

```
In [2]: import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
```

```
In [3]: df = pd.read_csv("C:/Users/Siddharth Singh/Downloads/uber.csv")
```

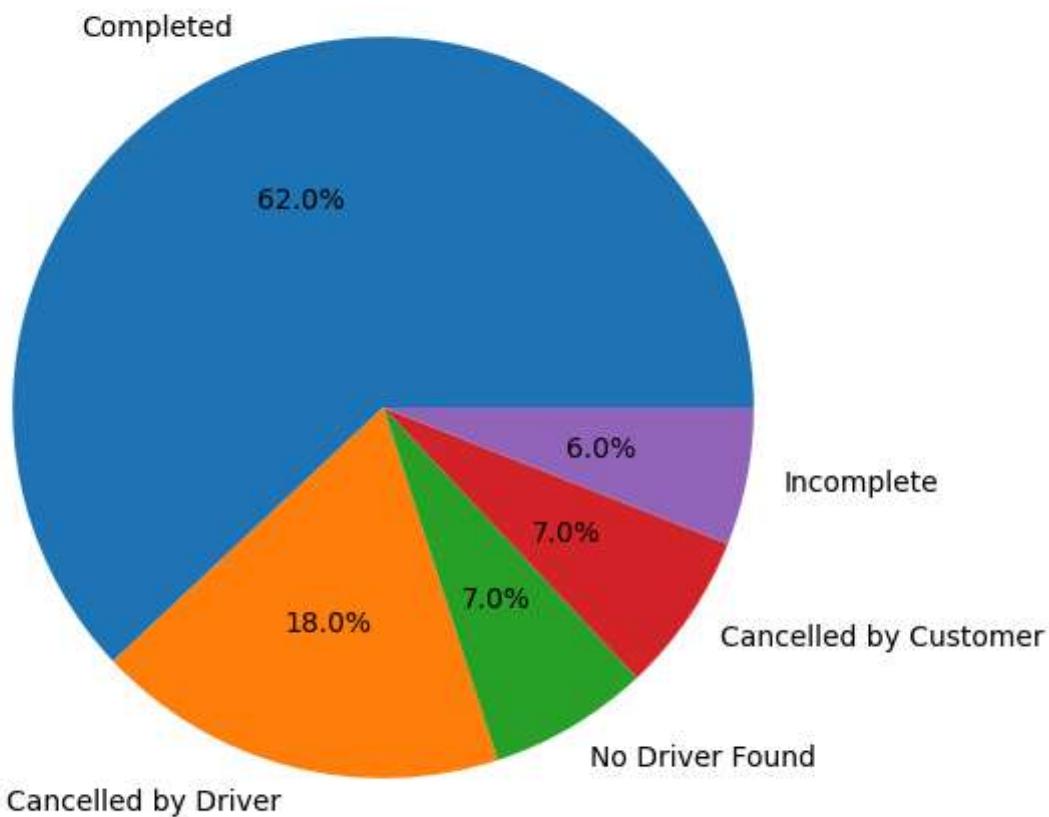
```
In [7]: total_bookings=len(df)
print("Total Bookings:",total_bookings)
```

Total Bookings: 150000

```
In [12]: status_counts = df['Booking Status'].value_counts()

plt.figure(figsize=(6,6))
plt.pie(status_counts, labels=status_counts.index, autopct='%1.1f%%')
plt.title("Booking Status Distribution")
plt.show()
```

