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

In [2]:
```
import numpy as np
```

In [7]:
```
ads = pd.read_csv("train_data_ads.csv")
```

In [8]:
```
feeds = pd.read_csv("train_data_feeds.csv")
```

In [9]:
```
ads = ads.drop_duplicates(subset='user_id', keep='first', inplace=False)
feeds = feeds.drop_duplicates(subset='u_userId', keep='first', inplace=False)
```

In [10]:
```
feeds.shape
```

Out[10]:
```
(180123, 28)
```

In [11]:
```
ads.shape
```

Out[11]:
```
(65297, 35)
```

In [12]:
```
feeds['user_id'] = feeds['u_userId']
```

In [13]:
```
feeds = feeds.drop('u_userId', axis = 1)
```

In [14]:
```
merged = pd.merge(ads, feeds, on = 'user_id', how = 'outer')
```

In [15]:
```
merged.head()
```

Out[15]:

	log_id	label_x	user_id	age	gender	residence	city	city_rank	series_dev	series_group	...	e
0	373250.0	0.0	100005	3.0	2.0	16.0	147.0	2.0	32.0	6.0	...	
1	101100.0	0.0	100006	5.0	2.0	13.0	191.0	4.0	32.0	6.0	...	
2	742637.0	0.0	100009	5.0	2.0	46.0	354.0	2.0	11.0	8.0	...	
3	744753.0	0.0	100010	3.0	4.0	33.0	319.0	3.0	31.0	3.0	...	
4	669191.0	0.0	100019	7.0	2.0	16.0	310.0	2.0	16.0	5.0	...	

5 rows × 62 columns

In [16]:
```
merged.shape
```

Out[16]:
```
(180123, 62)
```

In [17]:
```
76756 + 32278
```

Out[17]:
```
109034
```

In [18]:
```
merged.columns
```

Out[18]:
```
Index(['log_id', 'label_x', 'user_id', 'age', 'gender', 'residence', 'city',
       'city_rank', 'series_dev', 'series_group', 'emui_dev', 'device_name',
       'device_size', 'net_type', 'task_id', 'adv_id', 'creat_type_cd',
       'adv_prim_id', 'inter_type_cd', 'slot_id', 'site_id', 'spread_app_id',
       'hispace_app_tags', 'app_second_class', 'app_score',
       'ad_click_list_v001', 'ad_click_list_v002', 'ad_click_list_v003',
       'ad_close_list_v001', 'ad_close_list_v002', 'ad_close_list_v003',
       'pt_d', 'u_newsCatInterestsST_x', 'u_refreshTimes_x',
       'u_feedLifeCycle_x', 'u_phonePrice', 'u_browserLifeCycle',
       'u_browserMode', 'u_feedLifeCycle_y', 'u_refreshTimes_y',
       'u_newsCatInterests', 'u_newsCatDislike', 'u_newsCatInterestsST_y',
       'u_click_ca2_news', 'i_docId', 'i_s_sourceId', 'i_regionEntity',
       'i_cat', 'i_entities', 'i_dislikeTimes', 'i_upTimes', 'i_dtype', 'e_ch',
       'e_m', 'e_po', 'e_pl', 'e_rn', 'e_section', 'e_et', 'label_y',
       'cillabel', 'pro'],
      dtype='object')
```

In [19]:
```
df_cust = merged[merged['label_y'] == 1]
```

In [20]:
```
df_cust.shape
```

Out[20]:
```
(14450, 62)
```

In [21]:
```
df_cust
```

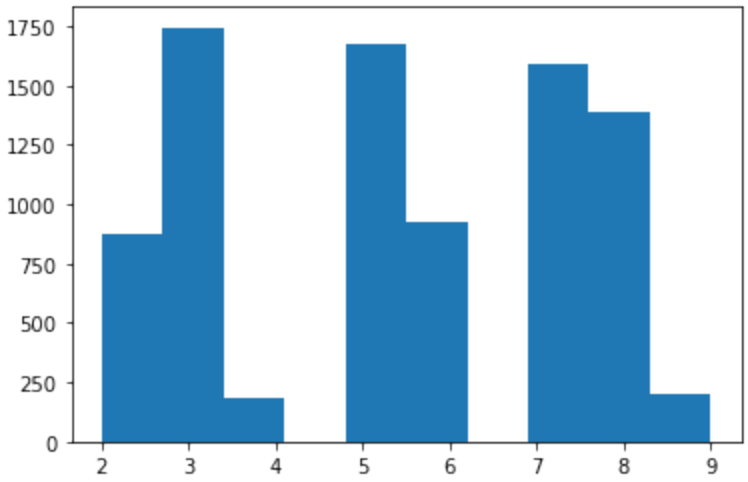
Out[21]:

	log_id	label_x	user_id	age	gender	residence	city	city_rank	series_dev	series_grou
18	583728.0	0.0	100108	2.0	3.0	21.0	200.0	4.0	30.0	3.
21	364370.0	0.0	100127	7.0	2.0	17.0	343.0	5.0	16.0	5.
23	588242.0	0.0	100149	8.0	2.0	16.0	425.0	2.0	34.0	7.
27	679513.0	0.0	100158	6.0	4.0	33.0	319.0	3.0	27.0	2.
29	1084910.0	0.0	100166	5.0	2.0	30.0	113.0	5.0	16.0	5.
...	...	...	...	...	...	...	...	...	...	.
179951	NaN	NaN	131907	NaN	NaN	NaN	NaN	NaN	NaN	Na
179959	NaN	NaN	123724	NaN	NaN	NaN	NaN	NaN	NaN	Na
179996	NaN	NaN	215157	NaN	NaN	NaN	NaN	NaN	NaN	Na
179997	NaN	NaN	107610	NaN	NaN	NaN	NaN	NaN	NaN	Na
180013	NaN	NaN	246671	NaN	NaN	NaN	NaN	NaN	NaN	Na

14450 rows × 62 columns

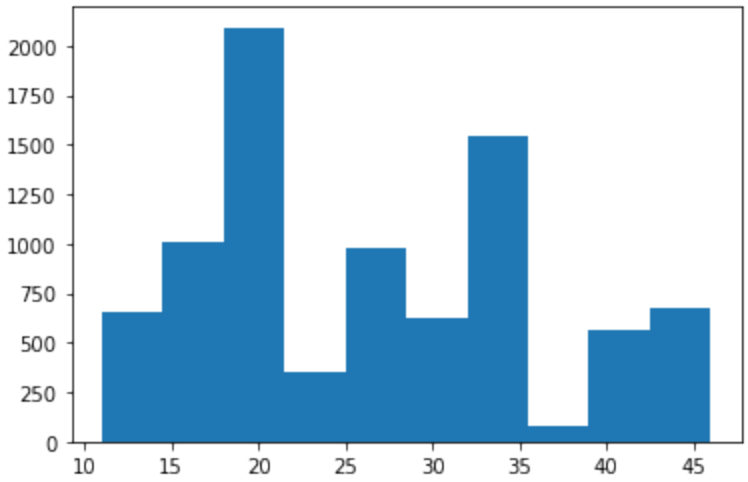
In [22]:
```
import matplotlib.pyplot as plt
plt.hist(df_cust['age'])
```

Out[22]:
```
(array([ 874., 1744.,  184.,    0., 1673.,  922.,    0., 1589., 1388.,
         204.]),
 array([2. , 2.7, 3.4, 4.1, 4.8, 5.5, 6.2, 6.9, 7.6, 8.3, 9. ]),
 <BarContainer object of 10 artists>)
```



