week2-hospitality

October 21, 2023

HOSPITALITY

[14]:
```
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
  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/Hospitality/check-in.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/hospitality/
 ↪check-in.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/hospitality/check-in.csv
```

[15]:
```
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/Hospitality/hospitality_2.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/hospitality/
 ↪hospitality_2.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/hospitality/hospitality_2.csv
```

[16]:
```
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/Hospitality/lounges.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/hospitality/
 ↪lounges.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/hospitality/lounges.csv
```

[17]:
```
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/Hospitality/waiting.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/hospitality/
 ↪waiting.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/hospitality/waiting.csv
```

[18]:
```python
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/Hospitality/waiting-area.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/hospitality/
 ↪waiting-area.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/hospitality/waiting-area.csv
```

[ ]: