```
import pandas as pd
import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt

data = pd.read_excel(r"C:\Users\saisn\Downloads\Onyx Data -DataDNA
Dataset Challenge - Email Analysis Dataset - July 2024\Onyx Data -
DataDNA Dataset Challenge - Email Analysis Dataset - July 2024.xlsx")

data.head()
```

```
   Email id     From Name From seniority       From Department  \
0         1  Leona McAree        C-level  Executive Management
1         2  Leona McAree        C-level  Executive Management
2         3  Leona McAree        C-level  Executive Management
3         4  Leona McAree        C-level  Executive Management
4         5  Leona McAree        C-level  Executive Management

               To Name             To seniority    To Department  \
0  Elsinore Waterland  Middle level management  Human Resources
1  Elsinore Waterland  Middle level management  Human Resources
2  Elsinore Waterland  Middle level management  Human Resources
3  Elsinore Waterland  Middle level management  Human Resources
4  Elsinore Waterland  Middle level management  Human Resources

          Email topic       Date Sentiment Is opened?   Device  \
0  Operational Issues 2024-03-14   neutral     opened  desktop
1   HR related topics 2024-03-22  positive     opened   mobile
2  Meeting Scheduling 2024-03-26   neutral   unopened  desktop
3   HR related topics 2024-03-11   neutral     opened   mobile
4   HR related topics 2024-03-11  positive     opened   mobile

  Within work hours Within workdays
0               yes              yes
1                no              yes
2               yes              yes
3               yes              yes
4               yes              yes
```

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1132 entries, 0 to 1131
Data columns (total 14 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   Email id           1132 non-null   int64
 1   From Name          1132 non-null   object
 2   From seniority     1132 non-null   object
 3   From Department    1132 non-null   object
 4   To Name            1132 non-null   object
```

```
 5   To seniority       1132 non-null   object
 6   To Department      1132 non-null   object
 7   Email topic        1132 non-null   object
 8   Date               1132 non-null   datetime64[ns]
 9   Sentiment          1132 non-null   object
 10  Is opened?         1132 non-null   object
 11  Device             1132 non-null   object
 12  Within work hours  1132 non-null   object
 13  Within workdays    1132 non-null   object
dtypes: datetime64[ns](1), int64(1), object(12)
memory usage: 123.9+ KB
```

```
data.describe()
```

```
          Email id                           Date
count  1132.000000                           1132
mean    624.666961  2024-03-16 11:07:50.671378176
min       1.000000            2024-03-01 00:00:00
25%     348.750000            2024-03-11 00:00:00
50%     631.500000            2024-03-15 00:00:00
75%     914.250000            2024-03-25 00:00:00
max    1197.000000            2024-03-31 00:00:00
std     337.942989                            NaN
```

```
data.tail()
```

```
      Email id       From Name           From seniority From
Department  \
1127      1193  Emmey Matoshin  Middle level management
Sales
1128      1194  Emmey Matoshin  Middle level management
Sales
1129      1195  Emmey Matoshin  Middle level management
Sales
1130      1196  Emmey Matoshin  Middle level management
Sales
1131      1197  Emmey Matoshin  Middle level management
Sales

              To Name             To seniority           To Department
\
1127  Blithe Arnaudet  Middle level management  Finance and Accounting

1128  Blithe Arnaudet  Middle level management  Finance and Accounting

1129  Blithe Arnaudet  Middle level management  Finance and Accounting

1130  Blithe Arnaudet  Middle level management  Finance and Accounting

1131  Blithe Arnaudet  Middle level management  Finance and Accounting
```

```
                 Email topic       Date Sentiment Is opened?   Device
\
1127  Finance and Accounting 2024-03-28  positive     opened   mobile

1128  Finance and Accounting 2024-03-07  negative     opened  desktop

1129      Operational Issues 2024-03-14   neutral     opened  desktop

1130      Meeting Scheduling 2024-03-08  positive     opened  desktop

1131  Finance and Accounting 2024-03-23   neutral     opened   mobile


     Within work hours Within workdays
1127               yes             yes
1128               yes             yes
1129                no             yes
1130               yes             yes
1131                no              no

data.isnull()

      Email id  From Name  From seniority  From Department  To Name  \
0        False      False           False            False    False
1        False      False           False            False    False
2        False      False           False            False    False
3        False      False           False            False    False
4        False      False           False            False    False
...        ...        ...             ...              ...      ...
1127     False      False           False            False    False
1128     False      False           False            False    False
1129     False      False           False            False    False
1130     False      False           False            False    False
1131     False      False           False            False    False

      To seniority  To Department  Email topic   Date  Sentiment  Is
opened?  \
0            False          False        False  False      False
False
1            False          False        False  False      False
False
2            False          False        False  False      False
False
3            False          False        False  False      False
False
4            False          False        False  False      False
False
...            ...            ...          ...    ...        ...
...
```

```
1127         False          False        False  False      False
False
1128         False          False        False  False      False
False
1129         False          False        False  False      False
False
1130         False          False        False  False      False
False
1131         False          False        False  False      False
False

      Device  Within work hours  Within workdays
0      False              False            False
1      False              False            False
2      False              False            False
3      False              False            False
4      False              False            False
...      ...                ...              ...
1127   False              False            False
1128   False              False            False
1129   False              False            False
1130   False              False            False
1131   False              False            False

[1132 rows x 14 columns]

data.describe().T

           count                            mean
min  \
Email id  1132.0                      624.666961                    1.0

Date        1132  2024-03-16 11:07:50.671378176  2024-03-01 00:00:00


                            25%                  50%
75%  \
Email id               348.75                631.5
914.25
Date      2024-03-11 00:00:00  2024-03-15 00:00:00  2024-03-25
00:00:00

                            max         std
Email id                 1197.0  337.942989
Date      2024-03-31 00:00:00         NaN

data

      Email id       From Name           From seniority       From
Department  \
0            1    Leona McAree                  C-level  Executive
```

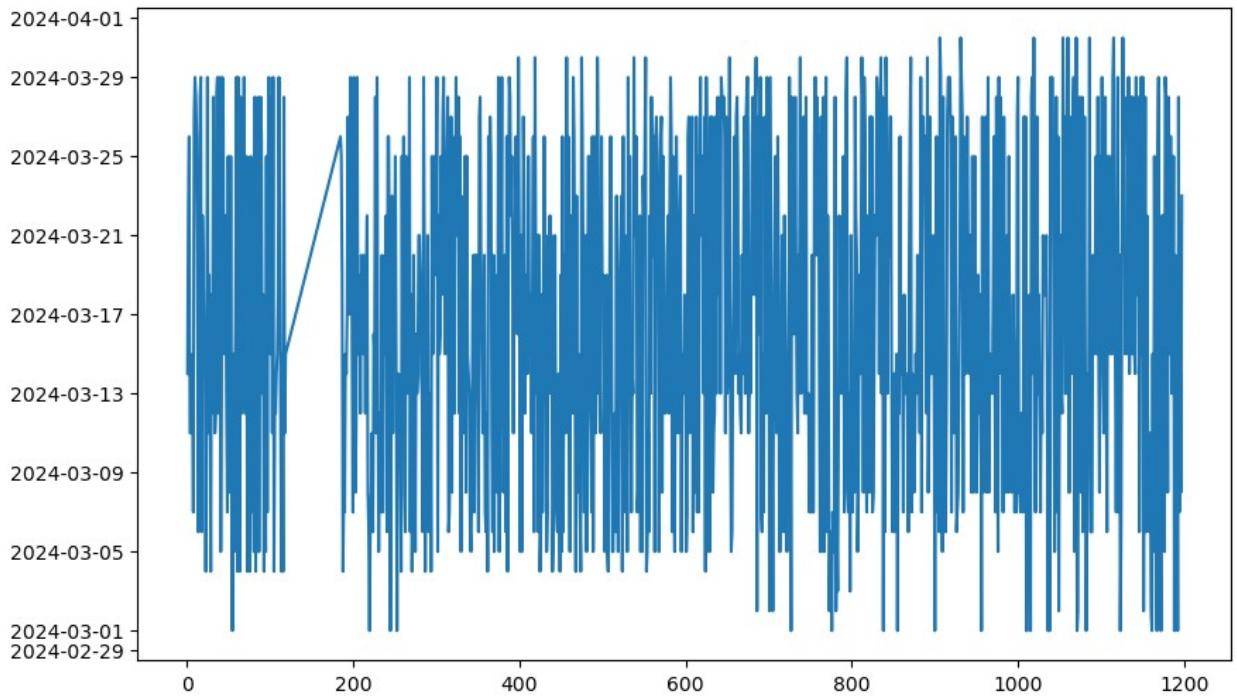
```
Management
1            2    Leona McAree                  C-level  Executive
Management
2            3    Leona McAree                  C-level  Executive
Management
3            4    Leona McAree                  C-level  Executive
Management
4            5    Leona McAree                  C-level  Executive
Management
...        ...             ...                       ...
...
1127      1193  Emmey Matoshin  Middle level management
Sales
1128      1194  Emmey Matoshin  Middle level management
Sales
1129      1195  Emmey Matoshin  Middle level management
Sales
1130      1196  Emmey Matoshin  Middle level management
Sales
1131      1197  Emmey Matoshin  Middle level management
Sales

