Questions:

1. What is the distribution of lung tissue classifications (benign, adenocarcinoma, squamous cell carcinoma) in the dataset?
   1. Generate simple count plots/bar charts separating by cancer types. A histogram or a pie chart could also be used to see the share of each class.
2. How does the count of benign colon tissues compare to colon adenocarcinomas in the dataset?
   1. A bar chart or pie chart to compare these two classes.
3. Are there any identifiable patterns or clusters when comparing features across different tissue classifications (eg: shape, size)?
   1. Image analysis to extract features from images like intensity distribution, texture features etc. after creating, perform EDA by creating scatterplots or maybe using some dimensionality reduction like PCA for visualization.
4. Identify any significant differences between variations of cancerous tissues (adenocarcinoma vs squamous cell carcinoma)?
   1. Similar as previous, implement image analysis techniques before EDA examine only those instances related with lung cancer and focus on highlighting differences between them using visualization from before, as well as adding box and violin plot comparisons too.
5. Is there a significant difference in the visual characteristics between benign tissue and adenocarcinoma for both lung and colon samples?
   1. t-SNE or PCA to reduce dimensionality of images, then visualize all data points on scatter plot. If groups are visually separable, this indicates physical differences between groups/categories.
6. How do the characteristics of lung squamous cell carcinoma differ from that of lung adenocarcinoma in these images?
   1. Similar to before, dimensionality reduction followed by visualization would help identifying patterns which are distinguishing these two types.
7. Identify any specific patterns or anomalies within the different classes through visualization?
   1. Correlation heatmap can be used with feature extraction methods to see if any features extracted from images strongly correlate with an anomaly or pattern across categories/classes. Techniques (like clustering) may reveal hidden relationships/patterns/anomalies.