The Data Ascent: Discovering Trends in Himalayan Mountaineering ¶

Exploring factors in expedition success and failure using Elizabeth Hawley's archives: The Himalayan Database

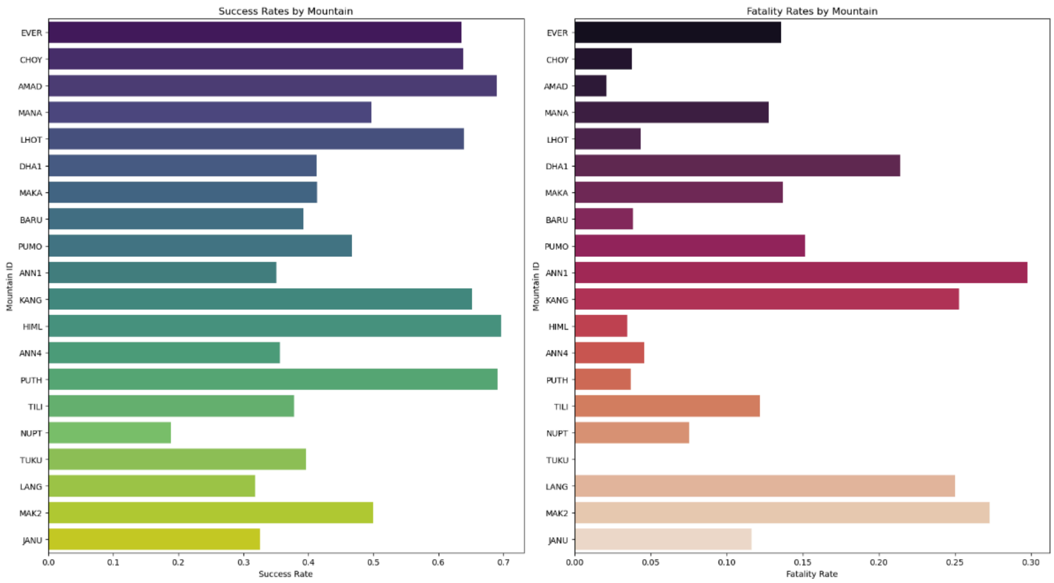
Section 1: Success and Fatality Rates on the various Mountains

Section 1: Success and Fatality Rates on the various Mountains

In [2]:

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

expeditions = pd.read_csv('cleaned_expeditions_data.csv')

expedition_aggregates = expeditions.groupby('peak_id').agg({
    'success1': 'sum',
    'mdeaths': 'sum',
    'hdeaths': 'sum',
    'exp_id': 'size'
}).reset_index()


expedition_aggregates.rename(columns={'exp_id': 'total_expeditions'}, inplace=True)


expedition_aggregates['total_deaths'] = expedition_aggregates['mdeaths'] + expedition_aggregates['hdeaths']


expedition_aggregates['success_rate'] = expedition_aggregates['success1'] / expedition_aggregates['total_expeditions']
expedition_aggregates['fatality_rate'] = expedition_aggregates['total_deaths'] / expedition_aggregates['total_expeditions']


top_mountains = expedition_aggregates.sort_values('total_expeditions', ascending=False).head(20)


plt.figure(figsize=(18, 10))


