Mark Middleton

Date: 11/15/23

In [5]:
```
# Imports the 'numpy', 'pandas', and 'matplotlib' libraries. This allows for the use more complex methods than
# in Python.

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

In [6]:
```
# This reads and imports the csv file, 'titanic.csv'.

titanic = pd.read_csv('titanic.csv')
```

In [7]:
```
# This displays the a portion of the information imported from the csv file, 'titanic.csv'.

titanic
```

Out[7]:

	Survived	Pclass	Name	Sex	Age	Siblings/Spouses Aboard	Parents/Children Aboard	Fare
0	0	3	Mr. Owen Harris Braund	male	22.0	1	0	7.2500
1	1	1	Mrs. John Bradley (Florence Briggs Thayer) Cum...	female	38.0	1	0	71.2833
2	1	3	Miss. Laina Heikkinen	female	26.0	0	0	7.9250
3	1	1	Mrs. Jacques Heath (Lily May Peel) Futrelle	female	35.0	1	0	53.1000
4	0	3	Mr. William Henry Allen	male	35.0	0	0	8.0500
...	...	...	...	...	...	...	...	...
882	0	2	Rev. Juozas Montvila	male	27.0	0	0	13.0000
883	1	1	Miss. Margaret Edith Graham	female	19.0	0	0	30.0000
884	0	3	Miss. Catherine Helen Johnston	female	7.0	1	2	23.4500
885	1	1	Mr. Karl Howell Behr	male	26.0	0	0	30.0000
886	0	3	Mr. Patrick Dooley	male	32.0	0	0	7.7500

887 rows × 8 columns

In [8]:
```
# This is a breakdown of the information from the csv file, 'titanic.csv'. It shows, per column, the number of
# average, the standard devivation, and the minimum to maximum value beginning with the minimum and increasing
# maximum value is achieved.

titanic.describe()
```

Out[8]:

	Survived	Pclass	Age	Siblings/Spouses Aboard	Parents/Children Aboard	Fare
count	887.000000	887.000000	887.000000	887.000000	887.000000	887.00000
mean	0.385569	2.305524	29.471443	0.525366	0.383315	32.30542
std	0.487004	0.836662	14.121908	1.104669	0.807466	49.78204
min	0.000000	1.000000	0.420000	0.000000	0.000000	0.00000
25%	0.000000	2.000000	20.250000	0.000000	0.000000	7.92500
50%	0.000000	3.000000	28.000000	0.000000	0.000000	14.45420
75%	1.000000	3.000000	38.000000	1.000000	0.000000	31.13750
max	1.000000	3.000000	80.000000	8.000000	6.000000	512.32920

In [43]:
```
# This breaks the data from the csv file into three different groups, organized by the passenger classes.

byClass = titanic.groupby('Pclass')
df1 = byClass.get_group(1)
df2 = byClass.get_group(2)
df3 = byClass.get_group(3)
```

In [60]:
```
# This is a breakdown of the data for first class passengers.

df1.describe()
```

Out[60]:
```
Pclass
1    512.3292
Name: Fare, dtype: float64
```

In [14]:
```
# This is a breakdown of the data for second class passengers.

df2.describe()
```

Out[14]:

	Survived	Pclass	Age	Siblings/Spouses Aboard	Parents/Children Aboard	Fare
count	184.000000	184.0	184.000000	184.000000	184.000000	184.000000
mean	0.472826	2.0	29.868641	0.402174	0.380435	20.662183
std	0.500623	0.0	13.756191	0.601633	0.690963	13.417399
min	0.000000	2.0	0.670000	0.000000	0.000000	0.000000
25%	0.000000	2.0	23.000000	0.000000	0.000000	13.000000
50%	0.000000	2.0	29.000000	0.000000	0.000000	14.250000
75%	1.000000	2.0	36.625000	1.000000	1.000000	26.000000
max	1.000000	2.0	70.000000	3.000000	3.000000	73.500000

In [15]:
```
# This is a breakdown of the data for third class passengers.

df3.describe()
```

Out[15]:

	Survived	Pclass	Age	Siblings/Spouses Aboard	Parents/Children Aboard	Fare
count	487.000000	487.0	487.000000	487.000000	487.000000	487.000000
mean	0.244353	3.0	25.188747	0.620123	0.396304	13.707707
std	0.430145	0.0	12.095084	1.379392	0.891793	11.817309
min	0.000000	3.0	0.420000	0.000000	0.000000	0.000000
25%	0.000000	3.0	19.000000	0.000000	0.000000	7.750000
50%	0.000000	3.0	24.000000	0.000000	0.000000	8.050000
75%	0.000000	3.0	31.000000	1.000000	0.000000	15.500000
max	1.000000	3.0	74.000000	8.000000	6.000000	69.550000

In [55]:
```
# This finds third class passengers under 10 years old, then finds the mean average survival rate as a decimal

thirdCChildren = titanic[(titanic['Pclass'] == 3) & (titanic['Age'] <= 10)]
survProb = thirdCChildren['Survived'].mean()
print(round(survProb, 2))
```

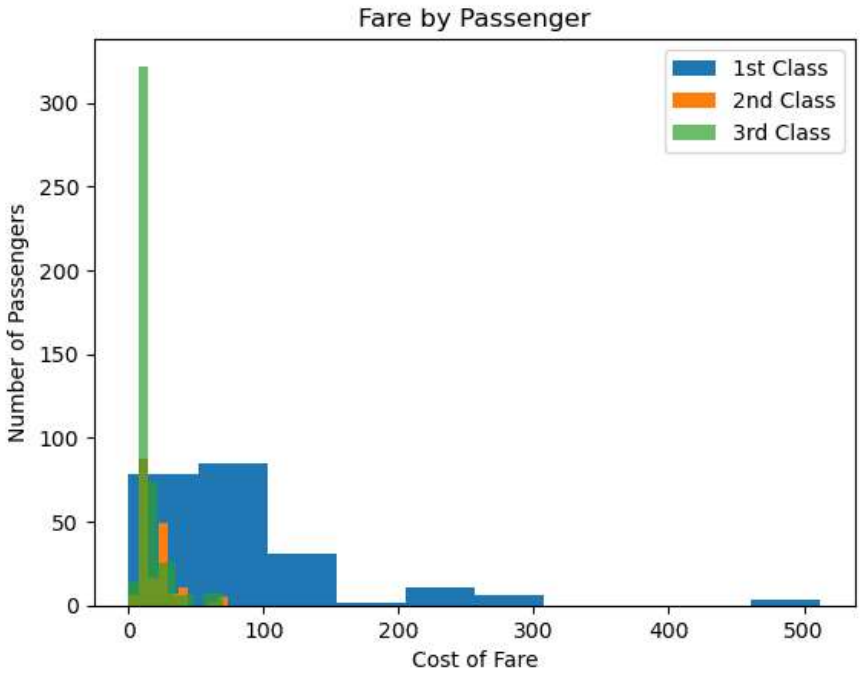
```
0.42
```

In [37]:
```
# This is a graph that sets the number of passengers opposed to the cost of their fare, with passenger class d
# different groups noted by color.

plt.hist(df1['Fare'])
plt.hist(df2['Fare'])
plt.hist(df3['Fare'], alpha=0.7)

plt.title('Fare by Passenger')
plt.xlabel('Cost of Fare')
plt.ylabel('Number of Passengers')
plt.legend(['1st Class', '2nd Class', '3rd Class'])
```

Out[37]: <matplotlib.legend.Legend at 0x24b81c49b50>



Data Analysis

The csv file, 'titanic.csv' provides the data for 887 passengers. Each row gives data for one indivdual passenger. Using this data, we can organize and focus the data to gain persepective on the event. Divided by age and class, we can see that third class children ages ten and younger have a 42% probability of survival. In contrast, the survival rate of first class passengers, regardless of age, had a substantially higher survival rate of 63%. However, as low as the survival rate of children is, the overall survival rate of third class passengers was 24%. The overall mean average ticket price was £32.31, while for first class is was £84.16, for second class £20.67, and third class £13.71. The highest prices paid by class, beginning with first class, were £512.33, £73.50, and £69.55, respectively. For the lowest ticket price for all three classes was £0.

In [ ]: