```
import pandas as pd
import numpy as np
from statsmodels.formula.api import ols
from scipy import stats
import statsmodels.api as sm
from datetime import datetime
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import StratifiedKFold
from sklearn.preprocessing import StandardScaler, OneHotEncoder, StandardScaler,
MinMaxScaler, LabelEncoder
from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestRegressor, RandomForestClassifier
from sklearn.metrics import accuracy_score, recall_score, precision_score, f1_score,
confusion_matrix, classification_report, roc_auc_score, roc_curve

df = pd.read_csv('MCI.csv', encoding='utf-8-sig')

# remove rows with null values.. there are 100 of null values as we saw in exploratory
analysis
df = df.dropna()
#remove trailing spaces & delete columns that are not needed
df.columns = df.columns.str.strip() #For column names
df.columns = [col.strip() for col in df.columns] #For data in each column
del df["X"]
del df["Y"]
del df["Index_"]
del df["event_unique_id"]
del df["Division"]
del df["occurrencedate"]
del df["reporteddate"]
del df["ucr_code"]
del df["ucr_ext"]
del df["reporteddayofyear"]
del df["occurrencedayofyear"]
del df["Hood_ID"]
del df["Longitude"]
del df["Latitude"]
del df["ObjectId"]
del df["Neighbourhood"]
del df["location_type"]

##################LOGISTIC REGRESSION MCI CATEGORY#########################
#logistics regression
print('\n')
# one hot encoding
print("creating a df_dummy")
#remove mci from df_dummy since its now encoded
```

```python
df_lr = pd.get_dummies(df, drop_first=False) #logisic regression
df_dummy=pd.get_dummies(df['mci_category'])
#changing each mci type to INT for one hot encoding
df_dummy['Assault']=df_dummy['Assault'].astype(int)
df_dummy['Auto Theft']=df_dummy['Auto Theft'].astype(int)
df_dummy['Break and Enter']=df_dummy['Break and Enter'].astype(int)
df_dummy['Robbery']=df_dummy['Robbery'].astype(int)
df_dummy['Theft Over']=df_dummy['Theft Over'].astype(int)
print("printing to see df1 with INT encoding")
print(df_dummy.info())
```

```
 creating a df_dummy
 printing to see df1 with INT encoding
 <class 'pandas.core.frame.DataFrame'>
 Int64Index: 299828 entries, 0 to 299827
 Data columns (total 5 columns):
  #   Column           Non-Null Count   Dtype
 ---  ------           --------------   -----
  0   Assault          299828 non-null  int64
  1   Auto Theft       299828 non-null  int64
  2   Break and Enter  299828 non-null  int64
  3   Robbery          299828 non-null  int64
  4   Theft Over       299828 non-null  int64
 dtypes: int64(5)
 memory usage: 13.7 MB
 None
```

```python
print("_____Now lets view the contents of df_dummy")
df_dummy=pd.concat([df, df_dummy], axis=1) #adding the df_dummy to df_premise
print(df_dummy.info())
```

```
 _____Now lets view the contents of df_dummy
 <class 'pandas.core.frame.DataFrame'>
 Int64Index: 299828 entries, 0 to 299827
 Data columns (total 18 columns):
  #   Column               Non-Null Count   Dtype
 ---  ------               --------------   -----
  0   premises_type        299828 non-null  object
  1   offence              299828 non-null  object
  2   reportedyear         299828 non-null  int64
  3   reportedmonth        299828 non-null  object
  4   reportedday          299828 non-null  int64
  5   reporteddayofweek    299828 non-null  object
  6   reportedhour         299828 non-null  int64
  7   occurrenceyear       299828 non-null  float64
  8   occurrencemonth      299828 non-null  object
  9   occurrenceday        299828 non-null  float64
  10  occurrencedayofweek  299828 non-null  object
  11  occurrencehour       299828 non-null  int64
  12  mci_category         299828 non-null  object
  13  Assault              299828 non-null  int64
  14  Auto Theft           299828 non-null  int64
  15  Break and Enter      299828 non-null  int64
  16  Robbery              299828 non-null  int64
  17  Theft Over           299828 non-null  int64
 dtypes: float64(2), int64(9), object(7)
 memory usage: 43.5+ MB
 None
```

```
#Since auto theft, break and enter, robbery and theft over are related to theft and
stealing, the ‘Assault’ category stands out.
#it will all be all considered as "theft/stealing" while the assault category will be
left as "Assault" and is independent.
#Therefore, we will compare "Assault" vs "Stealing"
#Here we remove all other categories so 1 is for Assault and 0 for "Stealing"

print("Regression Analysis of Assault mci-category with Occurrence Hour and Auto Theft
mci-category")
reg1 = sm.OLS(df_dummy["Assault"], sm.add_constant(df_dummy[["occurrencehour","Auto
Theft"]])).fit()
print(reg1.summary())
```

```
Regression Analysis of Assault mci-category with Occurrence Hour and Auto Theft mci-category
                            OLS Regression Results
==============================================================================
Dep. Variable:                Assault   R-squared:                       0.185
Model:                            OLS   Adj. R-squared:                  0.185
Method:                 Least Squares   F-statistic:                 3.401e+04
Date:                Sun, 02 Apr 2023   Prob (F-statistic):               0.00
Time:                        20:29:37   Log-Likelihood:            -1.8618e+05
No. Observations:              299828   AIC:                         3.724e+05
Df Residuals:                  299825   BIC:                         3.724e+05
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.5722      0.002    343.204      0.000       0.569       0.575
occurrencehour 0.0039      0.000     34.115      0.000       0.004       0.004
Auto Theft    -0.6271      0.002   -260.469      0.000      -0.632      -0.622
==============================================================================
Omnibus:                  1631805.431   Durbin-Watson:                   1.738
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            43933.709
Skew:                          -0.531   Prob(JB):                         0.00
Kurtosis:                       1.455   Cond. No.                         43.0
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

```
print("Regression Analysis of Assault mci-category with Occurrence Hour and Break and
Enter mci-category")
reg2 = sm.OLS(df_dummy["Assault"], sm.add_constant(df_dummy[["occurrencehour","Break
and Enter"]])).fit()
print(reg2.summary())
```

```
Regression Analysis of Assault mci-category with Occurrence Hour and Break and Enter mci-category
                            OLS Regression Results
==============================================================================
Dep. Variable:                Assault   R-squared:                       0.288
Model:                            OLS   Adj. R-squared:                  0.288
Method:                 Least Squares   F-statistic:                 6.050e+04
Date:                Sun, 02 Apr 2023   Prob (F-statistic):               0.00
Time:                        20:29:37   Log-Likelihood:            -1.6600e+05
No. Observations:              299828   AIC:                         3.320e+05
Df Residuals:                  299825   BIC:                         3.320e+05
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const              0.7121      0.002    432.978      0.000       0.709       0.715
occurrencehour    -0.0033      0.000    -31.114      0.000      -0.004      -0.003
Break and Enter   -0.6763      0.002   -347.582      0.000      -0.680      -0.672
==============================================================================
Omnibus:                   154553.639   Durbin-Watson:                   1.239
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            46939.067
Skew:                          -0.796   Prob(JB):                         0.00
Kurtosis:                       1.893   Cond. No.                         40.1
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

```
print("Regression Analysis of Assault mci-category with Occurrence Hour and Robbery
mci-category")
reg3= sm.OLS(df_dummy["Assault"],
sm.add_constant(df_dummy[["occurrencehour","Robbery"]])).fit()
print(reg3.summary())
```

```
Regression Analysis of Assault mci-category with Occurrence Hour and Robbery mci-category
                            OLS Regression Results
==============================================================================
Dep. Variable:                Assault   R-squared:                       0.126
Model:                            OLS   Adj. R-squared:                  0.126
Method:                 Least Squares   F-statistic:                 2.163e+04
Date:                Sun, 02 Apr 2023   Prob (F-statistic):               0.00
Time:                        20:29:37   Log-Likelihood:            -1.9662e+05
No. Observations:              299828   AIC:                         3.932e+05
Df Residuals:                  299825   BIC:                         3.933e+05
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const              0.5609      0.002    325.098      0.000       0.558       0.564
occurrencehour     0.0026      0.000     22.411      0.000       0.002       0.003
Robbery           -0.5970      0.003   -207.616      0.000      -0.603      -0.591
==============================================================================
Omnibus:                  1262824.646   Durbin-Watson:                   1.453
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            45204.863
Skew:                          -0.402   Prob(JB):                         0.00
Kurtosis:                       1.276   Cond. No.                         49.4
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

```
print('\n')
print("Regression Analysis of Assault mci-category with Occurrence Hour and  Theft Over mci-category")
reg4= sm.OLS(df_dummy["Assault"], sm.add_constant(df_dummy[["occurrencehour","Theft Over"]])).fit()
print(reg4.summary())
```

```
Regression Analysis of Assault mci-category with Occurrence Hour and  Theft Over mci-category
                            OLS Regression Results
==============================================================================
Dep. Variable:                Assault   R-squared:                       0.040
Model:                            OLS   Adj. R-squared:                  0.040
Method:                 Least Squares   F-statistic:                     6166.
Date:                Sun, 02 Apr 2023   Prob (F-statistic):               0.00
Time:                        20:54:03   Log-Likelihood:            -2.1078e+05
No. Observations:              299828   AIC:                         4.216e+05
Df Residuals:                  299825   BIC:                         4.216e+05
Df Model:                           2
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.5364      0.002    297.412      0.000       0.533       0.540
occurrencehour 0.0014      0.000     11.440      0.000       0.001       0.002
Theft Over    -0.5540      0.005   -110.399      0.000      -0.564      -0.544
==============================================================================
Omnibus:                  1077167.689   Durbin-Watson:                   1.503
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            48283.805
Skew:                          -0.222   Prob(JB):                         0.00
Kurtosis:                       1.085   Cond. No.                         82.0
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```