```python
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()
print("Viewing all columns")
print(df.info()) #to confirm its deleted for null values
print('\n')
#emove rows with null values.. there are 100 of null values as we saw in exploratory
analysis
df = df.dropna()
#remove trailing spaces & delete columns 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
print('\n')
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
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())
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.head())
print('\n')

#compare regression of each mci category with occurencehour


print('\n')
print("Regression Analysis of Assault mci-category with Occurrence Hour")
reg0 = sm.OLS(df_dummy["Assault"],
sm.add_constant(df_dummy[["occurrencehour"]])).fit()
print(reg0.summary())

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


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

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


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

```
The default interactive shell is now zsh.
To update your account to use zsh, please run `chsh -s /bin/zsh`.
For more details, please visit https://support.apple.com/kb/HT208050.
ndasprojectok:pandasproject royasalehzai$ cd /Users/royasalehzai/studysession/pa
/usr/local/bin/python3 /Users/royasalehzai/studysession/pandasproject/LogisticsRegression.py
Royas-MacBook:pandasproject royasalehzai$ /usr/local/bin/python3
/Users/royasalehzai/studysession/pandasproject/LogisticsRegression.py
Viewing all columns
<class 'pandas.core.frame.DataFrame'>
Int64Index: 299828 entries, 0 to 299827
Data columns (total 30 columns):
 #   Column             Non-Null Count   Dtype
---  ------             --------------   -----
 0   X                  299828 non-null  float64
 1   Y                  299828 non-null  float64
 2   Index_             299828 non-null  int64
 3   event_unique_id    299828 non-null  object
 4   Division           299828 non-null  object
 5   occurrencedate     299828 non-null  object
 6   reporteddate       299828 non-null  object
 7   location_type      299828 non-null  object
 8   premises_type      299828 non-null  object
 9   ucr_code           299828 non-null  int64
 10  ucr_ext            299828 non-null  int64
 11  offence            299828 non-null  object
 12  reportedyear       299828 non-null  int64
 13  reportedmonth      299828 non-null  object
 14  reportedday        299828 non-null  int64
 15  reporteddayofyear  299828 non-null  int64
 16  reporteddayofweek  299828 non-null  object
```

```
 17  reportedhour        299828 non-null  int64
 18  occurrenceyear      299828 non-null  float64
 19  occurrencemonth     299828 non-null  object
 20  occurrenceday       299828 non-null  float64
 21  occurrencedayofyear  299828 non-null  float64
 22  occurrencedayofweek  299828 non-null  object
 23  occurrencehour      299828 non-null  int64
 24  mci_category        299828 non-null  object
 25  Hood_ID             299828 non-null  object
 26  Neighbourhood       299828 non-null  object
 27  Longitude           299828 non-null  float64
 28  Latitude            299828 non-null  float64
 29  ObjectId            299828 non-null  int64
dtypes: float64(7), int64(9), object(14)
memory usage: 70.9+ MB
None
```

```
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
_______________________________Now lets view the contents of df_dummy
  premises_type           offence  reportedyear reportedmonth  reportedday  ... Assault  Auto
Theft  Break and Enter Robbery  Theft Over
0    Apartment           Assault        2014      January         3  ...      1        0            0      0
0
1       House              B&E        2014      January         3  ...      0        0            1      0         0
2     Outside           Assault        2014      January         3  ...      1        0            0      0         0
```

```
3   Commercial        Theft Over        2014      January         3  ...      0         0            0      0
1
4   Commercial  Robbery - Business        2014      January         3  ...      0         0            0
1         0

[5 rows x 18 columns]
```

Regression Analysis of Assault mci-category with Occurrence Hour

```
                         OLS Regression Results
==============================================================================
Dep. Variable:              Assault   R-squared:                       0.000
Model:                          OLS   Adj. R-squared:                  0.000
Method:               Least Squares   F-statistic:                     139.1
Date:              Wed, 12 Apr 2023   Prob (F-statistic):           4.24e-32
Time:                      02:28:27   Log-Likelihood:            -2.1676e+05
No. Observations:            299828   AIC:                       4.335e+05
Df Residuals:                299826   BIC:                       4.335e+05
Df Model:                         1
Covariance Type:          nonrobust
==============================================================================
====
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const            0.5174      0.002    282.491      0.000       0.514       0.521
occurrencehour    0.0015      0.000     11.794      0.000       0.001       0.002
==============================================================================
Omnibus:                1035622.654   Durbin-Watson:                   1.522
Prob(Omnibus):                0.000   Jarque-Bera (JB):            49884.582
Skew:                        -0.145   Prob(JB):                         0.00
Kurtosis:                     1.023   Cond. No.                         29.4
==============================================================================