plt.subplot(1, 2, 1)
sns.barplot(x='success_rate', y='peak_id', data=top_mountains, palette='viridis')
plt.title('Success Rates by Mountain')
plt.xlabel('Success Rate')
plt.ylabel('Mountain ID')
```

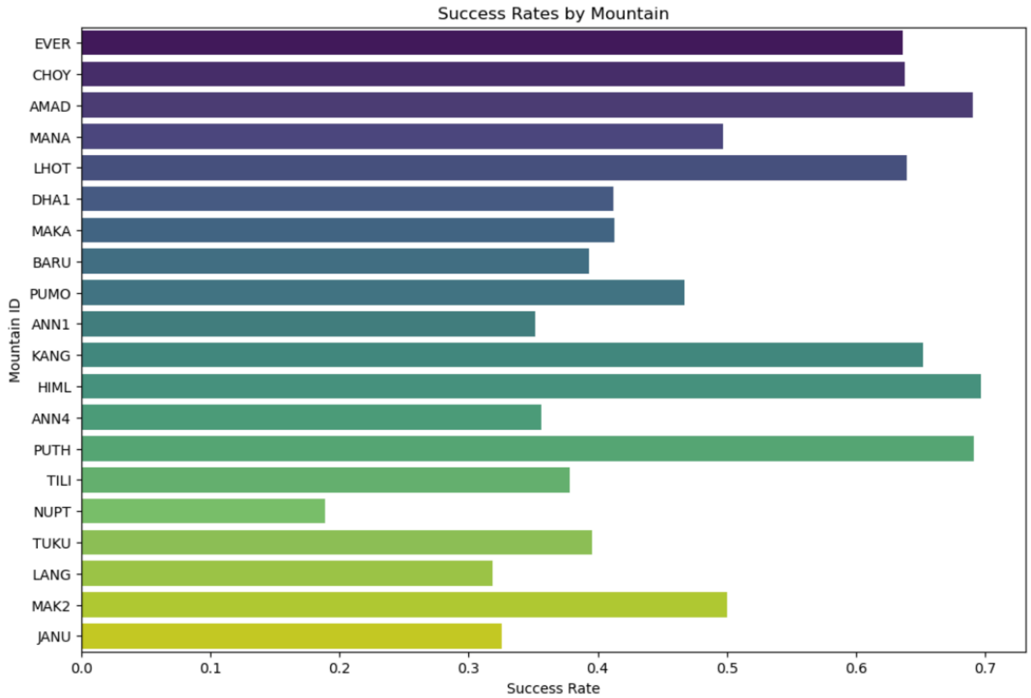
```python
plt.subplot(1, 2, 2)
sns.barplot(x='fatality_rate', y='peak_id', data=top_mountains, palette='rocket')
plt.title('Fatality Rates by Mountain')
plt.xlabel('Fatality Rate')
plt.ylabel('Mountain ID')

plt.tight_layout()
plt.show()
```

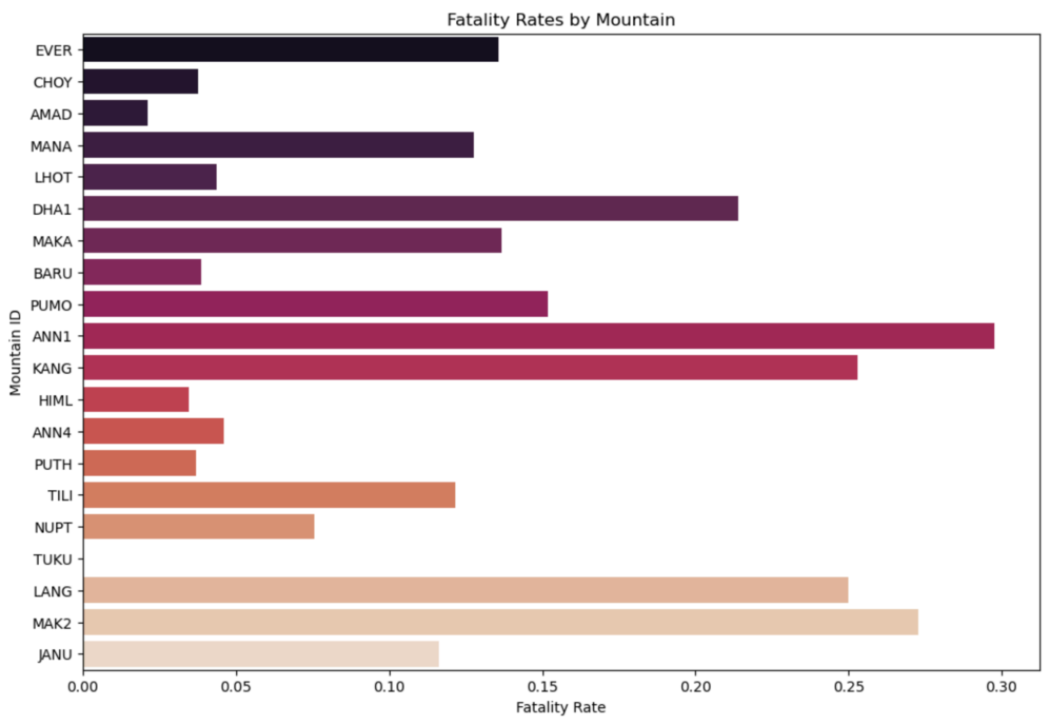


In [20]:

```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Load data
expeditions = pd.read_csv('cleaned_expeditions_data.csv')

# Calculate success and death counts
expedition_aggregates = expeditions.groupby('peak_id').agg({
    'success1': 'sum',  # Assuming 'success1' indicates a successful expedition
    'mdeaths': 'sum',
    'hdeaths': 'sum'
}).reset_index()

# Calculating the total deaths
expedition_aggregates['total_deaths'] = expedition_aggregates['mdeaths'] + expedition_aggregates['hdeaths']

# Calculate total expeditions correctly
total_expeditions = expeditions.groupby('peak_id').size().reset_index(name='total_expeditions')
expedition_aggregates = pd.merge(expedition_aggregates, total_expeditions, on='peak_id')

# Calculate success and fatality rates
expedition_aggregates['success_rate'] = expedition_aggregates['success1'] / expedition_aggregates['total_expeditions']
expedition_aggregates['fatality_rate'] = expedition_aggregates['total_deaths'] / expedition_aggregates['total_expeditions']

# Filter to top 20 mountains based on total expeditions
top_mountains = expedition_aggregates.sort_values('total_expeditions', ascending=False).head(20)



# Visualizing the data
plt.figure(figsize=(12, 8))

# Success rates plot
sns.barplot(x='success_rate', y='peak_id', data=top_mountains, palette='viridis')
plt.title('Success Rates by Mountain')
plt.xlabel('Success Rate')
plt.ylabel('Mountain ID')
plt.show()
```

