week2-food

October 21, 2023

FOOD AND BEVERAGES

FOR CLEANING DATA ABOUT CAFES

[1]:
```
import pandas as pd
import re

# Define a function to perform the data cleaning tasks
def clean_text(text):
    # Convert to lowercase
    text = text.lower()
    # Remove links (URLs)
    text = re.sub(r'http\S+', '', text)

    # Remove punctuation and symbols
    text = re.sub(r'[^\w\s]', ' ', text)

    # Remove @user mentions
    text = re.sub(r'@\w+', '', text)

    # Remove hashtags
    text = re.sub(r'#\w+', '', text)



    return text

# Load the data from an Excel file
file_path = "/Users/harshith/Downloads/WEEK2/Food and Beverages/cafe refresh.
 ↪csv"
df = pd.read_csv(file_path)

# Apply the cleaning function to the specified column
df['full_text'] = df['full_text'].apply(clean_text)
# Remove duplicates
df.drop_duplicates(subset=['full_text'], inplace=True)

# Remove rows with null values
```

```
df.dropna(subset=['full_text'], inplace=True)

# Save the cleaned data back to a new Excel file
cleaned_file_path = "//Users/harshith/Desktop/untitled folder 3/food and␣
 ↪beverages/cafe.csv"
df.to_csv(cleaned_file_path, index=False)

print(f"Cleaned data saved to {cleaned_file_path}")
```

```
Cleaned data saved to //Users/harshith/Desktop/untitled folder 3/food and
beverages/cafe.csv
```

[2]:
```
import pandas as pd
import re

# Define a function to perform the data cleaning tasks
def clean_text(text):
    # Convert to lowercase
    text = text.lower()
    # Remove links (URLs)
    text = re.sub(r'http\S+', '', text)

    # Remove punctuation and symbols
    text = re.sub(r'[^\w\s]', ' ', text)

    # Remove @user mentions
    text = re.sub(r'@\w+', '', text)

    # Remove hashtags
    text = re.sub(r'#\w+', '', text)



    return text

# Load the data from an Excel file
file_path = "/Users/harshith/Downloads/WEEK2/Food and Beverages/foodcourt.csv"
df = pd.read_csv(file_path)

# Apply the cleaning function to the specified column
df['full_text'] = df['full_text'].apply(clean_text)
# Remove duplicates
df.drop_duplicates(subset=['full_text'], inplace=True)

# Remove rows with null values
df.dropna(subset=['full_text'], inplace=True)
```

```
# Save the cleaned data back to a new Excel file
cleaned_file_path = "//Users/harshith/Desktop/untitled folder 3/food and␣
 ↪beverages/foodcourt.csv"
df.to_csv(cleaned_file_path, index=False)

print(f"Cleaned data saved to {cleaned_file_path}")
```

```
Cleaned data saved to //Users/harshith/Desktop/untitled folder 3/food and
beverages/foodcourt.csv
```

[3]:
```
import pandas as pd
import re

# Define a function to perform the data cleaning tasks
def clean_text(text):
    # Convert to lowercase
    text = text.lower()
    # Remove links (URLs)
    text = re.sub(r'http\S+', '', text)

    # Remove punctuation and symbols
    text = re.sub(r'[^\w\s]', ' ', text)

    # Remove @user mentions
    text = re.sub(r'@\w+', '', text)

    # Remove hashtags
    text = re.sub(r'#\w+', '', text)



    return text

# Load the data from an Excel file
file_path = "/Users/harshith/Downloads/WEEK2/Food and Beverages/restaurant␣
 ↪refresh.csv"
df = pd.read_csv(file_path)

# Apply the cleaning function to the specified column
df['full_text'] = df['full_text'].apply(clean_text)
# Remove duplicates
df.drop_duplicates(subset=['full_text'], inplace=True)

# Remove rows with null values
df.dropna(subset=['full_text'], inplace=True)

# Save the cleaned data back to a new Excel file
```

```
cleaned_file_path = "//Users/harshith/Desktop/untitled folder 3/food and␣
  ↪beverages/restaurant.csv"
df.to_csv(cleaned_file_path, index=False)

print(f"Cleaned data saved to {cleaned_file_path}")
```

```
Cleaned data saved to //Users/harshith/Desktop/untitled folder 3/food and
beverages/restaurant.csv
```

[ ]: