Question 1: Why is it important to handle missing values in a dataset before proceeding with data visualization? How might different methods of handling missing data (such as filling with zeros, mean imputation, or dropping rows/columns) affect the outcome of your analysis?

Anwser: As far as I am concernec, it is necessary to clean the dataset. The reason is that the unexpected value may influence the data visualization. For instance, the completeness of the data may influence the result of the data visualoization for the reason that the data with undefinred value or missing value might be the invalid value. The invalid value may cause the wrong statistic information. The way of filling the ,missing value is also crucial for the results of data visualization. The advantages and disadvantages are as follows:

1. **Filling with zeros:** The advantage of this method is that it is easy to fill the missing value with any computation. However, in many disciplines, it is cannot be used to fill the missing values. For example, as for the time series data, we cannot fill the value with zero since it will causes the distortion of the time series data.
2. **Filling with mean imputation:** It is a very common technique for researchers to fill the missing value with mean imputation. This method can be useful if the data is approximately normally distributed but can reduce the variance in the dataset and potentially bias the estimates if the data is not normally distributed.
3. **Dropping rows/columns**: This method is also common for us to use. However, it is not useful for us when there are to many samples with missing value in the datasets or there are not enough samples in the datasets. If we drop too much samples in the dataset, it will also cause the bias of distortion of the data visualization.

Question 2: How do visitor numbers change with the seasons, and what could be potential reasons for these fluctuations? Consider how setting specific intervals on the y-axis might aid in identifying these trends

Answer: From the figure, we can find that the visitor arrivals are cyclical over time during the whole year. The peak season of the Avila Adobe is around the spring and summer. The number of the visitors in peak season is lower from 2015 before a rising from 2014. There are two continuous month without the visitors in 2020, 2021, and 2023. The reason of that may because of the spread of the COVID-19. The specific interval on the y-axis is able to help researchers to find the patterns easier. Besides, it is also helpful for people to estimate the value of the line or the specific dot. For example, for the data point in February 2014, for the current y-axis interval, the estimated value would be 1,500 ~ 2,000. However, if we smaller the interval to 2,500 for this figure, we can find that the estimated value would be from 1,750 ~ 2,000. The smaller the interval, the estimate value is more accurate.

Question 3: What similarities and differences do you observe in the visitor patterns over time? Discuss possible reasons for these trends and how they might inform future decisions for museum management and marketing strategies

Answer: The similarity between two museums is that both of them have the same period during 2020, 2021, and 2022. I suppose that it may because of the break out of the COVIND-19. The difference is that the number of visitors in Chinese American Museum is lower than that of the Avila Adobe during these years. The reason may be that the reputation of the Chinese American Museum is not as high as Avila Adobe or there is the geography location problem for the Chinese American Museum. As for the Avila Adobe, I suppose that they need to propose more discount for the tourist so that more people are willing to visit there. As for the Chinese American Museum, they need more advocacy to let more people know about them.