In [1]:
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
```

Задание 1

In [2]:
```
df = pd.read_csv('frodo_users.csv')
```

In [3]:
```
compromised_funnels = df['count']
```

In [4]:
```
plt.hist(compromised_funnels, bins=5, color='skyblue', edgecolor='black')
plt.title('Distribution of Compromised Funnel Stages')
plt.xlabel('Number of Compromised Stages')
plt.ylabel('Frequency')
plt.show()
```

Задание 2

In [5]:
```
df = pd.read_csv('data.csv')
df = df.replace(r'^\s*$', np.nan, regex=True)
df = df.drop(columns=['Unnamed: 0'])
df
```

Out[5]:

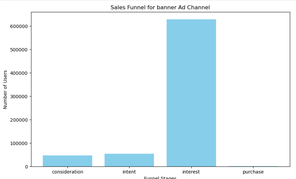
	user_id	funnel_stage	timestamp	profit	os	ad_channel	price
0	512217	interest	1500000	NaN	iOS	video	0.75
1	587720	interest	1500000	NaN	Android	banner	0.50
2	810648	interest	1500000	NaN	Android	video	0.75
3	47061	interest	1500000	NaN	Android	banner	0.50
4	1035593	interest	1500000	NaN	Android	partner_network	1.00
...	...	...	...	...	...	...	...
2737569	789473	purchase	201912	2136.0	Android	partner_network	1.00
2737570	1038000	purchase	202012	50.0	Android	traffic	0.00
2737571	539693	purchase	202053	50.0	iOS	traffic	0.00
2737572	583326	purchase	202140	1099.0	Android	traffic	0.00
2737573	590392	purchase	202152	2136.0	iOS	traffic	0.00

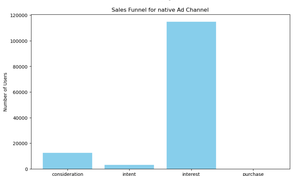
2737574 rows × 7 columns

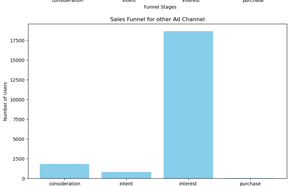
In [6]:
```
funnel_data = df.groupby(['ad_channel', 'funnel_stage'])['user_id'].count().reset_index()
```

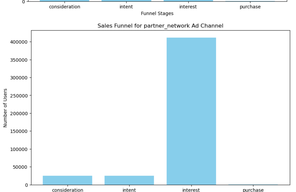
In [7]:
```
pivot_table = funnel_data.pivot(index='ad_channel', columns='funnel_stage', values='user_id').fillna(0)
```

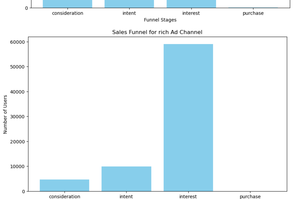
In [8]:
```
for channel in pivot_table.index:
    stages = pivot_table.loc[channel]
    plt.figure(figsize=(10, 6))
    plt.bar(stages.index, stages.values, color='skyblue')
    plt.title(f'Sales Funnel for {channel} Ad Channel')
    plt.xlabel('Funnel Stages')
    plt.ylabel('Number of Users')
    plt.show()
```

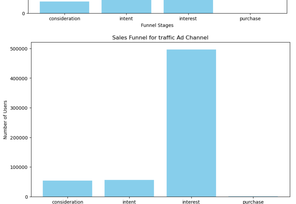












Задание 3

In [9]:
```
channel_data = pd.read_csv('data.csv')
channel_data = channel_data.replace(r'^\s*$', np.nan, regex=True)
channel_data = channel_data.drop(columns=['Unnamed: 0'])
channel_data
```

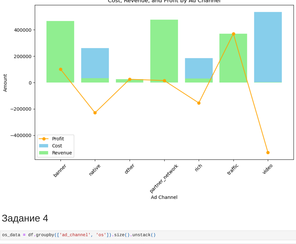
Out[9]:

	user_id	funnel_stage	timestamp	profit	os	ad_channel	price
0	512217	interest	1500000	NaN	iOS	video	0.75
1	587720	interest	1500000	NaN	Android	banner	0.50
2	810648	interest	1500000	NaN	Android	video	0.75
3	47061	interest	1500000	NaN	Android	banner	0.50
4	1035593	interest	1500000	NaN	Android	partner_network	1.00
...	...	...	...	...	...	...	...
2737569	789473	purchase	201912	2136.0	Android	partner_network	1.00
2737570	1038000	purchase	202012	50.0	Android	traffic	0.00
2737571	539693	purchase	202053	50.0	iOS	traffic	0.00
2737572	583326	purchase	202140	1099.0	Android	traffic	0.00
2737573	590392	purchase	202152	2136.0	iOS	traffic	0.00

2737574 rows × 7 columns

In [10]:
```
channel_data = df.groupby('ad_channel').agg({'profit': 'sum', 'price': 'sum'}).reset_index()
```

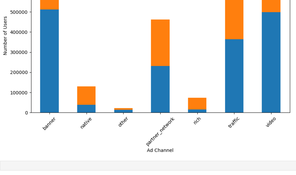
In [11]:
```
plt.figure(figsize=(10, 6))
plt.bar(channel_data['ad_channel'], channel_data['price'], label='Cost', color='skyblue')
plt.bar(channel_data['ad_channel'], channel_data['profit'] - channel_data['price'], bottom='aggregate')
plt.plot(channel_data['ad_channel'], channel_data['profit'] - channel_data['price'], marker='o', linestyle='-', color='orange', label='Profit')
plt.title('Cost, Revenue, and Profit by Ad Channel')
plt.xlabel('Ad Channel')
plt.ylabel('Amount')
plt.legend()
plt.xticks(rotation=45)
plt.show()
```



Задание 4

In [12]:
```
os_data = df.groupby(['ad_channel', 'os']).size().unstack()
```

In [13]:
```
os_data.plot(kind='bar', stacked=True, figsize=(10, 6))
plt.title('Distribution of User Operating Systems by Ad Channel')
plt.xlabel('Ad Channel')
plt.ylabel('Number of Users')
plt.xticks(rotation=45)
plt.show()
```



In [ ]:
```

```