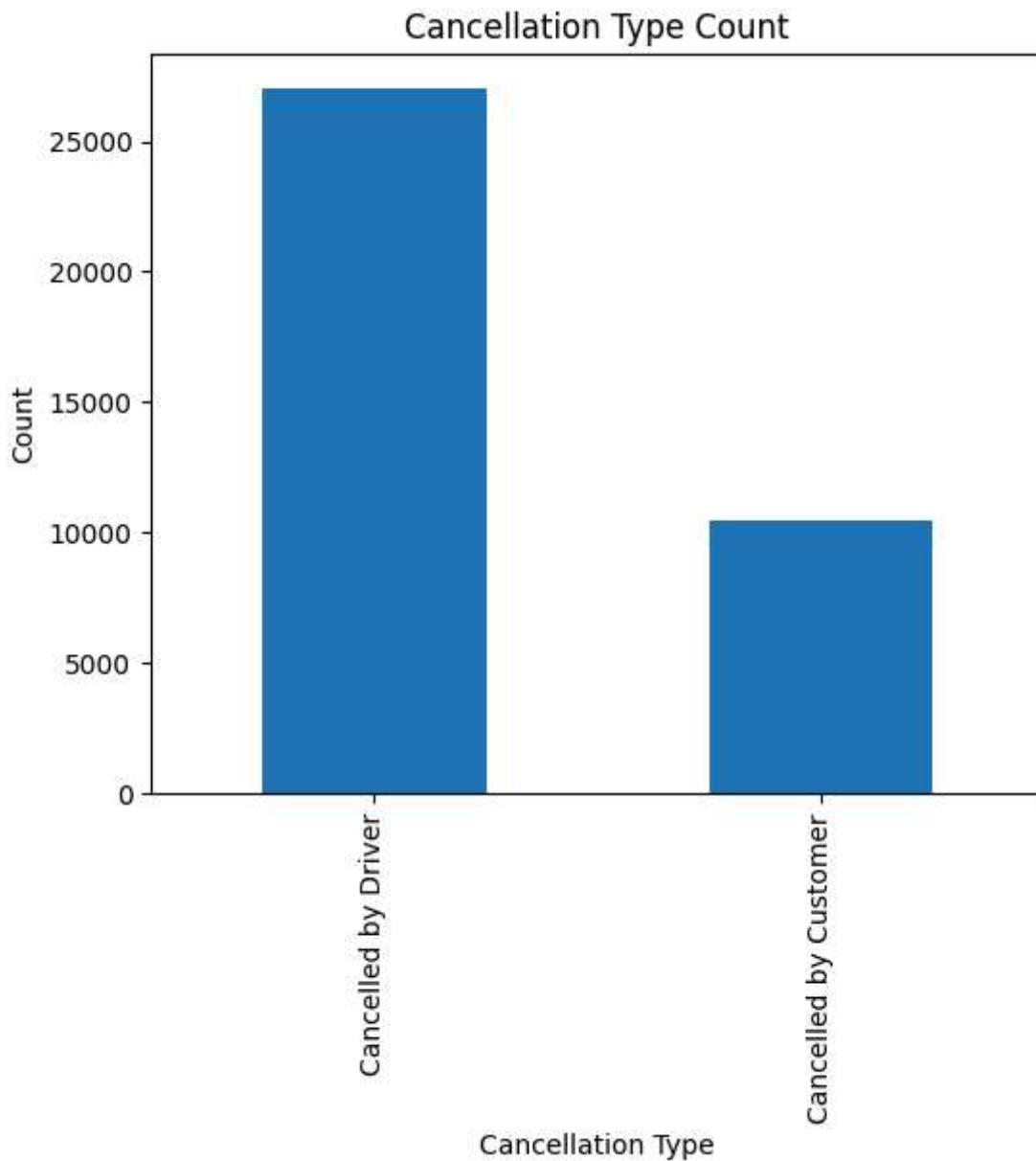
Booking Status Distribution



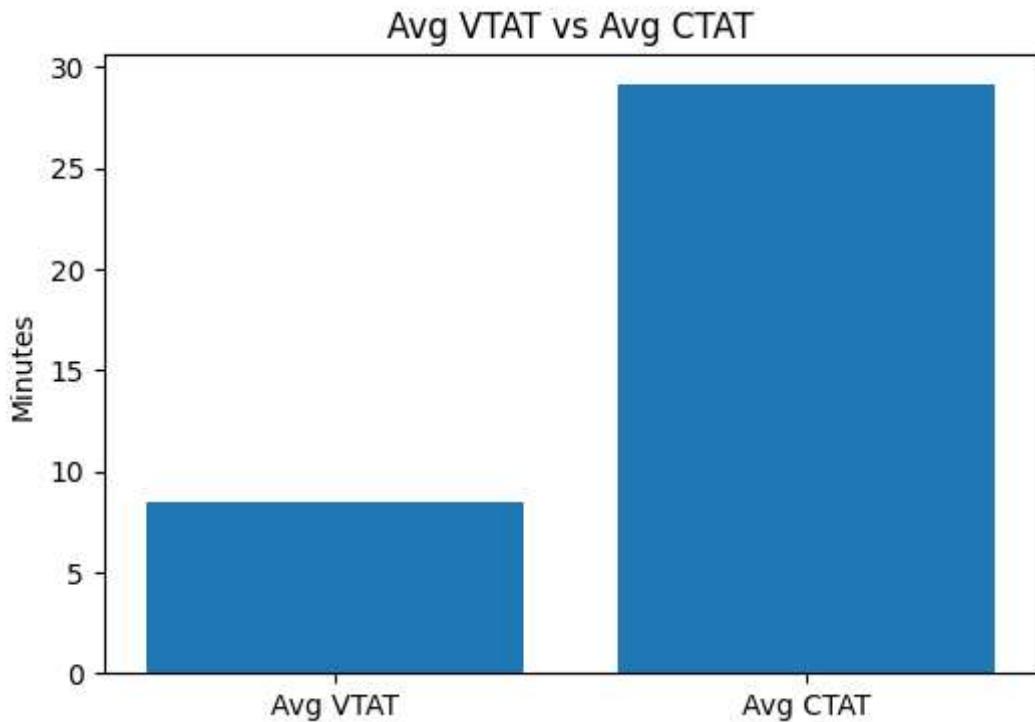
```
In [13]: completed_rate = (df['Booking Status']=='Completed').mean()
print("Completion Rate:", completed_rate)
```

Completion Rate: 0.62

```
In [14]: cancel_data = df[df['Booking Status'].isin(["Cancelled by Driver","Cancelled by Customer"])]  
cancel_counts = cancel_data['Booking Status'].value_counts()  
  
plt.figure(figsize=(6,5))  
cancel_counts.plot(kind='bar')  
plt.title("Cancellation Type Count")  
plt.xlabel("Cancellation Type")  
plt.ylabel("Count")  
plt.show()
```



```
In [15]: vtat = df['Avg VTAT'].mean()  
ctat = df['Avg CTAT'].mean()  
  
plt.figure(figsize=(6,4))  
plt.bar(['Avg VTAT','Avg CTAT'], [vtat, ctat])  
plt.title("Avg VTAT vs Avg CTAT")  
plt.ylabel("Minutes")  
plt.show()
```



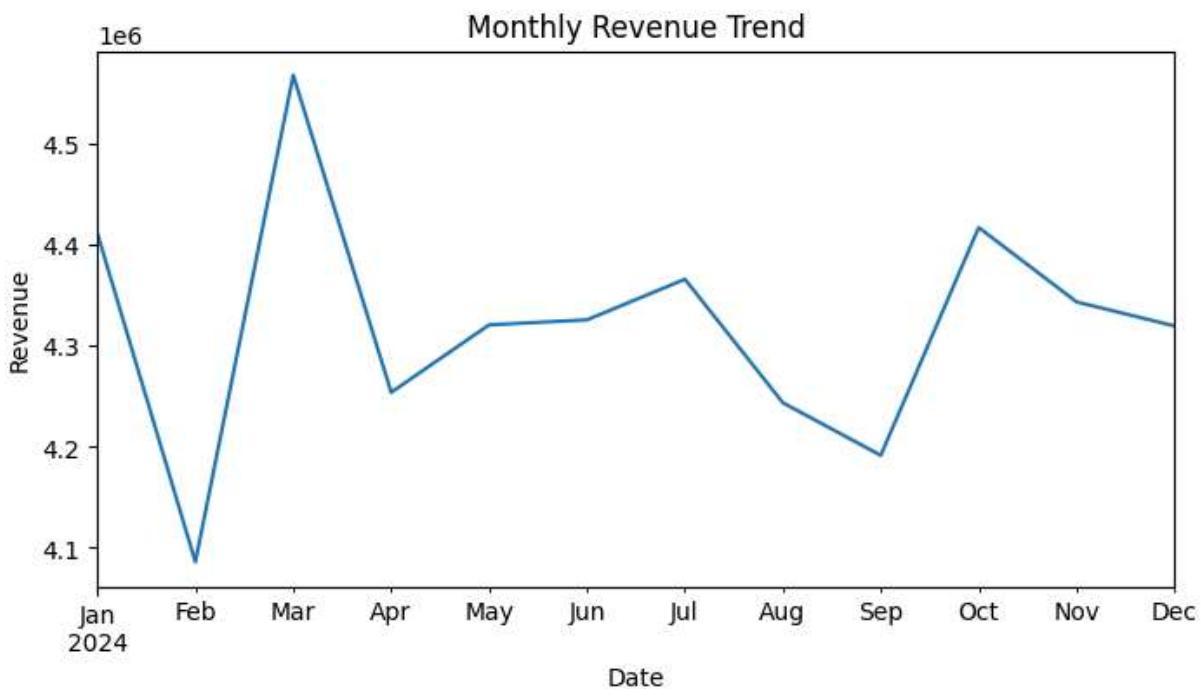
```
In [4]: plt.figure(figsize=(6,4))
plt.boxplot([df['Driver Ratings'].dropna(), df['Customer Rating'].dropna()],
            tick_labels=['Driver Ratings', 'Customer Rating'])
plt.title("Ratings Distribution")
plt.show()
```



```
In [20]: df['Date'] = pd.to_datetime(df['Date'], dayfirst=True)
rev = df.groupby(df['Date'].dt.to_period("M"))['Booking Value'].sum()

rev.plot(kind='line', figsize=(8,4))
```

```
plt.title("Monthly Revenue Trend")
plt.ylabel("Revenue")
plt.show()
```



In []: