**Preparation for DSA Exam**

PS: Files are available on the link provided below.

1. **Analysis of Scoring in the NFL.** The file *ScoringNFL* contains the points scored by each National Football League team for 10 seasons.
   1. NFL leadership is interested in seeing how the distribution of scoring across the league has changed over these 10 seasons. Construct a single chart that allows the audience to make these comparisons. Provide a thorough discussion of the distribution of scoring in the league over this decade.
   2. NFL leadership is also interested in the temporal patterns in scoring for individual teams. Unstack the data and use sparklines to show how scoring for each team has changed over these 10 seasons. List a team with a strong upward trend over these 10 seasons. List a team with a strong downward trend over these 10 seasons.
2. **Umbrella Sales.** The file *Umbrella* contains data on the quarterly sales revenue for a manufacturer of umbrellas and other weather-resistant gear.
   1. Create a line chart for the sales time series data. Add a four-period moving average to the chart to smooth the data.
   2. To investigate the possibility of seasonality on a quarterly basis, plot the data as a collection of five data series (one for each year). From this visualization, do you observe any indication of seasonality? If so, describe it.
3. **Market Capitalization and Profit By Sector.** The file *Fortune500Sector* contains data on the profits, market capitalizations, and industry sector for a recent sample of firms in the Fortune 500.
   1. Differentiating observations by using a different color for each industry sector, prepare a scatter chart to show the relationship between the variables Market Capitalization and Profit in which Market Capitalization is on the vertical axis and Profit is on the horizontal axis.
   2. Emphasize the relationship between Market Capitalization and Profit within the healthcare sector by formatting all other sectors with data points in gray with no fill. Create a trendline based only on the observations in the healthcare sector. What does the trendline indicate about this relationship between Market Capitalization and Profit within the healthcare sector?
4. **Refer to the EJB data file.** 
   1. Construct the relative frequency distribution of records over the values of the Category variable. Describe your findings.
   2. Construct the relative frequency distribution of records over the values of the New Customer variable. In the PivotTable, relabel a “No” value for New Customer as “Existing” and a “Yes” value as “New.” Describe your findings.
5. **Refer to the EJB data file.** To understand the implications of missing data, we must explore the patterns associated with the missing entries for the Product Satisfaction Rating variable in the EJB data used within the chapter.
   1. Construct the relative frequency distribution of records over the values of the Product Satisfaction Rating variable. What percentage of records are missing a value of Product Satisfaction Rating?
   2. Considering only records that report values of Product Satisfaction Rating, construct the relative frequency distribution of records over the different flavors.
   3. Considering records that are missing values of Product Satisfaction Rating, construct the relative frequency distribution of records over the different flavors.
   4. Compare the distributions in part (b) and part (c). What does this comparison suggest?
6. Refer to the attrition dataset.
   1. **Reflecting on our course and working on your Colab Notebook apply what you learned to this dataset (30 points)**
      1. **Data Cleaning - How would you validate that the data is clean. Include the code to validate this.**
      2. **Data Analysis:**
         1. **How would you explore the data for further analysis? Include code for the EDA**
         2. **Which visualizations would be most useful for this dataset?**
      3. **Machine Learning - How could machine learning be useful in exploring this dataset?**
7. Consider our discussion on data fluency. How could it be valuable to you as a new manager of a Data Science function in a Financial Services company?

Files for Exam Qn 1 Prep Questions:

<https://drive.google.com/drive/folders/1Hk2ZLF-InrlmEfbeNAOwwbkD8gIFOhTV?usp=sharing>