In [1]:
```
import warnings
warnings.filterwarnings("ignore")

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

In [2]:
```
df=pd.read_csv("baggagecomplaints.csv")
df
```

Out[2]:

	Airline	Date	Month	Year	Baggage	Scheduled	Cancelled	Enplaned
0	American Eagle	01/2004	1	2004	12502	38276	2481	992360
1	American Eagle	02/2004	2	2004	8977	35762	886	1060618
2	American Eagle	03/2004	3	2004	10289	39445	1346	1227469
3	American Eagle	04/2004	4	2004	8095	38982	755	1234451
4	American Eagle	05/2004	5	2004	10618	40422	2206	1267581
...	...	...	...	...	...	...	...	...
247	United	08/2010	8	2010	14099	30637	344	4263211
248	United	09/2010	9	2010	9435	28072	161	3679517
249	United	10/2010	10	2010	9565	29144	140	3952549
250	United	11/2010	11	2010	8597	27318	104	3573268
251	United	12/2010	12	2010	14415	27619	599	3493643

252 rows × 8 columns

```
- Our variable features


- Airline: Name of the Airline is given.
- Date: date has been mentioned here.
- Month: Month has been mentioned here.
- Year: Year has been mentioned here.
- Baggage: count of baggage is given here.
- Scheduled: Count of baggage scheduled is given here.
- Cancelled: Count of baggage which cancelled is given here.
- Enplaned: Count of baggage which is reached in plane.
```

EDA

In [3]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 252 entries, 0 to 251
Data columns (total 8 columns):
 #   Column     Non-Null Count  Dtype 
---  ------     --------------  ----- 
 0   Airline    252 non-null    object
 1   Date       252 non-null    object
 2   Month      252 non-null    int64 
 3   Year       252 non-null    int64 
 4   Baggage    252 non-null    int64 
 5   Scheduled  252 non-null    int64 
 6   Cancelled  252 non-null    int64 
 7   Enplaned   252 non-null    int64 
dtypes: int64(6), object(2)
memory usage: 15.9+ KB
```

here we have used df.info() to get detail about the columns how many columns present in the csv entries present in the columns and there data types.

In [4]:
```
df.head()
```

Out[4]:

	Airline	Date	Month	Year	Baggage	Scheduled	Cancelled	Enplaned
0	American Eagle	01/2004	1	2004	12502	38276	2481	992360
1	American Eagle	02/2004	2	2004	8977	35762	886	1060618
2	American Eagle	03/2004	3	2004	10289	39445	1346	1227469
3	American Eagle	04/2004	4	2004	8095	38982	755	1234451
4	American Eagle	05/2004	5	2004	10618	40422	2206	1267581

Here we have used df.head() we see first five columns of data in the present csv.

In [5]:
```
df.isnull().sum()
```

Out[5]:
```
Airline      0
Date         0
Month        0
Year         0
Baggage      0
Scheduled    0
Cancelled    0
Enplaned     0
dtype: int64
```

Here we have used df.isnull().sum() to identify that in how many there are null values but here there is no null values that's why it doesn't contain Data Cleaning part.

DATA VISUALIZATION

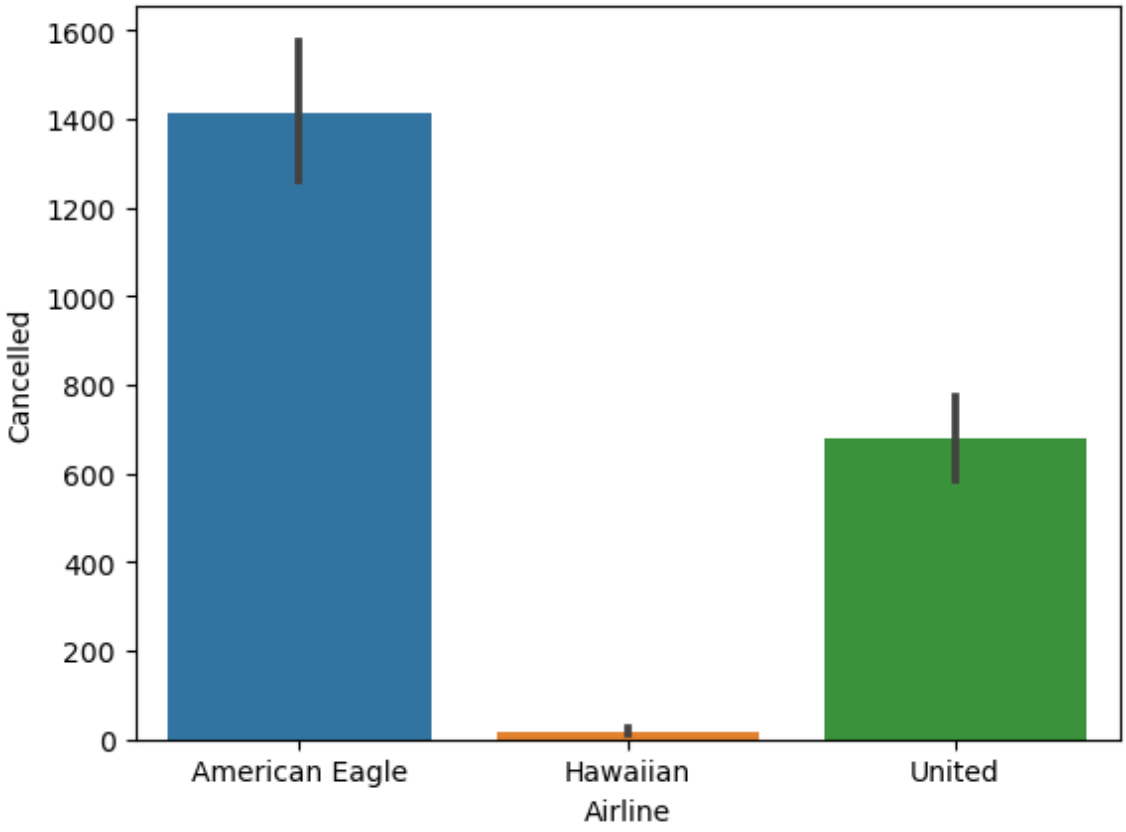
In [6]:
```
df["Airline"].value_counts()
```

Out[6]:
```
American Eagle    84
Hawaiian          84
United            84
Name: Airline, dtype: int64
```

Here we have used value_counts() on Airline columns to see that how many airline is there here.

