CREDIT CARD FRAUD DETECTION

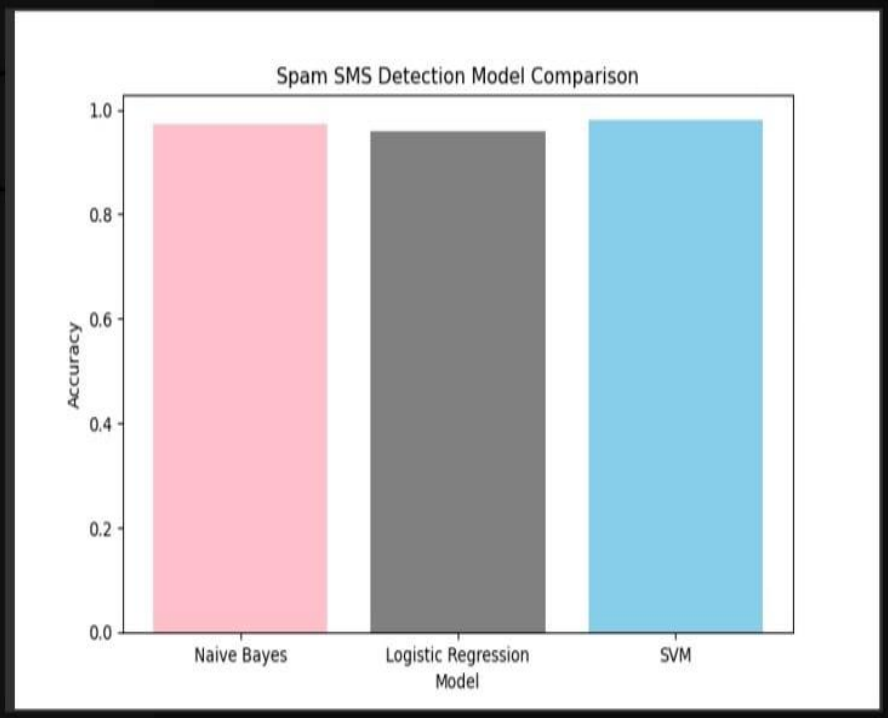
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1296675 entries, 0 to 1296674
Data columns (total 23 columns):
 #   Column                 Non-Null Count    Dtype
---  ------                 --------------    -----
 0   Unnamed: 0             1296675 non-null  int64
 1   trans_date_trans_time  1296675 non-null  object
 2   cc_num                 1296675 non-null  int64
 3   merchant               1296675 non-null  object
 4   category               1296675 non-null  object
 5   amt                    1296675 non-null  float64
 6   first                  1296675 non-null  object
 7   last                   1296675 non-null  object
 8   gender                 1296675 non-null  object
 9   street                 1296675 non-null  object
 10  city                   1296675 non-null  object
 11  state                  1296675 non-null  object
 12  zip                    1296675 non-null  int64
 13  lat                    1296675 non-null  float64
 14  long                   1296675 non-null  float64
 15  city_pop               1296675 non-null  int64
 16  job                    1296675 non-null  object
 17  dob                    1296675 non-null  object
 18  trans_num              1296675 non-null  object
 19  unix_time              1296675 non-null  int64
 20  merch_lat              1296675 non-null  float64
 21  merch_long             1296675 non-null  float64
 22  is_fraud               1296675 non-null  int64
dtypes: float64(5), int64(6), object(12)
memory usage: 227.5+ MB
None
```

```
None
   Unnamed: 0 trans_date_trans_time            cc_num  \
0           0   2019-01-01 00:00:18  2703186189652095
1           1   2019-01-01 00:00:44      630423337322
2           2   2019-01-01 00:00:51    38859492057661
3           3   2019-01-01 00:01:16  3534093764340240
4           4   2019-01-01 00:03:06   375534208663984

                            merchant       category     amt      first  \
0          fraud_Rippin, Kub and Mann       misc_net    4.97   Jennifer
1     fraud_Heller, Gutmann and Zieme    grocery_pos  107.23  Stephanie
2                  fraud_Lind-Buckridge  entertainment  220.11     Edward
3  fraud_Kutch, Hermiston and Farrell  gas_transport   45.00     Jeremy
4                  fraud_Keeling-Crist       misc_pos   41.96      Tyler

     last gender                        street  ...      lat      long  \
0   Banks      F                561 Perry Cove  ...  36.0788  -81.1781
1    Gill      F  43039 Riley Greens Suite 393  ...  48.8878 -118.2105
2  Sanchez      M      594 White Dale Suite 530  ...  42.1808 -112.2620
3   White      M  9443 Cynthia Court Apt. 038  ...  46.2306 -112.1138
4  Garcia      M              408 Bradley Rest  ...  38.4207  -79.4629

   city_pop                                job         dob  \
0      3495          Psychologist, counselling  1988-03-09
1       149  Special educational needs teacher  1978-06-21
2      4154        Nature conservation officer  1962-01-19
3      1939                    Patent attorney  1967-01-12
4        99     Dance movement psychotherapist  1986-03-28

