Since Age is a continuous variable, the line plot from code cell 23 is not very informative - the survival rate for each individual point is based on a very small number of passengers. A better option would be to compare age histograms for surviving and deceased passengers in the same plot:

age\_bins = np.arange(0, 100, 4)

sns.distplot(titanic\_df.loc[(titanic\_df['Survived']==0) & (~titanic\_df['Age'].isnull()),'Age'], bins=age\_bins)

sns.distplot(titanic\_df.loc[(titanic\_df['Survived']==1) & (~titanic\_df['Age'].isnull()),'Age'], bins=age\_bins)

plt.title('age distribution among survival classes')

plt.ylabel('frequency')

plt.legend(['did not survive', 'survived']);

One thing you could do to improve your project even further is take advantage of the many plots from the Seaborn package, such as [heatmaps](https://stanford.edu/~mwaskom/software/seaborn/generated/seaborn.heatmap.html), [swarm plots](http://seaborn.pydata.org/generated/seaborn.swarmplot.html), [violin plots](http://seaborn.pydata.org/generated/seaborn.violinplot.html), and [kernel density estimates](https://stanford.edu/~mwaskom/software/seaborn/generated/seaborn.kdeplot.html).

For instance, here's a group of swarm plots showing survival status according to age, gender, and ticket class:

import seaborn as sns

sns.factorplot(data=titanic\_df, col='Pclass', x='Sex', y='Age', hue='Survived',

kind='swarm', size=10, aspect=0.5, s=10)

plt.yticks(np.arange(0,81,10));

**Suggestion**

Instead of just discussing the results of the calculations and visualizations in the conclusion, a better idea is to discuss each result right after they are presented. Then the conclusion would be used to recapitulate what has already been explored and discussed.

A good template to follow is this:

* Write a short paragraph explaining what you will do
* Present the code and its results
* Write a short paragraph discussing the code's output, and set up the question/idea that will lead to the next piece of code