**Visualizing Hotel Data**

This project will guide you through an analysis of [hotel cancellation rate data](https://www.kaggle.com/datasets/jessemostipak/hotel-booking-demand) using Excel. This analysis will focus on lead time: the amount of time between making a reservation and the arrival date.

You’ll start by exploring the data using conditional formatting and sparklines. Based on patterns from this exploration, you’ll explore the effect of lead time on cancellations using histograms and line charts. Lastly, you’ll break this analysis down by distribution channel using column charts.

**First Group of Checkboxes (Initial Exploration):**

* Start by using conditional formatting and sparklines to get a general sense of the data. On the Explore sheet, you’ll find a table breaking down the number of reservations by planned arrival month, reservation status, and lead time. Apply a color scale to the table of data in D9:O17.
* Create a sparkline in P9 that draws a trend line for the Cancelled and Long row.
* Drag the sparkline down through P17.
* Record any observations you make in the Notes section.

**Second Group of Checkboxes (Lead Time):**

* Let’s look at Lead Time on its own. In particular, let’s take a look at the raw numbers instead of the short, medium, and long categories. These numbers are all listed on the Lead Time sheet. Create a histogram of the lead time data.
* Take a note of any patterns you notice in the Notes section.

**Third Group of Checkboxes (Lead Time by Status):**

* Let’s break this analysis of lead time down by reservation status (not that we’ve grouped No Show in with Cancellation). On the Lead Time by Status sheet, create a histogram of the Cancelled column.
* On the same sheet, create a histogram of the Kept column.
* Make a note of any observations you have in the Notes section.

**Fourth Group of Checkboxes (Lead Time to Cancellations):**

* Since it seems as if lead time might be connected somehow with cancellations, let’s look at a scatter-plot to see if there’s any correlation. The Lead Time to Cancellations sheet has a table of average lead time and number of cancellations for each month in the dataset. Create a scatterplot comparing these two columns.
* Record any observations you have in the Notes section.

**Fifth Group of Checkboxes (Lead Time and Cancellations over Time):**

* A correlation is just a correlation. Let’s see if lead time and cancellations are both just being influenced by seasonal fluctuations. The same monthly data is on the Over Time sheet, although the numbers have been normalized to fit in the same range. Create a line graph plotting both average lead time and the number of cancellations.
* Record any observations you make in the Notes section.

**Sixth Group of Checkboxes (Distribution Channels):**

* Let’s end by breaking our analysis down by distribution channel: direct sales, corporate sales, and travel agent sales. There are three tables on the By Distribution Channel sheet. The first table on the left has the average lead time by distribution channel. Create a bar chart of this data.
* The middle table has the breakdown of kept/cancelled reservations by each distribution channel. Create a stacked bar chart of this data.
* The final table on the right has the percentage of reservations that get cancelled for each distribution type and lead time category. Create a clustered column chart of this data.
* Record any observations you have in the Notes section.