```python
# Fatality rates plot
plt.figure(figsize=(12, 8))
sns.barplot(x='fatality_rate', y='peak_id', data=top_mountains, palette='rocket')
plt.title('Fatality Rates by Mountain')
plt.xlabel('Fatality Rate')
plt.ylabel('Mountain ID')
plt.show()
```





In the comprehensive analysis of success and fatality rates across the various mountain expeditions, we can see distinct patterns emerging with such a high number of expeditions with success rates varying significantly among different mountains. Everest and Cho Oyo offer the highest success rates, which is connected to how popular they are. This is because Everest is the mountain that has the most tourism every year and so the expedition experience is a lot more advanced because it has been done so many times and there are well established routes. Cho Oyu is also not surprising as this is a very popular mountain in the training before climbing Everest due to its unique position of still being an 8000m peak (thus giving you a flavor of what the altitude effects are like) whilst having relatively easier terrain compared to other 8000m peaks.

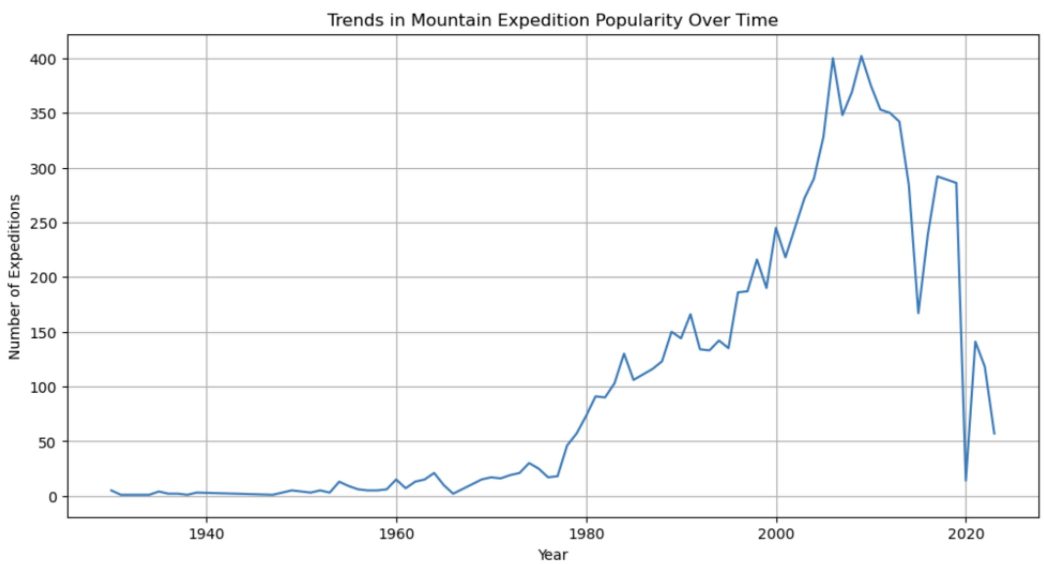
On the flip side, this analysis also reveals some of the more sobering realities of expeditions. Some mountains have relatively low fatality rates however some mountains like Kangchenjunga and Annapurna show high rates indicating the extreme challenges and severe conditions encountered. However some mountains like Tuku show no fatalities on records which indicates either fewer expeditions or safer climbing conditions and routes.

These graphs help us understand the different dynamics of mountain expeditions suggesting that while some mountains offer higher chances of summiting, they also carry a proportionally higher risk of death. This data could be critical for future climbers who are preparing challenges.

Section 2: Changes in popularity over time

In [23]:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
expeditions = pd.read_csv('cleaned_expeditions_data.csv')

yearly_trends = expeditions['year'].value_counts().sort_index()
plt.figure(figsize=(12, 6))
sns.lineplot(x=yearly_trends.index, y=yearly_trends.values)
plt.title('Trends in Mountain Expedition Popularity Over Time')
plt.xlabel('Year')
plt.ylabel('Number of Expeditions')
plt.grid(True)
plt.show()
```

