```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from keras.models import Sequential
from keras.layers.core import Dense,Activation,Dropout
from keras.layers import LSTM
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
import seaborn as sns
from sklearn.ensemble import RandomForestClassifier
from sklearn import svm
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from xgboost import XGBClassifier
from sklearn.neural_network import MLPClassifier
from sklearn.metrics import accuracy_score
from sklearn.tree import DecisionTreeClassifier
from keras.callbacks import EarlyStopping
import math
import os
from sklearn.feature_extraction.text import TfidfVectorizer, CountVectorizer
from sklearn.metrics.pairwise import linear_kernel, cosine_similarity
from nltk.stem.snowball import SnowballStemmer
from nltk.stem.wordnet import WordNetLemmatizer
from nltk.corpus import wordnet
from nltk.stem import WordNetLemmatizer
```

```
hotel_details=pd.read_csv('/content/Hotel_Room_attributes.csv',delimiter=',')
hotel_rooms=pd.read_csv('/content/Hotel_Room_attributes.csv',delimiter=',')
hotel_cost=pd.read_csv('/content/Hotel_Room_attributes.csv',delimiter=',')
```

```
hotel_details.head()
```

	id	hotelcode	roomamenities	roomtype	
0	50677497	634876	Air conditioning: ;Alarm clock: ;Carpeting: ;C...	Double Room	Room size: 15
1	50672149	8328096	Air conditioning: ;Closet: ;Fireplace: ;Free W...	Vacation Home	Shower, Kitch
2	50643430	8323442	Air conditioning: ;Closet: ;Dishwasher: ;Firep...	Vacation Home	Shower, Kitch
[illegible]	[illegible]	[illegible]	Air conditioning: ;Clothes rack:	Standard Triple	Room size: 20

```
hotel_rooms.head()
```

	hotelcode	roomamenities	roomtype	
0	634876	Air conditioning: ;Alarm clock: ;Carpeting: ;C...	Double Room	Room size: 15 m²/1
1	8328096	Air conditioning: ;Closet: ;Fireplace: ;Free W	Vacation Home	Shower, Kitchenette,
2	8323442	;Firep	Vacation Home	

```
del hotel['hotelid']
del hotel['url']
del hotel['curr']
del hotel['Source']
```

```
hotel.columns
```

```
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
```

```
def plotPerColumnDistribution(df, nGraphShown, nGraphPerRow):
    nunique = df.nunique()
    df = df[[col for col in df if nunique[col] > 1 and nunique[col] < 50]] # For displayir
    nRow, nCol = df.shape
    columnNames = list(df)
    nGraphRow = (nCol + nGraphPerRow - 1) / nGraphPerRow
    plt.figure(num = None, figsize = (6 * nGraphPerRow, 8 * nGraphRow), dpi = 80, facecolc
    for i in range(min(nCol, nGraphShown)):
        plt.subplot(nGraphRow, nGraphPerRow, i + 1)
        columnDf = df.iloc[:, i]
        if (not np.issubdtype(type(columnDf.iloc[0]), np.number)):
            valueCounts = columnDf.value_counts()
            valueCounts.plot.bar()
        else:
            columnDf.hist()
        plt.ylabel('counts')
        plt.xticks(rotation = 90)
        plt.title(f'{columnNames[i]} (column {i})')
    plt.tight_layout(pad = 1.0, w_pad = 1.0, h_pad = 1.0)
    plt.show()
```

```
def plotCorrelationMatrix(df, graphWidth):
    filename = df.dataframeName
    df = df.dropna('columns') # drop columns with NaN
    df = df[[col for col in df if df[col].nunique() > 1]] # keep columns where there are m
    if df.shape[1] < 2:
        print(f'No correlation plots shown: The number of non-NaN or constant columns ({df
        return
    corr = df.corr()
    plt.figure(num=None, figsize=(graphWidth, graphWidth), dpi=80, facecolor='w', edgecolc
    corrMat = plt.matshow(corr, fignum = 1)
    plt.xticks(range(len(corr.columns)), corr.columns, rotation=90)
    plt.yticks(range(len(corr.columns)), corr.columns)
    plt.gca().xaxis.tick_bottom()
    plt.colorbar(corrMat)
    plt.title(f'Correlation Matrix for {filename}', fontsize=15)
    plt.show()

def plotScatterMatrix(df, plotSize, textSize):
    df = df.select_dtypes(include =[np.number]) # keep only numerical columns
    # Remove rows and columns that would lead to df being singular
    df = df.dropna('columns')
    df = df[[col for col in df if df[col].nunique() > 1]] # keep columns where there are m
    columnNames = list(df)
    if len(columnNames) > 10: # reduce the number of columns for matrix inversion of kerne
        columnNames = columnNames[:10]
    df = df[columnNames]
    ax = pd.plotting.scatter_matrix(df, alpha=0.75, figsize=[plotSize, plotSize], diagonal
    corrs = df.corr().values
    for i, j in zip(*plt.np.triu_indices_from(ax, k = 1)):
        ax[i, j].annotate('Corr. coef = %.3f' % corrs[i, j], (0.8, 0.2), xycoords='axes fr
    plt.suptitle('Scatter and Density Plot')
    plt.show()


nRowsRead = 1000 # specify 'None' if want to read whole file
# Hotel_Room_attributes.csv may have more rows in reality, but we are only loading/preview
df1 = pd.read_csv('Hotel_Room_attributes.csv', delimiter=',', nrows = nRowsRead)
df1.dataframeName = 'Hotel_Room_attributes.csv'
nRow, nCol = df1.shape
print(f'There are {nRow} rows and {nCol} columns')
```

```
There are 1000 rows and 5 columns
There are 1000 rows and 5 columns
```

```
df1.head(5)
```

