In [95]:
```
import pandas as pd
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.metrics import mean_squared_error, r2_score
import numpy as np
from sklearn.ensemble import RandomForestRegressor
```

In [31]:
```
df=pd.read_csv('C:/Users/NITISH BOKKA/Downloads/archive (11)/uber.csv')
```

In [32]:
```
df.isnull().sum()
```

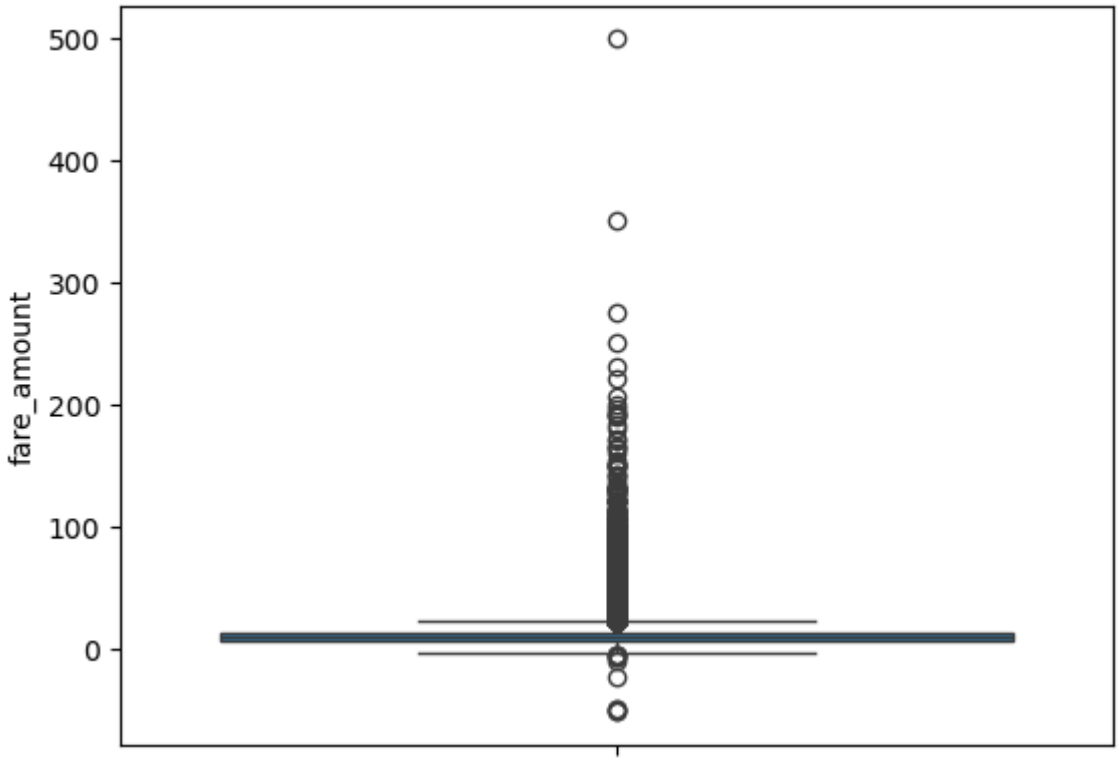
Out[32]:
```
Unnamed: 0           0
key                  0
fare_amount          0
pickup_datetime      0
pickup_longitude     0
pickup_latitude      0
dropoff_longitude    1
dropoff_latitude     1
passenger_count      0
dtype: int64
```

In [33]:
```
df.dropna(inplace=True)
```

In [34]:
```
df.isnull().sum()
```

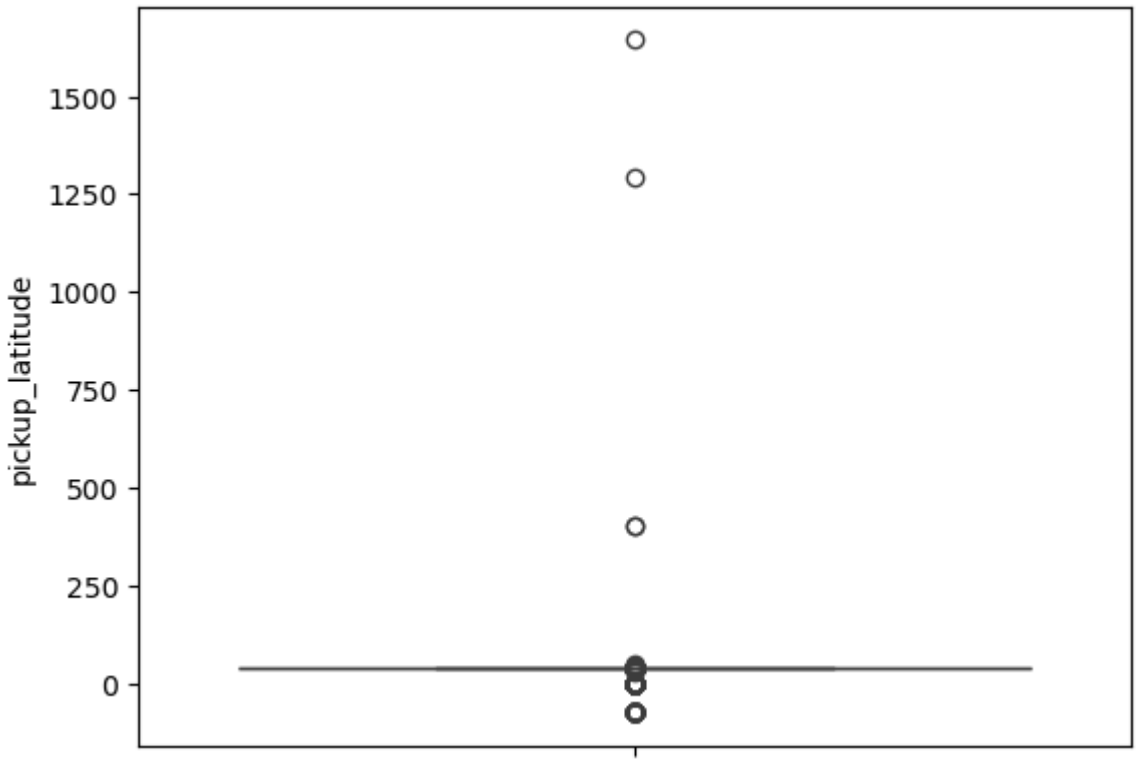
Out[34]:
```
Unnamed: 0           0
key                  0
fare_amount          0
pickup_datetime      0
pickup_longitude     0
pickup_latitude      0
dropoff_longitude    0
dropoff_latitude     0
passenger_count      0
dtype: int64
```

In [35]:
```
sns.boxplot(df['fare_amount'])
plt.show()
```



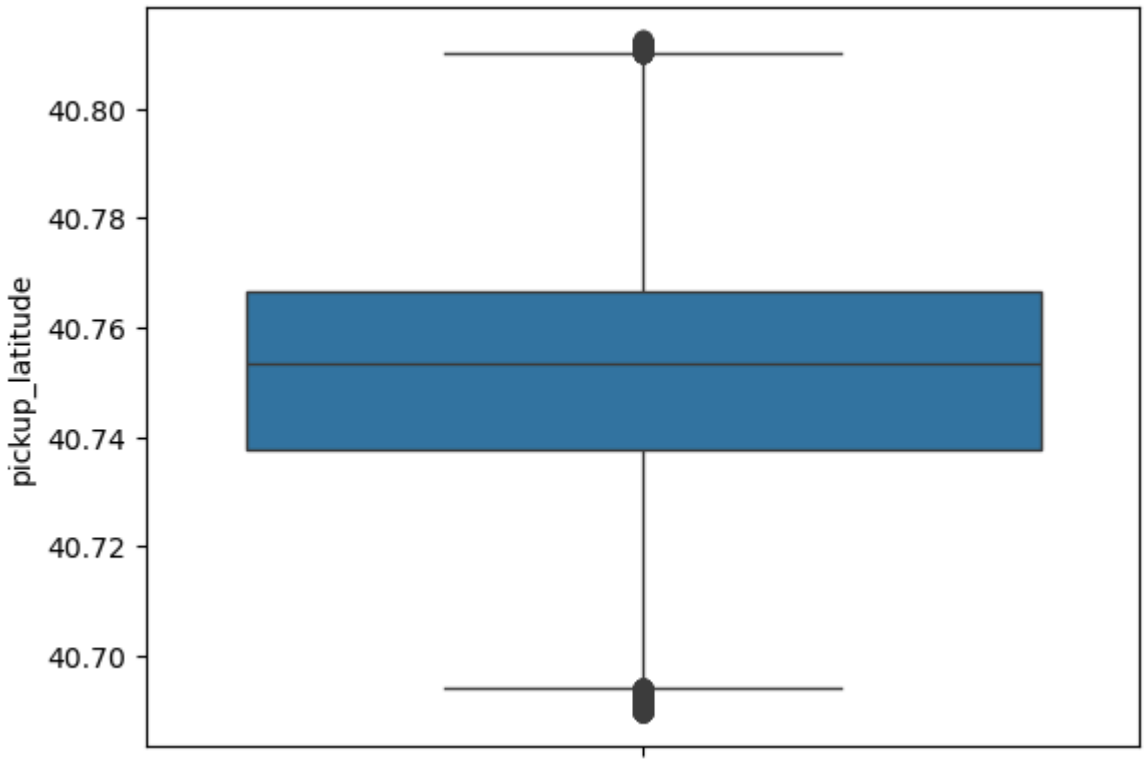
In [45]:
```
q1=df['fare_amount'].quantile(0.25)
q3=df['fare_amount'].quantile(0.75)
IQR=q3-q1
lower_bound=q1-1.5*IQR
upper_bound=q3+1.5*IQR
df=df[(df['fare_amount']>=lower_bound) & (df['fare_amount']<=upper_bound)]
```

In [49]:
```
sns.boxplot(df['pickup_latitude'])
plt.show()
```

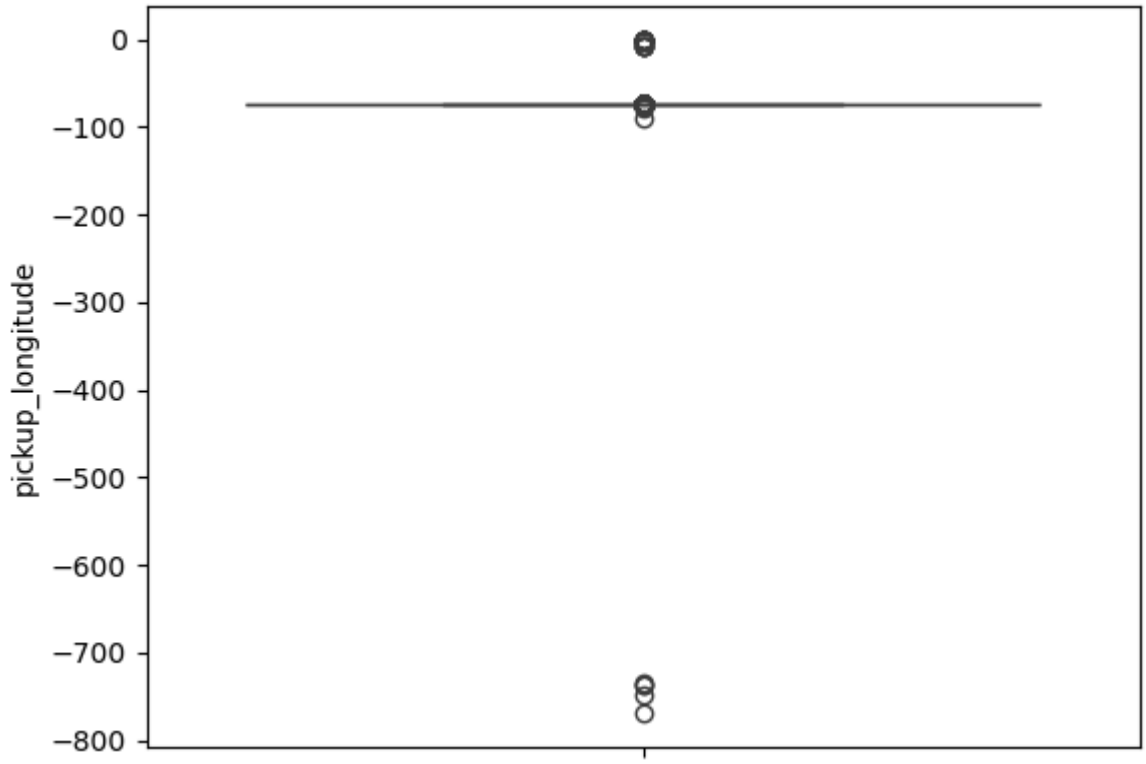


In [53]:
```
q1=df['pickup_latitude'].quantile(0.25)
q3=df['pickup_latitude'].quantile(0.75)
IQR=q3-q1
lower_bound=q1-1.5*IQR
upper_bound=q3+1.5*IQR
df=df[(df['pickup_latitude']>=lower_bound) & (df['pickup_latitude']<=upper_bound
```

In [55]:
```
sns.boxplot(df['pickup_latitude'])
plt.show()
```

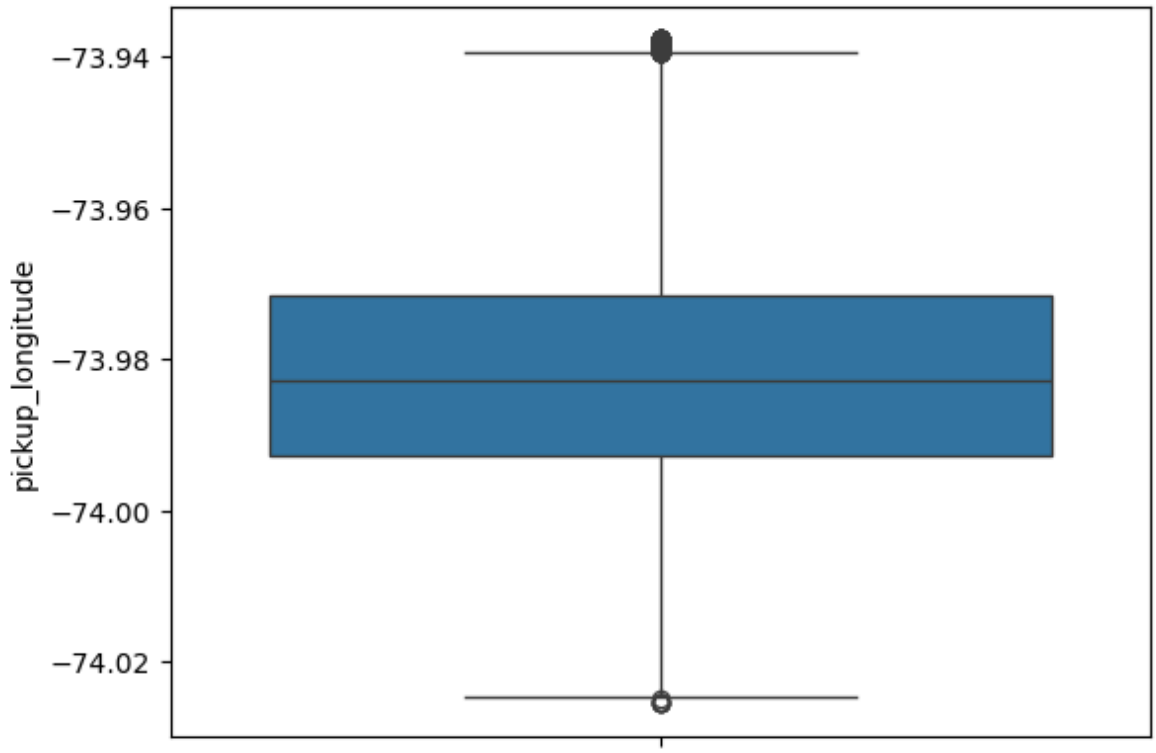


In [57]:
```
sns.boxplot(df['pickup_longitude'])
plt.show()
```



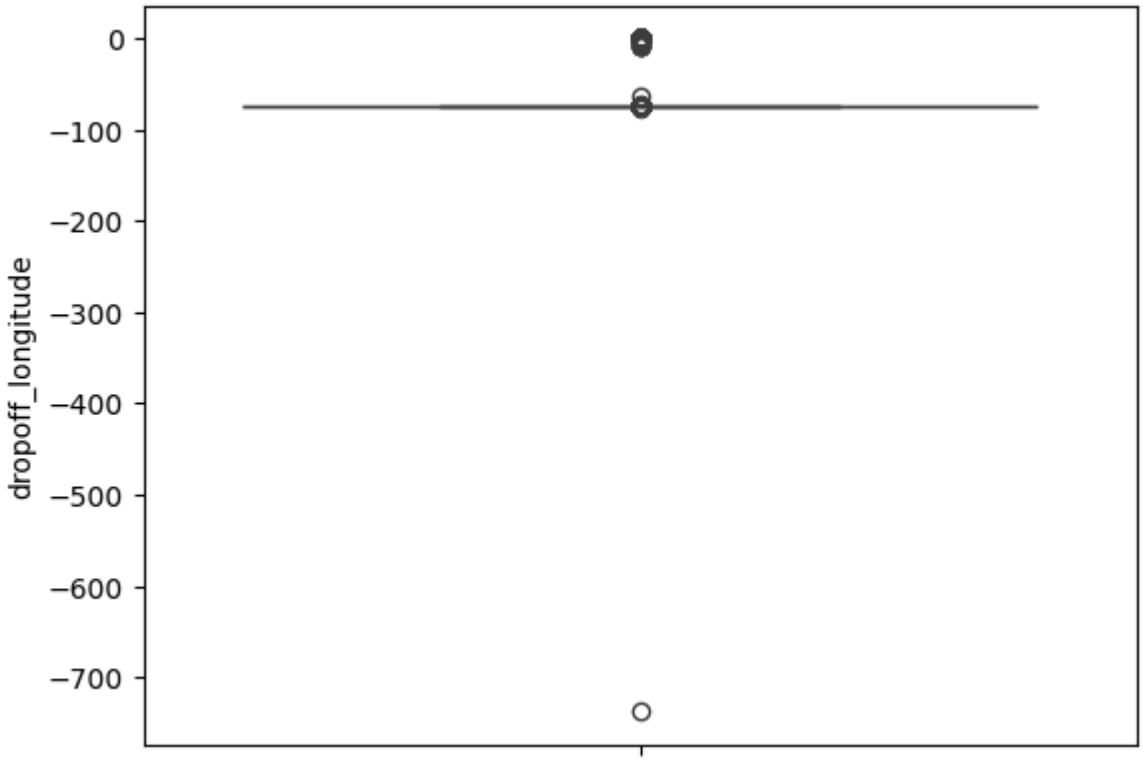
In [59]:
```
q1=df['pickup_longitude'].quantile(0.25)
q3=df['pickup_longitude'].quantile(0.75)
IQR=q3-q1
lower_bound=q1-1.5*IQR
upper_bound=q3+1.5*IQR
df=df[(df['pickup_longitude']>=lower_bound) & (df['pickup_longitude']<=upper_bou
```

In [61]:
```
sns.boxplot(df['pickup_longitude'])
plt.show()
```



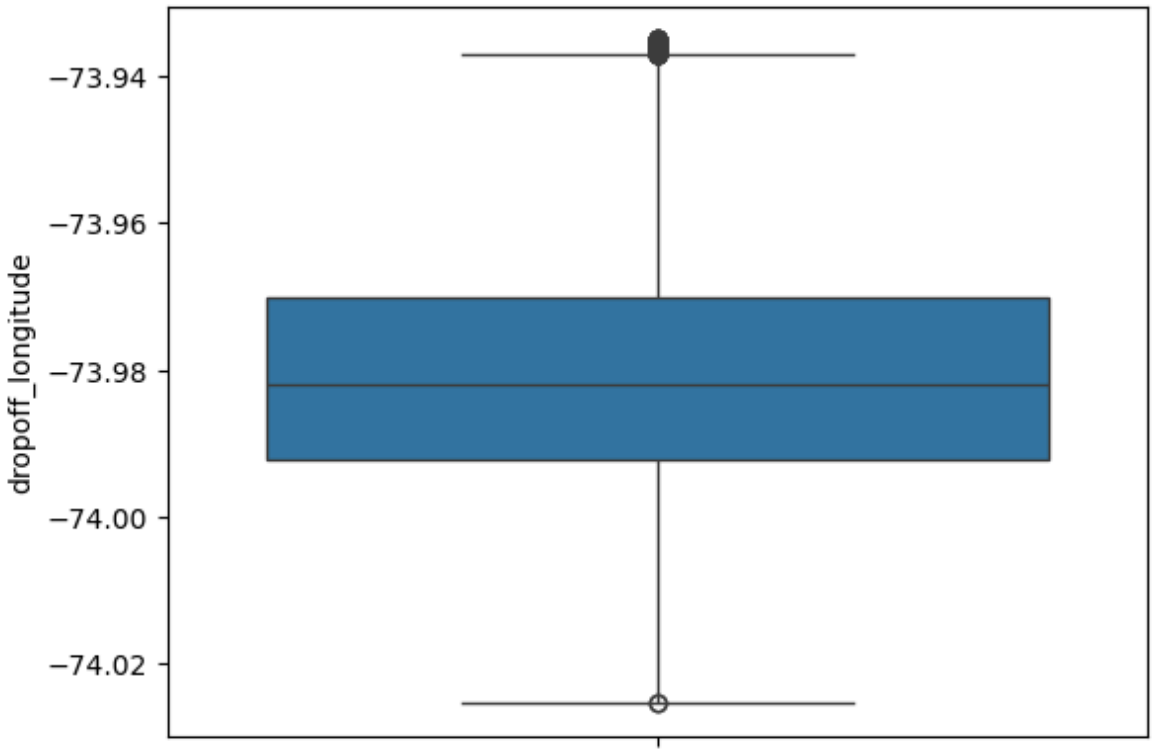
In [63]:
```
sns.boxplot(df['dropoff_longitude'])
```

```
plt.show()
```

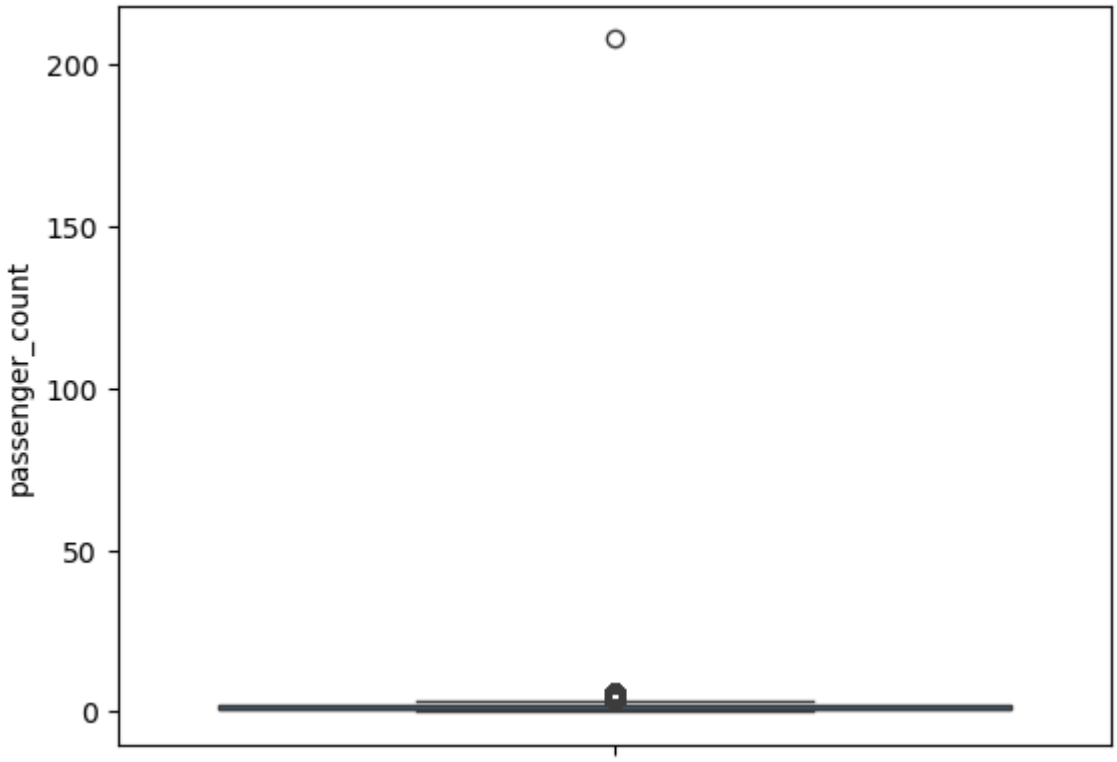


In [65]:
```
q1=df['dropoff_longitude'].quantile(0.25)
q3=df['dropoff_longitude'].quantile(0.75)
IQR=q3-q1
lower_bound=q1-1.5*IQR
upper_bound=q3+1.5*IQR
df=df[(df['dropoff_longitude']>=lower_bound) & (df['dropoff_longitude']<=upper_b
```

In [67]:
```
sns.boxplot(df['dropoff_longitude'])
plt.show()
```



In [69]:
```
sns.boxplot(df['passenger_count'])
plt.show()
```



In [75]:
```
q1=df['passenger_count'].quantile(0.25)
q3=df['passenger_count'].quantile(0.75)
IQR=q3-q1
lower_bound=q1-1.5*IQR
upper_bound=q3+1.5*IQR
df=df[(df['passenger_count']>=lower_bound) & (df['passenger_count']<=upper_bound
```

In [77]:
```
sns.boxplot(df['passenger_count'])
plt.show()
```

In [79]:
```
X=df[['pickup_latitude','pickup_longitude','dropoff_latitude','dropoff_longitude
y=df['fare_amount']
```

In [81]:
```
X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.2,random_state=42
```

In [87]:
```
linear=LinearRegression()
linear.fit(X_train,y_train)
y_pred=linear.predict(X_test)
```

In [93]:
```
mse=mean_squared_error(y_test,y_pred)
print(mse)
print(np.sqrt(mse))
print(r2_score(y_test,y_pred))
```

```
15.939420085833529
3.992420329303207
0.011738606407233454
```

In [98]:
```
random=RandomForestRegressor()
random.fit(X_train,y_train)
y_pred=random.predict(X_test)
```

In [101…
```
print(mean_squared_error(y_test,y_pred))
print(np.sqrt((mean_squared_error(y_test,y_pred))))
print((r2_score(y_test,y_pred)))
```

```
4.5948745486328955
2.143565848914583
0.7151127769791495
```

In [ ]:
```

```