PROYECT 4

Credit Card Fraud Detection 2019 & 2020

Our scope was

To find the accuracy in scam detection within the usage of credit cards across USA (2019-2020).

Tools applied:

pandas

matplotlib

tableau

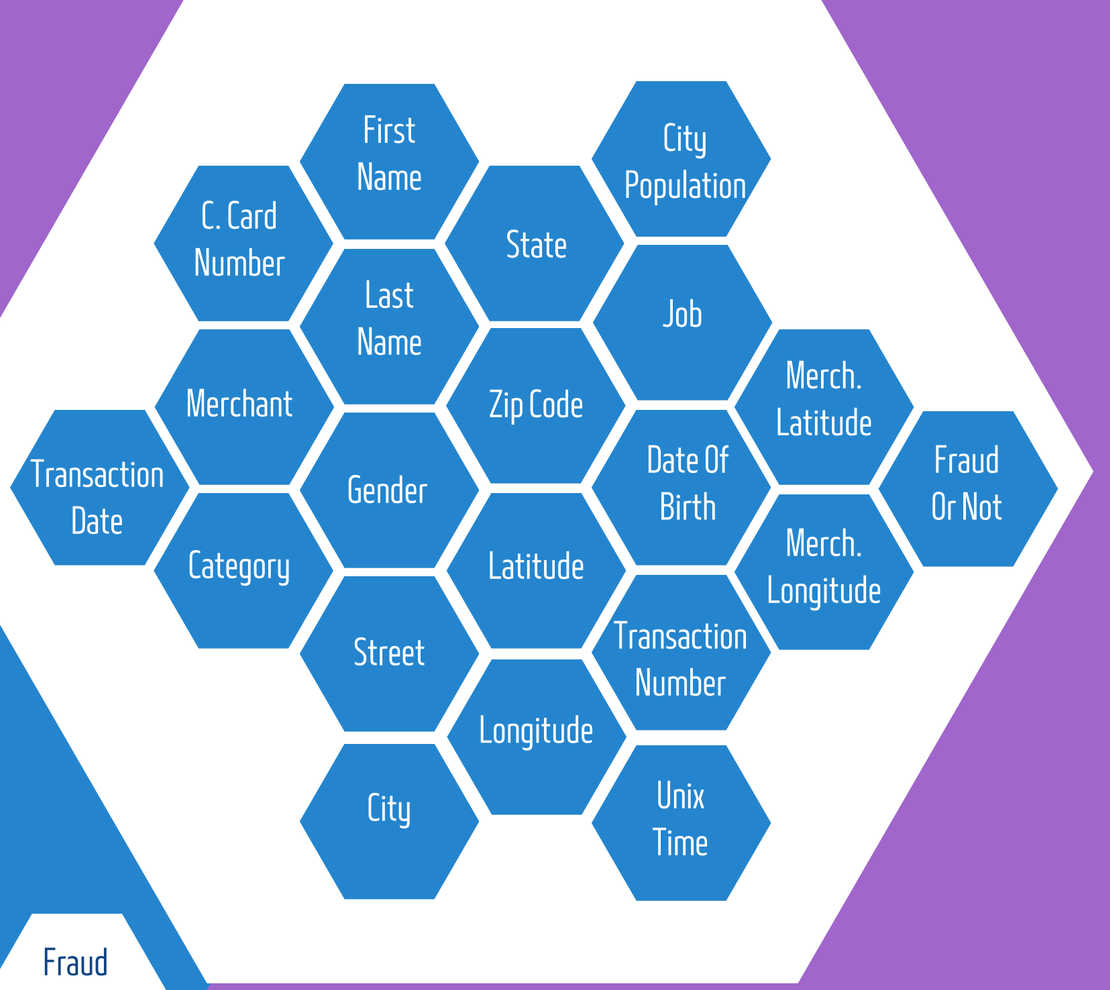
kaggle

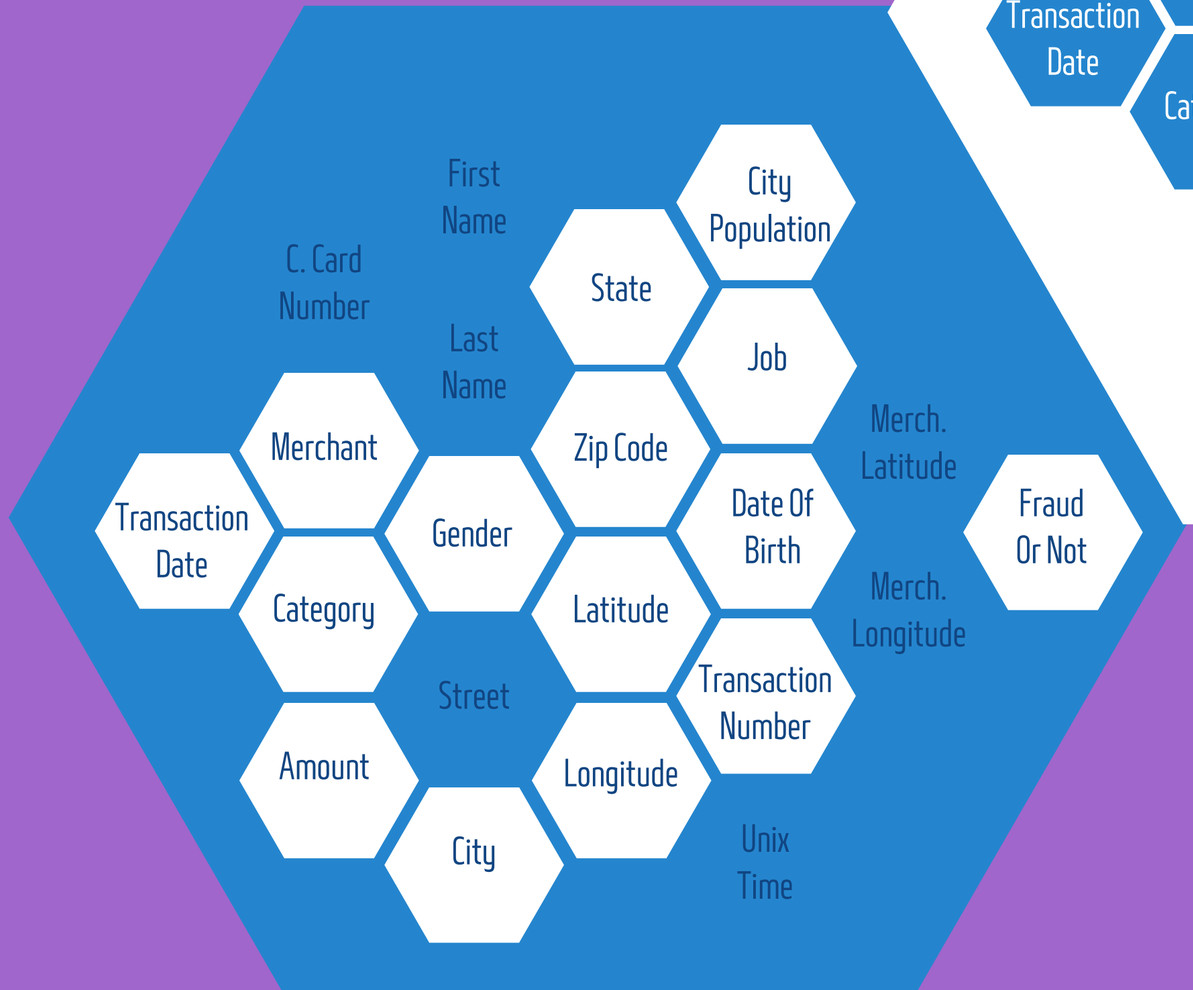
Our dataset was taken from Kaggle for learning purposes.

It is compound for 14 columns and 1,604,296 rows.

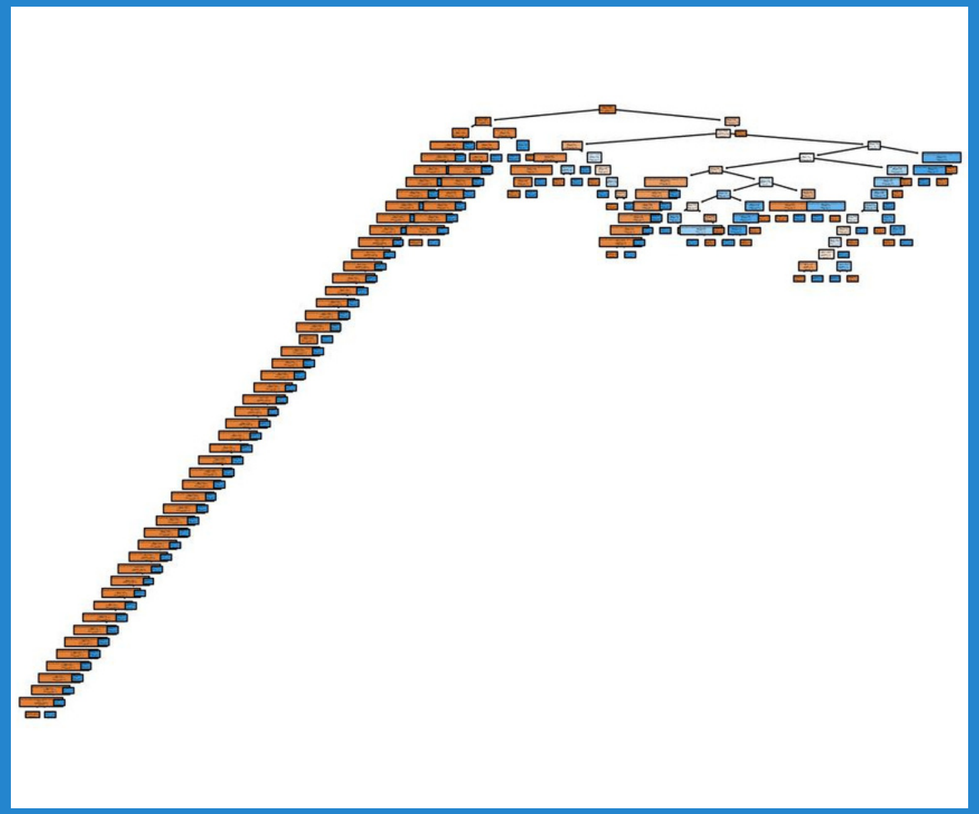
Our data set was cleaned, normalized, and standardized with SQL, prior to modeling.

Data Selection

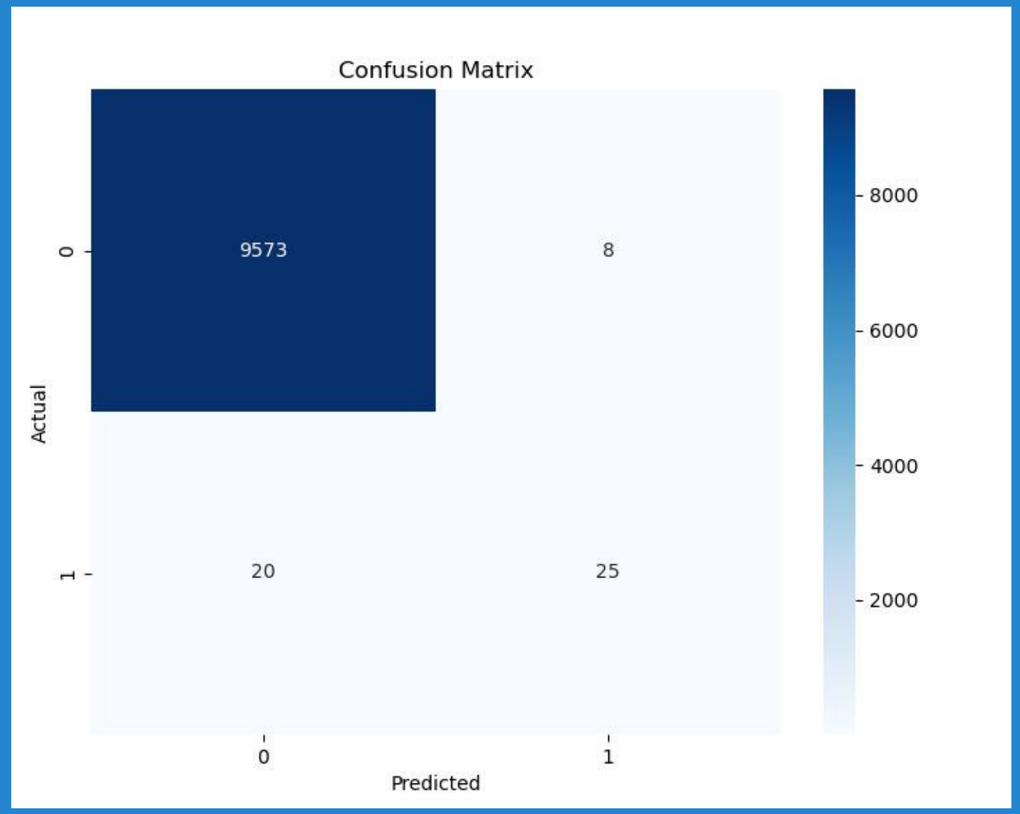




Decision Tree Model

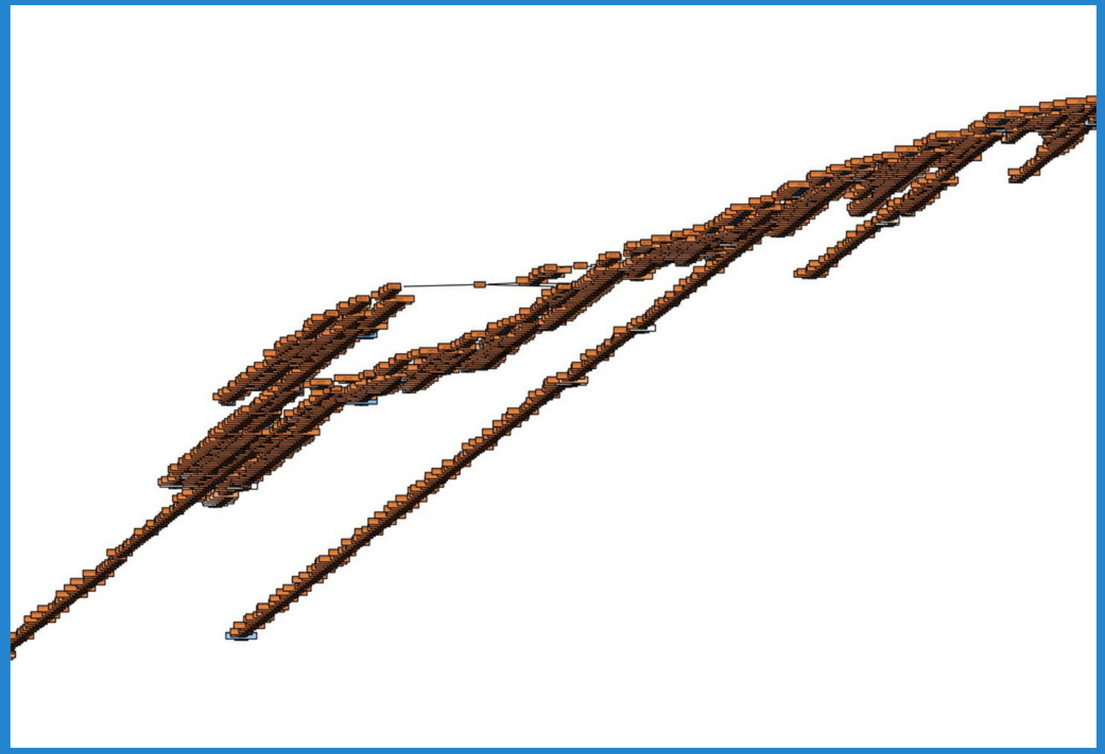


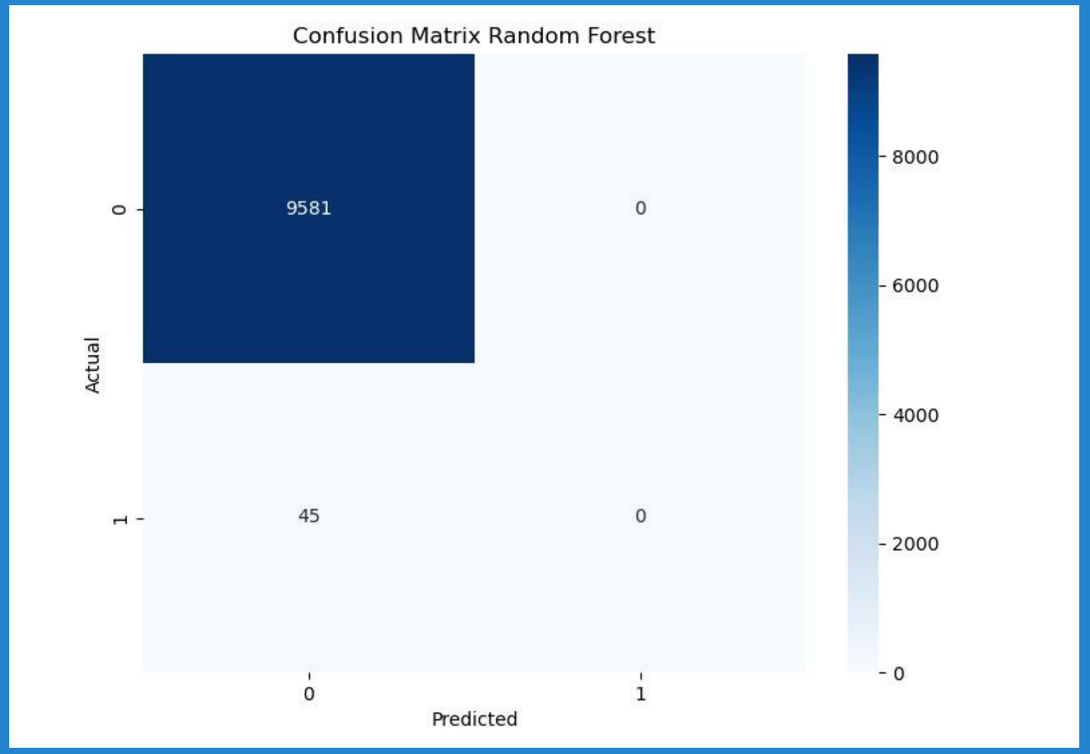
Visualization



Matrix

Random Forest Model





Visualization **Matrix**

Neural Network

```python
# Define the model - deep neural net, i.e., the number of input features and hidden nodes for each layer.
input_features = len(X_train[0])
hidden_nodes_layer_1 = 80
hidden_nodes_layer_2 = 30

nn = tf.keras.models.Sequential()

# First hidden layer
nn.add(tf.keras.layers.Dense(units=hidden_nodes_layer_1, input_dim = input_features, activation = "relu"))

# Second hidden layer
nn.add(tf.keras.layers.Dense(units=hidden_nodes_layer_2, activation = "relu"))


# Output layer
nn.add(tf.keras.layers.Dense(units=1, activation="sigmoid"))

# Check the structure of the model
nn.summary()
```

```
Model: "sequential_6"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense_18 (Dense)            (None, 80)                1516320

 dense_19 (Dense)            (None, 30)                2430

 dense_20 (Dense)            (None, 1)                 31

=================================================================
Total params: 1518781 (5.79 MB)
Trainable params: 1518781 (5.79 MB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
```

```python
# Evaluate the model using the test data
model_loss, model_accuracy = nn.evaluate(X_test_scaled,y_test,verbose=2)
print(f"Loss: {model_loss}, Accuracy: {model_accuracy}")
```

```
63/63 - 0s - loss: 0.1516 - accuracy: 0.9940 - 55ms/epoch - 873us/step
Loss: 0.15157388150691986, Accuracy: 0.9940179586410522
```

Does the records of frauds increased over time?

Does probability of fraud change through the days of the week?

Which are the hours with a higher records of frauds?

Which category is more likely to have more incidents?

Which are the top 10 states in USA to have frauds?

DASHBOARD

THANK FOR YOUR ATTENTION.