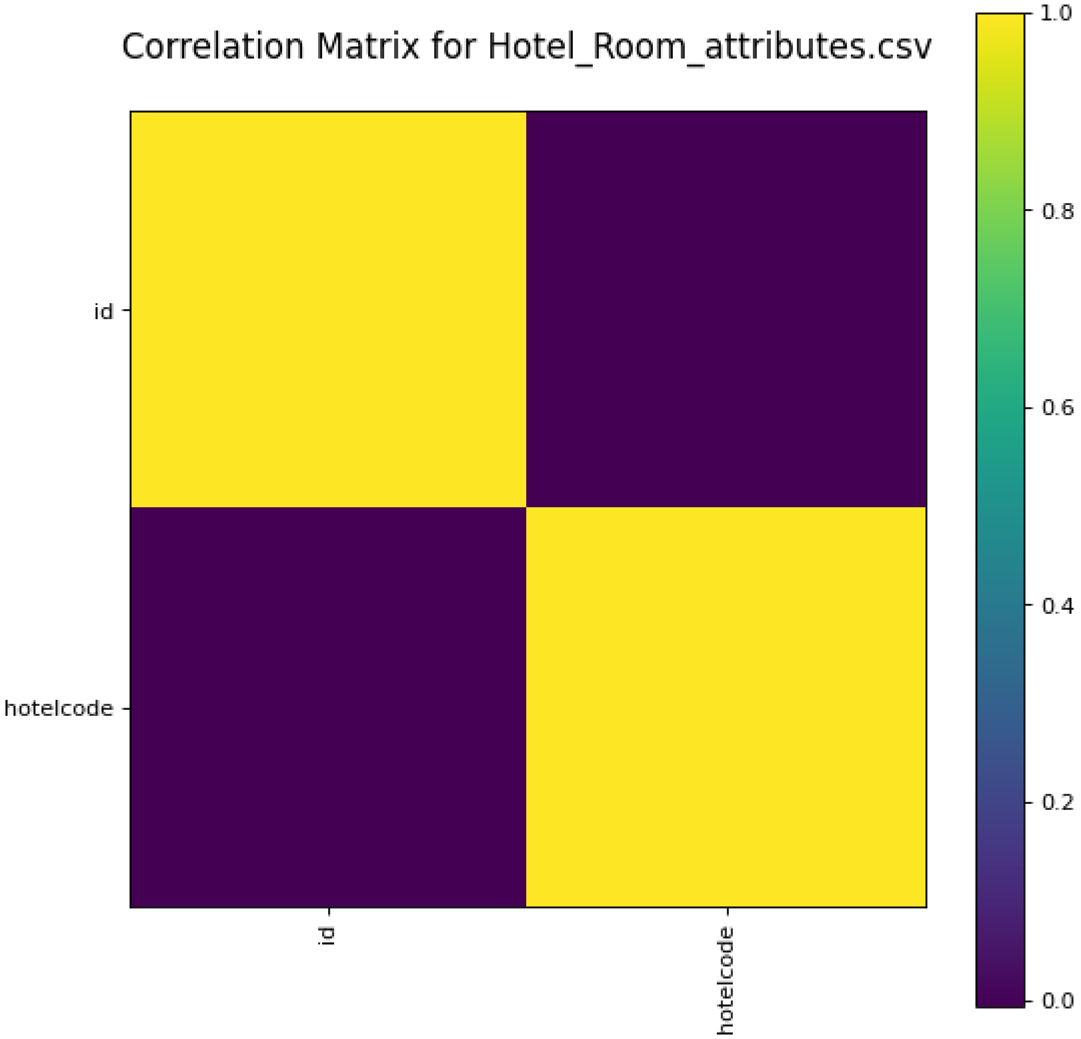
	id	hotelcode	roomamenities	roomtype	
0	50677497	634876	Air conditioning: ;Alarm clock: ;Carpeting: ;C...	Double Room	Room size: 15
1	50672149	8328096	Air conditioning: ;Closet: ;Fireplace: ;Free W...	Vacation Home	Shower, Kitch
2	50643430	8323442	Air conditioning: ;Closet: ;Dishwasher: ;Firep...	Vacation Home	Shower, Kitch
[illegible]	[illegible]	[illegible]	Air conditioning: ;Clothes rack:	Standard Triple	Room size: 20

```
plotPerColumnDistribution(df1, 10, 5)
```

```
<Figure size 2400x512 with 0 Axes>
```

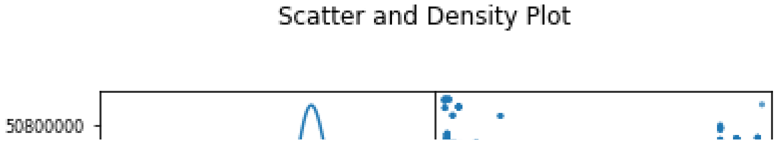
```
plotCorrelationMatrix(df1, 8)
```

```
usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:3: FutureWarning: In a fu
  This is separate from the ipykernel package so we can avoid doing imports until
```



```
plotScatterMatrix(df1, 6, 15)
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:4: FutureWarning: In a
  after removing the cwd from sys.path.
```



```
nRowsRead = 1000 # specify 'None' if want to read whole file
# Hotel_details.csv may have more rows in reality, but we are only loading/previewing the
df2 = pd.read_csv('/content/Hotel_Room_attributes.csv', delimiter=',', nrows = nRowsRead)
df2.dataframeName = '/content/Hotel_Room_attributes.csv'
nRow, nCol = df2.shape
print(f'There are {nRow} rows and {nCol} columns')
```

```
There are 1000 rows and 5 columns
```

```
df2.head(5)
```

	id	hotelcode	roomamenities	roomtype	
0	50677497	634876	Air conditioning: ;Alarm clock: ;Carpeting: ;C...	Double Room	Room size: 15
1	50672149	8328096	Air conditioning: ;Closet: ;Fireplace: ;Free W...	Vacation Home	Shower, Kitch
2	50643430	8323442	Air conditioning: ;Closet: ;Dishwasher: ;Firep...	Vacation Home	Shower, Kitch
[illegible]	[illegible]	[illegible]	Air conditioning: ;Clothes rack:	Standard Triple	Room size: 20

```
plotPerColumnDistribution(df2, 10, 5)
```

```
<Figure size 2400x512 with 0 Axes>
```