                          trans_num  unix_time  merch_lat  merch_long  \
0  0b242abb623afc578575680df30655b9  1325376018  36.011293  -82.048315
1  1f76529f8574734946361c461b024d99  1325376044  49.159047 -118.186462
2  a1a22d70485983eac12b5b88dad1cf95  1325376051  43.150704 -112.154481
3  6b849c168bdad6f867558c3793159a81  1325376076  47.034331 -112.561071
4  a41d7549acf90789359a9aa5346dcb46  1325376186  38.674999  -78.632459

   is_fraud
0         0
1         0
2         0
3         0
4         0
```

```
[5 rows x 23 columns]
Training Logistic Regression...
Logistic Regression Accuracy: 0.9955211176871764
[[553230    344]
 [  2145      0]]
              precision    recall  f1-score   support

           0       1.00      1.00      1.00    553574
           1       0.00      0.00      0.00      2145

    accuracy                           1.00    555719
   macro avg       0.50      0.50      0.50    555719
weighted avg       0.99      1.00      0.99    555719

---
Training Random Forest...
Random Forest Accuracy: 0.9972115448650116
[[553399    175]
 [  1428    717]]
              precision    recall  f1-score   support

           0       1.00      1.00      1.00    553574
           1       0.80      0.33      0.47      2145

    accuracy                           1.00    555719
   macro avg       0.90      0.67      0.74    555719
weighted avg       1.00      1.00      1.00    555719

---
```

SPAM SMS DETECTION

```
SPAM SMS DETECTION

Index(['v1', 'v2', 'Unnamed: 2', 'Unnamed: 3', 'Unnamed: 4'], dtype='object')

Naive Bayes Accuracy: 0.9722

Naive Bayes Classification Report:
              precision    recall  f1-score   support

           0       0.97      1.00      0.98       965
           1       1.00      0.79      0.88       150

    accuracy                           0.97      1115
   macro avg       0.98      0.90      0.93      1115
weighted avg       0.97      0.97      0.97      1115


----------

Logistic Regression Accuracy: 0.9578

Logistic Regression Classification Report:
              precision    recall  f1-score   support

           0       0.96      1.00      0.98       965
           1       0.97      0.71      0.82       150

    accuracy                           0.96      1115
   macro avg       0.96      0.85      0.90      1115
weighted avg       0.96      0.96      0.95      1115


----------

SVM Accuracy: 0.9794

SVM Classification Report:
              precision    recall  f1-score   support

           0       0.98      1.00      0.99       965
           1       0.97      0.87      0.92       150

    accuracy                           0.98      1115
   macro avg       0.98      0.93      0.95      1115
weighted avg       0.98      0.98      0.98      1115
```



CUSTOMER CHURN PREDICTION

```
CUSTOMER CHURN PREDICTION

dataset information:

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 14 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   RowNumber       10000 non-null  int64
 1   CustomerId      10000 non-null  int64
 2   Surname         10000 non-null  object
 3   CreditScore     10000 non-null  int64
 4   Geography       10000 non-null  object
 5   Gender          10000 non-null  object
 6   Age             10000 non-null  int64
 7   Tenure          10000 non-null  int64
 8   Balance         10000 non-null  float64
 9   NumOfProducts   10000 non-null  int64
 10  HasCrCard       10000 non-null  int64
 11  IsActiveMember  10000 non-null  int64
 12  EstimatedSalary 10000 non-null  float64
 13  Exited          10000 non-null  int64
dtypes: float64(2), int64(9), object(3)
memory usage: 1.1+ MB
None
   RowNumber  CustomerId   Surname  CreditScore Geography  Gender  Age  \
0          1    15634602  Hargrave          619    France  Female   42
1          2    15647311      Hill          608     Spain  Female   41
2          3    15619304      Onio          502    France  Female   42
3          4    15701354      Boni          699    France  Female   39
4          5    15737888  Mitchell          850     Spain  Female   43

   Tenure    Balance  NumOfProducts  HasCrCard  IsActiveMember  \
0       2       0.00              1          1               1
1       1   83807.86              1          0               1
2       8  159660.80              3          1               0
3       1       0.00              2          0               0
4       2  125510.82              1          1               1

   EstimatedSalary  Exited
0        101348.88       1
1        112542.58       0
2        113931.57       1
3         93826.63       0
4         79084.10       0
(10000, 14)

Logistic Regression Accuracy: 0.8155
Random Forest Accuracy: 0.864
Gradient Boosting Accuracy: 0.866
```

```
Logistic Regression Accuracy: 0.8155
Random Forest Accuracy: 0.864
Gradient Boosting Accuracy: 0.866

CUSTOMER CHURN PREDICTION
```