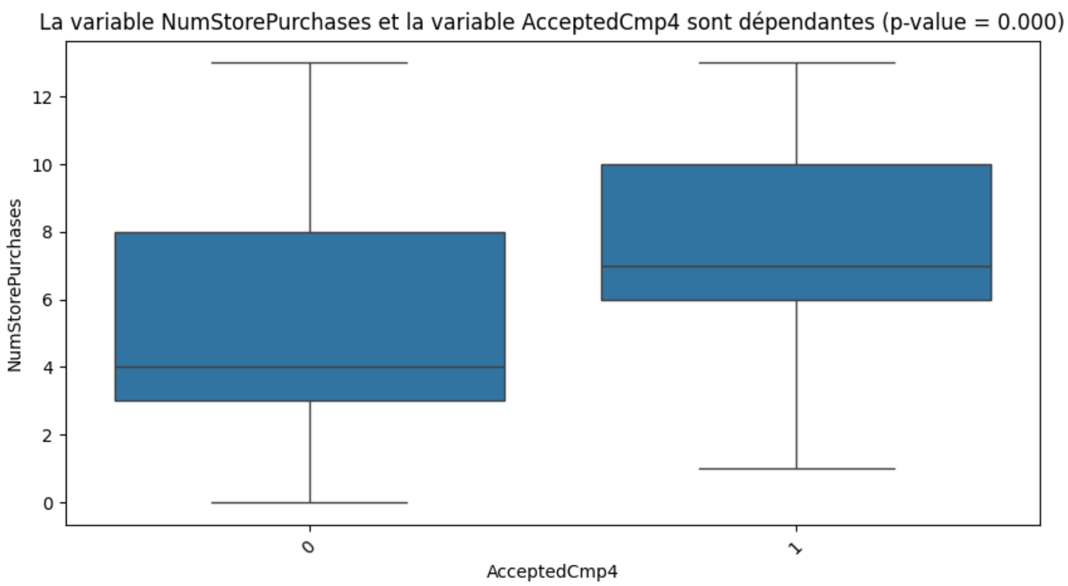
```
 La moyenne de la variable NumStorePurchases par rapport à la variable
Marital_Status
 Marital_Status
Widow       6.275
Absurd      6.000
YOLO        6.000
Married     5.788
Divorced    5.707
Together    5.607
Single      5.519
Alone       4.000
Name: NumStorePurchases, dtype: float64
```




```
 La moyenne de la variable NumStorePurchases par rapport à la variable
AcceptedCmp3
 AcceptedCmp3
0    5.751
1    4.896
Name: NumStorePurchases, dtype: float64
```

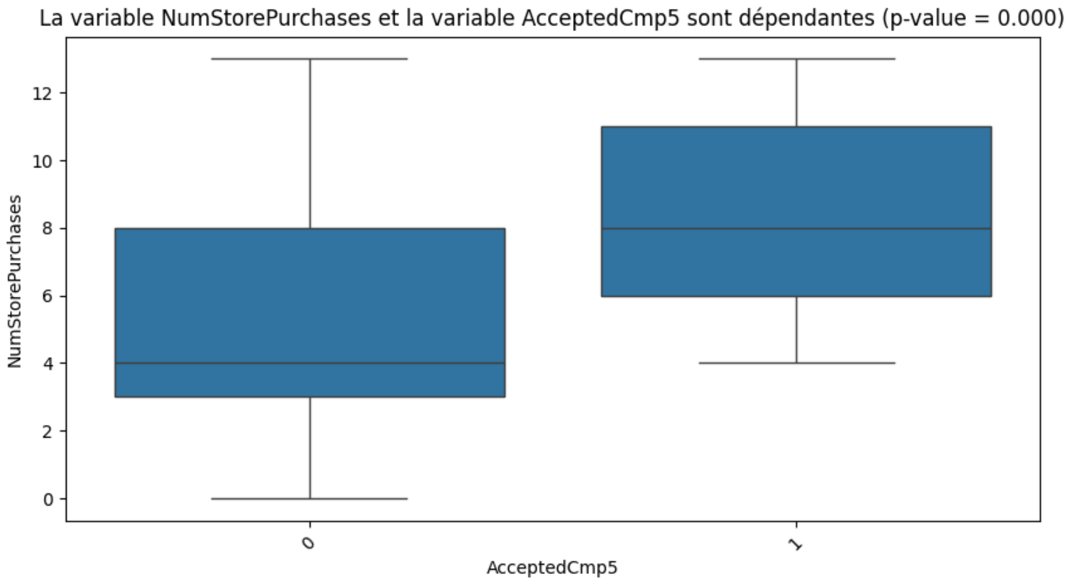



```
 La moyenne de la variable NumStorePurchases par rapport à la variable
AcceptedCmp4
 AcceptedCmp4
1    7.768
0    5.526
Name: NumStorePurchases, dtype: float64
```

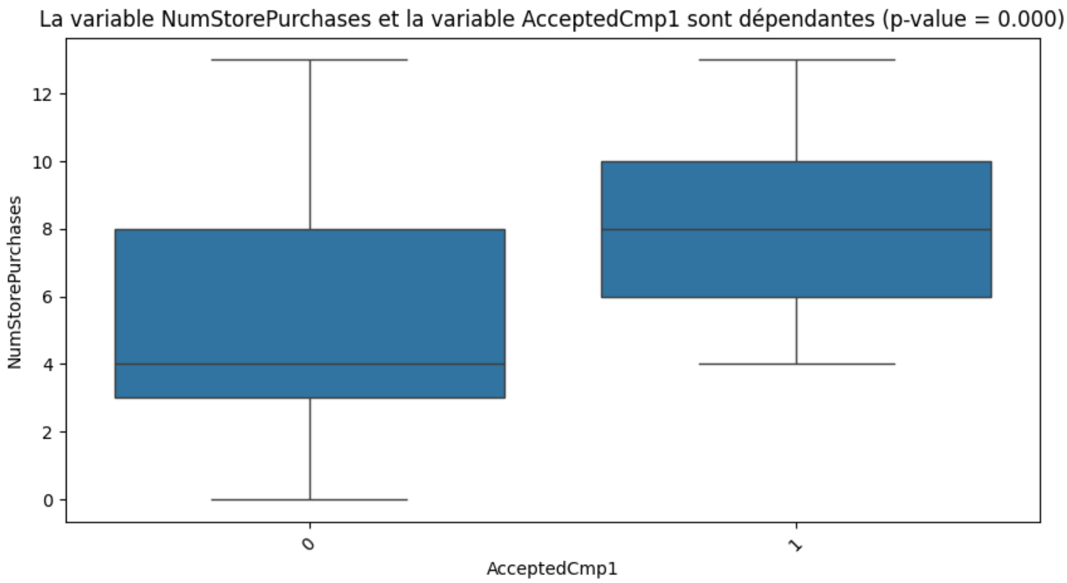



```
 La moyenne de la variable NumStorePurchases par rapport à la variable
AcceptedCmp5
 AcceptedCmp5
```

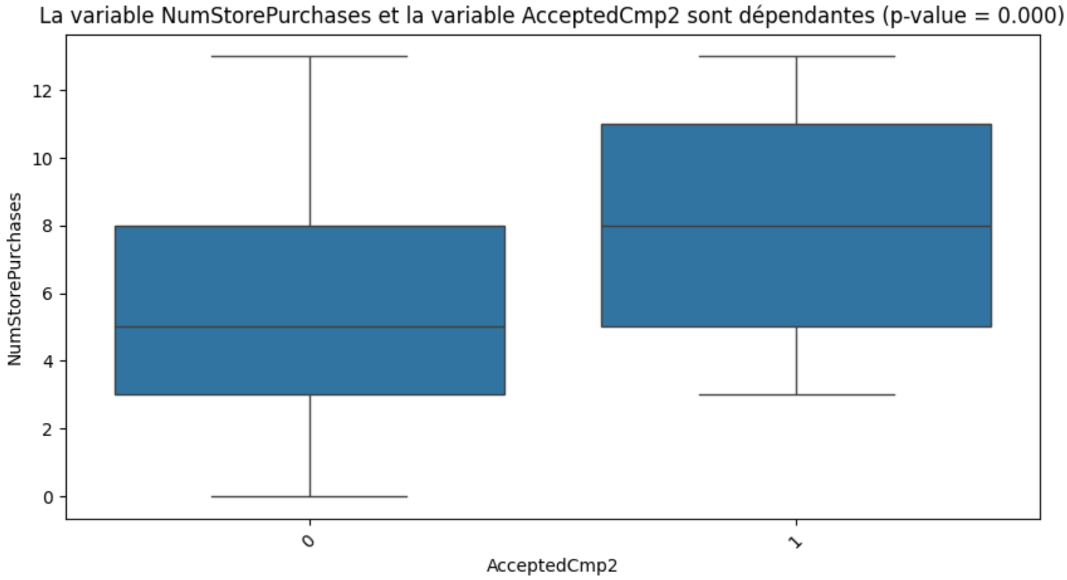
```
1    8.328
0    5.520
Name: NumStorePurchases, dtype: float64
```




```
 La moyenne de la variable NumStorePurchases par rapport à la variable
AcceptedCmp1
 AcceptedCmp1
1    7.948
0    5.560
Name: NumStorePurchases, dtype: float64
```



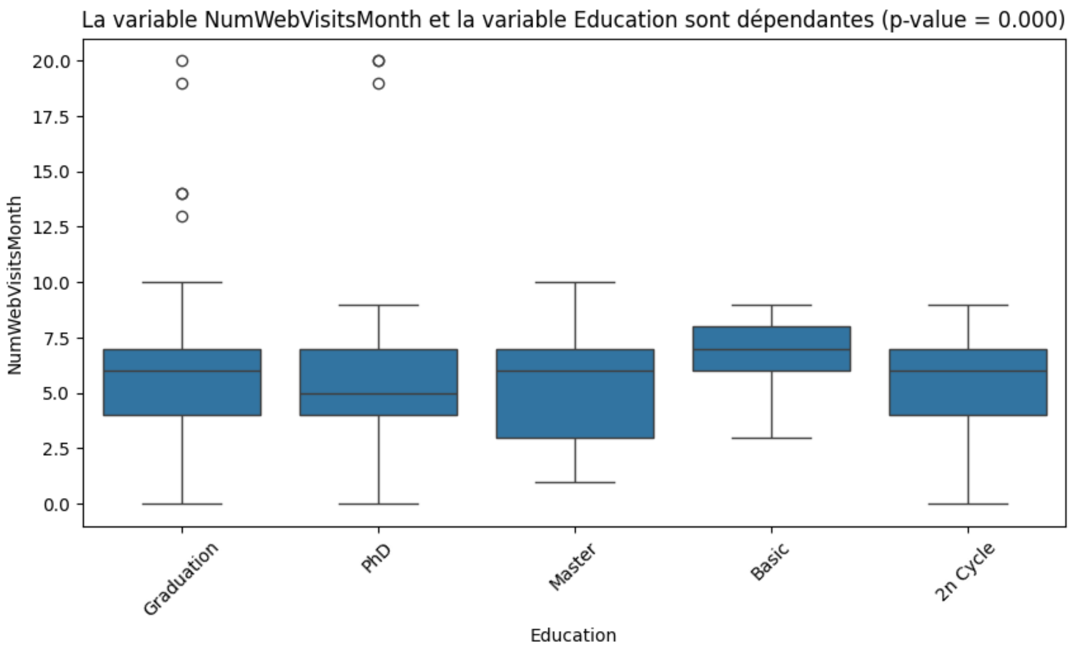
 La moyenne de la variable NumStorePurchases par rapport à la variable AcceptedCmp2
```
 AcceptedCmp2
1    7.926
0    5.661
Name: NumStorePurchases, dtype: float64
```



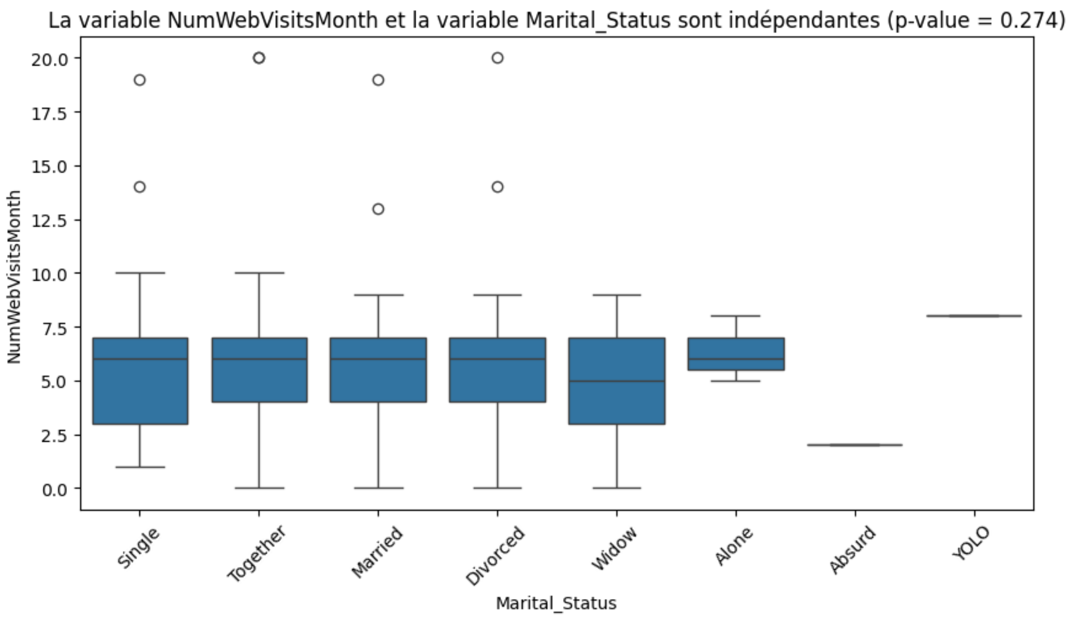
 La moyenne de la variable NumStorePurchases par rapport à la variable Complain
```
 Complain
0    5.694
1    5.238
Name: NumStorePurchases, dtype: float64
```

```
 La moyenne de la variable NumStorePurchases par rapport à la variable Response
 Response
1    5.896
0    5.656
Name: NumStorePurchases, dtype: float64
```

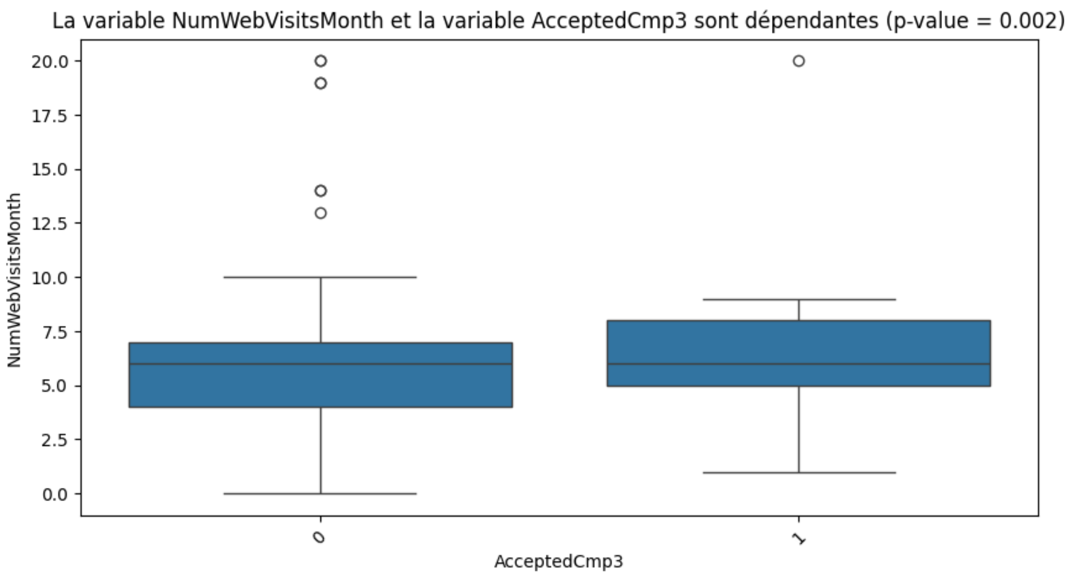

```
 La moyenne de la variable NumWebVisitsMonth par rapport à la variable Education
 Education
Basic          6.830
2n Cycle       5.513
Graduation     5.432
PhD            5.361
Master         5.221
Name: NumWebVisitsMonth, dtype: float64
```




```
 La moyenne de la variable NumWebVisitsMonth par rapport à la variable
Marital_Status
 Marital_Status
YOLO        8.000
Alone       6.333
Divorced    5.568
Married     5.462
Single      5.406
Together    5.366
Widow       4.986
Absurd      2.000
Name: NumWebVisitsMonth, dtype: float64
```

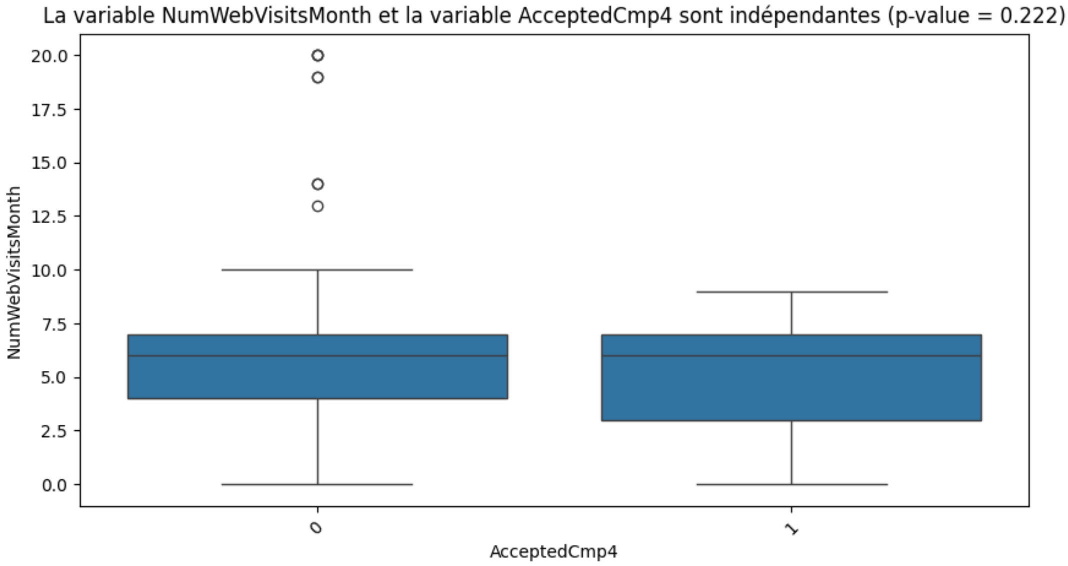



```
 La moyenne de la variable NumWebVisitsMonth par rapport à la variable
AcceptedCmp3
 AcceptedCmp3
1    5.994
0    5.378
Name: NumWebVisitsMonth, dtype: float64
```

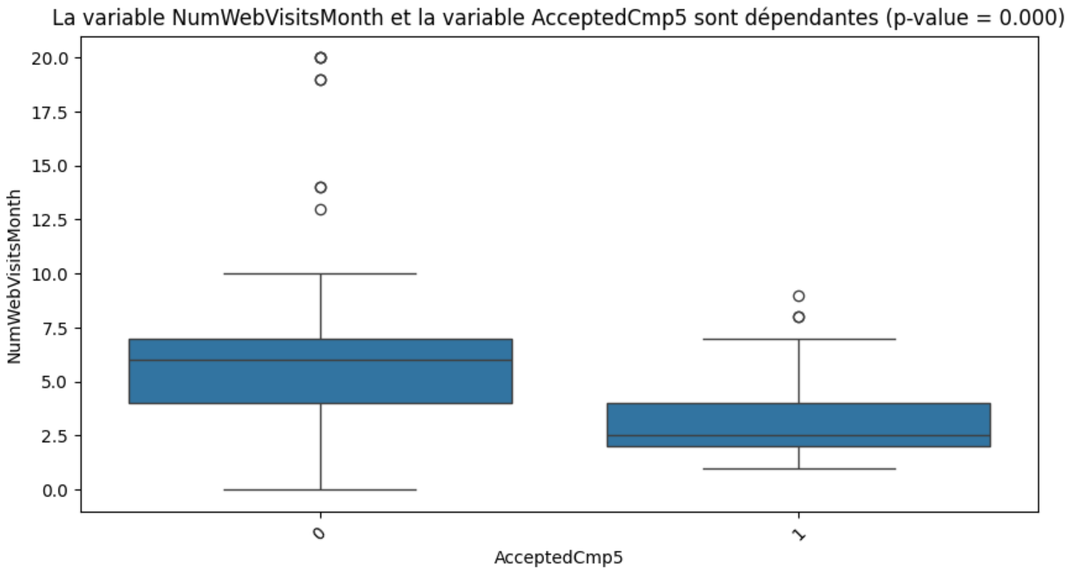



```
 La moyenne de la variable NumWebVisitsMonth par rapport à la variable
AcceptedCmp4
 AcceptedCmp4
```

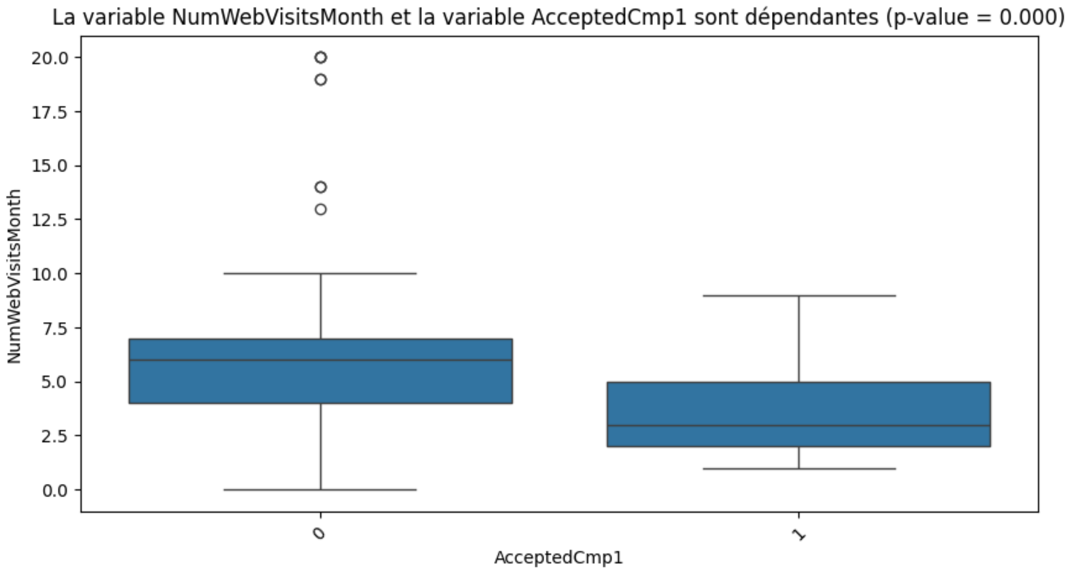
```
0    5.44
1    5.20
Name: NumWebVisitsMonth, dtype: float64
```



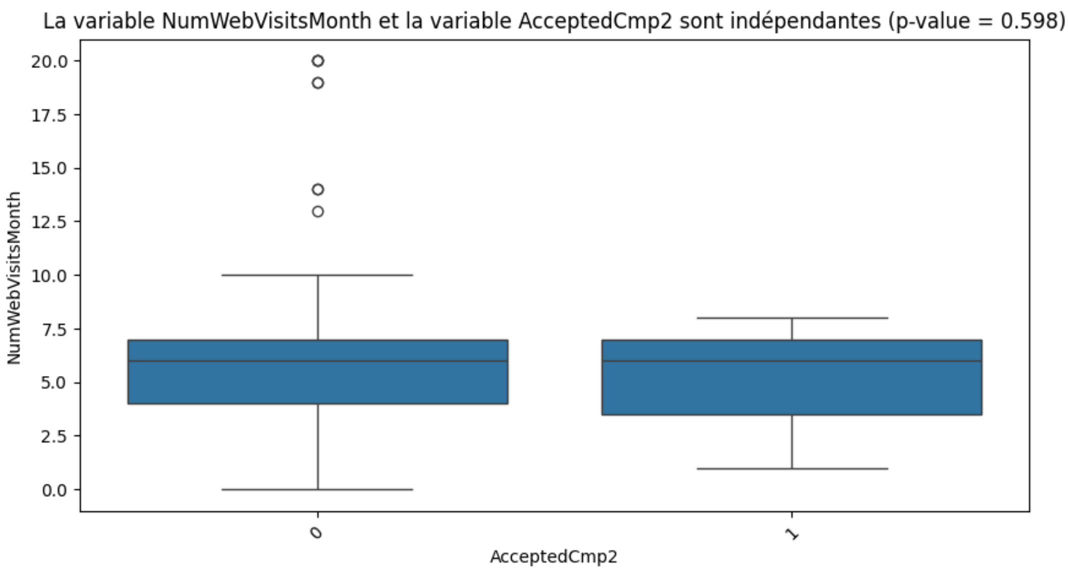

```
 La moyenne de la variable NumWebVisitsMonth par rapport à la variable
AcceptedCmp5
 AcceptedCmp5
0    5.572
1    3.094
Name: NumWebVisitsMonth, dtype: float64
```



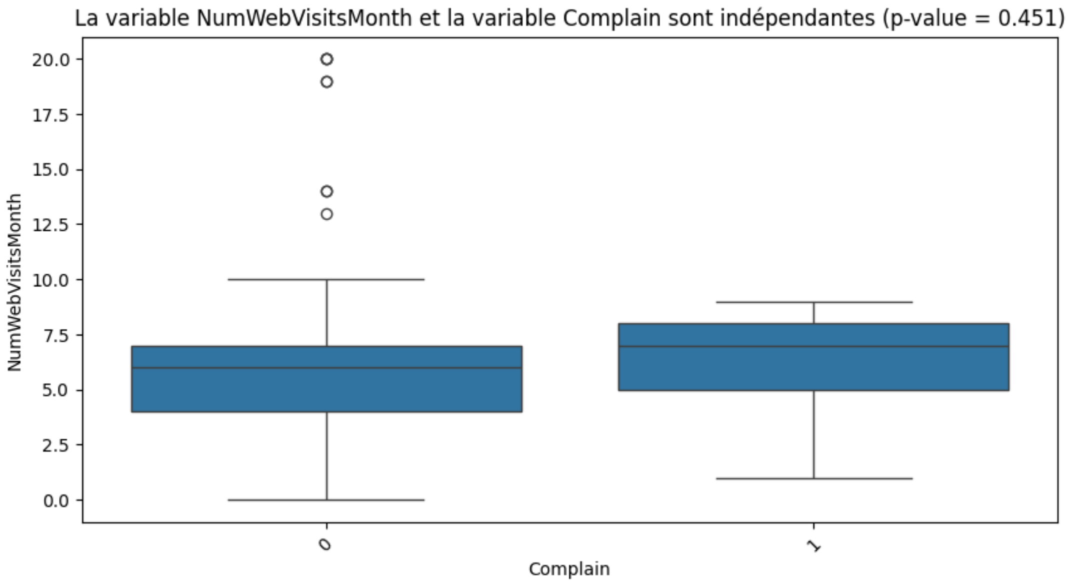
```
 La moyenne de la variable NumWebVisitsMonth par rapport à la variable
AcceptedCmp1
 AcceptedCmp1
0    5.520
1    3.722
Name: NumWebVisitsMonth, dtype: float64
```



