```python
# Exploratory Data Analysis
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
# Import the datasets in jupyter notebook
data = pd.read_csv(r"C:\Users\ADMIN\Downloads\healthcare_dataset.csv",encoding="latin1")
df=pd.DataFrame(data)
# Data cleaning.....
# Check null values in the dataset
check_null=df.isnull().sum() # their is no null values in the dataset
# Calculate the correlation matrix for the Age and Billing Amount columns
correlation_matrix = df[['Age', 'Billing Amount']].corr()
# Create a heatmap to visualize the correlation
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt='.2f')
# Add a title
plt.title('Correlation between Age and Billing Amount')
# Show the heatmap
plt.show()
# Correlation between the admission type and billing amount
# Firstly we have to encode the admission type as emergency=1,urgent=2,elective=3
df["Admission_Type_Encode"]=df["Admission Type"].map({"Emergency":1,"Urgent":2,"Elective":3})
# Calculate the correlation matrix for the encoded Admission Type and Billing Amount
correlation_matrix=df[["Admission_Type_Encode","Billing Amount"]].corr()
# Create a heatmap to visualize the correlation
sns.heatmap(correlation_matrix,annot=True,fmt=".2f")
#Add a title
plt.title("Correlation Between Admission Type and Billing Amount")
#Show the heatmap
plt.show()
#Correlation Between Medical Condition and Billing Amount
#Firstly we have to encode the Medical Condition as Cancer=1,Hypertension=2,Asthma=3,Arthritis=4,Diabetes=5
df["Medical Condition Encode"]=df["Medical Condition"].map({"Cancer":1,"Hypertension":2,"Asthma":3,"Arthritis":4,"Diabetes":5})
# Calcualte the correlation matrix for the encoded medical condition and billing amount
correlation_matrix=df[["Medical Condition Encode","Billing Amount"]].corr()
# Create a heatmap to visualize the correlation
sns.heatmap(correlation_matrix,annot=True,fmt=".2f")
#Add a title
plt.title("Correlation Between Medical Condition and Billing Amount")
#Show the heatmap
plt.show()
#['Blue Cross' 'Medicare' 'Aetna' 'UnitedHealthcare' 'Cigna']
# Correlation between Insurance Provider and Billing Amount
# Firsty we have to endcode insurane provider companies Blur Cross=1,Medicare=2,Atena = 3,Unitedhealthcare=4,Cigna=5
df["Insurance Provider Encode"]=df["Insurance Provider"].map({"Blue Cross":10,"Medicare":20,"Aetna":30,"UnitedHealthcare":40,"Cigna":50})
