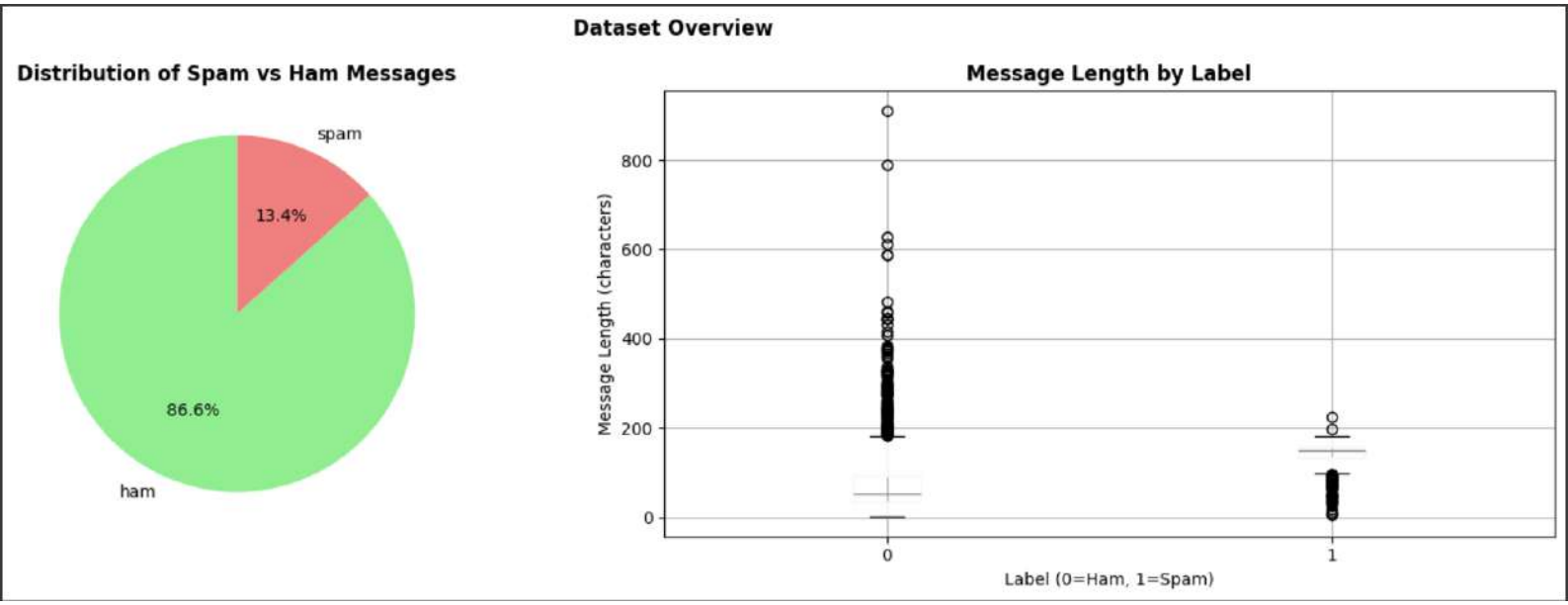
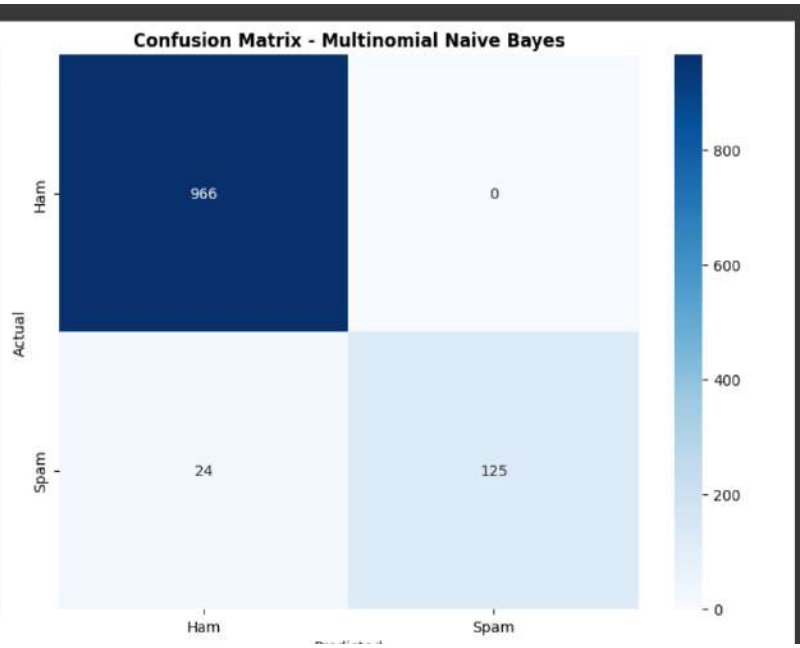
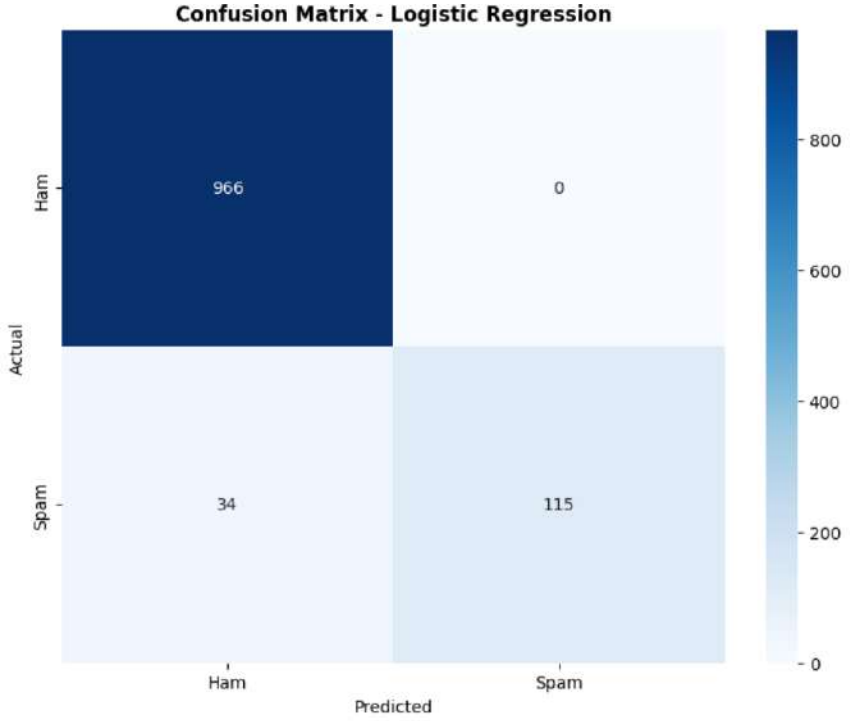
```
Dataset shape: (5572, 5)

Dataset columns: ['v1', 'v2', 'Unnamed: 2', 'Unnamed: 3', 'Unnamed: 4']

First 5 rows of the dataset:
     v1                                                 v2  Unnamed: 2  \
0   ham  Go until jurong point, crazy.. Available only ...         NaN
1   ham                      Ok lar... Joking wif u oni...         NaN
2  spam  Free entry in 2 a wkly comp to win FA Cup fina...         NaN
3   ham  U dun say so early hor... U c already then say...         NaN
4   ham  Nah I don't think he goes to usf, he lives aro...         NaN

   Unnamed: 3  Unnamed: 4
0         NaN         NaN
1         NaN         NaN
2         NaN         NaN
3         NaN         NaN
4         NaN         NaN
```