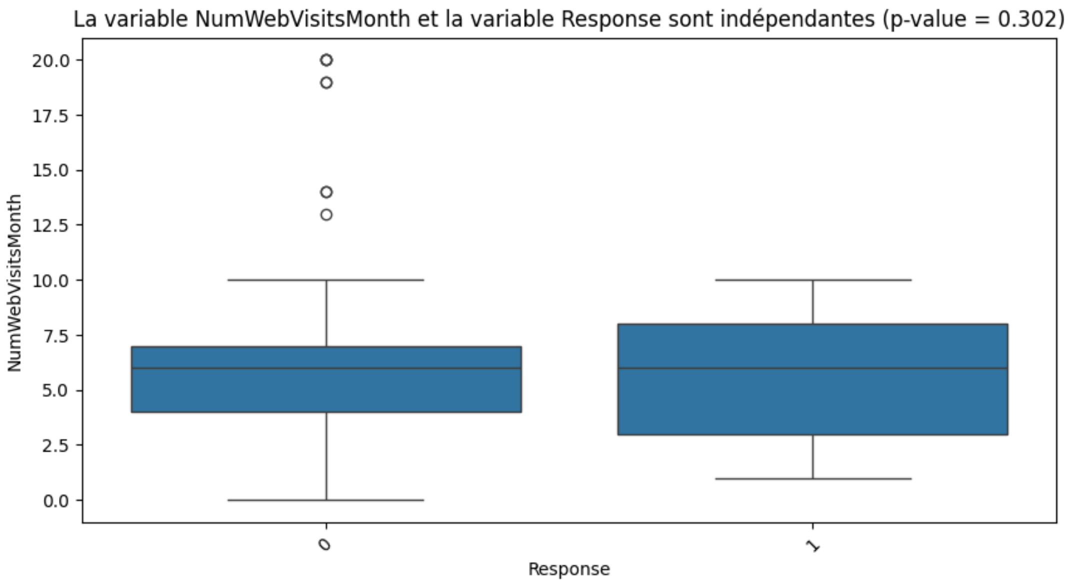

```
 La moyenne de la variable NumWebVisitsMonth par rapport à la variable
AcceptedCmp2
 AcceptedCmp2
0    5.426
1    5.185
Name: NumWebVisitsMonth, dtype: float64
```



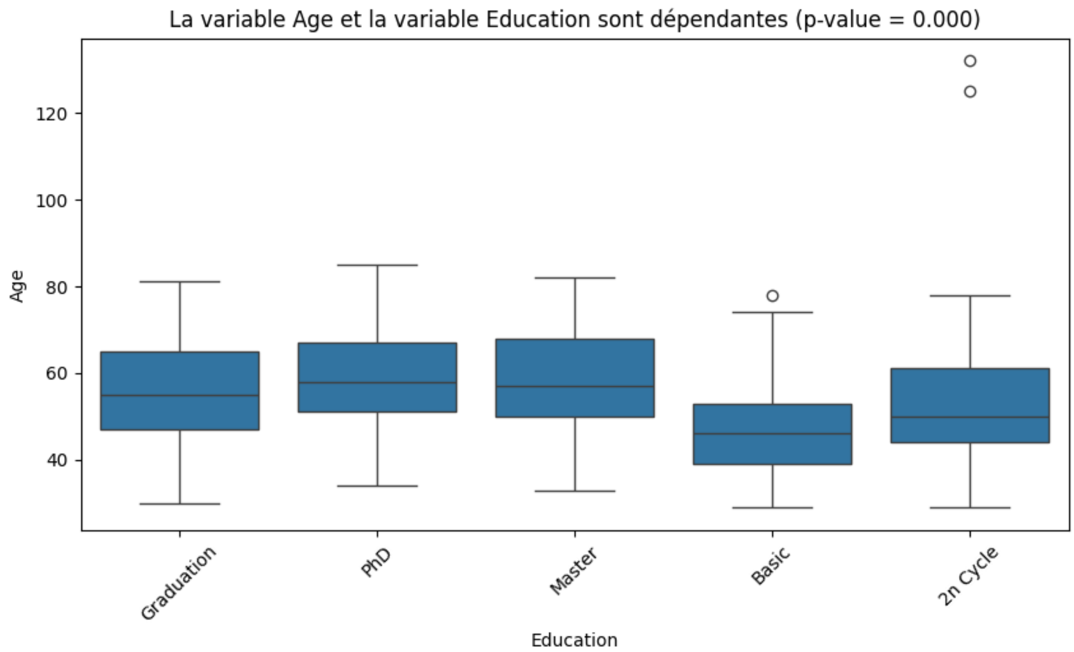
```
 La moyenne de la variable NumWebVisitsMonth par rapport à la variable Complain
 Complain
1    5.810
0    5.419
Name: NumWebVisitsMonth, dtype: float64
```



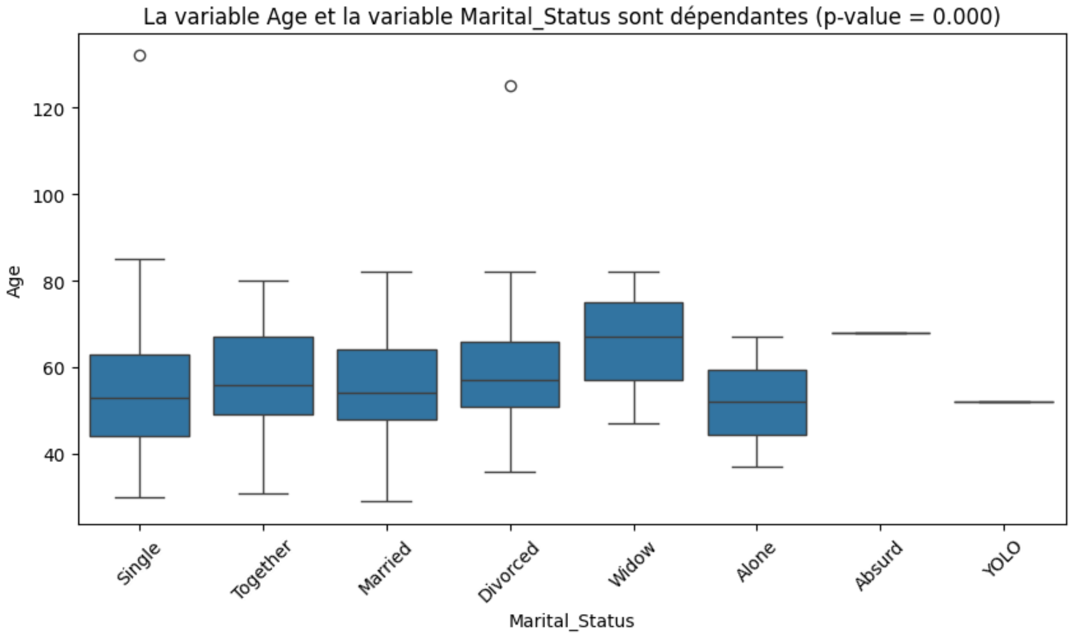

```
 La moyenne de la variable NumWebVisitsMonth par rapport à la variable Response
 Response
1    5.554
0    5.402
Name: NumWebVisitsMonth, dtype: float64
```



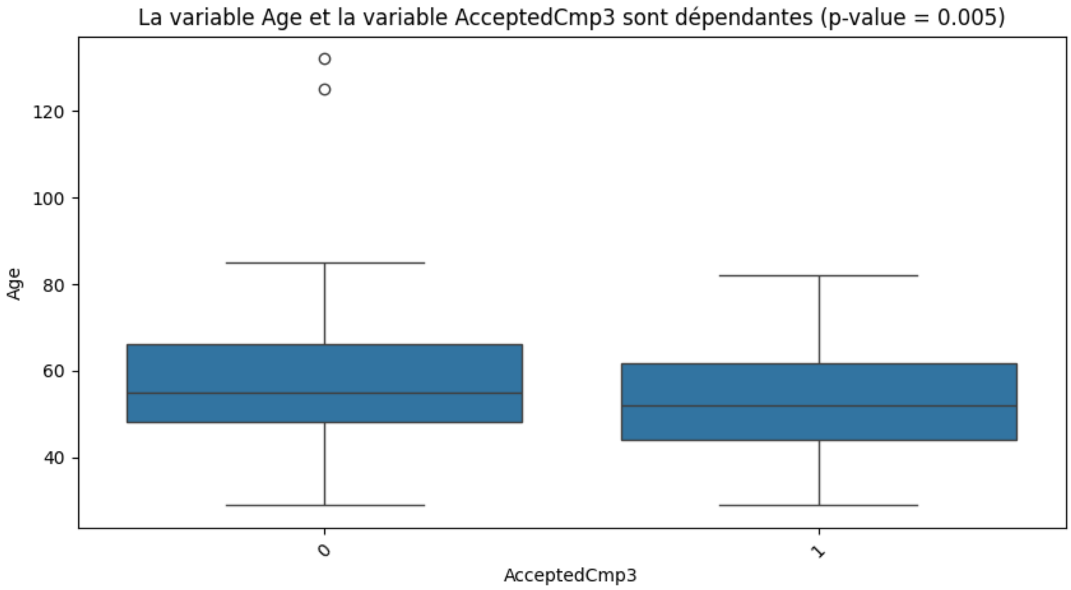
```
 La moyenne de la variable Age par rapport à la variable Education
 Education
PhD            58.769
Master         58.165
Graduation     55.490
2n Cycle       52.916
Basic          47.604
Name: Age, dtype: float64
```




```
 La moyenne de la variable Age par rapport à la variable Marital_Status
 Marital_Status
Absurd      68.000
Widow       66.130
Divorced    58.752
Together    57.281
Married     55.496
Single      53.585
Alone       52.000
YOLO        52.000
Name: Age, dtype: float64
```

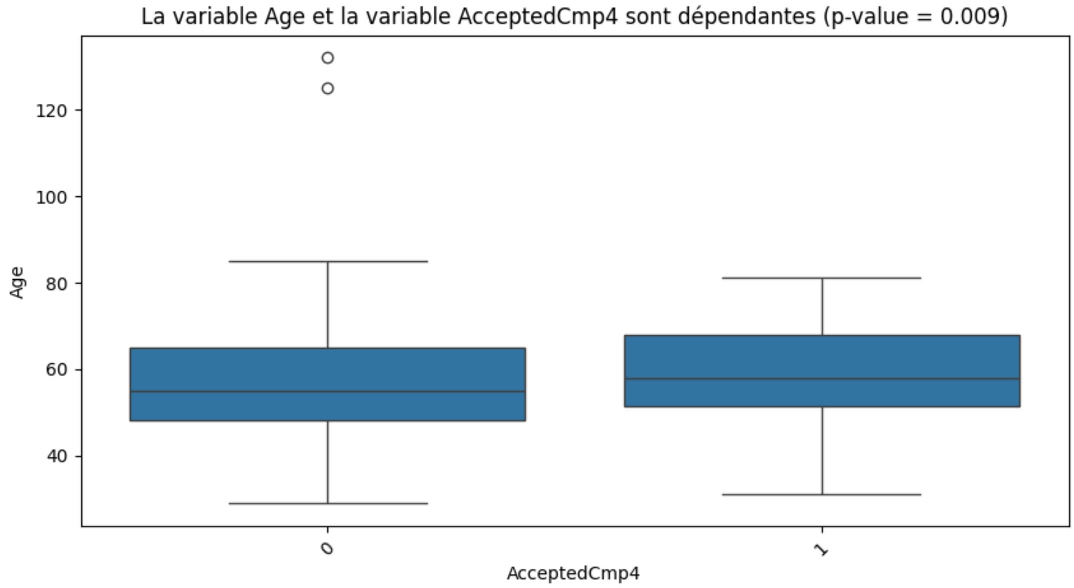



```
 La moyenne de la variable Age par rapport à la variable AcceptedCmp3
 AcceptedCmp3
0    56.436
1    53.643
Name: Age, dtype: float64
```

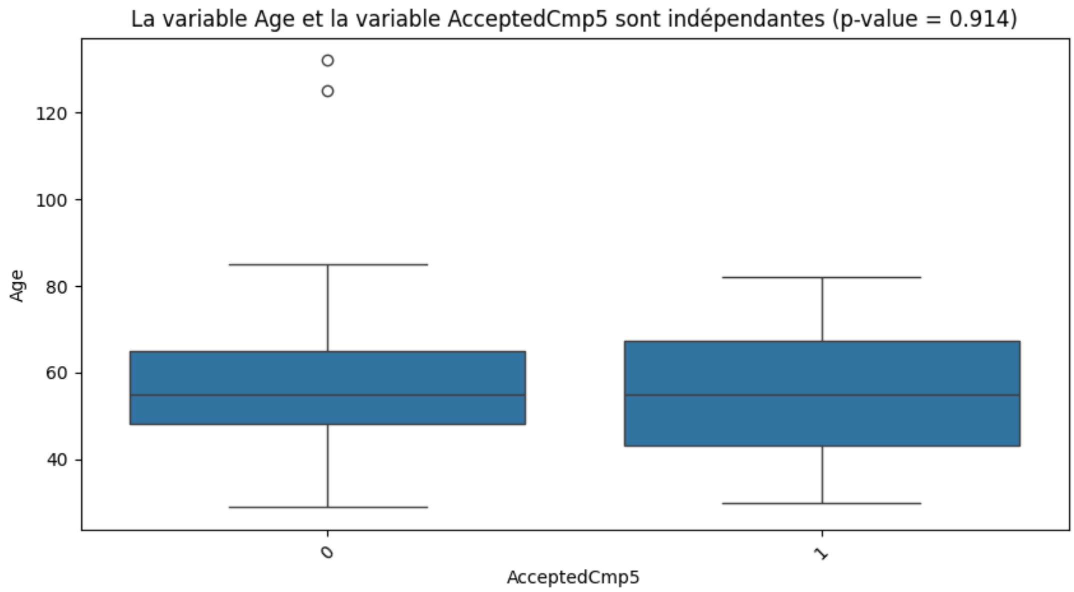



```
 La moyenne de la variable Age par rapport à la variable AcceptedCmp4
 AcceptedCmp4
```

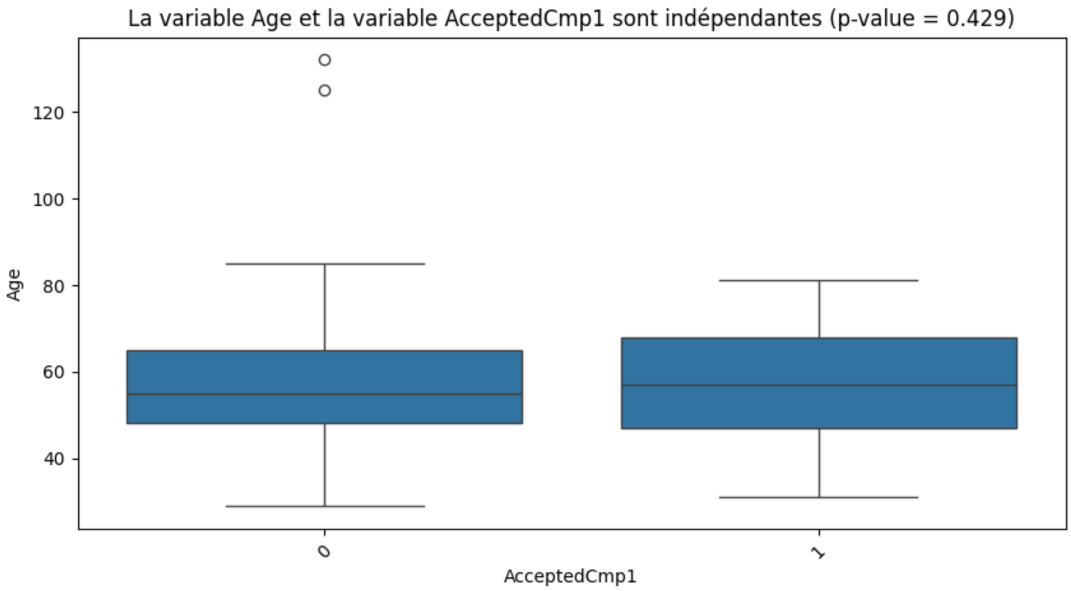
```
1    58.606
0    56.047
Name: Age, dtype: float64
```



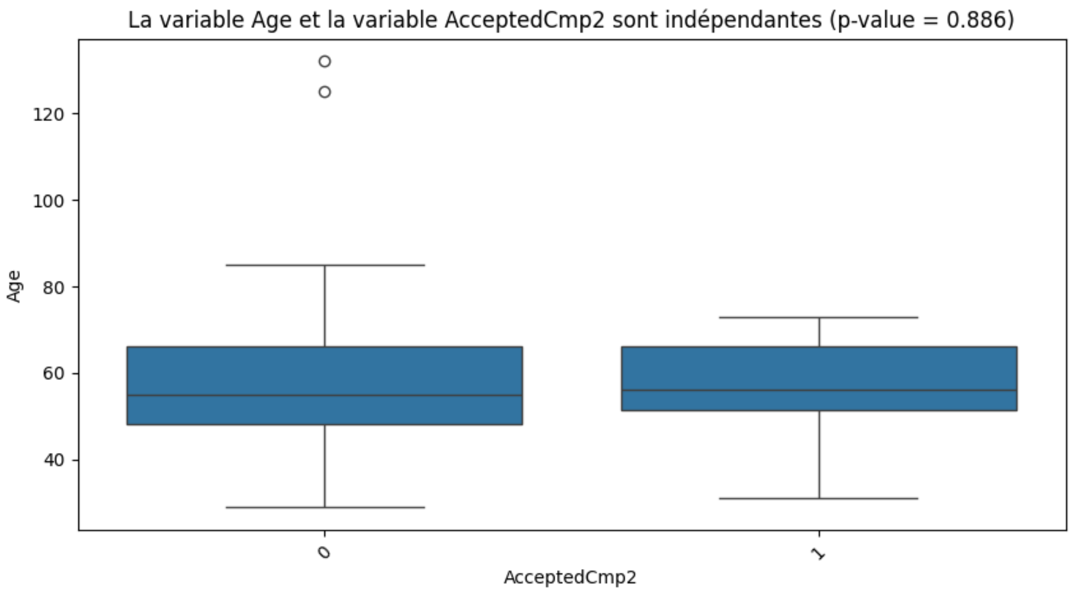

```
 La moyenne de la variable Age par rapport à la variable AcceptedCmp5
 AcceptedCmp5
0    56.240
1    56.125
Name: Age, dtype: float64
```



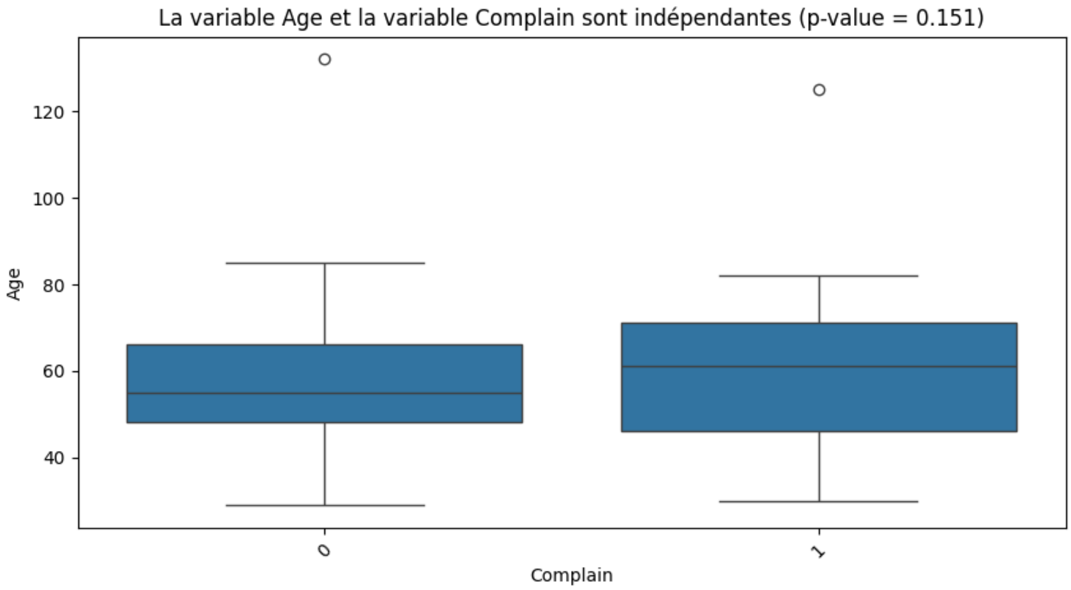
```
 La moyenne de la variable Age par rapport à la variable AcceptedCmp1
 AcceptedCmp1
1    57.078
0    56.185
Name: Age, dtype: float64
```



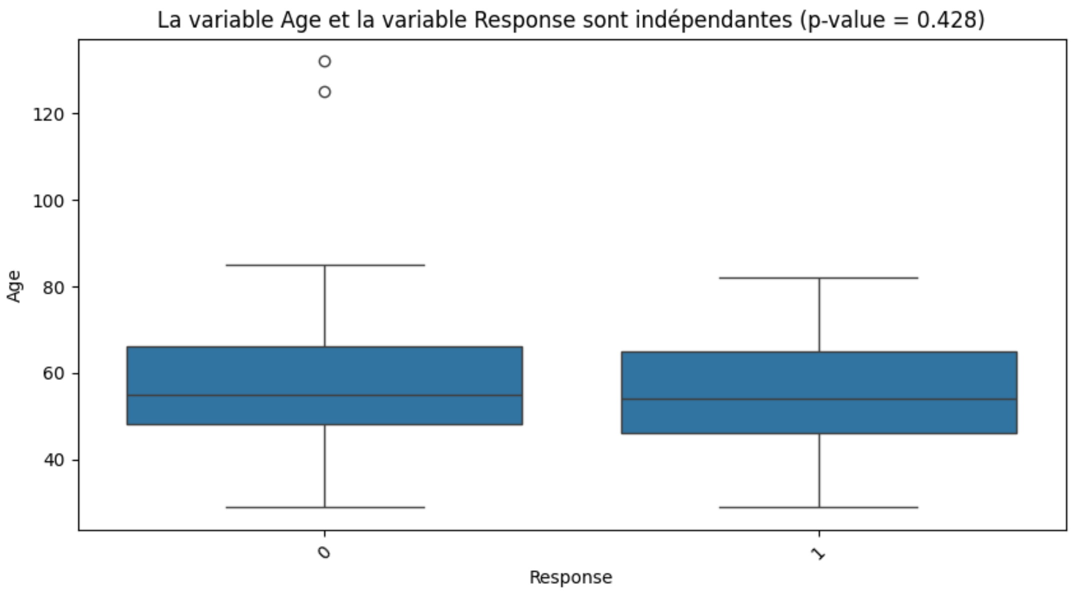

```
 La moyenne de la variable Age par rapport à la variable AcceptedCmp2
 AcceptedCmp2
1    56.556
0    56.229
Name: Age, dtype: float64
```



```
 La moyenne de la variable Age par rapport à la variable Complain
 Complain
1    59.905
0    56.197
Name: Age, dtype: float64
```



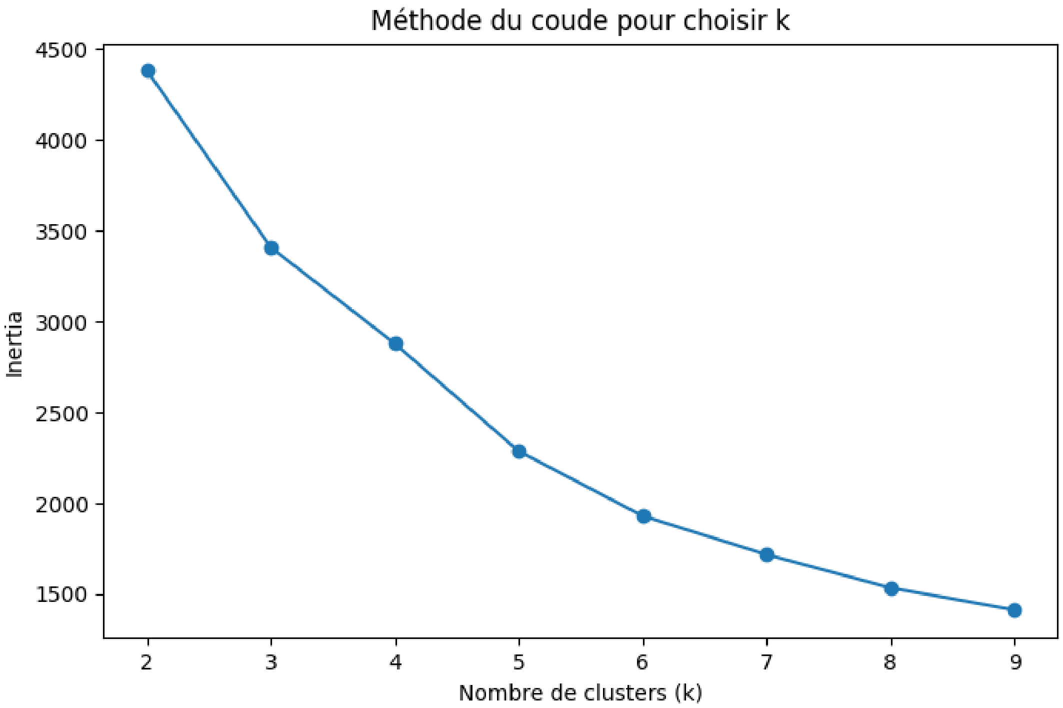

```
 La moyenne de la variable Age par rapport à la variable Response
 Response
0    56.315
1    55.732
Name: Age, dtype: float64
```



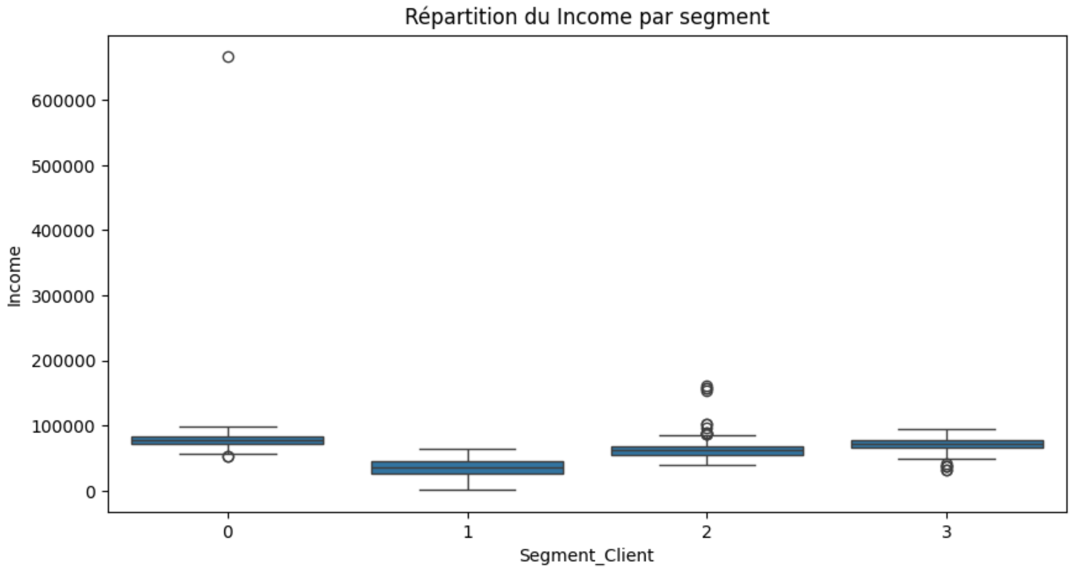
0.2 CLUSTERING

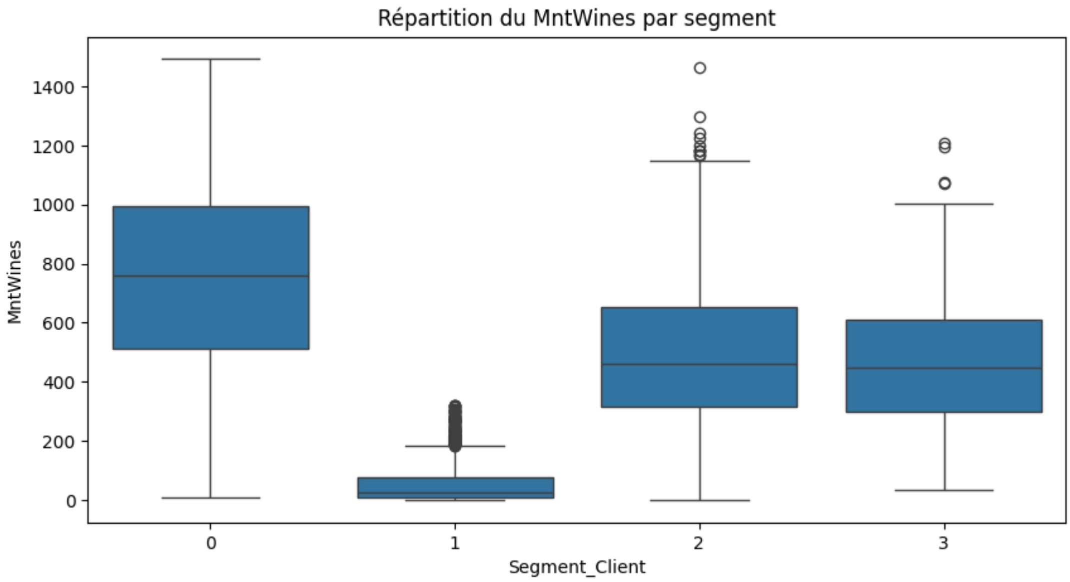
Il s'agit de segmenter les clients en fonction non seulement de leur revenu mais aussi des l'historique de leur achat.

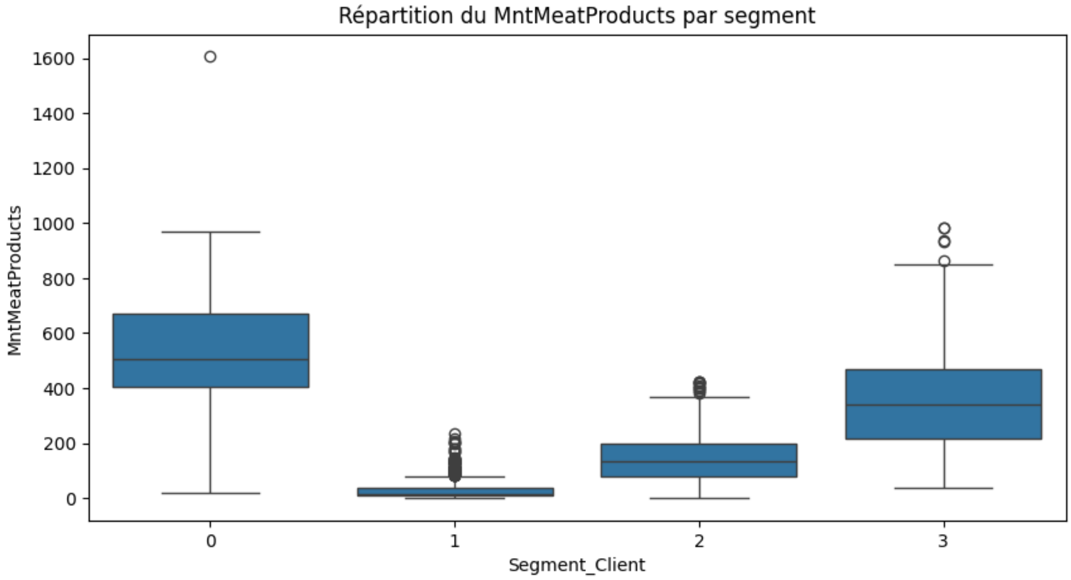
[97]:
```
# Sélection des variables pour le clustering
clustering_features = data[["Income", "MntWines", "MntMeatProducts",␣
↪"MntFishProducts"]]

# Normalisation des données
scaler = StandardScaler()
clustering_scaled = scaler.fit_transform(clustering_features)

# Trouver le bon nombre de clusters (Méthode du coude)
inertia = []
K_range = range(2, 10)
for k in K_range:
    kmeans = KMeans(n_clusters=k, random_state=42, n_init=10)
    kmeans.fit(clustering_scaled)
    inertia.append(kmeans.inertia_)

# Affichage de la courbe
plt.figure(figsize=(8, 5))
plt.plot(K_range, inertia, marker="o")
plt.xlabel("Nombre de clusters (k)")
plt.ylabel("Inertia")
plt.title("Méthode du coude pour choisir k")
plt.show()
```

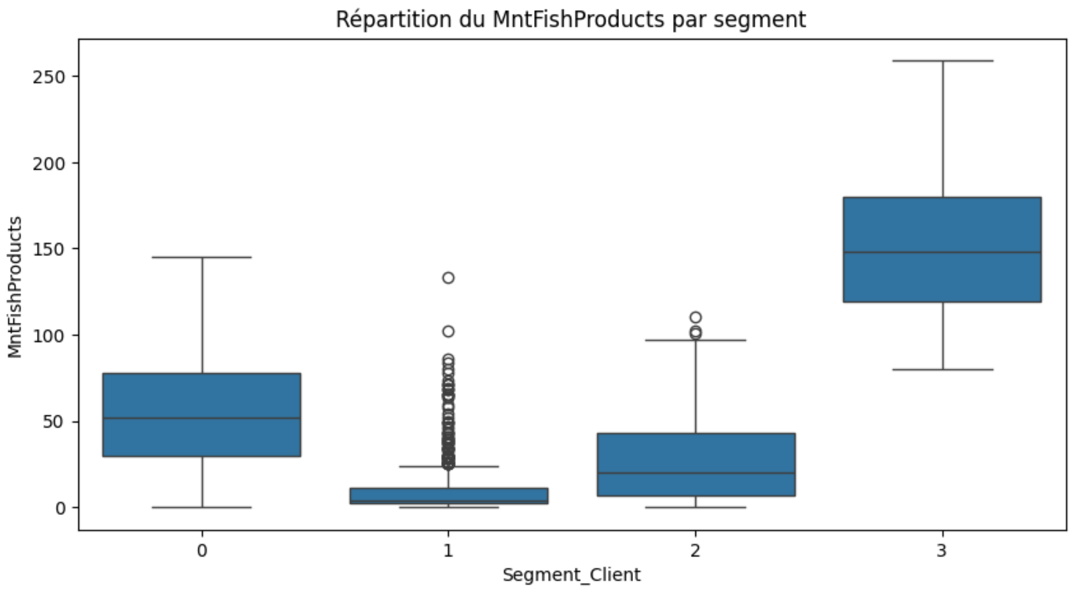


[98]:
```python
# Appliquer K-Means avec le k optimal trouvé
optimal_k = 4
kmeans = KMeans(n_clusters=optimal_k, random_state=42, n_init=10)
data["Segment_Client"] = kmeans.fit_predict(clustering_scaled)
for i in ["Income", "MntWines", "MntMeatProducts", "MntFishProducts"]:
    # Visualisation des segments
    plt.figure(figsize=(10,5))
    sns.boxplot(x="Segment_Client", y=i, data=data)
    plt.title(f"Répartition du {i} par segment")
    plt.show()
```






Répartition du MntMeatProducts par segment
MntMeatProducts
1600
1400
1200
1000
800
600
400
200
0
0
1
2
3
Segment_Client


Répartition du MntFishProducts par segment
MntFishProducts
250
200
150
100
50
0
0
1
2
3
Segment_Client

Resultat

- ***Revenu (income)*** :

  Les outliers dans les segments 0 et 2 indiquent une certaine hétérogénéité dans les revenus au sein de ces groupes, car quelque éléments du groupe détient un revenu très important.

  Les segments 1 et 3 présentent une distribution plus cohérente, avec peu ou pas d'écarts marqués.

  Ces informations peuvent guider les décisions basées sur les revenus, comme la personnalisation d'offres.

- ***Fruits (MntFishProducts)***:

  Le graphique met en évidence des comportements très différents entre les segments. En effet,

  Les clients du segment 3 dépensent le plus pour les produits de mer, tandis que ceux du segment 1 dépensent généralement très peu, mais avec quelques exceptions.

- ***Vin (MntWines)***:

  Les segments 0 et 1 représentent des extrêmes : dépenses très élevées en vin (segment 0) contre dépenses très faibles (segment 1).

  Les segments 2 et 3 sont modérés, avec des comportements légèrement différents en termes d'homogénéité.

- ***Viande (MntMeatProducts)***

  Le segment 1 et 2 sont ceux qui consomment plus la viande contraire au segment 0 qui a une consommation modérée en viande.

0.3 Modélisation

L'objectif principal, ici, c'est de prédire si un client acceptera ou non une offre lors de la dernière campagne après 5 campagne.

Instancier le modèle

[99]:
```python
# Variable cible et caractéristique
features = data.drop(columns=["Segment_Client","Dt_Customer","Response"])
target = data["Response"]

# Pretraitement
prepros = ColumnTransformer([
    ("cat", OneHotEncoder(),["Education","Marital_Status"]),
    ("norm", StandardScaler(), var_quanti)
], remainder='passthrough')

# Parametrage du modele
pipe_model = Pipeline([
    ('Preprocessing',prepros),
    ('model',␣
 ↪RandomForestClassifier(n_estimators=100,criterion='gini',max_features=int(np.
 ↪sqrt(features.shape[1]))))
])
```

[100]:
```python
# Séparation du data
X_train, X_test, y_train, y_test =␣
 ↪train_test_split(features,target,test_size=0.2, random_state=42)

print(f"La dimension du set de train (feature): {X_train.shape}")
print(f"La dimension du set de train (target): {y_train.shape}")
print(f"La dimension du set de test (feature): {X_test.shape}")
print(f"La dimension du set de test (target): {y_test.shape}")
```

```
La dimension du set de train (feature): (1702, 24)
La dimension du set de train (target): (1702,)
La dimension du set de test (feature): (426, 24)
La dimension du set de test (target): (426,)
```

Entrainement du modèle

[101]:
```python
# Entrainement du modèle
pipe_model.fit(X_train,y_train)
```

[101]:
```
Pipeline(steps=[('Preprocessing',
                 ColumnTransformer(remainder='passthrough',
                                   transformers=[('cat', OneHotEncoder(),
                                                  ['Education',
                                                   'Marital_Status']),
                                                 ('norm', StandardScaler(),
                                                  Index(['Income', 'Kidhome',
'Teenhome', 'Recency', 'MntWines', 'MntFruits',
       'MntMeatProducts', 'MntFishProducts', 'MntSweetProducts',
```

```
        'MntGoldProds', 'NumDealsPurchases', 'NumWebPurchases',
        'NumCatalogPurchases', 'NumStorePurchases', 'NumWebVisitsMonth',␣
↪'Age'],
       dtype='object'))])),
                ('model', RandomForestClassifier(max_features=4))])
```

Prediction, évaluation du modèle, identification des variable influencant l'efficacité du modèle

[102]:
```
# Prediction
y_pred = pipe_model.predict(X_test)
print(classification_report(y_test,y_pred))

# Detection des variables les plus importantes
model = pipe_model.named_steps["model"]
var_importante = np.round(model.feature_importances_,3)
var_importante_name = prepros.get_feature_names_out()

df_important = pd.DataFrame(
    {
        "variable importante": var_importante_name,
        "variable" : var_importante
    }
)

df_important.sort_values(by="variable",ascending=False)
```

```
              precision    recall  f1-score   support

           0       0.91      0.97      0.94       375
           1       0.58      0.29      0.39        51

    accuracy                           0.89       426
   macro avg       0.74      0.63      0.66       426
weighted avg       0.87      0.89      0.87       426
```

[102]:

	variable importante	variable
16	norm__Recency	0.084
13	norm__Income	0.083
17	norm__MntWines	0.081
19	norm__MntMeatProducts	0.079
22	norm__MntGoldProds	0.064
28	norm__Age	0.052
21	norm__MntSweetProducts	0.048
18	norm__MntFruits	0.047
20	norm__MntFishProducts	0.043
26	norm__NumStorePurchases	0.042
27	norm__NumWebVisitsMonth	0.040
25	norm__NumCatalogPurchases	0.038
24	norm__NumWebPurchases	0.036
23	norm__NumDealsPurchases	0.035

```
31        remainder__AcceptedCmp5     0.035
29        remainder__AcceptedCmp3     0.035
32        remainder__AcceptedCmp1     0.023
9      cat__Marital_Status_Single     0.017
15                 norm__Teenhome     0.014
10   cat__Marital_Status_Together     0.013
4              cat__Education_PhD     0.012
8     cat__Marital_Status_Married     0.011
30        remainder__AcceptedCmp4     0.010
2       cat__Education_Graduation     0.010
14                  norm__Kidhome     0.010
33        remainder__AcceptedCmp2     0.009
7    cat__Marital_Status_Divorced     0.008
3           cat__Education_Master     0.007
11      cat__Marital_Status_Widow     0.005
0         cat__Education_2n Cycle     0.004
1            cat__Education_Basic     0.002
12       cat__Marital_Status_YOLO     0.001
6       cat__Marital_Status_Alone     0.001
34            remainder__Complain     0.001
5      cat__Marital_Status_Absurd     0.000
```

Optimisation des hyper-parametres

[103]:
```
param = {
    'model__n_estimators' : [100,200,300],
    'model__max_features' : [3,4,5,6,7,8],
    'model__criterion': ['gini','entropy','log_loss']
}

grid = GridSearchCV(pipe_model, param_grid=param,cv = 5, n_jobs=-1)

grid.fit(X_train,y_train)

# Afficher les meilleurs paramètres
print("Meilleurs paramètres : \n", grid.best_params_)
```

```
Meilleurs paramètres :
 {'model__criterion': 'gini', 'model__max_features': 3, 'model__n_estimators':
100}
```

[104]:
```
# Pretraitement
prepros = ColumnTransformer([
    ("cat", OneHotEncoder(),["Education","Marital_Status"]),
    ("norm", StandardScaler(), var_quanti)
], remainder='passthrough')

# Parametrage du modele
pipe_model_opti = Pipeline([
    ('Preprocessing',prepros),
    ('model',␣
 ↪RandomForestClassifier(n_estimators=100,criterion='gini',max_features=3))
```

```
])

X_train, X_test, y_train, y_test =␣
↪train_test_split(features,target,test_size=0.2, random_state=42)
# Entrainement du modèle
pipe_model_opti.fit(X_train,y_train)
```

[104]:
```
Pipeline(steps=[('Preprocessing',
                 ColumnTransformer(remainder='passthrough',
                                   transformers=[('cat', OneHotEncoder(),
                                                  ['Education',
                                                   'Marital_Status']),
                                                 ('norm', StandardScaler(),
                                                  Index(['Income', 'Kidhome',
'Teenhome', 'Recency', 'MntWines', 'MntFruits',
       'MntMeatProducts', 'MntFishProducts', 'MntSweetProducts',
       'MntGoldProds', 'NumDealsPurchases', 'NumWebPurchases',
       'NumCatalogPurchases', 'NumStorePurchases', 'NumWebVisitsMonth',␣
↪'Age'],
      dtype='object'))])),
                ('model', RandomForestClassifier(max_features=3))])
```

[105]:

```python
y_pred_opti = pipe_model.predict(X_test)
print(classification_report(y_test,y_pred_opti))

print(f"La matrice de confusion :  \n ",confusion_matrix(y_test,y_pred_opti))
```

```
              precision    recall  f1-score   support

           0       0.91      0.97      0.94       375
           1       0.58      0.29      0.39        51

    accuracy                           0.89       426
   macro avg       0.74      0.63      0.66       426
weighted avg       0.87      0.89      0.87       426

La matrice de confusion :
   [[364  11]
 [ 36  15]]
```