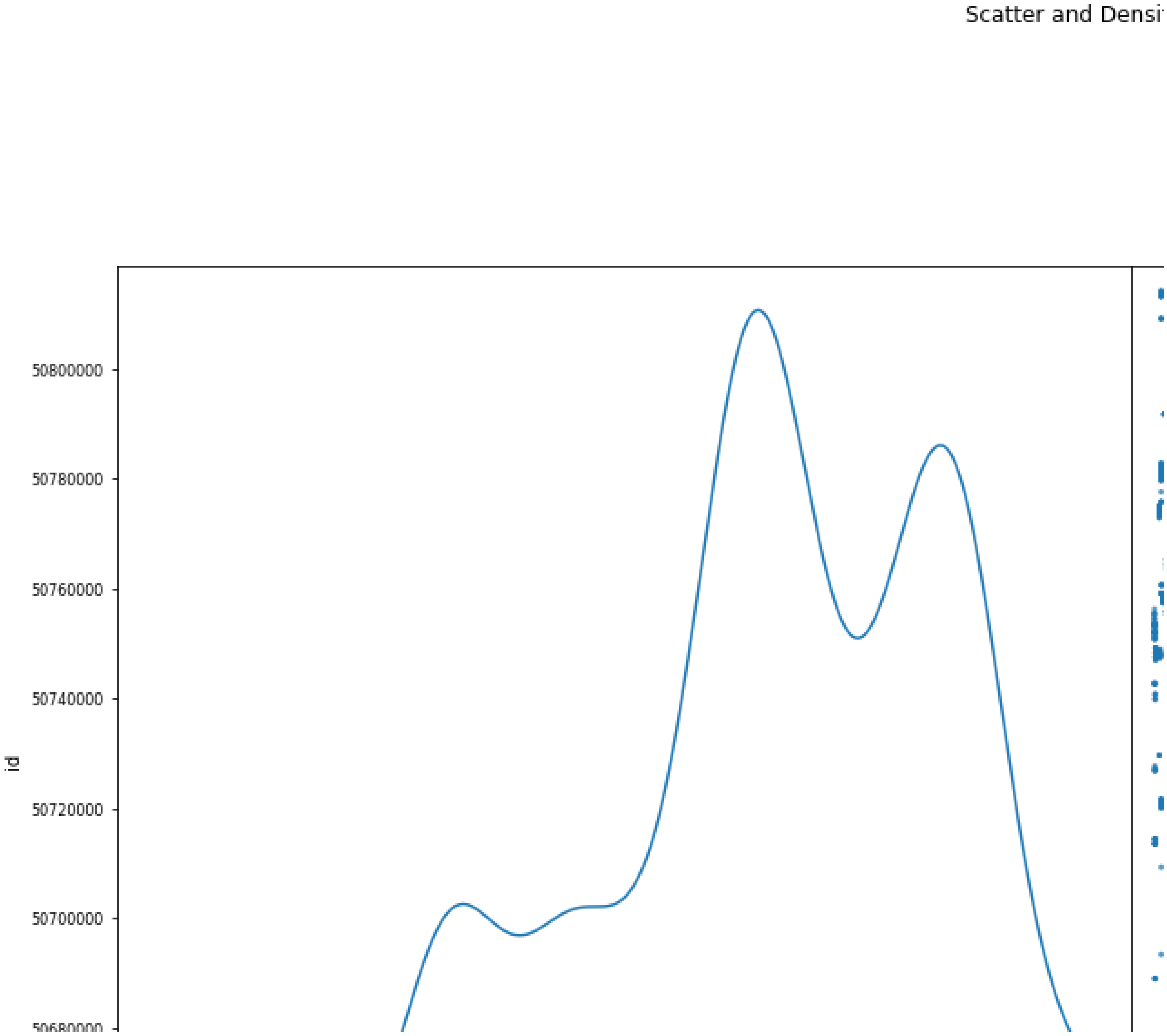
```
plotCorrelationMatrix(df2, 8)
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:3: FutureWarning: In a
  This is separate from the ipykernel package so we can avoid doing imports until
```

Correlation Matrix for /content/Hotel_Room_attributes.cs
id
1.0
0.8

```
plotScatterMatrix(df2, 20, 10)
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:4: FutureWarning: In a
  after removing the cwd from sys.path.
```



```
nRowsRead = 1000 # specify 'None' if want to read whole file
# hotels_RoomPrice.csv may have more rows in reality, but we are only loading/previewing t
df3 = pd.read_csv('/content/Hotel_Room_attributes.csv', delimiter=',', nrows = nRowsRead)
df3.dataframeName = '/content/Hotel_Room_attributes.csv'
nRow, nCol = df3.shape
print(f'There are {nRow} rows and {nCol} columns')
```

```
There are 1000 rows and 5 columns
```

```
df3.head(5)
```