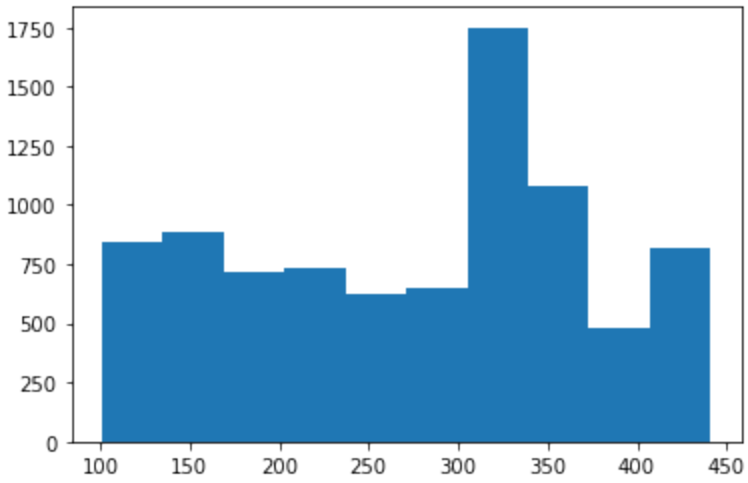
In [23]:
```
plt.hist(df_cust['residence'])
```

Out[23]:
```
(array([ 651., 1013., 2093.,  353.,  983.,  626., 1544.,   77.,  562.,
         676.]),
 array([11. , 14.5, 18. , 21.5, 25. , 28.5, 32. , 35.5, 39. , 42.5, 46. ]),
 <BarContainer object of 10 artists>)
```



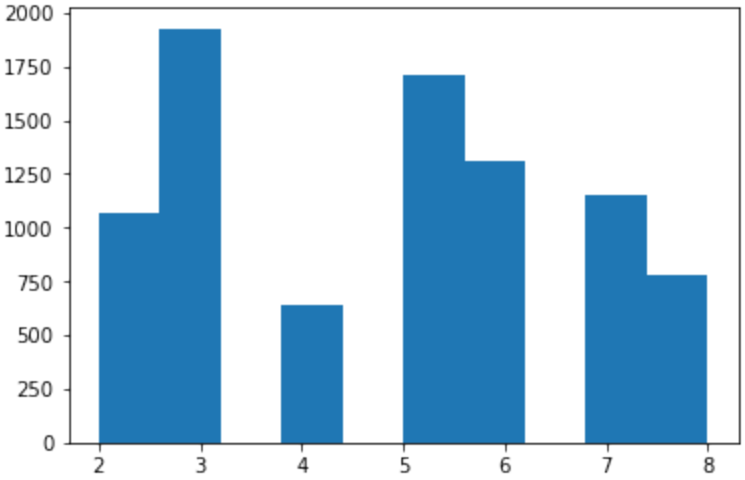
In [24]:
```
plt.hist(df_cust['city'])
```

Out[24]:
```
(array([ 843.,  888.,  717.,  733.,  621.,  651., 1750., 1079.,  477.,
         819.]),
 array([101., 135., 169., 203., 237., 271., 305., 339., 373., 407., 441.]),
 <BarContainer object of 10 artists>)
```



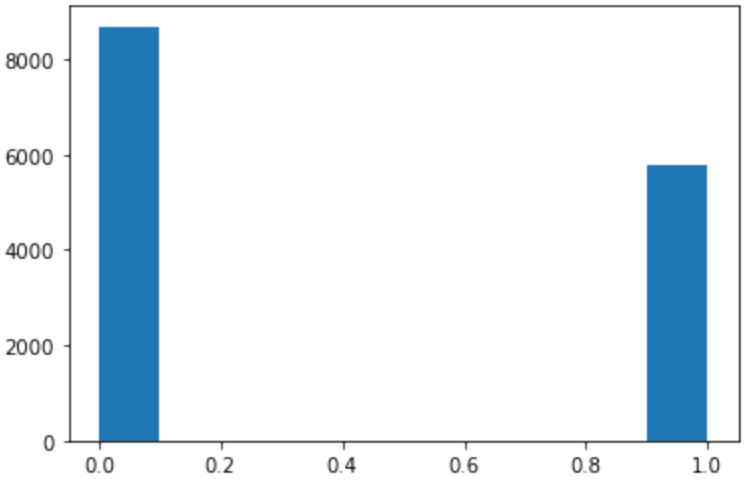
In [25]:
```
plt.hist(df_cust['series_group'])
```

Out[25]:
```
(array([1065., 1926.,    0.,  643.,    0., 1705., 1311.,    0., 1153.,
         775.]),
 array([2. , 2.6, 3.2, 3.8, 4.4, 5. , 5.6, 6.2, 6.8, 7.4, 8. ]),
 <BarContainer object of 10 artists>)
```



In [26]:
```
plt.hist(df_cust['e_section'])
```

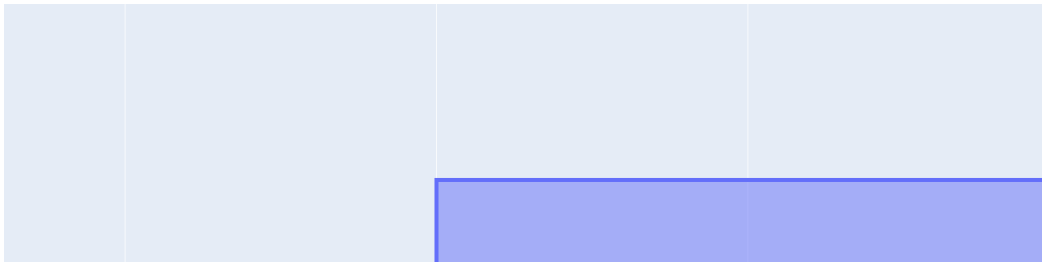
Out[26]:
```
(array([8685.,    0.,    0.,    0.,    0.,    0.,    0.,    0.,    0.,
        5765.]),
 array([0. , 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1. ]),
 <BarContainer object of 10 artists>)
```



In [27]:
```
import plotly.express as px
fig = px.pie(df_cust, values='age', names='age', title = "Potential Customer Age Distr
fig.show()
```

In [28]:
```
fig = px.box(df_cust, x="age", title = "Potential Customer Age Distribution (Boxplot)"
fig.show()
```

Potential Customer Age Distribution (Boxplot)



In [29]:
```
import plotly.express as px
fig = px.pie(df_cust, values='residence', names='residence', title = "Potential Custom
fig.show()
```

In [30]:
```
import plotly.express as px
fig = px.pie(df_cust, values='city', names='city')
fig.show()
```

In [52]:
```
merged['e_section'].value_counts()
```

Out[52]:
```
1    109898
0     70225
Name: e_section, dtype: int64
```

In [53]:
```
df_cust['e_section'].value_counts()
```

Out[53]:
```
0    8685
1    5765
Name: e_section, dtype: int64
```

In [61]:
```
value_counts = df_cust['e_section'].value_counts()
count_0 = value_counts.get(0, 0)
count_1 = value_counts.get(1, 0)

labels = ['0', '1']
values = [count_0, count_1]

fig = px.pie(values=values, names=labels, title='Distribution of content preferences a
fig.show()
```

Distribution of content preferences among Potential Customer

In [32]:
```
df_cust.columns
```

Out[32]:
```
Index(['log_id', 'label_x', 'user_id', 'age', 'gender', 'residence', 'city',
       'city_rank', 'series_dev', 'series_group', 'emui_dev', 'device_name',
       'device_size', 'net_type', 'task_id', 'adv_id', 'creat_type_cd',
       'adv_prim_id', 'inter_type_cd', 'slot_id', 'site_id', 'spread_app_id',
       'hispace_app_tags', 'app_second_class', 'app_score',
       'ad_click_list_v001', 'ad_click_list_v002', 'ad_click_list_v003',
       'ad_close_list_v001', 'ad_close_list_v002', 'ad_close_list_v003',
       'pt_d', 'u_newsCatInterestsST_x', 'u_refreshTimes_x',
       'u_feedLifeCycle_x', 'u_phonePrice', 'u_browserLifeCycle',
       'u_browserMode', 'u_feedLifeCycle_y', 'u_refreshTimes_y',
       'u_newsCatInterests', 'u_newsCatDislike', 'u_newsCatInterestsST_y',
       'u_click_ca2_news', 'i_docId', 'i_s_sourceId', 'i_regionEntity',
       'i_cat', 'i_entities', 'i_dislikeTimes', 'i_upTimes', 'i_dtype', 'e_ch',
       'e_m', 'e_po', 'e_pl', 'e_rn', 'e_section', 'e_et', 'label_y',
       'cillabel', 'pro'],
      dtype='object')
```

In [33]:
```
import plotly.express as px
fig = px.pie(df_cust, values='series_dev', names='series_dev', title = "Potential Cust
fig.show()
```

In [34]:
```
import plotly.express as px
fig = px.pie(df_cust, values='series_group', names='series_group', title = "Potential
fig.show()
```

In [35]:
```
import plotly.express as px
fig = px.pie(df_cust, values='emui_dev', names='emui_dev', title = "Potential Customer
fig.show()
```

In [36]:
```
import plotly.express as px
fig = px.pie(df_cust, values='device_name', names='device_name', title = "Potential Cu
fig.show()
```

Potential Customer Device Name Distribution

In [37]:
```
import plotly.express as px
fig = px.pie(df_cust, values='device_size', names='device_size', title = "Potential Cu
fig.show()
```

Potential Customer Device Size Distribution

In [38]:
```
df_cust.columns
```

Out[38]:
```
Index(['log_id', 'label_x', 'user_id', 'age', 'gender', 'residence', 'city',
       'city_rank', 'series_dev', 'series_group', 'emui_dev', 'device_name',
       'device_size', 'net_type', 'task_id', 'adv_id', 'creat_type_cd',
       'adv_prim_id', 'inter_type_cd', 'slot_id', 'site_id', 'spread_app_id',
       'hispace_app_tags', 'app_second_class', 'app_score',
       'ad_click_list_v001', 'ad_click_list_v002', 'ad_click_list_v003',
       'ad_close_list_v001', 'ad_close_list_v002', 'ad_close_list_v003',
       'pt_d', 'u_newsCatInterestsST_x', 'u_refreshTimes_x',
       'u_feedLifeCycle_x', 'u_phonePrice', 'u_browserLifeCycle',
       'u_browserMode', 'u_feedLifeCycle_y', 'u_refreshTimes_y',
       'u_newsCatInterests', 'u_newsCatDislike', 'u_newsCatInterestsST_y',
       'u_click_ca2_news', 'i_docId', 'i_s_sourceId', 'i_regionEntity',
       'i_cat', 'i_entities', 'i_dislikeTimes', 'i_upTimes', 'i_dtype', 'e_ch',
       'e_m', 'e_po', 'e_pl', 'e_rn', 'e_section', 'e_et', 'label_y',
       'cillabel', 'pro'],
      dtype='object')
```

In [39]:
```
df_cust['pt_d'] = pd.to_datetime(df_cust['pt_d'], format='%Y%m%d%H%M')
df_cust['e_et'] = pd.to_datetime(df_cust['e_et'], format='%Y%m%d%H%M')
```

```
C:\Users\anime\AppData\Local\Temp\ipykernel_7156\194800331.py:1: SettingWithCopyWarni
ng:


A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/us
er_guide/indexing.html#returning-a-view-versus-a-copy

C:\Users\anime\AppData\Local\Temp\ipykernel_7156\194800331.py:2: SettingWithCopyWarni
ng:


A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/us
er_guide/indexing.html#returning-a-view-versus-a-copy
```

In [40]:
```
df_cust['ads_hour'] = df_cust['pt_d'].dt.hour
df_cust['feeds_hour'] = df_cust['e_et'].dt.hour
```

```
C:\Users\anime\AppData\Local\Temp\ipykernel_7156\2629954945.py:1: SettingWithCopyWarn
ing:


A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/us
er_guide/indexing.html#returning-a-view-versus-a-copy

C:\Users\anime\AppData\Local\Temp\ipykernel_7156\2629954945.py:2: SettingWithCopyWarn
ing:


A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/us
er_guide/indexing.html#returning-a-view-versus-a-copy
```

In [41]:
```
df_cust['ads_day'] = df_cust['pt_d'].dt.dayofweek
df_cust['feeds_day'] = df_cust['e_et'].dt.dayofweek
```

```
C:\Users\anime\AppData\Local\Temp\ipykernel_7156\1720565557.py:1: SettingWithCopyWarn
ing:


A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/us
er_guide/indexing.html#returning-a-view-versus-a-copy

C:\Users\anime\AppData\Local\Temp\ipykernel_7156\1720565557.py:2: SettingWithCopyWarn
ing:


A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/us
er_guide/indexing.html#returning-a-view-versus-a-copy
```

In [42]:
```
df_cust['ads_dayname'] = df_cust['pt_d'].dt.day_name()
df_cust['feeds_dayname'] = df_cust['e_et'].dt.day_name()
```

```
C:\Users\anime\AppData\Local\Temp\ipykernel_7156\1132003763.py:1: SettingWithCopyWarn
ing:


A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/us
er_guide/indexing.html#returning-a-view-versus-a-copy

C:\Users\anime\AppData\Local\Temp\ipykernel_7156\1132003763.py:2: SettingWithCopyWarn
ing:


A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/us
er_guide/indexing.html#returning-a-view-versus-a-copy
```

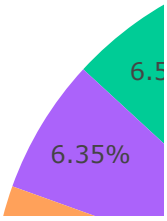
In [43]:
```
df_cust.columns
```

Out[43]:
```
Index(['log_id', 'label_x', 'user_id', 'age', 'gender', 'residence', 'city',
       'city_rank', 'series_dev', 'series_group', 'emui_dev', 'device_name',
       'device_size', 'net_type', 'task_id', 'adv_id', 'creat_type_cd',
       'adv_prim_id', 'inter_type_cd', 'slot_id', 'site_id', 'spread_app_id',
       'hispace_app_tags', 'app_second_class', 'app_score',
       'ad_click_list_v001', 'ad_click_list_v002', 'ad_click_list_v003',
       'ad_close_list_v001', 'ad_close_list_v002', 'ad_close_list_v003',
       'pt_d', 'u_newsCatInterestsST_x', 'u_refreshTimes_x',
       'u_feedLifeCycle_x', 'u_phonePrice', 'u_browserLifeCycle',
       'u_browserMode', 'u_feedLifeCycle_y', 'u_refreshTimes_y',
       'u_newsCatInterests', 'u_newsCatDislike', 'u_newsCatInterestsST_y',
       'u_click_ca2_news', 'i_docId', 'i_s_sourceId', 'i_regionEntity',
       'i_cat', 'i_entities', 'i_dislikeTimes', 'i_upTimes', 'i_dtype', 'e_ch',
       'e_m', 'e_po', 'e_pl', 'e_rn', 'e_section', 'e_et', 'label_y',
       'cillabel', 'pro', 'ads_hour', 'feeds_hour', 'ads_day', 'feeds_day',
       'ads_dayname', 'feeds_dayname'],
      dtype='object')
```

In [44]:
```
import plotly.express as px
fig = px.pie(df_cust, values='ads_hour', names='ads_hour', title = "Potential Customer
fig.show()
```

In [68]:
```
import plotly.express as px
fig = px.pie(df_cust, values= df_cust['feeds_hour'].value_counts().values, names=df_cu
fig.show()
```

Potential Customer Feeds Hour Viewed Distribution



In [64]:
```
import plotly.express as px
fig = px.pie(df_cust, values= df_cust['ads_day'].value_counts().values, names=df_cust[
fig.show()
```

In [59]:
```
value_counts = df_cust['e_section'].value_counts()
count_0 = value_counts.get(0, 0)
count_1 = value_counts.get(1, 0)

labels = ['0', '1']
values = [count_0, count_1]

fig = px.pie(values=values, names=labels, title='Distribution of 0s and 1s')
fig.show()
```

```
1    14450
Name: label_y, dtype: int64
14450
```

In [47]:
```
df_cust['feeds_day']
```

Out[47]:
```
18        4
21        2
23        2
27        4
29        4
         ..
179951    6
179959    6
179996    6
179997    6
180013    6
Name: feeds_day, Length: 14450, dtype: int64
```

In [67]:
```
import plotly.express as px
fig = px.pie(df_cust, values= df_cust['feeds_day'].value_counts().values, names=df_cus
fig.show()
```

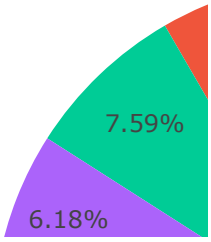
In [50]:
```
df_noncust = merged[merged['label_y'] == -1.0]
```

In [51]:
```
merged.to_csv("merged_dataframe.csv")
```

In [70]:
```
import plotly.express as px
fig = px.pie(df_noncust, values='age', names='age', title = "Non-Potential Customer Ag
fig.show()
```

In [74]:
```
import plotly.express as px
fig = px.pie(df_noncust, values='residence', names='residence', title = "Non-Potential
fig.show()
```

Non-Potential Customer Residence Distribution



In [73]:
```
value_counts = df_noncust['e_section'].value_counts()
count_0 = value_counts.get(0, 0)
count_1 = value_counts.get(1, 0)

labels = ['0', '1']
values = [count_0, count_1]

fig = px.pie(values=values, names=labels, title='Distribution of content preferences a
fig.show()
```

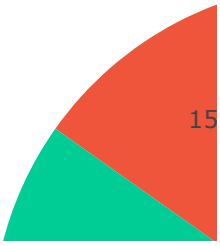
Distribution of content preferences among Non-Potential Custo

In [75]:
```
import plotly.express as px
fig = px.pie(df_noncust, values='series_group', names='series_group', title = "Non-Pot
fig.show()
```

Non-Potential Customer Device Series Group Distribution

In [76]:
```
import plotly.express as px
fig = px.pie(df_noncust, values='series_dev', names='series_dev', title = "Non-Potenti
fig.show()
```

In [77]:
```
import plotly.express as px
fig = px.pie(df_noncust, values='emui_dev', names='emui_dev', title = "Non-Potential C
fig.show()
```



In [78]:
```
df_noncust['pt_d'] = pd.to_datetime(df_noncust['pt_d'], format='%Y%m%d%H%M')
df_noncust['e_et'] = pd.to_datetime(df_noncust['e_et'], format='%Y%m%d%H%M')
```

```
C:\Users\anime\AppData\Local\Temp\ipykernel_7156\372196138.py:1: SettingWithCopyWarni
ng:


A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/us
er_guide/indexing.html#returning-a-view-versus-a-copy

C:\Users\anime\AppData\Local\Temp\ipykernel_7156\372196138.py:2: SettingWithCopyWarni
ng:


A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/us
er_guide/indexing.html#returning-a-view-versus-a-copy
```

In [79]:
```
df_noncust['ads_hour'] = df_noncust['pt_d'].dt.hour
df_noncust['feeds_hour'] = df_noncust['e_et'].dt.hour
```

```
C:\Users\anime\AppData\Local\Temp\ipykernel_7156\137544654.py:1: SettingWithCopyWarni
ng:


A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/us
er_guide/indexing.html#returning-a-view-versus-a-copy

C:\Users\anime\AppData\Local\Temp\ipykernel_7156\137544654.py:2: SettingWithCopyWarni
ng:


A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/us
er_guide/indexing.html#returning-a-view-versus-a-copy
```

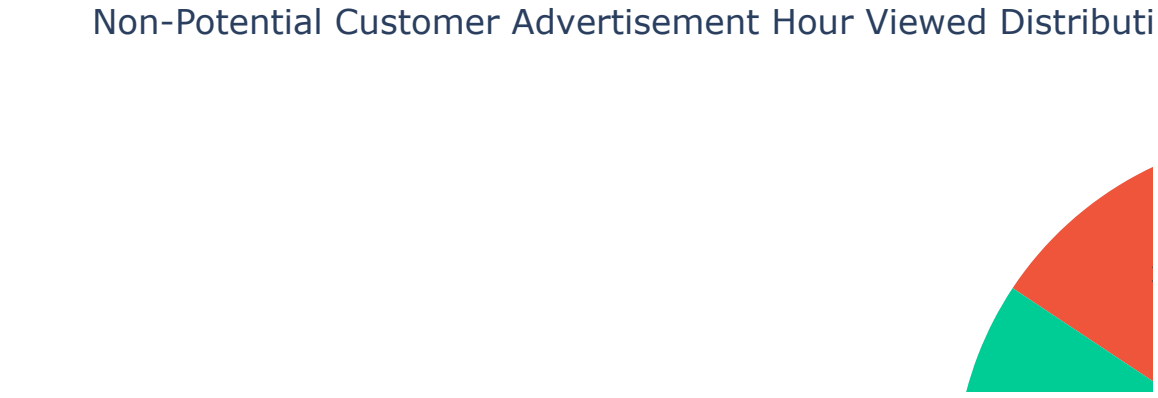
In [80]:
```
df_noncust['ads_day'] = df_noncust['pt_d'].dt.dayofweek
df_noncust['feeds_day'] = df_noncust['e_et'].dt.dayofweek
```

```
C:\Users\anime\AppData\Local\Temp\ipykernel_7156\1650693043.py:1: SettingWithCopyWarn
ing:


A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/us
er_guide/indexing.html#returning-a-view-versus-a-copy

C:\Users\anime\AppData\Local\Temp\ipykernel_7156\1650693043.py:2: SettingWithCopyWarn
ing:


A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/us
er_guide/indexing.html#returning-a-view-versus-a-copy
```

