**Seaborn Exercises**

**Instructions:**

• Complete the tasks using Python (pandas, numpy, seaborn).

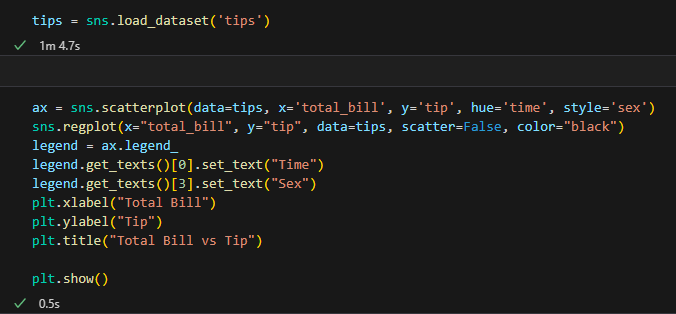
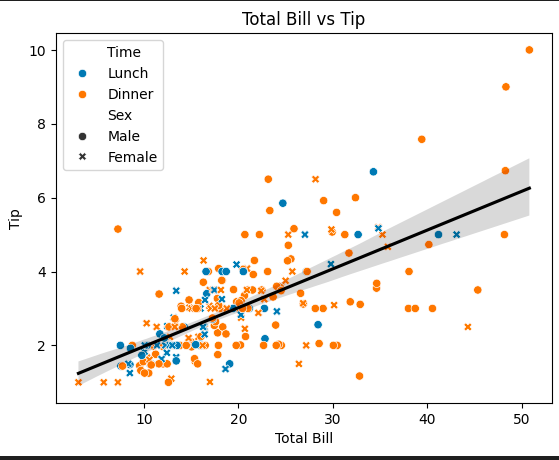
• Include code cells, resulting plots, and short interpretations where requested.

• Submit a single notebook with all answers and save final figures as PNG files.

• This worksheet contains 10 exercises arranged from easy → hard.

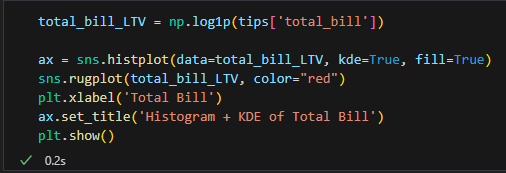
1. **1) Load & visualize: `tips` scatter**

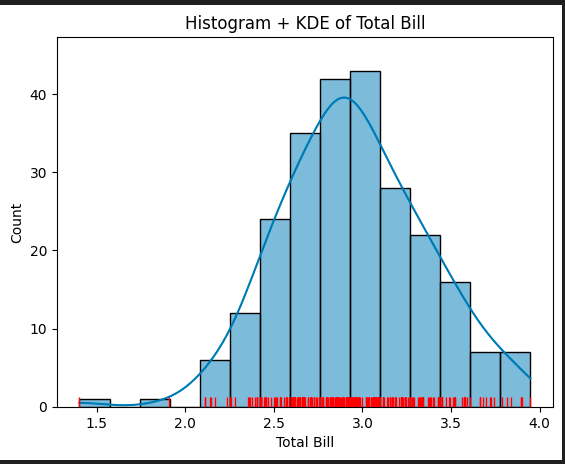
Dataset: sns.load\_dataset('tips')  
  
Task: Create a scatter plot of total\_bill vs tip. Color points by time and use different markers for sex. Add axis labels and a title.  
  
*Bonus: Add a regression line for each time (lunch/dinner) on the same axes.*



1. **2) Distribution + NumPy: histogram & KDE**

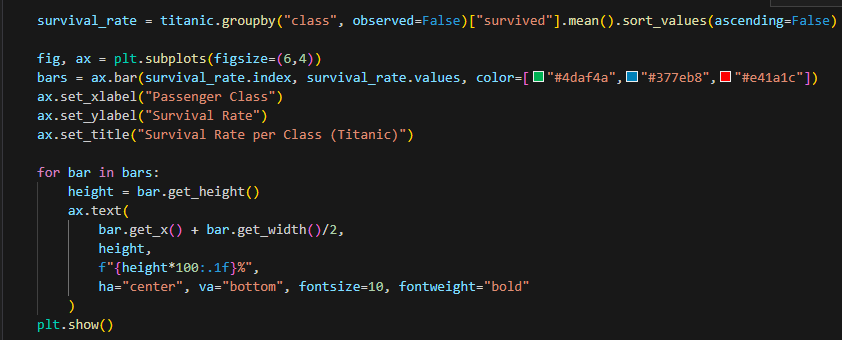
Dataset: tips (use total\_bill)  
  
Task: Using numpy, compute a log-transformed version of total\_bill (np.log1p). Plot its histogram and KDE on the same axes. Briefly state whether the transform made the distribution more symmetric.  
  
*Bonus: Overlay a rug plot.*

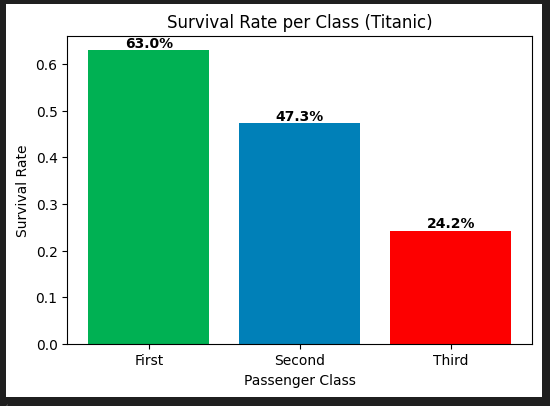




1. **3) Categorical aggregation: barplot with pandas groupby**

Dataset: titanic  
  
Task: Using pandas, group by class and compute the survival rate; show a bar plot of survival rate per class. Annotate bars with percentages.  
  
*Bonus: Split bars by sex using hue.*





1. **4) Pairwise relationships: pairplot vs PairGrid**

Dataset: iris  
  
Task:  
 a) Produce sns.pairplot(iris, hue='species').  
 b) Create the same using PairGrid with scatterplots in the upper triangle and KDEs on the diagonal. Explain one advantage of PairGrid.

1. **5) Time-series & heatmap: flights pivot**

Dataset: flights (year, month, passengers)  
  
Task: Pivot into a month x year matrix and plot a heatmap with annotations and colorbar. Interpret the main pattern you observe.  
  
*Bonus: Ensure months appear in chronological order.*

1. **6) Regression with groups & matplotlib tweak**

Dataset: mpg (model\_year, mpg, origin)  
  
Task: For each origin, plot a linear regression (mpg vs model\_year) on the same axes with different line styles. Add a legend outside the plot and a vertical dashed line at the median model year.  
  
*Bonus: Compute and display the slope for each origin as text on the plot.*

1. **7) Faceting & custom aggregation with FacetGrid**

Dataset: tips  
  
Task: Create a FacetGrid faceted by day (columns) and smoker (rows). In each facet show a violin plot of total\_bill and overlay the mean point (white dot with black edge). Annotate each facet with the mean value.

1. **8) Joint distributions + conditional coloring**

Dataset: penguins (or substitute)  
  
Task: Create a jointplot of bill\_length\_mm vs bill\_depth\_mm with hex bins. Color hexes by the average species-encoded-as-integer inside each bin (compute binned-statistic). Explain your binning approach.  
  
*Bonus: Provide a legend mapping species encoding back to names.*

1. **9) Build a custom seaborn helper function**

Task: Implement pretty\_violin(df, x, y, hue=None, title=None) that:  
 - sets a publication-style seaborn theme  
 - draws a violin plot with inner quartiles  
 - overlays swarm points (max 200 per group, sample if needed)  
 - adds a text box with mean, median, std for each group  
Demonstrate on a synthetic dataset (3 groups).  
  
*Bonus: Return fig, ax.*

1. **10) Exploratory mini-project**

Dataset: choose one (titanic, penguins, mpg, or tips)  
  
Deliverables: For the chosen dataset produce cells that:  
 1. Clean missing data and describe strategy.  
 2. Show a correlation heatmap and explain three strongest associations.  
 3. Compare a numerical outcome across two categorical variables.  
 4. Create a multi-panel facet visualization that reveals an interaction effect.  
 5. Create one advanced visualization (annotated regression + residuals OR hexbin + marginals OR clustermap).  
 6. Summarize your 3 most important findings in bullet points.  
  
*Bonus: Add an interactive widget to choose hue/col (optional).*