                  To Name             To seniority            To
Department  \
0     Elsinore Waterland  Middle level management         Human
Resources
1     Elsinore Waterland  Middle level management         Human
Resources
2     Elsinore Waterland  Middle level management         Human
Resources
3     Elsinore Waterland  Middle level management         Human
Resources
4     Elsinore Waterland  Middle level management         Human
Resources
...                  ...                         ...
...
1127     Blithe Arnaudet  Middle level management  Finance and
Accounting
1128     Blithe Arnaudet  Middle level management  Finance and
Accounting
1129     Blithe Arnaudet  Middle level management  Finance and
Accounting
1130     Blithe Arnaudet  Middle level management  Finance and
Accounting
1131     Blithe Arnaudet  Middle level management  Finance and
Accounting

                  Email topic       Date Sentiment Is opened?   Device
\
```

```
0          Operational Issues 2024-03-14   neutral     opened  desktop
1           HR related topics 2024-03-22  positive     opened   mobile
2          Meeting Scheduling 2024-03-26   neutral   unopened  desktop
3           HR related topics 2024-03-11   neutral     opened   mobile
4           HR related topics 2024-03-11  positive     opened   mobile
...                       ...        ...       ...        ...      ...
1127  Finance and Accounting 2024-03-28  positive     opened   mobile
1128  Finance and Accounting 2024-03-07  negative     opened  desktop
1129       Operational Issues 2024-03-14   neutral     opened  desktop
1130       Meeting Scheduling 2024-03-08  positive     opened  desktop
1131  Finance and Accounting 2024-03-23   neutral     opened   mobile


     Within work hours Within workdays
0                  yes             yes
1                   no             yes
2                  yes             yes
3                  yes             yes
4                  yes             yes
...                ...             ...
1127               yes             yes
1128               yes             yes
1129                no             yes
1130               yes             yes
1131                no              no

[1132 rows x 14 columns]

data.head(5)

   Email id     From Name From seniority       From Department  \
0         1  Leona McAree        C-level  Executive Management
1         2  Leona McAree        C-level  Executive Management
2         3  Leona McAree        C-level  Executive Management
3         4  Leona McAree        C-level  Executive Management
4         5  Leona McAree        C-level  Executive Management

               To Name             To seniority    To Department  \
0  Elsinore Waterland  Middle level management  Human Resources
1  Elsinore Waterland  Middle level management  Human Resources
2  Elsinore Waterland  Middle level management  Human Resources
```

```
3  Elsinore Waterland  Middle level management  Human Resources
4  Elsinore Waterland  Middle level management  Human Resources
```

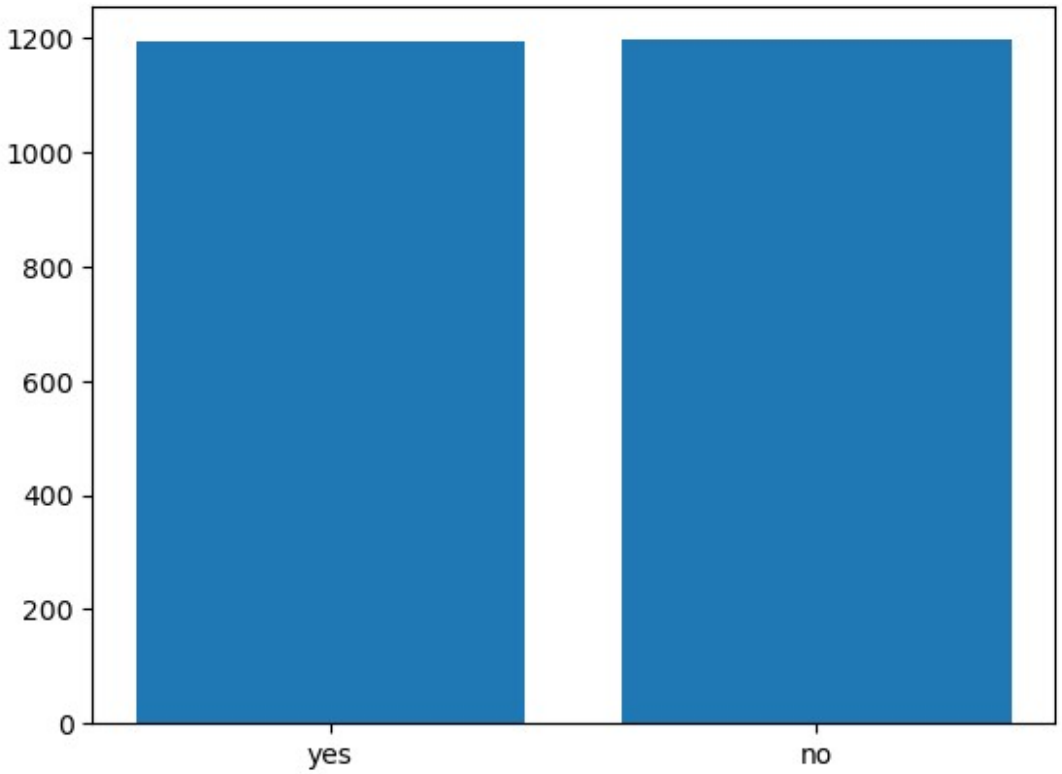
```
          Email topic       Date Sentiment Is opened?   Device  \
0  Operational Issues 2024-03-14   neutral     opened  desktop
1   HR related topics 2024-03-22  positive     opened   mobile
2  Meeting Scheduling 2024-03-26   neutral   unopened  desktop
3   HR related topics 2024-03-11   neutral     opened   mobile
4   HR related topics 2024-03-11  positive     opened   mobile

  Within work hours Within workdays
0               yes              yes
1                no              yes
2               yes              yes
3               yes              yes
4               yes              yes
```

```
plt.figure(figsize=(10, 6))
plt.plot(data['Email id'],data['Date'])
plt.show()
```

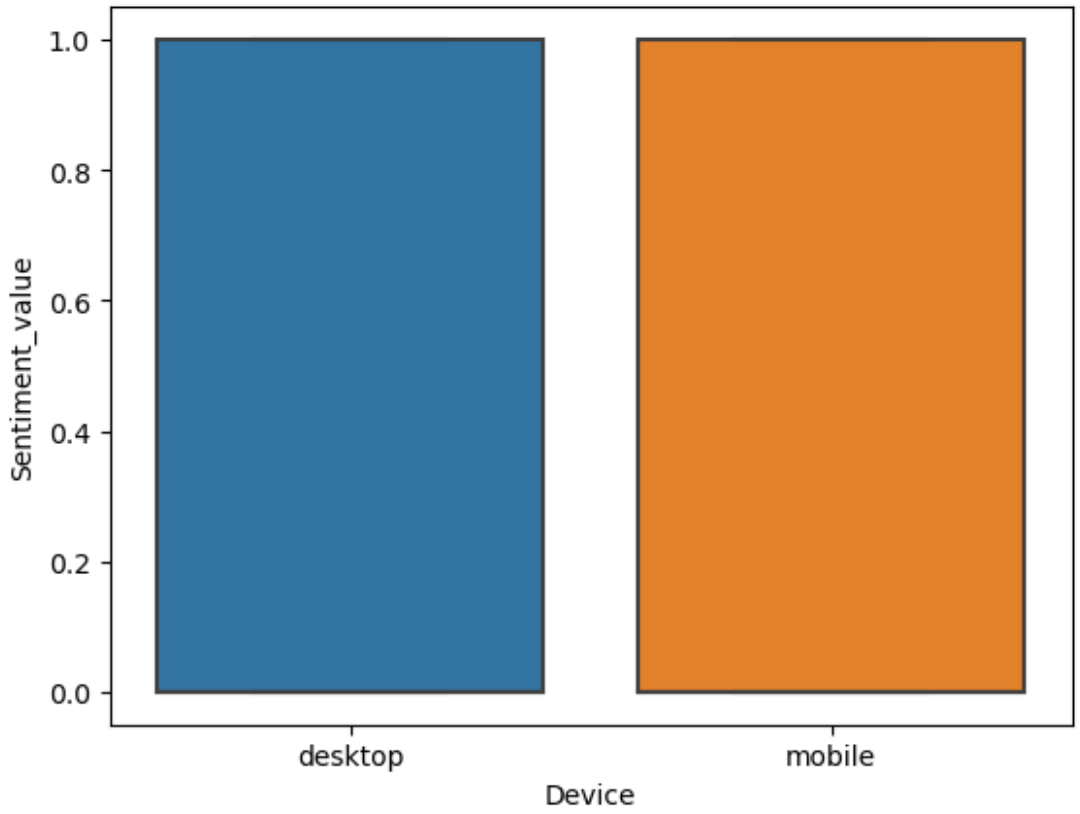