In [7]:
```
sns.barplot(data=df, x="Airline", y="Cancelled")
```

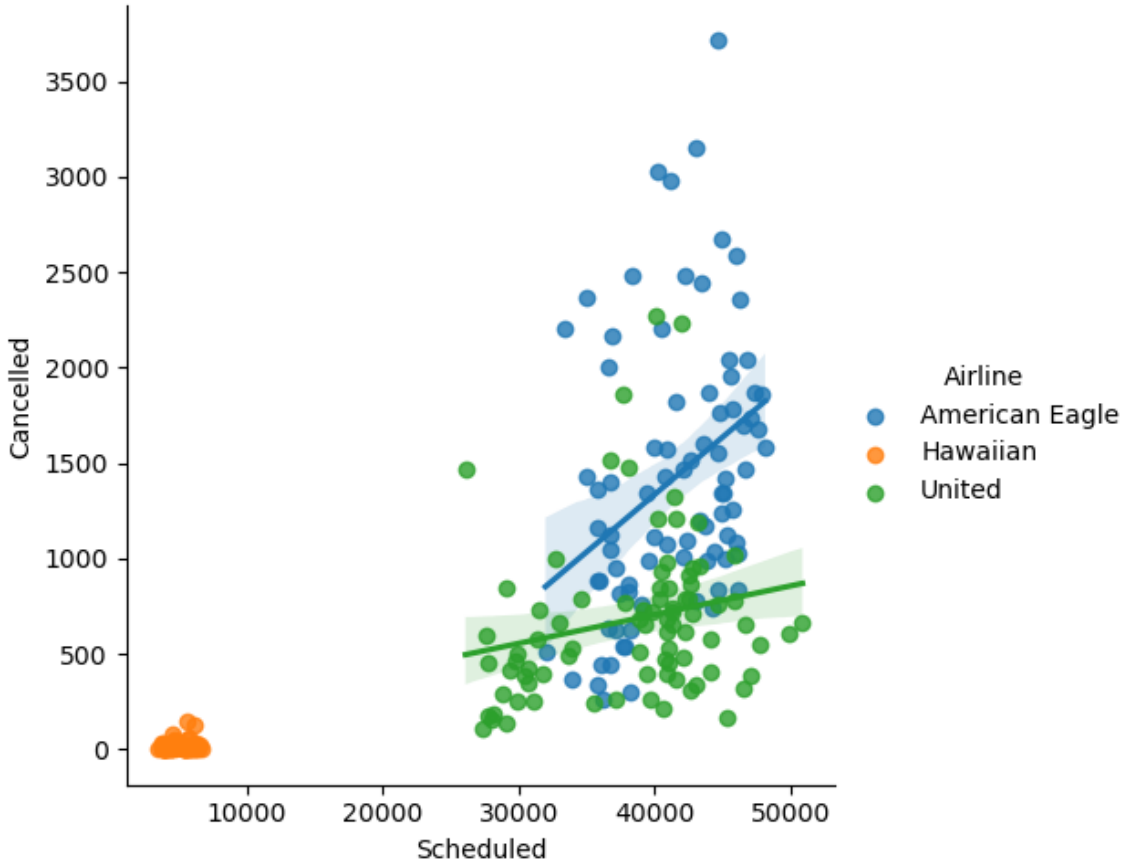
Out[7]:
```
<Axes: xlabel='Airline', ylabel='Cancelled'>
```



Here we have made barplot to identify that there are how many airline where baggage has been cancelled.

In [9]:
```
sns.lmplot(data=df, x="Scheduled", y="Cancelled", hue="Airline")
```

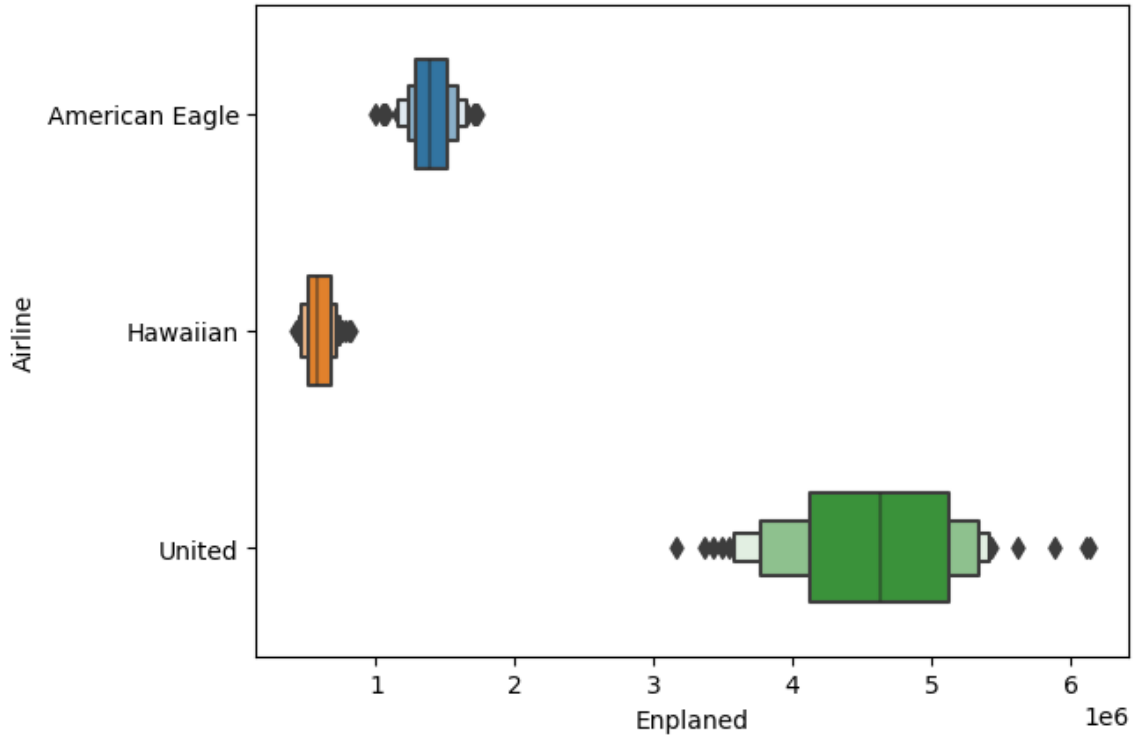
Out[9]:
```
<seaborn.axisgrid.FacetGrid at 0x237449c9d50>
```



Here we have made lmplot to identify that there are how many airline which have scheduled or cancelled there baggages.

In [10]:
```
sns.boxenplot(data=df, x="Enplaned", y="Airline", width=.5)
```

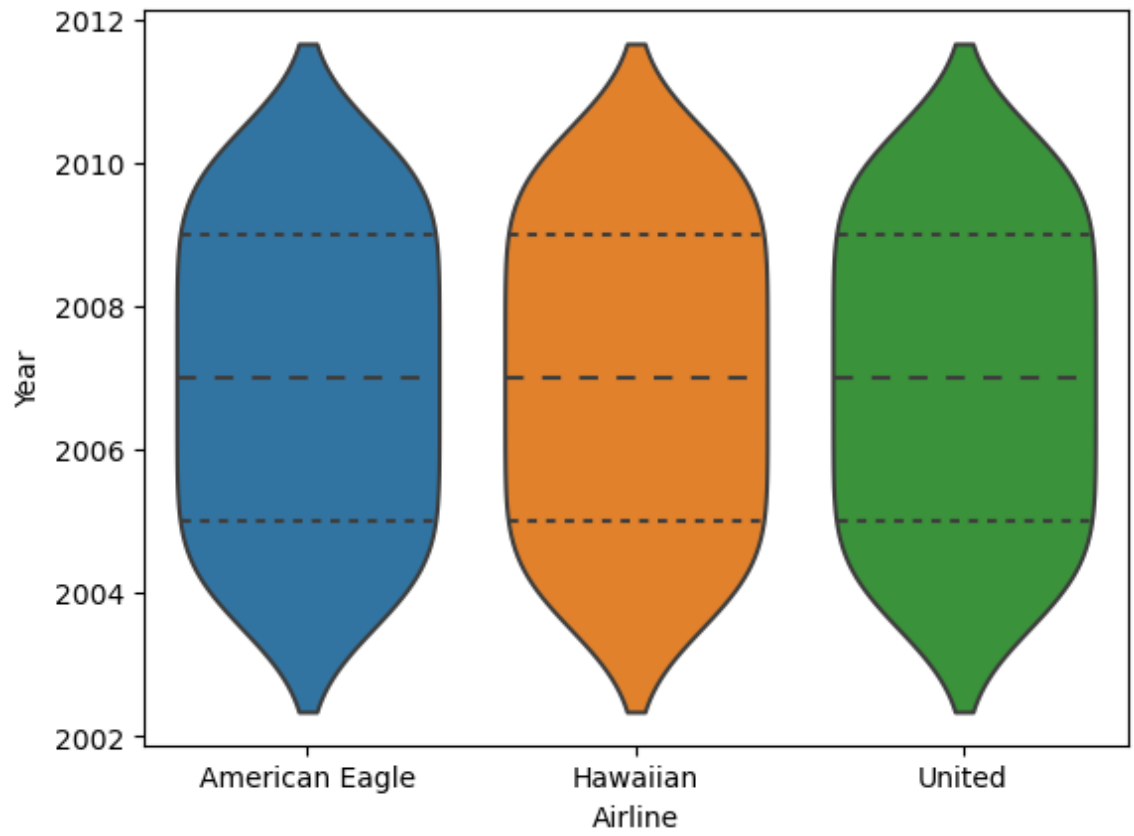
Out[10]:
```
<Axes: xlabel='Enplaned', ylabel='Airline'>
```



Here we have made boxplot to identify that how many baggages has been enplaned in Airline.

In [15]:
```
sns.violinplot(data=df, x="Airline", y="Year", split=True, inner="quart")
```

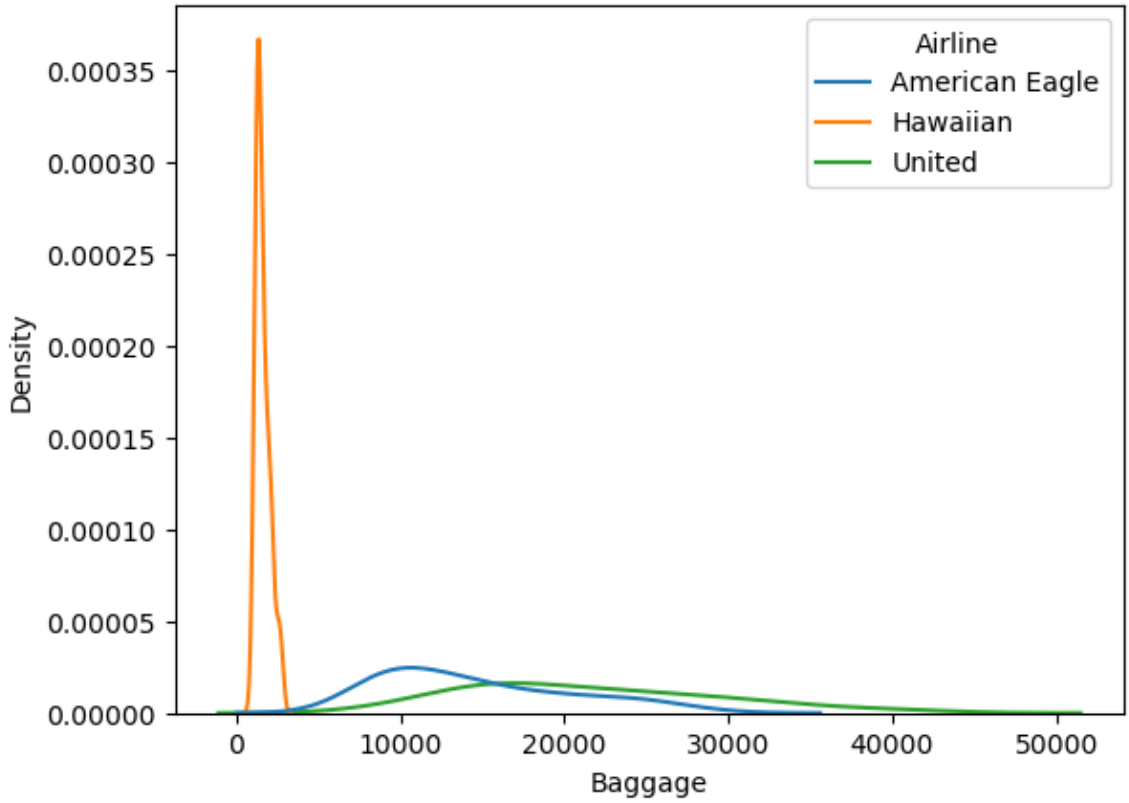
Out[15]:
```
<Axes: xlabel='Airline', ylabel='Year'>
```



Here we have made violinplot to identify that there how many baggages contain in Airline in a Year

In [12]:
```
sns.kdeplot(data=df, x="Baggage", hue="Airline")
```

Out[12]: <Axes: xlabel='Baggage', ylabel='Density'>



In [ ]:
```
Here we have made kdeplot to identify that which airline contains highest nu
```