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

Regression Analysis of Auto Theft mci-category with Occurrence Hour

```
                         OLS Regression Results
==============================================================================
Dep. Variable:           Auto Theft   R-squared:                       0.007
Model:                          OLS   Adj. R-squared:                  0.007
```

```
Method:             Least Squares   F-statistic:                 1980.
Date:            Wed, 12 Apr 2023   Prob (F-statistic):           0.00
Time:                    02:28:27   Log-Likelihood:         -1.0332e+05
No. Observations:          299828   AIC:                     2.066e+05
Df Residuals:              299826   BIC:                     2.067e+05
Df Model:                       1
Covariance Type:        nonrobust
==============================================================================
====
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const            0.0874      0.001     69.687      0.000       0.085       0.090
occurrencehour     0.0038   8.62e-05     44.497      0.000       0.004       0.004
==============================================================================
Omnibus:                111083.844   Durbin-Watson:                0.511
Prob(Omnibus):               0.000   Jarque-Bera (JB):         298697.413
Skew:                        2.105   Prob(JB):                       0.00
Kurtosis:                    5.487   Cond. No.                       29.4
==============================================================================

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


Regression Analysis of Break and Enter mci-category with Occurrence Hour
                            OLS Regression Results
==============================================================================
Dep. Variable:       Break and Enter   R-squared:                       0.017
Model:                           OLS   Adj. R-squared:                  0.017
Method:             Least Squares   F-statistic:                 5104.
Date:            Wed, 12 Apr 2023   Prob (F-statistic):           0.00
Time:                    02:28:27   Log-Likelihood:         -1.4701e+05
No. Observations:          299828   AIC:                     2.940e+05
Df Residuals:              299826   BIC:                     2.940e+05
Df Model:                       1
Covariance Type:        nonrobust
==============================================================================
====
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const            0.2879      0.001    198.360      0.000       0.285       0.291
occurrencehour    -0.0071   9.97e-05    -71.439      0.000      -0.007      -0.007
==============================================================================
Omnibus:                 61969.760   Durbin-Watson:                1.757
```

```
Prob(Omnibus):              0.000   Jarque-Bera (JB):         110102.548
Skew:                       1.477   Prob(JB):                      0.00
Kurtosis:                   3.290   Cond. No.                      29.4
==============================================================================
```

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

Regression Analysis of Robbery mci-category with Occurrence Hour

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                Robbery   R-squared:                       0.002
Model:                            OLS   Adj. R-squared:                  0.002
Method:                 Least Squares   F-statistic:                     671.0
Date:                Wed, 12 Apr 2023   Prob (F-statistic):          8.63e-148
Time:                        02:28:28   Log-Likelihood:                -60513.
No. Observations:              299828   AIC:                         1.210e+05
Df Residuals:                  299826   BIC:                         1.211e+05
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
====
                 coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
const              0.0729      0.001     67.012      0.000       0.071       0.075
occurrencehour     0.0019   7.48e-05     25.904      0.000       0.002       0.002
==============================================================================
Omnibus:                   156646.412   Durbin-Watson:                   1.622
Prob(Omnibus):              0.000   Jarque-Bera (JB):         725078.786
Skew:                       2.708   Prob(JB):                      0.00
Kurtosis:                   8.358   Cond. No.                      29.4
==============================================================================
```

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

Regression Analysis of Theft Over mci-category with Occurrence Hour

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             Theft Over   R-squared:                       0.000
Model:                            OLS   Adj. R-squared:                  0.000
Method:                 Least Squares   F-statistic:                     8.611
```

```
Date:               Wed, 12 Apr 2023   Prob (F-statistic):          0.00334
Time:                       02:28:28   Log-Likelihood:              92260.
No. Observations:             299828   AIC:                    -1.845e+05
Df Residuals:                 299826   BIC:                    -1.845e+05
Df Model:                          1
Covariance Type:           nonrobust
==================================================================================
====
                  coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
const           0.0344      0.001     52.606      0.000       0.033       0.036
occurrencehour   -0.0001   4.49e-05     -2.935      0.003      -0.000   -4.38e-05
==================================================================================
Omnibus:                298552.524   Durbin-Watson:                   1.870
Prob(Omnibus):               0.000   Jarque-Bera (JB):          9567890.777
Skew:                        5.254   Prob(JB):                         0.00
Kurtosis:                   28.602   Cond. No.                         29.4
==================================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
Royas-MacBook:pandasproject royasalehzai$
```