```
plt.bar(data['Within work hours'],data['Email id'])
```

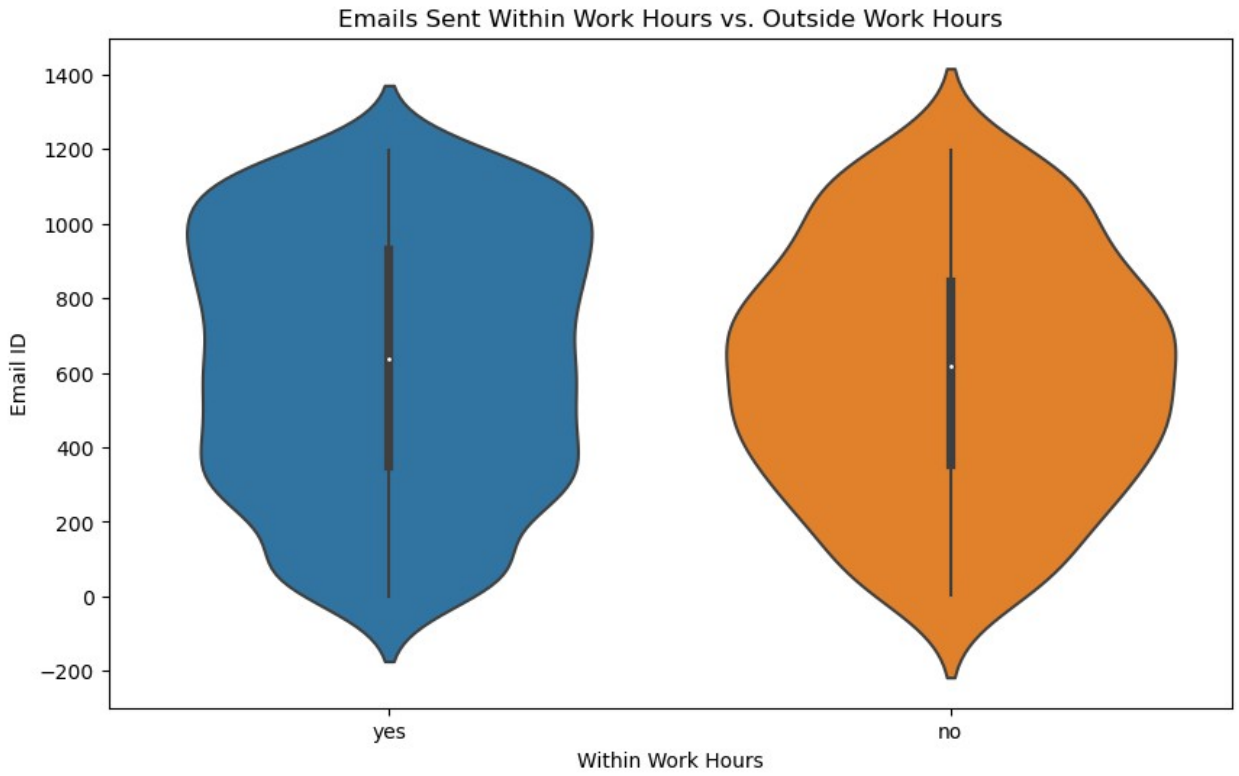
```
<BarContainer object of 1132 artists>
```




```
import seaborn as sns
```

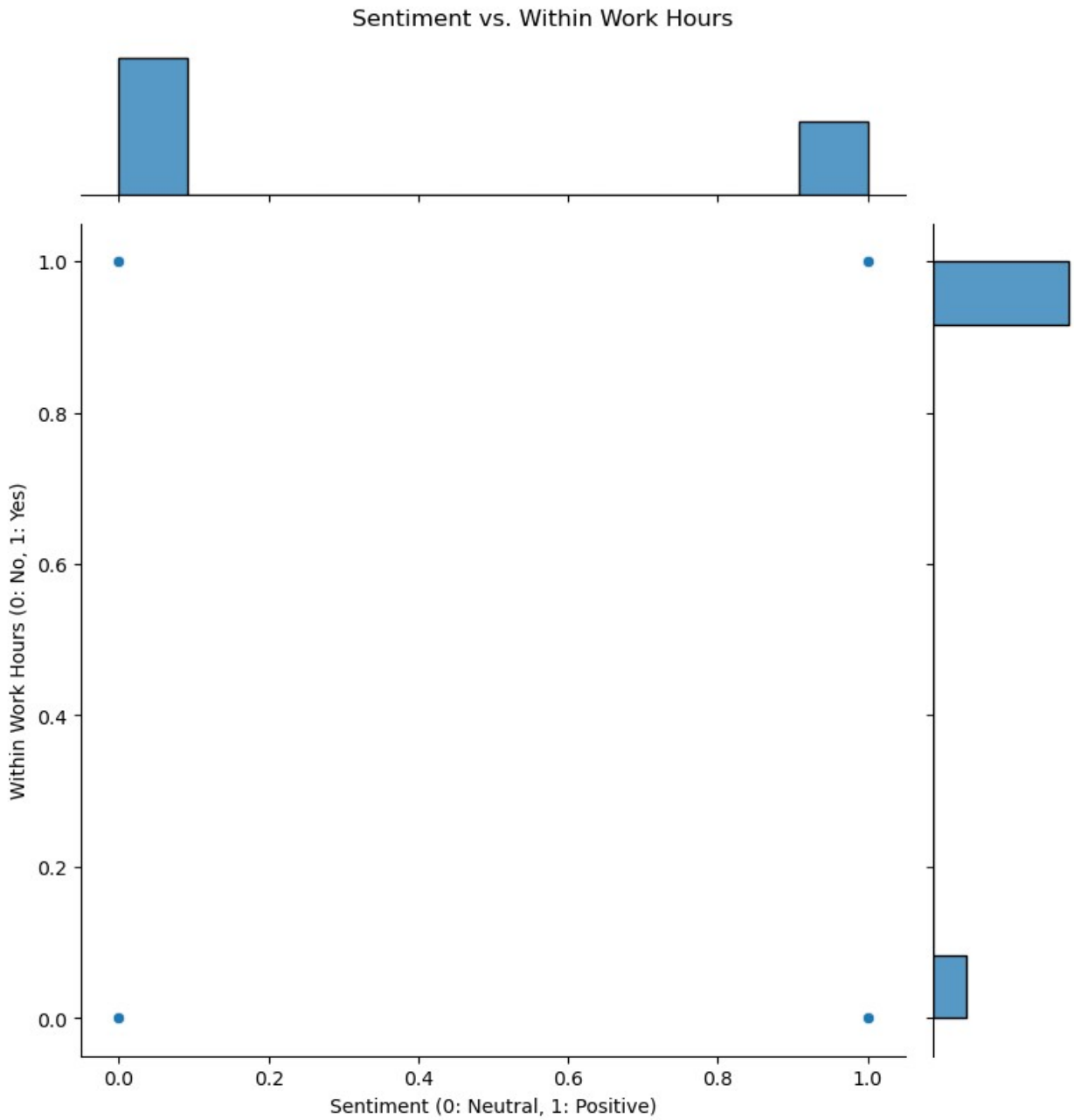



```
# Violin Plot: Emails Sent Within Work Hours
plt.figure(figsize=(10, 6))
sns.violinplot(x='Within work hours', y='Email id', data=df)
plt.title('Emails Sent Within Work Hours vs. Outside Work Hours')
plt.xlabel('Within Work Hours')
plt.ylabel('Email ID')
plt.show()
```



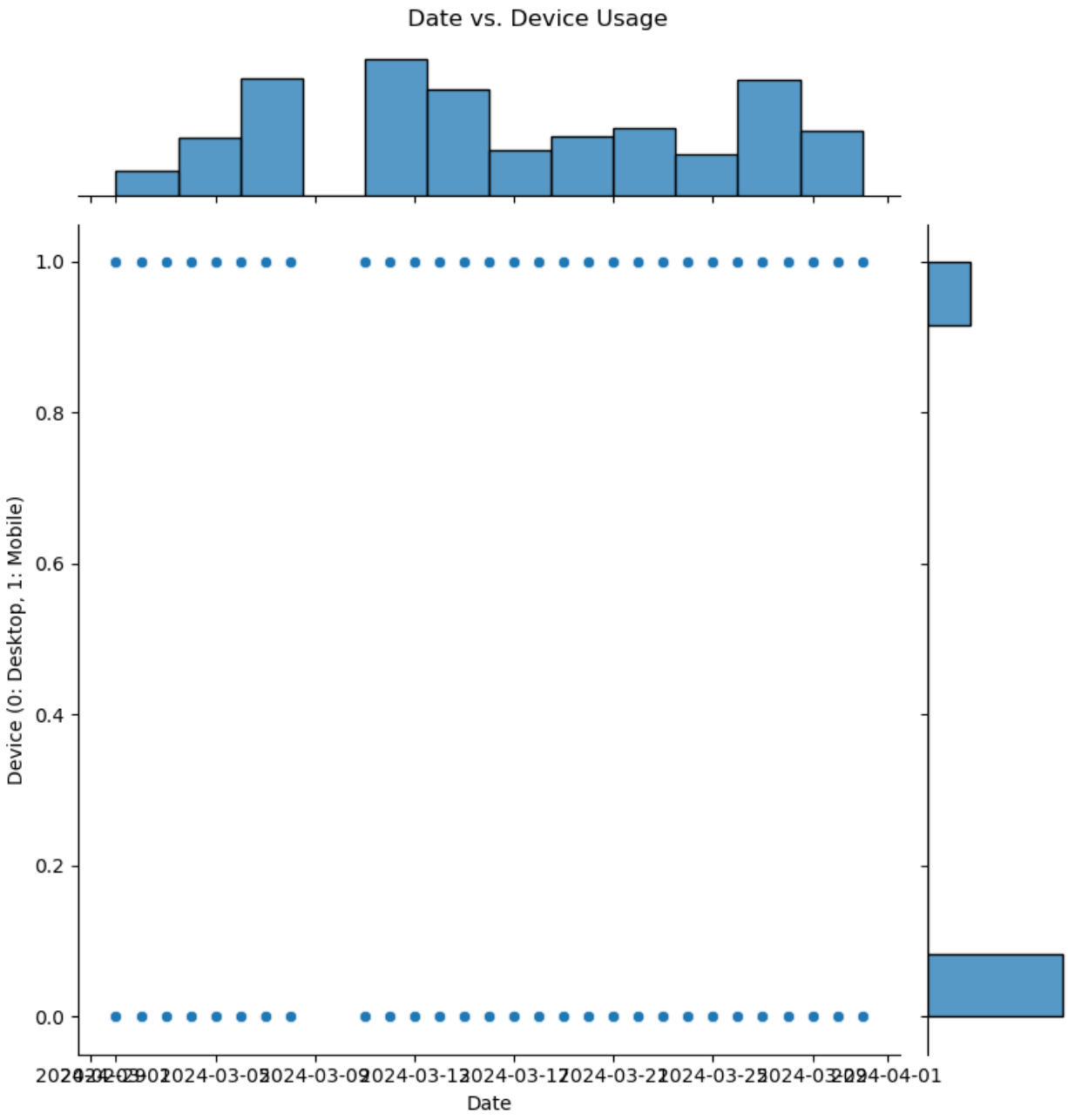

```
# Mapping sentiment to numerical values for plotting
sentiment_map = {'neutral': 0, 'positive': 1}
df['Sentiment_value'] = df['Sentiment'].map(sentiment_map)
plt.figure(figsize=(8, 8))
# Convert 'Within work hours' to numerical values for plotting
work_hours_map = {'yes': 1, 'no': 0}
df['Within_work_hours_value'] = df['Within work
hours'].map(work_hours_map)

sns.jointplot(x='Sentiment_value', y='Within_work_hours_value',
data=df, kind='scatter', height=8)
plt.suptitle('Sentiment vs. Within Work Hours', y=1.02)
plt.xlabel('Sentiment (0: Neutral, 1: Positive)')
plt.ylabel('Within Work Hours (0: No, 1: Yes)')
plt.show()
```

```
<Figure size 800x800 with 0 Axes>
```



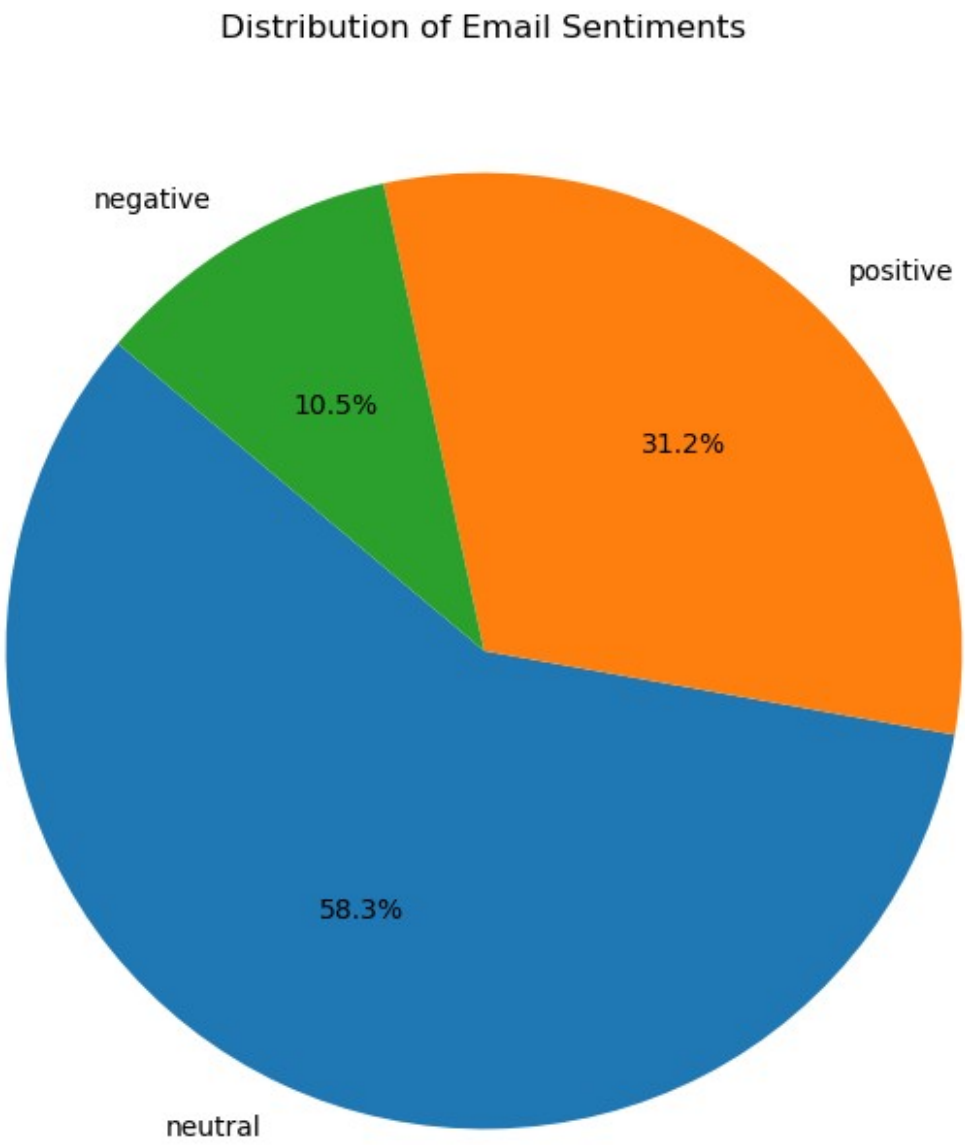

```
# Joint Plot: Date vs. Device
# Convert 'Device' to numerical values for plotting
device_map = {'desktop': 0, 'mobile': 1}
df['Device_value'] = df['Device'].map(device_map)
plt.figure(figsize=(8, 8))
sns.jointplot(x='Date', y='Device_value', data=df, kind='scatter',
height=8)
plt.suptitle('Date vs. Device Usage', y=1.02)
plt.xlabel('Date')
plt.ylabel('Device (0: Desktop, 1: Mobile)')
plt.show()
```

```
<Figure size 800x800 with 0 Axes>
```




```
df = pd.DataFrame(data)

# Count the occurrences of each sentiment
sentiment_counts = df['Sentiment'].value_counts()

# Plotting the pie chart
plt.figure(figsize=(8, 8))
```

```
plt.pie(sentiment_counts, labels=sentiment_counts.index,
autopct='%1.1f%%', startangle=140)
plt.title('Distribution of Email Sentiments')
plt.show()
```




```
df = pd.DataFrame(data)

# Filter to only include numerical columns
numerical_df = df.select_dtypes(include=['number'])

# Generate descriptive statistics for numerical data
```

```
desc = numerical_df.describe()

# Plot the heatmap
plt.figure(figsize=(10, 6))
sns.heatmap(desc, annot=True, linewidths=0.5, cmap='coolwarm')
plt.title('Summary Statistics Heatmap')
plt.show()
```