Dataset Overview
Distribution of Spam vs Ham Messages
spam
13.4%
86.6%
ham
Message Length by Label
Message Length (characters)
0
200
400
600
800
0
1
Label (0=Ham, 1=Spam)

```
🔡 Performing feature extraction with TF-IDF...
✅ Feature extraction completed! Shape: (5572, 3000)

🤖 Training machine learning models...
Training set: 4457 samples
Testing set: 1115 samples

🔧 Training Multinomial Naive Bayes...
✅ Multinomial Naive Bayes -> Accuracy: 0.9785, Precision: 1.0000
Confusion Matrix:
 [[966   0]
 [ 24 125]]

🔧 Training Logistic Regression...
✅ Logistic Regression -> Accuracy: 0.9695, Precision: 1.0000
Confusion Matrix:
 [[966   0]
 [ 34 115]]
```

Generating model performance visualizations...





Model Performance Summary

Best Model: Multinomial Naive Bayes

Best Accuracy: 0.9785

Best Precision: 1.0000

```
🧪 Testing the model with example messages...
============================================================
Message 1: Congratulations! You've won a $1000 Walmart gift card. Text YES to claim.
Prediction: SPAM (Confidence: 0.7023)
Spam Probability: 0.7023
Ham Probability: 0.2977
------------------------------------------------
Message 2: Hey, are we still meeting for lunch tomorrow?
Prediction: HAM (Confidence: 0.9974)
Spam Probability: 0.0026
Ham Probability: 0.9974
------------------------------------------------
Message 3: URGENT: Your bank account has been suspended. Click here to verify your details.
Prediction: HAM (Confidence: 0.5047)
Spam Probability: 0.4953
Ham Probability: 0.5047
------------------------------------------------
Message 4: Ok, see you later. Thanks for the update.
Prediction: HAM (Confidence: 0.9900)
Spam Probability: 0.0100
Ham Probability: 0.9900
------------------------------------------------
Message 5: FREE entry to win a new car! Reply NOW to claim your prize!
Prediction: SPAM (Confidence: 0.9647)
Spam Probability: 0.9647
Ham Probability: 0.0353
------------------------------------------------
Message 6: Hi mom, can you pick me up from school today?
Prediction: HAM (Confidence: 0.9906)
Spam Probability: 0.0094
Ham Probability: 0.9906
```