**EX:No.1 221501051**

**25/01/25**

**PROGRAM TO IMPLEMENT TIME SERIES DATA FOR IMPORT LIBRARY, LOAD DATA, PREPROCESSING AND VISUALISING**

**PROGRAM AND OUTPUT:**

import pandas as pd

import numpy as np

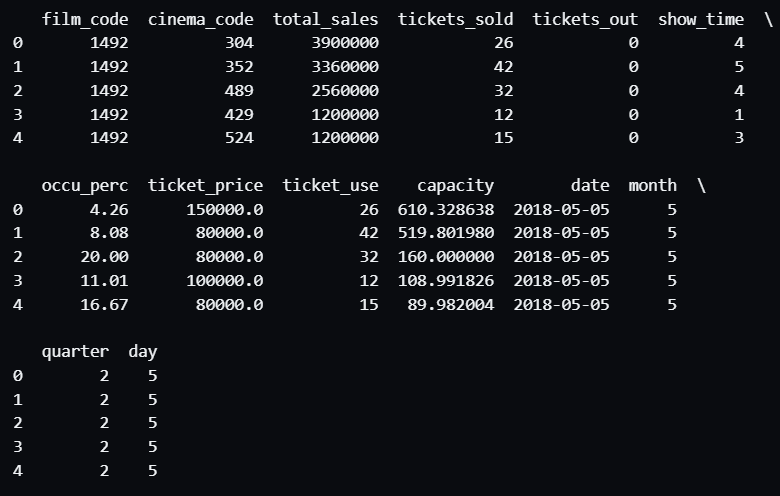
import matplotlib.pyplot as plt

import seaborn as sns

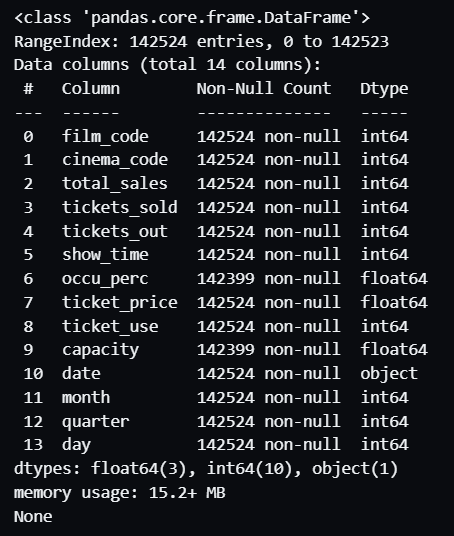
file\_path = 'cinemaTicket\_Ref.csv'

data = pd.read\_csv(file\_path)

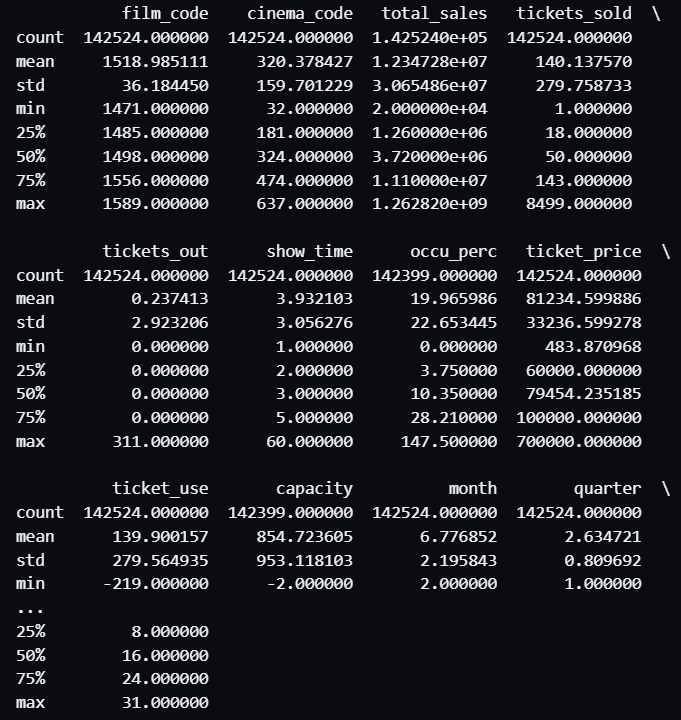
print(data.head())



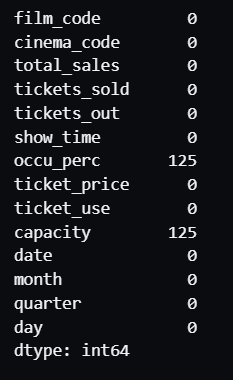
print(data.info())



print(data.describe())



print(data.isnull().sum())



for col in data.select\_dtypes(include=[np.number]).columns:

    plt.figure(figsize=(6, 4))

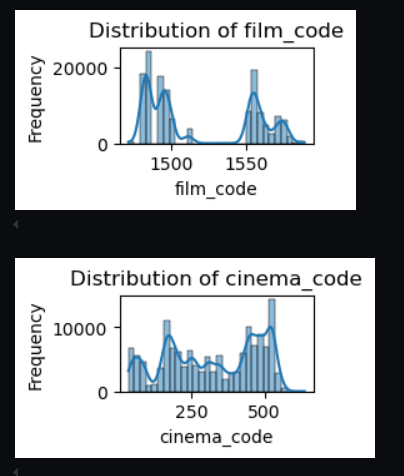
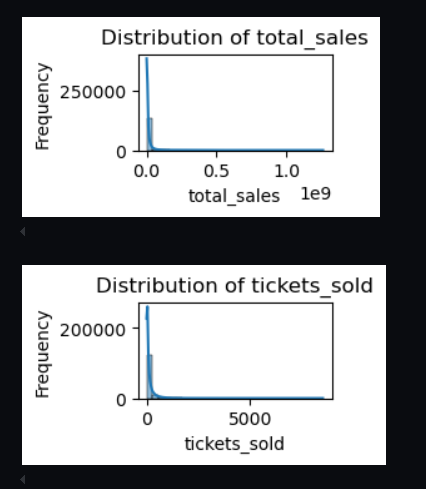
    sns.histplot(data[col], kde=True, bins=30)

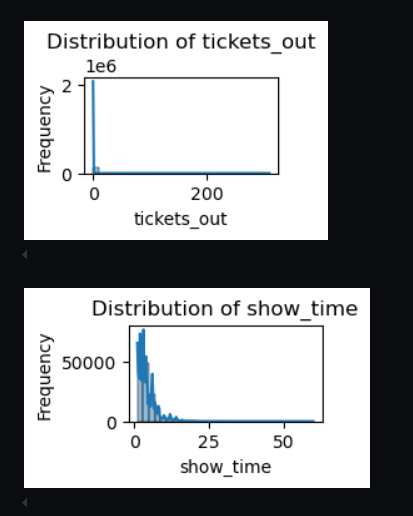
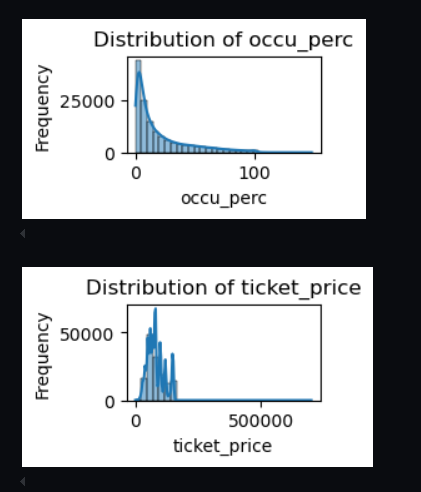
    plt.title(f"Distribution of {col}")

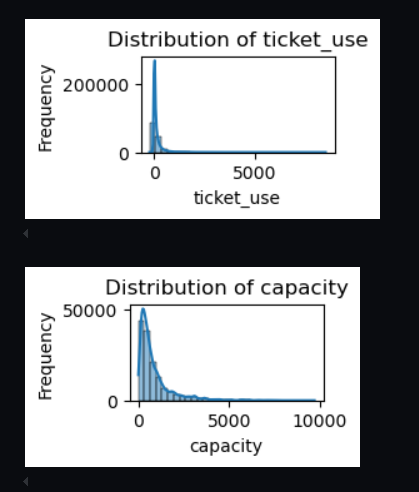
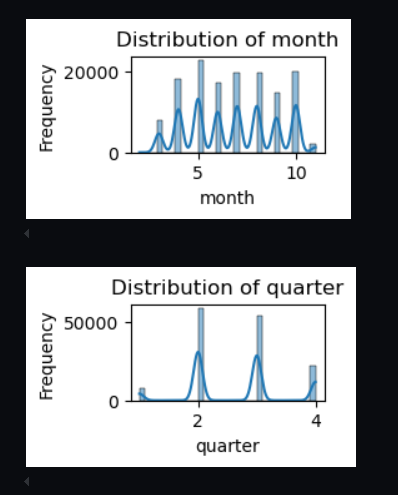
    plt.xlabel(col)

    plt.ylabel("Frequency")

    plt.show()

  
  
data.fillna(data.mean(numeric\_only=True), inplace=True)

for col in categorical\_columns:

    data[col].fillna(data[col].mode()[0], inplace=True)

cleaned\_file\_path = 'data/cleaned\_cinemaTicket\_Ref.csv'

data.to\_csv(cleaned\_file\_path, index=False)