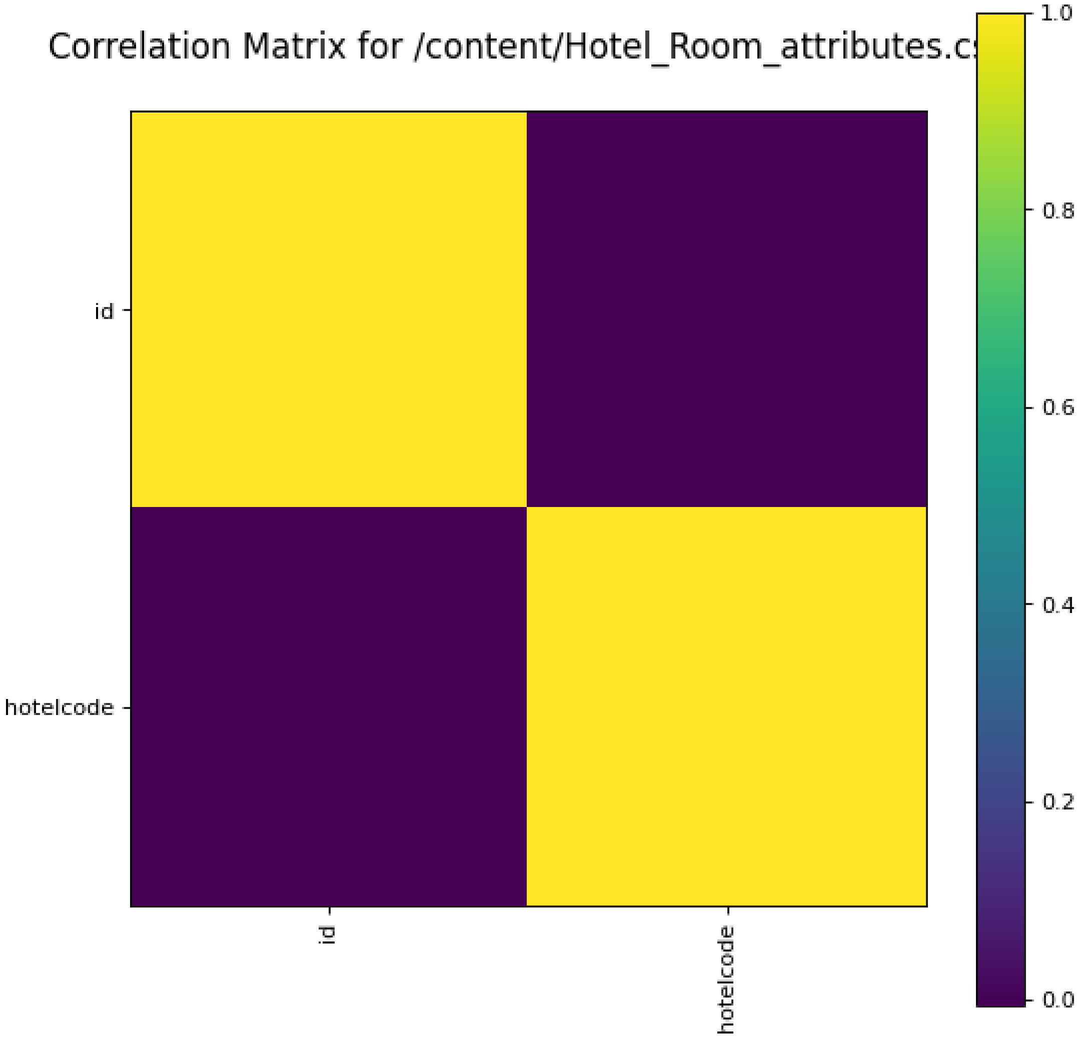
	id	hotelcode	roomamenities	roomtype	
0	50677497	634876	Air conditioning: ;Alarm clock: ;Carpeting: ;C...	Double Room	Room size: 15
1	50672149	8328096	Air conditioning: ;Closet: ;Fireplace: ;Free W...	Vacation Home	Shower, Kitch
2	50643430	8323442	Air conditioning: ;Closet: ;Dishwasher: ;Firep...	Vacation Home	Shower, Kitch
3	[illegible]	[illegible]	Air conditioning: ;Clothes rack:	Standard Triple	Room size: 20

```
plotPerColumnDistribution(df3, 10, 5)
```

```
<Figure size 2400x512 with 0 Axes>
```

```
plotCorrelationMatrix(df3, 8)
```

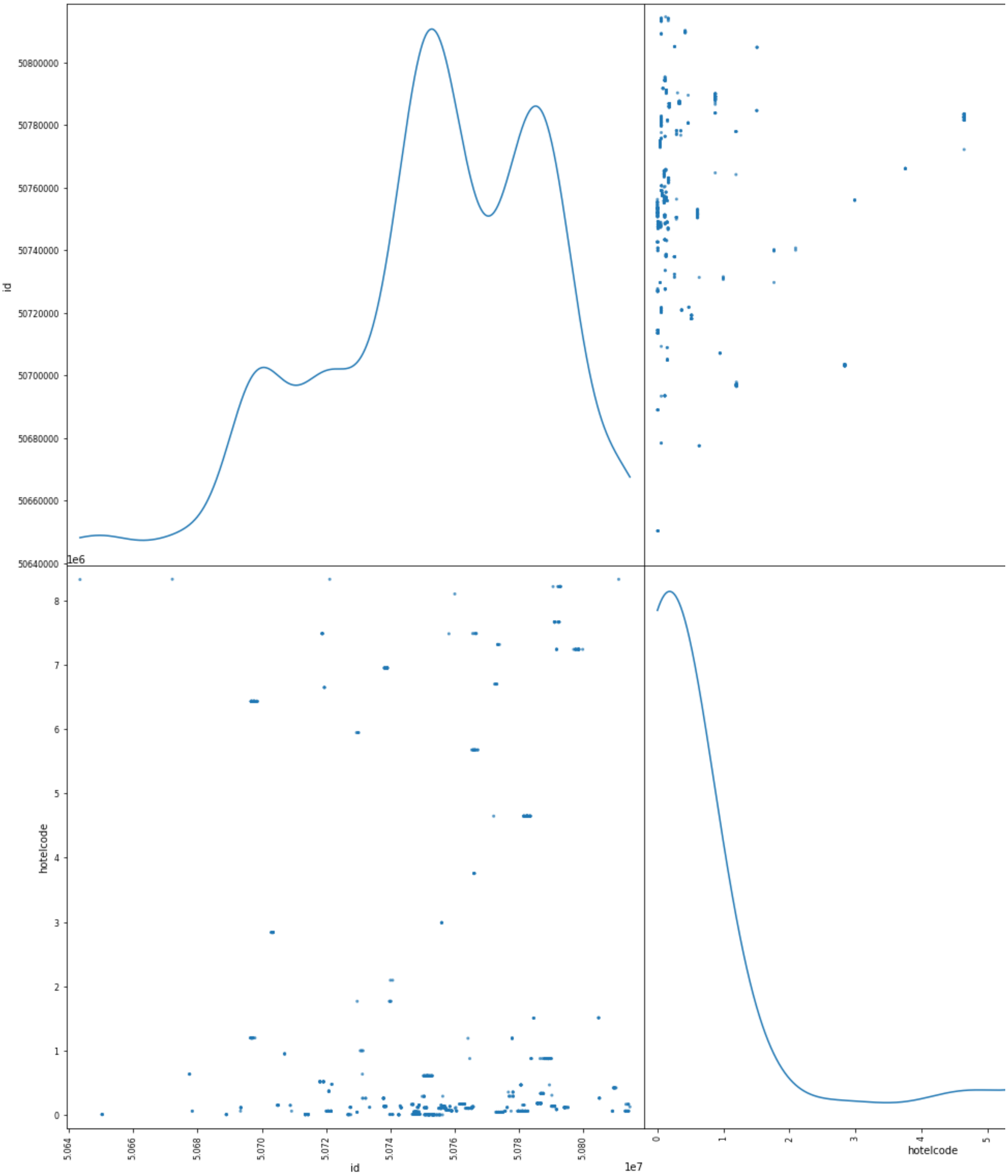
```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:3: FutureWarning: In a
  This is separate from the ipykernel package so we can avoid doing imports until
```



```
plotScatterMatrix(df3, 20, 10)
```

```
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:4: FutureWarning: In a
  after removing the cwd from sys.path.
```

Scatter and Density Plot



! 0s completed at 12:11 AM ● ×