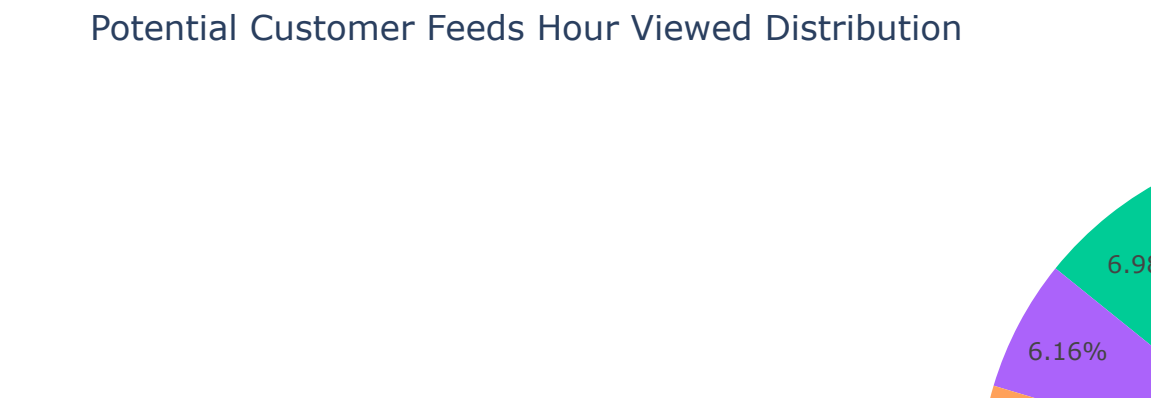
In [81]:
```
df_noncust['ads_dayname'] = df_noncust['pt_d'].dt.day_name()
df_noncust['feeds_dayname'] = df_noncust['e_et'].dt.day_name()
```

```
C:\Users\anime\AppData\Local\Temp\ipykernel_7156\4293718302.py:1: SettingWithCopyWarn
ing:


A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/us
er_guide/indexing.html#returning-a-view-versus-a-copy

C:\Users\anime\AppData\Local\Temp\ipykernel_7156\4293718302.py:2: SettingWithCopyWarn
ing:


A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/us
er_guide/indexing.html#returning-a-view-versus-a-copy
```

In [82]:
```
import plotly.express as px
fig = px.pie(df_noncust, values='ads_hour', names='ads_hour', title = "Non-Potential C
fig.show()
```



In [83]:
```
import plotly.express as px
fig = px.pie(df_noncust, values= df_noncust['feeds_hour'].value_counts().values, names
fig.show()
```



In [84]:
```
import plotly.express as px
fig = px.pie(df_noncust, values= df_noncust['ads_day'].value_counts().values, names=df
fig.show()
```

In [85]:
```
import plotly.express as px
fig = px.pie(df_noncust, values= df_noncust['feeds_day'].value_counts().values, names=
fig.show()
```

In [86]:
```
df_cust.to_csv("customer_df.csv")
```

In [87]:
```
df_noncust.to_csv("noncustomer_df.csv")
```

In [ ]:
```

```