Email spam filtering

The email spam filtering model and visualize the results using Power BI, you can follow these steps:

1. Prepare Your Dataset

Ensure your dataset contains columns for the email text and labels indicating whether each email is spam or not. If you don't have a dataset, you can find one on platforms like Kaggle.

2. Build the Spam Filtering Model

You can use Python to build a spam filtering model. Here's a basic example using the Naive Bayes classifier:

```
# Import necessary libraries

import pandas as pd

from sklearn.feature_extraction.text import CountVectorizer

from sklearn.model_selection import train_test_split

from sklearn.naive_bayes import MultinomialNB

from sklearn.metrics import accuracy_score


# Load the dataset with an alternative encoding

data = pd.read_csv(r'C:\power bi\Email Spam Filtering\spam.csv', encoding='latin1')


# Check the first few rows of the dataset

print(data.head())


# Check the column names and their count

print(data.columns)

print(f"Number of columns: {len(data.columns)}")
```

```python
# Assuming the columns are 'label' and 'text', if not, rename them accordingly

# Adjust the number of column names based on the actual number of columns

if len(data.columns) == 2:

    data.columns = ['label', 'text']

else:

    print("Unexpected number of columns. Please check the dataset.")

    print(data.columns)  # Print the actual column names for further inspection


# Preprocess the data if the columns are correctly named

if 'text' in data.columns and 'label' in data.columns:

    data['text'] = data['text'].str.lower()

    data['text'] = data['text'].str.replace('[^\w\s]', '', regex=True)

    # Ensure there are no missing values

    data.dropna(subset=['text', 'label'], inplace=True)

    # Extract features

    vectorizer = CountVectorizer(stop_words='english')

    X = vectorizer.fit_transform(data['text'])

    y = data['label']

    # Split the data

    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)


    # Train the model

    model = MultinomialNB()
```

```python
    model.fit(X_train, y_train)

    # Predict and evaluate

    y_pred = model.predict(X_test)

    accuracy = accuracy_score(y_test, y_pred)

    print(f'Accuracy: {accuracy}')

    # Add predictions to the dataset

    data['predicted_label'] = model.predict(vectorizer.transform(data['text']))

    # Output the data for Power BI

    output = data[['text', 'label', 'predicted_label']]

    print(output.head())

else:

    print("Required columns 'text' and 'label' are not present in 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
Index(['v1', 'v2', 'Unnamed: 2', 'Unnamed: 3', 'Unnamed: 4'], dtype='object')
Number of columns: 5
Unexpected number of columns. Please check the dataset.
Index(['v1', 'v2', 'Unnamed: 2', 'Unnamed: 3', 'Unnamed: 4'], dtype='object')
Required columns 'text' and 'label' are not present in the dataset.
```
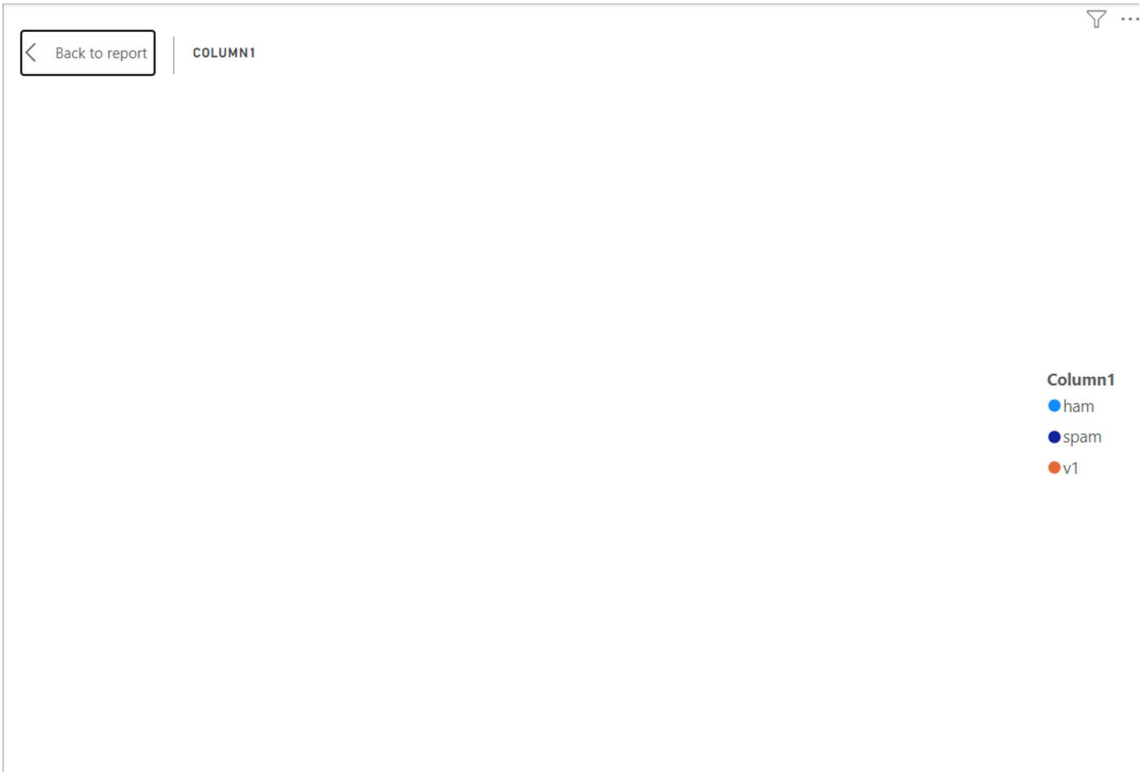
3. Visualize in Power BI