#Calculate correaltion matrix for the encoded insurance provider and billing amount
correlation_matrix=df[["Insurance Provider Encode","Billing Amount"]].corr()
# Create a heatmap to visualize the correlation
sns.heatmap(correlation_matrix,annot=True,fmt=".2f",cmap="coolwarm")
#Add a title
plt.title("Correaltion Between insurance provider and billing amount")
#show the heatmap
plt.show()
# Most No of time hospital Name using mode
most_no_of_time_hospital=df["Hospital"].mode()
print(most_no_of_time_hospital) # LLC SMITH HOSPITAL
# Most No of time Doctor name
most_no_of_time_doctor_name=df["Doctor"].mode()
print(most_no_of_time_doctor_name) # Michael Smith
# Most_No_of_time_blood_group
most_no_of_time_blood_group=df["Blood Type"].mode()
print(most_no_of_time_blood_group) # A-
# Most insurance provider
most_no_of_insurance_provider=df["Insurance Provider"].mode()
print(most_no_of_insurance_provider) # Cigna
# Most Admission Type
most_admission_type=df["Admission Type"].mode()
print(most_admission_type) # Elective
# Most Medication
most_medication=df["Medication"].mode()
print(most_medication) # Lipitor
# most of medical_condition
most_no_of_medical_condition=df["Medical Condition"].mode()
print(most_no_of_medical_condition) # Arthritis
# Avg billing amount
avg_billing_amount=df["Billing Amount"].mean()
print(avg_billing_amount) # 25544.306284383965
# Avg age
avg_age_of_patients=df["Age"].mean()
print(avg_age_of_patients) # 51.53518538733035
#middle billing amount
middle_billing_amount=df["Billing Amount"].median()
print(middle_billing_amount) #25542.749144555833
#middle age
middile_age=df["Age"].median()
print(middile_age) # 52
```
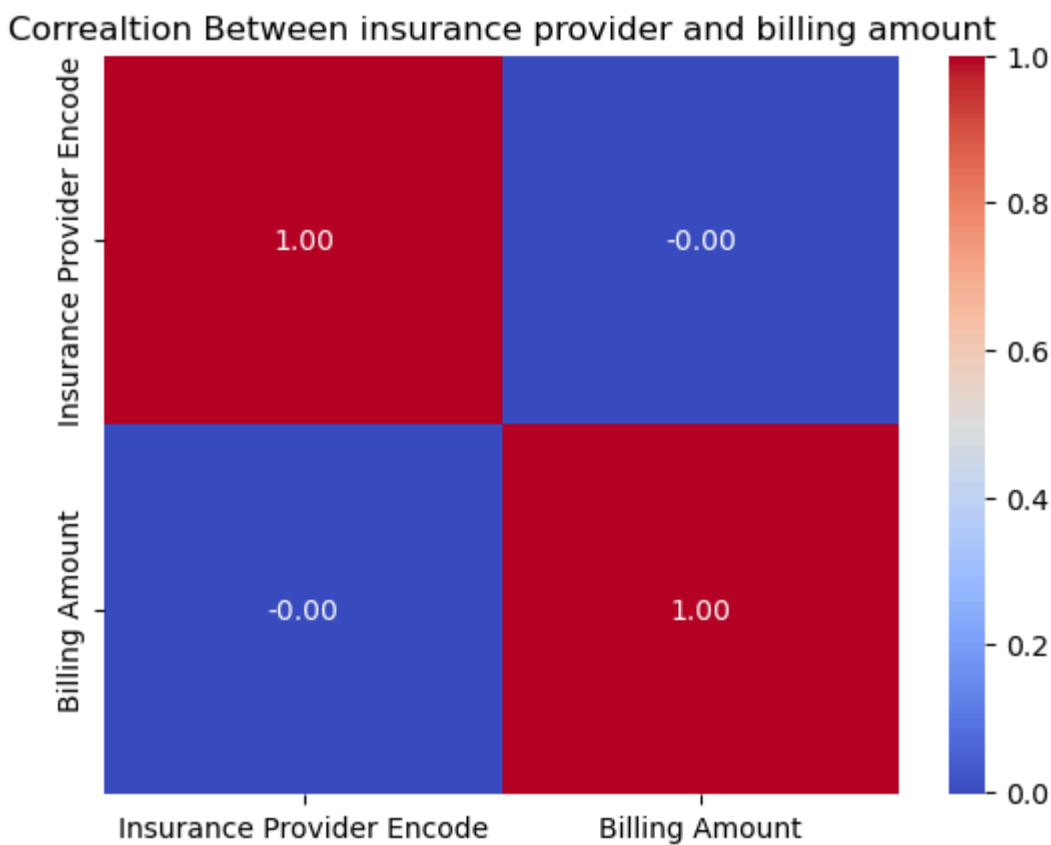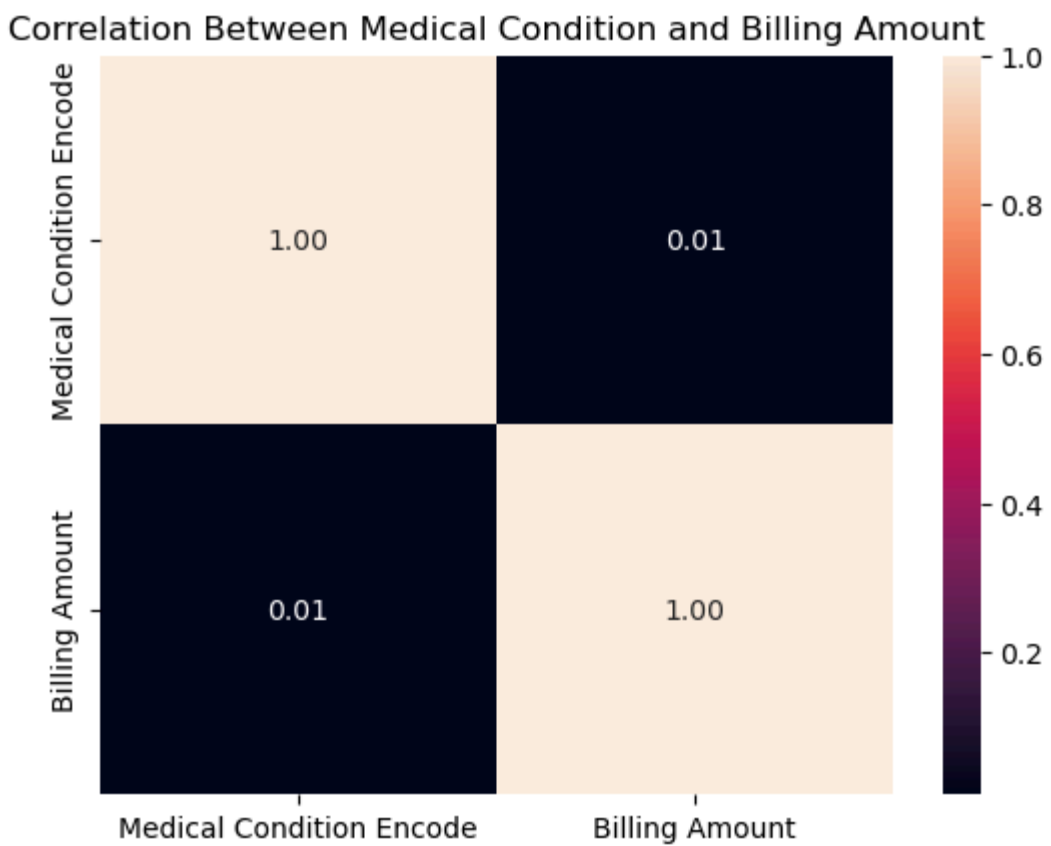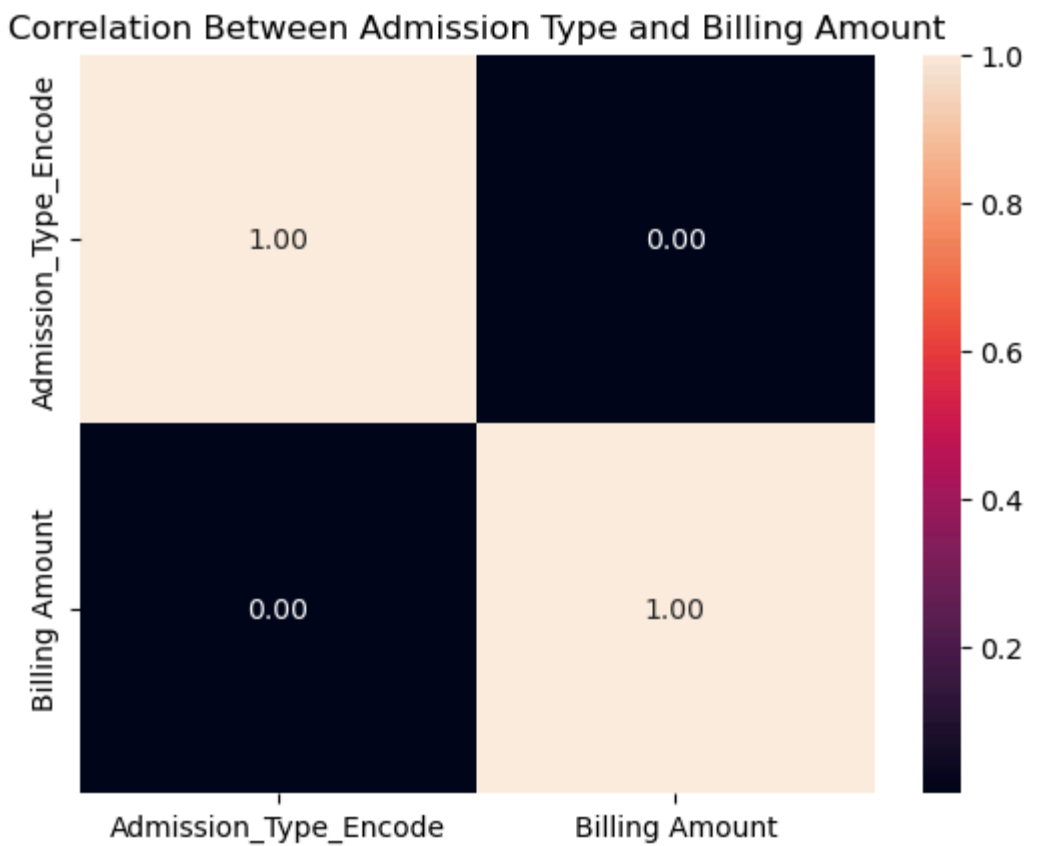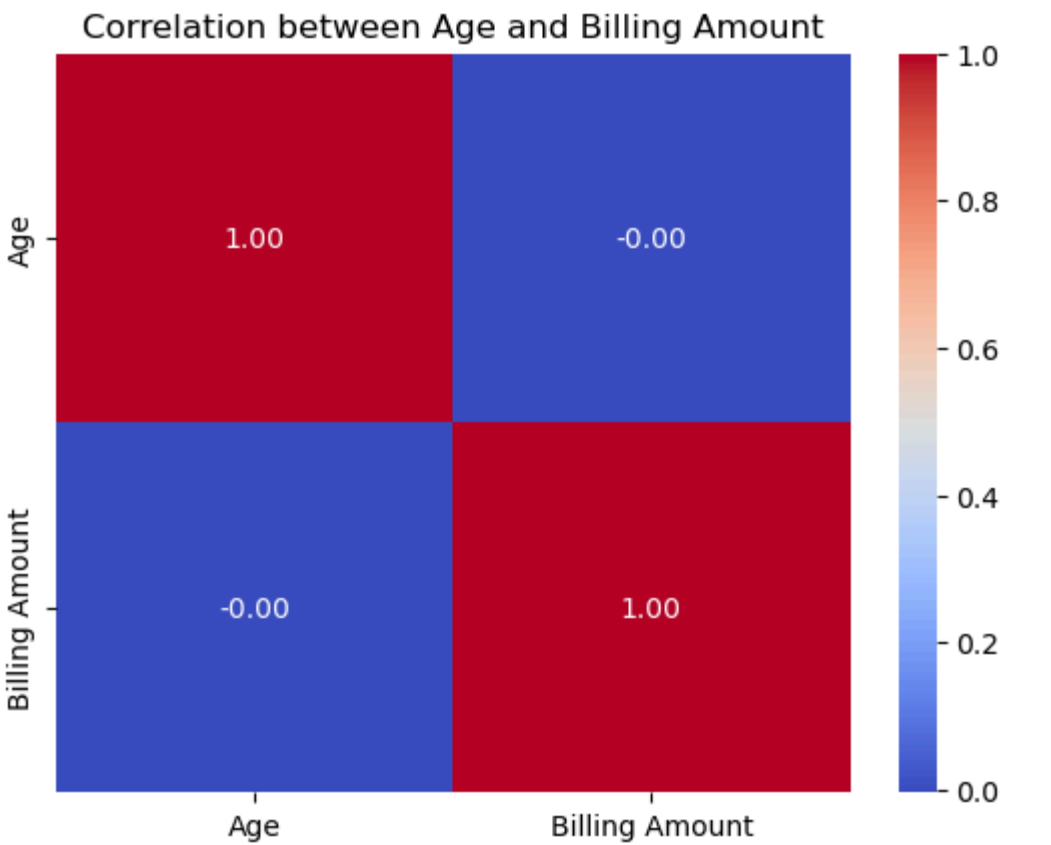

Correlation between Age and Billing Amount

Correlation Between Admission Type and Billing Amount

Correlation Between Medical Condition and Billing Amount

Correaltion Between insurance provider and billing amount

```
0    LLC Smith
Name: Hospital, dtype: object
0    Michael Smith
Name: Doctor, dtype: object
0    A-
Name: Blood Type, dtype: object
0    Cigna
Name: Insurance Provider, dtype: object
0    Elective
Name: Admission Type, dtype: object
0    Lipitor
Name: Medication, dtype: object
0    Arthritis
Name: Medical Condition, dtype: object
25539.316097211795
51.53945945945946
```

25538.069375965664
52.0

In [ ]: