Question 3: The output shown below is from python code **Assignment1_Question3.py**

PART A.

➔ The empirical fraud rate is approximately **0.1042**, which means approximately **10.42%** of investigations are found to be frauds.

```
Empirical Fraud Rate: 0.1995
```

PART B.

➔ In the Training Partition, there are **4768 observations**.
➔ In the Testing Partition, there are **1192 observations**.

```
Number of Observations in Training Partition: 4768
Number of Observations in Testing Partition: 1192
```

PART C.

➔ Misclassification Rates for Different Numbers of Neighbors in the Testing Partition:

```
Neighbors: 2, Train Misclassification Rate: 0.1424, Test Misclassification Rate: 0.1636
Neighbors: 3, Train Misclassification Rate: 0.1395, Test Misclassification Rate: 0.1980
Neighbors: 4, Train Misclassification Rate: 0.1674, Test Misclassification Rate: 0.1678
Neighbors: 5, Train Misclassification Rate: 0.1665, Test Misclassification Rate: 0.1787
Neighbors: 6, Train Misclassification Rate: 0.1749, Test Misclassification Rate: 0.1628
Neighbors: 7, Train Misclassification Rate: 0.1743, Test Misclassification Rate: 0.1703
```

PART D.

➔ The number of neighbors that yields the lowest misclassification rate in the Testing Partition is **6.**

PART E.

➔ Neighbor Observation Values:

	CASE_ID	FRAUD	TOTAL_SPEND	DOCTOR_VISITS	NUM_CLAIMS	MEMBER_DURATION	OPTOM_PRESC	NUM_MEMBERS
2973	2974	0	16300	11	0	180	1	1
2967	2968	0	16300	2	0	193	0	2
2980	2981	0	16300	1	0	162	3	1
2972	2973	1	16300	2	0	200	0	2
2962	2963	1	16300	12	5	125	1	1
2971	2972	0	16300	5	0	246	3	2

Predicted Probability of Fraud: 0.3333333333333333