1. **Import the Data**: Load the output_for_power_bi.csv file into Power BI.
2. **Create Visualizations**: Use Power BI to create visualizations that show the distribution of spam vs. non-spam emails, accuracy of the model, and other insights.

Column2

- <#> in mca. But not conform.
- <#> mins but i had to stop somew...
- <DECIMAL> m but its not a commo...
- and picking them up from various points
- came to look at the flat, seems ok, in his ...
- gonna let me know cos comes bak from ...
- said kiss, kiss, i can't do the sound effect...
- says that he's quitting at least5times a d...
- what number do u live at? Is it 11?
- <#> am I think? Should say on sylla...
- <#> great loxahatchee xmas tree b...
- <#> is fast approaching. So, Wish u...
- <#> w jetton ave if you forgot
- <#> %of pple marry with their lover...
- <#> , that's all? Guess that's easy en...
- <#> ISH MINUTES WAS 5 MINUTES ...
- (And my man carlos is definitely coming ...
- (Bank of Granite issues Strong-Buy) EXPL...
- (I should add that I don't really care and i...
- (No promises on when though, haven't e...

▼

Column3

-
- \"It is d wonderful fruit that a tree gives ...
- always give response 2 who cares 4 U\"......
- b'coz nobody will fight for u. Only u &a...
- bt not his girlfrnd... G o o d n i g h t . . .@"
- But at d end my love compromised me f...
- but dont try to prove it..\" .Gud noon...."
- but dont try to prove\" ..... Gud mrng..."
- but watever u shared should be true\"...."
- Dont Come Near My Body..!! Bcoz My H...
- don't miss ur best life for anything... Gud...
- ENJOYIN INDIANS AT THE MO..yeP. SaLL...
- GOD said
- Gud night...."
- HAD A COOL NYTHO
- HOPE UR OK... WILL GIVE U A BUZ WEDL...
- hopeSo hunny. i amnow feelin ill & ithin...
- HOWU DOIN? FOUNDURSELF A JOBYET ...
- I don't mind
- I'll come up"

▼

Column4

●
● \"OH No! COMPETITION\". Who knew
● ILLSPEAK 2 U2MORO WEN IM NOT ASLE...
● MK17 92H. 450Ppw 16"
● TX 4 FONIN HON
● why to miss them
● GE
● i wil tolerat.bcs ur my someone..... But
● IåÕL CALL U\""
● U NO THECD ISV.IMPORTANT TOME 4 2...
● whoever is the KING\"!... Gud nyt"

Column5

●
● CALL 2MWEN IM BK FRMCLOUD 9! J X\""
● just Keep-in-touch\" gdeve.."
● Never comfort me with a lie\" gud ni8 an...
● one day these two will become FREINDS ...
● GNT:-)"