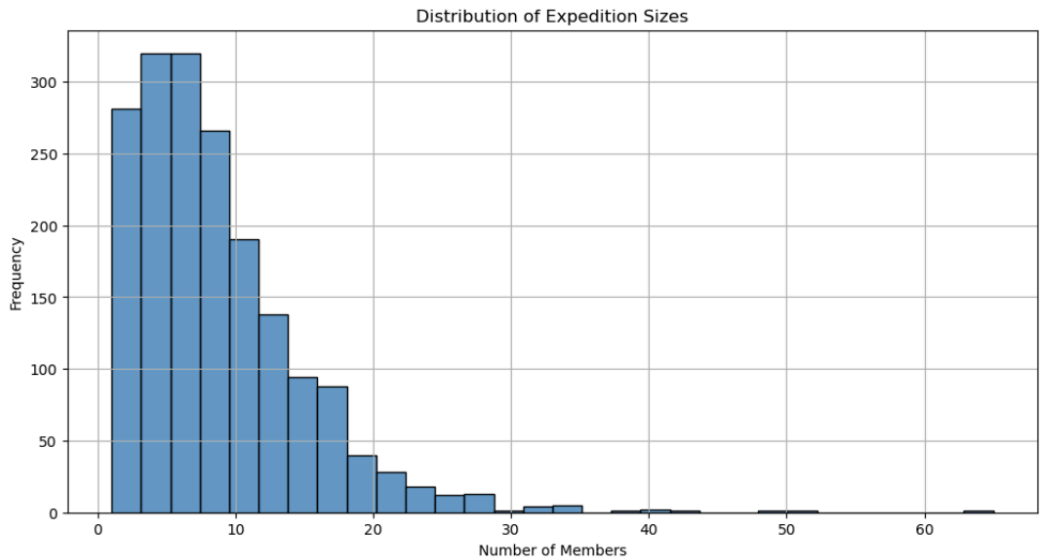


The historical trend in mountain expeditions visualised by this graph can show us a compelling narrative of the changing interests in high altitude climbing over time. From the 40s-70s the number of expeditions was low and stable reflecting a niche activity where only a handful of enthusiasts and professionals ventured into high altitude expeditions. But from the 80s we see a large surge that continues into the early noughties and this can be attributed to several factors such as advancements in climbing tech/gear, increased coverage by the media and a growing global interest in these activities.

However there is a striking decline post 2010 and particularly around 2020 which corresponds to global economic stresses and the unprecedented impact of Covid which restricted travel and expeditions. This trend analysis highlights the fluctuating popularity in mountaineering and how global external events can significantly influence engagement in this sport. I would suggest that the resilience in climbing interest post pandemic suggests a resilient interest in climbing and points toward future resurgence as conditions normalise.

In [24]:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
members = pd.read_csv('cleaned_members_data.csv')

members_count = members.groupby('expid').size().reset_index(name='num_members')
plt.figure(figsize=(12, 6))
sns.histplot(members_count['num_members'], bins=30, kde=False)
plt.title('Distribution of Expedition Sizes')
plt.xlabel('Number of Members')
plt.ylabel('Frequency')
plt.grid(True)
plt.show()
```



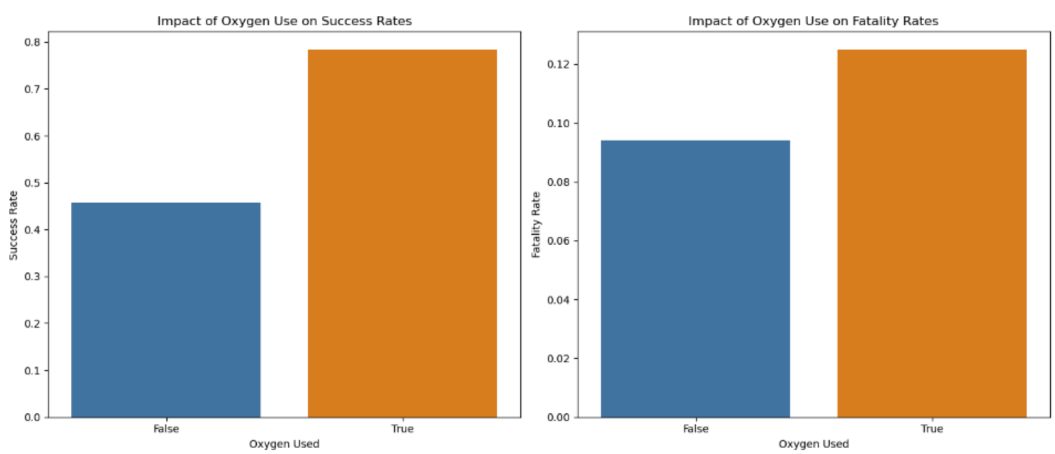
This histogram shows the distribution of expedition sizes and reveals that the majority of expeditions are small consisting of 5 to 15 members. This reflects a balance being struck between logistical manageability and a need for sufficient manpower to handle the challenges of high altitude climbing, such as carrying group kit, so if there are more members in the group it's easier to spread out. Then around the 20 member mark we see a rapid decline which suggests that while larger groups benefit from greater resources and capabilities, the complexity of managing such groups make them unfeasible.

Section 3: Oxygen Use and Outcomes

In [5]:

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


expeditions = pd.read_csv('cleaned_expeditions_data.csv')

print(expeditions.columns)


oxygen_stats = expeditions.groupby('o2used').agg({
    'success1': 'mean',
    'mdeaths': 'sum',
    'hdeaths': 'sum',
    'exp_id': 'count'
}).reset_index()


oxygen_stats['total_deaths'] = oxygen_stats['mdeaths'] + oxygen_stats['hdeaths']

oxygen_stats['death_rate'] = oxygen_stats['total_deaths'] / oxygen_stats['exp_id']

plt.figure(figsize=(14, 6))

plt.subplot(1, 2, 1)
sns.barplot(x='o2used', y='success1', data=oxygen_stats)
plt.title('Impact of Oxygen Use on Success Rates')
plt.xlabel('Oxygen Used')
plt.ylabel('Success Rate')

plt.subplot(1, 2, 2)
sns.barplot(x='o2used', y='death_rate', data=oxygen_stats)
plt.title('Impact of Oxygen Use on Fatality Rates')
plt.xlabel('Oxygen Used')
plt.ylabel('Fatality Rate')

plt.tight_layout()
plt.show()
```

```
Index(['exp_id', 'peak_id', 'year', 'season', 'host_country', 'route1',
       'route2', 'route3', 'nation', 'leaders', 'sponsor', 'success1',
       'success2', 'success3', 'success4', 'claimed', 'disputed', 'countries',
       'approach', 'bcdate', 'smtdate', 'smttime', 'smtdays', 'totdays',
       'termdate', 'termreason', 'termnote', 'highpoint', 'traverse', 'ski',
       'parapente', 'camps', 'rope', 'totmembers', 'smtmembers', 'mdeaths',
       'tothired', 'smthired', 'hdeaths', 'nohired', 'o2used', 'o2none',
       'o2climb', 'o2descent', 'o2sleep', 'o2medical', 'o2taken', 'o2unkwn',
       'othersmts', 'campsites', 'accidents', 'achievment', 'agency', 'comrte',
       'stdrte', 'primrte', 'primmem', 'primref', 'primid', 'chksum', 'pkname',
       'heightm'],
      dtype='object')
```



Oxygen Used
Oxygen Used

In [29]:

```python
# Print the column names of the expeditions DataFrame
expeditions = pd.read_csv('cleaned_expeditions_data.csv')
print(expeditions.columns)
```

```
Index(['exp_id', 'peak_id', 'year', 'season', 'host_country', 'route1',
       'route2', 'route3', 'nation', 'leaders', 'sponsor', 'success1',
       'success2', 'success3', 'success4', 'claimed', 'disputed', 'countries',
       'approach', 'bcdate', 'smtdate', 'smttime', 'smtdays', 'totdays',
       'termdate', 'termreason', 'termnote', 'highpoint', 'traverse', 'ski',
       'parapente', 'camps', 'rope', 'totmembers', 'smtmembers', 'mdeaths',
       'tothired', 'smthired', 'hdeaths', 'nohired', 'o2used', 'o2none',
       'o2climb', 'o2descent', 'o2sleep', 'o2medical', 'o2taken', 'o2unkwn',
       'othersmts', 'campsites', 'accidents', 'achievment', 'agency', 'comrte',
       'stdrte', 'primrte', 'primmem', 'primref', 'primid', 'chksum', 'pkname',
       'heightm'],
      dtype='object')
```

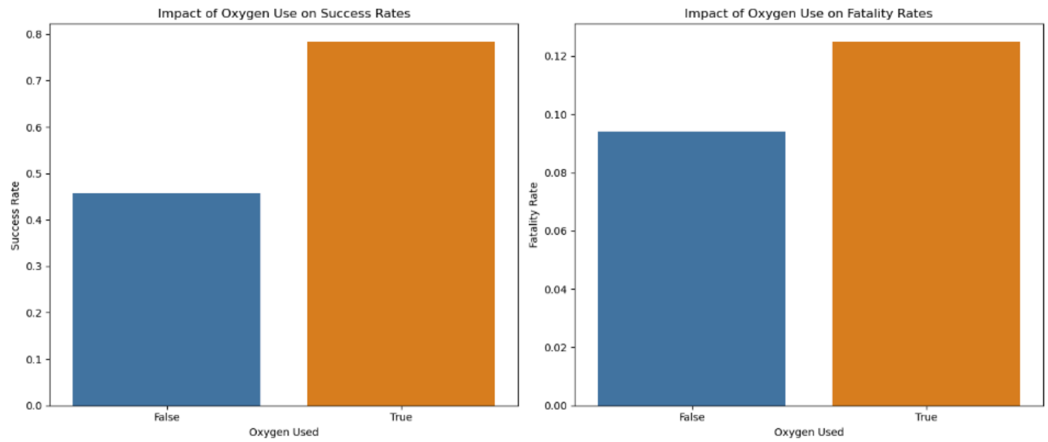
In [38]:

```python
import matplotlib.pyplot as plt
```

In [38]:

```python
import matplotlib.pyplot as plt
import seaborn as sns
expeditions = pd.read_csv('cleaned_expeditions_data.csv')
oxygen_stats = expeditions.groupby('o2used').agg({
    'success1': 'mean',
    'mdeaths': 'sum',
    'hdeaths': 'sum',
    'exp_id': 'count'
}).reset_index()

oxygen_stats = expeditions.groupby('o2used').agg({
    'success1': 'mean',
    'mdeaths': 'sum',
    'hdeaths': 'sum',
}).reset_index()

oxygen_stats['total_expeditions'] = expeditions.groupby('o2used').size().values


oxygen_stats['total_deaths'] = oxygen_stats['mdeaths'] + oxygen_stats['hdeaths']
oxygen_stats['death_rate'] = oxygen_stats['total_deaths'] / oxygen_stats['total_expeditions']


plt.figure(figsize=(14, 6))
plt.subplot(1, 2, 1)
sns.barplot(x='o2used', y='success1', data=oxygen_stats)
plt.title('Impact of Oxygen Use on Success Rates')
plt.xlabel('Oxygen Used')
plt.ylabel('Success Rate')
plt.subplot(1, 2, 2)
sns.barplot(x='o2used', y='death_rate', data=oxygen_stats)
plt.title('Impact of Oxygen Use on Fatality Rates')
plt.xlabel('Oxygen Used')
plt.ylabel('Fatality Rate')

plt.tight_layout()
plt.show()
```

Impact of Oxygen Use on Success Rates
Impact of Oxygen Use on Fatality Rates



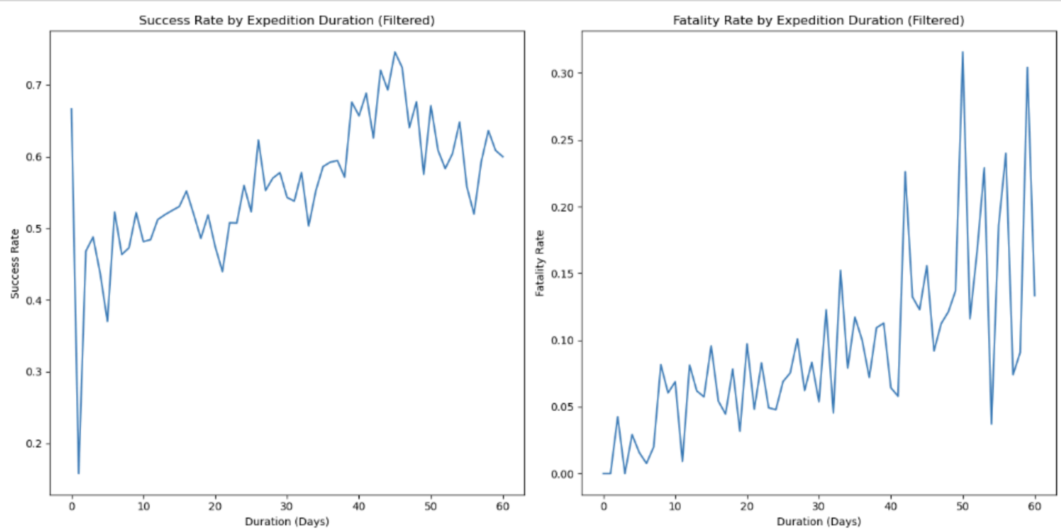
Here we can see that the use of oxygen during expeditions significantly impacts success and fatality rates. Expeditions using oxygen had substantially higher success rates, doubling that of those which did not use supplemental oxygen. Suggesting it enhances the climbers ability to cope in the thin air of the 'death zone' at 8000m where oxygen is 1/3 of that at sea level.

Paradoxically the fatality rates were also higher when using oxygen, this could be due to the fact that expeditions opting for supplemental oxygen may target more challenging routes where the risks are greater. Often expeditions will take oxygen with them but not use it, so it could be that oxygen and fatalities are linked because these expeditions end up using the oxygen because they are in trouble already but it is not enough to prevent fatalities at that point.

In [51]:

```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns


expeditions = pd.read_csv('cleaned_expeditions_data.csv')


expeditions['bcdate'] = pd.to_datetime(expeditions['bcdate'], errors='coerce')
expeditions['termdate'] = pd.to_datetime(expeditions['termdate'], errors='coerce')



expeditions['duration'] = (expeditions['termdate'] - expeditions['bcdate']).dt.days

duration_stats = expeditions.groupby('duration').agg({
    'success1': 'mean',
    'mdeaths': 'sum',
    'hdeaths': 'sum',
    'exp_id': 'count'
}).reset_index()


filtered_expeditions = expeditions[expeditions['duration'] <= 60]

filtered_duration_stats = filtered_expeditions.groupby('duration').agg({
    'success1': 'mean',
    'mdeaths': 'sum',
    'hdeaths': 'sum',
    'exp_id': 'count'
}).reset_index()


filtered_duration_stats['fatality_rate'] = (filtered_duration_stats['mdeaths'] + filtered_duration_stats['hdeaths']) / filter


duration_stats['fatality_rate'] = (duration_stats['mdeaths'] + duration_stats['hdeaths']) / duration_stats['exp_id']

plt.figure(figsize=(14, 7))
```

```python
plt.subplot(1, 2, 1)
sns.lineplot(x='duration', y='success1', data=filtered_duration_stats)
plt.title('Success Rate by Expedition Duration (Filtered)')
plt.xlabel('Duration (Days)')
plt.ylabel('Success Rate')

plt.subplot(1, 2, 2)
sns.lineplot(x='duration', y='fatality_rate', data=filtered_duration_stats)
plt.title('Fatality Rate by Expedition Duration (Filtered)')
plt.xlabel('Duration (Days)')
plt.ylabel('Fatality Rate')

plt.tight_layout()
plt.show()
```



The analysis of expedition duration reveals a complex relationship with success and fatality rates among mountain expeditions. From the filtered data, we observe a fluctuating but generally positive trend in success rates as the duration increases, peaking around mid-duration before a slight decline. This suggests that a moderate amount of time spent on expeditions, likely allowing for adequate acclimatization and preparation, correlates with higher chances of success. However, success rates show variability, indicating other factors, such as changing weather conditions, expedition experience, and mountain difficulty, also play significant roles.

In contrast, the fatality rates exhibit a less predictable pattern, with several peaks throughout the duration range. This volatility could reflect the inherent risks associated with particular days or phases within expeditions, such as crossing difficult terrains or summit attempts, which are not solely dependent on the overall duration but rather specific circumstances encountered during those times.

These insights emphasize the importance of strategic planning and risk management in expedition scheduling. They suggest that while longer durations can enhance success by allowing climbers to acclimatize and prepare adequately, they also require careful consideration of the risks involved at different stages of the expedition.

In [ ]:

```
Git hub: https://github.com/mattwj1/Him_Database
```